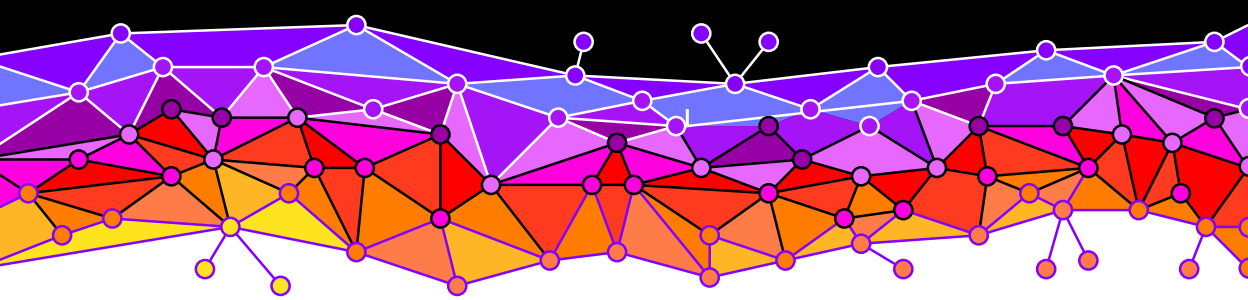
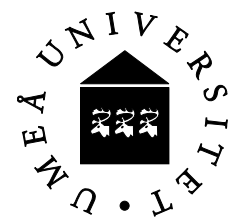
FORM AND FUNCTION OF COMPLEX NETWORKS

Petter Holme





Department of Physics
Umeå University
Umeå 2004

Department of Physics

Umeå University

901 87 Umeå, Sweden

This online version differs from the printed version only in that the figures are in colour, the text is hyperlinked and that the Acknowledgement section is omitted.

ISBN 91-7305-629-4

Printed by Print & Media, Umeå 2004

Abstract

Networks are all around us, all the time. From the biochemistry of our cells to the web of friendships across the planet. From the circuitry of modern electronics to chains of historical events. A network is the result of the forces that shaped it. Thus the principles of network formation can be, to some extent, deciphered from the network itself. All such information comprises the structure of the network. The study of network structure is the core of modern network science. This thesis centres around three aspects of network structure: What kinds of network structures are there and how can they be measured? How can we build models for network formation that give the structure of networks in the real world? How does the network structure affect dynamical systems confined to the networks? These questions are discussed using a variety of statistical, analytical and modelling techniques developed by physicists, mathematicians, biologists, chemists, psychologists, sociologists and anthropologists. My own research touches all three questions. In this thesis I present works trying to answer: What is the best way to protect a network against sinister attacks? How do groups form in friendship networks? Where do traffic jams appear in a communication network? How is cellular metabolism organised? How do Swedes flirt on the Internet? ...and many other questions.

Sammanfattning (summary in Swedish)

Vi är omgivna av nätverk. Från biokemiska processer i våra celler till nätverket av vänskapsband mellan alla planetens invånare. Från elektroniska kretsar i elektriska apparater till kedjor av historiska händelser. Ett nätverk är resultatet av de krafter som skapat det. Alltså kan man studera de bakomliggande mekanismerna genom att studera nätverket i sig själv. Denna information utgör nätverkets struktur. Studiet av nätverksstruktur är kärnan i modern nätverksforskning. Denna avhandling behandlar nätverksstruktur ur tre synvinklar: Vilka typer av nätverksstruktur finns det och hur kan den mätas? Hur kan vi modellera tillkomsten av nätverk med samma struktur som verkliga nätverk? Hur påverkar den underliggande nätverksstrukturen dynamiska system på nätverket? Dessa frågor diskuteras med hjälp av statistiska, analytiska och modelleringsbaserade tekniker utvecklade av fysiker, matematiker, biologer, kemister, psykologer, sociologer och antropologer. Min egen forskning berör alla dessa tre frågor. I denna avhandling presenterar jag arbeten som försöker besvara: Hur kan man bäst skydda ett nätverk mot ondsinta attacker? Hur bildas sociala grupper i vänskapsnätverk? Var kommer trafikstockningar uppstå i kommunikationsnätverk? Hur är metabolismen i celler organiserad? Hur flörtar svenskar med varandra på internet? . . . och många andra frågor.

Contents

Papers

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

I P. Holme, M. Huss, and H. Jeong, *Subnetwork hierarchies of biochemical pathways*, Bioinformatics **19** (2003), pp. 533-538.

II P. Holme, F. Liljeros, C. R. Edling, and B. J. Kim, *Network bipartivity*, Phys. Rev. E. **68** (2003), art. no. 027101.

III P. Holme, C. R. Edling, and F. Liljeros, *Structure and time-evolution of an Internet dating community*, to appear in Soc. Netw.

IV P. Holme, *Network dynamics of ongoing social relationships*, Europhys. Lett. **64** (2003), pp. 427-433.

V P. Holme and B. J. Kim, *Growing scale-free networks with tunable clustering*, Phys. Rev. E **65** (2002), art. no. 026107.

VI A. Grönlund and P. Holme, *Networking the seceder model: Group formation in social and economic systems*, submitted but not yet accepted for publication.

VII P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, *Attack vulnerability of complex networks*, Phys. Rev. E **65** (2002), art. no. 056109.

VIII P. Holme, *Efficient local strategies for vaccination and network attack*, submitted but not yet accepted for publication.

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

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

XI B. J. Kim, H. Hong, P. Holme, G. S. Jeon, P. Minnhagen, and M. Y. Choi, *XY model in small-world networks*, Phys. Rev. E **64** (2001), art. no. 056135.

XII K. Medvedyeva, P. Holme, P. Minnhagen, and B. J. Kim, *Dynamic critical behavior of the XY model in small-world networks*, Phys. Rev. E **67** (2003), art. no. 036118.

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

XIV B. J. Kim, A. Trusina, P. Holme, P. Minnhagen, J. S. Chung, and M. Y. Choi, *Dynamic instabilities induced by asymmetric influence: Prisoners' dilemma game in small-world networks*, Phys. Rev. E **66** (2002), art. no. 021907

XV P. Holme, A. Trusina, B. J. Kim, and P. Minnhagen, *Prisoners' dilemma in real-world acquaintance networks: Spikes and quasi-equilibria induced by the interplay between structure and dynamics*, Phys. Rev. E **68** (2003), art. no. 030901.

Other papers by the author:

- P. Holme and M. Huss, *Discovery and analysis of biochemical subnetwork hierarchies*, 3rd Workshop on Computation of Biochemical Pathways and Genetic Networks, R. Gauges, U. Kummer, J. Pahle, and U. Rost, eds., European Media Lab Proceedings, Logos, Berlin, 2003, pp. 3-9.
- P. Holme, B. J. Kim, and P. Minnhagen, *Phase Transitions in the Two-Dimensional Random Gauge XY Model*, Phys. Rev. B **67** (2003), art. no. 104510.
- P. Holme and P. Olsson, *A zero-temperature study of vortex mobility in two-dimensional vortex glass models*, Europhys. Lett. **60** (2002), pp. 439-445.
- P. Olsson and P. Holme, *Transition in the two-dimensional step model: A Kosterlitz-Thouless transition in disguise*, Phys. Rev. B **63** (2000), art. no. 052407.
- P. Olsson and P. Holme, *Comment on "Structure and Phase Transition of Josephson Vortices in Anisotropic High-Tc Superconductors"*, Phys. Rev. Lett. **85** (2000), p. 2651.

Preface

Interdisciplinary science is cool—no doubt about that—but it took me a while to get that insight. When I came to Umeå for my Master's thesis work I was at heart an orthodox sinologist that reluctantly acted a physics student. Now, a few years later, I am a physicist that eagerly plays sociologist and biologist. The things I really like with the work behind this thesis is that the things we study is a part of the world around us. E-mail viruses, influenzas, power outages, meeting new friends and getting trapped in traffic jams are things we experience all our lives—it was a while ago that mainstream physics could say the same.

Defining a problem is usually the hard part, solving it is easier. This is probably true for all science. Another thing I like with interdisciplinary science is that the simpler the solution is, the better. This is just the opposite of the "science as sports" mentality that exists in some areas of mostly theoretical physics and mathematics. I am not saying that reductionist science is not needed, not at all, just that if a problem neither has an impact on the everyday world around us, nor is entertaining, then it is not worth the effort to solve it. On the other hand, I personally find a lot of rather bizarre mathematics quite entertaining, and other people probably enjoy the things I don't like. Ideally, all forms of intellectual entertainment should be supported, so I am not arguing for any redistribution of financial support, just explaining why I like the science of everyday life. (Well, I am not saying that everything in this thesis is very entertaining or relevant for our everyday experience. But, at least, that is always the goal when I start a new project.)

OK, there is also a darker side to what physicists do in the interdisciplinary arena. We might be good at various analytical and simulation techniques, but as we seldom actually prove thing as mathematicians do, we are able to make much headway in many fields of natural and social science. The only trouble is that we tend to loose ourselves in an anything-goes mentality. As Duncan Watts (himself a mathematician) writes in his *Six Degrees*:[356]

> Physicists, it turns out, are almost perfectly suited to invading other people's disciplines, being not only extremely clever but also generally much less fussy than most about the problems they choose to study. Physicists tend to see themselves as the lords of the academic jungle, loftily regarding their own methods as above the ken of anybody else and jealously guarding their own terrain. But their alter egos are closer to scavengers, happy to borrow ideas and techniques from anywhere if they seem like they might be useful, and delighted to stomp all over someone else's problem. As irritating as this attitude can be to everybody else, the arrival of the physicists into a previously non-physics area of research often presages a period of great discovery and excitement. Mathematicians do the same thing occasionally, but no one descends with such fury and in so great a number as a pack of hungry physicists, adrenalized by the scent of a new problem.

I laughed a lot reading this passage. Although I disagree that physicists regards "their own methods as above the ken of anybody else" (at least I don't). I believe the criticism in this quotation should be taken serious. There is no point in publishing a paper about e.g. biology if biologists cannot assimilate the results. To keep our reputation and scientific *savoir faire* there is, I believe, one method better than any other—to collaborate with scientists of other disciplines. Sure, I am far from the best in this respect, but more cross-disciplinary collaborations is what I hope for in the future . . . for myself and others.

You might wonder why I call myself a physicist when none[1] of the systems studied, and few of the methods I use, can be found in a physics course text book. Well, it is probably more of an old habit rather than a good description. The division of science into the disciplines of physics, chemistry, social science and so on, is so deeply rooted in the common consciousness that it is easier to say one belongs to one of them than to something outside.

One might argue that methods of statistical physics are useful for studying networks. I agree that there are good examples of that, but most network papers published in physics journals are based on Monte Carlo simulations, master equations, generating functions and other techniques that are frequently used in chemistry, statistics, and so on. My first choice would be to defy all classification as much as possible; or if it is really necessary to have a name, follow László Barabási[31] and call it the "science of networks." As for rigid

[1] Papers XI and XII is about the *XY* model of statistical physics, but in a setting that hardly can be realised experimentally and, in the context of physics, only is of academic interest.

institutions we probably have to wait to get departments of network studies, in that context I guess "interdisciplinary physics" has to do.

chapter 1

Introduction

This introduction starts with a three-section essay about the methodology, scope, usefulness and future of network science. No mathematical knowledge is required to follow these sections. To widen the perspectives (and maybe make the text more entertaining) I try to take examples from a broader area than the usual applications, I hope this will not confuse you too much. The last two sections are more technical, concerning the basic terminology and the various types of networks.

1.1 Network science

Networks are all around us. They can be concrete such as subway systems, power grids, the nervous system or the electric circuitry of a washing machine; or more nebulous such as networks of acquaintances, chemical reactions or chains of historical events. One of the central themes in network science is the extraction of information hidden within the wiring of a network. In this section I will try to give you an idea how that might work. Our first example is the network shown in fig. 1.1(a). This is a network of romantic relationships between celebrities as presented in the Swedish tabloid Aftonbladet.[127][i] The circles represent actors and actresses and a line between two circles means that the corresponding actors have had a relationship. One observation you might make if you contemplate the network is that there are circuits—you can follow edges (lines) from an vertex (a circle representing actress or actor) back to itself without traversing the same edge twice. Furthermore, there are only circuits of even number of edges. If that would not be the case, some of the contacts

[i] The network itself may be unscientifically acquired and even incorrect in some parts, but it works well as an example.

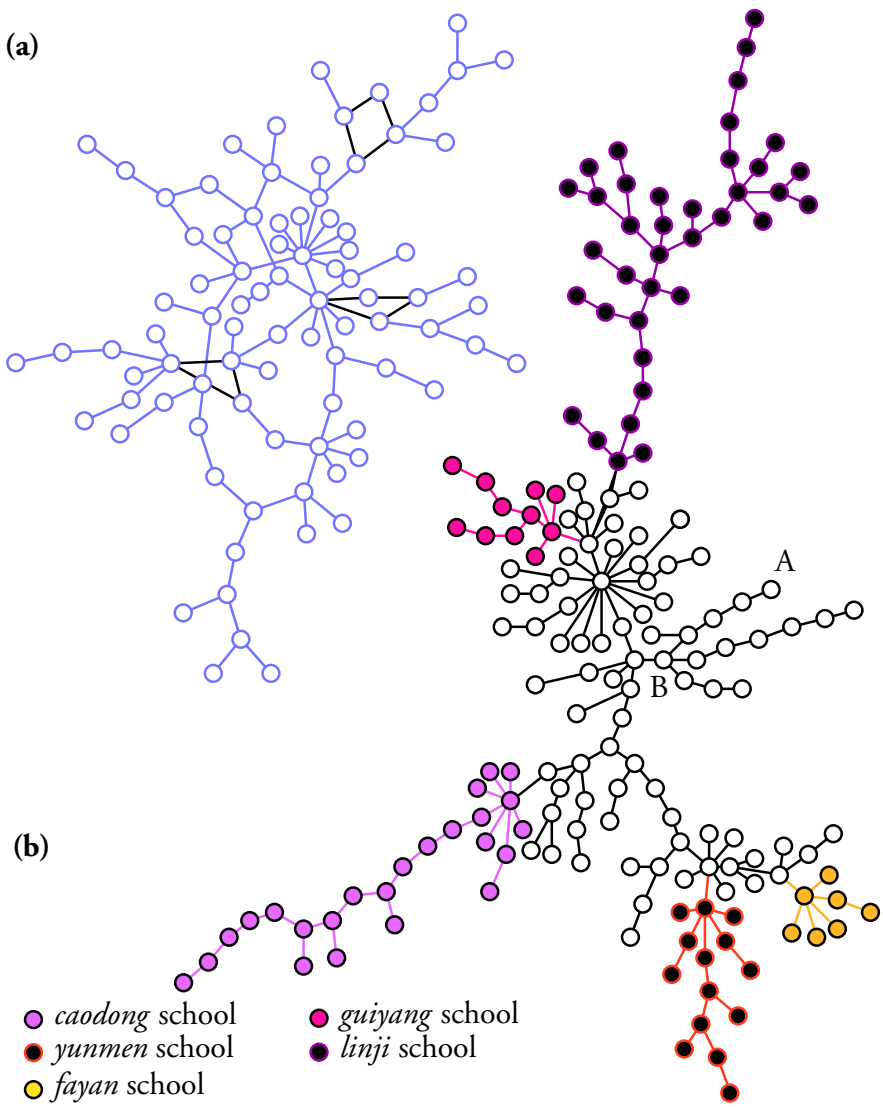


Figure 1.1: *Examples of social networks: (a) shows the network of relationships between Hollywood celebrities as reported in the Swedish tabloid Aftonbladet.*[127] *The black lines shows the circuits of length four. (b) shows a lineage chart of Chinese Zen masters.*[121] *This network is directed—in a tie (line) the master is the one fewest steps away from the root (indicated by A), the disciple is the one further away from the root.*

must have been homosexual.[ii] Yet another peculiarity one can observe is that there are three circuits of length four. Why is that so special? Well, it means that one actor must have dated the ex-girlfriend of his ex-girlfriends present partner. One might think that this would be ticklish situation, and indeed there is a study[42] arguing that high-school students avoid this kind of dating. So celebrities seem to choose their partners differently from high-school students. This might not shake you up very much—the important thing is that we were able to conclude this by just inspecting the network structure.

Let us take another example. Fig. 1.1(b) shows a lineage chart of Chinese masters and disciples of Zen Buddhism.[121] Zen sources claim this branch of Buddhism to be brought from India to China by the monk Bodhidharma (around 600 AD). The many influences by the Chinese Taoist tradition[305] that can be seen in Zen makes most scholars of today believe that Zen was developed in China and consider Bodhidharma to be fictitious or of little historical significance.[84] Nevertheless, the "root" of fig. 1.1(b) marked by *A* is Bodhidharma. (A network with no circuits is called a tree. If the links of a tree have a specific direction and the whole network can be reached from a node, this node is called the *root*.) If one follows the lineage down from Bodhidharma nothing dramatic happens until the monk labelled *B* is reached. From *B* the tree forks into two main branches. Indeed the sixth patriarch Huineng marked by *B* was a figure of paramount importance in the development of Zen.[108] The nodes of different colours represent different branches of Zen. We note that the first monks of a colour, the founders of the branch, all have more disciples than the average. This is perhaps quite natural: The monks founding a new branch are probably renowned already during their lifetime, thus attracting many disciples. Another perspective is that if a master does not have many disciples than his teaching is likely to be forgotten rather than form a new school. By now I hope you have an idea how network science can be informative in many situations. If you are able to rephrase such observations in mathematical terms, then you are a network scientist. It needs not to be very difficult. In my personal opinion, the simpler idea the better science.[iii]

[ii] Indeed there are no homosexual contacts in this data set. I will not speculate about if this is less than expected from a random sample of the population.

[iii] As I see it, the impediment for the interdisciplinary network scientist to overcome, is to know the literature in diverse fields of natural and social science. To be able to tie results from different fields together requires scholarly knowledge rather than the traditional analytic skills of prominent natural scientists.

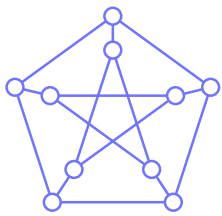


Figure 1.2: *The Petersen graph. A graph where no pair of vertices are separated by more than two steps and the shortest circuit has length five.*

We conclude with a few words about mathematics: Probabilistic graph theory (see refs. 176,55,54) the branch of mathematics closest to the themes of this thesis, has a long tradition. As many times before the mathematicians watch with frustration when physicists rush ahead without caring much about details. In some cases (e.g. the Barabási-Albert model of sect. 3.4.1) physicists' network models do not define a probability space, and hence not a model in the eyes of mathematicians. In this thesis I do not pay very much attention to details and I do sometimes use physicists' hand-waving arguments. By doing this one can often reach further than one would with a strict theorem and proof style, but there are certainly features of graphs that one cannot really deduce from approximation arguments: One example is the four colour theorem—given any graph you can draw on a paper without any edges crossing, four colours are enough to colour every vertex such that no edge leads between two vertices of the same colour.[19] Another (less practical) example goes as follows: Draw five dots on a circle. This is a graph with the following two features: 1. No pair of dots are more than two steps away from each other. 2. The shortest circuit has five steps. Are there other graphs with these two properties? After thinking a while you will probably realise that the answer is yes and that the graph in fig. roves this. We cannot get much further than this without mathematical proofs. The rest of the story is that a graph has our two properties, then every vertex has the same number k of neighbours. Now apart from our two examples with $k = 2$ and $k = 3$ there is also a graph with $k = 7$ that have the two properties. Yet any? Yes, there is possibly one with $k = 57$ too, but for almost half a century people searched for it without result.[161]

1.2 Form follows function (and vice versa)

The dictum "form follows function" was coined 1896 by the American architect Louis Sullivan.[331] Ever since, it has been seen as the nub of the functionalist philosophy of architecture and design. It might be argued to be an "empty jingle"[89] in its original sense, but it makes a good starting point for a further discussion of the aims of network science: Many of the infrastructural networks around us are designed by humans; other networks can be altered by

design—biochemical networks can be manipulated to cure diseases;[306] your local World Wide Web domain can be made more efficient and visible by careful hyper-linking (cf. ref. 275). In such examples one benefits from knowing just which form (network structure) that gives the intended function.

Apart from giving tools for engineering, many branches of science strive for ability to predict the behaviour of a system. To be able to make predictions, one needs to pin down the precepts of a system's response to the influences of its surroundings. In terms of networks this works just as in section 1.1: From fig. 1.1(a) we know to some (in this case rather small) degree of certainty that famous Hollywood actors do not decline a date with a Hollywood actress just because she happens to be the ex-girlfriend of his ex-girlfriend's new boyfriend. Now, this is a possible ingredient of a model for dating among Hollywood celebrities. Such a model can in turn be combined with a model for the spread of sexually transmitted diseases, to give predictions about how susceptible celebrities are to, say, chlamydia. This question might not be the most pressing issue mankind faces today, but the way to predict—to go from observations, via modelling to prediction—is used ubiquitously in network science. To continue the simile to Sullivan's motto, this methodology is to first see how form follows function (that celebrity dating gives many four-circuits), and then by modelling be able to see how function follows form (how the many four-circuits affects the epidemics of chlamydia). One shall keep in mind that, to make predictions, we do not only need to understand the network structure but also the dynamical system on top of it.

1.3 Complexity and the other buzz words

During the last decades complex systems science has emerged as a common interdisciplinary theme linking traditional branches like physics, chemistry, biology, ecology, and even social sciences such as economy and sociology. The establishments of complex system research centres such as the Santa Fe Institute and journals like *Complexity*, *Advances in Complex Systems* and *Complexus*, have fueled this movement. Nevertheless, the phrase "complex systems" is not very much more well-defined in science than in the vernacular. A common definition is that a complex system is a system consisting of many interacting units whose collective behaviour cannot be explained from the behaviour of the individual units alone. The different traditional disciplines have developed models of complex systems relatively independently. One may say the social sciences are nothing but complex systems research—remember Adam Smith's 1776 idea that an individual who "intends only his own gain" acts as if "led

by an invisible hand to promote [···] the public interest."[315] Or in Thomas Schelling's concrete example about air travel:[303]

> Somehow all of the activities seems to be coordinated. There's a taxi to get you to the airport. There's butter and cheese for lunch on the airplane. There are refineries to make the airport fuel and trucks to transport it, cement for the runways, electricity for the escalators, and, most important of all, passengers who want to fly where the airplanes are going.

As in this example, the subject systems are most often complicated, but to understand the global behaviour of these one uses very simple models. The branch of physics dealing with collective behaviour of many interacting elements is called "statistical mechanics" or "statistical physics." The traditional goal of this discipline is the study of phase transitions—when the collective behaviour is qualitatively altered due to changed environmental conditions (shifts in temperature, pressure, etc.).[326,322,170] Similar effects in complex systems in biology and social sciences can be (and have been) studied. But physicists' (and others) models in these areas serve to identify the minimal requirements for the behaviour of individuals for a certain collective behaviour to emerge.[iv] To take one concrete example, the study of social insects has the last decades become one of the most vital sub-disciplines of entomology.[58,76,145] Many species of insects live in colonies that operate without central control. The fitness of the colony hinges on the number of individual insects—an individual, or a too small group of insects, fails to forage efficiently;[44] but if the number is large enough, the colony will *self-organise*[v] to a state of efficient collective behaviour. By modelling the individual behaviour of ants one can single out the ingredients necessary for an efficient collective behaviour to emerge. In general, the common scientific question, asked in modelling complex systems as outlined above, is what ingredients of the individual behaviour that are necessary to produce the observed collective behaviour; or in other words, what is the minimal model that can reproduce the system's behaviour. Another concrete example of a phenomenon where the understanding is meager is the spontaneous emergence of fads and (in a greater perspective) sub-cultures: Japanese youths have a

[iv] Yes, *emergence* is another buzz-word of turn-of-the millennium science; and it used to signify when collective behaviour is something else from what can be expected from the individual entities. Some good introductions are refs. 15, 135, 162, 349.

[v] Here is another buzz-word. "Self-organisation" shares much of the meaning of "emergence." The social connotations of the word "organise" makes this, I think, the preferable term for systems where the interacting entities are living organisms.

reputation for joining sub-cultures with great zeal and commitment. In 1979 a small boutique called *Takenoko* (bamboo shoot) opened in Harajuku, Tokyo. It featured unisex clothing, including billowing pantaloons and baggy coats that came down to the ankles, in a range of bright colours. By the autumn, youths wearing Takenoko fashions started dancing at the nearby Omotesando street. These gatherings grew rapidly and the kids, known as *takenoko-zoku* (bamboo shoot tribe), became the media darlings of 1980.[336,304] At its peak there were several hundred dancers, and the Takenoko fashion had spread to other major cities of Japan. The interesting thing is that all this happened without much advertisement from the Takenoko-shop itself, yet million dollar advertisement campaigns often fail to stir such commotion. The takenoko-zoku is just one example, raising the following question: How can some habits turn from insignificant into nation-wide fad, when others die out even if backed by massive advertisement? Indeed, it has been argued that, with the advent of modern information technology, spontaneously formed communities will not only change character but also be an increasingly important social force.[291] There are theories and models around,[139,147,355] but a full understanding demands a joint description from both psychology (see e.g. ref. 184, 284) and complex system science.

Now let us turn to the phrase "complex networks" as in the title of this thesis. Now a network is, in the classical sense, a static structure; can it still be so that the "complex" of "complex networks" is the same as in "complex systems"? Yes, because a network is often the cast of a complex system. So by studying the network one can study the underlying complex system itself. Yet, many kinds of networks, biochemical networks being one example, do not have any directly corresponding systems of interacting agents. In these cases we can motivate the epithet "complex" by another argument: These networks are complex in that any comprehensive description of their nature has to be lengthy. They contain much randomness, but to model them as purely random graphs (cf. sect. 3.1) would not be very apt since they also contain structure. This latter kind of complexity can be quantified by the length of a computer program to regenerate a representative member of the class of networks in question.[216]

To conclude this section on complexity we note that there are other ways of modelling complex systems than breaking it down to simple components. One can more bluntly simulate the whole system with all its components to the best of ones ability. In this way one has, for example, created a test bed for the mass flow of chemical substances in a cell,[338,337] with which one can put the cell in starvation and other extreme situations. This approach has also been

used to simulate people's daily transportations within a whole city.[vi] This latter approach to complex system modelling is appropriate when the system consists of very different entities that cannot be disintegrated and direct experiments are difficult to execute. Much criticism has been raised against these projects by more traditional-minded modellers, but to their defense one can say that this kind of modelling will be more important in the future as databases get more comprehensive.

1.4 Graph theoretical preliminaries

Before embarking on the exposé of complexity and networks, and the introduction to my work, I will go through the basic definitions of graph theory. If this section is hard to digest at a first reading, you can use it as a glossary for the rest of the thesis. For further reading, my personal favourites are ref. 343 and 70.

As mentioned, a *network* is a system that can be modelled as a set of entities that are connected pairwise. We will call the entities *vertices*[vii] and denote the set of vertices V. Pairs of connected vertices are called *edges* and the set of edges is denoted E. The number of elements in V and E are denoted N and M respectively. When we talk specifically about directed connections we talk about *arcs*. The set of vertices together with the set of edges (or arcs) defines a *graph* $G = (V, E)$. The term "graph" refers only to such mathematical representation, whereas "network" can refer both to the actual system and the mathematical representation. I will not make any effort to be more precise in my usage of the term "network." Normally one does not allow self-edges (or *loops*) from a vertex to itself; neither are more than one edge between a vertex pair. Without these restrictions one talks about a *multigraph*. The matrix with entries $A_{vw} = 1$ if $(v, w) \in E$ and $A_{vw} = 0$ otherwise is called the *adjacency matrix*. A subset W of vertices along with a subset of edges within W is called a *subgraph*. A subgraph is said to be *maximal* with respect to some property if it cannot be extended without loosing this property. If one just talks about a *maximal subgraph* of a vertex set $V' \subseteq V$, one refers to the graph (V', E'), where E' is the set of all $(v, w) \in E$ such that $v, w \in V$. A maximal subgraph

[vi] See the homepage of the TRANSIMS project: transims.tsasa.lanl.gov.

[vii] The nomenclature is not clear in this matter. Vertices are called "nodes" in computer science, "sites" in mainstream physics, "actors" in social science. Edges are called "links" in computer science, "bonds" in physics, "ties" in social science. Well, actually all terms might occur in all disciplines along with others like "items," "points," "connections," "wires" and so on ...

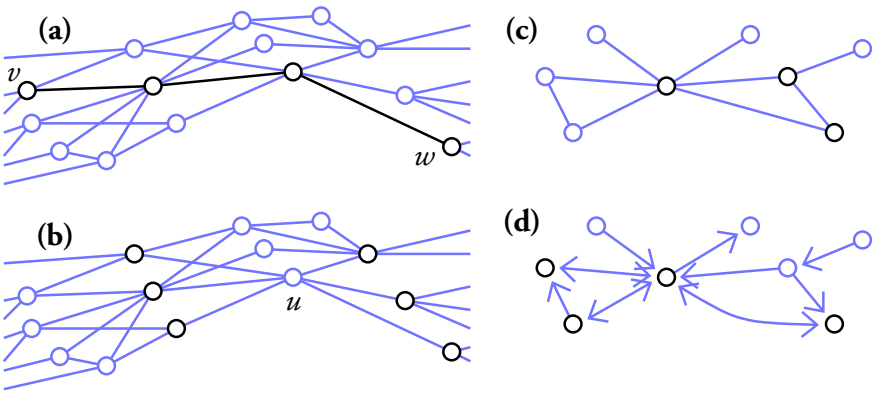


Figure 1.3: *In (a) one (of the two) geodesics between v and w is drawn in black. The distance between v and w is thus three. In (b) the degree of u is six and u's neighbourhood is marked black. (c) shows a graph with radius* 2 *(from the black vertices with eccentricity* 2*), diameter* 3*, and characteristic path length* $25/14 \approx 1.78$*. In (d) we see a directed graph. The black vertices marks the largest strongly connected component. The graph itself is the largest weakly connected component.*

of two vertices is called a *dyad*; a maximal subgraph of three vertices is called a *triad*. A subgraph whose every pair of vertices are connected is called a *clique*. Two graphs $G = (V, E)$ and $G' = (V', E')$ are said to be *isomorphic* if there is an one-to-one mapping between V and V' such that a pair of vertices forms an edge in E if and only if the corresponding pair of vertices in G' forms an edge in E'. Quite commonly, numbers are assigned to the vertices and edges (or arcs) of graphs, in such cases the graph is said to be *weighted* or *valued*.

A central concept in graph theory is a *path*, mathematically defined as a sequence of vertices $\{v_1, v_2, \cdots, v_n\}$ such that (v_i, v_{i+1}) is an edge (or arc) for every i. A *circuit* is a path that ends where it starts (i.e. $v_1 = v_n$). A circuit of three edges is called a *triangle*. An *elementary path* is a path where no vertex appears twice; and similarly an *elementary circuit* is a path where only the first and last vertex appears twice.[viii] The *length* of a path is the number of edges in it, or the number of vertices minus one. The path, starting at vertex v and ending at w, with smallest possible length is called a *geodesic* between v and w. The length of a geodesic between v and u is called the *distance* between v and w and denoted $d(v, w)$. The set of vertices at distance 1 from v is called the

[viii] Even here there is much ambiguity in the literature. Sometimes a circuit is called a *cycle*, sometimes "cycle" denotes an elementary circuit. Please be cautious!

neighbourhood of v (denoted Γ_v). The maximal distance from v to any other vertex in G is called the *eccentricity* of v. The largest eccentricity value among all vertices is the graph's *diameter*, the minimal eccentricity is *radius* of the graph. The average distance is often referred to as the *characteristic pathlength* (and sometimes, regrettably, as "diameter"). A graph without circuits is called a *tree* if it is connected or *forest* otherwise. A graph without circuits of odd length is called *bipartite*. A bipartite graph can be bisected into two sets such that no edge goes between vertices of the same set. If one explicitly discriminates the vertices of the two sets, so the graph can be written $G = (V_1, V_2, E)$, the graph is said to be in a *two-mode* representation.

A maximal subset W of vertices such that there is a path between every pair of vertices in W is called a *connected component* of vertices. In directed graphs we talk about a *strongly connected component* W if there is a path from every vertex in W to every other vertex in W. Moreover, one talks about a *weakly connected component* if there is an undirected path[ix] between every vertex pair. The *reflexive closure* of a directed graph $G = (V, A)$ is G extended with an arc (v, w) for every arc (w, v) such that $(v, w) \notin A$. The property whether or not G itself is a connected component is the *connectivity* of G.

The number of edges containing the vertex v is called the *degree* of v.[x] If G is a directed graph one discriminates between *in-degree*—the number of incoming arcs, and *out-degree*—the number of outgoing arcs; and correspondingly the *in-neighbourhood* Γ_v^{in} (other vertices with arcs pointing at the vertex in question) and *out-neighbourhood* Γ_v^{out}. A list of the degrees of a particular graph is called a *degree sequence*. A vertex of degree zero is called an *isolate* and a vertex of degree one is a *leaf*.[xi] At last: If M is bounded by a function linear in N it is said to be *sparse* (otherwise G is *dense*). Many of the concepts introduced in this section are illustrated in fig. 1.3.

1.5 Types of networks

In this section we take a closer look at systems for which networks are relevant models. There are many ways to classify networks. For example, a good way would be to distinguish between networks where the edges are physical objects and those where edges represent interaction of some kind. In this section we follow the distinction from the scientific branches to which the networks be-

[ix] In an undirected path the arcs can also be followed backwards.

[x] To add to the confusion, degree is sometimes referred to as connectivity.

[xi] This term is most common in the context of trees. (This is not a bad joke.)

long and distinguish between social networks of humans, other social networks, economical, infrastructural, biological, technological, and other networks.

1.5.1 Social networks of individuals

The study of the social network[308,351] of persons, and the ties between them, is probably the oldest discipline of applied graph theory.[248,92,296] The nature of the ties can be any imaginable social interaction, from romance among adolescents[42,68] to professional ties between corporate executives,[93,94] from interaction on and off the track at a NASCAR raceway[297] to sex[219,218,202,283] and pair-formation among tango dancers.[212] (We have already had two examples—the Hollywood celebrities and the Zen monks of section 1.1.) Yet other studied networks include networks of terrorists,[207] musicians,[140,96] Mormons[49] and karate practitioners.[370] The questions that social network studies try to answer have much in common with complex network studies in general:

What are the relevant structures in the networks? Many structural measures have been developed to describe social networks, and have later been brought over to other network topics. Examples of such quantities are centrality measures that will be discussed in sects. 2.1.2, 2.1.3, 2.1.4 and structural equivalence of sect. 2.2.2, and many more not mentioned in this thesis (the interested are recommended refs. 308, 351).

What processes give rise to these structures? The most distinguishing feature of the physicists' character—as I see it—is the idea that everything that can be analysed quantitatively should be modelled[xii] (and if it cannot be analysed it should be modelled anyway). Thus it is not surprising that social network studies by physicists to a rather large extent present models for network growth, while this kind of studies is rarely seen in sociologists' papers. See e.g. refs. 30, 42, 122 for sociologists' network models and e.g. refs 180, 229, 111, 80, 17, 358 for network models by non-sociologists. An important aspect of social networks is how social groups are formed and how the individual traits of persons relates to the groups they belong to. Such topics form an interface between psychology, sociology and economy.[284,184,323,273,303] A notable subgroup of such studies are the study of how race and ethnicity of the actors relate to the network structure.[68,365,285,246]

[xii] Note that physicists use the word model in a slightly different way than statisticians and mathematicians, see the introduction to chapter 3.

How are dynamical processes affected by the network structure? The dynamical processes that are relevant in the context of social networks are for example disease spreading[28,136,122,218,202,283] (see also sect. 4.1) the spread of rumours, fads and ideas,[26,355] and game theoretical models of social stability and other states of society[372,1] (see also sect. 4.5).

Except these points there are a few issues quite unique for social network studies. One can for example study how much the actors know about their social surrounding[191,48] and how they use this knowledge to find other persons,[357,102] search for jobs,[146,371] and so on.

1.5.2 Other social networks

Apart from the social network of individual people tied by social interaction one can imagine many other types of networks formed by social forces. One can consider the two-mode representation of the books in a library and the borrowers. Such a network contains no information on any interaction between people, but one can imagine that such a network has a similar type of community structure[xiii] as an acquaintance network.[xiv] Another type of networks in this category that has been studied are the networks of families linked through marriage[64,277] or gift exchange.[336] Similarly, one can study ties between tribes (in ref. 153 ties between tribes of the eastern New Guinean highland are investigated).

Other more uncommon social networks include interaction between animals[90,78,301,160] and fictional characters in Leo Tolstoy's *Anna Karenina*[203] and Marvel Comics comic books.[8]

1.5.3 Economical networks

A network based modelling is appropriate for many types of economical systems. A good example is trade networks, where the vertices represent organisations, companies or individuals and an arc (yes, these are networks that are naturally modelled as directed graphs) represent that one vertex has bought something from another vertex.[362,369] One can also study more abstract networks of business contacts (conveying control or cooperation).[228,242] The social net-

[xiii] I.e., that the network can be decomposed into distinct subgroups that are more connected (cohesive) within than to the outside. See also sects. 2.2.4 and 2.2.5.

[xiv] I do not know of any study of such networks, but it would indeed be very interesting, much because of the large databases of such information that exist.

work of CEO's[134] or boards of directors[228,93,94] forms an interface between social and economical networks studies.

Another example researchers have considered is the network of joint-stock companies, where there is a link between two companies if the time development of their shares values are correlated[239,227,210,292,61] or if they are connected in some more general way.[235]

One can also think of the money flow among bank accounts as a network. This kind of networks could be reconstructed from the vast databases within banking corporations. There are of course many legal issues to be cleared in order to get hold of such material (which is probably the reason that there has not been any studies, to my knowledge, of such systems).

1.5.4 Infrastructural networks

Power-grids[358,354,272] and railroads[309,214,310] and the wirings of the Internet[118,280,347] are examples of networks where the edges and vertices are physical objects. This property puts many constraints on infrastructural networks that give them many common properties. The Internet is hard to map out cable by cable, instead one often looks at either the networks of so called Autonomous Systems[171]—itself networks of computers and servers roughly corresponding to an Internet provider; sometimes a vertex is associated with an IP address.[67]

Network considerations will probably be very important in future design of infrastructural networks. One reason for this is that modern infrastructural networks tend to get increasingly intertwined—power-grids need an operating telecommunication network for control and the telecommunication needs electric power.[292] Such issues need to be integrated in future models of infrastructural networks.

1.5.5 Biochemical networks

There is biochemistry behind all life. When we eat a portion of Russian caviar receptors on our tongues spark a biochemical chain-reaction, or signal transduction, and when the caviar turns into energy in our body it does so through a series of interconversions between chemical substances called the metabolism. To understand biochemistry is essential for the design of medicines. Viagra, as a well-known example, functions by obstructing a signal transduction pathway (or path, in mathematical parlance). For a complete modelling of the biochemistry of an organism, i.e. to find the concentrations of the chemical substances

as functions of time for different conditions, requires that the reactions are represented by differential equations. It would often be a too complex task to set up and integrate the entire set of coupled differential equations (although there are exceptions[338,337]). Here network studies can help in two respects—one can either try to find small subnetworks representing some function, or one can try to extract information from the networks as a whole. One approach of the first category is to see what types of small subgraphs that are present in an abnormally large or small number in a biochemical network,[312] another is to try to demarcate the boundaries between relatively autonomously functioning subnetworks[306,307]

A major issue with biochemical networks is how to choose the appropriate network representation. Should they be directed or not? In principle all reactions are reversible, but they can, under normal conditions, be almost infinitely skewed to one side. Should they be valued or not? For sure, the concentration of some substrates (chemical substances), like water, is way over the average. Should they be an one-mode network of substrates or two-mode network of substrates and reactions? The set of substrates is in principle the only interesting vertex set, but with an one-mode representation one looses information such as what substrates that needs to be present for a reaction to occur.

1.5.6 Other network of biology

Organisms have, apart from biochemical networks, much more tangible networks too. The nervous system is one important example where both edges and vertices are nerve cells. To map out the neural network of a larger animal would be an enormously tedious task. Instead one can either study the network of some microscopical organism[272,364] or study the network of functional correlations, e.g., the human brain.[325] In such studies the brain scanned during a period of time and the three-dimensional compartments of the scan whose activity are strongly correlated are considered to be linked. Such networks are not an attempt to recreate the physical wirings of the neural network (they have indeed some features that neural networks are known not to have), but to be a tool for analysing the brain's functionality.

Another concrete, and well-understood, class of biological networks is the vascular systems of animals and plants.[359,29]

1.5.7 Ecological networks

Networks of species that prey on other species—so called food webs[282]—are ecology's contribution to the zoo of real-world complex networks. It is a tedious project to map out a food web. There are a few larger studies of well-defined ecosystems such as lakes,[230] bays[27] and estuaries,[154,172] which all have been studied from different structural perspectives.[319,245,224,225,367,110,109]

1.5.8 Technological networks

Technological networks such as the wiring of electronic circuits[123] and dependency graph of large-scale computer software systems[263,345] have been studied. Such networks are designed both to optimise computational performance and to be easy to develop, and conflicting objectives of these design principles can lead to interesting network structures. Furthermore—technological networks often serve as a metaphor for other types of networks (the "modularity" concept being one example, see the Summary of paper I).

1.5.9 Information networks

The vertices of information networks can be any kind of information source (newspapers, scientific monographs, web pages, etc.) and the edges are references from one source to another. Two thoroughly studied networks in this category are the World Wide Web (WWW) and networks of citations between scientific papers. Both these networks are naturally represented as directed graphs. The main difference is that the citation graph has no circuits.[xv] Both the World Wide Web and citation networks can be mapped out by "network crawlers"—computer programs that follow the links of the network and report the network structure back. The WWW graph is the biggest mapped out network to date, an early study reports over 2×10^8 vertices and 2×10^9 edges.[66][xvi] The building of efficient crawlers is an area to which network science has given a substantial contribution (more about that in sect. 2.1.4).

[xv] Well, at least almost none. The only chance of a circuit to occur is if one paper refers to an unpublished manuscript—something that is not uncommon in natural sciences. (This is, I am sure, quite shocking for all ex-sinologists in science; and yet more shocking is the fact that some of us adapt to this barbarian tradition—see e.g. paper XV.)

[xvi] In reality it is much more: The search engine Google reports that it (the 15th March 2004) index 4.3×10^9 pages.

1.5.10 Other networks

In this section I give examples of a few other types of networks that do not fall into any of the above categories.

A historical event can often be described as a chain of smaller events. Historians refer to such descriptions as "cases." It can be useful to make a map of how the events are related—a graph where vertices are events and an arc means that one event is triggered (or influenced) by another.[69,41] By analysing this network one can estimate which narratives that are most crucial to the overall description; how reliable the account is; and, if there are contradictive narratives, which ones that are likely to be phony. Ref. 41 performs such a network analysis of cases derived from the Swedish leftist intellectual Jan Myrdal's book *Report from a Chinese village*.[253]

Other rather unusual applications of network science are linguistic networks where the vertices are words or word classes and the edges link words that are related somehow. The relation can be constructed from a relations listed in a thesaurus[203,196,329,251] or words occurring juxtaposed in texts.[125]

Yet another category is preference networks which link people to items (books, films, music, etc.) they like.[290,311,143] This is the structure behind recommender systems typically found at online shopping sites.

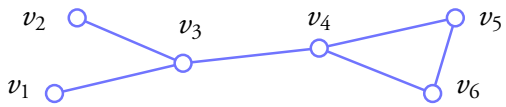
chapter 2

Measuring network structure

A network is a reflection of the forces that formed it. Think of networks where the vertices and edges are physical objects—like the network of roads or railroads,[309,214,310,183] or vascular systems like the human blood vessels[359,29] in the human body; for this kind of networks the number of edges to a vertex is limited by the space, so you never see an one-hundred-road highway crossing or the confluence of a thousand blood vessels. (Quantities such as the degree will be discussed in sect. 2.1.) With a few assumptions on the dynamics that forms the network we can go the other way—given a network we can study the network-forming forces. Such studies can be qualitative: If a large fraction of an acquaintance network consists of isolated subnetworks, then we can conclude that the society is very segregated. We can also try to quantify the network-forming forces from the structure of the network: Given an acquaintance network we can estimate how strong the segregation in society is by measuring (roughly speaking) how few links that have to be cut to split the network into well-connected, isolated, communities.[267]

In many economical and biological networks the individual vertices have different roles and functions. Another approach to network structure is therefore to label and classify the vertices with respect to each other. For this purpose there are two principal methods: Either one starts from the whole vertex set and splits it into finer and finer partitions; or one measures the pairwise similarity between vertices and group the vertices together according to mutual similarity. This kind of methods will be discussed further in sect. 2.2.

We finish with a few words of caution: A final goal of almost all network studies is to find in what way the network structure affects the dynamical systems that are confined to the networks. Many quantities in this section take their names from good guesses of what would be the situation in a dynamical



	k_v	eccentricity	C_C	C_B	w_v	C_I	c_v
v_1, v_2	1	3	0.455	0	0.183	0.581	0
v_3	3	2	0.714	14	0.417	0.947	0
v_4	3	2	0.714	12	0.584	0.947	0.333
v_5, v_6	2	3	0.455	0	0.457	0.720	1

Figure 2.1: *A tiny graph and some vertex quantities: Degree k_v and eccentricity of sect. 1.4, closeness centrality C_C of sect. 2.1.2, betweenness centrality C_B of sect. 2.1.3, eigenvector w_v and information centralities C_I of sect. 2.1.4 and the local clustering coefficient c_v of sect. 2.1.5. In almost all networks the centrality measures are, on average, quite correlated. (I do not know of any network where there is a negative covariance of the above measures.) However, individual vertices might be central in one measure and peripheral in another.*

system. For example "information centrality" of sect. 2.1.4 does not necessarily correspond to the accumulation of information at a vertex (but it surely is a good guess as we will see in sect. 4.4); the average reciprocal geodesic length (see sect. 2.3.1) has been called "efficiency,"[213] betweenness (see sect. 2.1.3) has been called "load."[141,142i] But to really investigate the essence of these names, one has to look at each dynamical system separately. The situation gets even worse for buzz words like "robustness" and "hierarchy" where an ample number of definitions exists, each trying to capture a certain aspect of the colloquial word, and each suitable for different dynamical contexts.

2.1 Vertex and edge quantities

In this section we discuss quantities defined for individual vertices and edges.

2.1.1 Degree

The structure that has received most attention the last few years is the degree, or rather the probability density function of the degree—the *degree distribution*. This popularity is due to the discovery of Barabási and coworkers[32,179,33] and

[i] I use betweenness as a load assessment too (in papers IX and X).

the Faloutsos brothers[118] that many real-world networks have a power-law[241] degree distribution—i.e., the probability density function $p(k)$ (k is the degree) has the form

$$p(k) = ak^{-\gamma} \tag{2.1}$$

where a and γ are constants. This kind of distribution is very far from what would be expected from a purely random network formation process. Furthermore, power-laws have a special role in statistical physics, which is probably the reason for much of the excitement around power-law degree distribution: The traditional models of statistical physics usually exhibit a phase transition between ordered and random states. Exactly at the transition the system has "self-similar" or "fractal" or "scale-free" properties—no matter if one zooms in on a snapshot of the system it will look qualitatively the same. The process of zooming in or out—*rescaling* in physicist parlance—on a real to real function f is mathematically treated by tuning a scale parameter $a \in (0, \infty)$; so f is scale-free (has the same form for all scales) if it obeys

$$f(ax) = bf(x)\,, \; x \in \mathbb{R} \tag{2.2}$$

(b is a constant). The only solution to eq. 2.2 is a power-law as eq. 2.1. Networks with a power-law degree distribution are, for this reason, often called "scale-free networks." Given some class of graphs with power-law degree distribution it is, however, not necessarily true that the network is scale-free. There might be scales visible in other structural measures.

Another issue with many alleged power-laws in nature and society is that they often must have cut-offs, not only by the finiteness of the real world but also by the constraints of the environment (other interacting systems). These cut-offs represent scales that are both telling accounts of the process of network formation and probably very relevant to dynamical systems. In the physicist literature, so far, the regions fitting to power-laws have attracted more attention than the cut-offs. The maybe most conspicuous example is physicists models of social networks: Degree distribution for such include power-laws,[112,219,7] power-laws with cut-offs[272,260,316,42] (see also paper III), exponential distributions[94,266,151] and Gaussian distributions[119,49]—most likely different social networks have different types of degree distributions. Despite this observation some authors[194,80,111] present models for social network having emergent power-law degree distributions without discussing this point (even though they obviously regard the degree distribution as a quantity important enough to

measure).[ii]

Degree can also be a building block in more complex schemes for analysing networks. For example, it has been used as a proxy for "importance" in constructing a method for quantifying "hierarchical organisation" in networks,[342] and for identifying chemical substances that exist in large concentrations in cell.[307] For such auxiliary uses of degree one needs to have a good understanding of the networked system—degree of a high-way map probably make an useful measure of "importance" of cities, but the CEO needs not to have more subordinates than mid-level executives.

For many types of communication, or information flow, networks a node of high degree has the role of a centre in its surrounding, so degree can be said to be a local centrality measure.

2.1.2 Closeness centrality

A vertex of high degree is central in its surrounding, but can be peripheral in the graph as a whole in the sense that the average distance to other vertices is large. In remedy one can use the *closeness centrality*[43,300,40]

$$C_C(v) = \left[\frac{1}{N-1} \sum_{w \in V \setminus \{v\}} d(v,w) \right]^{-1} . \tag{2.3}$$

This is the simplest possible *global* centrality measure in the sense that the closeness centrality for an individual vertex is influenced by the wiring of the entire graph. A fast approximation scheme can be found in ref. 114. The closeness centrality for the vertices of a small graph is shown in fig. 2.1.

[ii] In my eyes, all these models are relevant and very interesting. But they illustrate the slightly too warm sentiments towards power-laws among many physicists. The abundance of power-law generators at least tells us that there probably is not one underlying mechanism behind power-laws in different systems. Note also that the examples of social networks with a good fit to a power-law must probably have an upper cut-off of the degree distribution if the network was extended to the whole population of the Earth: If the network of sexual contacts (a year back in time) had a $p(k) \sim k^{-3.5}$ as reported in ref. 219 then, assuming $N = 6.4 \times 10^9$ the most active person would have around $N^{-1/2.5}(\text{year})^{-1} \approx 23$ new sexual contacts per day.[86] If the number of phone-calls per person and day is distributed as $p(k) \sim k^{-2.1}$ as reported in ref. 7 the person calling most would dial $\sim 5 \times 10^5$ phone-calls a second, 24 hours a day ... In the spirit of keeping a model simple one does not need to make it realistic in every detail, but I think it would be good if one points out how the cut-offs enters the modelled real-world system. I am not the best in this respect myself—see paper V—but I will try not to forget that if I ever present a power-law generator again ...

2.1.3 Betweenness centrality

The *betweenness centrality*[18,129] (or, *betweenness*, for short) is designed to capture how central a vertex is in the communication flow in networks like the Internet, social networks,[39] or any other networks where short pathlengths matter in the routing of the communication. Briefly, the betweenness of a vertex v is the number of geodesics between other pairs of vertices that passes v. More mathematically the betweenness centrality $C_B(v)$ is defined as

$$C_B(v) = \sum_{u \in V} \sum_{w \in V \setminus \{u\}} \frac{\sigma_{uw}(v)}{\sigma_{uw}} , \tag{2.4}$$

where σ_{uw} is the number of geodesics between u and w, and $\sigma_{uw}(v)$ is the number of geodesics between u and w that passes v. Similarly one can define *edge betweenness*

$$C_B(e) = \sum_{u \in V} \sum_{w \in V \setminus \{u\}} \frac{\sigma_{uw}(e)}{\sigma_{uw}} , \tag{2.5}$$

where $\sigma_{uw}(e)$ is the number of distinct geodesics between u and w containing the edge e. The fastest algorithms for calculating the betweenness of the whole network runs in $O(MN)$ time and can be found in refs. 62, 259.

It is important to remember the approximation behind the betweenness centrality—that all communication travels along shortest paths. For real networks this is often only true in a "free-flow" state (i.e. when the traffic never has to stop enroute). For example, Internet can often be quite jammed[335,133,132] and by congestion control implemented in routing protocols[171,223] data packages makes detours to their targets. This tends to congest not only the most central vertices but also their neighbours (cf. paper XIII). Notwithstanding, betweenness is a reasonable first approximation of the traffic load in networked communication systems and is frequently used as such (see refs. 141,142,37,36 and papers IX and X).

To amend some of the shortcomings listed above a few extensions of the betweenness concept have been proposed. One extension is the *flow betweenness*,[130] defined as the sum of flows through a vertex when a maximal flow[343,6] goes between all vertex-pairs. This algorithm has the obvious advantage that it does not only take the shortest paths into account. Furthermore, it needs no generalization to work with valued graphs. The drawback is that the calculation of flow betweenness is quite time demanding—$O(M^2N)$ to be precise.[6] Another extension is *random walk betweenness*[256,267] defined as the fraction of

time steps a random walker[iii] visits a certain vertex averaged over time and starting vertices. In practice, the random walk betweenness is calculated through matrix algebra. Just as the flow betweenness, this measure also accounts for suboptimal routes but is also affected by a $O(M^2N)$ running time on sparse graphs. The betweenness centralities of a tiny example graph are shown in fig. 2.1.

2.1.4 Eigenvector and information centralities

In many types of networks, being close to the most powerful (or important) vertex implies power (or importance). If the CEO turns to a subordinate for advise then that employee has a great deal of power; or if a well-known webpage links to your homepage than your homepage is probably famous as well. The method designed to capture these observations is called *eigenvector centrality*[59,60,308,351] w_v: The vertices $v \in V$ are assigned positive weights w_v proportional to the sum of weights of the in-neighbourhood:[iv]

$$w_v = \lambda^{-1} \sum_{u \in \Gamma_v^{\text{in}}} w_u, \text{ or in matrix form } \mathbf{Aw} = \lambda \mathbf{w}\,, \tag{2.6}$$

where $\mathbf{A}$ is the adjacency matrix of G, λ is the largest eigenvalue and $\mathbf{w}$ is the corresponding normalised eigenvector.

The eigenvector centrality measure has found an important application in the design of WWW search engines.[65,278] Since a hyperlink indicates that the web-page referred to has some relevance for the reader, the eigenvector centrality gives an estimate of the overall importance of the page. To sort the entries to a query, search engines use eigenvector centrality.

A similar measure is the *information centrality*[328] of social network analysis: Consider the matrix $\mathbf{B}$ with elements

$$B_{vw} = \begin{cases} 1 + k_v & \text{if } v = w \\ 1 - A_{vw} & \text{if } v \neq w \end{cases} \tag{2.7}$$

[iii] The random walk is a process in discrete time. The random walker $v(t)$ is a function from time to the vertex set, with the property that $v(t+1) \in \Gamma_{v(t)}$.

[iv] The logic behind not considering the out-neighbourhood is that if your personal homepage is referred to by some important Web site, say cnn.com, you must be important in some respect. But if you refer to cnn.com, it does not say much about your importance. A counter argument is that if a very important web page is only one click away from your web page, then your page is quite useful too. A way to include the out-neighbourhood is mentioned in sect. 4.4.

Then, provided $\mathbf{B}$ has an inverse $\mathbf{B}^{-1}$, the information centrality $C_I(v)$ is

$$C_I(v) = \frac{N}{NB^{-1}_{vv} + \mathrm{Tr}\,\mathbf{B}^{-1} - 2\sum_{w\in V} B^{-1}_{vw}} \tag{2.8}$$

(where Tr denotes trace). Very loosely (for details, see ref. 328), the idea behind information centrality is that if "signals" travel along the paths between vertices, then one can assume that the information is inversely proportional to the variance of the signal strength. The information centrality is then the harmonic mean of the signal strengths of paths between a vertex and the rest of the graph. Examples of eigenvector and information centralities can be found in fig. 2.1.

2.1.5 Local clustering coefficient

Acquaintance networks have the property that there are many triangles (three-circuits)—two of your friends are likely to know each other. To measure to what extent this holds for a specific vertex one can measure the local *clustering coefficient* c_v,[358] defined as:

$$c_v = \frac{|\Gamma_v|_E}{\binom{k_v}{2}} \tag{2.9}$$

where $|\Gamma_v|_E$ denotes the number of edges in maximal subgraph spanned by the neighbourhood Γ_v.

c_v can be an useful quantity in other than social networks: A high c_v tells us that v is nested in a highly connected cluster of the network (provided the network is sparse). In biological networks, such regions are often interpreted as relatively independently functioning subsystems (so called modules).

The fastest algorithm (based ref. 13) to calculate c_v runs in $O(M^{1.41})$ in combination with fast matrix multiplication[159] (or in $O(M^{1.5})$ with regular matrix multiplication). c_v of our small test graph is shown in fig. 2.1.

2.2 Vertex-pair relations, hierarchies and clustering schemes

This section concerns measures for how one vertex relates to the other vertices and how such relationships can be represented as partial orderings or partitionings of the vertex set.

2.2.1 Connection strength weights

Given a graph, directed or not, how strong is the relationship between the two vertices v and w? To give quantitative answers to such question one has to keep the particularities of the networked system in mind. One possible measure is the number of vertex- or edge-disjoint paths between v and w.[360] This weight is rapidly calculated by "preflow relabeling" or "augmenting path" algorithms.[6] Rather than counting only vertex- or edge-disjoint paths one can also include all paths between v and w weighted such that short paths contribute more to the weight $W(v, w)$. A way to realise this idea is to calculate the weight matrix as

$$\mathbf{W} = (\mathbf{I} - a\mathbf{A})^{-1} \tag{2.10}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{A}$ is the adjacency matrix and $a \in (0, 1/\lambda)$ (λ is the largest absolute value of $\mathbf{A}$'s eigenvalues) is a parameter controlling the decay of weight-contribution from longer paths.[185]

2.2.2 Structural equivalence

Structural equivalence measures are designed to quantify how similar the positions of two vertices in the network are. In the dichotomous sense, two vertices v and w are structurally equivalent if they have the edges[v] to the same vertices in $V' = V \setminus (\{v, w\})$.[222] One possible way to create a continuous measure S_t of structural equivalence is to measure the Euclidean distance between the in- and out-neighbourhoods:[72]

$$S_s(v, w) = \sqrt{\left|\Gamma_v^{\text{in}} \cap \Gamma_w^{\text{in}}\right|^2 + \left|\Gamma_v^{\text{out}} \cap \Gamma_w^{\text{out}}\right|^2}\,. \tag{2.11}$$

A normalised version $s_s \in [0, 1]$ of S_s is easily constructed:[vi]

$$s_s(v, w) = \frac{S_s(v, w)}{\sqrt{\min(k_v^{\text{in}}, k_w^{\text{in}})^2 + \min(k_v^{\text{out}}, k_w^{\text{out}})^2}}\,. \tag{2.12}$$

[v] In general one can consider systems with different types of relations. I.e. graphs with more than one edge set.

[vi] Structural equivalence has been used—or rather, reinvented—a few times in the context of biochemical networks: In ref. 144 the Manhattan distance equivalent of eq. 2.12 is used to construct improved inference methods for empirical biochemical networks. Ref. 289 uses roughly the same measure as ref. 144 under the name "topological overlap" (the difference is that the neighbourhoods Γ_v are replaced by $\Gamma_v \cup \{v\}$). Yet another very similar measure is found in ref. 174. In this paper, the Manhattan distance (in this case equivalent to the Hamming distance) between the two neighbourhoods Γ_v and Γ_u are normalised by dividing with $\sqrt{k_v k_w}$.

In our small example graph of fig. 2.1 the $s_s(v, w)$ matrix becomes:

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 1 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 1/2 & 1 \end{pmatrix} . \tag{2.13}$$

For valued graphs one can generalise eq. 2.11 as follows:

$$S_s(v, w) = \sqrt{\sum_{u \in V'} \left[(W_{uv} - W_{uw})^2 + (W_{vu} - W_{wu})^2 \right]} . \tag{2.14}$$

Another way to measure structural equivalence is by calculating Pearson's correlation coefficient between the rows and columns of the weight matrix.[351]

2.2.3 Regular equivalence

In general, it is not necessarily the case that two vertices have the same function just because their neighborhood overlaps to a great extent. A less specific assumption is two vertices are likely to have a similar role if the rest of the network (or at least the closest surroundings) looks similar from the perspectives of both vertices. This is the premise of regular equivalence measures. One possibility is to find vertices that are (exactly) regularly equivalent: For a directed valued graph $G = (V, A_1, \cdots, A_{n_r})$ (with n_r relations, or arc sets) a relation $\equiv$ is a regular equivalence if, and only if, for all $v, w, v' \in V$, $v \equiv w$ implies the following two clauses:[38]

1. $(v, w) \in A_r$ implies that there exists an $w' \in V$ such that $(v', w') \in A_r$ and $w' \equiv w$.

2. $(w, v) \in A_r$ implies that there exists an $w' \in V$ such that $(w', v') \in A_r$ and $w \equiv w'$.

It is often more interesting to find a continuous measure of how regularly equivalent a pair of vertices are. The basic algorithm for a continuous regular equivalence measure, White and Reitz's *REGE*,[361] is an iterative method that returns a value 1 for the pair v and w if $v \equiv w$, otherwise the output goes to zero as the iterations proceed. Let the arc sets be represented by the matrices $\mathbf{R}^r$. Let superscript n denote the nth iteration and let M^n_{vw} be the value of our

regular equivalence measure after n iterations. Then REGE is defined through the recursion formula

$$M_{vw}^{n+1} = \frac{\sum_{v' \in V} \max_{w' \in V} \sum_{r=1}^{N_r} M_{v'w'}^{n} ({}_{vw}\mathbf{M}_{v'w'}^{r} + {}_{wv}\mathbf{M}_{v'w'}^{r})}{\sum_{v' \in V} \max_{w'}^{*} \sum_{r=1}^{N_r} ({}_{vw}\mathbf{X}_{v'w'}^{r} + {}_{wv}\mathbf{X}_{v'w'}^{r})} \tag{2.15}$$

where

$${}_{vw}\mathbf{M}_{v'w'}^{r} = \min(R_{vv'}^{r}, R_{ww'}^{r}) + \min(R_{v'v}^{r}, R_{w'w}^{r}) \tag{2.16}$$

and

$${}_{vw}\mathbf{X}_{v'w'}^{r} = \max(R_{vv'}^{r}, R_{ww'}^{r}) + \max(R_{v'v}^{r}, R_{w'w}^{r}) \,, \tag{2.17}$$

and $\max_{w'}^{*}(X)$ denotes the X value for the w' value maximising the r-sum of the numerator. The numerator picks out the best matching counterparts for all vertices v' and w' in the in- or out-neighbourhoods of v and w. The denominator ensures that all $M_{vw}^{n} \in [0, 1]$. It equals he numerator if all ties between v and Γ_v can be matched to all ties between w and Γ_w. As the iterations proceed the relations of increasingly distant vertices affect M_{vw}^{n}. To use REGE one has to fix the number of iterations. It must not be too small, so not only the very closest vertices affects the measure; and not too large, so that not all values are either one or very close to zero. Following ref. 361, $s_r(v, w) = M_{vw}^{3}$ has become the standard measure of regular equivalence.

The applications of regular equivalence outside social network analysis are few. To my knowledge it has only been applied to food webs.[224,225] Maybe the running time of REGE, $O(nN_rN^4)$, limits the applicability.

2.2.4 Hierarchical clustering schemes

In sections 2.2.1, 2.2.2, and 2.2.3 we have presented measures of different types of vertex-pair relations. In a larger network one would often like to forsake some information to gain clarity. A way to do this is by constructing a hierarchical clustering tree, or *dendrogram* (see fig. 2.2). After weights $W(u, v)$ are assigned to each vertex pair (u, v) one applies a *hierarchical clustering scheme*.[181] This method starts from letting the all vertices being isolates. Then one iteratively merges vertices or vertex sets in order of increasing weight. If, to start with, v and w constitutes the vertex pair with highest weight, then they are merged and $\{v, w\}$ are from then on always treated as a single object. The vertical dimension of the dendrogram represents the weight between the two clusters merged at that particular level of the hierarchy. The problem is how to assign a value to the relation between $\{v, w\}$ and an arbitrary other vertex u. The two common methods (for obvious reasons termed the maximum and minimum methods) are to choose either $\min(W(v, u), W(w, u))$ or $\max(W(v, u), W(w, u))$. In most cases these methods produce similar results. For the maximum method the dendrogram is connected faster (more nodes are joined at a given W in fig. 2.2) than for the minimum method.

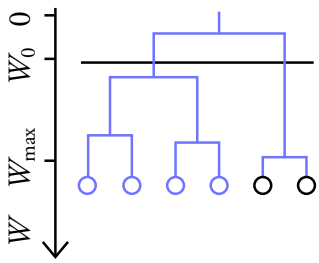


Figure 2.2: *Example of a dendrogram. The vertices are at the bottom. At level W_0 the grey and black vertices forms two clusters.*

If the objective is to decompose the graph into cohesive subgroups, one can use the vertex-pair weights given in sect. 2.2.1. If one uses the structural equivalence (sect. 2.2.2) or regular equivalence (sect. 2.2.3) vertices connected close to the base of the dendrogram have a high degree of role similarity (in the sense of these two different measures respectively).

2.2.5 Girvan and Newman's algorithm

Algorithms to trace cohesive subgroups based on hierarchical clustering schemes (sect. 2.2.4) along with vertex-pair connection weights (sect. 2.2.1) have a troublesome pathology. In many cases they do work reasonably well; but, for example, leaves are often isolated very high in the dendrogram (i.e. forming one-vertex subgroups for most hierarchical levels). The remedy, proposed by Girvan and Newman,[137] is based on the idea that one can construct a dendro-

gram without going through a hierarchical clustering scheme. Instead one can start from the whole network and iteratively delete edges and let the dendrogram reflect the disintegration of the network throughout the iterations. The edge to remove is the one with highest betweenness (see sect. 2.1.3). To sum up, the algorithm works by repeating the following steps until no edge remains.

1. Calculate the edge betweenness for the whole graph.

2. Delete the edge with highest betweenness.

3. Save information about the current network to construct the dendrogram.

With fast betweenness algorithms[62,259] this scheme runs in time $O(NM^2)$. The algorithm has been demonstrated to output a very reasonable decomposition of the networks into subnetworks that are more tightly connected within than to the outside. In our tiny example graph, fig. 2.1, $V' = \{v_1, v_2, v_3\}$ and $V'' = \{v_4, v_5, v_6\}$ would first be separated, then V' would split into isolates before V''. The drawbacks with this algorithm is that it is rather time consuming—more recent works have improved this.[286,255,368] Moreover, the vertical dimension of the dendrogram does not have as clear interpretation as the hierarchical clustering schemes (where it represents the vertex pair weight).

The dendrogram represents a partial ordering[vii] of subsets of V, a subnetwork hierarchy in the mathematical sense. Sometimes one likes to reduce the dendrogram to just a partition of V into subnetworks. This issue is discussed in refs. 286, 255, 38. If such a partition of V is supplemented by ties between the subnetworks one has a *blockmodel*[363]—a frequently used concept in social network analysis.

Summary of paper I

In paper I we adapt Girvan and Newman's algorithm to biochemical networks. The main idea is to represent the biochemical network as a directed two-mode network (see sect. 1.5.5) and deconstruct the set of substrates. Instead of removing edges with high edge-betweenness we remove reaction vertices of high

[vii] A binary relation $\sqsubseteq$ is a partial ordering on the set X if the following three conditions hold: 1. $x \sqsubseteq x$. 2. $x \sqsubseteq x'$ and $x' \sqsubseteq x$ if and only if $x' = x$. 3. If $x \sqsubseteq x'$ and $x' \sqsubseteq x''$ then $x \sqsubseteq x''$. (x, x' and x'' are all members of X.)

betweenness with respect to the substrates. Moreover, we also divide the betweenness by the in-degree and consider an "effective betweenness"

$$c_B(v) = \frac{C_B(v)}{k_{\text{in}}(v)} \tag{2.18}$$

of the reaction vertex v. The logic behind this is that if a reaction takes place, then there are conversions of all in-neighbours, so one path of mass-flow through a reaction vertex indicates the simultaneous existence of paths through the other in-neighbours. Note that $c_B(v)$ is the average edge-betweenness of the edges from Γ_v^{in} to v, so in this sense the efficient betweenness makes the algorithm more like the original. The upshot of this is that the metabolic networks of 43 different organisms from all domains of life are very similar and can be characterised by a core of highly connected vertices surrounded by increasingly weakly connected substrates. In addition there are a few well-defined subnetworks at all levels of the dendrograms representing the subnetwork hierarchy.[viii] Maybe this structure could be anticipated by the fairly power-law like degree distributions (see sect. 2.1.1) of biochemical networks. Nevertheless it tells us that biochemical networks do not share social networks' conspicuous community structure.[ix]

[viii] If the metabolic networks are extended with other reaction categories, such as signal transduction and other information processes, these additional functions often shows up as such small well-defined subnetworks. Thus our algorithm can, at least, single out the major functional parts of biochemical networks.

[ix] There is a consensus that biochemical networks—well, biological systems in general—are, to a substantial degree, "modular." I.e., that they can be decomposed into subunits that performs specific tasks, just like electronic machinery or computer programs. The picture from our paper seems to contradict this paradigm a little—so either the modules are not cohesive subgroups, or the modularity picture is wrong? Well, I would say that there is a little truth in both these claims. First of all, some substrates (like water) exist in so high concentrations that a module needs not to contain them to be independent. A more appropriate module-detection would maybe disregard the substrates of highest degree. (Indeed such decomposition schemes has been proposed.[307]) If the stoichiometry is taken into account the picture may also be much different than the graph structures we detect. Another issue is localisation—many substrates exist only at well-defined places (the nucleus, the mitochondria, etc.) in the cell. Nevertheless, the modularity paradigm is, in my eyes, a little exaggerated. When faced with a complex system such as metabolism, one either has to decompose it (whether it is naturally done or not), or rely on large scale simulations. Traditional biochemistry has chosen the former alternative and made modules such as the citric acid cycle or glycolysis so well-established that they adorn the walls of biochemists' laboratories throughout the planet. This may cause a conviction that the whole biochemistry really is modular, but few has really tried to prove this explicitly. The last few years have seen a new paradigm "systems biology" counterbalancing the modularity idea. Systems biology is the study of the interaction of many biological entities—well, modules. If

2.3 Graph quantities

This section concerns quantities that are functions of whole graphs.

2.3.1 Average geodesic length

A good measure of how tightly knit networks are is the average geodesic length l. Just as the name suggests this quantity is defined as

$$l = \frac{2}{N(N-1)} \sum_{v \in V} \sum_{w \in V \setminus \{v\}} d(v,w) \,, \tag{2.19}$$

where $d(v,w)$ is the distance between v and w (see sect. 1.4). As a network often is disconnected one rather studies the l of the largest connected component l_{GC}, or the sum of reciprocal distances[213]

$$l^{-1} = \frac{2}{N(N-1)} \sum_{v \in V} \sum_{w \in V \setminus \{v\}} \frac{1}{d(v,w)} \,. \tag{2.20}$$

Based on the average geodesic length one can define a dichotomy of graphs in terms of small-world and large-world networks: If a growing graph with a constant mean degree has an l that increases logarithmically, or slower, with N, then the network is said to be a *small-world* network, otherwise it is *large-world*.[265] The origin of this terminology comes from Milgram's "small world" experiments in the 1960s.[240,341,97] In these experiments people were instructed to send a folder with some documents to a person described by name, profession and place of living; but not to him directly, rather to the acquaintance (known on a first name basis) that were closest to the recipient (in geographical or social space). The outcome was that the average steps for the letters to reach their target, extrapolated to the population of the whole Earth, was about six, something that since ref. 150 is known colloquially as the "six degrees of separation." Milgram's result has recently been confirmed by an e-mail based study.[102]

2.3.2 Density of short circuits

The clustering coefficient (see sect. 2.1.5) measures the fraction (out of the maximal possible number) of triangles centered at a vertex. In this section we generalise this measure to the graph as a whole.

---

the interaction is strong, the modularity is weak; so systems biology is, in a way, insensitive to the organisation of the biological system.

In general sociologists define a triad (v_1, v_2, v_3) as *transitive* if whenever (v_1, v_2) and (v_2, v_3) are arcs (or edges) then (v_1, v_3) is also an arc (edge).[351] If every triad of a graph is transitive, then the graph itself is said transitive. In real world data one would rather like to measure the fraction of transitive triads of all connected triples. In the post Watts-Strogatz[358] literature this fraction

$$C = \frac{n_c(3)}{n_p(3)} \tag{2.21}$$

(where $n_c(l)$ is the number of representations of cycles of length l, and $n_p(l)$ is the number of representations of paths of length l[x]) is called the *clustering coefficient* of the graph, or sometimes the *transitivity* (both these terms have other meanings too, and are thus a little unfortunate). The fact the acquaintance networks have a bias towards transitivity (in any definition of the word) was noted early,[287] and has been regarded as one of the most important social structures since ref. 163. In the sociological literature much emphasis has been put on counts of the 16 possible different directed triads.[163,95,350xi] Studies of triads have also been conducted in social psychological experiments on coalition formation[77,126] and attitude change.[276] More recent studies have measured the clustering coefficient of many types of graphs, not only social, using eq. 2.21 or the local clustering coefficient (see eq. 2.9) averaged over the vertex set. Since the degree distribution has come to be regarded as the most basic structure[xii] many studies have focused on relating the clustering coefficient to the degree distribution. In general, a broader degree distribution will give a higher clustering coefficient; if the tail is a power-law with a slow enough decay C will tend to a positive value as $N \to \infty$[265] even if there is no explicit bias for triangle formation in the network forming process.

Fast algorithms are based on the same triangle counting as for the local clustering coefficient (see sect. 2.1.5).

2.3.3 Degree-degree correlations

If one can relate the quantities of a vertex with the quantities of its neighbours, one is close to the logic of the network forming forces. The simplest such corre-

[x] With representation I mean that the different ways of listing a path (or cycle) is to be taken into account, so one triangle have six representations and a specific path have two representations.

[xi] This kind of statistics has recently been used to analyse biochemical networks.[312]

[xii] Without much debate, I must add. Personally I believe that the most basic structure might be different for different networks.

lation is, naturally, the most studied—that of the degrees of two neighbouring vertices. This can be done in lesser or greater detail:

1. One can plot the two-dimensional histogram of degrees at either side of an edge[232,231,342]—the *correlation profile*.

2. One can plot the mean degree of the neighbours of a vertex v as a function of k_v.[280,347]

3. One can calculate the correlation coefficient r (often called *assortative mixing coefficient*[xiii]) of the degrees at either side of an edge:[261]

$$r = \frac{4\langle k_1\, k_2\rangle - \langle k_1 + k_2\rangle^2}{2\langle k_1^2 + k_2^2\rangle - \langle k_1 + k_2\rangle^2} \tag{2.22}$$

   where k_i is the degree of the i'th argument of the edges as they appear in an enumeration of hte edges.

These three methods correspond to three different dimensionalities: Method 1 yields a matrix, method 2 a vector and method 3 a number. What method one should select is much a matter of the quality of the data: Method 1 needs large sample sizes to get good precision, and error estimates are notoriously difficult to represent in three-dimensional plots. On the other hand method 1 gives the most detailed picture and could thus reveal informative biases of the network shaping process. The concise method 3 is less prone to errors from small data sets, while method 2 is a good compromise between precision and information content.

For directed networks the assortative mixing coefficient can be extended to correlation between in- and out-degrees in four combinations (see paper III). Directed edges lead to an asymmetric correlation profile for the undirected degree (the degrees of the reflexive closure of the network). This probably makes method 1 comparatively more useful for directed networks.

Summary of paper II

In this paper we develop two measures for *bipartivity*—how close to bipartite a graph is. For the first quantity, we assign a sign (+ or −) to every vertex and minimise the number of edges between vertices of the same sign. Our measure

[xiii] In sociology the term assortative mixing refers to the phenomenon that similar people tend to associate with each other. Yet the the term "assortative mixing coefficient" does commonly refer to degree-degree correlations in particular.

b_1 is then defined as the fraction of edges between vertices of the same sign. This problem is NP-complete, i.e. we need to find an approximation scheme in order to evaluate b_1 for large graphs. The approximation is performed by a so called simulated annealing scheme[197] that we develop specifically for this problem. The second quantity b_2 is intended to capture the same structure as b_1 but to be computationally tractable. It is based on counting odd-circuits (the graph elements inducing the disparity from a bipartite graph). We measure b_1 and b_2 on model networks with an intuitively tunable bipartivity and conclude that our measures are proportional to the bipartivity controlling model parameters. Then we measure b_1 and b_2 for a host of real-world networks. In many cases the measures produce sensible results, but (to our disappointment) they fail to discriminate an Internet-community network with more romantic interaction (see paper III) from an Internet-community network with more friendship oriented interaction.[316]

Summary of paper III

In this paper we present a longitudinal case-study of an Internet community. The community, pussokram.com, is Swedish and targeted at romantic interaction among adolescents. We study the time development of structural quantities such as the degree distribution (see sect. 2.1.1), clustering coefficient (see sect. 2.3.2), average geodesic lengths (see sect. 2.3.1), assortative mixing coefficients (see sect. 2.3.3) and more. Among other things, we find that the clustering coefficient is highly skewed, that there is no or a small negative preference for assortative mixing, and that the clustering is low.

Summary of paper IV

Normally, the edge set of social networks is constructed by accumulating the social ties over a period of time. In this paper I look at the structure of social networks as the cessation of ties is taken into account. I do this by first showing, for three different data sets, that the average contact rate of a relationship decays roughly exponentially, and thus that there exists a characteristic duration time of a relationship t_{dur}. Then I argue that if the sampling time of the network is larger than t_{dur} the network of ongoing—started and unfinished—relationships can (at a particular time) be approximated by the network of ties that have occurred and will occur again. For one of my data sets (indeed the one presented in paper III) t_{dur} is much smaller than the sampling time. For this data set I measure some structural quantities at time interval of the order

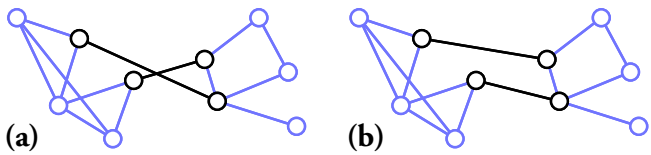


Figure 2.3: *Conditional uniform graph test with respect to degree sequences. In (a) two edges are chosen randomly (or one edge sequentially and one edge randomly). In (b) the edges are rewired.*

$t_{\rm dur}$, and note e.g. that the degree distribution of the approximated network of ongoing contacts is closer to a power-law than that of the network of accumulated contacts.

2.4 Conditional uniform graph tests

Imagine you have invented a measure of network structure and applied it to a real world network. Without something to compare the value to your effort would not be worth much. One possibility is to compare it to the corresponding value for maximally random graphs (see sect. 3.1). Often, however, one may argue that some other network structure is more basic than the newly invented one. In that case one rather averages the quantity over an ensemble of graphs conditioned to have the more basic network structure. This method, called *conditional uniform graph tests*, has ever since ref. 187 been well established in social network analysis. Examples of such more fundamental structures are degree-degree correlations,[103] degree distributions,[295,232,231,312] reciprocity.[164]

A general technique to construct sample networks conditionally uniformly is to make small random changes to the network in such a way that the basic network quantity is preserved. If one aims to preserve the reciprocity in a directed network one can pick an arc (v, w) and a vertex w' (such that $v \neq w'$ and $(v, w') \notin A$) replace (v, w) by (v, w'). If one wants to conduct a conditional uniform graph test with respect to the degree sequence one can (as in papers III and XV) perform the following:

1. Construct the set E' of edges that would, if rewired, gives rise to loops or multiple edges or cause no change: It is easy to convince oneself that E' is the union of $\{(v, u) : u \in \Gamma_v\}$, $\{(u, u') : u \in \Gamma_v\}$, $\{(u, w) : u \in \Gamma_w\}$ and $\{(u, u') : u \in \Gamma_w\}$. An edge is an unordered pair of vertices, but

here the order of the vertex pairs is important (we consider a particular representation of the edges).

2. Choose randomly an edge $(v', w') \in E \setminus E'$.

3. Replace (v, w) and (v', w') by (v, w') and (v', w).

This rewiring procedure is illustrated in fig. 2.4. The idea of constructing the set E' in step 1 is that a rewiring is guaranteed in (almost) every step even for dense networks. An alternative way would be to pick two edges at each time step and not rewire these if loops or multiple edges would be introduced. Ref. 295 gives another algorithm to sample graphs with given degree sequences, suited for matrix representation of the network.

Another option (than using Monte Carlo sampling) is to use recursive combinatorial formulae to explicitly calculate the number of graphs with a specific degree sequence.[186,317]

chapter 3

Network models

"Model" is an almost omnipresent concept in the natural and social sciences and a word frequent as few other in this thesis. Unfortunately, my usage varies slightly with the context: Sometimes I talk about network models as definitions of a probability spaces, but most often I mean a scheme for generating networks. When a model, in the second sense, is to be constructed, there are (I think) two sensible ways to go ahead: Either one looks for the minimal model generating one (or many) specific network structure, or one bases the model on measurements or observation-founded assumptions.

The traditional way of statistical physicists (and much of the natural sciences) is to hunt minimal models. The benefit of these is that they present one single candidate for the mechanism generating the structure, and point other events of the real world situation out as irrelevant. For example, in the study of social insects (see sect. 1.3) one might ask what is required from an individual ant's behaviour for the whole colony to forage efficiently. For networks one can ask what the minimal requirements are for a certain degree distribution to emerge in a growing network.

The models of statistical mechanics are most often designed to study phase transitions—the phenomenon that a system, at a critical value of a model parameter, changes the behaviour qualitatively. It turns out that very different models can behave similarly in the vicinity of a critical point (in parameter space)—a phenomenon is called "universality." Quite naturally, structural phase transitions and universality have been studied many times in physicists' network literature, a few examples being refs. 52, 233, 124, 74. There are also works on structural phase transitions in early mathematical literature. The most celebrated example being the emergence of a "giant component" as M

grows beyond $N/2$ in a random graph.[321,116] (I will discuss this further in sect. 3.1.) Since I have not dealt with such problems myself, and since there are few (if any?) real-world examples of structural phase transitions through changes in environment variables,[i] I will not dwell much longer on this topic. However, structural changes can occur due to other effects than shifts in model-parameter values: In sect. 3.5.1 we will discuss the network resilience to attacks (i.e. to what extent the network is affected by deletion of nodes and edges). Section 3.5.2 concerns another type of structural changes—cascading failures and overload breakdowns.

The other way of modelling networks mentioned above, to build the model on observations and measurements, is quite rare in the literature at the time of writing. Sometimes simple models are retrospectively backed by empirical evidence,[257,34] but most often it is the structure itself, rather than the observed network dynamics, that is the incentive for the modelling. This type of modelling is, however, existent in other areas. Two of the most extreme examples (mentioned in sect. 1.3) are projects such as e-cell platform for simulation of cellular metabolism and the TRANSIM (see the URL: transims.tsasa.lanl.gov) project. This is a possible future approach for modelling networks.

In this section I will present a dozen simple and generic network models, and discuss applications of these. An alternative outline would perhaps be to present the different networks first followed by the corresponding models of their formation.

3.1 Random graphs

Consider the following simple model of random networks: Start with N vertices. For each pair of distinct vertices, add an edge with probability p. This model, that we (following ref. 176) denote $\mathbb{G}(N, p)$, was proposed by Solomonoff and Rapoport[321] and later reinvented by Erdős and Rényi.[115] Erdős and Rényi also proposed another model $\mathbb{G}(N, M)$: Start with N vertices. Add iteratively, randomly, M edges such that no loops or multiple edges are introduced. (Fans of statistical mechanics may notice the exact analogy between the two models $\mathbb{G}(N, M)$ and $\mathbb{G}(N, p)$ on the one side, and the canonical and grand canonical ensemble[170] on the other.[265]) We will call members of the ensembles $\mathbb{G}(N, p)$ and $\mathbb{G}(N, M)$ *random graphs*. All properties of these models that we will need

[i] This is not at all to say that such studies are worthless. Actually, the mentioned emergence of a "giant component" is a concept that can be used in percolation on networks (see sect. 3.5.1), and further brought into epidemiology (see sect. 4.1). That is an example of a phase transition in a dynamical system on the network.

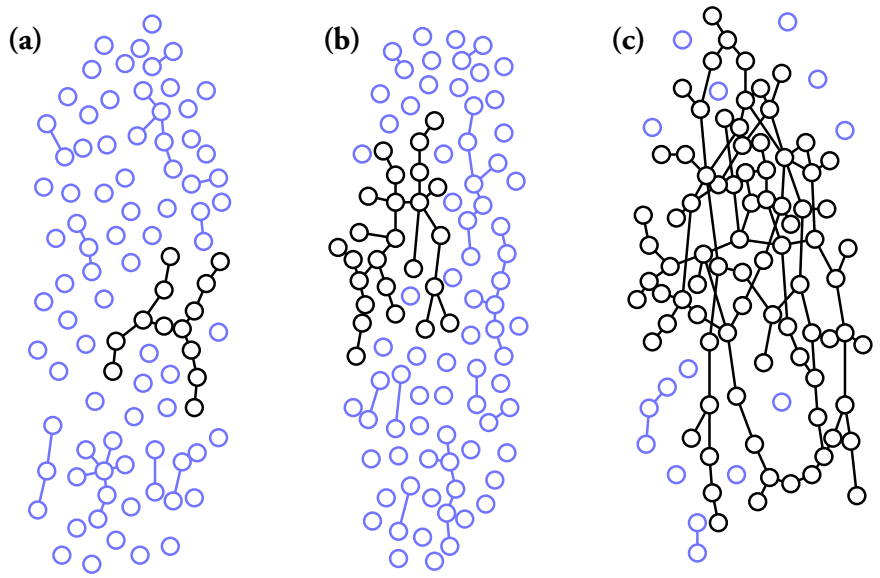


Figure 3.1: *Three realizations of the random graph model* $\mathbb{G}(N, M)$ *where* $N = 100$. *(a) is in the subcritical phase* ($M = 40$)*, (b) is at the jump* ($M = 50$)*, (c) is in the supercritical phase* ($M = 100$) *where a giant component has formed. The largest component is marked black, smaller components are grey.*

are qualitatively the same, so by "random graphs" I refer to both models unless I state it explicitly.

For a newcomer to random graph models it is a good exercise to write a program producing random graphs, and look at these for different parameter values. The most conspicuous feature in such an experiment is that, when increasing the number of edges (and keeping N fixed), the graphs all of a sudden goes from consisting of small isolated clusters, to a state dominated by one large connected component (see fig. 3). In more mathematical terms (the early references being ref. 321 and ref. 116) there are three regions with different behaviours:[176]

The subcritical phase In this region—occurring when M is of the order $N/2 - s$ where $N^{2/3} \ll s \ll N$—the expectation value of the size of the largest component S_1 is $O(\log N)$. The connected components are unlikely to contain more than one circuit.

The jump For $M \in O(N/2 \pm s)$, $0 \leqslant s \ll N^{2/3}$ the expectation value of

S_1 is $O(N^{2/3})$ and the graph will most likely contain more than one component containing more than one circuit.

The supercritical phase Here the expectation value for S_1 is of the order of N, i.e. almost all vertices will be connected into a single *giant component*. Except the giant component, the graph looks like in the subcritical phase, with small components of one or no circuits.

Note that in the $N \to \infty$ limit, the jump occurs precisely at $M = N/2$.

Real-world networks have very often one component much larger than any other, and a few disconnected vertices. In such cases, even if one does not know how S_1 scales with N, one commonly speaks of the largest connected component as the "giant component." For this reason, when random graphs are used as models for real-world networks, one use to sample the supercritical phase.[ii] The problem with using random graphs as model networks is that it lacks the structures of real-world networks discussed in Chapter 2, and the structures thought to be the key for understanding the behaviour of the dynamical systems which exist on networks. In fact, the random graph models are often used as a zero-gauge against which the structural biases are defined. Even objectively speaking, one can say that random graphs in many respects lack structure: Two examples are that the clustering and assortative mixing coefficients tend to zero in the $N \to \infty$ limit.[261] Another example is the degree distribution that, in the large-N limit, has a smooth Poissonian form:[10,105,265]

$$p(k) \approx \exp(-\langle k \rangle) \frac{\langle k \rangle^k}{k!} , \tag{3.1}$$

where $\langle k \rangle$ denotes the average degree

$$\langle k \rangle = \begin{cases} 2M/N & \text{for } \mathbb{G}(N, M) \\ p(N-1) & \text{for } \mathbb{G}(N, p) \end{cases} . \tag{3.2}$$

To conclude, random graphs are as random as graphs can be—agents forming a network without any preferences at all would form a random graph. Thus random graphs are useful as reference points against which structure can be defined, but not very good as model network for structured real-world networks.

[ii] Or rather in the "early supercritical phase," i.e. when M is bounded by a function linear in N.

3.2 Analytical tools for network models

We take a short break in our exhibition of network models and present two useful tools to analytically derive network properties.

3.2.1 Generating functions

The generating function formalism is a general technique of probability theory that can be very helpful for some certain types of network models.[366][iii] Given a probability distribution, like the degree distribution $p(k)$, the generating function is

$$g_0(x) = \sum_k p(k)x^k \,, \tag{3.3}$$

where $-1 \leqslant x \leqslant 1$. Since

$$p(k) = \frac{1}{k!}\left.\frac{d^k g_0}{dx^k}\right|_{x=0} \tag{3.4}$$

all the information of the probability distribution is encoded in the generating function. Furthermore, the moments of k are easily calculated:

$$\langle k^n \rangle = \left[\left(x\frac{d}{dx}\right)^n g_0(x)\right]_{x=1} \,. \tag{3.5}$$

Another important property is that if the probability distribution concerns certain values associated to the members of a set—like the degrees of vertices—then the sum of the property of m uniformly randomly picked such items is generated by $[g_0(x)]^m$.

To give an example of how generating functions work, we consider the random graphs of vertices with degree one or three and ask what the fraction f of degree-three vertices the graph must have to form a giant component. To do this we need to know that the average number $\bar{k}_2$ of next-nearest neighbours[iv] is[264]

$$\bar{k}_2 = \langle k^2 \rangle - \langle k \rangle \tag{3.6}$$

and that this formula can be iterated so that the number $\bar{k}_l$ of vertices at distance l is

$$\bar{k}_l = \left(\frac{\bar{k}_2}{\bar{k}}\right)^{l-1} \bar{k} \,, \tag{3.7}$$

[iii] This section, the example aside, is based on refs. 264, 265, 268.

[iv] In a graph with vanishing clustering in the $N \to \infty$ limit.

where $\bar{k} = 2M/N$ is the average degree. For this formula we see that if the ratio $\bar{k}_2/\bar{k}$ is larger than one, the number of neighbours at distance l will grow exponentially as $l \to \infty$, but if $\bar{k}_2/\bar{k} < 1$ the number of vertices at distance l goes to zero.[243] This means that the requirement for a graph (with a vanishing number of short cycles) to have a giant component is that $\bar{k}_2 > \bar{k}$. So for our class of graphs, the generating function is

$$g_0(x) = (1-f)x + fx^3 \,. \tag{3.8}$$

The average degree is (of course)

$$\bar{k} = \langle k \rangle = (1-f) + 3f = 1 + 2f \,. \tag{3.9}$$

Using equation 3.5 we get

$$\langle k^2 \rangle = 1 + 8f \tag{3.10}$$

and using eq. 3.6 we get

$$\bar{k}_2 = (1 + 8f) - (1 + 2f) = 6f \,. \tag{3.11}$$

So, from equations 3.9 and 3.11 we see that if

$$\frac{\bar{k}_2}{\bar{k}} = \frac{6f}{1+2f} > 1 \;\Leftrightarrow\; f > \frac{1}{4} \tag{3.12}$$

the graph has a giant component. We note that, at the critical f-value the average degree is 3/2 which is higher than the corresponding value ($\bar{k} = 1$) for the random graphs of sect. 3.1.

3.2.2 Master equations

Constructing *master equations* (or *rate equations*) is a powerful technique to calculate properties such as degree-distribution, degree-degree correlations and so on, for dynamic network models. The general idea is to express the change of a quantity per iteration of the network construction scheme. Then, if one puts this change equal to the average change,[v] one can extract the expression for the quantity in question.

[v] I.e. if the the sizes, N and M, are fixed, then the change of structural quantities should, on average, be zero. If M grows $2\bar{k}N$ one still demands the average value of the quantity to be constant, but since the master equation does not involve mean values one has to calculate the condition for stationarity explicitly. An example of this will follow.

Imagine a network model as follows: Start with m_0 isolates, and for each iteration add one vertex v and m edges between v and vertices present in the graph with a probability

$$P_w = \frac{k_w + 1}{\sum_{v \in V}(k_v + 1)} . \tag{3.13}$$

where w is the new vertex to attach to. (This a version[9] of the so called Barabási-Albert model that will be discussed further in sect. 3.4.1.) Now we will derive an expression for the degree distribution using master equations.[vi] Let k be a number bigger than m and denote the number of vertices of degree k by N_k. During the addition of an edge, N_k increases if a vertex of degree $k-1$ is attached to, and decreases if a vertex of degree k is attached to. Using eq. 3.13 we can write down the change in N_k:

$$\begin{aligned} \Delta N_k &= \frac{mkN_{k-1}}{\sum_{v \in V}(k_v + 1)} - \frac{m(k+1)N_k}{\sum_{v \in V}(k_v + 1)} \\ &= m\frac{kp(k-1) - (k+1)p(k)}{2m+1} , \end{aligned} \tag{3.14}$$

where we use that $\sum_{v \in V} k_v = 2M$. Moreover, we require the degree distribution to be stationary, i.e. we need (on average) $p^t(k) = p^{t+1}(k)$ for all k and for every iteration t. Since N grows with one each iteration, we get (after a little algebra) the condition

$$\Delta N_k = p(k) . \tag{3.15}$$

Now combining eq. 3.15 with eq. 3.14 gives:

$$\begin{aligned} mkp(k-1) - m(k+1)p(k) &= (2m+1)p(k) \;\Rightarrow \\ mkp(k-1) &= (3m + 1 + mk)p(k) . \end{aligned} \tag{3.16}$$

Next step is to approximate k by a continuous variable x and $p(k) - p(k-1)$ by p's derivative p'. We get (after some more algebra):

$$xp' = -(3 + 1/m)p \tag{3.17}$$

which has a solution of the form $p(x) = ax^{-(3+1/m)}$ where a is a normalising constant. The degree distribution is thus a power-law with exponent $-3 - 1/m$.

[vi] This is done for generalisations of this model in refs. 206 and 106.

3.3 Static models

In this section we discuss network models intended to sample random networks with some structure, but without trying to imitate some network forming dynamics.

3.3.1 The configuration model

The configuration model[45] is simple generalisation of random graphs to sample graphs a given degree sequence: Start with N disconnected vertices and assign each vertex a number of "half-edges" or "spokes" corresponding to the degree sequence; then pair the half-edges together randomly to form edges (see fig. 3.2). The important feature of this method is that it samples the ensemble of isomorphically distinct multigraphs with a given degree sequence uniformly.[243,244]

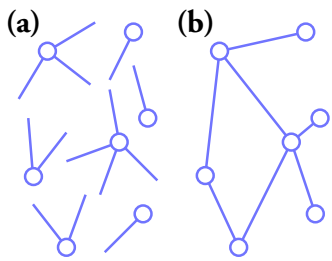


Figure 3.2: *A realization of the configuration model for the degree sequence* {3, 1, 1, 4, 2, 1, 2}. *(a) shows the vertices and half-edges. (b) shows the resulting graph.*

Now let $p(k)$ be the degree distribution from which the degree sequence is taken (i.e. the expectation value of k is $Np(k)$). Just as for random graphs the configuration model has a phase transition between phases with and without giant components. The condition is that a giant component exists when the sum

$$\sum_k k(k-2)p(k) \tag{3.18}$$

is finite.[265]

As mentioned in sect. 2.1.1 many real-world networks have a stretched degree-distribution, in several cases it is close to a power-law. If the degree distribution of the configuration model is chosen as a power law

$$p(k) = \begin{cases} 0 & \text{for } k = 0 \\ k^{-\alpha}/\zeta(\alpha) & \text{for } k > 0 \end{cases} \tag{3.19}$$

(here $\zeta(\alpha)$ is the Riemann ζ-function included as a normalising factor); then a giant component exists for $\alpha < \alpha_c$, where α_c is the solution to

$$\zeta(\alpha-2) = 2\zeta(\alpha-1) \;\Rightarrow\; \alpha_c \approx 3.48 \tag{3.20}$$

(see ref. 7).

The drawback with the configuration model is that, as formulated above, it can generate loops and multiple edges, i.e. it produces multigraphs and not graphs. This happens very seldom though, so for sparse and moderately sized graphs one can iterate the algorithm until one gets a graph without loops or multiple edges.

3.3.2 Exponential random graphs

The traditional models of statistical mechanics interpolate between order and disorder through the (real or abstract) inverse temperature β: A more ordered configuration is associated with a lower energy ϵ; a configuration of energy ϵ occurs with a probability proportional to $\exp(-\beta\epsilon)$. Thus, the lower the temperature (the higher the β) the more likely it is that the system is in a low energy configuration. As you might have already guessed, this framework has been utilised by physicists to interpolate between more random and more ordered networks, see refs. 46, 71, 233, 124. These studies are more or less similar to *exponential random graphs* originally proposed by statisticians[165, 128, 330] (and, ironically enough, not physicists).

Now we turn to the definition of exponential random graphs. Let $\{\epsilon_i\}$ be a set of measurable quantities of the graph (e.g. the number of triangles), and let $\{\beta_i\}$ be a set of inverse temperatures. Then the exponential random graphs are the set of all graphs of N vertices such that the graph G occurs with the probability

$$P(G) = \frac{1}{Z}\exp\left(-\sum_i \beta_i \epsilon_i\right), \tag{3.21}$$

where the normalising constant Z (the "partition function") is

$$Z = \sum_G \exp\left(-\sum_i \beta_i \epsilon_i\right). \tag{3.22}$$

Given a set of inverse temperatures the average values of graph quantities are simply

$$\langle \epsilon_i \rangle = -\frac{\partial}{\partial \beta_i}\log Z = \frac{\partial f}{\partial \beta_i}. \tag{3.23}$$

The "reduced free energy" f thus serves as a generating function (see sect. 3.2.1) for the distribution of the ϵ_i quantities. The next step would be to construct a

Gaussian field theory for f and a diagrammatic perturbation theory. The Feynman diagrams in such an expansion will in fact be isomorphic to the networks themselves.[265] However, this is too complicated to be really rewarding—an intractability of exponential random graphs is to some extent compensated by the fact that they are easy to generate by a Markov chain Monte Carlo scheme as follows: Define a method to reconfigure a network so that, if this method is applied repeatedly, any graph (of N vertices) could be generated from any other. Then iterate the method, and after each generation of a modified (trial) network G' from an old network G, if $P(G) < P(G')$, or if $P(G) > P(G')$ with probability $P(G')/P(G)$, replace G by G'.

In social network analysis one often uses a variant of exponential random graphs, p^*-models, to extract the structural biases of a network.[14,352,318] The model parameters are structural quantities that can be extracted by maximum likelihood estimation.

3.3.3 Hidden variable models

So called *hidden variable models* work by assigning a trait-vector (or number) to each vertex and then link vertices that fulfil some condition dependent on the trait-vector. Variants of this kind of modelling can be found in refs. 141, 194, 73, 252. As an example I present a special case of the model of Caldarelli *et al.*:[73]

1. Assign real numbers $\{x_v : v \in V\}$ to the (isolated) vertices. The x_v's can for example have an exponentially distributed probability density function.

2. Add an edge between vertices v and w with a probability $\Theta(x_v + x_w - U)$ where $\Theta(\cdot)$ is the Heaviside function and U is a constant.

This algorithm generates network with a power-law degree distribution with an exponent ~ -2.

3.3.4 Small-world network models

With exponential random graphs of sect. 3.3.2 one can obtain networks with logarithmically increasing path-length and an arbitrary clustering. A simpler method to get the same result is the Watts-Strogatz small world network model (see refs. 358, 354, 353) (almost equivalent models were independently proposed in refs. 28, 56, see also ref. 85).[vii] One can motivate it as model of

[vii] The name "small world" refers to Milgram's experiment. (See sect. 2.3.1.)

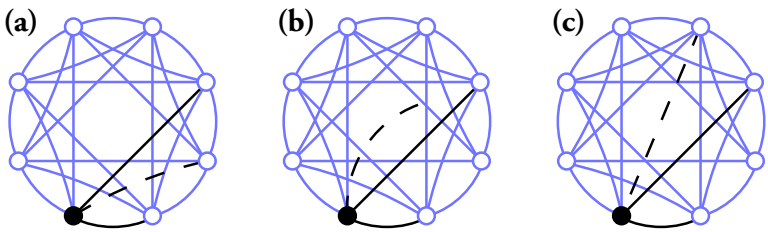


Figure 3.3: *The construction of a Watts-Strogatz model network. One starts from a regular one-dimensional lattice (a). For every vertex (consider the black vertex specifically) one goes through the edges on one side (for the black vertex these are black). Then, with probability P one detaches the other end (b) and re-attach it (c) with the condition that no loops or multiple edges must be formed.*

acquaintance networks by the assumption that a person's set of friends consists of a number of close friends who all know each other (presumably people the person meets frequently) and maybe also a few persons that these close friends do not know (presumably some geographically distant person). As a mathematical model this can be reproduced as follows: A regular one-dimensional network (a *circulant*) is first constructed by arranging N vertices in a circularly fashion and linking each vertex to $2k$ neighbours. Then one goes through each edge one at a time, and with the rewiring probability P detaches the far side of the edge and reconnect it to a randomly chosen other vertex (with the restriction that loops and multiple edges must not be formed). This procedure is illustrated fig. 3.3.

3.4 Dynamic models

3.4.1 The de Solla Price and Barabási-Albert models

Right-skewed degree distributions with a high-degree tail that decreases slower than exponential are abundant in real-world networks. A popular explanation for the emergence of such degree distributions is that vertices with high degrees get new edges at a higher rate than low-degree vertices. That such processes can give rise to power-law distributions is known from Simon's work on wealth distributions.[313] In this section we describe two very similar network models of this kind.

The model of de Solla Price[99] was intended to describe the emergence of power-law degree distributions of citation networks that he discovered a decade earlier.[98] A citation network is naturally described as a directed network, where

an arc (v, w) represents that paper v cites paper w. The out-degree of a paper is the number of other papers it cites and is fixed at its creation; the in-degree, however, is strictly increasing in time. In de Solla Price's model, the network is grown from a set of isolated vertices. The out-degree of a newly added vertex has the expectation value m in some distribution, and the probability an incoming link is attached to a particular vertex w of in-degree k_w^{in} is proportional to

$$P_w = \frac{k_w^{\text{in}} + k_0}{\sum_{v \in V}(k_v^{\text{in}} + k_0)} . \tag{3.24}$$

This model gives an in-degree distribution of[265]

$$p(k_{\text{in}}) \sim k_{\text{in}}^{-(2+1/m)} . \tag{3.25}$$

with natural average degree values this model also produces the exponents of real-world networks with degree distributions close to power-laws.[10,105] de Solla Price uses $k_0 = 1$ in his examples and motivates the choice from that the paper itself can be seen as its first citation (and can also be justified *a posteriori* by the fact that the degree distribution is not dependent on the choice of k_0).

To explain the apparent power-law degree distribution of the World Wide Web Barabási and Albert (BA) introduced a model very similar to that of de Solla Price.[32] The work of Barabási and Albert has been the source of inspiration for a huge number of network papers (something de Solla Price's model failed to be). The BA model differs on two points from the de Solla Price model—it considers the edges to be undirected and it fixes the number of edges added to each new vertex to a constant, m. More precisely BA model networks are constructed as follows: Start with m_0 isolated vertices. Grow the network by adding one vertex at a time ($N - m_0$ times). At each time step, add m edges between the new vertex and old vertices. The cumulative advantage rule of de Solla Price's model is called the "preferential attachment" rule in context of the BA model: An old vertex w is attached to with a probability[viii]

$$P_w = \frac{k_w}{\sum_{v \in V} k_v} , \tag{3.26}$$

[viii] There are various ways to implement the preferential attachment rule in a computer program. In my program I keep a list L of length $2M$ (if one use the definition eq. 3.26) or $2M + N$ (if one use the definition eq. 3.13) of vertex labels. As an edge (v, u) is added to the network I add v and u to the L. (If one follows the definition eq. 3.13 one also has to add every vertex as it enters the network. When one attaches a new edge from a new vertex v one chooses one of the $2M - \sum_{w \in \Gamma_v} k_w$ (or $2M + N - \sum_{w \in \Gamma_v}(k_w + 1)$ if one follows eq. 3.13) vertices not a member of Γ_v.

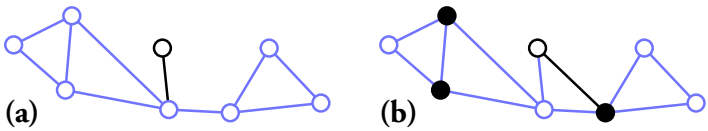


Figure 3.4: *The model of paper V. (a) shows a PA step where the new (black) vertex selects a neighbour with probability proportional to its degree. The attachment in (b) is according to the triangle formation rule where the new vertex links to a vertex in the neighbourhood (solid black vertices) of the vertex attached to in the previous PA step.*

or, if one allows isolated vertices one may use the same offset $k_0 = 1$ as in de Solla Price's model which gives equation 3.13.

The BA model generates networks with power-law degree distribution with exponent -3 (sect. 2.1.1), an assortative mixing coefficient (sect. 2.3.3) that tends to zero from below, and a clustering (sect. 2.3.2) that slowly converges to zero. The BA model is not the only growth-model that generates power-law degree distributions; it has some features that real-world systems hardly share. For example there will be a strong correlation between age and degree, and furthermore only linear preferential attachment will give a power-law degree distribution. Other dynamic mechanisms generating power-law like degree distributions can be found in refs. 299, 192 and sect. 3.4.2.

Summary of paper V

In this paper we present a network model that generates a power-law degree distribution and a non-vanishing, tunable, clustering coefficient.[ix] The model is similar to the BA model (see sect. 3.4.1) except that we, after an initial preferential attachment (PA) step, with probability $P_t = m_t/(m-1)$ replace preferential attachment by a triangle formation step: The new edge is linked to one of the neighbours of the vertex attached to in the previous PA step. (See fig. 3.4.1.) An alternative way to describe the model is that it interpolates between the BA model and the following very simple model by Dorogovtsev, Mendes and Samukhin (DMS):[107]

1. Start with two vertices connected by one edge.
2. Pick an edge (v, w).

[ix] This was actually precisely our goal when we started the project, which is something unique—normally my projects tend to shift focus enroute.

3. Add a new vertex with one new edge leading to v and one to w.
4. Go to Step 2.

This procedure generates graphs with power-law degree distributions with an exponent -3. One can reformulate the BA model for the case $m_0 = m = 2$ in a very similar way:

1. Start with three vertices connected by two edges.
2. Pick two edges (v, w) and (v', w').
3. Add a new vertex with one new edge leading to v or w and one to v' or w'.
4. Go to Step 2.

This is so because a vertex v occurs in k_v number of edges, so selecting an edge and attaching with equal probability to one of its vertices gives, effectively, preferential attachment.[x] Now our algorithm interpolates between the DMS and BA models by executing Step 2 and 3 of the DMS model with a certain probability p_t and Step 2 and 3 of the BA model with probability $1 - p_t$.[xi]

Paper V contains mostly simulation statistics, but it does not mean that the model is completely analytically intractable. An analytic expression for the expected value of the local clustering coefficient of a k-degree vertex was derived by Szabó, Alava and Kertész:[334]

$$c_k \approx \frac{4m_t}{k} + \frac{m-1}{8}\frac{(\ln N)^2}{N} \,. \tag{3.27}$$

The first term comes from the triangle formation steps, the second from the PA steps. For few triangle formation steps, there will be a clear cross-over between the behaviour for low degree vertices (where the first term dominates) and high degree vertices (where the second term dominates). This crossover happens around degree[334]

$$k_{\text{cross over}} \approx \frac{32}{m(\ln N)^2} p_t N \,. \tag{3.28}$$

A few comments and errata: Note that we measure clustering by averaging c_v (see eq. 2.9) over V rather than using C defined in eq. 2.21. Since dividing

[x] This defines another efficient way for computer implementations of the BA model.

[xi] Note that this is only exactly true for $m = m_0 = 2$ as one can define ours and the BA model for other average degrees.

by $k(k-1)/2$ and taking the average over all vertices do not commute, the two measures give different results. Clearly C is more intuitive as a measure for the clustering of a whole graph; since, in effect, low-degree vertices contribute more to the averaged c_v than vertices of higher degree do. An error in the paper is our claim that the average geodesic length scales as $\log N$. This scaling is known to be $\log N/\log\log N$ for power-law graphs with $\gamma = 3$.[57] In fact one can see a slight deviation from the exponential form in fig. 4 of the paper. Note also that the term "triad" for a clique of three vertices is unfortunate (see sect. 1.4).

Summary of paper VI

This paper presents a network version of an agent based model[xii] of group formation—the "seceder model."[100,324] The idea behind the seceder model is that people wants to be different than the average, but not to the price of being completely different (and regarded as eccentric by others).[184] What I personally have in mind is the search for an identity among adolescents and its relation to the formation of youth subcultures (cf. the takenoko-zoku described in sect. 1.3) To investigate the community structure we use the algorithm of ref. 255, and we find that the networked seceder model indeed produces networks with a very pronounced community structure.

3.4.2 The FKP model

To give an example of a model specifically targeted at explaining the power-law degree distribution of the Internet we consider the following model by Fabrikant, Koutsoupias and Papadimitrou (FKP):[117] Let $\mathbf{r}_1, \cdots, \mathbf{r}_N$ be points in the unit square (which will be the vertices of our network). Let T_1 be the graph consisting solely of $\mathbf{r}_1$ and consider a sequence of trees $T_1, \cdots, T_N$ defined by the recursion relation

$$V_{i+1} = V_n \cup \{\mathbf{r}_{i+1}\} \tag{3.29}$$

$$E_{i+1} = E_n \cup \{(\mathbf{r}_{i+1}, \mathbf{r}_j)\} \tag{3.30}$$

where V_i (E_i) is the vertex (edge) set of T_i and $\mathbf{r}_j$ is the point that minimises

$$\alpha|\mathbf{r}_{i+1} - \mathbf{r}_j| + d(\mathbf{r}_j, \mathbf{r}_1)\ . \tag{3.31}$$

[xii] I.e., a model with many entities (so called *agents*) that individually respond to events in the environment. Two (of many) readable references are 204 and 279.

The vertices represent servers with a specific location in space. A new vertex tries to get close to the centre $\mathbf{r}_1$ in graph distance and tries simultaneously to minimise the length of new cables (represented by the Euclidean distance to the selected neighbor). α is a parameter that sets the balance between the two competing objectives for the optimisation.

This model is simple enough to be treated analytically. For small and constant α the degree distribution is a star (i.e., $E = \{(\mathbf{r}_i, \mathbf{r}_1) : 2 \leqslant i \leqslant N\}$). For large α (if α grows faster than $\sqrt{N}\log N$) the degree distribution is exponential. For small but more than logarithmically increasing α ($(\log N)^4 \leqslant \alpha \leqslant \sqrt{N}/(\log N)^4$) the $N^{1/6}$ highest-degree vertices will follow a power-law degree distribution.[47]

3.4.3 Vertex copying models

Another mechanism producing networks with power-law degree distributions is given by so called vertex copying models. These model networks like the World Wide Web (see sect. 1.5.9) and biochemical networks (see sect. 1.5.5), where a plausible growth mechanism can be that a new vertex form edges to the same neighborhood as an old vertex (i.e. copies the neighborhood).[211,201] To be specific, the growth mechanism is as follows: At a time step, m edges are added to a randomly chosen vertex v: One iteratively link all neighbourhood vertices of randomly selected vertices until m more edges are linked to v. The probability to attach to a specific vertex v by this procedure equals the fraction of edges containing v; in other words it is proportional to v's degree. Thus this procedure leads to an effective preferential attachment (see sect. 3.4.1), and therefore a power-law degree distribution. A number of more specific models of the evolution of protein interaction networks based on vertex copying have been proposed.[320,195,346]

3.5 Models of structure-changing events

This section deals with models of events that changes the structure of a network. Such changes can be the result effects from dynamical systems confined to the network or outer events inflicted on vertices or edges. As the models of this section reshape the network they go under Chapter 3 and not 4.

3.5.1 Percolation, random failures and network attacks

This section expounds on a very basic question: What happens with the network if we remove vertices or edges? If one deletes vertices or edges uniformly

randomly then one copes with the *percolation* problem.

One major application of percolation studies is to the robustness of communication networks and biological networks.[75] Throughout the world, Internet servers crash every once in a while, but the Internet (or the World Wide Web) as a whole is not badly affected of a crash.[11,66,86] Systems of ecology[319] and cellular biochemistry[179,178] have the same feature. This can be explained by a high *percolation threshold*—a major part of the vertices (or edges) has to be removed to break the network into pieces. More formally one defines the percolation threshold as the fraction p_c of vertices (or edges) that has to be removed for the system to lose its giant component.[xiii] The percolation problem has an important role in epidemiology—if pN vertices are removed, the connected clusters correspond to the potential outbreak sizes of a disease in a population where a fraction p are immune.[247]

A problem akin to percolation is that of *attack vulnerability* of networks. In this case one looks at how the network responds to removal of vertices (or edges) performed to inflict the worst possible damage. There is no prevalent way to quantify "damage." People have looked at the size of the largest connected component, the average geodesic length,[11] or the fraction of vertices (or edges) that has to be removed to qualitatively alter the degree distribution.[87] Consequently, there is no unique way to damage the network (different strategies would be required to maximise different damage measures). Most frequently, people have studied deletion of vertices based on their initial degree[11] or their subsequent degree during the deletion process.[66,86]

The attack vulnerability problem is maybe most relevant for Internet and similar communication networks,[226,113] but it can also be interpreted in a friendlier way: If a population is to be vaccinated against some malicious pathogen one would save both money and side-effect cases by vaccinating only the potentially biggest spreaders (which arguably are the vertices that would cause most damage if removed in an attack).[188,158,221,88]

Summary of paper VII

In this paper we investigate the efficiency of four different attack strategies and their impact on the size of the largest connected subnetwork and the inverse geodesic length. The four strategies are removal of edges in order of their original degree[11] and their original betweenness (see sect. 2.1.3), and the instanta-

[xiii] This definition works well with model networks but is rather awkward to use for real-world data. (Unless one have a time stamp on the edges so the network growth can be reconstructed.) For regular lattices there are simpler percolation criteria.[327]

neous degree[66] and betweenness during the attacking process. Furthermore, we investigate not only vertex attacks but also edge attacks. To implement degree based attacks of edges we test different definitions of "edge-degree," i.e., purely local centrality measures for edges. The product of the degrees proves to be the one of our candidates with the strongest correlation to edge-betweenness and is thus the definition we use.

We examine two model networks—the Internet at the Autonomous System level and a network of scientific collaborations in high-energy physics[258]—and four types of model networks—random graphs (see sect. 3.1), the Watts-Strogatz model (see sect. 3.3.4), the Barabási-Albert model (see sect. 3.4.1, and the model of paper V. For all cases of edge attacks, and all except one case of vertex attacks, the most devastating strategy is to remove edges or vertices with currently highest betweenness. The one exception is the vertex attack on random graphs where the recalculated degree strategy is most efficient.

Summary of paper VIII

When performing a vaccination campaign one can often not vaccinate the whole population. An efficient attack strategy is a good vaccination strategy—the only difference between the attack and vaccination problem is that the disease dynamics can potentially make the static measures of attack efficiency less relevant. If the attack procedures of paper VII is to be implemented in practice one has to know the structure of the whole network. In practice this is seldom the case for the networks diseases spread over. Paper VIII deals with strategies for vaccination that are—based on the assumption that people have a good understanding of their social neighbourhood—possible to implement in practice. I find that the best strategies in real-world networks are to select the vertices iteratively in a chain-like fashion. The person to vaccinate is the neighbour of the previous vaccinee with highest degree or the neighbour with most edges out of the neighbourhood. Apart from static efficiency measures I use disease spread models that will be discussed in sect. 4.1.

3.5.2 Overload breakdown cascades

The power-grid is a network that is very concrete to us all. We see the edges—the power-lines—ever so often, like the blood vessels of the modern society, transferring electricity from power plants to electronic machinery, household appliances, and so on. Just like the blood, if you block one power-line the electricity has to go somewhere else—the amount that flows into a junction has to

flow out too. If a power-line breaks, the electricity has to reroute its flow. The power-lines that get the flow from the dysfunctional line might become overloaded and break, giving rise to new overloaded power-lines, and eventually a cascading failure and a power-outage that is very hard to control.[101,83] By now you have already guessed that the sensitivity to overload breakdown cascades is related to the network structure—just like the random failures and network attacks of sect. 3.5.1. But, unlike the case of random failures and attacks, to model overload breakdown phenomena the dynamical system as a whole plays a greater role. In fact, the dynamical system on top of the network is what reshapes the network during the cascade.

Not only power-grids are sensitive to cascading failures triggered by overload. Similar events might also happen to other infrastructural networks like the Internet (although the effects might be severe congestion rather than hardware crashes). As usual, physicists have searched for simple, hopefully generic, models for these phenomena.[xiv] One of these model is the *fibre bundle model*.[91] In the context of networks it can be defined as follows: Each vertex is associated with a random number κ_v of a given distribution (the "capacity"), and a parameter $\sigma_v \geqslant 0$ (the "load"). The total load $\sum_v \sigma_v$ is initialised to zero. Then one increases the load continuously, with the same rate at all vertices, until one vertex is overloaded, i.e. if

$$\sigma_v > \kappa_v \,. \tag{3.32}$$

Then v is removed and its load is redistributed to its neighbors:

$$\sigma_w = \sigma'_w + \sigma_v/k_v \text{ for all } w \in \Gamma_v \,, \tag{3.33}$$

where σ'_w denotes w's load just before the breakdown of v.

When the fibre bundle model is placed on a regular lattice it breaks down continuously as the load increases. On a Barabási-Albert model network (see sect. 3.4.1) the breakdown is more violent where the first cascading failure immediately shatters the whole network.[249]

Summary of paper IX

In this paper we present a model for cascading overload failures very different from the fibre bundle model of sect. 3.5.2. We use betweenness as an approximation of the vertex load, so instead of the local load of the fibre bundle model we use a load measure related to the flow on the network. One can question

[xiv] For a model from the engineer literature see ref. 79.

the validity of this approximation (see sect. 2.1.3), but for systems where distance economy is important, the load is probably at least an increasing function of betweenness. In our simulation a vertex breaks down if its betweenness exceeds a threshold value. The threshold is reached by letting the network grow with Barabási-Albert style (see sect. 3.4.1) preferential attachment or uniformly random attachment. We consider two load cases: One corresponding to that every vertex communicates with a fixed number of others (we call it "extrinsic communication activity"). Another case corresponds to the situation when the average user increases its communication activity proportional to the size of the network ("intrinsic communication activity"). The reality for most communication networks, such as the Internet, probably lies between these two cases.

Both load cases show similar behaviours: First the network grows steadily. Then some vertices become overloaded but the network can recover and continue to grow. Eventually the network crashes completely to a state of small isolated subnetworks, reminiscent of the subcritical phase of random graphs (see sect. 3.1). The intrinsic communication activity case has the additional feature that sometimes a giant component does not form. This occurs more often when the degree of the new vertices is higher.

Our picture is an argument that the Internet is facing the "tragedy of the commons."[155][xv] The good news is that the first breakdowns works as a precursor of an immanent collapse.

Summary of paper X

The topic of paper X is similar to that of paper IX; but instead of letting the vertices be sensitive to overloading this paper concerns overload breakdown of edges. The result is in stark contrast to the vertex overload problem. For edge overload the network responds much less violently. There can be some smaller cascades, but the big picture is that the network (for most parameter settings) smoothly reaches a steady state where, at a time step, as many edges break as new edges are added.

[xv] The tragedy of the commons is a metaphor for situations when the common good will suffer if people acts according to their own interests. The metaphor takes its name from when shepherds of a village, in ancient days, could keep their sheep in a commons. For a particular shepherd it is always a good idea, if he can afford, to add another sheep to the commons, but if every one does so, the pasture will be soon be overgrazed.

chapter 4

Dynamical systems on networks

In the end almost all network studies strive for one goal: To elucidate the effects of the underlying network structure on the dynamical systems confined to them. This final chapter concerns just this. In most works in this area one uses a model network for the underlying network. By tuning a network-model parameter and study the effect of the dynamical model one can draw conclusions on the networks structure's impact on the dynamics.

With Chapters 2 and 3 in mind one can outline a program[i] for network science as follows:

1. Find the structures of real-world networks.

2. Construct minimal models for the emergence (or time-evolution) of the structures, and motivate these models by empirical measurements.

3. Study the effect of the underlying network structure on dynamical systems.

If the time-scale of the network evolution is of the same order as the dynamics of the dynamical system confined to the network, then there are further complications to point 3, and we end up in the most complex situation of complex network studies. To my knowledge there is not much done in this area (refs. 189, 293, 81, 372, 314 are the few examples I know).

[i] This is not meant to be restrictive in any way, more like a way to structure (at least my own) network research.

4.1 Epidemiological models

Diseases spread over social networks with more or less well-defined edges.[ii] The easier the chains of infection can be reconstructed, the more useful is network epidemiology. Examples of such infections (appropriate for network studies) are sexually transmitted diseases[218] and infections that primarily spread within wards of hospital (by pathogens such as *Mychoplasma pneumoniae*,[238] Meticillin-resistant *Staphylococcus aureus* or vancomycin-resistant enterococci[120]). Traditional epidemiological models assume a well-mixed population, so that every individual can infect every other. The traditional key quantity is the *basic reproduction rate* R_0,[298] defined as the estimated number of people of an uninfected population that will be infected from one carrier of the pathogen. If $R_0 > 1$ the disease will spread in the population. A crude assessment (what physicists would call a "mean field approximation") is that

$$R_0 = r\beta\tau \,, \tag{4.1}$$

where r is the average number of contacts per time and person, β is the probability of transmitting the infection from an infective to an susceptible person, and τ is the average time of the infection. R_0 tells us if a pathogen is above or below the epidemic threshold in a given population, but says nothing about how many that will be infected or about dynamical questions such as when an epidemic will die out, if it will stay endemic in the population, or when the peak of the epidemic will be reached. To answer such questions one needs to understand the time evolution *effective reproduction rate* R—the expected number of uninfected an infective will spread the disease to. But, rather than explicitly calculating R, one use to take the answers to dynamical questions directly from the models of disease spreading.

Two very simple dynamical models for disease propagation are the SIS ("S" meaning "susceptible" and "I" denoting "infective") and SIR ("R" meaning "removed"[iii]) models, the names giving the cycle of states individuals can go through. The model parameters are β and τ (as defined above). Sometimes τ is replaced by λ—the rate by which infective individuals cease being infective. If one supposes that every individual can infect (or be infected by) everyone else the R_0 for the SIS and SIR models can be written on the form of eq. 4.1. But this assumption is an oversimplification in almost all cases. (To assume that β and τ are the same for all individuals are also oversimplifications that

[ii] For general reviews on this subject, see refs. 16, 136.

[iii] In practice "removed" can mean that a person recovers and get immune, or dies.

may lead to erroneous conclusions.) If the structure of the contact network is known, one can put the SIS and SIR model on top of a network with that typical structure to get a better picture of the disease dynamics. One of the most spectacular findings of network epidemiology is that if the underlying network has a power-law degree distribution with an exponent $\gamma > -3$, then $R_0 > 1$ for any non-zero β and τ;[281] i.e., a disease would be able to spread no matter how un-infectious it is or how short incubation time it has. Subsequent studies[53,250] have, however, toned down the relevance of this result—the degree-degree correlations (see sect. 2.3.3) and cut-offs of real-world fat-tail degree distributions (see sect. 2.1.1) make the epidemic threshold positive.

To derive epidemic threshold criteria for dynamic models on networks one can study the percolation problem on a particular network (see sect. 3.5.1).[302,149] If β and γ are distributed according to P_β and P_γ respectively, then the disease can spread over edge percolation clusters with the edge occupation probability[262]

$$T = \int P_\beta(x) P_\gamma(x') e^{-x/x'} \, dx \, dx' \, . \tag{4.2}$$

The distribution of percolation clusters corresponds to the distribution of outbreak sizes, and the percolation threshold corresponds to the epidemic threshold.

4.2 Spin models of statistical mechanics

In the spin models of statistical mechanics each vertex v is associated with a "spin" σ_v (a number of some bounded set). The probability for the system to be in the configuration $S = \{\sigma_v : v \in V\}$ is determined by the *Boltzmann distribution*:

$$P(S) = \frac{1}{Z} \exp\left(-\frac{H(S)}{k_B T}\right) , \tag{4.3}$$

where k_B is a constant, Z is a normalising constant (cf. eq. 3.22), T is the "temperature," and H maps the state to a real number. In model systems k_B and $H(S)$ are often made dimensionless. In general, the temperature is controlling the randomness (entropy) of the system. In the $N \to \infty$ limit, many systems have a *critical* temperature T_c such that if $T > T_c$ the system behaves qualitatively different than if $T < T_c$. In such a situation the system is said to exhibit a *phase transition* at T_c. Normally, spin models are placed on regular lattices, but the definitions above work for general graphs as well.

Two examples of spin systems are the Ising and XY models: The Ising model was originally a model for so called ferromagnetic materials. It is defined by $\sigma \in \{-1, 1\}$ and

$$H(S) = -\sum_{(v,w)\in E} \sigma_v \sigma_w \,. \tag{4.4}$$

The Ising model has been used in modelling systems very different than magnets. For example, it can be used in analysing the Hopfield model of neural computation,[169,157] opinion formation,[26,25] models of the canopy height of forests,[82] and so on. The XY model is a generalisation of the Ising model to a continuous spin variable $\sigma \in [-\pi, \pi)$:

$$H(S) = -\sum_{(v,w)\in E} \cos(\sigma_v - \sigma_w) \,. \tag{4.5}$$

Traditionally the XY model is used for various physical systems like high-T_c superconductors,[274,236] lattice gauge theory of quantum chromodynamics,[237,332] and more.[iv] Just like the Ising model, the XY model has been used outside the study of physical systems, one example is the formation of flying birds.[339]

Since spin models are close to the heart of statistical physicists there has been many studies of spin models on various types of model networks (see refs. 173, 156, 177, 168, 167, 166, 104, 51, 215, 12, 35, 138). Such studies are mostly of academic interest, but there are actually practical applications (see Summary of paper XI).

Summary of paper XI

In this paper we put XY spins on the vertices of Watts-Strogatz small-world network model (see sect. 3.3.4). We see that (just as for the Ising model[35]) the phase transition is of mean-field character[v] for finite rewiring probability.

[iv] If you are not familiar with these terms, just remember that the XY model exhibits a phase-transition for some value of $T \geqslant 0$ between a more ordered low-temperature phase and a more random high-temperature phase.

[v] When a spin is governed by the average values of the other spins (i.e. their fluctuations does not matter), the system is said to allow a *mean-field* description. For the XY on regular lattices this happens in four (or more) dimensions. Note that a four dimensional hyper-cubic lattice has an average degree $\langle k_v \rangle = 8$ but in our case we can have mean-field behaviour for graphs as sparse as (at least) $\langle k_v \rangle = 4$.

An erratum: In the paper we claim both that any finite rewiring ($P > 0$) gives a finite critical temperature and that the critical temperature T_c can be approximated by

$$T_c = 0.41 \ln P + 2.89 \, . \tag{4.6}$$

This is of course contradictory and my belief is that eq. 4.6 does hold quite good for medium range P, but not for $P \lesssim 0.001$.

At the time of writing this, paper XI is the most cited article in this thesis. This is of course mostly due to physicists' love of spin models, but our work has found its way to a totally different field too: An approach to the modelling of complex systems of many interacting entities (or *agents*) is to take advantage of parallel programming and let each parallel process represent one agent and follow the dynamics of that agent.[131,269] Internally each process progresses according to its *virtual time*. Communication between agents occurs at discrete instants that can, in general, be asynchronous. To ensure causality only processes with a virtual time earlier than the next (local) communication call can be updated. The set of local simulation times at a given real time is called the *virtual time horizon*. For the simulation to be computationally tractable two criteria need to be fulfilled: First, the virtual time horizon should, on average, progress at a nonzero rate. Second, the spread of the time horizon should be bounded as the number of agents grows. One has shown that for certain problems the synchronisation scheme benefits from having a topology as the Watts-Strogatz model (see sect. 3.3.4). In that case the fluctuations of the virtual time horizon should be very close to those of the "order parameter" in the XY model on Watts-Strogatz networks.[340,205]

Summary of paper XII

Just like paper XI this paper concerns a spin model from statistical physics confined to Watts-Strogatz model networks (see sect. 3.3.4). The basis for this study is the same XY model as in paper XI, but this time we look at the *dynamical* critical behaviour. The XY model itself states the probability for a specific configuration $S = \{\sigma_v\}$ to occur, but does not give any answer to questions. For example: Given a configuration S at time 0, what is the probability that the configuration S' occurs at time t. The dynamics we use is a sequential Metropolis Monte Carlo update: We go through the vertices one by one and replace σ_v by $\sigma'_v = \sigma_v + \delta\sigma$ (where $\delta\sigma$ is an uniformly random

number in the interval $[-\pi/6, \pi/6]$) with the probability:

$$P(\sigma_v, \sigma'_v) = \begin{cases} 1 & \text{if } \Delta E(S, S') < 0 \\ \exp(-\Delta E(S, S')/T) & \text{otherwise} \end{cases} \tag{4.7}$$

where

$$\Delta E(S, S') = \sum_{w \in \Gamma_v} \left[\cos(\sigma_w - \sigma'_v) - \cos(\sigma_w - \sigma_v) \right] . \tag{4.8}$$

When the temperature gets closer to its critical value, the relaxation time[vi] diverges as $N^{\bar{z}}$. We show that $\bar{z}$ is constant for finite values of the rewiring probability P, and has the same value as a square grid of dimension four or more (just like the static critical exponents studied in paper XI).

4.3 Traffic flow

In any system consisting of some kind of traffic flow over a network, the structure of the network will be an important factor for the dynamics. A car commuter will know that the traffic jams typically occur at certain points. To understand why and where these points of congestions occur are perhaps the most important questions regarding traffic flow (regardless if the traffic consists of vehicles, data packages or something else). There are two major classes of traffic flow models—differential equation[152] and particle hopping models.[254] In vehicular traffic studies a frequently used particle hopping model is the *asymmetric stochastic exclusion process*:[209] Let the road be represented by an one-dimensional chain of cells where each cell can accommodate one or zero particles (vehicles). At a time step a particle is picked at random and, if the right neighbour cell is unoccupied, moved to the right. The advantage of this model is that it is simple enough to be solved analytically. (Whether or not it is too simple to give any interesting results is still debated.) Simulations of computer network flow has a long tradition and is quite naturally based on protocols such as the TCP/IP.[182,175] However, both for vehicular and computer traffic, large scale simulations of structured complex networks are infrequent.

[vi] Loosely speaking, the relaxation time τ it takes for the system to get back to a normal configuration from an out-of-equilibrium situation. In our case we start from a configuration where $\sigma_v = 0$ for all v and estimate τ from the decay of the correlations between configurations of independent runs (a so called *short time relaxation* scheme)[217]).

Summary of paper XIII

As hinted above, real world communication networks are sensitive to congestion—if an Internet server gets many simultaneous requests to route a package, then the delivery is slowed down. In the rush hours urban traffic gets jammed. In this paper I try to identify congestion-prone vertices by considering a simple particle-hopping system where the items of communication (the "agents" or "particles") are not allowed to occupy the same vertex.[vii] (Hence it is obviously congestion sensitive.) Each agent has a start and a target vertex and a strategy for how to navigate between these points. I give statistics for how the speed of the communication dynamics depends on the underlying network structure. The perhaps most interesting conclusion is the comparison between the occupation ratio (the fraction of the time steps that there is an agent on a specific vertex) and the betweenness (see sect. 2.1.3)—these are, in some cases, far from proportional to each other. This is interesting since betweenness, as mentioned in sect. 2.1.3, has been used as a measure for the "load" on a vertex in communication networks, see refs. 141, 142, 37, 36 and papers IX and X. To optimise speed, agents travel along shortest paths. In a free-flow state (where a particle rarely waits for another) betweenness is a good approximation for the congestion (measured by the occupation ratio). But it is precisely in the opposite, congested, limit that a load measure is interesting. My main result is that non central vertices get a higher occupation ratio than one would expect from the betweenness. More specifically, a neighbour to a high-degree vertex is also likely to be congested.

4.4 Search processes

Related to the information-flow problem is that of network search processes. If one in peer-to-peer file-sharing networks such as Gnutella[5, 294, 2] or SoulSeek[viii] types in a search term, then a recursive chain of requests is started. One does not only ask "do you have 'The Homerun Tiger in a Bush' by Denki Groove?" but rather "if you have 'The Homerun Tiger in a Bush' by Denki Groove please tell me, otherwise please forward this question to your neighbours." Now if you know a little more about your neighbours than just the name and address,

[vii] This restriction may seem rather abstract. In a more realistic model one may consider the agent to be put in a queue upon arrival to the vertex. One example of a real world system with such an exclusion constraint is railway networks represented as a so called "line graph"[208] where the vertices are railway segments and the edges go between segments that leads to the same junction.

[viii] See www.gnutella.com and www.slsknet.org respectively.

and if one runs a depth-first search rather than a breadth-first search (which is reasonable if one does not want to overload the network), then one can choose the order of neighbours to question smarter: If one queries the neighbours in order of their degree, then it turns out that the search, on networks with power-law degree distributions, is as fast as a breadth first search.[3,4,193] This is rather remarkable since the number of requests in parallel grows exponentially with time for a breath-first search (but is constantly one for a depth-first search).

For searches on information networks, as mentioned in sect. 2.1.4, the problem to rank pages with the required search word can be solved by ordering the pages according to increasing information centrality. This centrality measure can be yet more suitable for page ranking by replacing the adjacency matrix $\mathbf{A}$ by the symmetric matrix $\mathbf{A}\mathbf{A}^T$, so eq. 2.6 is replaced by[198]

$$\mathbf{A}\mathbf{A}^T\mathbf{w} = \lambda\mathbf{w}\,. \tag{4.9}$$

Another situation of network search occurs in social networks when someone is searching for e.g. a job[148] or an old classmate. Milgram's experiment (see sect. 2.3.1) shows not only that the paths of social networks are very short but also that humans are very good at finding them. This idea—that social networks are *searchable*—was modelled by Kleinberg in refs. 199, 200: Consider an underlying square lattice and complete the network by (in Watts-Strogatz style, see sect. 3.3.4) adding long-range shortcuts. The additional edges are added with higher probability between vertices that are close in the Euclidean metrics of the underlying square lattice—the probability of adding a short-cut between v_{ij} and $v_{i'j'}$ (subscript denotes coordinates on the lattice) is proportional to

$$\left[(i-i')^2 + (j-j')^2\right]^{\alpha/2} \tag{4.10}$$

with a constant α. The lower time-bound for a greedy algorithm to find its target is proportional to

$$N^\beta \text{ where } \beta = \begin{cases} (2-\alpha)/3 & \text{if } \alpha \in [0,2) \\ (2-\alpha)/(1-\alpha) & \text{if } \alpha \in (2,\infty) \end{cases}\,. \tag{4.11}$$

But when $\alpha = 2$ the time bound scales logarithmically.[ix] This model shows that greedy search based on local information can be efficient for some special types of networks, but the small parameter space where the logarithmic scaling occurs is disappointing, and the resemblance to the structures of social networks is

[ix] In d dimensions the finding time scales logarithmically at $\alpha = 2$.

not very convincing. A subsequent model[357] for a searchable social network has improved these discrepancies: Suppose the characteristics of individuals can be encoded in a trait vector $(s_1, \cdots, s_n)$ where s_i (of some discrete index set) is to be interpreted as "in category i the person belongs to group s_i." The categories can be anything like profession, place of residence, hobby interests, age, ethnicity, and so on. Moreover, the groups of a category form a partial ordering (see sect. 2.2.5), so that the distance between two individuals can be measured by the number of levels to the closest common superior in the tree representing the partial order. Now suppose that the probability that there is an edge between two individuals is decaying exponentially with their distance in the multidimensional trait-space, then the social network is searchable for a wide parameter range.

4.5 Spatial games

The systems of game theory[348] consist of at least two *players* that interact with each others following the *rules* of the game. The players use their own *strategies* that are supposed to maximise the player's *gain*. (The four italicised words above are the crucial ingredients of game theoretical models.) The strategies may or may not be the same for the players; and may or may not change (i.e., the model can have, but need not, a temporal dimension). Quite naturally, game theory has been used to model games (usually with cards or dice, as a basis for artificial intelligence systems) but also systems of economy,[20,21] society,[333] ecology[234] and the evolution of microorganisms.[344,190]

One of the most celebrated examples of game theory is the prisoners' dilemma.[288] The anecdote behind goes like this: Imagine that two criminals A and B are kept in custody, isolated from each other, and being interrogated having the choice to turn the other in, or deny any knowledge of the crime. If neither one gives the other in ("cooperates") they will both get one month in prison. If both give each other in ("defects") they will both get three months in prison. If A gives B in, but not vice versa, B gets five months in prison and A is freed. (The last sentence with A and B exchanged is also true.) The catch of the dilemma is that in the short perspective it is always best for a player to defect, but if one can establish mutual trust with the opponent one could do better cooperating. So, if the criminals believe that this situation will never happen again, then the only rational thing is to defect and turn the other in. But if they know that they will face the same situation again, but not how many more times, then what is the best strategy? This is a very complex optimisation problem that Robert Axelrod attacked in an unusual way—he arranged a

computer program competition where scientists of different disciplines where invited to submit programs trying to beat the rest. The fifteen programs met each other in a round robin tournament over 200 iterations of the game.[22,24] The winner was the simplest of all strategies "Tit For Tat"—start cooperating and then do whatever the opponent did in the previous time step. The virtue of Tit For Tat is that it can establish cooperation while at the same time it avoids exploitation. The efficiency of Tit For Tat has subsequently been confirmed by genetic algorithm based optimisation.[23]

If this theory was only applicable to accomplices facing a trial it would not receive much attention. The truth is that you play the prisoners' dilemma game yourself ever so often: To take an example for the world of science—imagine that you are involved in many projects with different colleagues, to maximise your productivity you would like to contribute just so little to each project that your colleagues allow you to be listed as a coauthor, not more. This might make your colleagues feeling deceived and retaliate in some way in the future. Examples can be found from the Westernisation of Central Africa[50] to the evolution of RNA viruses,[344] from trench warfare[22] to NASCAR racing.[297]

The prisoners' dilemma game, just as most other games, can be generalised to many players. In such systems, in both nature and society, the players often interact in a restricted neighbourhood only. This situation can lead to spatial effects—in the context of the prisoners' dilemma game there might be regions (in social space) of cooperation or defection. In Nowak and May's original version of the spatial prisoners' dilemma game[271] the deceived will always get zero gain (regardless whether or not the opponent cooperates). The players follow the strategy of the best scorer of itself and its neighbours the previous time step (all players are updated synchronously). So the question one asks with this setting is not what the best strategy is, but what the conditions are for spatial stability of cooperation. The dynamics of Nowak and May's game will soon reach a cyclic behaviour. To give the dynamics a touch of reality (and make the time evolution less repetitive and boring) one can add the occasional counter-rational choice of strategy,[220] a so called "mutation."

Summary of paper XIV

In this paper we put Nowak and May's prisoners' dilemma game on a two-dimensional Watts-Strogatz type network (see sect. 3.3.4). Apart from the symmetric connections we have one vertex with out-going arcs modelling an influential person, like a media celebrity. With this additional ingredient the time evolution of the cooperator density shows downward spikes. We give

statistics for the frequency and duration of this instability.

Summary of paper XV

This paper we put the Nowak-May game on empirical real-world social networks (based on an e-mail exchange study,[112] the pussokram.com network of paper III and the acquaintance network of ref. 370). The time evolution of the cooperator density has an even more complex behaviour than the model of paper XIV despite the lack of directed links. (However, this is not the only spatial games with a complex time evolution.[372,63,270]) As the networks have the structure of real-world social networks we argue that this is the most realistic prisoners' dilemma study to date. In future studies one can extend to use real-world dynamic too. But then the spatial prisoners' dilemma is pushed to the limit—to get more realistic simulations of cooperation and exploitation one needs to include deletion, or rewiring, of links as a response to low gain[372] as the really unrealistic feature of the model is that some agent gets exploited repeatedly without doing anything about their situation. Anyway, I believe that the main conclusion is correct and interesting—that if a socially central person changes starts to exploit his, or her, environment, then it might trigger a period of social unrest in a larger community than expected. But the change of one person's mind is not always enough—the minds of the people in his, or her, social surrounding might need to be set in a specific way. The social structure is just as important as the rules of the game for the sustentation of cooperative regimes.

Bibliography

1 G. Abramson and M. Kuperman, *Social games in a social network*, Phys. Rev. E **63** (2001), art. no. 030901.

2 L. A. Adamic, *Network Dynamics: The World Wide Web*, PhD thesis, Stanford University, Stanford, 2001.

3 L. A. Adamic, R. M. Lukose, and B. A. Huberman, *Local search in unstructured networks*, in Handbook of Graphs and Networks, S. Bornholdt and H. G. Schuster, eds., Wiley-VCH, Berlin, 2003.

4 L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, *Search in power-law networks*, Phys. Rev. E **64** (2001), art. no. 046135.

5 E. Adar and B. A. Huberman, *Free riding on gnutella*, First Monday **5** (2000).

6 R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice-Hall, Upper Saddle River, 1993.

7 W. Aiello, F. Chung, and L. Lu, *A random graph model for massive graphs*, in Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, New York, 2000, Association of Computing Machinery, pp. 171-180.

8 R. Alberich, J. Miro-Julia, and F. Rossello, *Marvel universe looks almost like a real social network*. e-print cond-mat/0202174.

9 R. Albert and A.-L. Barabási, *Topology of evolving networks: Local events and universality*, Phys. Rev. Lett. **85** (2000), pp. 5234-5237.

10 ——, *Statistical mechanics of complex networks*, Rev. Mod. Phys **74** (2002), pp. 47-98.

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

12 A. Aleksiejuk, J. A. Hołyst, and D. Stauffer, *Ferromagnetic phase transition in Barabási-Albert networks*, Physica A **310** (2002), pp. 260-266.

13 N. Alon, R. Yuster, and U. Zwick, *Finding and counting given length cycles*, Algorithmica **17** (1997), pp. 209-223.

14 C. Anderson, S. Wasserman, and B. Crouch, *A p^* primer: Logit models for social networks*, Soc. Netw. **21** (1999), pp. 37-66.

15 P. W. Anderson, *More is different*, Science **177** (1972), pp. 393-396.

16 R. M. Anderson and R. M. May, *Infectious diseases in humans*, Oxford University Press, Oxford, 1992.

17 M. Anghel, Z. Toroczkai, K. E. Bassler, and G. Korniss, *Competition in social networks: Emergence of a scale-free leadership structure and collective efficiency*. e-print cond-mat/0307740.

18 J. M. Anthonisse, *The rush in a directed graph*, Tech. Rep. BN 9/71, Stichting Mathematisch Centrum, 1971.

19 K. Appel and W. Haken, *Every planar map is four colorable*, Am. Math. Soc. Bull. **82** (1976), pp. 711-712.

20 R. J. Aumann and S. Hart, eds., *Handbook of game theory with economic applications*, vol. **1**, North-Holland, Amsterdam, 1992.

21 ——, eds., *Handbook of game theory with economic applications*, vol. **2**, North-Holland, Amsterdam, 1994.

22 R. Axelrod, *The evolution of cooperation*, Basic Books, New York, 1984.

23 ——, *The evolution of strategies in the iterated prisoner's dilemma*, in Genetic Algorithms and Simulated Annealing, L. Davis, ed., Morgan Kaufman, Los Altos, 1987, pp. 32-41.

24 R. Axelrod and W. D. Hamilton, *The evolution of cooperation*, Science **211** (1981), pp. 1390-1396.

25 D. B. Bahr and E. Passerini, *Statistical mechanics of collective behavior: Macro-sociology*, J. Math. Sociol. **23** (1998), pp. 29-49.

26 ——, *Statistical mechanics of opinion formation and collective behavior*, J. Math. Sociol. **23** (1998), pp. 1-27.

27 D. Baird and R. E. Ulanowicz, *The seasonal dynamics of the Chesapeake Bay ecosystem*, Ecological Monographs **59** (1989), pp. 329-364.

28 F. Ball, D. Mollison, and G. Scala-Tomba, *Epidemics with two levels of mixing*, Ann. Appl. Probab. **7** (1997), pp. 46-89.

29 J. R. Banavar, A. Maritan, and A. Rinaldo, *Size and form in efficient transportation networks*, Nature **399** (1999), pp. 130-132.

30 D. L. Banks and K. M. Carley, *Models for network evolution*, J. Math. Sociol. **21** (1996), pp. 173-196.

31 A.-L. Barabási, *Linked: The New Science of Networks*, Perseus, Cambridge, MA, 2002.

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

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

34 A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, *Evolution of the social network of scientific collaborations*, Physica A **311** (2002), pp. 590-614.

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

36 M. Barthélemy, *Betweenness centrality in large complex networks*. e-print cond-mat/0309436.

37 ——, *Comment on: "Universal behavior of load distribution in scale-free networks"*, Phys. Rev. Lett. **91** (2003), art. no. 189803.

38 V. Batagelj, P. Doreian, and A. Ferligoj, *An optimizational approach to regular equivalence*, Soc. Netw. **14** (1992), pp. 121-135.

39 A. Bavelas, *A mathematical model for group structure*, Human Organization **7** (1948), pp. 16-30.

40 ——, *Communication patterns in task oriented groups*, J. Acoust. Soc. Amer. **22** (1950), pp. 271-282.

41 P. Bearman, J. Moody, and R. Faris, *Networks and history*, Complexity **8** (2003), pp. 61-71.

42 P. Bearman, J. Moody, and K. Stovel, *Chains of affection: The structure of adolescent romantic and sexual networks*. preprint submitted to Am. J. Soc.

43 M. A. Beauchamp, *An improved index of centrality*, Behav. Sci. **10** (1965), pp. 161-163.

44 M. Beekman, D. J. T. Sumpter, and F. L. W. Ratnieks, *Phase transition between disordered and ordered foraging in pharaoh's ants*, Proc. Natl. Acad. Sci. USA **98** (2001), pp. 9703-9706.

45 A. Bekessy, P. Bekessy, and J. Komolos, *Asymptotic enumeration of regular matrices*, Stud. Sci. Math. Hung. **7** (1972), pp. 343-353.

46 J. Berg and M. Lässig, *Correlated random networks*, Phys. Rev. Lett. **89** (2002), art. no. 228701.

47 N. Berger, B. Bollobás, C. Borgs, J. Chayes, and O. Riordan, *Degree distribution of the FKP network model*, in Proceedings of the 30th International Conference on Automata, Languages, and Programming, vol. **2719** of Lecture nodes in Computer science, Heidelberg, 2003, Springer, pp. 725-738.

48 H. Bernard and B. Killworth, *Informant accuracy in social network data II*, Human Comm. Res. **4** (1977), pp. 3-18.

49 H. R. Bernard, P. D. Kilworth, M. J. Evans, C. McCarty, and G. A. Selley, *Studying social relations cross-culturally*, Ethnology **27** (1988), pp. 155-179.

50 D. W. Bethlehem, *The effect of Westernization on cooperative behavior in Central Africa*, Int. J. Psychol. **10** (1975), pp. 219-224.

51 G. Bianconi, *Mean field solution of the Ising model on a Barabási-Albert network*, Phys. Lett. A **303** (2002), pp. 166-168.

52 G. Bianconi and A.-L. Barabási, *Bose-Einstein condensation in complex networks*, Phys. Rev. Lett. **86** (2001), pp. 5632-5635.

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

54 B. Bollobás, *Modern Graph Theory*, Springer, Berlin, 1999.

55 ——, *Random Graphs*, Academic Press, New York, 2 ed., 2001.

56 B. Bollobás and F. R. K. Chung, *The diameter of a cycle plus a random matching*, SIAM J. Discrete Math. **1** (1988), pp. 328-333.

57 B. Bollobás and O. Riordan, *The diameter of a scale-free random graph*. preprint, Department of Mathematical Sciences, University of Memphis, 2002.

58 E. Bonabeau, G. Theraulaz, J.-L. Deneubourg, S. Aron, and S. Camazine, *Self-organization in social-insects*, Trends Ecol. Evol. **12** (1997), pp. 188-193.

59 P. F. Bonacich, *A technique for analyzing overlapping memberships*, in Sociological Methodology, H. Costner, ed., Jossey-Bass, San Francisco, 1972.

60 ——, *Power and centrality: A family of measures*, Amer. J. Sociol. **92** (1987), pp. 1170-1182.

61 G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna, *Topology of correlation-based minimal spanning trees in real and model markets*, Phys. Rev. E **68** (2003), art. no. 046130.

62 U. Brandes, *A faster algorithm for betweenness centrality*, J. Math. Sociol. **25** (2001), pp. 163-177.

63 K. Brauchli, T. Killingback, and M. Doebeli, *Evolution of cooperation in spatially structured populations*, J. Theor. Biol. **200** (1999), pp. 405-417.

64 R. Breiger and P. Pattison, *Cumulated social roles: The duality of persons and their algebras*, Soc. Netw. **8** (1986), pp. 215-256.

65 S. Brin and L. Page, *The anatomy of a large-scale hypertextual websearch engine*, Comput. Netw. **30** (1998), pp. 107-117.

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

67 A. Broida and K. C. Claffy, *Internet topology: Connectivity of IP graphs*, in Scalability and Traffic Control in IP Networks, S. Fahmy and K. Park, eds., vol. **4526** of Proc. SPIE, Bellingham, 2001, International Society for Optical Engineering, pp. 172-187.

68 L. M. Brown, P. S. McNatt, and G. D. Cooper, *Ingroup romantic preferences among Jewish and non-Jewish White undergraduates*, Int. J. Intercult. Rel. **27** (2003), pp. 335-354.

69 L. Brudner and D. White, *Class, property and structural endogamy: Visualizing networked histories*, Theory and Society **25** (1997), pp. 161-208.

70 F. Buckley and F. Harary, *Distance in graphs*, Addison-Wesley, Redwood City, 1989.

71 Z. Burda, J. D. Correia, and A. Krzywicki, *Statistical ensemble of scale-free random graphs*, Phys. Rev. E **64** (2001), art. no. 046118.

72 R. S. Burt, *Positions in networks*, Soc. Forces **55** (1976), pp. 93-122.

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

74 D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz, *Are randomly grown graphs really random?*, Phys. Rev. E **64** (2001), art. no. 041902.

75 D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Network robustness and fragility: Percolation on random graphs*, Phys. Rev. Lett. **85** (2000), pp. 5468-5471.

76 S. Camazine, J.-L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau, eds., *Self-organization in biological systems*, Princeton University Press, Princeton, 2001.

77 T. A. Caplow, *A theory of coalitions in the triad*, Am. Sociol. Rew. **21** (1956), pp. 489-493.

78 C. R. Carpenter, *Concepts and problems of primate sociometry*, Sociometry **8** (1945), pp. 56-61.

79 B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman, *Critical points and transitions in an electric power transmission model for cascading failure blackouts*, Chaos **12** (2002), pp. 985-994.

80 M. Catanzaro, G. Caldarelli, and L. Pietronero, *Assortative model for social networks*. e-print cond-mat/0308073.

81 S. Catterall, J. Kogut, and R. Renken, *The XY model on a dynamical random lattice*, Nucl. Phys. B **408** (1993), pp. 427-444.

82 J. Chave, *Spatial patterns and persistence of woody plant species in ecological communities*, Am. Nat. **157** (2001).

83 H.-D. Chiang, I. Dobson, R. Thomas, J. Thorp, and L. Fekih-Ahmed, *On voltage collapse in electric power systems*, IEEE Trans. Power Systems **5** (1990), pp. 601-611.

84 J. Ching, *Chinese Religions*, Macmillan, London, 1993.

85 F. R. K. Chung and M. R. Garey, *Diameter bounds for altered graphs*, J. Graph Theory **8** (1984), pp. 511-534.

86 R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, *Resilience of the Internet to random breakdowns*, Phys. Rev. Lett. **85** (2000), pp. 4626-4628.

87 ——, *Breakdown of the Internet under intentional attack*, Phys. Rev. Lett. **86** (2001), pp. 3682-3685.

88 R. Cohen, S. Havlin, and D. ben Avraham, *Efficient immunization strategies for computer networks and populations*, Phys. Rev. Lett. **91** (2003), art. no. 247901.

89 P. Collins, *Changing Ideals in Modern Architecture 1750-1950*, McGill-Queen's University Press, Montreal, London, 2 ed., 1998.

90 R. C. Connor, M. R. Heithaus, and L. M. Barre, *Superalliance of bottlenose dolphins*, Nature **397** (1999), pp. 571-572.

91 H. E. Daniels, *The statistical theory of the strength of bundles of threads I*, Proc. R. Soc. A **183** (1945), pp. 405-435.

92 A. Davis, B. B. Gardner, and M. R. Gardner, *Deep South*, University of Chicago Press, Chicago, 1941.

93 G. F. Davis and H. R. Greve, *Corporate elite networks and governance changes in the 1980s*, Amer. J. Sociol. **103** (1997), pp. 1-37.

94 G. F. Davis, M. Yoo, and W. E. Baker, *The small world of the American corporate elite, 1982-2001*, Strategic Organization **1** (2003), pp. 301-326.

95 J. A. Davis and S. Leinhardt, *The structure of positive interpersonal relations in small groups*, in Sociological Theories in Progress, J. Berger, ed., vol. **2**, Houghton Mifflin, Boston, 1972, pp. 218-251.

96 D. de Lima e Silva, M. Medeiros Soares, M. V. C. Henriques, M. T. Schivani Alves, S. G. de Aguiar, T. P. de Carvalho, G. Corso, and L. S. Lucena, *The complex network of the Brazilian popular music*, Physica A **332,** (2004), pp. 559-565.

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

98 D. J. de Solla Price, *Networks of scientific papers*, Science **149** (1965), pp. 510-515.

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

100 P. Dittrich, F. Liljeros, A. Soulier, and W. Banzhaf, *Spontaneous group formation in the seceder model*, Phys. Rev. Lett. **84** (2000), pp. 3205-3208.

101 I. Dobson, H. Glavitsch, C. Liu, Y. Tamura, and K. Vu, *Voltage collapse in power systems*, IEEE Circ. Dev. Mag. **8** (1992), pp. 40-45.

102 P. S. Dodds, R. Muhamad, and D. J. Watts, *An experimental study of search in global social networks*, Science **301** (2003), pp. 827-830.

103 S. N. Dorogovtsev, *Networks with given correlations*. e-print cond-mat/0308336.

104 S. N. Dorogovtsev, V. Goltsev, and J. F. F. Mendes, *Ising model on networks with an arbitrary distribution of connections*, Phys. Rev. E **66** (2002), art. no. 016104.

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

106 S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, *Structure of growing networks with preferential linking*, Phys. Rev. Lett. **85** (2000), pp. 4633-4636.

107 ——, *Size-dependent degree distribution of a scale-free growing network*, Phys. Rev. E **63** (2001), art. no. 062101.

108 H. Dumoulin, ed., *The development of Chinese Zen after the Sixth Patriarch: In the light of Mumonkan*, First Zen Institute, New York, 1953.

109 J. A. Dunne, R. J. Williams, and N. D. Martinez, *Food-web structure and network theory: The role of connectance and size*, Proc. Natl. Acad. Sci. USA **99** (2002), pp. 12917-12922.

110 ——, *Network structure and biodiversity loss in food webs: Robustness increases with connectance*, Ecology Lett. **5** (2002), pp. 558-567.

111 H. Ebel, J. Davidsen, and S. Bornholdt, *Dynamics of social networks*, Complexity **8** (2002), pp. 24-27.

112 H. Ebel, L.-I. Mielsch, and S. Bornholdt, *Scale-free topology of e-mail networks*, Phys. Rev. E **66** (2002), art. no. 035103.

113 J. Ellis, D. Fisher, T. Longstaff, L. Pesante, and R. Pethia, *Report to the President's Commission on Critical Infrastructure Protection*, Tech. Rep. CMU/SEI-97-SR-003, Carnegie Mellon University, 1997.

114 D. Eppstein and J. Y. Wang, *Fast approximation of centrality*, in Proceedings of the 12th Symposium on Discrete Algorithms, New York, 2001, Association for Computing Machinery, pp. 228-229.

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

116 ——, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci. **5** (1960), pp. 17-61.

117 A. Fabrikant, Koutsoupias, and C. H. Papadimitriou, *Heuristically optimized trade-offs: A new paradigm for power laws in the Internet*, in Proceedings of the 29th International Conference on Automata, Languages, and Programming, vol. **2380** of Lecture notes in Computer science, Heidelberg, 2002, Springer, pp. 110-122.

118 M. Faloutsos, P. Faloutsos, and C. Faloutsos, *On power-law relationships of the Internet topology*, Comput. Commun. Rev. **29** (1999), pp. 251-262.

119 T. J. Fararo and M. H. Sunshine, *A study of a biased friendship net*, Syracuse University Press, Syracuse, NY, 1964.

120 B. M. Farr, C. D. Salgado, T. B. Karchmer, and R. J. Sherertz, *Can antibiotic-resistant nosocomial infections be controlled?*, Lancet Infect. Dis. **1** (2001), pp. 38-45.

121 A. Ferguson, *Zen's Chinese Heritage: The Masters and Their Teachings*, Wisdom Publications, Somerville, 2000.

122 N. M. Ferguson and G. P. Garnett, *More realistic models of sexually transmitted disease transmission dynamics: Sexual partnership networks, pair models, and moment closure*, Sex. Transm. Dis. **27** (2000), pp. 600-609.

123 R. Ferrer i Cancho, C. Janssen, and R. V. Solé, *Topology of technology graphs: Small world patterns in electronic circuits*, Phys. Rev. E **64** (2001), art. no. 046119.

124 R. Ferrer i Cancho and R. V. Solé, *Optimization in complex networks*, Tech. Rep. 01-11-068, Santa Fe Institute, November 2001.

125 ——, *The small world of human language*, Proc. Roy. Soc. Lond. B **268** (2001), pp. 2261-2265.

126 L. Festinger, *A Theory of Cognitive Dissonance*, Row, Peterson & Co., Evanston, 1957.

127 K. Fjällborg, *Alla vägar bär till Gwyneth*, Aftonbladet (2003).

128 O. Frank and D. Strauss, *Markov graphs*, J. Amer. Stat. Assoc. **81** (1986), pp. 832-842.

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

130 L. C. Freeman, S. P. Borgatti, and D. R. White, *Centrality in valued graphs: A measure of betweenness based on network flow*, Soc. Netw. **13** (1991), pp. 141-154.

131 R. M. Fujimoto, *Parallel discrete event simulation*, Commun. ACM **33** (1990), pp. 30-53.

132 K. Fukuda, H. Takayasu, and M. Takayasu, *Observation of phase transition behaviors in Internet traffic*, Adv. Perf. Anal. **2** (1999), pp. 45-66.

133 ——, *Spatial and temporal behavior of congestion in Internet traffic*, Fractals **7** (1999), pp. 23-31.

134 J. Galaskiewicz and P. V. Marsden, *Interorganizational resource networks: Formal patterns of overlap*, Social Sci. Res. **7** (1978), pp. 89-107.

135 M. Gell-Mann, *The Quark and the Jaguar: Adventures in the Simple and the Complex*, W. H. Freeman, New York, 1994.

136 J. Giesecke, *Modern infectious disease epidemiology*, Arnold, London, 2 ed., 2002.

137 M. Girvan and M. E. J. Newman, *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. USA **99** (2002), pp. 7821-7826.

138 M. Gitterman, *Small-world phenomena in physics: The Ising model*, J. Phys. A: Math. Gen. **33** (2000), pp. 8373-8381.

139 M. Gladwell, *The tipping point: How small things can make a big difference*, Little Brown, New York, 2000.

140 P. Gleiser and L. Danon, *Community structure in Jazz*, Adv. Complex. Syst. **6** (2003), pp. 565-573.

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

142 K.-I. Goh, E. Oh, H. Jeong, B. Kahng, and D. Kim, *Classification of scale-free networks*, Proc. Natl. Acad. Sci. USA **99** (2002), pp. 12583-12588.

143 D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, *Using collaborative filtering to weave an information tapestry*, Comm. ACM **35** (1992), pp. 61-70.

144 D. S. Goldberg and F. P. Roth, *Assessing experimentally derived interactions in a small world*, Proc. Natl. Acad. Sci. USA **100** (2003), pp. 4372-4376.

145 D. M. Gordon, *The organization of work in social insect colonies: Task allocation as an ad hoc, dynamical network*, Complexity **8** (2003), pp. 43-46.

146 M. S. Granovetter, *The strength of weak ties*, Am. J. Sociol. **78** (1973), pp. 1360-1380.

147 ——, *Threshold models of collective behavior*, Am. J. Sociol. **83** (1978), pp. 1420-1443.

148 ——, *Getting a Job: A Study of Contacts and Careers*, University of Chicago Press, 2 ed., 1995.

149 P. Grassberger, *On the critical behavior of the general epidemic process and dynamical percolation*, Math. Biosci. **63** (1983), pp. 157-172.

150 J. Guare, *Six degrees of separation: A play*, Vintage, New York, 1990.

151 R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, *Self-similar community structure in organisations*, Phys. Rev. E **65** (2003), art. no. 065103.

152 R. Haberman, *Mathematical models in mechanical vibrations, population dynamics, and traffic flow*, Prentice-Hall, Upper Saddle River, 1977.

153 P. Hage and F. Harary, *Structural models in anthropology*, Cambridge University Press, Cambridge, 1983.

154 S. J. Hall and D. Raffaelli, *Food web patterns: Lessons from a species-rich web*, J. Anim. Ecol. **60** (1991), pp. 823-842.

155 G. Hardin, *The tragedy of the commons*, Science **162** (1968), pp. 1243-1248.

156 C. P. Herrero, *Ising model in small-world networks*, Phys. Rev. E **65** (2002), art. no. 066110.

157 J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the theory of neural computation*, Addison-Wesley, Redwood City, 1991.

158 H. Hethcote, *The mathematics of infectious diseases*, SIAM Rev. **42** (20010), pp. 599-653.

159 N. Higham, *Exploiting fast matrix multiplication within the level 3 BLAS*, ACM Trans. Math. Soft. **16** (1990), pp. 352-368.

160 R. A. Hinde, ed., *Primate Social Relationships: An Integrated Approach*, Blackwell Science, Oxford, 1983.

161 A. J. Hoffman and R. R. Singleton, *On Moore graphs of diameter two and three*, IBM J. Res. Develop. **4** (1960), pp. 497-504.

162 J. H. Holland, *Hidden order: How adaptation builds complexity*, Perseus, Cambridge, MA, 1996.

163 P. W. Holland and S. Leinhardt, *Some evidence on the transitivity of positive interpersonal sentiment*, Am. J. Sociol. **72** (1972), pp. 1205-1209.

164 ——, *The statistical analysis of local structure in social networks*, in Sociological Methodology, D. R. Heise, ed., Jossey-Bass, San Fransico, 1975, pp. 1-45.

165 ——, *An exponential family of probability distribuions for directed graphs*, J. Amer. Stat. Assoc. **76** (1981), pp. 33-65.

166 H. Hong, B. J. Kim, and M. Y. Choi, *Comment on "Ising model on a small world network"*, Phys. Rev. E **66** (2002), art. no. 018101.

167 ——, *Stochastic resonance in the driven Ising model on small-world networks*, Phys. Rev. E **66** (2002), art. no. 011107.

168 ——, *Optimal size of a complex network*, Phys. Rev. E **67** (2003), art. no. 046101.

169 J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. USA **79** (1982), pp. 2554-2558.

170 K. Huang, *Statistical mechanics*, John Wiley, New York, 2 ed., 1987.

171 C. Huitema, *Routing in the Internet*, Prentice-Hall, Upper Saddle River, 2 ed., 2000.

172 M. Huxam, S. Beaney, and D. Raffaelli, *Do parasites reduce the chances of triangulation in real food webs?*, Oikos **76** (1996), pp. 284-300.

173 F. Iglói and L. Turban, *First- and second-order phase transitions in scale-free networks*, Phys. Rev. E **66** (2002), art. no. 036140.

174 J. Ihmels, G. Friedlander, S. Bergmann, O. Sarig, Y. Ziv, and N. Barkai, *Revealing modular organization in the yeast transcriptional network*, Nature Genetics **31** (2002), pp. 370-377.

175 V. Jacobson and M. J. Karels, *Congestion avoidance and control*, ACM Computer Comm. Rev. **18** (1988), pp. 314-329.

176 S. Janson, T. Łuczac, and A. Ruciński, *Random Graphs*, Whiley, New York, 1999.

177 D. Jeong, H. Hong, B. J. Kim, and M. Y. Choi, *Phase transition in the Ising model on a small-world network with distance-dependent interactions*, Phys. Rev. E **68** (2003), art. no. 027101.

178 H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, *Lethality and centrality in protein networks*, Nature **411** (2001), pp. 41-42.

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

180 E. M. Jin, M. Girvan, and M. E. J. Newman, *The structure of growing social networks*, Phys. Rev. E **64** (2001), art. no. 046132.

181 S. C. Johnson, *Hierarchical clustering schemes*, Psychometrika **32** (1976), pp. 241-253.

182 L. Kalampoukas and A. Varma, *Performance of TCP over multi-hop ATM networks: A comparative study of ATM-layer congestion control schemes*, Tech. Rep. UCSC-CRL-95-13, University of California, Santa Cruz, 1995.

183 V. K. Kalapala, V. Sanwalani, and C. Moore, *The structure of the United States road network*. University of New Mexico, 2003.

184 C. Kampmeier and B. Simon, *Individuality and group formation: The role of independence and differentiation*, J. Person. Soc. Psychol. **81** (2001), pp. 448-462.

185 L. Katz, *A new status index derived from sociometric analysis*, Psychometrika **18** (1953), pp. 39-43.

186 L. Katz and J. H. Powell, *The number of locally restricted directed graphs*, Proc. Am. Math. Soc. **5** (1954), pp. 621-626.

187 ——, *Probability distributions of random variables associated with a structure of the sample space of sociometric investigations*, Ann. Math. Stat. **28** (1957), pp. 442-448.

188 M. J. Keeling, M. E. J. Woolhouse, R. M. May, G. Davies, and B. T. Grenfell, *Modelling vaccination strategies against foot-and-mouth disease*, Nature **421** (2003), pp. 136-142.

189 D. Kempe, J. Kleinberg, and A. Kumar, *Connectivity and inference problems for temporal networks*, J. Comput. Syst. Sci. **64** (2002), pp. 820-842.

190 B. Kerr, M. A. Riley, M. W. Feldman, and B. J. M. Bohannan, *Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors*, Nature **418** (2002), pp. 171-174.

191 B. Killworth and H. Bernard, *Informant accuracy in social network data*, Human Organization **35** (1976), pp. 269-286.

192 B. J. Kim, A. Trusina, P. Minnhagen, and K. Sneppen, *Self organized scale-free networks from merging and regeneration*. preprint, Department of Molecular Science and technology, Ajou University.

193 B. J. Kim, C. N. Yoon, S. K. Han, and H. Jeong, *Path finding strategies in scale-free networks*, Phys. Rev. E **65** (2002), art. no. 027103.

194 D.-H. Kim, B. Kahng, and D. Kim, *The q-component static model: Modeling social networks*. e-print cond-mat/0307184.

195 J. Kim, P. L. Krapivsky, B. Kahng, and S. Redner, *Infinite-order percolation and giant fluctuations in a protein interaction network*, Phys. Rev. E **66** (2002), art. no. 055101.

196 O. Kinouchi, A. S. Martinez, G. F. Lima, G. M. Lourenço, and S. Risau-Gusman, *Deterministic walks in random networks: An application to thesaurus graphs*, Physica A **315** (2002), pp. 665-676.

197 S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Optimization by simulated annealing*, Science **220** (1983), pp. 671-680.

198 J. M. Kleinberg, *Authoritative sources in a hyperlinked environment*, J. ACM **46** (1999), pp. 604-632.

199 ——, *Navigation in a small world*, Nature **406** (2000), p. 845.

200 ——, *The small-world phenomenon: An algorithmic perspective*, in Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, New York, 2000, Association of Computing Machinery, pp. 163-170.

201 J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, *The Web as a graph: Measurements, models and methods*, in Proceedings of the International Conference on Combinatorics and Computing, vol. **1627** of Lecture notes in computer science, Berlin, 1999, Springer, pp. 1-18.

202 A. S. Klovdahl, J. J. Potterat, D. E. Woodhouse, J. B. Muth, S. Q. Muth, and W. W. Darrow, *Social networks and infectious disease: The Colorado Springs study*, Soc. Sci. Med. **38** (1994), pp. 79-88.

203 D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing*, Addison-Wesley, Reading, 1993.

204 T. A. Kohler and G. J. Gumerman, eds., *Dynamics in Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes*, Santa Fe Institute Studies on the Sciences of Complexity, Oxford University Press, Oxford, 2000.

205 G. Korniss, M. A. Novotny, H. Guclu, Z. Toroczkai, and P. A. Rikvold, *Suppressing roughness of virtual times in parallel discrete-event simulations*, Science **299** (2003), pp. 677-679.

206 P. L. Krapivsky, S. Redner, and F. Leyvraz, *Connectivity of growing random networks*, Phys. Rev. Lett. **85** (2000), pp. 4629-4632.

207 V. E. Krebs, *Mapping networks of terrorist cells*, Connections **24** (2002), pp. 43-52.

208 M. Kretschmar and M. Morris, *Measures of concurrency in networks and the spread of infectious disease*, Math. Biosci. **133** (1996), pp. 165-195.

209 J. Krug and H. Spohn, *Universality classes for deterministic surface growth*, Phys. Rev. A **38** (1988), pp. 4271-4283.

210 L. Kullmann, J. Kertesz, and K. Kaski, *Time dependent cross correlations between different stock returns: A directed network of influence*. e-print cond-mat/0203256.

211 R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. S. Tomkins, and E. Upfal, *Stochastic models for the Web graph*, in Proceedings of the 42st Annual IEEE Symposium on the Foundations of Computer Science, New York, 2000, Institute of Electrical and Electronics Engineers, pp. 57-65.

212 C. Kuttler and R. Blossey, *Individualistic networks and networking individuals in Tango Argentino*. e-print cond-mat/0308455.

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

214 ——, *Is the Boston subway a small-world network?*, Physica A **314** (2002), pp. 109-113.

215 M. Leone, A. Vázquez, A. Vespignani, and R. Zecchina, *Ferromagnetic ordering in graphs with arbitrary degree distribution*, Eur. Phys. J. B **28** (2002), pp. 191-197.

216 M. Li and P. M. B. Vitanyi, *An Introduction to Kolmogorov Complexity and Its Applications*, Springer-Verlag, New York, 1993.

217 Z. B. Li, L. Schülke, and B. Zheng, *Dynamic Monte Carlo measurement of critical exponents*, Phys. Rev. Lett. **74** (1995), pp. 3396-3398.

218 F. Liljeros, C. R. Edling, and L. A. Nunes Amaral, *Sexual networks: Implication for the transmission of sexually transmitted infection*, Microbes Infect. **5** (2003), pp. 189-196.

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

220 K. Lindgren and M. G. Nordahl, *Evolutionary dynamics of spatial games*, Physica D **75** (1994), pp. 292-309.

221 A. L. Lloyd and R. M. May, *How viruses spread among computers and people*, Science **292** (2001), pp. 1316-1317.

222 F. Lorrain and H. C. White, *Structural equivalence of individuals in social networks*, J. Math. Sociol. **1** (1971), pp. 49-80.

223 S. H. Low, F. Paganini, and J. C. Doyle, *Internet congestion control*, IEEE Control Syst. Mag. **22** (2002), pp. 28-43.

224 J. J. Luczkovich, S. P. Borgatti, J. C. Johnson, and M. G. Everett, *Measuring trophic role similarity in food webs using regular equivalence*, J. Theor. Biol. **220** (2003), pp. 303-321.

225 J. J. Luczkovich, G. C. Ward, J. C. Johnson, R. C. Christian, D. Baird, H. Neckles, and B. Rizzo, *Determining the trophic guilds of fishes and macroinvertebrates in a seagrass food web*, Estuaries **25** (2002), pp. 1143-1164.

226 D. Magoni, *Tearing down the Internet*, IEEE J. Sel. Areas Commun. **21** (2003), pp. 949-960.

227 R. N. Mantegna, *Hierarchical structure in financial markets*, Eur. Phys. J. B **11** (1999), pp. 193-197.

228 P. Mariolis, *Interlocking directorates and control of corporations: The theory of bank control*, Social Sci. Quart. **56** (1975), pp. 425-439.

229 M. Marsili, F. Vega-Redondo, and F. Slanina, *The rise and fall of a networked society*. e-print cond-mat/0307321.

230 N. D. Martinez, *Artifacts or attributes? Effects of resolution on the Little Rock Lake food web*, Eco. Monogr. **61** (1991), pp. 367-392.

231 S. Maslov and K. Sneppen, *Specificity and stability in topology of protein networks*, Science **296** (2002), pp. 910-913.

232 S. Maslov, K. Sneppen, and A. Zaliznyak, *Detection of topological patterns in complex networks: Correlation profile of the Internet*, Physica A **333** (2004), pp. 529-540.

233 N. Mathias and V. Gopal, *Small worlds: How and why*, Phys. Rev. E **63** (2001), art. no. 021117.

234 J. Maynard Smith, *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, 1982.

235 J. M. McPherson and J. R. Ranger-Moore, *Evolution on a dancing landscape: Organizations and networks in dynamic Blau space*, Social Forces **70** (1991), pp. 19-42.

236 K. Medvedyeva, *Characteristic properties of two-dimensional superconductors close to the phase transition in zero magnetic field*, PhD thesis, Umeå University, Umeå, 2003.

237 H. Meyer-Ortmanns, *Phase transitions in quantum chromodynamics*, Rev. Mod. Phys. **68** (1996), pp. 473-598.

238 L. A. Meyers, M. E. J. Newman, M. Martin, and S. Schrag, *Applying network theory to epidemics: Control measures for outbreaks of Mycoplasma pneumoniae*, Emerg. Infect. Dis. **9** (2003), pp. 204-210.

239 S. Miccichè, G. Bonanno, F. Lillo, and R. N. Mantegna, *Degree stability of a minimum spanning tree of price return and volatility*, Physica A **324** (2003), pp. 66-73.

240 S. Milgram, *The small world problem*, Psycol. Today **2** (1967), pp. 60-67.

241 M. Mitzenmacher, *A brief history of generative models for power law and lognormal distributions*. to appear in Internet Mathematics.

242 M. S. Mizruchi, *The American Corporate Network, 1904-1974*, Sage, Beverly Hills, 1982.

243 M. Molloy and B. Reed, *A critical point for random graphs with a given degree sequence*, Random Struct. Algo. **6** (1995), pp. 161-180.

244 ——, *The size of the giant component of a random graph with a given degree sequence*, Combin. Probab. Comput. **7** (1998), pp. 295-305.

245 J. M. Montoya and R. V. Solé, *Small world patterns in food webs*, J. Theor. Bio. **214** (2002), pp. 405-412.

246 J. Moody, *Race, school integration, and friendship segregation in America*, Am. J. Sociol. **107** (2001), pp. 679-716.

247 C. Moore and M. E. J. Newman, *Epidemics and percolation in small-world networks*, Phys. Rev. E **61** (2000), pp. 5678-5682.

248 J. L. Moreno, *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy, and Sociodrama*, Nervous and Mental Disease Publishing Co., Washington, 1934.

249 Y. Moreno, J. B. Gómez, and A. F. Pacheco, *Instability of scale-free networks under node-breaking avalanches*, Europhys. Lett. **58** (2002), pp. 630-636.

250 Y. Moreno and A. Vázquez, *Disease spreading in structured scale-free networks*, Eur. Phys. J. B **31** (2003), pp. 265-271.

251 A. E. Motter, A. P. de Moura, Y.-C. Lai, and P. Dasgupta, *Topology of the conceptual network of language*, Phys. Rev. E **65** (2002), art. no. 065102.

252 A. E. Motter, T. Nishikawa, and Y.-C. Lai, *Large-scale structural organization of social networks*, Phys. Rev. E **68** (2003), art. no. 036105.

253 J. Myrdal, *Report from a Chinese village*, Pantheon, New York, 1965.

254 K. Nagel, *Particle hopping models and traffic flow theory*, la-ur 95-2908, Los Alamos National Laboratory, 1996.

255 M. E. J. Newman, *Fast algorithm for detecting community structure in networks*. e-print cond-mat/0309508.

256 ——, *A measure of betweenness centrality based on random walks*. e-print cond-mat/0309045.

257 ——, *Clustering and preferential attachment in growing networks*, Phys. Rev. E **64** (2001), art. no. 025102.

258 ——, *Scientific collaboration networks. I. Network construction and fundamental results*, Phys. Rev. E **64** (2001), art. no. 016131.

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

260 ——, *The structure of scientific collaboration networks*, Proc. Natl. Acad. Sci. USA **98** (2001), pp. 404-409.

261 ——, *Assortative mixing in networks*, Phys. Rev. Lett. **89** (2002), art. no. 208701.

262 ——, *Spread of epidemic disease on networks*, Phys. Rev. E **66** (2002), art. no. 016128.

263 ——, *Mixing patterns in networks*, Phys. Rev. E **67** (2003), art. no. 026126.

264 ——, *Random graphs as models of networks*, in Handbook of Graphs and Networks, S. Bornholdt and H. G. Schuster, eds., Wiley-VCH, Berlin, 2003, pp. 35-68.

265 ——, *The structure and function of complex networks*, SIAM Rev. **45** (2003), pp. 167-256.

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

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

268 M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Random graphs with arbitrary degree distributions and their applications*, Phys. Rev. E **64** (2001), art. no. 026118.

269 D. M. Nicol and R. M. Fujimoto, *Parallel simulation today*, Ann. Oper. Res. **53** (1994), pp. 249-286.

270 M. Nowak and K. Sigmund, *A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game*, Nature **364** (1992), pp. 56-58.

271 M. A. Nowak and R. M. May, *Evolutionary games and spatial chaos*, Nature **359** (1992), pp. 826-829.

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

273 R. L. Nwankwo and H. A. Regis, *The perception of transcendent interest: Communication and identification in an emergent community*, Int. J. Intercult. Rel. **7** (1983), pp. 361-375.

274 P. Olsson, *Vortex fluctuations in superconductors*, PhD thesis, Umeå University, Umeå, 1992.

275 A. Orlowski, *Google to fix blog noise problem*, The Register **6** (2003).

276 C. E. Osgood and P. H. Tannenbaum, *The principle of congruity in the prediction of attitude change*, Psych. Rev. **62** (1955), pp. 42-55.

277 J. F. Padgett and C. K. Ansell, *Robust action and the rise of the Medici, 1400-1434*, Amer. J. Sociol. **98** (1993), pp. 1259-1319.

278 L. Page, S. Brin, R. Motwani, and T. Winograd, *The Pagerank citation ranking: Bringing order to the Web*, tech. rep., Stanford University, 1998.

279 S. Parsons, P. J. Gmytrasiewicz, and M. Woolridge, eds., *Game Theory and Decision Theory in Agent-Based Systems (Multiagent Systems, Artificial Societies, and Simulated Organizations)*, Kluwer Academic Publishers, New York, 2002.

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

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

282 S. L. Pimm, *Food Webs*, University of Chicago Press, Chicago, 2 ed., 2002.

283 J. J. Potterat, L. Phillips-Plummer, S. Q. Muth, R. B. Rothenberg, D. E. Woodhouse, T. S. Maldonado-Long, H. P. Zimmerman, and J. B. Muth, *Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs*, Sexually Transmitted Infections **78** (2002), pp. 159-163.

284 D. A. Prentice, D. T. Miller, and J. R. Lightdale, *Asymmetries in attachments to groups and to their members: Distinguishing between common-identity and common-bond groups*, Pers. Soc. Psych. Bull. **20** (1994), pp. 484-493.

285 K. Pugliesi and S. L. Shook, *Gender ethnicity and network characteristics: Variation in social support resources*, Sex Roles **38** (1998), pp. 215-238.

286 F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, *Defining and identifying communities in networks*, Proc. Natl. Acad. Sci. USA **101** (2004), pp. 2658-2663.

287 A. Rapoport, *Spread of information through a population with a sociostructural bias: I. Assumption of transitivity*, Bull. Math. Biophys. **15** (1953), pp. 523-533.

288 A. Rapoport and A. M. Chammah, *Prisoner's dilemma*, University of Michigan Press, Ann Arbor, 1965.

289 E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, *Hierarchical organization of modularity in metabolic networks*, Science **297** (2002), pp. 1553-1555.

290 P. Resnick and H. R. Varian, *Recommender systems*, Comm. ACM **40** (1997), pp. 56-58.

291 H. Rheingold, *Smart mobs: The next social revolution*, Perseus, Cambridge, MA, 2002.

292 S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, *Identifying, understanding, and analyzing critical infrastructure interdependencies*, IEEE Control Syst. Mag. **21** (2001), pp. 11-25.

293 C. S. Riolo, J. S. Koopman, and S. E. Chick, *Methods and measures for the description of epidemiological contact networks*, J. Urban Health **78** (2001), pp. 446-456.

294 M. Ripeanu, I. Foster, and A. Iamnitchi, *Mapping the Gnutella network: Properties of large-scale peer-to-peer networks and implications for system design*, IEEE Internet Comput. **6** (2002), pp. 50-57.

295 J. M. Roberts Jr., *Simple methods for simulating sociomatrices with given marginal totals*, Soc. Netw. **22** (2000), pp. 273-283.

296 F. J. Roethlisberger and W. J. Dickson, *Management and the Worker*, Harvard University Press, Cambridge, MA, 1939.

297 D. Ronfeldt, *Social science at 190 MPH on NASCAR's biggest superspeedways*, First Monday **5** (2000).

298 R. Ross, *The Prevention of Malaria*, Murray, London, 1909.

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

300 G. Sabidussi, *The centrality index of a graph*, Psychometrika **31** (1966), pp. 581-603.

301 D. S. Sade, *Sociometrics of Macaca mulatta I: Linkages and cliques in grooming matrices*, Folia Primatol. **17** (1972), pp. 196-223.

302 L. M. Sander, C. P. Warren, I. Sokolov, C. Simon, and J. Koopman, *Percolation on disordered networks as a model for epidemics*, Math. Biosci. **180** (2002), pp. 293-305.

303 T. C. Schelling, *Micromotives and macrobehavior*, W. W. Norton & Company, New York, 1978.

304 M. Schilling, *The encyclopedia of Japanese pop-culture*, Weatherhill, New York, 1997.

305 K. M. Schipper, *The Taoist Body*, University of California Press, Berkeley, 1993.

306 S. Schuster, T. Dandekar, and D. Fell, *Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering*, Trends Biotech. **17** (1999), pp. 53-60.

307 S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar, *Exploring the pathway structure of metabolism: Decomposition into subnetworks and application to Mycoplasma pneumoniae*, Bioinformatics **18** (2002), pp. 351-361.

308 J. Scott, *Social Network Analysis: A Handbook*, Sage, London, 2 ed., 2000.

309 K. A. Seaton and L. M. Hackett, *Stations, trains and small-world networks*. e-print cond-mat/0311254.

310 P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna, *Small-world properties of the Indian railway network*, Phys. Rev. E **67** (2003), art. no. 036106.

311 U. Shardanand and P. Maes, *Social information filtering: Algorithms for automating word of mouth*, in Proceedings of ACM Conference on Human Factors and Computing Systems, New York, 1995, Association of Computing Machinery, pp. 210-217.

312 S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, *Network motifs in the transcriptional regulation network of Escherichia coli*, Nature Genetics **31** (2002), pp. 64-68.

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

314 B. Skyrms and R. Freemantle, *A dynamic model of social network formation*, Proc. Natl. Acad. Sci. USA **97** (2000), pp. 9340-9346.

315 A. Smith, *The Wealth of Nations*, Modern Library, New York, 1937.

316 R. Smith, *Instant messaging as a scale-free network*. e-print cond-mat/0206378, June 2002.

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

318 ——, *Markov chain Monte Carlo estimation of exponential random graph models*, J. Soc. Struct. **2** (2002).

319 R. V. Solé and J. M. Montoya, *Complexity and fragility in ecological networks*, Proc. Roy. Soc. Lond. B **268** (2001), pp. 2039-2045.

320 R. V. Solé, R. Pastor-Satorras, E. Smith, and T. B. Kepler, *A model of large-scale proteome evolution*, Adv. Complex Systems **5** (2002), pp. 43-54.

321 R. Solomonoff and A. Rapoport, *Connectivity of random nets*, Bull. Math. Biophysics **13** (1951), pp. 107-117.

322 D. Sornette, *Critical Phenomena in Natural Sciences*, Springer, Berlin, 2000.

323 A. Soubeyran and S. Weber, *District formation and local social capital: A (tacit) co-opetition approach*, J. Urban Econ. **52** (2002), pp. 65-92.

324 A. Soulier and T. Halpin-Healy, *The dynamics of multidimensional secession: Fixed points and ideological condensation*, Phys. Rev. Lett. **90** (2003), art. no. 258103.

325 O. Sporns, G. Tononi, and G. M. Edelman, *Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices*, Cerebral Cortex **10** (2000), pp. 127-141.

326 H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, Oxford, 1971.

327 D. Stauffer and A. Aharony, *Introduction to percolation theory*, Taylor and Francis, London, 2 ed., 1992.

328 K. Stephenson and M. Zelen, *Rethinking centrality: Methods and applications*, Soc. Netw. **11** (1989), pp. 1-37.

329 M. Steyvers, J. B. Tenenbaum, E.-J. Wagenmakers, and B. Blum, *Inferring causal networks from observations and interventions*, Cognitive Science **27** (2003), pp. 453-489.

330 D. Strauss, *On a general class of models for interaction*, SIAM Rev. **28** (1986), pp. 513-527.

331 L. H. Sullivan, *Kindergarten Chats (revised 1918) and Other Writings*, Wittenborn Art Books, New York, 1947, ch. The Tall Office Building Artistically Considered, pp. 202-213.

332 L. Susskind, *Lattice fermions*, Phys. Rev. D **16** (1977), pp. 3031-3039.

333 R. Swedberg, *Sociology and game theory: Contemporary and historical perspectives*, Theory and Society **30** (2001), pp. 301-335.

334 G. Szabó, M. Alava, and J. Kertész, *Structural transitions in scale-free networks*, Phys. Rev. E **67** (2003), art. no. 056102.

335 M. Takayasu, H. Takayasu, and T. Sato, *Critical behavior and 1/f noise in information traffic*, Physica A **233** (1996), pp. 824-834.

336 K. Taro Greenfeld, *Speed tribes: Days and nights with Japan's next generation*, Perennial, New York, 1995.

337 M. Tomita, *Whole cell simulation: A grand challenge of the 21st century*, Trends Biotech. **19** (2001), pp. 205-210.

338 M. Tomita, K. Hashimoto, K. Takahashi, T. Simon Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J. C. Venter, and C. A. Hutchison III, *E-cell: Software environment for whole-cell simulation*, Bioinformatics **15** (1999), pp. 72-84.

339 J. Toner and Y. Tu, *Long-range order in a two-dimensional dynamical XY model: How birds fly together*, Phys. Rev. Lett. **75** (1995), pp. 4326-4329.

340 Z. Toroczkai, G. Korniss, M. A. Novotny, and H. Guclu, *Virtual time horizon control via communication network design*, in Computational Complexity and Statistical Physics, A. Percus, G. Istrate, and C. Moore, eds., Santa Fe Institute Studies in the Sciences of Complexity, Oxford University Press, Oxford, 2003.

341 J. Travers and S. Milgram, *An experimental study of the small world problem*, Sociometry **32** (1969), pp. 425-443.

342 A. Trusina, S. Maslov, P. Minnhagen, and K. Sneppen, *Hierarchy and anti-hierarchy in real and scale free networks*. e-print cond-mat/0308339.

343 A. Tucker, *Applied Combinatorics*, John Wiley, New York, 3 ed., 1995.

344 P. E. Turner and L. Chao, *Prisoner's dilemma in an RNA virus*, Nature **398** (1999), pp. 441-443.

345 S. Valverde, R. Ferrer i Cancho, and R. V. Solé, *Scale-free networks from optimal design*, Europhys. Lett. **60** (2002), pp. 512-517.

346 A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, *Modeling of protein interaction networks*, Complexus **1** (2003), pp. 38-44.

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

348 J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*, Princeton University Press, Princeton, 2 ed., 1947.

349 M. M. Waldrop, *Complexity: The emerging science at the edge of order and chaos*, Touchstone, New York, 1992.

350 S. Wasserman, *Random directed graph distributions and the triad census in social networks*, J. Math. Sociol. **5** (1977), pp. 61-86.

351 S. Wasserman and K. Faust, *Social network analysis: Methods and applications*, Cambridge University Press, Cambridge, 1994.

352 S. Wasserman and P. Pattison, *Logit models and logistic regressions for social networks: An introduction to Markov random graphs and p**, Psychometrika **61** (1996), pp. 401-426.

353 D. J. Watts, *Networks, dynamics, and the small world phenomenon*, Am. J. Sociol. **105** (1999), pp. 493-592.

354 ——, *Small Worlds: The dynamics of networks between order and Randomness*, Princeton University Press, Princeton, 1999.

355 ——, *A simple model of global cascades on random networks*, Proc. Natl. Acad. Sci. USA **99** (2002), pp. 5766-5771.

356 ——, *Six Degrees: The Science of a Connected Age*, W. W. Norton & Co., New York, 2003.

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

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

359 G. B. West, J. H. Brown, and B. J. Enquist, *A general model for the structure and allometry of plant vascular systems*, Nature **400** (1999), pp. 664-667.

360 D. R. White and F. Harary, *The cohesiveness of blocks in social networks: Node connectivity and conditional density*, Sociol. Methodol. **31** (2001), pp. 305-359.

361 D. R. White and K. P. Reitz, *Measuring role distance: Stuctural, regular and relational equivalence*. unpublished manuscript, University of California, 1985.

362 H. C. White, *Where do markets come from?*, Am. J. Sociol. **87** (1981), pp. 517-547.

363 H. C. White, S. A. Boorman, and R. L. Breiger, *Social structure from multiple networks. I. Blockmodels of roles and positions*, Am. J. Sociol. **81** (1976), pp. 730-779.

364 J. G. White, E. Southgate, J. N. Thompson, and S. Brenner, *The structure of the nervous system of the nematode C. Elegans*, Philos. Trans. Roy. Soc. Lond. **314** (1986), pp. 1-340.

365 I. Wiklund, *Social networks and proffiiciency in Swedish: A study of bilingual adolescents in both mono- and multicultural contexts in Sweden*, PhD thesis, Stockholm University, Stockholm, 2002.

366 H. S. Wilf, *Generatingfunctionology*, Academic Press, Boston, 1990.

367 R. J. Williams, E. L. Berlow, J. A. Dunne, A.-L. Barabási, and N. D. Martinez, *Two degrees of separation in complex food webs*, Proc. Natl. Acad. Sci. USA **99** (2002), pp. 12913-12916.

368 F. Wu and B. A. Huberman, *Finding communities in linear time: A physics approach*. e-print cond-mat/0310600.

369 I. Yang, H. Jeong, B. Kahng, and A.-L. Barabási, *Emerging behavior in electronic bidding*, Phys. Rev. E **68** (2003), art. no. 016102.

370 W. Zachary, *An information flow model for conflict and fission in small groups*, J. Anthropol. Res. **33** (1977), pp. 452-473.

371 X. Zhang and G. Li, *Does guanxi matter to nonfarm employment?*, J. Compar. Econ. **31** (2003), pp. 315-331.

372 M. G. Zimmerman and V. M. Eguíluz, *Cooperation, social networks and the emergence of leadership in a prisoners dilemma with adaptive local interactions*. submitted to J. Econ. Dynam. Control, December 2001.