

# The Structure and Function of Complex Networks\*

M. E. J. Newman<sup>†</sup>

**Abstract.** Inspired by empirical studies of networked systems such as the Internet, social networks, and biological networks, researchers have in recent years developed a variety of techniques and models to help us understand or predict the behavior of these systems. Here we review developments in this field, including such concepts as the small-world effect, degree distributions, clustering, network correlations, random graph models, models of network growth and preferential attachment, and dynamical processes taking place on networks.

**Key words.** networks, graph theory, complex systems, computer networks, social networks, random graphs, percolation theory

**AMS subject classifications.** 05C75, 05C90, 94C15

**PII.** S0036144503424804

## Contents.

<b>I</b>	<b>Introduction</b>	<b>168</b>
1.1	Types of Networks . . . . .	171
1.2	Other Resources . . . . .	172
1.3	Outline of the Review . . . . .	174
<b>2</b>	<b>Networks in the Real World</b>	<b>174</b>
2.1	Social Networks . . . . .	174
2.2	Information Networks . . . . .	176
2.3	Technological Networks . . . . .	178
2.4	Biological Networks . . . . .	179
<b>3</b>	<b>Properties of Networks</b>	<b>180</b>
3.1	The Small-World Effect . . . . .	180
3.2	Transitivity or Clustering . . . . .	183
3.3	Degree Distributions . . . . .	185
3.3.1	Scale-Free Networks . . . . .	186
3.3.2	Maximum Degree . . . . .	188
3.4	Network Resilience . . . . .	189
3.5	Mixing Patterns . . . . .	190

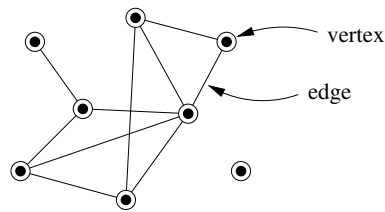
\*Received by the editors January 20, 2003; accepted for publication (in revised form) March 17, 2003; published electronically May 2, 2003. This work was supported in part by the U.S. National Science Foundation under grants DMS-0109086 and DMS-0234188 and by the James S. McDonnell Foundation and the Santa Fe Institute.

<http://www.siam.org/journals/sirev/45-2/42480.html>

<sup>†</sup>Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109 (mejn@umich.edu).

3.6	Degree Correlations . . . . .	192
3.7	Community Structure . . . . .	193
3.8	Network Navigation . . . . .	195
3.9	Other Network Properties . . . . .	196
<b>4</b>	<b>Random Graphs</b>	<b>196</b>
4.1	Poisson Random Graphs . . . . .	197
4.2	Generalized Random Graphs . . . . .	200
4.2.1	The Configuration Model . . . . .	200
4.2.2	Example: Power-Law Degree Distribution . . . . .	202
4.2.3	Directed Graphs . . . . .	203
4.2.4	Bipartite Graphs . . . . .	204
4.2.5	Degree Correlations . . . . .	205
<b>5</b>	<b>Exponential Random Graphs and Markov Graphs</b>	<b>206</b>
<b>6</b>	<b>The Small-World Model</b>	<b>208</b>
6.1	Clustering Coefficient . . . . .	210
6.2	Degree Distribution . . . . .	210
6.3	Average Path Length . . . . .	211
<b>7</b>	<b>Models of Network Growth</b>	<b>212</b>
7.1	Price's Model . . . . .	213
7.2	The Model of Barabási and Albert . . . . .	215
7.3	Generalizations of the Model of Barabási and Albert . . . . .	219
7.4	Other Growth Models . . . . .	221
7.5	Vertex Copying Models . . . . .	223
<b>8</b>	<b>Processes Taking Place on Networks</b>	<b>224</b>
8.1	Percolation Theory and Network Resilience . . . . .	225
8.2	Epidemiological Processes . . . . .	229
8.2.1	The SIR Model . . . . .	229
8.2.2	The SIS Model . . . . .	232
8.3	Search on Networks . . . . .	233
8.3.1	Exhaustive Network Search . . . . .	234
8.3.2	Guided Network Search . . . . .	235
8.3.3	Network Navigation . . . . .	236
8.4	Phase Transitions on Networks . . . . .	238
8.5	Other Processes on Networks . . . . .	239
<b>9</b>	<b>Summary and Directions for Future Research</b>	<b>240</b>

**1. Introduction.** A *network* is a set of items, which we will call *vertices* or sometimes nodes, with connections between them, called *edges* (Figure 1.1). Systems taking the form of networks (also called “graphs” in much of the mathematical literature) abound in the world. Examples include the Internet, the World Wide Web, social networks of acquaintance or other connections between individuals, organizational networks and networks of business relations between companies, neural networks, metabolic networks, food webs, distribution networks such as blood vessels or



**Fig. 1.1** *A small example network with eight vertices and ten edges.*

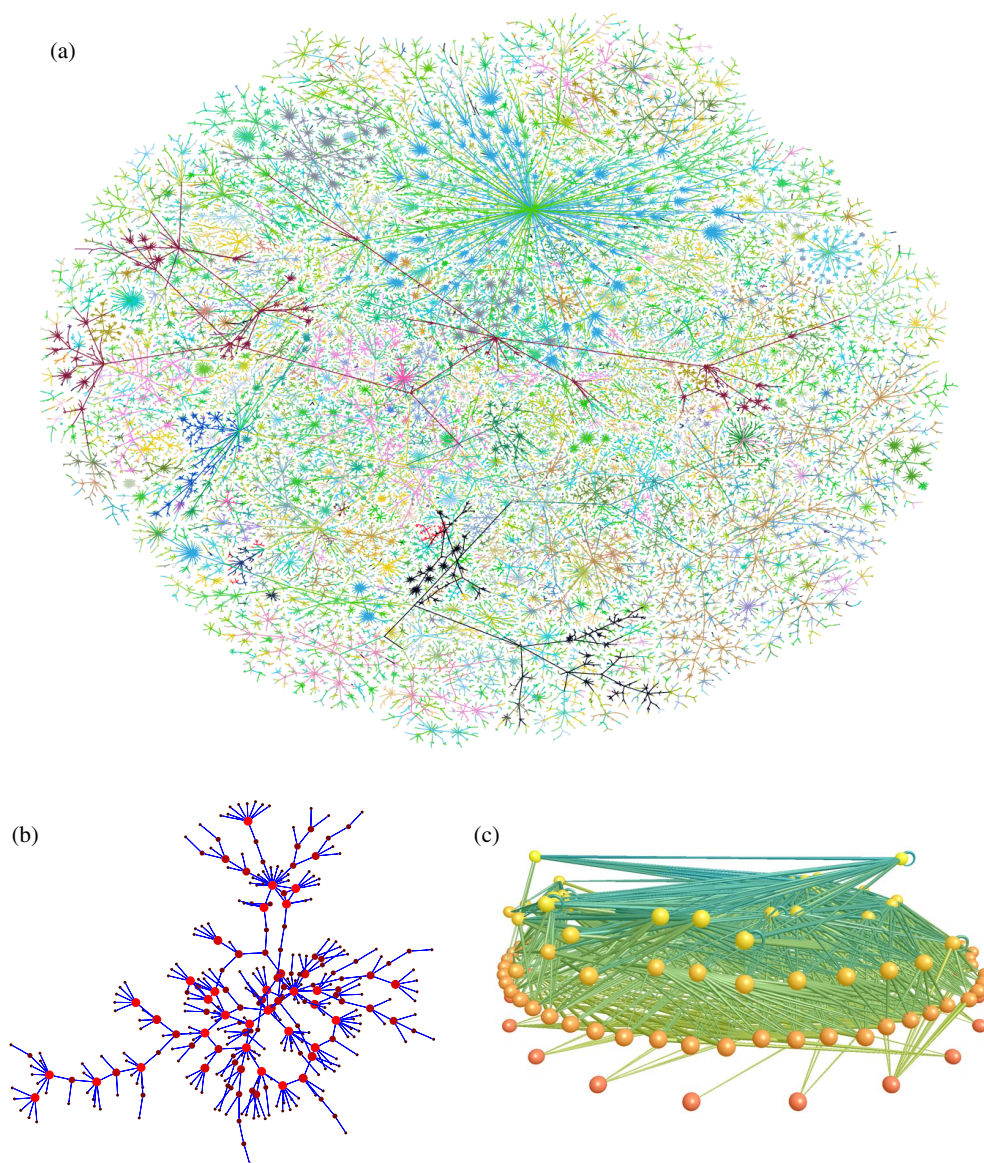
postal delivery routes, networks of citations between papers, and many others (Figure 1.2). This paper reviews recent (and some not-so-recent) work on the structure and function of networked systems such as these.

The study of networks, in the form of mathematical graph theory, is one of the fundamental pillars of discrete mathematics. Euler's celebrated 1735 solution of the Königsberg bridge problem is often cited as the first true proof in the theory of networks, and during the twentieth century graph theory has developed into a substantial body of knowledge.

Networks have also been studied extensively in the social sciences. Already in the 1930s, sociologists realized the importance of the patterns of connection between people to the understanding of the functioning of human society (Figure 1.3). Typical network studies in sociology involve the circulation of questionnaires, asking respondents to detail their interactions with others. One can then use the responses to reconstruct a network in which vertices represent individuals and edges the interactions between them. Typical social network studies address issues of centrality (which individuals are best connected to others or have most influence) and connectivity (whether and how individuals are connected to one another through the network).

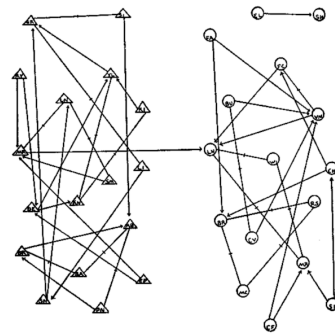
Recent years, however, have witnessed a substantial new movement in network research, with the focus shifting away from the analysis of single small graphs and the properties of individual vertices or edges within such graphs to consideration of large-scale statistical properties of graphs. This new approach has been driven largely by the availability of computers and communication networks that allow us to gather and analyze data on a scale far larger than previously possible. Where studies used to look at networks of maybe tens or in extreme cases hundreds of vertices, it is not uncommon now to see networks with millions or even billions of vertices. This change of scale forces upon us a corresponding change in our analytic approach. Many of the questions that might previously have been asked in studies of small networks are simply not useful in much larger networks. A social network analyst might have asked, "Which vertex in this network would prove most crucial to the network's connectivity if it were removed?" But such a question has little meaning in most networks of a million vertices—no single vertex in such a network will have much effect at all when removed. On the other hand, one could reasonably ask a question like, "What percentage of vertices need to be removed to substantially affect network connectivity in some given way?" and this type of statistical question has real meaning even in a very large network.

However, there is another reason why our approach to the study of networks has changed in recent years, a reason whose importance should not be underestimated, although it often is. For networks of tens or hundreds of vertices, it is a relatively straightforward matter to draw a picture of the network with actual points and lines,



**Fig. 1.2** Three examples of the kinds of networks that are the topic of this review. (a) A visualization of the network structure of the Internet at the level of “autonomous systems”—local groups of computers each representing hundreds or thousands of machines. Picture by Hal Burch and Bill Cheswick, courtesy of Lumeta Corporation. (b) A social network, in this case of sexual contacts, redrawn from the HIV data of Potterat et al. [341]. (c) A food web of predator-prey interactions between species in a freshwater lake [271]. Picture courtesy of Richard Williams.

and to answer specific questions about network structure by examining this picture. This has been one of the primary methods of network analysts since the field began (see Figure 1.3). The human eye is an analytic tool of remarkable power, and eyeballing pictures of networks is an excellent way to gain an understanding of their

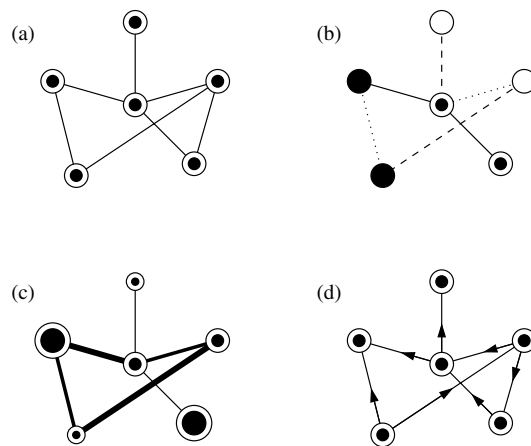


**Fig. 1.3** An early hand-drawn social network from 1934 representing friendships between school children. After Moreno [295]. Reprinted with permission from ASGPP.

structure. With a network of a million or a billion vertices, however, this approach is useless. (See Figure 1.2a for an example of a network that lies at the upper limit of what can usefully be drawn on a piece of paper or computer screen.) One simply cannot draw a meaningful picture of a million vertices, even with modern 3D computer rendering tools, and therefore direct analysis by eye is hopeless. The recent development of statistical methods for quantifying large networks is to a large extent an attempt to find something to play the part played by the eye in the network analysis of the twentieth century. Statistical methods answer the question, “How can I tell what this network looks like, when I can’t actually look at it?”

The body of theory that is the primary focus of this review aims to do three things. First, it aims to find and highlight statistical properties, such as path lengths and degree distributions, that characterize the structure and behavior of networked systems, and to suggest appropriate ways to measure these properties. Second, it aims to create models of networks that can help us to understand the meaning of these properties—how they came to be as they are, and how they interact with one another. Third, it aims to predict what the behavior of networked systems will be on the basis of measured structural properties and the local rules governing individual vertices. How, for example, will network structure affect traffic on the Internet, or the performance of a Web search engine, or the dynamics of social or biological systems? As we will see, the scientific community has, by drawing on ideas from a broad variety of disciplines, made an excellent start on the first two of these aims, the characterization and modeling of network structure. Studies of the effects of structure on system behavior on the other hand are still in their infancy. It remains to be seen what the crucial theoretical developments will be in this area.

**1.1. Types of Networks.** A set of vertices joined by edges is only the simplest type of network; there are many ways in which networks may be more complex than this (Figure 1.4). For instance, there may be more than one different type of vertex in a network, or more than one different type of edge. And vertices or edges may have a variety of properties, numerical or otherwise, associated with them. Taking the example of a social network of people, the vertices may represent men or women, people of different nationalities, locations, ages, incomes, or many other things. Edges may represent friendship, but they could also represent animosity, or professional acquaintance, or geographical proximity. They can carry weights, representing, say,



**Fig. 1.4** Examples of various types of networks: (a) an undirected network with only a single type of vertex and a single type of edge; (b) a network with a number of discrete vertex and edge types; (c) a network with varying vertex and edge weights; (d) a directed network in which each edge has a direction.

how well two people know each other. They can also be directed, pointing in only one direction. Graphs composed of directed edges are themselves called directed graphs or sometimes *digraphs*, for short. A graph representing telephone calls or email messages between individuals would be directed, since each message only goes in one direction. Directed graphs can be either cyclic, meaning they contain closed loops of edges, or acyclic, meaning they do not. Some networks, such as food webs, are approximately but not perfectly acyclic.

One can also have *hyperedges*—edges that join more than two vertices together. Graphs containing such edges are called *hypergraphs*. Hyperedges could be used to indicate family ties in a social network for example— $n$  individuals connected to each other by virtue of belonging to the same immediate family could be represented by an  $n$ -edge joining them. Graphs may also be naturally partitioned in various ways. We will see a number of examples in this review of *bipartite graphs*: graphs that contain vertices of two distinct types, with edges running only between unlike types. So-called *affiliation networks* in which people are joined together by common membership of groups take this form, the two types of vertices representing the people and the groups. Graphs may also evolve over time, with vertices or edges appearing or disappearing, or values defined on those vertices and edges changing. And there are many other levels of sophistication one can add. The study of networks is by no means a complete science yet, and many of the possibilities have yet to be explored in depth, but we will see examples of at least some of the variations described here in the work reviewed in this paper.

The jargon of the study of networks is unfortunately confused by differing usages among investigators from different fields. To avoid (or at least reduce) confusion, we give in Box 1 a short glossary of terms as they are used in this paper.

**1.2. Other Resources.** A number of other reviews of this area have appeared recently, which the reader may wish to consult. Albert and Barabási [13] and Dorogovtsev and Mendes [120] have given extensive pedagogical reviews focusing on the physics literature. Both devote the larger part of their attention to the models of

*Vertex (pl. vertices):* The fundamental unit of a network, also called a site (physics), a node (computer science), or an actor (sociology).

*Edge:* The line connecting two vertices. Also called a bond (physics), a link (computer science), or a tie (sociology).

*Directed/undirected:* An edge is directed if it runs in only one direction (such as a one-way road between two points), and undirected if it runs in both directions. Directed edges, which are sometimes called *arcs*, can be thought of as sporting arrows indicating their orientation. A graph is directed if all of its edges are directed. An undirected graph can be represented by a directed one having two edges between each pair of connected vertices, one in each direction.

*Degree:* The number of edges connected to a vertex. Note that the degree is not necessarily equal to the number of vertices adjacent to a vertex, since there may be more than one edge between any two vertices. In a few recent articles, the degree is referred to as the “connectivity” of a vertex, but we avoid this usage because the word connectivity already has another meaning in graph theory. A directed graph has both an in-degree and an out-degree for each vertex, which are the numbers of incoming and outgoing edges respectively.

*Component:* The component to which a vertex belongs is that set of vertices that can be reached from it by paths running along edges of the graph. In a directed graph a vertex has both an in-component and an out-component, which are the sets of vertices from which the vertex can be reached and which can be reached from it.

*Geodesic path:* A geodesic path is the shortest path through the network from one vertex to another. Note that there may be and often is more than one geodesic path between two vertices.

*Diameter:* The diameter of a network is the length (in number of edges) of the longest geodesic path between any two vertices. A few authors have also used this term to mean the *average* geodesic distance in a graph, although strictly the two quantities are quite distinct.

---

**Box 1** *A short glossary of terms.*

---

growing graphs that we describe in section 7. Shorter reviews taking other viewpoints have been given by Newman [308] and Hayes [188, 189], who both concentrate on the so-called small-world models (see section 6), and by Strogatz [386], who includes an interesting discussion of the behavior of dynamical systems on networks.

A number of books also make worthwhile reading. Dorogovtsev and Mendes [122] have expanded their above-mentioned review into a book, which again focuses on models of growing graphs. The edited volumes by Bornholdt and Schuster [70] and by Pastor-Satorras and Rubi [329] both contain contributed essays on various topics by leading researchers. Detailed treatments of many of the topics covered in the present work can be found there. The book by Newman, Barabási, and Watts [319] is a collection of previously published papers and also contains some review material by the editors.

Three popular books on the subject of networks merit a mention. Barabási's *Linked* [31] gives a personal account of recent developments in the study of networks, focusing particularly on Barabási's work on scale-free networks. Watts's *Six Degrees* [413] gives a sociologist's view, partly historical, of discoveries old and new. Buchanan's *Nexus* [76] gives an entertaining portrait of the field from the point of view of a science journalist.

Farther afield, there are a variety of books on the study of networks in particular fields. Within graph theory the books by Harary [187] and by Bollobás [62] are widely cited as are, among social network theorists, the books by Wasserman and Faust [408] and by Scott [362]. The book by Ahuja, Magnanti, and Orlin [7] is a useful source for information on network algorithms.

**1.3. Outline of the Review.** The outline of this paper is as follows. In section 2 we describe empirical studies of the structure of networks, including social networks, information networks, technological networks, and biological networks. In section 3 we describe some of the common properties that are observed in many of these networks, how they are measured, and why they are believed to be important for the functioning of networked systems. Sections 4 to 7 form the heart of the review. They describe work on the mathematical modeling of networks, including random graph models and their generalizations, exponential random graphs,  $p^*$  models and Markov graphs, the small-world model and its variations, and models of growing graphs including preferential attachment models and their many variations. In section 8 we discuss the progress, such as it is, that has been made on the study of processes taking place on networks, including epidemic processes, network failure, models displaying phase transitions, and dynamical systems like random Boolean networks and cellular automata. In section 9 we give our conclusions and point to directions for future research.

**2. Networks in the Real World.** In this section we look at what is known about the structure of networks of different types. Recent work on the mathematics of networks has been driven largely by observations of the properties of actual networks and attempts to model them, so network data are the obvious starting point for a review such as this. It also makes sense to examine simultaneously data from different kinds of networks. One of the principal thrusts of recent work in this area, inspired particularly by a groundbreaking 1998 paper by Watts and Strogatz [415], has been the comparative study of networks from different branches of science, with emphasis on properties that are common to many of them and the mathematical developments that mirror those properties. We here divide our summary into four loose categories of networks: social networks, information networks, technological networks, and biological networks.

**2.1. Social Networks.** A social network is a set of people or groups of people with some pattern of contacts or interactions between them [362, 408]. The patterns of friendships between individuals [295, 347], business relationships between companies [268, 285], and intermarriages between families [326] are all examples of networks that have been studied in the past.<sup>1</sup> Of the academic disciplines, the social sciences have the longest history of the substantial quantitative study of real-world networks [162, 362]. Of particular note among the early works on the subject are the following: Moreno's networks of friendships within small groups [295], of which Figure 1.3 is an example; the so-called southern women study of Davis, Gardner, and Gardner [103], which focused on the social circles of women in an unnamed city in the American south in 1936; the study by Elton Mayo and colleagues of social networks of factory workers in the late 1930s in Chicago [356]; the mathematical models of Rapoport [345], who was one of the first theorists, perhaps *the* first, to stress

<sup>1</sup>Occasionally social networks of animals have been investigated also, such as dolphins [96], not to mention networks of fictional characters, such as the protagonists of Tolstoy's *Anna Karenina* [243] or Marvel Comics superheroes [10].



the importance of the degree distribution in networks of all kinds, not just social networks; and the studies of friendship networks of school children by Rapoport and others [347, 149]. In more recent years, studies of business communities [167, 168, 268] and of patterns of sexual contacts [45, 217, 242, 265] have attracted particular attention.

Another important set of experiments are the famous “small-world” experiments of Milgram [282, 392]. No actual networks were reconstructed in these experiments, but nonetheless they tell us about network structure. The experiments probed the distribution of path lengths in an acquaintance network by asking participants to pass a letter<sup>2</sup> to one of their first-name acquaintances in an attempt to get it to an assigned target individual. Most of the letters in the experiment were lost, but about a quarter reached the target and passed on average through the hands of only about six people in doing so. This experiment was the origin of the popular concept of the “six degrees of separation,” although that phrase did not appear in Milgram’s writing, being coined some decades later by Guare [182]. A brief but useful early review of Milgram’s work and work stemming from it was given by Garfield [169].

Traditional social network studies often suffer from problems of inaccuracy, subjectivity, and small sample size. With the exception of a few ingenious indirect studies such as Milgram’s, data collection is usually carried out by querying participants directly using questionnaires or interviews. Such methods are labor-intensive and therefore limit the size of the network that can be observed. Survey data are, moreover, influenced by subjective biases on the part of respondents; how one respondent defines a friend, for example, could be quite different from how another does. Although much effort is put into eliminating possible sources of inconsistency, it is generally accepted that there are large and essentially uncontrolled errors in most of these studies. A review of the issues has been given by Marsden [270].

Because of these problems many researchers have turned to other methods for probing social networks. One source of copious and relatively reliable data is collaboration networks. These are typically affiliation networks in which participants collaborate in groups of one kind or another, and links between pairs of individuals are established by common group membership. A classic, though rather frivolous, example of such a network is the collaboration network of film actors, which is thoroughly documented in the online Internet Movie Database.<sup>3</sup> In this network, actors collaborate in films and two actors are considered connected if they have appeared in a film together. Statistical properties of this network have been analyzed by a number of authors [4, 20, 322, 415]. Other examples of networks of this type are networks of company directors, in which two directors are linked if they belong to the same board of directors [104, 105, 268]; networks of coauthorship among academics, in which individuals are linked if they have coauthored one or more papers [36, 43, 68, 107, 181, 278, 291, 310, 311, 312]; and coappearance networks, in which individuals are linked by mention in the same context, particularly on Web pages [3, 226] or in newspaper articles [99] (see Figure 1.2b).

Another source of reliable data about personal connections between people is communication records of certain kinds. For example, one could construct a network in which each (directed) edge between two people represented a letter or package sent by mail from one to the other. No study of such a network has been published as far as we are aware, but some similar things have. Aiello, Chung, and Lu [8, 9] have analyzed

<sup>2</sup>Actually a folder containing several documents.

<sup>3</sup><http://www.imdb.com/>

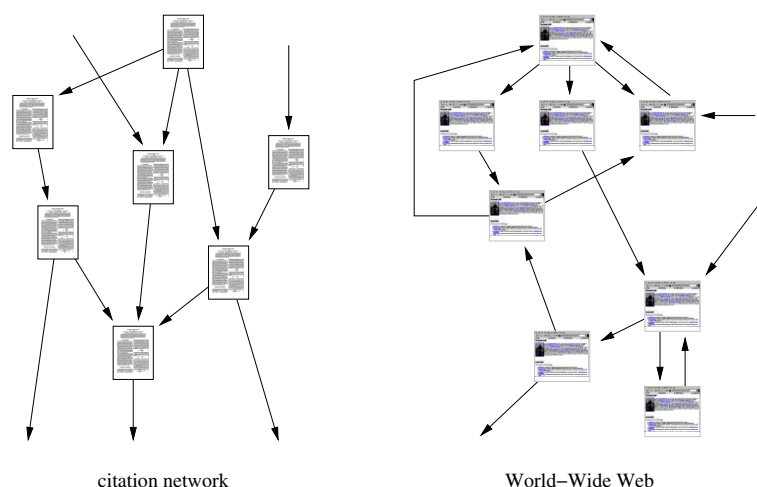
a network of telephone calls made over the AT&T long-distance network on a single day. The vertices of this network represent telephone numbers and the directed edges calls from one number to another. Even for just a single day this graph is enormous, having about 50 million vertices, one of the largest graphs yet studied after the graph of the World Wide Web. Ebel, Mielsch, and Bornholdt [136] have reconstructed the pattern of email communications between five thousand students at Kiel University from logs maintained by email servers. In this network the vertices represent email addresses and directed edges represent a message passing from one address to another. Email networks have also been studied by Newman, Forrest, and Balthrop [320] and by Guimerà et al. [184], and similar networks have been constructed for an “instant messaging” system by Smith [370], and for an Internet community Web site by Holme, Edling, and Liljeros [195]. Dodds, Muhamad, and Watts [110] have carried out an email version of Milgram’s small-world experiment in which participants were asked to forward an email message to one of their friends in an effort to get the message ultimately to some chosen target individual. Response rates for the experiment were quite low, but a few hundred completed chains of messages were recorded, enough to allow various statistical analyses.

**2.2. Information Networks.** Our second network category is what we will call *information networks* (also sometimes called “knowledge networks”). The classic example of an information network is the network of citations between academic papers [138]. Most learned articles cite previous works by others on related topics. These citations form a network in which the vertices are articles and a directed edge from article A to article B indicates that A cites B. The structure of the citation network then reflects the structure of the information stored at its vertices, hence the term “information network,” although certainly there are social aspects to the citation patterns of papers too [419].

Citation networks are acyclic (see section 1.1) because papers can only cite other papers that have already been written, not those that have yet to be written. Thus all edges in the network point backwards in time, making closed loops impossible, or at least extremely rare (see Figure 2.1).

As an object of scientific study, citation networks have a great advantage in the copious and accurate data available for them. Quantitative study of publication patterns stretches back at least as far as Alfred Lotka’s groundbreaking 1926 discovery of the so-called Law of Scientific Productivity, which states that the distribution of the numbers of papers written by individual scientists follows a power law. That is, the number of scientists who have written  $k$  papers falls off as  $k^{-\alpha}$  for some constant  $\alpha$ . (In fact, this result extends to the arts and humanities as well.) The first serious work on citation patterns was conducted in the 1960s as large citation databases became available through the work of Eugene Garfield and other pioneers in the field of bibliometrics. The network formed by citations was discussed in an early paper by Price [342], in which, among other things, the author points out for the first time that both the in- and out-degree distributions of the network follow power laws, a far-reaching discovery which we discuss further in section 3.3. Many other studies of citation networks have been performed since then, using the ever better resources available in citation databases. Of particular note are the studies by Seglen [363] and Redner [350].<sup>4</sup>

<sup>4</sup>An interesting development in the study of citation patterns has been the arrival of automatic citation “crawlers” that construct citation networks from online papers. Examples include Cite-seer (<http://citeseer.nj.nec.com/>), SPIRES (<http://www.slac.stanford.edu/spires/hep/>), and Cite-base (<http://citebase.eprints.org/>).



**Fig. 2.1** *The two best studied information networks. Left: the citation network of academic papers in which the vertices are papers and the directed edges are citations of one paper by another. Since papers can only cite those that came before them (lower down in the figure) the graph is acyclic—it has no closed loops. Right: the World Wide Web, a network of text pages accessible over the Internet, in which the vertices are pages and the directed edges are hyperlinks. There are no constraints on the Web that forbid cycles and hence it is in general cyclic.*

Another very important example of an information network is the World Wide Web, which is a network of Web pages containing information, linked together by hyperlinks from one page to another [202]. The Web should not be confused with the Internet, which is a physical network of computers linked together by optical fiber and other data connections.<sup>5</sup> Unlike a citation network, the World Wide Web is cyclic; there is no natural ordering of sites and no constraints that prevent the appearance of closed loops (Figure 2.1). The Web has been very heavily studied since its first appearance in the early 1990s, with the studies by Albert et al. [14, 34], Kleinberg et al. [240], and Broder et al. [74] being particularly influential. The Web also appears to have power-law in- and out-degree distributions (section 3.3), as well as a variety of other interesting properties [2, 14, 74, 158, 240, 253].

One important point to notice about the Web is that our data about it come from “crawls” of the network, in which Web pages are found by following hyperlinks from other pages [74]. Our picture of the network structure of the World Wide Web is therefore necessarily biased. A page will only be found if another page points to it,<sup>6</sup> and in a crawl that covers only a part of the Web (as all crawls do at present) pages are more likely to be found the more other pages point to them [262]. This suggests, for instance, that our measurements of the fraction of pages with low in-degree might be an underestimate.<sup>7</sup> This behavior contrasts with that of a citation network. A

<sup>5</sup>While the Web is primarily an information network, it, like citation networks, has social aspects to its structure also [3].

<sup>6</sup>This is not always strictly true. Some Web search engines allow the submission of pages by members of the public for inclusion in databases, and such pages need not be the target of links from any other pages. However, such pages also form a very small fraction of all Web pages, and certainly the biases discussed here remain very much present.

<sup>7</sup>The degree distribution for the Web shown in Figure 3.2 falls off slightly at low values of the in-degree, which may perhaps reflect this bias.

paper can appear in the citation indices even if it has never been cited (and in fact a plurality of papers in the indices are never cited).

A few other examples of information networks have been studied to a lesser extent. Jaffe and Trajtenberg [206], for instance, have studied the network of citations between U.S. patents, which is similar in some respects to citations between academic papers. A number of authors have looked at peer-to-peer networks [5, 6, 204], which are virtual networks of computers that allow sharing of files between computer users over local- or wide-area networks. The network of relations between word classes in a thesaurus has been studied by Knuth [243] and more recently by various other authors [233, 383, 303]. This network can be looked upon as an information network—users of a thesaurus “surf” the network from one word to another looking for the particular word that perfectly captures the idea they have in mind. However, it can also be looked at as a conceptual network representing the structure of the language, or possibly even the mental constructs used to represent the language. A number of other semantic word networks have also been investigated [119, 157, 368, 383].

Preference networks provide an example of a bipartite information network. A preference network is a network with two kinds of vertices representing individuals and the objects of their preference, such as books or films, with an edge connecting each individual to the books or films they like. (Preference networks can also be weighted to indicate strength of likes or dislikes.) A widely studied example of a preference network is the *EachMovie* database of film preferences.<sup>8</sup> Networks of this kind form the basis for *collaborative filtering* algorithms and *recommender systems*, which are techniques for predicting new likes or dislikes based on comparison of individuals’ preferences with those of others [175, 351, 366]. Collaborative filtering has found considerable commercial success for product recommendation and targeted advertising, particularly with online retailers. Preference networks can also be thought of as social networks, linking not only people to objects, but also people to other people with similar preferences. This approach has been adopted occasionally in the literature [226].

**2.3. Technological Networks.** Our third class of networks is technological networks, man-made networks designed typically for distribution of some commodity or resource, such as electricity or information. The electric power grid is a good example. This is a network of high-voltage three-phase transmission lines that spans a country or a portion of a country (as opposed to the local low-voltage AC power delivery lines that span individual neighborhoods). Statistical studies of power grids have been made by, for example, Watts and Strogatz [415], Watts [411], and Amaral et al. [20]. Other distribution networks that have been studied include the network of airline routes [20] and networks of roads [220], railways [261, 365], and pedestrian traffic [87]. River networks could be regarded as a naturally occurring form of distribution network (actually a collection network) [111, 269, 352, 355], as could the vascular networks discussed in section 2.4. The telephone network and delivery networks such as those used by the post office or parcel delivery companies also fall into this general category and are presumably studied within the relevant corporations, if not yet by academic researchers. (We distinguish here between the physical telephone network of wires and cables and the network of who calls whom, discussed in section 2.1.) Electronic circuits [155] fall somewhere between distribution and communication networks.

<sup>8</sup><http://research.compaq.com/SRC/eachmovie/>

Another very widely studied technological network is the Internet, i.e., the network of physical connections between computers. Since there is a large and ever-changing number of computers on the Internet, the structure of the network is usually examined at a coarse-grained level, either the level of routers, special-purpose computers on the network that control the movement of data, or “autonomous systems,” which are groups of computers within which networking is handled locally, but between which data flows over the public Internet (Figure 1.2a). The computers at a single company or university would probably form a single autonomous system—autonomous systems often correspond roughly with domain names.

In fact, the network of physical connections on the Internet is not easy to discover since the infrastructure is maintained by many separate organizations. Typically, therefore, researchers reconstruct the network by reasoning from large samples of point-to-point data routes. So-called “traceroute” programs can report the sequence of network nodes that a data packet passes through when traveling between two points, and if we assume an edge in the network between any two consecutive nodes along such a path, then a sufficiently large sample of paths will give us a fairly complete picture of the entire network. There may, however, be some edges that never get sampled, so the reconstruction is typically a good, but not perfect, representation of the true physical structure of the Internet. Studies of Internet structure have been carried out by, among others, Faloutsos et al. [148], Broida and Claffy [75], and Chen et al. [86].

One of the interesting features of all of these technological networks is that their structure is clearly governed to some extent by space and geography. Power grids, the Internet, air, road, and rail networks all span continents, and which vertices in the network are connected to which others is presumably both a function of what is technologically desirable and what is geographically feasible [425]. It is not yet well understood what the interplay of these factors is.

**2.4. Biological Networks.** A number of biological systems can be usefully represented as networks. Perhaps the classic example of a biological network is the network of metabolic pathways, which is a representation of metabolic substrates and products with directed edges joining them if a known metabolic reaction exists that acts on a given substrate and produces a given product. Most of us will probably have seen at some point the giant maps of metabolic pathways that many molecular biologists pin to their walls.<sup>9</sup> Studies of the statistical properties of metabolic networks have been performed by, for example, Jeong et al. [213, 339], Fell and Wagner [153, 404], and Stelling et al. [382]. A separate network is the network of mechanistic physical interactions between proteins (as opposed to chemical reactions among metabolites), which is usually referred to as a protein interaction network. Interaction networks have been studied by a number of authors [205, 211, 273, 375, 393].

Another important class of biological network is the genetic regulatory network. The expression of a gene, i.e., the production by transcription and translation of the protein for which the gene codes, can be controlled by the presence of other proteins, both activators and inhibitors, so that the genome itself forms a switching network with vertices representing the proteins and directed edges representing dependence of protein production on the proteins at other vertices. The statistical structure of regulatory networks has been studied recently by various authors [152, 183, 367]. Genetic regulatory networks were in fact one of the first networked dynamical sys-

<sup>9</sup>The standard chart of the metabolic network is somewhat misleading. For reasons of clarity and aesthetics, many metabolites appear in more than one place on the chart, so that some pairs of vertices are actually the same vertex.

tems for which large-scale modeling attempts were made. The early work on random Boolean nets by Kauffman [223, 224, 225] is a classic in this field and anticipated recent developments by several decades.

Another much-studied example of a biological network is the food web, in which the vertices represent species in an ecosystem and a directed edge from species A to species B indicates that A preys on B [91, 338]—see Figure 1.2c. (Sometimes the relationship is drawn the other way around, because ecologists tend to think in terms of energy or carbon flows through food webs; a predator-prey interaction is thus drawn as an arrow pointing from prey to predator, indicating energy flow from prey to predator when the prey is eaten.) Construction of complete food webs is a laborious business, but a number of quite extensive data sets have become available in recent years [27, 176, 203, 271]. Statistical studies of the topologies of food webs have been carried out by Solé and Montoya [289, 374], Camacho, Guimerà, and Amaral [82], and Dunne et al. [132, 133, 422], among others. A particularly thorough study of webs of plants and herbivores has been conducted by Jordano, Bascompte, and Olesen [218], which includes statistics for no less than 53 different networks.

Neural networks are another class of biological networks of considerable importance. Measuring the topology of real neural networks is extremely difficult, but has been done successfully in a few cases. The best known example is the reconstruction of the 282-neuron neural network of the nematode *C. elegans* by White et al. [420]. The network structure of the brain at larger scales than individual neurons—functional areas and pathways—has been investigated by Sporns [378] and Sporns, Tononi, and Edelman [379].

Blood vessels and the equivalent vascular networks in plants form the foundation for one of the most successful theoretical models of the effects of network structure on the behavior of a networked system, the theory of biological allometry [29, 416, 417], although we are not aware of any quantitative studies of their statistical structure.

Finally we mention two examples of networks from the physical sciences, the network of free energy minima and saddle points in glasses [130] and the network of conformations of polymers and the transitions between them [360], both of which appear to have some interesting structural properties.

**3. Properties of Networks.** Perhaps the simplest useful model of a network is the random graph, first studied by Rapoport [345, 346], Solomonoff and Rapoport [377], and Erdős and Rényi [141, 142, 143], which we describe in section 4.1. In this model, undirected edges are placed at random between a fixed number  $n$  of vertices to create a network in which each of the  $\frac{1}{2}n(n-1)$  possible edges is independently present with some probability  $p$ , and the number of edges connected to each vertex—the degree of the vertex—is distributed according to a binomial distribution, or a Poisson distribution in the limit of large  $n$ . The random graph has been well studied by mathematicians [63, 210, 222], and many results, both approximate and exact, have been proved rigorously. Most of the interesting features of real-world networks that have attracted the attention of researchers in the last few years, however, concern the ways in which networks are *not* like random graphs. Real networks are nonrandom in some revealing ways that suggest both possible mechanisms that could be guiding network formation, and possible ways in which we could exploit network structure to achieve certain aims. In this section we describe some features that appear to be common to networks of many different types.

**3.1. The Small-World Effect.** In section 2.1 we described the famous experiments carried out by Stanley Milgram in the 1960s, in which letters passed from

person to person were able to reach a designated target individual in only a small number of steps—around six in the published cases. This result is one of the first direct demonstrations of the *small-world effect*, the fact that most pairs of vertices in most networks seem to be connected by a short path through the network.

The existence of the small-world effect had been speculated upon before Milgram's work, notably in a remarkable 1929 short story by the Hungarian writer Frigyes Karinthy [221] and more rigorously in the mathematical work of Pool and Kochen [340], which, although published after Milgram's studies, was in circulation in preprint form for a decade before Milgram took up the problem. Nowadays, the small-world effect has been studied and verified directly in a large number of different networks.

Consider an undirected network, and let us define  $\ell$  to be the mean geodesic (i.e., shortest) distance between vertex pairs in a network:

$$(3.1) \quad \ell = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij},$$

where  $d_{ij}$  is the geodesic distance from vertex  $i$  to vertex  $j$ . Notice that we have included the distance from each vertex to itself (which is zero) in this average. This is mathematically convenient for a number of reasons, but not all authors do it. In any case, its inclusion simply multiplies  $\ell$  by  $(n-1)/(n+1)$  and hence gives a correction of order  $n^{-1}$ , which is often negligible for practical purposes.

The quantity  $\ell$  can be measured for a network of  $n$  vertices and  $m$  edges in time  $O(mn)$  using simple breadth-first search [7], also called a “burning algorithm” in the physics literature. In Table 3.1, we show values of  $\ell$  taken from the literature for a variety of different networks. As the table shows, the values are in all cases quite small—much smaller than the number  $n$  of vertices, for instance.

The definition (3.1) of  $\ell$  is problematic in networks that have more than one component. In such cases, there exist vertex pairs that have no connecting path. Conventionally one assigns infinite geodesic distance to such pairs, but then the value of  $\ell$  also becomes infinite. To avoid this problem one usually defines  $\ell$  on such networks to be the mean geodesic distance between all pairs that have a connecting path. Pairs that fall in two different components are excluded from the average. The figures in Table 3.1 were all calculated in this way. An alternative and perhaps more satisfactory approach is to define  $\ell$  to be the “harmonic mean” geodesic distance between all pairs, i.e., the reciprocal of the average of the reciprocals:

$$(3.2) \quad \ell^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}^{-1}.$$

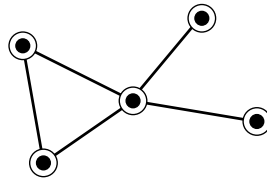
Infinite values of  $d_{ij}$  then contribute nothing to the sum. This approach has been adopted only occasionally in network calculations [259], but perhaps should be used more often.

The small-world effect has obvious implications for the dynamics of processes taking place on networks. For example, if one considers the spread of information, or indeed anything else, across a network, the small-world effect implies that that spread will be fast on most real-world networks. If it takes only six steps for a rumor to spread from any person to any other, for instance, then the rumor will spread much faster than if it takes a hundred steps, or a million. This affects the number of “hops” a packet must make to get from one computer to another on the Internet, the number

**Table 3.1** Basic statistics for a number of published networks. The properties measured are as follows: total number of vertices  $n$ ; total number of edges  $m$ ; mean degree  $z$ ; mean vertex-vertex distance  $\ell$ ; type of graph, directed or undirected; exponent  $\alpha$  of degree distribution if the distribution follows a power law (or “–” if not; in/out-degree exponents are given for directed graphs); clustering coefficient  $C^{(1)}$  from (3.3); clustering coefficient  $C^{(2)}$  from (3.6); degree correlation coefficient  $r$ , section 3.6. The last column gives the citation for the network in the bibliography. Blank entries indicate unavailable data.

Network	Type	$n$	$m$	$z$	$\ell$	$\alpha$	$C^{(1)}$	$C^{(2)}$	$r$	Ref(s).
Social	film actors	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	[20, 415]
	company directors	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	[105, 322]
	math coauthorship	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	[107, 181]
	physics coauthorship	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	[310, 312]
	biology coauthorship	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	[310, 312]
	telephone call graph	47 000 000	80 000 000	3.16	–	2.1	–	–	–	[8, 9]
	email messages	59 912	86 300	1.44	4.95	1.5/2.0	0.17	0.16	0.092	[136]
	email address books	16 881	57 029	3.38	5.22	–	–	0.13	–	[320]
	student relationships	573	477	1.66	16.01	–	0.005	0.001	–0.029	[45]
	sexual contacts	2 810	–	–	–	3.2	–	–	–	[264, 265]
Information	WWW nd. edu	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	[14, 34]
	WWW Altavista	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7	–	–	–	[74]
	citation network	783 339	6 716 198	8.57	–	3.0/–	–	–	–	[350]
	Roget's Thesaurus	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	[243]
Technological	word co-occurrence	460 902	17 000 000	70.13	–	2.7	–	0.44	–	[119, 157]
	Internet	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	[86, 148]
	power grid	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	[415]
	train routes	587	19 603	66.79	2.16	–	–	0.69	–0.033	[365]
Biological	software packages	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	[317]
	software classes	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	[394]
	electronic circuits	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	[155]
	peer-to-peer network	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	[6, 353]
Biological	metabolic network	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	[213]
	protein interactions	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	[211]
	marine food web	135	598	4.43	2.05	–	0.16	0.23	–0.263	[203]
	freshwater food web	92	997	10.84	1.90	–	0.40	0.48	–0.326	[271]
Biological	neural network	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	[415, 420]





**Fig. 3.1** Illustration of the definition of the clustering coefficient  $C$ , (3.3). This network has one triangle and eight connected triples, and therefore has a clustering coefficient of  $3 \times \frac{1}{8} = \frac{3}{8}$ . The individual vertices have local clustering coefficients, Eq. (3.5), of 1, 1,  $\frac{1}{6}$ , 0, and 0, for a mean value, (3.6), of  $C = \frac{13}{30}$ .

of legs of a journey for an air or train traveler, the time it takes for a disease to spread throughout a population, and so forth. The small-world effect also underlies some well-known parlor games, particularly the calculation of Erdős numbers [107] and Bacon numbers.<sup>10</sup>

On the other hand, the small-world effect is also mathematically obvious. If the number of vertices within a distance  $r$  of a typical central vertex grows exponentially with  $r$ —and this is true of many networks, including the random graph (section 4.1)—then the value of  $\ell$  will increase as  $\log n$ . In recent years the term “small-world effect” has thus taken on a more precise meaning: networks are said to show the small-world effect if the value of  $\ell$  scales logarithmically or slower with network size for fixed mean degree. Logarithmic scaling can be proved for a variety of network models [61, 63, 88, 127, 164] and has also been observed in various real-world networks [13, 311, 312]. Some networks have mean vertex–vertex distances that increase slower than  $\log n$ . Bollobás and Riordan [64] have shown that networks with power-law degree distributions (section 3.3) have values of  $\ell$  that increase no faster than  $\log n / \log \log n$  (see also [164]), and Cohen and Havlin [95] have given arguments that suggest that the actual variation may be slower even than this.

**3.2. Transitivity or Clustering.** A clear deviation from the behavior of the random graph can be seen in the property of network transitivity, sometimes also called clustering, although the latter term also has another meaning in the study of networks (see section 3.7) and so can be confusing. In many networks it is found that if vertex A is connected to vertex B and vertex B to vertex C, then there is a heightened probability that vertex A will also be connected to vertex C. In the language of social networks, the friend of your friend is likely also to be your friend. In terms of network topology, transitivity means the presence of a heightened number of triangles in the network—sets of three vertices each of which is connected to each of the others. It can be quantified by defining a clustering coefficient  $C$  thus:

$$(3.3) \quad C = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of vertices}},$$

where a “connected triple” means a single vertex with edges running to an unordered pair of others (see Figure 3.1).

In effect,  $C$  measures the fraction of triples that have their third edge filled in to complete the triangle. The factor of three in the numerator accounts for the fact that each triangle contributes to three triples and ensures that  $C$  lies in the range

<sup>10</sup><http://www.cs.virginia.edu/oracle/>

$0 \leq C \leq 1$ . In simple terms,  $C$  is the mean probability that two vertices that are network neighbors of the same other vertex will themselves be neighbors. It can also be written in the form

$$(3.4) \quad C = \frac{6 \times \text{number of triangles in the network}}{\text{number of paths of length two}},$$

where a path of length two refers to a directed path starting from a specified vertex. This definition shows that  $C$  is also the mean probability that the friend of your friend is also your friend.

The definition of  $C$  given here has been widely used in the sociology literature, where it is referred to as the “fraction of transitive triples.”<sup>11</sup> In the mathematical and physical literature it seems to have been first discussed by Barrat and Weigt [40].

An alternative definition of the clustering coefficient, also widely used, has been given by Watts and Strogatz [415], who proposed defining a local value

$$(3.5) \quad C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}.$$

For vertices with degree 0 or 1, for which both numerator and denominator are zero, we put  $C_i = 0$ . Then the clustering coefficient for the whole network is the average

$$(3.6) \quad C = \frac{1}{n} \sum_i C_i.$$

This definition effectively reverses the order of the operations of taking the ratio of triangles to triples and of averaging over vertices—one here calculates the mean of the ratio, rather than the ratio of the means. It tends to weight the contributions of low-degree vertices more heavily, because such vertices have a small denominator in (3.5) and hence can give quite different results from (3.3). In Table 3.1 we give both measures for a number of networks (denoted  $C^{(1)}$  and  $C^{(2)}$  in the table). Normally our first definition (3.3) is easier to calculate analytically, but (3.6) is easily calculated on a computer and has found wide use in numerical studies and data analysis. It is important when reading (or writing) literature in this area to be clear about which definition of the clustering coefficient is in use. The difference between the two is illustrated in Figure 3.1.

The local clustering  $C_i$  above has been used quite widely in its own right in the sociological literature, where it is referred to as the “network density” [362]. Its dependence on the degree  $k_i$  of the central vertex  $i$  has been studied by Dorogovtsev, Goltsev, and Mendes [113] and Szabó, Alava, and Kertész [388]; both groups found that  $C_i$  falls off with  $k_i$  approximately as  $k_i^{-1}$  for certain models of scale-free networks (section 3.3.1). Similar behavior has also been observed empirically in real-world networks [348, 349, 396].

In general, regardless of which definition of the clustering coefficient is used, the values tend to be considerably higher than for a random graph with a similar number of vertices and edges. Indeed, it is suspected that for many types of networks the probability that the friend of your friend is also your friend should tend to a nonzero limit as the network becomes large, so that  $C = O(1)$  as  $n \rightarrow \infty$ .<sup>12</sup> On the random

<sup>11</sup>For example, the standard network analysis program UCInet includes a function to calculate this quantity for any network.

<sup>12</sup>An exception is scale-free networks with  $C_i \sim k_i^{-1}$ , as described above. For such networks (3.3) tends to zero as  $n \rightarrow \infty$ , although (3.6) is still finite.

graph, by contrast,  $C = O(n^{-1})$  for large  $n$  (either definition of  $C$ ) and hence the real-world and random graph values can be expected to differ by a factor of order  $n$ . This point is discussed further in section 4.1.

The clustering coefficient measures the density of triangles in a network. An obvious generalization is to ask about the density of longer loops also: loops of length four and above. A number of authors have looked at such higher order clustering coefficients [54, 79, 165, 172, 316], although there is so far no clean theory, similar to a cumulant expansion, that separates the independent contributions of the various orders from one another. If more than one edge is permitted between a pair of vertices, then there is also a lower order clustering coefficient that describes the density of loops of length two. This coefficient is particularly important in directed graphs where the two edges in question can point in opposite directions. The probability that two vertices in a directed network point to each other is called the *reciprocity* and is often measured in directed social networks [362, 408]. It has been examined occasionally in other contexts too, such as the World Wide Web [3, 137] and email networks [320].

**3.3. Degree Distributions.** Recall that the degree of a vertex in a network is the number of edges incident on (i.e., connected to) that vertex. We define  $p_k$  to be the fraction of vertices in the network that have degree  $k$ . Equivalently,  $p_k$  is the probability that a vertex chosen uniformly at random has degree  $k$ . A plot of  $p_k$  for any given network can be formed by making a histogram of the degrees of vertices. This histogram is the degree distribution for the network. In a random graph of the type studied by Erdős and Rényi [141, 142, 143], each edge is present or absent with equal probability, and hence the degree distribution is, as mentioned earlier, binomial, or Poisson in the limit of large graph size. Real-world networks are mostly found to be very unlike the random graph in their degree distributions. Far from having a Poisson distribution, the degrees of the vertices in most networks are highly right-skewed, meaning that their distribution has a long right tail of values that are far above the mean.

Measuring this tail is somewhat tricky. Although in theory one just has to construct a histogram of the degrees, in practice one rarely has enough measurements to get good statistics in the tail, and direct histograms are thus usually rather noisy (see the histograms in [74, 148, 342], for example). There are two accepted ways to get around this problem. One is to construct a histogram in which the bin sizes increase exponentially with degree. For example, the first few bins might cover degree ranges 1, 2–3, 4–7, 8–15, and so on. The number of samples in each bin is then divided by the width of the bin to normalize the measurement. This method of constructing a histogram is often used when the histogram is to be plotted with a logarithmic degree scale, so that the widths of the bins will appear even. Because the bins get wider as we get out into the tail, the problems with statistics are reduced, although they are still present to some extent as long as  $p_k$  falls off faster than  $k^{-1}$ , which it must if the distribution is to be integrable.

An alternative way of presenting degree data is to make a plot of the cumulative distribution function

$$(3.7) \quad P_k = \sum_{k'=k}^{\infty} p_{k'},$$

which is the probability that the degree is greater than or equal to  $k$ . Such a plot has the advantage that all the original data are represented. When we make a conventional histogram by binning, any differences between the values of data points that fall in

the same bin are lost. The cumulative distribution function does not suffer from this problem. The cumulative distribution also reduces the noise in the tail. On the downside, the plot doesn't give a direct visualization of the degree distribution itself, and adjacent points on the plot are not statistically independent, making correct fits to the data tricky.

In Figure 3.2 we show cumulative distributions of degree for a number of the networks described in section 2. As the figure shows, the distributions are indeed all right-skewed. Many of them follow power laws in their tails:  $p_k \sim k^{-\alpha}$  for some constant exponent  $\alpha$ . Note that such power-law distributions show up as power laws in the cumulative distributions also, but with exponent  $\alpha - 1$  rather than  $\alpha$ :

$$(3.8) \quad P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)}.$$

Some of the other distributions have exponential tails:  $p_k \sim e^{-k/\kappa}$ . These also give exponentials in the cumulative distribution, but with the *same* exponent:

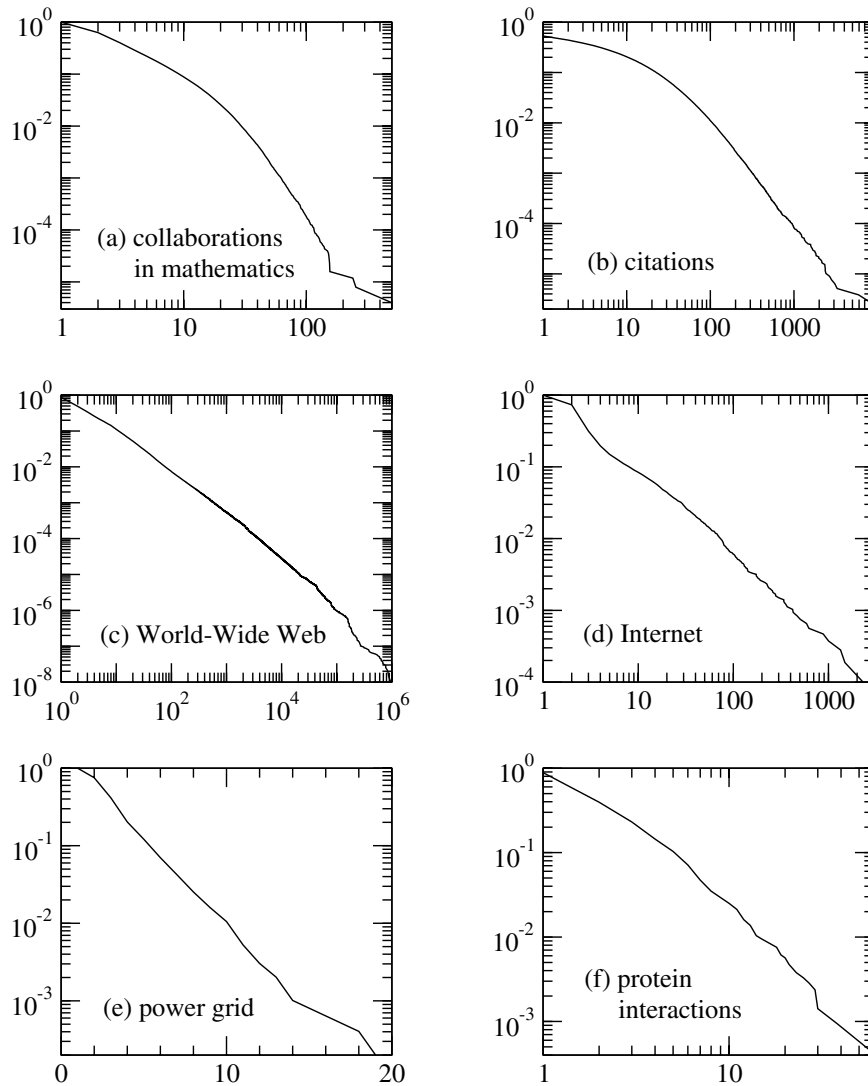
$$(3.9) \quad P_k = \sum_{k'=k}^{\infty} p_{k'} \sim \sum_{k'=k}^{\infty} e^{-k'/\kappa} \sim e^{-k/\kappa}.$$

This makes power-law and exponential distributions particularly easy to spot experimentally, by plotting the corresponding cumulative distributions on logarithmic scales (for power laws) or semilogarithmic scales (for exponentials).

For other types of networks degree distributions can be more complicated. For bipartite graphs, for instance (section 1.1), there are two degree distributions, one for each type of vertex. For directed graphs each vertex has both an in-degree and an out-degree, and the degree distribution therefore becomes a function  $p_{jk}$  of two variables, representing the fraction of vertices that simultaneously have in-degree  $j$  and out-degree  $k$ . In empirical studies of directed graphs like the Web, researchers have usually given only the individual distributions of in- and out-degree [14, 34, 74], i.e., the distributions derived by summing  $p_{jk}$  over one or other of its indices. This, however, discards much of the information present in the joint distribution. It has been found that in- and out-degrees are quite strongly correlated in some networks [320], which suggests that there is more to be gleaned from the joint distribution than is normally appreciated.

**3.3.1. Scale-Free Networks.** Networks with power-law degree distributions have been the focus of a great deal of attention in the literature [13, 120, 386]. They are sometimes referred to as *scale-free networks* [32], although it is only their degree distributions that are scale-free;<sup>13</sup> one can and usually does have scales present in other network properties. The earliest published example of a scale-free network is probably Price's network of citations between scientific papers [342] (see section 2.2). He quoted a value of  $\alpha = 2.5$  to 3 for the exponent of his network. In a later paper he quoted a more accurate figure of  $\alpha = 3.04$  [343]. He also found a power-law distribution for the out-degree of the network (number of bibliography entries in each paper), although later work has called this into question [395]. More recently, power-law degree distributions have been observed in a host of other networks, including

<sup>13</sup>The term "scale-free" refers to any functional form  $f(x)$  that remains unchanged to within a multiplicative factor under a rescaling of the independent variable  $x$ . In effect this means power-law forms, since these are the only solutions to  $f(ax) = bf(x)$ , and hence "power-law" and "scale-free" are, for our purposes, synonymous.



**Fig. 3.2** Cumulative degree distributions for six different networks. The horizontal axis for each panel is vertex degree  $k$  (or in-degree for the citation and Web networks, which are directed) and the vertical axis is the cumulative probability distribution of degrees, i.e., the fraction of vertices that have degree greater than or equal to  $k$ . The networks shown are the following: (a) the collaboration network of mathematicians [181]; (b) citations between 1981 and 1997 to all papers cataloged by the Institute for Scientific Information [350]; (c) a 300 million vertex subset of the World Wide Web, circa 1999 [74]; (d) the Internet at the level of autonomous systems, April 1999 [86]; (e) the power grid of the western United States [415]; (f) the interaction network of proteins in the metabolism of the yeast *S. cerevisiae* [211]. Of these networks, three of them, (c), (d), and (f), appear to have power-law degree distributions, as indicated by their approximately straight-line forms on the doubly logarithmic scales, and one (b) has a power-law tail but deviates markedly from power-law behavior for small degree. Network (e) has an exponential degree distribution (note the log-linear scales used in this panel) and network (a) appears to have a truncated power-law degree distribution of some type, or possible two separate power-law regimes with different exponents.

notably other citation networks [350, 363], the World Wide Web [14, 34, 74], the Internet [86, 148, 400], metabolic networks [211, 213], telephone call graphs [8, 9], and the network of human sexual contacts [217, 265]. The degree distributions of some of these networks are shown in Figure 3.2.

Other common functional forms for the degree distribution are exponentials, such as those seen in the power grid [20] and railway networks [365], and power laws with exponential cutoffs, such as those seen in the network of movie actors [20] and some collaboration networks [312]. Note also that while a particular form may be seen in the degree distribution for the network as a whole, specific subnetworks within the network can have other forms. The World Wide Web, for instance, shows a power-law degree distribution overall but unimodal distributions within domains [337].

**3.3.2. Maximum Degree.** The maximum degree  $k_{\max}$  of a vertex in a network will in general depend on the size of the network. For some calculations on networks the value of this maximum degree matters (see, for example, section 8.3.2). In work on scale-free networks, Aiello, Chung, and Lu [8] assumed that the maximum degree was approximately the value above which there is less than one vertex of that degree in the graph on average, i.e., the point where  $np_k = 1$ . This means, for instance, that  $k_{\max} \sim n^{1/\alpha}$  for the power-law degree distribution  $p_k \sim k^{-\alpha}$ . This assumption, however, can give misleading results; in many cases there will be vertices in the network with significantly higher degree than this, as discussed by Adamic et al. [6].

Given a particular degree distribution (and assuming all degrees to be sampled independently from it, which may not be true for networks in the real world), the probability of there being exactly  $m$  vertices of degree  $k$  and no vertices of higher degree is  $\binom{n}{m} p_k^m (1 - P_k)^{n-m}$ , where  $P_k$  is the cumulative probability distribution, (3.7). Hence the probability  $h_k$  that the highest degree on the graph is  $k$  is

$$(3.10) \quad h_k = \sum_{m=1}^n \binom{n}{m} p_k^m (1 - P_k)^{n-m} = (p_k + 1 - P_k)^n - (1 - P_k)^n,$$

and the expected value of the highest degree is  $k_{\max} = \sum_k k h_k$ .

For both small and large values of  $k$ ,  $h_k$  tends to zero, and the sum over  $k$  is dominated by the terms close to the maximum. Thus, in most cases, a good approximation to the expected value of the maximum degree is given by the modal value. Differentiating and observing that  $dP_k/dk = p_k$ , we find that the maximum of  $h_k$  occurs when

$$(3.11) \quad \left( \frac{dp_k}{dk} - p_k \right) (p_k + 1 - P_k)^{n-1} + p_k (1 - P_k)^{n-1} = 0,$$

or  $k_{\max}$  is a solution of

$$(3.12) \quad \frac{dp_k}{dk} \simeq -np_k^2,$$

where we have made the (fairly safe) assumption that  $p_k$  is sufficiently small for  $k \gtrsim k_{\max}$  that  $np_k \ll 1$  and  $P_k \ll 1$ .

For example, if  $p_k \sim k^{-\alpha}$  in its tail, then we find that

$$(3.13) \quad k_{\max} \sim n^{1/(\alpha-1)}.$$

As shown by Cohen et al. [93], a simple rule of thumb that leads to the same result is that the maximum degree is roughly the value of  $k$  that solves  $np_k = 1$ . Note, however, that, as shown by Dorogovtsev and Samukhin [129], the fluctuations in the tail of the degree distribution are very large for the power-law case.

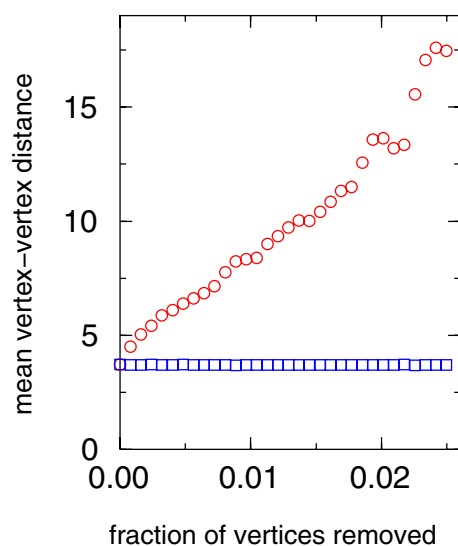
Dorogovtsev, Mendes, and Samukhin [126] have also shown that (3.13) holds for networks generated using the “preferential attachment” procedure of Barabási and Albert [32] described in section 7.2, and a detailed numerical study of this case has been carried out by Moreira, Andrade, and Amaral [294].

**3.4. Network Resilience.** Related to degree distributions is the property of resilience of networks to the removal of their vertices, which has been the subject of a good deal of attention in the literature. Most of the networks we have been considering rely for their function on their connectivity, i.e., the existence of paths leading between pairs of vertices. If vertices are removed from a network, the typical length of these paths will increase, and ultimately vertex pairs will become disconnected and communication between them through the network will become impossible. Networks vary in their level of resilience to such vertex removal.

There are also a variety of different ways in which vertices can be removed and different networks show varying degrees of resilience to these also. For example, one could remove vertices at random from a network, or one could target some specific class of vertices, such as those with the highest degrees. Network resilience is of particular importance in epidemiology, where “removal” of vertices in a contact network might correspond, for example, to vaccination of individuals against a disease. Because vaccination not only prevents the vaccinated individuals from catching the disease but may also destroy paths between other individuals by which the disease might have spread, it can have a wider reaching effect than one might at first think, and careful consideration of the efficacy of different vaccination strategies could lead to substantial advantages for public health.

Recent interest in network resilience has been sparked by the work of Albert, Jeong, and Barabási [15], who studied the effect of vertex deletion in two example networks, a 6000-vertex network representing the topology of the Internet at the level of autonomous systems (see section 2.3) and a 326,000-page subset of the World Wide Web. Both the Internet and the Web have been observed to have degree distributions that are approximately power-law in form [14, 74, 86, 148, 400] (section 3.3.1). The authors measured average vertex–vertex distances as a function of number of vertices removed, both for random removal and for progressive removal of the vertices with the highest degrees.<sup>14</sup> In Figure 3.3 we show their results for the Internet. They found for both networks that distance was almost entirely unaffected by random vertex removal, i.e., the networks studied were highly resilient to this type of removal. This is intuitively reasonable, since most of the vertices in these networks have low degree and therefore lie on few paths between others; thus their removal rarely affects communications substantially. On the other hand, when removal is targeted at the highest degree vertices, it is found to have devastating effect. Mean vertex–vertex distance increases very sharply with the fraction of vertices removed, and typically only a few percent of vertices need be removed before essentially all communication through the network is destroyed. Albert et al. expressed their results in terms of failure or sabotage of network nodes. The Internet (and the Web) they suggest, is highly resilient against the random failure of vertices in the network but highly vulnerable to deliberate attack on its highest-degree vertices.

<sup>14</sup>In removing the vertices with the highest degrees, Albert et al. recalculated degrees following the removal of each vertex. Most other authors who have studied this issue have adopted a slightly different strategy of removing vertices in order of their *initial* degree in the network before any removal.



**Fig. 3.3** Mean vertex-vertex distance on a graph representation of the Internet at the autonomous system level, as vertices are removed one by one. If vertices are removed in random order (squares), distance increases only very slightly, but if they are removed in order of their degrees, starting with the highest degree vertices (circles), then distance increases sharply. After Albert, Jeong, and Barabási [15]. Reprinted with permission from Nature Publishing Group ([www.nature.com](http://www.nature.com)).

Results similar to those of Albert et al. were found independently by Broder et al. [74] for a much larger subset of the Web graph. Interestingly, however, Broder et al. gave an entirely opposite interpretation of their results. They found that in order to destroy connectivity in the Web one has to remove all vertices with degree greater than five, which seems like a drastic attack on the network, given that some vertices have degrees in the thousands. They thus concluded that the network was very *resilient* against targeted attack. In fact, however, there is not such a conflict between these results as at first appears. Because of the highly skewed degree distribution of the Web, the fraction of vertices with degree greater than five is only a small fraction of all vertices.

Following these studies, many authors have looked into the question of resilience for other networks. In general the picture seems to be consistent with that seen in the Internet and Web. Most networks are robust against random vertex removal but considerably less robust to targeted removal of the highest-degree vertices. Jeong et al. [211] have looked at metabolic networks, Dunne, Williams, and Martinez [132, 133] at food webs, Newman, Forrest, and Balthrop [320] at email networks, and a variety of authors at resilience of model networks [15, 81, 93, 94, 199], which we discuss in more detail in later sections of the review. A particularly thorough study of the resilience of both real-world and model networks has been conducted by Holme et al. [199], who looked not only at vertex removal but also at removal of edges and considered some additional strategies for selecting vertices based on so-called betweenness (see sections 3.7 and 3.9).

**3.5. Mixing Patterns.** Delving a little deeper into the statistics of network structure, one can ask about which vertices pair up with which others. In most kinds



**Table 3.2** *Couples in the study of Catania et al. [85] tabulated by race of either partner. After Morris [301].*

		Women			
		Black	Hispanic	White	Other
Men	Black	506	32	69	26
	Hispanic	23	308	114	38
	White	26	46	599	68
	Other	10	14	47	32

of networks there are at least a few different types of vertices, and the probabilities of connection between vertices often depends on types. For example, in a food web representing which species eat which in an ecosystem (section 2.4) one sees vertices representing plants, herbivores, and carnivores. Many edges link the plants and herbivores, and many more the herbivores and carnivores. But there are few edges linking herbivores to other herbivores, or carnivores to plants. For the Internet, Maslov, Sneppen, and Zaliznyak [274] have proposed that the structure of the network reflects the existence of three broad categories of nodes: high-level connectivity providers who run the Internet backbone and trunk lines, consumers who are end users of Internet service, and ISPs who join the two. Again there are many links between end users and ISPs, and many between ISPs and backbone operators, but few between ISPs and other ISPs, or between backbone operators and end users.

In social networks this kind of selective linking is called *assortative mixing* or *homophily* and has been widely studied, as it has also in epidemiology. (The term “assortative matching” is also seen in the ecology literature, particularly in reference to mate choice among animals.) A classic example of assortative mixing in social networks is mixing by race. Table 3.2, for example, reproduces results from a study of 1,958 couples in the city of San Francisco, California. Among other things, the study recorded the race (self-identified) of study participants in each couple. As the table shows, participants appear to draw their partners preferentially from those of their own race, and this is believed to be a common phenomenon in many social networks: we tend to associate preferentially with people who are similar to ourselves in some way.

Assortative mixing can be quantified by an “assortativity coefficient,” which can be defined in a couple of different ways. Let  $E_{ij}$  be the number of edges in a network that connect vertices of types  $i$  and  $j$ , with  $i, j = 1, \dots, N$ , and let  $\mathbf{E}$  be the matrix with elements  $E_{ij}$ , as depicted in Table 3.2. We define a normalized mixing matrix by

$$(3.14) \quad \mathbf{e} = \frac{\mathbf{E}}{\|\mathbf{E}\|},$$

where  $\|\mathbf{x}\|$  means the sum of all the elements of the matrix  $\mathbf{x}$ . The elements  $e_{ij}$  measure the *fraction* of edges that fall between vertices of types  $i$  and  $j$ . One can also ask about the conditional probability  $P(j|i)$  that my network neighbor is of type  $j$  given that I am of type  $i$ , which is given by  $P(j|i) = e_{ij} / \sum_j e_{ij}$ . These quantities satisfy the normalization conditions

$$(3.15) \quad \sum_{ij} e_{ij} = 1, \quad \sum_j P(j|i) = 1.$$

Gupta, Anderson, and May [185] have suggested that assortative mixing be quantified by the coefficient

$$(3.16) \quad Q = \frac{\sum_i P(i|i) - 1}{N - 1}.$$

This quantity has the desirable properties that it is 1 for a perfectly assortative network (every edge falls between vertices of the same type), and 0 for randomly mixed networks, and it has been quite widely used in the literature. But it suffers from two shortcomings [317]: (1) for an asymmetric matrix like the one in Table 3.2,  $Q$  has two different values, depending on whether we put the men or the women along the horizontal axis, and it is unclear which of these two values is the “correct” one for the network; (2) the measure weights each vertex type equally, regardless of how many vertices there are of each type, which can give rise to misleading figures for  $Q$  in cases where community size is heterogeneous, as it often is.

An alternative assortativity coefficient that remedies these problems is defined by [317]

$$(3.17) \quad r = \frac{\text{Tr } \mathbf{e} - \|\mathbf{e}^2\|}{1 - \|\mathbf{e}^2\|}.$$

This quantity is also 0 in a randomly mixed network and 1 in a perfectly assortative one. But its value is not altered by transposition of the matrix and it weights vertices equally rather than communities, so that small communities make an appropriately small contribution to  $r$ . For the data of Table 3.2 we find  $r = 0.621$ .

Another type of assortative mixing is mixing by scalar characteristics such as age or income. Again it is usually found that people prefer to associate with others of similar age and income to themselves, although of course age and income, like race, may be proxies for other driving forces, such as cultural differences. Garfinkel, Glei, and McLanahan [170] and Newman [317], for example, have analyzed data for unmarried and married couples, respectively, to show that there is strong correlation between the ages of partners. Mixing by scalar characteristics can be quantified by calculating a correlation coefficient for the characteristic in question.

In theory assortative mixing according to vector characteristics should also be possible. For example, geographic location probably affects individuals’ propensity to become acquainted. Location could be viewed as a two-vector, with the probability of connection between pairs of individuals being assortative on the values of these vectors.

**3.6. Degree Correlations.** A special case of assortative mixing according to a scalar vertex property is mixing according to vertex degree, also commonly referred to simply as degree correlation. Do the high-degree vertices in a network associate preferentially with other high-degree vertices? Or do they prefer to attach to low-degree ones? Both situations are seen in some networks, as it turns out. The case of assortative mixing by degree is of particular interest because, since degree is itself a property of the graph topology, degree correlations can give rise to some interesting network structure effects.

Several different ways of quantifying degree correlations have been proposed. Maslov et al. [273, 274] have simply plotted the two-dimensional histogram of the degrees of vertices at either ends of an edge. They have shown results for protein interaction networks and the Internet. A more compact representation of the situation is that proposed by Pastor-Satorras et al. [330, 400], who in studies of the Internet

calculated the mean degree of the network neighbors of a vertex as a function of the degree  $k$  of that vertex. This gives a one-parameter curve which increases with  $k$  if the network is assortatively mixed. For the Internet in fact it is found to decrease with  $k$ , a situation we call disassortativity. Newman [313, 317] reduced the measurement still further to a single number by calculating the Pearson correlation coefficient of the degrees at either ends of an edge. This gives a single number that should be positive for assortatively mixed networks and negative for disassortative ones. In Table 3.1 we show results for a number of different networks. An interesting observation is that essentially all social networks measured appear to be assortative, but other types of networks (information networks, technological networks, biological networks) appear to be disassortative. It is not clear what the explanation for this result is, or even if there is any one single explanation. (Probably there is not.)

**3.7. Community Structure.** It is widely assumed [362, 408] that most social networks show “community structure,” i.e., groups of vertices that have a high density of edges within them, with a lower density of edges between groups. It is a matter of common experience that people do divide into groups along lines of interest, occupation, age, and so forth, and the phenomenon of assortativity discussed in section 3.5 certainly suggests that this might be the case. (It is possible for a network to have assortative mixing but no community structure. This can occur, for example, when there is assortative mixing by age or other scalar quantities. Networks with this type of structure are sometimes said to be “stratified.”)

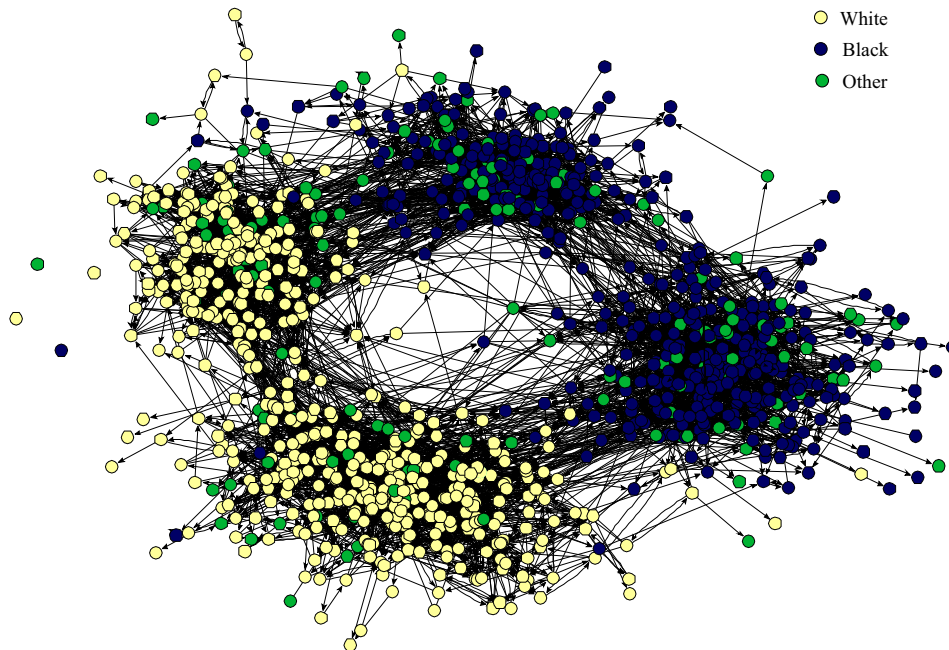
In Figure 3.4 we show a visualization of the friendship network of children in a U.S. school taken from a study by Moody [290].<sup>15</sup> The figure was created using a “spring embedding” algorithm, in which linear springs are placed between vertices and the system is relaxed using a first-order energy minimization. We have no special reason to suppose that this very simple algorithm would reveal anything particularly useful about the network, but the network appears to have strong enough community structure that in fact the communities appear clearly in the figure. Moreover, when Moody colors the vertices according to the race of the individuals they represent, as shown in the figure, it becomes immediately clear that one of the principal divisions in the network is by individuals’ race, and this is presumably what is driving the formation of communities in this case. (The other principal division visible in the figure is between middle school and high school, which are age divisions in the American education system.)

It would be of some interest, and indeed practical importance, were we to find that other types of networks, such as those listed in Table 3.1, show similar group structure also. One might well imagine, for example, that citation networks would divide into groups representing particular areas of research interest, and a good deal of energy has been invested in studies of this phenomenon [101, 138]. Similarly communities in the World Wide Web might reflect the subject matter of pages, communities in metabolic, neural, or software networks might reflect functional units, communities in food webs might reflect subsystems within ecosystems, and so on.

The traditional method for extracting community structure from a network is *cluster analysis* [147], sometimes also called *hierarchical clustering*.<sup>16</sup> In this method, one assigns a “connection strength” to vertex pairs in the network of interest. In general each of the  $\frac{1}{2}n(n-1)$  possible pairs in a network of  $n$  vertices is assigned such

<sup>15</sup>This image does not appear in the paper cited, but it and a number of other images of friendship networks from the same study can be found at <http://www.sociology.ohio-state.edu/jwm/>.

<sup>16</sup>Not to be confused with the entirely different use of the word *clustering* introduced in section 3.2.

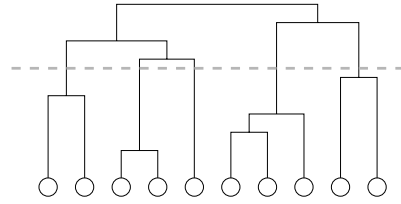


**Fig. 3.4** *Friendship network of children in a U.S. school. Friendships are determined by asking the participants, and hence are directed, since A may say that B is their friend but not vice versa. Vertices are color coded according to race, as marked, and the split from left to right in the figure is clearly primarily along lines of race. The split from top to bottom is between middle school and high school, i.e., between younger and older children. Picture courtesy of James Moody.*

a strength, not just those that are connected by an edge, although there are versions of the method where not all pairs are assigned a strength; in that case one can assume the remaining pairs to have a connection strength of zero. Then, starting with  $n$  vertices with no edges between any of them, one adds edges in order of decreasing vertex–vertex connection strength. One can pause at any point in this process and examine the component structure formed by the edges added so far; these components are taken to be the communities (or “clusters”) at that stage in the process. When all edges have been added, all vertices are connected to all others, and there is only one community. The entire process can be represented by a tree or *dendrogram* of union operations between vertex sets in which the communities at any level correspond to a horizontal cut through the tree—see Figure 3.5.<sup>17</sup>

Clustering is possible according to many different definitions of the connection strength. Reasonable choices include various weighted vertex–vertex distance measures, the sizes of minimum cut-sets (i.e., maximum flow) [7], and weighted path counts between vertices. Recently a number of authors have had success with methods based on “edge betweenness,” which is the count of how many geodesic paths between vertices run along each edge in the network [171, 184, 196, 421]. Results ap-

<sup>17</sup>For some reason such trees are conventionally depicted with their “root” at the top and their “leaves” at the bottom, which is not the natural order of things for most trees.



**Fig. 3.5** *An example of a dendrogram showing the hierarchical clustering of ten vertices. A horizontal cut through the dendrogram, such as that denoted by the dotted line, splits the vertices into a set of communities, five in this case.*

pear to show that, for social and biological networks at least, community structure is a common network property, although some food webs are found not to break up into communities in any simple way. (Food webs may be different from other networks in that they appear to be dense: mean vertex degree increases roughly linearly with network size, rather than remaining constant as it does in most networks [132, 272]. The same may be true of metabolic networks also [P. Holme, personal communication].)

Network clustering should not be confused with the technique of data clustering, which is a way of detecting groupings of data-points in high-dimensional data spaces [207]. The two problems do have some common features, however, and algorithms for one can be adapted for the other, and vice versa. For example, high-dimensional data can be converted into a network by placing edges between closely spaced data points, and then network clustering algorithms can be applied to the result. On balance, however, one normally finds that algorithms specially devised for data clustering work better than such borrowed methods, and the same is true in reverse.

In the social networks literature, network clustering has been discussed to a great extent in the context of so-called *block models* [71, 418], which are essentially just divisions of networks into communities or blocks according to one criterion or another. Sociologists have concentrated particularly on *structural equivalence*. Two vertices in a network are said to be structurally equivalent if they have all of the same neighbors. Exact structural equivalence is rare, but approximate equivalence can be used as the basis for a hierarchical clustering method such as that described above.

Another slightly different question about community structure, but related to the one discussed here, has been studied by Flake et al. [158]: if one is given an example vertex drawn from a known network, can one identify the community to which it belongs? Algorithmic methods for answering this question would clearly be of some practical value for searching networks such as the World Wide Web and citation networks. Flake et al. give what appears to be a very successful algorithm, at least in the context of the Web, based on a maximum flow method.

**3.8. Network Navigation.** Stanley Milgram's famous small-world experiment (section 2.1), in which letters were passed from person to person in an attempt to get them to a desired target individual, showed that there exist short paths through social networks between apparently distant individuals. However, there is another conclusion that can be drawn from this experiment which Milgram apparently failed to notice; it was pointed out in 2000 by Kleinberg [237, 238]. Milgram's results demonstrate that there exist short paths in the network, but they also demonstrate that ordinary people are good at finding them. This is, upon reflection, perhaps an

even more surprising result than the existence of the paths in the first place. The participants in Milgram's study had no special knowledge of the network connecting them to the target person. Most people know only who their friends are and perhaps a few of their friends' friends. Nonetheless it proved possible to get a message to a distant target in only a small number of steps. This indicates that there is something quite special about the structure of the network. On a random graph, for instance, as Kleinberg pointed out, short paths between vertices exist but no one would be able to find them given only the kind of information that people have in realistic situations. If it were possible to construct artificial networks that were easy to navigate in the same way that social networks appear to be, it has been suggested they could be used to build efficient database structures or better peer-to-peer computer networks [5, 6, 414] (see section 8.3.3).

**3.9. Other Network Properties.** In addition to the heavily studied network properties of the preceding sections, a number of others have received some attention. In some networks the size of the largest component is an important quantity. For example, in a communication network like the Internet the size of the largest component represents the largest fraction of the network within which communication is possible and hence is a measure of the effectiveness of the network at doing its job [74, 81, 93, 94, 125, 322]. The size of the largest component is often equated with the graph theoretical concept of the "giant component" (see section 4.1), although technically the two are only the same in the limit of large graph size. The size of the second-largest component in a network is also measured sometimes. In networks well above the density at which a giant component first forms, the largest component is expected to be much larger than the second largest (section 4.1).

Goh et al. [174] have made a statistical study of the distribution of the "betweenness centrality" of vertices in networks. The betweenness centrality of a vertex  $i$  is the number of geodesic paths between other vertices that run through  $i$  [161, 362, 408]. Goh et al. show that betweenness appears to follow a power law for many networks and propose a classification of networks into two kinds based on the exponent of this power law. Betweenness centrality can also be viewed as a measure of network resilience [199, 311]—it tells us how many geodesic paths will get longer when a vertex is removed from the network. Latora and Marchiori [259, 260] have considered the harmonic mean distance between a vertex and all others, which they call the "efficiency" of the vertex. This, like betweenness centrality, can be viewed as a measure of network resilience, indicating how much effect on path length the removal of a vertex will have. A number of authors have looked at the eigenvalue spectra and eigenvectors of the graph Laplacian (or equivalently the adjacency matrix) of a network [55, 146, 151], which tells us about diffusion or vibration modes of the network, and about vertex centrality [66, 67] (see also the discussion of network search strategies in section 8.3.1).

Milo et al. [283, 367] have presented a novel analysis that picks out recurrent motifs—small subgraphs—from complete networks. They apply their method to genetic regulatory networks, food webs, neural networks and the World Wide Web, finding different motifs in each case. They have also made suggestions about the possible function of these motifs within the networks. In regulatory networks, for instance, they identify common subgraphs with particular switching functions in the system, such as gates and other feed-forward logical operations.

**4. Random Graphs.** The remainder of this review is devoted to our primary topic of study, the mathematics of model networks of various kinds. Recent work has focused on models of four general types, which we treat in four following sections. In

this section we look at random graph models, starting with the classic Poisson random graph of Rapoport [345], Solomonoff and Rapoport [377], and Erdős and Rényi [141, 142], and concentrating particularly on the generalized random graphs studied by Molloy and Reed [286, 287] and others. In section 5 we look at the somewhat neglected but potentially very useful Markov graphs and their more general forms, exponential random graphs and  $p^*$  models. In section 6 we look at the “small-world model” of Watts and Strogatz [415] and its generalizations. Then in section 7 we look at models of growing networks, particularly the models of Price [343] and Barabási and Albert [32], and generalizations. Finally, in section 8 we look at a number of models of processes occurring on networks, such as search and navigation processes, and network transmission and epidemiology.

The first serious attempt at constructing a model for large and (apparently) random networks was the “random net” of Rapoport and collaborators [345, 377], which was independently rediscovered a decade later by Erdős and Rényi [141], who studied it exhaustively and rigorously, and who gave it the name “random graph,” by which it is most often known today. Where necessary, we will here refer to it as the “Poisson random graph” to avoid confusion with other random graph models. It is also sometimes called the “Bernoulli graph.” As we will see in this section, the random graph, while illuminating, is inadequate to describe some important properties of real-world networks, and so has been extended in a variety of ways. In particular, the random graph’s Poisson degree distribution is quite unlike the highly skewed distributions of section 3.3 and Figure 3.2. Extensions of the model to allow for other degree distributions lead to the class of models known as “generalized random graphs,” “random graphs with arbitrary degree distributions,” and the “configuration model.”

Here we look first at the Poisson random graph, and then at its generalizations. Our treatment of the Poisson case is brief. A much more thorough treatment can be found in the books by Bollobás [63] and Janson, Łuczak, and Rucinski [210] and the review by Karoński [222].

**4.1. Poisson Random Graphs.** Solomonoff and Rapoport [377] and independently Erdős and Rényi [141] proposed the following extremely simple model of a network. Take some number  $n$  of vertices and connect each pair (or not) with probability  $p$  (or  $1 - p$ ).<sup>18</sup> This defines the model that Erdős and Rényi called  $G_{n,p}$ . In fact, technically,  $G_{n,p}$  is the *ensemble* of all such graphs in which a graph having  $m$  edges appears with probability  $p^m(1 - p)^{M-m}$ , where  $M = \frac{1}{2}n(n - 1)$  is the maximum possible number of edges. Erdős and Rényi also defined another, related model, which they called  $G_{n,m}$ , which is the ensemble of all graphs having  $n$  vertices and exactly  $m$  edges, each possible graph appearing with equal probability.<sup>19</sup> Here we will discuss  $G_{n,p}$ , but most of the results carry over to  $G_{n,m}$  in a straightforward fashion.

Many properties of the random graph are exactly solvable in the limit of large graph size, as was shown by Erdős and Rényi in a series of papers in the 1960s [141, 142, 143]. Typically the limit of large  $n$  is taken holding the mean degree  $z = p(n - 1)$  constant, in which case the model clearly has a Poisson degree distribution, since the

<sup>18</sup>Slight variations on the model are possible depending on whether one allows self-edges or not (i.e., edges that connect a vertex to itself), but this distinction makes a negligible difference to the average behavior of the model in the limit of large  $n$ .

<sup>19</sup>Those familiar with statistical mechanics will notice a similarity between these two models and the so-called canonical and grand canonical ensembles. In fact, the analogy is exact, and one can define equivalents of the Helmholtz and Gibbs free energies, which are generating functions for moments of graph properties over the distribution of graphs and which are related by a Lagrange transform with respect to the “field”  $p$  and the “order parameter”  $m$ .

presence or absence of edges is independent, and hence the probability of a vertex having degree  $k$  is

$$(4.1) \quad p_k = \binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{z^k e^{-z}}{k!},$$

with the last approximate equality becoming exact in the limit of large  $n$  and fixed  $k$ . This is the reason for the name “Poisson random graph.”

The expected structure of the random graph varies with the value of  $p$ . The edges join vertices together to form components, i.e., (maximal) subsets of vertices that are connected by paths through the network. Both Solomonoff and Rapoport and also Erdős and Rényi demonstrated what is for our purposes the most important property of the random graph, that it possesses what we would now call a phase transition, from a low-density, low- $p$  state in which there are few edges and all components are small, having an exponential size distribution and finite mean size, to a high-density, high- $p$  state in which an extensive (i.e.,  $O(n)$ ) fraction of all vertices are joined together in a single *giant component*, the remainder of the vertices occupying smaller components with again an exponential size distribution and finite mean size.

We can calculate the expected size of the giant component from the following simple heuristic argument. Let  $u$  be the fraction of vertices on the graph that do not belong to the giant component, which is also the probability that a vertex chosen uniformly at random from the graph is not in the giant component. The probability of a vertex not belonging to the giant component is also equal to the probability that none of the vertex's network neighbors belong to the giant component, which is just  $u^k$  if the vertex has degree  $k$ . Averaging this expression over the probability distribution of  $k$ , (4.1), we then find the following self-consistency relation for  $u$  in the limit of large graph size:

$$(4.2) \quad u = \sum_{k=0}^{\infty} p_k u^k = e^{-z} \sum_{k=0}^{\infty} \frac{(zu)^k}{k!} = e^{z(u-1)}.$$

The fraction  $S$  of the graph occupied by the giant component is  $S = 1 - u$  and hence

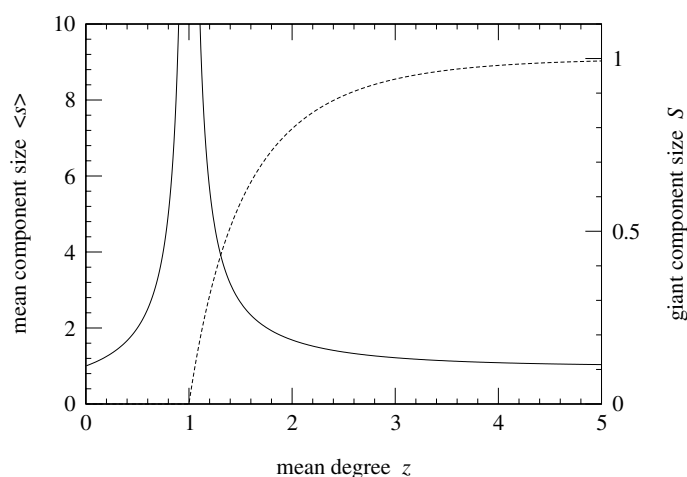
$$(4.3) \quad S = 1 - e^{-zS}.$$

By an argument only slightly more complex, which we give in the following section, we can show that the mean size  $\langle s \rangle$  of the component to which a randomly chosen vertex belongs (for nongiant components) is

$$(4.4) \quad \langle s \rangle = \frac{1}{1 - z + zS}.$$

The form of these two quantities is shown in Figure 4.1. Equation (4.3) is transcendental and has no closed-form solution, but it is easy to see that for  $z < 1$  its only nonnegative solution is  $S = 0$ , while for  $z > 1$  there is also a nonzero solution, which is the size of the giant component. The phase transition occurs at  $z = 1$ . This is also the point at which  $\langle s \rangle$  diverges, a behavior that will be recognized by those familiar with the theory of phase transitions:  $S$  plays the role of the order parameter in this transition and  $\langle s \rangle$  the role of the order-parameter fluctuations. The corresponding critical exponents, defined by  $S \sim (z - 1)^\beta$  and  $\langle s \rangle \sim |z - 1|^{-\gamma}$ , take the values  $\beta = 1$  and  $\gamma = 1$ . Precisely at the transition,  $z = 1$ , there is a “double jump”—the mean size





**Fig. 4.1** The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line), for the Poisson random graph, (4.3) and (4.4).

of the largest component in the graph goes as  $O(n^{2/3})$  for  $z = 1$ , rather than  $O(n)$  as it does above the transition. The components at the transition have a power-law size distribution with exponent  $\tau = \frac{5}{2}$  (or  $\frac{3}{2}$  if one asks about the component to which a randomly chosen vertex belongs). We look at these results in more detail in the next section for the more general “configuration model.”

The random graph reproduces well one of the principal features of real-world networks discussed in section 3, namely, the small-world effect. The mean number of neighbors a distance  $\ell$  away from a vertex in a random graph is  $z^\ell$ , and hence the value of  $\ell$  needed to encompass the entire network is  $z^\ell \simeq n$ . Thus a typical distance through the network is  $\ell = \log n / \log z$ , which satisfies the definition of the small-world effect given in section 3.1. Rigorous results to this effect can be found in, for instance, [61] and [63]. However, in almost all other respects, the properties of the random graph do not match those of networks in the real world. It has a low clustering coefficient: the probability of connection of two vertices is  $p$  regardless of whether they have a common neighbor, and hence  $C = p$ , which tends to zero as  $n^{-1}$  in the limit of large system size [415]. The model also has a Poisson degree distribution, quite unlike the distributions in Figure 3.2. It has entirely random mixing patterns, no correlation between degrees of adjacent vertices, and no community structure, and navigation is impossible on a random graph using local algorithms [237, 238, 313, 317, 400]. In short it makes a good straw man but is rarely taken seriously in the modeling of real systems.

Nonetheless, much of our basic intuition about the way networks behave comes from the study of the random graph. In particular, the presence of the phase transition and the existence of a giant component are ideas that underlie much of the work described in this review. One often talks about the giant component of a network, meaning in fact the largest component; one looks at the sizes of smaller components, often finding them to be much smaller than the largest component; one sees a giant component transition in many of the more sophisticated models that we will look at in the coming sections. All of these are ideas that started with the Poisson random graph.

**4.2. Generalized Random Graphs.** Random graphs can be extended in a variety of ways to make them more realistic. The property of real graphs that is simplest to incorporate is the property of non-Poisson degree distributions, which leads us to the so-called configuration model. Here we examine this model in detail; in sections 4.2.3–4.2.5 we describe further generalizations of the random graph to add other features.

**4.2.1. The Configuration Model.** Consider the model defined in the following way. We specify a degree distribution  $p_k$ , such that  $p_k$  is the fraction of vertices in the network having degree  $k$ . We choose a *degree sequence*, which is a set of  $n$  values of the degrees  $k_i$  of vertices  $i = 1, \dots, n$ , from this distribution. We can think of this as giving each vertex  $i$  in our graph  $k_i$  “stubs” or “spokes” sticking out of it, which are the ends of edges-to-be. Then we choose pairs of stubs at random from the network and connect them together. It is straightforward to demonstrate [286] that this process generates every possible topology of a graph with the given degree sequence with equal probability.<sup>20</sup> The *configuration model* is defined as the ensemble of graphs so produced, with each having equal weight.<sup>21</sup>

Since the 1970s the configuration model has been studied by a number of authors [46, 47, 60, 88, 89, 267, 286, 287, 322, 424]. An exact condition is known in terms of  $p_k$  for the model to possess a giant component [286], the expected size of that component is known [287], and the average size of nongiant components both above and below the transition is known [322], along with a variety of other properties, such as mean numbers of vertices a given distance away from a central vertex and typical vertex–vertex distances [88]. Here we give a brief derivation of the main results using the generating function formalism of Newman, Strogatz, and Watts [322]. More rigorous treatments of the same results can be found in [88, 89, 286, 287].

There are two important points to grasp about the configuration model. First,  $p_k$  is, in the limit of large graph size, the distribution of degrees of vertices in our graph, but the degree of the vertex we reach by following a randomly chosen edge on the graph is not given by  $p_k$ . Since there are  $k$  edges that arrive at a vertex of degree  $k$ , we are  $k$  times as likely to arrive at that vertex as we are at some other vertex that has degree 1. Thus the degree distribution of the vertex at the end of a randomly chosen edge is proportional to  $kp_k$ . In most cases, we are interested in how many edges there are leaving such a vertex other than the one we arrived along, i.e., in the so-called *excess degree*, which is one less than the total degree of the vertex. In the configuration model, the excess degree has a distribution  $q_k$  given by

$$(4.5) \quad q_k = \frac{(k+1)p_{k+1}}{\sum_k kp_k} = \frac{(k+1)p_{k+1}}{z},$$

where  $z = \sum_k kp_k$  is, as before, the mean degree in the network.

<sup>20</sup>Each possible graph can be generated  $\prod_i k_i!$  different ways, since the stubs around each vertex are indistinguishable. This factor is a constant for a given degree sequence, and hence each graph appears with equal probability.

<sup>21</sup>An alternative model has recently been proposed by Chung and Lu [88, 89]. In their model, each vertex  $i$  is assigned a *desired* degree  $k_i$  chosen from the distribution of interest, and then  $m = \frac{1}{2} \sum_i k_i$  edges are placed between vertex pairs  $(i, j)$  with probability proportional to  $k_i k_j$ . This model has the disadvantage that the final degree sequence is not in general precisely equal to the desired degree sequence, but it has some significant calculational advantages that make the derivation of rigorous results easier. It is also a logical generalization of the Poisson random graph, in a way that the configuration model is not. The approach has been further developed by Dorogovtsev, Mendes, and Samukhin [128] and Caldarelli et al. [78].

The second important point about the model is that the chance of finding a loop in a small component of the graph goes as  $n^{-1}$ . The number of vertices in a non-giant component is  $O(n^{-1})$ , and hence the probability of there being more than one path between any pair of vertices is also  $O(n^{-1})$  for suitably well-behaved degree distributions.<sup>22</sup> This property is crucial to the solution of the configuration model but is definitely not true of most real-world networks (see section 3.2). It is an open question how much the predictions of the model would change if we were able to incorporate the true loop structure of real networks into it.

We now proceed by defining two generating functions for the distributions  $p_k$  and  $q_k$ :<sup>23</sup>

$$(4.6) \quad G_0(x) = \sum_{k=0}^{\infty} p_k x^k, \quad G_1(x) = \sum_{k=0}^{\infty} q_k x^k.$$

Note that, using (4.5), we also find that  $G_1(x) = G'_0(x)/z$ , which is occasionally convenient. Then the generating function  $H_1(x)$  for the total number of vertices reachable by following an edge satisfies the self-consistency condition

$$(4.7) \quad H_1(x) = xG_1(H_1(x)).$$

This equation says that when we follow an edge, we find at least one vertex at the other end (the factor of  $x$  on the right-hand side), plus some other clusters of vertices (each represented by  $H_1$ ) which are reachable by following other edges attached to that one vertex. The number of these other clusters is distributed according to  $q_k$ , hence the appearance of  $G_1$ . A detailed derivation of (4.7) is given in [322].

The total number of vertices reachable from a randomly chosen vertex, i.e., the size of the component to which such a vertex belongs, is generated by  $H_0(x)$ , where

$$(4.8) \quad H_0(x) = xG_0(H_1(x)).$$

The solution of (4.7) and (4.8) gives us the entire distribution of component sizes. Mean component size below the phase transition in the region where there is no giant component is given by

$$(4.9) \quad \langle s \rangle = H'_0(1) = 1 + \frac{G'_0(1)}{1 - G'_1(1)} = 1 + \frac{z_1^2}{z_1 - z_2},$$

where  $z_1 = z = \langle k \rangle = G'_0(1)$  is the average number of neighbors of a vertex and  $z_2 = \langle k^2 \rangle - \langle k \rangle = G'_0(1)G'_1(1)$  is the average number of second neighbors. We see that this diverges when  $z_1 = z_2$ , or equivalently when

$$(4.10) \quad G'_1(1) = 1.$$

This point marks the phase transition at which a giant component first appears. Substituting (4.6) into (4.10), we can also write the condition for the phase transition as

$$(4.11) \quad \sum_k k(k-2)p_k = 0.$$

<sup>22</sup>Using arguments similar to those leading to (4.14), we can show that the density of loops in small components will tend to zero as graph size becomes large, provided that  $z$  is finite and  $\langle k^2 \rangle$  grows slower than  $n^{1/2}$ . See also footnote 25.

<sup>23</sup>Traditionally, the independent variable in a generating function is denoted  $z$ , but here we use  $x$  to avoid confusion with the mean degree  $z$ .

Indeed, since this sum increases monotonically as edges are added to the graph, it follows that the giant component exists if and only if this sum is positive. A more rigorous derivation of this result has been given by Molloy and Reed [286].

Above the transition there is a giant component which occupies a fraction  $S$  of the graph. If we define  $u$  to be the probability that a randomly chosen edge leads to a vertex that is not a part of this giant component, then, by an argument precisely analogous to the one preceding (4.3), this probability must satisfy the self-consistency condition  $u = G_1(u)$  and  $S$  is given by the solution of

$$(4.12) \quad S = 1 - G_0(u), \quad u = G_1(u).$$

An equivalent result is derived in [287]. Normally the equation for  $u$  cannot be solved in closed form, but once the generating functions are known a solution can be found to any desired level of accuracy by numerical iteration. And once the value of  $S$  is known, the mean size of small components above the transition can be found by subtracting off the giant component and applying the arguments that led to (4.9) again, giving

$$(4.13) \quad \langle s \rangle = 1 + \frac{zu^2}{[1 - S][1 - G'_1(u)]}.$$

The result is a behavior qualitatively similar to that of the Poisson random graph, with a continuous phase transition at a point defined by (4.11), characterized by the appearance of a giant component and the divergence of the mean size of nongiant components. The ratio  $z_2/z_1$  of the mean number of vertices two steps away to the number one step away plays the role of the independent parameter governing the transition, as the mean degree  $z$  does in the Poisson case, and one can again define critical exponents for the transition, which take the same values as for the Poisson case,  $\beta = \gamma = 1$ ,  $\tau = \frac{5}{2}$ .

We can also find an expression for the clustering coefficient, (3.3), of the configuration model. A simple calculation shows that [136, 318]

$$(4.14) \quad C = \frac{1}{nz_1} \left[ \frac{z_2}{z_1} \right]^2 = \frac{z}{n} \left[ \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle^2} \right]^2,$$

which is the value  $C = z/n$  for the Poisson random graph times an extra factor that depends on  $z$  and on the ratio  $\langle k^2 \rangle / \langle k \rangle^2$ . Thus  $C$  will normally go to zero as  $n^{-1}$  for large graphs, but for highly skewed degree distributions, like some of those in Figure 3.2, the factor of  $\langle k^2 \rangle / \langle k \rangle^2$  can be quite large, so that  $C$  is not necessarily negligible for the graph sizes seen in empirical studies of networks (see below).

**4.2.2. Example: Power-Law Degree Distribution.** As an example of the application of these results, consider the much-studied case of a network with a power-law degree distribution:

$$(4.15) \quad p_k = \begin{cases} 0 & \text{for } k = 0, \\ k^{-\alpha} / \zeta(\alpha) & \text{for } k \geq 1 \end{cases}$$

for given constant  $\alpha$ . Here  $\zeta(\alpha)$  is the Riemann  $\zeta$ -function, which functions as a normalizing constant. Substituting into (4.6) we find that

$$(4.16) \quad G_0(x) = \frac{\text{Li}_\alpha(x)}{\zeta(\alpha)}, \quad G_1(x) = \frac{\text{Li}_{\alpha-1}(x)}{x\zeta(\alpha-1)},$$

where  $\text{Li}_n(x)$  is the  $n$ th polylogarithm of  $x$ . Then (4.10) tells us that the phase transition occurs at the point

$$(4.17) \quad \zeta(\alpha - 2) = 2\zeta(\alpha - 1),$$

which gives a critical value for  $\alpha$  of  $\alpha_c = 3.4788\dots$ . Below this value a giant component exists; above it there is no giant component. For  $\alpha < \alpha_c$ , the value of the variable  $u$  of (4.12) is

$$(4.18) \quad u = \frac{\text{Li}_{\alpha-1}(u)}{u\zeta(\alpha-1)},$$

which gives  $u = 0$  below  $\alpha = 2$  and hence  $S = 1$ . Thus the giant component occupies the entire graph below this point, or more strictly, a randomly chosen vertex belongs to the giant component with probability 1 in the limit of large graph size (but see the following discussion of the clustering coefficient and footnote 25). In the range  $2 < \alpha < \alpha_c$  we have a nonzero giant component whose size is given by (4.12). All of these results were first shown by Aiello, Chung, and Lu [8].

We can also calculate the clustering coefficient for the power-law case using (4.14). For  $\alpha < 3$  we have  $\langle k^2 \rangle \sim k_{\max}^{3-\alpha}$ , where  $k_{\max}$  is the maximum degree in the network. Using (3.13) for  $k_{\max}$ , equation (4.14) then gives

$$(4.19) \quad C \sim n^{-\beta}, \quad \beta = \frac{3\alpha - 7}{\alpha - 1}.$$

This gives interesting behavior for the typical values  $2 \leq \alpha \leq 3$  of the exponent  $\alpha$  seen in most networks (see Table 3.1). If  $\alpha > \frac{7}{3}$ , then  $C$  tends to zero as the graph becomes large, although it does so slower than the  $C \sim n^{-1}$  of the Poisson random graph provided  $\alpha < 3$ . At  $\alpha = \frac{7}{3}$ ,  $C$  becomes constant (or logarithmic) in the graph size, and for  $\alpha < \frac{7}{3}$  it actually increases with increasing system size.<sup>24</sup> Thus for scale-free networks with smaller exponents  $\alpha$ , we would not be surprised to see quite substantial values of the clustering coefficient, even if the pattern of connections were completely random.<sup>25</sup> This mechanism can, for instance, account for much of the clustering seen in the World Wide Web [318].

**4.2.3. Directed Graphs.** Substantially more sophisticated extensions of random graph models are possible than the simple first example given above. In this and the next few sections we list some of the many possibilities, starting with directed graphs.

Each vertex in a directed graph has both an in-degree  $j$  and an out-degree  $k$ , and the degree distribution therefore becomes, in general, a double distribution  $p_{jk}$  over both degrees, as discussed in section 3.3. The generating function for such a distribution is a function of two variables

$$(4.20) \quad \mathcal{G}(x, y) = \sum_{j,k} p_{jk} x^j y^k.$$

<sup>24</sup>For sufficiently large networks this implies that the clustering coefficient will be greater than 1. Physically this means that there will be more than one edge on average between two vertices that share a common neighbor.

<sup>25</sup>This means in fact that the generating function formalism breaks down for  $\alpha < \frac{7}{3}$ , invalidating some of the preceding results for the power-law graph, since a fundamental assumption of the method is that there are no short loops in the network. Aiello, Chung, and Lu [8] get around this problem by assuming that the degree distribution is cut off at  $k_{\max} \sim n^{1/\alpha}$  (see section 3.3.2), which gives  $C \rightarrow 0$  as  $n \rightarrow \infty$  for all  $\alpha > 2$ . This, however, is somewhat artificial; in real power-law networks there is normally no such cutoff.

Each vertex  $A$  also belongs to an *in-component* and an *out-component*, which are, respectively, the set of vertices from which  $A$  can be reached and the set that can be reached from  $A$ , by following directed edges only in their forward direction. There is also the *strongly connected component*, which is the set of vertices which can both reach and be reached from  $A$ . In a random directed graph with a given degree distribution, the giant in-, out-, and strongly connected components can all be shown [322] to form at a single transition that takes place when

$$(4.21) \quad \sum_{jk} (2jk - j - k) p_{jk} = 0.$$

Defining generating functions for in- and out-degree separately and their excess-degree counterparts,

$$(4.22a) \quad F_0(x) = \mathcal{G}(x, 1), \quad F_1(x) = \frac{1}{z} \frac{\partial \mathcal{G}}{\partial y} \bigg|_{y=1},$$

$$(4.22b) \quad G_0(y) = \mathcal{G}(1, y), \quad G_1(y) = \frac{1}{z} \frac{\partial \mathcal{G}}{\partial x} \bigg|_{x=1},$$

the sizes of the giant out-, in-, and strongly connected components are given by [322, 125]

$$(4.23a) \quad S_{\text{out}} = 1 - F_0(u),$$

$$(4.23b) \quad S_{\text{in}} = 1 - G_0(v),$$

$$(4.23c) \quad S_{\text{str}} = 1 - \mathcal{G}(u, 1) - \mathcal{G}(1, v) + \mathcal{G}(u, v),$$

where

$$(4.24) \quad u = F_1(u), \quad v = G_1(v).$$

**4.2.4. Bipartite Graphs.** Another class of generalizations of random graph models is to networks with more than one type of vertex. One of the simplest and most important examples of such a network is the bipartite graph, which has two types of vertices and edges running only between vertices of unlike types. As discussed in section 1.1, many social networks are bipartite, forming what the sociologists call *affiliation networks*, i.e., networks of individuals joined by common membership of groups. In such networks the individuals and the groups are represented by the two vertex types with edges between them representing group membership. Networks of CEOs [167, 168], boards of directors [104, 105, 268], and collaborations of scientists [312] and film actors [415] are all examples of affiliation networks. Some other networks, such as the railway network studied by Sen et al. [365], are also bipartite, and bipartite graphs have been used as the basis for models of sexual contact networks [144, 314].

Bipartite graphs have two degree distributions, one each for the two types of vertices. Since the total number of edges attached to each type of vertex is the same, the means  $\mu$  and  $\nu$  of the two distributions are related to the numbers  $M$  and  $N$  of the types of vertices by  $\mu/M = \nu/N$ . One can define generating functions as before for the two types of vertices, generating both the degree distribution and the excess degree distribution, and denoted  $f_0(x)$ ,  $f_1(x)$ ,  $g_0(x)$ , and  $g_1(x)$ . Then, for example, we can show that there is a phase transition at which a giant component appears when

$f'_1(1)g'_1(1) = 1$ . Expressions for the expected size of giant and nongiant components can easily be derived [322].

In many cases, graphs that are fundamentally bipartite are actually studied by projecting them down onto one set of vertices or the other—so called one-mode projections. For example, in the study of boards of directors of companies, it has become standard to look at board “interlocks.” Two boards are said to be interlocked if they share one or more common members, and the graph of board interlocks is the one-mode projection of the full board graph onto the vertices representing just the boards. Many results for these one-mode projections can also be extracted from the generating function formalism. To give one example, the projected networks do not have a vanishing clustering coefficient  $C$  in the limit of large system size but instead can be shown to obey [322]

$$(4.25) \quad \frac{1}{C} - 1 = \frac{(\mu_2 - \mu_1)(\nu_2 - \nu_1)^2}{\mu_1\nu_1(2\nu_1 - 3\nu_2 + \nu_3)},$$

where  $\mu_n$  and  $\nu_n$  are the  $n$ th moments of the degree distributions of the two vertex types.

More complicated types of network structure can be introduced by increasing the number of different types of vertices beyond two, and by relaxing the patterns of connection between vertex types. For example, one can define a model with the type of mixing matrix shown in Table 3.2 and solve exactly for many of the standard properties [317, 373].

**4.2.5. Degree Correlations.** The type of degree correlations discussed in section 3.6 can also be introduced into a random graph model [313]. Extending the formalism of section 3.5, we can define the probability distribution  $e_{jk}$  to be the probability that a randomly chosen edge on a graph connects vertices of excess degrees  $j$  and  $k$ . On an undirected graph, this quantity is symmetric and satisfies

$$(4.26) \quad \sum_{jk} e_{jk} = 1, \quad \sum_j e_{jk} = q_k.$$

Then the equivalent of (4.12) is

$$(4.27) \quad S = 1 - p_0 - \sum_{k=1}^{\infty} p_k u_{k-1}^k, \quad u_j = \frac{\sum_k e_{jk} u_k^k}{\sum_k e_{jk}},$$

which must be solved self-consistently for the entire set  $\{u_k\}$  of quantities, one for each possible value of the excess degree. The phase transition at which a giant component appears takes place when  $\det(\mathbf{I} - \mathbf{m}) = 0$ , where  $\mathbf{m}$  is the matrix with elements  $m_{jk} = ke_{jk}/q_j$ . Matrix conditions of this form appear to be the typical generalization of the criterion for the appearance of a giant component to graphs with nontrivial mixing patterns [58, 317, 399].

Two other random graph models for degree correlations are also worth mentioning. One is the exponential random graph, which we study in more detail in the following section. This is a general model, which has been applied to the particular problem of degree correlations by Berg and Lässig [48].

A more specialized model that aims to explain the degree anticorrelations seen in the Internet has been put forward by Maslov, Sneppen, and Zaliznyak [274]. They suggest that these anticorrelations are a simple result of the fact that the Internet

graph has at most one edge between any vertex pair. Thus they are led to consider the ensemble of all networks with a given degree sequence and no double edges. (The configuration model, by contrast, allows double edges, and typical graphs usually have at least a few such edges, which would disqualify them from membership in the ensemble of Maslov et al.) The ensemble with no duplicate edges, it turns out, is hard to treat analytically [47, 406], so Maslov et al. instead investigate it numerically, sampling the ensemble at random using a Monte Carlo algorithm. Their results appear to indicate that anticorrelations of the type seen in the Internet do indeed arise as a finite-size effect within this model. (An alternative explanation of the same observations has been put forward by Capocci, Caldarelli, and De Los Rios [83], who use a modified version of the model of Barabási and Albert discussed in section 7.2 to show that correlations can arise through network growth processes.)

**5. Exponential Random Graphs and Markov Graphs.** The generalized random graph models of the previous sections effectively address one of the principal shortcomings of early network models such as the Poisson random graph: their unrealistic degree distribution. However, they have a serious shortcoming in that they fail to capture the common phenomenon of transitivity described in section 3.2. The only solvable random graph models that currently incorporate transitivity are the bipartite and community-structured models of section 4.2.4 and certain dual-graph models [344], and these cover rather special cases. For general networks we currently have no idea how to incorporate transitivity into random graph models; the crucial property of independence between the neighbors of a vertex is destroyed by the presence of short loops in a network, invalidating all the techniques used to derive solutions. Some approximate methods may be useful in limited ways [316], or perhaps some sort of perturbative analysis will prove possible, but no progress has yet been made in this direction.

The main hope for progress in understanding the effects of transitivity, which are certainly substantial, seems to lie in formulating a completely different model or models based on some alternative ensemble of graph structures. In this and the following section we describe two candidate models: the Markov graphs of Holland and Leinhardt [193], Strauss [384], and Frank and Strauss [160], and the small-world model of Watts and Strogatz [415].

Strauss [384] considers *exponential random graphs*, also (in a slightly generalized form) called *p\* models* [22, 409], which are a class of graph ensembles of fixed vertex number  $n$  defined by analogy with the Boltzmann ensemble of statistical mechanics.<sup>26</sup> Let  $\{\epsilon_i\}$  be a set of measurable properties of a single graph, such as the number of edges, the number of vertices of given degree, or the number of triangles of edges in the graph. These quantities play a role similar to energy in statistical mechanics. And let  $\{\beta_i\}$  be a set of inverse-temperature or field parameters, whose values we are free to choose. We then define the exponential random graph model to be the set of all possible graphs (undirected in the simplest case) of  $n$  vertices in which each graph  $G$  appears with probability

$$(5.1) \quad P(G) = \frac{1}{Z} \exp\left(-\sum_i \beta_i \epsilon_i\right),$$

<sup>26</sup>Indeed, in a development typical of this highly interdisciplinary field, exponential random graphs have recently been rediscovered, apparently quite independently, by physicists [48, 77].



where the partition function  $Z$  is

$$(5.2) \quad Z = \sum_G \exp\left(-\sum_i \beta_i \epsilon_i\right).$$

For a sufficiently large set of temperature parameters  $\{\beta_i\}$ , this definition can encompass any probability distribution over graphs that we desire, although its practical application requires that the size of the set be limited to a reasonably small number.

The calculation of the ensemble average of a graph observable  $\epsilon_i$  is then found by taking a suitable derivative of the (reduced) free energy  $f = -\log Z$ :

$$(5.3) \quad \langle \epsilon_i \rangle = \sum_G \epsilon_i(G) P(G) = \frac{1}{Z} \sum_G \epsilon_i \exp\left(-\sum_i \beta_i \epsilon_i\right) = \frac{\partial f}{\partial \beta_i}.$$

Thus, the free energy is a generating function for the expectation values of the observables, in a manner familiar from statistical field theory. If a particular observable of interest does not appear in the exponent of (5.1) (the “graph Hamiltonian”), then one can simply introduce it, with a corresponding temperature  $\beta_i$  which is set to zero.

While these preliminary developments appear elegant in principle, little real progress has been made. One would like to find the appropriate Gaussian field theory for which  $f$  can be expressed in closed form, and then perturb around it to derive a diagrammatic expansion for the effects of higher-order graph operators. In fact, one can show that the Feynman diagrams for the expansion *are* the networks themselves. Unfortunately, carrying through the entire field-theoretic program has not proved easy. The general approach one should take is clear [48, 77], but the mechanics appear intractable for most cases of interest. Some progress can be made by restricting ourselves to *Markov graphs*, which are the subset of graphs in which the presence or absence of an edge between two vertices in the graph is correlated only with those edges that share one of the same two vertices—edge pairs that are disjoint (have no vertices in common) are uncorrelated. Overall, however, the question of how to carry out calculations in exponential random graph ensembles is an open one.

In the absence of analytic progress on the model, therefore, researchers have turned to Monte Carlo simulation, a technique to which the exponential random graph lends itself admirably. Once the values of the parameters  $\{\beta_i\}$  are specified, the form (5.1) of  $P(G)$  makes generation of graphs correctly sampled from the ensemble straightforward using a Metropolis–Hastings-type Markov chain method. One defines an ergodic move-set in the space of graphs with given  $n$ , and then repeatedly generates moves from this set, accepting them with probability

$$(5.4) \quad p = \begin{cases} 1 & \text{if } P(G') > P(G), \\ P(G')/P(G) & \text{otherwise,} \end{cases}$$

and rejecting them with probability  $1 - p$ , where  $G'$  is the graph after performance of the move. Because of the particular form, (5.1), assumed for  $P(G)$ , this acceptance probability is particularly simple to calculate:

$$(5.5) \quad \frac{P(G')}{P(G)} = \exp\left(-\sum_i \beta_i [\epsilon'_i - \epsilon_i]\right).$$

This expression is independent of the value of the partition function, and its evaluation involves calculating only the differences  $\epsilon'_i - \epsilon_i$  of the energy-like graph properties  $\epsilon_i$ , which for local move-sets and local properties can often be accomplished in

time independent of graph size. The following are suitable move-sets: (a) addition and removal of edges between randomly chosen vertex pairs for the case of variable edge numbers; (b) movement of edges randomly from one place to another for the case of fixed edge numbers but variable degree sequence; (c) edge swaps of the form  $\{(v_1, w_1), (v_2, w_2)\} \rightarrow \{(v_1, v_2), (w_1, w_2)\}$  for the case of fixed degree sequence, where  $(v_1, w_1)$  denotes an edge from vertex  $v_1$  to vertex  $w_1$ . Monte Carlo algorithms of this type are straightforward to implement and appear to converge quickly, allowing us to study quite large graphs.

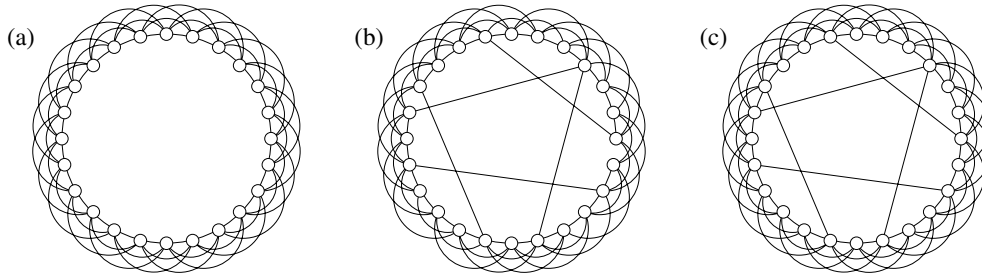
There is, however, one unfortunate pathology of the exponential random graph that plagues numerical work, and particularly affects Markov graphs as they are used to model transitivity. If, for example, we include a term in the graph Hamiltonian that is linear in the number of triangles in the graph, with an accompanying positive temperature favoring these triangles, then the model has a tendency to “condense,” forming regions of the graph that are essentially complete cliques—subsets of vertices within which every possible edge exists. It is easy to see why the model shows this behavior: cliques have the largest number of triangles for the number of edges they contain, and are therefore highly energetically favored, while costing the system a minimum in entropy by virtue of leaving the largest possible number of other edges free to contribute to the (presumably extensive) entropy of the rest of the graph. Networks in the real world, however, do not seem to have this sort of “clumpy” transitivity—regions of cliquishness contributing heavily to the clustering coefficient, separated by other regions with few triangles. It is not clear how this problem is to be circumvented, although for higher temperatures (lower values of the parameters  $\{\beta_i\}$ ) it is less problematic, since higher temperatures favor entropy over energy.

Another area in which some progress has been made is in techniques for extracting appropriate values for the temperature parameters in the model from real-world network data. Procedures for doing this have been particularly important for social network applications. Parameters so extracted can be fed back into the Monte Carlo graph generation methods described above to generate model graphs which have similar statistical properties to their real-world counterparts and which can be used for hypothesis testing or as a substrate for further network simulations. Reviews of parameter extraction techniques can be found in [22] and [371].

**6. The Small-World Model.** A less sophisticated but more tractable model of a network with high transitivity is the *small-world model* proposed by Watts and Strogatz [410, 411, 415].<sup>27</sup> As discussed in section 2.3, many networks have a geographical component to them; the vertices of the network have positions in space, and in many cases it is reasonable to assume that geographical proximity will play a role in deciding which vertices are connected to which others. The small-world model starts from this idea by positing a network built on a low-dimensional regular lattice and then adding or moving edges to create a low density of “shortcuts” that join remote parts of the lattice to one another.

Small-world models can be built on lattices of any dimension or topology, but the best studied case by far is one-dimensional. If we take a one-dimensional lattice of  $L$  vertices with periodic boundary conditions, i.e., a ring, and join each vertex to its neighbors  $k$  or fewer lattice spacings away, we get a system like Figure 6.1a, with  $Lk$  edges. The small-world model is then created by taking a small fraction of the edges in this graph and “rewiring” them. The rewiring procedure involves going through

<sup>27</sup> An equivalent model was proposed by Ball, Mollison, and Scalia-Tomba [28] some years earlier, as a model of the spread of disease between households, but appears not to have been widely adopted.

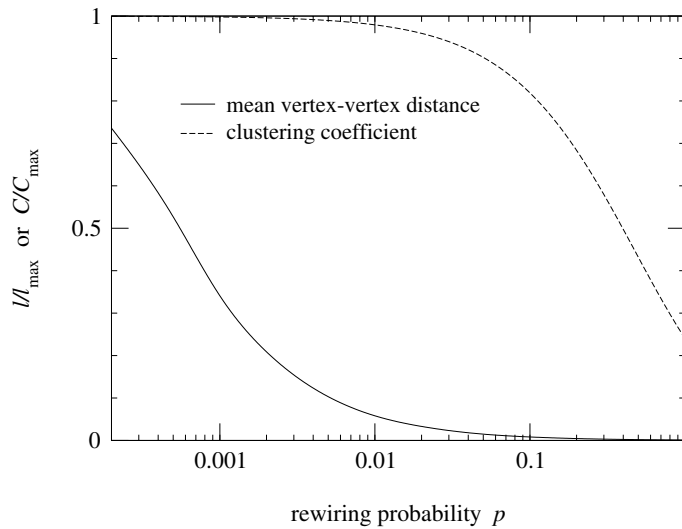


**Fig. 6.1** (a) A one-dimensional lattice with connections between all vertex pairs separated by  $k$  or fewer lattice spacing, with  $k = 3$  in this case. (b) The small-world model [415, 411] is created by choosing at random a fraction  $p$  of the edges in the graph and moving one end of each to a new location, also chosen uniformly at random. (c) A slight variation on the model [323, 288] in which shortcuts are added randomly between vertices, but no edges are removed from the underlying one-dimensional lattice.

each edge in turn and, with probability  $p$ , moving one end of that edge to a new location chosen uniformly at random from the lattice, except that no double edges or self-edges are ever created. This process is illustrated in Figure 6.1b.

The rewiring process allows the small-world model to interpolate between a regular lattice and something which is similar, though not identical (see below), to a random graph. When  $p = 0$ , we have a regular lattice. It is not hard to show that the clustering coefficient of this regular lattice is  $C = (3k - 3)/(4k - 2)$ , which tends to  $\frac{3}{4}$  for large  $k$ . The regular lattice, however, does not show the small-world effect. Mean geodesic distances between vertices tend to  $L/4k$  for large  $L$ . When  $p = 1$ , every edge is rewired to a new random location and the graph is almost a random graph, with typical geodesic distances on the order of  $\log L / \log k$  but very low clustering  $C \simeq 2k/L$  (see section 4.1). As Watts and Strogatz showed by numerical simulation, however, there exists a sizable region in between these two extremes for which the model has both low path lengths and high transitivity—see Figure 6.2.

The original model proposed by Watts and Strogatz is somewhat baroque. The fact that only one end of each chosen edge is rewired, not both, that no vertex is ever connected to itself, and that an edge is never added between vertex pairs where there is already one makes it quite difficult to enumerate or average over the ensemble of graphs. For the purposes of mathematical treatment, the model can be simplified considerably by rewiring both ends of each chosen edge, and by allowing both double and self-edges. This results in a system that genuinely interpolates between a regular lattice and a random graph. Another variant of the model that has become popular was proposed independently by Monasson [288] and by Newman and Watts [323]. In this variant, no edges are rewired. Instead “shortcuts” joining randomly chosen vertex pairs are added to the low-dimensional lattice—see Figure 6.1c. The parameter  $p$  governing the density of these shortcuts is defined so as to make it as similar as possible to the parameter  $p$  in the first version of the model:  $p$  is defined as the probability per edge on the underlying lattice of there being a shortcut anywhere in the graph. Thus the mean total number of shortcuts is  $Lkp$  and the mean degree is  $2Lk(1 + p)$ . This version of the model has the desirable property that no vertices ever become disconnected from the rest of the network, and hence the mean vertex–vertex distance is always formally finite. Both this version and the original have been studied at some length in the mathematical and physical literature [308].



**Fig. 6.2** The clustering coefficient  $C$  and mean vertex-vertex distance  $\ell$  in the small-world model of Watts and Strogatz [415] as a function of the rewiring probability  $p$ . For convenience, both  $C$  and  $\ell$  are divided by their maximum values, which they assume when  $p = 0$ . Between the extremes  $p = 0$  and  $p = 1$ , there is a region in which clustering is high and mean vertex-vertex distance is simultaneously low.

**6.1. Clustering Coefficient.** The clustering coefficient for both versions of the small-world model can be calculated relatively easily. For the original version, Barrat and Weigt [40] showed that

$$(6.1) \quad C = \frac{3(k-1)}{2(2k-1)}(1-p)^3,$$

while for the version without rewiring, Newman [315] showed that

$$(6.2) \quad C = \frac{3(k-1)}{2(2k-1) + 4kp(p+2)}.$$

**6.2. Degree Distribution.** The degree distribution of the small-world model does not match most real-world networks very well, although this is not surprising, since this was not a goal of the model in the first place. For the version without rewiring, each vertex has degree at least  $2k$ , for the edges of the underlying regular lattice, plus a binomially distributed number of shortcuts. Hence the probability  $p_j$  of having degree  $j$  is

$$(6.3) \quad p_j = \begin{cases} 0 & \text{if } j < 2k, \\ \binom{L}{j-2k} \left[\frac{2kp}{L}\right]^{j-2k} \left[1 - \frac{2kp}{L}\right]^{L-j+2k} & \text{otherwise.} \end{cases}$$

For the rewired version of the model, the distribution has a lower cutoff at  $k$  rather than  $2k$  and is rather more complicated. The full expression is [40]

$$(6.4) \quad p_j = \begin{cases} 0 & \text{if } j < k, \\ \sum_{n=0}^{\min(j-k, k)} \binom{k}{n} (1-p)^n p^{k-n} \frac{(pk)^{j-k-n}}{(j-k-n)!} e^{-pk} & \text{otherwise.} \end{cases}$$

**6.3. Average Path Length.** By far the most attention has been focused on the average geodesic path length of the small-world model. We denote this quantity  $\ell$ . We do not have any exact solution for the value of  $\ell$  yet, but a number of partial exact results are known, including scaling forms, as well as some approximate solutions for its behavior as a function of the model's parameters.

In the limit  $p \rightarrow 0$ , the model is a “large world”—the typical path length tends to  $\ell = L/4k$ , as discussed above. Small-world behavior, by contrast, is typically characterized by logarithmic scaling  $\ell \sim \log L$  (see section 3.1), which we see for large  $p$ , where the model becomes like a random graph. In between these two limits there is presumably some sort of crossover from large- to small-world behavior. Barthélemy and Amaral [42] conjectured that  $\ell$  satisfies a scaling relation of the form

$$(6.5) \quad \ell = \xi g(L/\xi),$$

where  $\xi$  is a correlation length that depends on  $p$ , and  $g(x)$  is an unknown but universal scaling function that depends only on system dimension and lattice geometry but not on  $L$ ,  $\xi$ , or  $p$ . The variation of  $\xi$  defines the crossover from large- to small-world behavior; the known behavior of  $\ell$  for small and large  $L$  can be reproduced by having  $\xi$  diverge as  $p \rightarrow 0$  and

$$(6.6) \quad g(x) \sim \begin{cases} x & \text{for } x \gg 1, \\ \log x & \text{for } x \ll 1. \end{cases}$$

Barthélemy and Amaral conjectured that  $\xi$  diverges as  $\xi \sim p^{-\tau}$  for small  $p$ , where  $\tau$  is a constant exponent. These conjectures have all turned out to be correct. Barthélemy and Amaral also conjectured on the basis of numerical results that  $\tau = \frac{2}{3}$ , which turned out not to be correct [39, 41, 323].

Equation (6.5) has been shown to be correct by a renormalization group treatment of the model [323]. From this treatment one can derive a scaling form for  $\ell$  of

$$(6.7) \quad \ell = \frac{L}{k} f(Lkp),$$

which is equivalent to (6.5), except for a factor of  $k$ , if  $\xi = 1/kp$  and  $g(x) = xf(x)$ . Thus we immediately conclude that the exponent  $\tau$  defined by Barthélemy and Amaral is 1, as was also argued by Barrat [39] using a mixture of scaling ideas and numerical simulation.

The scaling form (6.7) shows that we can go from the large-world regime to the small-world one either by increasing  $p$  or by increasing the system size  $L$ . Indeed, the crucial scaling variable  $Lkp$  that appears as the argument of the scaling function is simply equal to the mean number of shortcuts in the model, and hence  $\ell$  as a fraction of system size depends only on how many shortcuts there are for given  $k$ .

Making any further progress has proved difficult. We would like to be able to calculate the scaling function  $f(x)$ , but this turns out not to be easy. The calculation is possible, though complicated, for a variant model in which there are no short cuts but random sites are connected to a single central “hub” vertex [115]. But for the normal small-world model no exact solution is known, although some additional exact scaling forms have been found [19, 252]. Accurate numerical measurements have been carried out for system sizes up to about  $L = 10^7$  [39, 42, 109, 305, 323, 324], and quite good results can be derived using series expansions [324]. A mean-field treatment of the model has been given by Newman, Moore, and Watts [321], which shows that

$f(x)$  is approximately

$$(6.8) \quad f(x) = \frac{1}{2\sqrt{x^2 + 2x}} \tanh^{-1} \sqrt{\frac{x}{x+2}},$$

and Barbour and Reinert [38] have further shown that this result is the leading order term in an expansion for  $\ell$  that can be used to derive more accurate results for  $f(x)$ .

The primary use of the small-world model has been as a substrate for the investigation of various processes taking place on graphs, such as percolation [293, 324, 325, 359], coloring [387, 405], coupled oscillators [37, 200, 415], iterated games [1, 135, 230, 415], diffusion processes [150, 173, 215, 257, 258, 288, 328], epidemic processes [28, 234, 254, 292, 427, 428], and spin models [40, 190, 201, 255, 336, 429]. Some of this work is discussed further in section 8.

A few of variations of the small-world model have been proposed. Several authors have studied the model in dimensions higher than one [109, 305, 323, 324, 325]—the results are qualitatively similar to the one-dimensional case and follow the expected scaling laws. Various authors have also studied models in which shortcuts preferentially join vertices that are close together on the underlying lattice [214, 237, 238, 306, 364]. Of particular note is the work by Kleinberg [237, 238], which is discussed in section 8.3.3. Rozenfeld et al. [358] and independently Warren, Sander, and Sokolov [407] have studied models in which there are *only* shortcuts and no underlying lattice, but the signature of the lattice still remains, guiding shortcuts to fall with higher probability between more closely spaced vertices (see section 8.1).

**7. Models of Network Growth.** All of the models discussed so far take observed properties of real-world networks, such as degree sequences or transitivity, and attempt to create networks that incorporate those properties. The models do not, however, help us to understand how networks come to have those properties in the first place. In this section we examine a class of models whose primary goal is to explain network properties. In these models, the networks typically grow by the gradual addition of vertices and edges in some manner intended to reflect growth processes that might be taking place on the real networks, and it is these growth processes that lead to the characteristic structural features of the network.<sup>28</sup> For example, a number of authors [30, 102, 197, 216, 219, 241, 396, 397, 410, 411] have studied models of network transitivity that make use of “triadic closure” processes. In these models, edges are added to the network preferentially between pairs of vertices that have another third vertex as a common neighbor. In other words, edges are added so as to complete triangles, thereby increasing the denominator in (3.3) and so increasing the amount of transitivity in the network. (There is some empirical evidence from collaboration networks in support of this mechanism [309].)

But the best studied class of network growth models by far, and the class on which we concentrate primarily in this section, is the class of models aimed at explaining the origin of the highly skewed degree distributions discussed in section 3.3. Indeed these models are some of the best studied in the whole of the networks literature, having been the subject of an extraordinary number of papers in the last few years. In this section we describe first the archetypal model of Price [343], which was based in turn on previous work by Simon [369]. Then we describe the highly influential model of

<sup>28</sup>An alternative and intriguing idea, which has so far not been investigated in much depth, is that features such as power-law degree distributions may arise through network optimization. See, for instance, [29, 156, 166, 394, 416, 417].

Barabási and Albert [32], which has been the driving force behind much of the recent work in this area. We also describe a number of variations and generalizations of these models due to a variety of authors.

**7.1. Price's Model.** As discussed in section 3.3, the physicist-turned-historian-of-science Derek de Solla Price described in 1965 probably the first example of what would now be called a scale-free network; he studied the network of citations between scientific papers and found that both in- and out-degrees (number of times a paper has been cited and number of other papers a paper cites) have power-law distributions [342]. Apparently intrigued by the appearance of these power laws, Price published another paper some years later [343] in which he offered what is now the accepted explanation for power-law degree distributions. Like many after him, his work built on ideas developed in the 1950s by Herbert Simon [69, 369], who showed that power laws arise when “the rich get richer,” when the amount you get goes up with the amount you already have. In sociology this is referred to as the *Matthew effect* [281], after the biblical edict, “For to every one that hath shall be given...” (Matthew 25:29).<sup>29</sup> Price called it *cumulative advantage*. Today it is usually known under the name *preferential attachment*, coined by Barabási and Albert [32].

The important contribution of Price's work was to take the ideas of Simon and apply them to the growth of a network. Simon was thinking of wealth distributions in his early work, and although he later gave other applications of his ideas, none of them were to networked systems. Price appears to have been the first to discuss cumulative advantage specifically in the context of networks, and in particular in the context of the network of citations between papers and its in-degree distribution. His idea was that the rate at which a paper gets new citations should be proportional to the number that it already has. This is easy to justify in a qualitative way. The probability that one comes across a particular paper whilst reading the literature will presumably increase with the number of other papers that cite it, and hence the probability that you cite it yourself in a paper that you write will increase similarly. The same argument can be applied to other networks also, such as the Web. It is not clear that the dependence of citation probability on previous citations need be strictly linear, but certainly this is the simplest assumption one could make, and it is the one that Price, following Simon, adopts. We now describe in detail Price's model and his exact solution of it, which uses what we would now call a *master-equation* or *rate-equation* method.

Consider a directed graph of  $n$  vertices, such as a citation network. Let  $p_k$  be the fraction of vertices in the network with in-degree  $k$ , so that  $\sum_k p_k = 1$ . New vertices are continually added to the network, though not necessarily at a constant rate. Each added vertex has a certain out-degree—the number of papers that it cites—and this out-degree is fixed permanently at the creation of the vertex. The out-degree may vary from one vertex to another, but the mean out-degree, which is denoted  $m$ , is a constant over time.<sup>30</sup> (Certain conditions on the distribution of  $m$  about the mean

<sup>29</sup>In fact, this is really only half of the Matthew effect, since the same verse continues, “...but from him that hath not, that also which he seemeth to have shall be taken away.” In the processes studied by Simon and Price nothing is taken away from anyone. The full Matthew effect, with both the giving and the taking away, corresponds more closely to the Polya urn process than to Price's cumulative advantage. Price points out this distinction in his paper [343].

<sup>30</sup>Elsewhere in this review we have used the letter  $z$  to denote mean degree. While it would make sense in many ways to use the same notation here, we have opted instead to change notation and use  $m$  because this is the notation used in most of the recent papers on growing networks. The reader should bear in mind therefore that  $m$  is not, as previously, the total number of edges in the graph.

must hold; see, for instance, [134].) The value  $m$  is also the mean in-degree of the network:  $\sum_k k p_k = m$ . Since the out-degree can vary between vertices,  $m$  can take noninteger values, including values less than 1.

In the simplest form of cumulative advantage process the probability of attachment of one of our new edges to an old vertex—i.e., the probability that a newly appearing paper cites a previous paper—is simply proportional to the in-degree  $k$  of the old vertex. This, however, immediately gives us a problem, since each vertex starts with in-degree zero and hence would forever have zero probability of gaining new edges. To circumvent this problem, Price suggests that the probability of attachment to a vertex should be proportional to  $k + k_0$ , where  $k_0$  is a constant. Although he discusses the case of general  $k_0$ , all his mathematical developments are for  $k_0 = 1$ , which he justifies for the citation network by saying that one can consider the initial publication of a paper to be its first citation (of itself by itself). Thus the probability of a new citation is proportional to  $k + 1$ .

The probability that a new edge attaches to *any* of the vertices with degree  $k$  is thus

$$(7.1) \quad \frac{(k+1)p_k}{\sum_k (k+1)p_k} = \frac{(k+1)p_k}{m+1}.$$

The mean number of new citations per vertex added is simply  $m$ , and hence the mean number of new citations to vertices with current in-degree  $k$  is  $(k+1)p_k m/(m+1)$ . The number  $n p_k$  of vertices with in-degree  $k$  decreases by this amount, since the vertices that get new citations become vertices of degree  $k+1$ . However, the number of vertices of in-degree  $k$  increases because of influx from the vertices previously of degree  $k-1$  that have also just acquired a new citation, except for vertices of degree zero, which have an influx of exactly 1. If we denote by  $p_{k,n}$  the value of  $p_k$  when the graph has  $n$  vertices, then the net change in  $n p_k$  per vertex added is

$$(7.2) \quad (n+1)p_{k,n+1} - n p_{k,n} = \begin{cases} [k p_{k-1,n} - (k+1)p_{k,n}]m/(m+1) & \text{for } k \geq 1, \\ 1 - p_{0,n}m/(m+1) & \text{for } k = 0. \end{cases}$$

Looking for stationary solutions  $p_{k,n+1} = p_{k,n} = p_k$ , we then find

$$(7.3) \quad p_k = \begin{cases} [k p_{k-1} - (k+1)p_k]m/(m+1) & \text{for } k \geq 1, \\ 1 - p_0 m/(m+1) & \text{for } k = 0. \end{cases}$$

Rearranging, we find  $p_0 = (m+1)/(2m+1)$  and  $p_k = p_{k-1}k/(k+2+1/m)$  or

$$(7.4) \quad p_k = \frac{k(k-1)\dots 1}{(k+2+1/m)\dots(3+1/m)} p_0 = (1+1/m)B(k+1, 2+1/m),$$

where  $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$  is Legendre's beta-function, which goes asymptotically as  $a^{-b}$  for large  $a$  and fixed  $b$ , and hence

$$(7.5) \quad p_k \sim k^{-(2+1/m)}.$$

In other words, in the limit of large  $n$ , the degree distribution has a power-law tail with exponent  $\alpha = 2 + 1/m$ . This will typically give exponents in the interval between 2 and 3, which is in agreement with the values seen in real-world networks—see Table 3.1. (Bear in mind that the mean degree  $m$  need not take an integer value and can be less than 1.) Price gives a comparison between his model and citation network



data from the Science Citation Index, making a plausible case that the parameter  $m$  has about the right value to give the observed power-law citation distribution.

Note that Price's assumption that the offset parameter  $k_0 = 1$  can be justified a posteriori because the value of the exponent does not depend on  $k_0$ . (This contrasts with the behavior of the model of Barabási and Albert [32], which is discussed in section 7.3.) The argument above is easily generalized to the case  $k_0 \neq 1$ , and we find that

$$(7.6) \quad p_k = \frac{m+1}{m(k_0+1)+1} \frac{B(k+k_0, 2+1/m)}{B(k_0, 2+1/m)},$$

and hence  $\alpha = 2 + 1/m$  again for large  $k$  and fixed  $k_0$ . See section 7.3 and [123] and [244] for further discussion of the effects of offset parameters. Thorough reviews of master-equation methods for grown graph models have been given by Dorogovtsev and Mendes [120] and Krapivsky and Redner [247].

The analytic solution above was the extent of the progress Price was able to make in understanding his model network. Unlike present-day authors, for instance, he did not have computational resources available to simulate the model, and so could give no numerical results. In recent years, a great deal more progress has been made in understanding cumulative advantage processes and the growth of networks. Most of this work, however, has been carried out using a slightly different model—the model of Barabási and Albert, which we now describe.

**7.2. The Model of Barabási and Albert.** The mechanism of cumulative advantage proposed by Price [343] is now widely accepted as the probable explanation for the power-law degree distribution observed not only in citation networks but in a wide variety of other networks also, including the World Wide Web, collaboration networks, and possibly the Internet and other technological networks also. The work of Price himself, however, is largely unknown in the scientific community, and cumulative advantage did not achieve currency as a model of network growth until its rediscovery some decades later by Barabási and Albert [32], who gave it the new name of *preferential attachment*. In a highly influential paper published—like Price's first paper on citation networks—in the journal *Science*, they proposed a network growth model of the Web that is very similar to Price's, but with one important difference.

The model of Barabási and Albert [32, 33] is the same as Price's in having vertices that are added to the network with degree  $m$ , which is never changed thereafter, the other end of each edge being attached to ("citing") another vertex with probability proportional to the degree of that vertex. The difference between the two models is that in the model of Barabási and Albert edges are undirected, so there is no distinction between in- and out-degree. This has pros and cons. On the one hand, both citation networks and the Web are in reality directed graphs, so any undirected graph model is missing a crucial feature of these networks. On the other hand, by ignoring the directed nature of the network, the model of Barabási and Albert gets around Price's problem of how a paper gets its first citation or a Web site gets its first link. Each vertex in the graph appears with initial degree  $m$ , and hence automatically has a nonzero probability of receiving new links. (Note that for the model to be solvable using the master-equation approach as demonstrated below, the number of edges added with each vertex must be exactly  $m$ —it cannot vary around the mean value as in the model of Price. Hence it must also be an integer and must always have a value  $m \geq 1$ .)

Another way of looking at the model of Barabási and Albert is to say the network *is* directed, with edges going from the vertex just added to the vertex that it is citing or linking to, but that the probability of attachment of a new edge is proportional to the sum of the in- and out-degrees of the vertex. This, however, is perhaps a less satisfactory viewpoint, since it is difficult to conjure up a mechanism, either for citation networks or the Web, which would give rise to such an attachment process. Overall, perhaps the best way to look at the model of Barabási and Albert is as a model that sacrifices some of the realism of Price's model in favor of simplicity. As we will see, the main result of this sacrifice is that the model produces only a single value  $\alpha = 3$  for the exponent governing the degree distribution, although this has been remedied in later generalizations of the model, which we discuss in section 7.3.

The model of Barabási and Albert can be solved exactly in the limit of large graph size<sup>31</sup> using the master-equation method, and such a solution has been given by Krapivsky, Redner, and Leyvraz [248] and independently by Dorogovtsev, Mendes, and Samukhin [123]. (Barabási and Albert themselves gave an approximate solution based on the assumption that all vertices of the same age have the same degree [32, 33]. The method of Krapivsky et al. and Dorogovtsev et al. does not make this assumption.)

The probability that a new edge attaches to a vertex of degree  $k$ —the equivalent of (7.1)—is

$$(7.7) \quad \frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{2m}.$$

The sum in the denominator is equal to the mean degree of the network, which is  $2m$ , since there are  $m$  edges for each vertex added, and each edge, being now undirected, contributes two ends to the degrees of network vertices. Now the mean number of vertices of degree  $k$  that gain an edge when a single new vertex with  $m$  edges is added is  $m \times kp_k / 2m = \frac{1}{2}kp_k$ , independent of  $m$ . The number  $np_k$  of vertices with degree  $k$  thus decreases by this same amount, since the vertices that get new edges become vertices of degree  $k + 1$ . The number of vertices of degree  $k$  also increases because of influx from vertices previously of degree  $k - 1$  that have also just acquired a new edge, except for vertices of degree  $m$ , which have an influx of exactly 1. If we denote by  $p_{k,n}$  the value of  $p_k$  when the graph has  $n$  vertices, then the net change in  $np_k$  per vertex added is

$$(7.8) \quad (n+1)p_{k,n+1} - np_{k,n} = \begin{cases} \frac{1}{2}(k-1)p_{k-1,n} - \frac{1}{2}kp_{k,n} & \text{for } k > m, \\ 1 - \frac{1}{2}mp_{m,n} & \text{for } k = m, \end{cases}$$

and there are no vertices with  $k < m$ .

Looking for stationary solutions  $p_{k,n+1} = p_{k,n} = p_k$  as before, the equations equivalent to (7.3) for the model are

$$(7.9) \quad p_k = \begin{cases} \frac{1}{2}(k-1)p_{k-1} - \frac{1}{2}kp_k & \text{for } k > m, \\ 1 - \frac{1}{2}mp_m & \text{for } k = m. \end{cases}$$

Rearranging for  $p_k$  once again, we find  $p_m = 2/(m+2)$  and  $p_k = p_{k-1}(k-1)/(k+2)$ , or [123, 248]

$$(7.10) \quad p_k = \frac{(k-1)(k-2)\dots m}{(k+2)(k+1)\dots(m+3)} p_m = \frac{2m(m+1)}{(k+2)(k+1)k}.$$

<sup>31</sup>The behavior of the model at finite system sizes has been investigated by Krapivsky and Redner [245].

In the limit of large  $k$  this gives a power law degree distribution  $p_k \sim k^{-3}$ , with only the single fixed exponent  $\alpha = 3$ . A more rigorous derivation of this result has been given by Bollobás et al. [65].

In addition to the basic solution of the model for its degree distribution, many other results are now known about the model of Barabási and Albert. Krapivsky and Redner [244] have conducted a thorough analytic study of the model, showing among other things that the model has two important types of correlations. First, there is a correlation between the age of vertices and their degrees, with older vertices having higher mean degree. For the case  $m = 1$ , for instance, they find that the probability distribution of the degree of a vertex  $i$  with age  $a$ , measured as the number of vertices added after vertex  $i$ , is

$$(7.11) \quad p_k(a) = \sqrt{1 - \frac{a}{n}} \left( 1 - \sqrt{1 - \frac{a}{n}} \right)^k.$$

Thus for specified age  $a$  the distribution is exponential, with a characteristic degree scale that diverges as  $(1 - a/n)^{-1/2}$  as  $a \rightarrow n$ ; the earliest vertices added have substantially higher expected degree than those added later, and the overall power-law degree distribution of the whole graph is a result primarily of the influence of these earliest vertices.

This correlation between degree and age has been used by Adamic and Huberman [4] to argue against the model as a model of the World Wide Web—they show, using actual Web data, that there is no such correlation in the real Web. This does not mean that preferential attachment is not the explanation for power-law degree distributions in the Web, only that the dynamics of the Web must be more complicated than this simple model to account also for the observed age distribution [35]. An extension of the model that may explain why age and degree are not correlated has been given by Bianconi and Barabási [52, 53] and is discussed in section 7.3.

Second, Krapivsky and Redner [244] show that there are correlations between the degrees of adjacent vertices in the model, of the type discussed in section 3.6. Looking again at the special case  $m = 1$ , they show that the quantity  $e_{jk}$  defined in section 4.2.5, which is the number of edges that connect vertex pairs with (excess) degrees  $j$  and  $k$ , is

$$(7.12) \quad e_{jk} = \frac{4j}{(k+1)(k+2)(j+k+2)(j+k+3)(j+k+4)} + \frac{12j}{(k+1)(j+k+1)(j+k+2)(j+k+3)(j+k+4)}.$$

Note that this quantity is asymmetric. This is because Krapivsky and Redner regard the network as being directed, with edges leading from the vertex just added to the preexisting vertex to which they attach. In the expression above, however,  $j$  and  $k$  are total degrees of vertices, not in- and out-degree.

Although (7.12) shows that the vertices of the model have nontrivial correlations, the correlation coefficient of the degrees of adjacent vertices in the network is asymptotically zero as  $n \rightarrow \infty$  [313]. This is because the correlation coefficient measures correlations relative to a linear model, and no such correlations are present in this case.

One of the main advantages that we have today over early workers such as Price is the widespread availability of powerful computer resources. Quite a number of numerical studies have been performed of the model of Barabási and Albert, which

would have been entirely impossible thirty years earlier. It is worth mentioning here how simulations of these types of models are conducted. We consider the Barabási–Albert model. The exact same ideas can be applied to Price’s model also.

A naive simulation of the preferential attachment process is quite inefficient. In order to attach to a vertex in proportion to its degree we normally need to examine the degrees of all vertices in turn, a process that takes  $O(n)$  time for each step of the algorithm. Thus the generation of a graph of size  $n$  would take  $O(n^2)$  steps overall. A much better procedure, which works in  $O(1)$  time per step and  $O(n)$  time overall, is the following. We maintain a list, in an integer array for instance, that includes  $k_i$  entries of value  $i$  for each vertex  $i$ . Thus, for example, a network of four vertices labeled 1, 2, 3, and 4 with degrees 2, 1, 1, and 3, respectively, could be represented by the array (1, 1, 2, 3, 4, 4, 4). Then in order to choose a target vertex for a new edge with the correct preferential attachment, one simply chooses a number at random from this list. Of course, the list must be updated as new vertices and edges are added, but this is simple. Notice that there is no requirement that the items in the list be in any particular order. If we add a new vertex 5 to our network above, for example, with degree 1 and one edge that connects it to vertex 2, the list can be updated by adding new items to the end, so that it reads (1, 1, 1, 2, 3, 4, 4, 4, 5, 2). And so forth. Models such as Price’s, in which there is an offset  $k_0$  in the probability of selecting a vertex (so that the total probability goes as  $k + k_0$ ), can be treated with the same method—the offset merely means that with some probability one chooses a vertex with preferential attachment, and otherwise one chooses it uniformly from the set of all vertices.

An alternative method for simulating the model of Barabási and Albert has been described by Krapivsky and Redner [244]. Their method uses the network structure itself in place of the list of vertices above and works as follows. The model is regarded as a directed network in which there are exactly  $m$  edges running out of each vertex, pointing to others. We first pick a vertex at random from the graph and then with some probability either we keep that vertex or we “redirect” to one of its neighbors, meaning that we pick at random one of the vertices it points to. Since each vertex has exactly  $m$  outgoing edges, the latter operation is equivalent to choosing an edge at random from the graph and following it, and hence alights on a target vertex with probability proportional to the in-degree  $j$  of that target (because there are  $j$  ways to arrive at a vertex of in-degree  $j$ —see section 4.2.1). Thus the total probability of selecting any given vertex is proportional to  $j + c$ , where  $c$  is some constant. However, since the out-degree of all vertices is simply  $m$ , the total degree is  $k = j + m$  and the selection probability is therefore also proportional to  $k + c - m$ . By choosing the probability of redirection appropriately, we can arrange for the constant  $c$  to be equal to  $m$ , and hence for the probability of selecting a vertex to be simply proportional to  $k$ . Since it does not require an extra array for the vertex list, this method of simulation is more memory efficient than the previous method, although it is slightly more complicated to implement.

In their original paper on their model, Barabási and Albert [32] gave simulations showing the power-law distribution of degrees. A number of authors have subsequently published more extensive simulation results. Of particular note is the work by Dorogovtsev and Mendes [116, 114] and by Krapivsky and Redner [245].

A crucial element of both the models of Price and of Barabási and Albert is the assumption of linear preferential attachment. It is worth asking whether there is any empirical evidence in support of this assumption. (We discuss in the next section some work on models that relax the linearity assumption.) Two studies indicate that

it may be a reasonable approximation to the truth. Jeong, Néda, and Barabási [212] looked at the time evolution of citation networks, the Internet, and actor and scientist collaboration networks and measured the number of new edges a vertex acquires in a single year as a function of the number of previously existing edges. They found that the one quantity was roughly proportional to the other, and hence concluded that linear preferential attachment was at work in these networks. Newman [309] performed a similar study for scientific collaboration networks, but with finer time resolution, measured by the publication of individual papers, and came to similar conclusions.

**7.3. Generalizations of the Model of Barabási and Albert.** The model of Barabási and Albert [32] has attracted an exceptional amount of attention in the literature. In addition to analytic and numerical studies of the model itself, many authors have suggested extensions or modifications of the model that alter its behavior or make it a more realistic representation of processes taking place in real-world networks. We discuss a few of these here. A more extensive review of developments in this area has been given by Albert and Barabási [13] (see particularly Table III in that paper).

Dorogovtsev, Mendes, and Samukhin [123] and Krapivsky and Redner [244] have examined the model in which the probability of attachment to a vertex of degree  $k$  is proportional to  $k + k_0$ , where the offset  $k_0$  is a constant. Note that  $k_0$  is allowed to be negative—it can fall anywhere in the range  $-m < k_0 < \infty$  and the probability of attachment will be positive. The equations for the stationary state of the degree distribution of this model, analogous to (7.9), are

$$(7.13) \quad p_k = \begin{cases} [(k-1)p_{k-1} - kp_k]m/(2m + k_0) & \text{for } k > m, \\ 1 - p_m m^2/(2m + k_0) & \text{for } k = m, \end{cases}$$

which gives  $p_m = (2m + k_0)/(m^2 + 2m + k_0)$ , and

$$(7.14) \quad p_k = \frac{(k-1) \dots m}{(k+2+k_0/m) \dots (m+3+k_0/m)} p_m = \frac{B(k, 3+k_0/m)}{B(m, 2+k_0/m)},$$

where  $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$  is again the Legendre beta-function. This gives a power law for large  $k$  once more, with exponent  $\alpha = 3 + k_0/m$ . It is proposed that negative values of  $k_0$  could be the explanation for the values  $\alpha < 3$  seen in real-world networks.<sup>32</sup> A longer discussion of the effects of offset parameters is given in [244].

Krapivsky et al. [244, 248] also consider another important generalization of the model, to the case where the probability of attachment to a vertex is not linear in the degree  $k$  of the vertex but goes instead as some general power of degree  $k^\gamma$ . Again this model is solvable using methods similar to those above, and the authors find three general classes of behavior. For  $\gamma = 1$  exactly, we recover the normal linear preferential attachment and power-law degree sequences. For  $\gamma < 1$ , the degree distribution is a power law multiplied by a stretched exponential, whose exponent is a complicated function of  $\gamma$ . (In fact, in most cases there is no known analytic solution for the equations governing the exponent; they must be solved numerically.) For  $\gamma > 1$  there is a “condensation” phenomenon, in which a single vertex gets a finite fraction of all the connections in the network, and for  $\gamma > 2$  there is a nonzero probability that this

<sup>32</sup>Price’s result  $\alpha = 2 + 1/m$  [343] corresponds to  $k_0 = -(m-1)$  so that the “attractiveness” of a new vertex is 1. The model of Barabási and Albert corresponds to  $k_0 = 0$ , so that  $\alpha = 3$ .

“gel node” will be connected to every other vertex on the graph. The remainder of the vertices have an exponentially decaying degree distribution.

Another variation on the basic growing network theme is to make the mean degree change over time. There is evidence to suggest that in the World Wide Web the average degree of a vertex is increasing with time, i.e., the parameter  $m$  appearing in the models is increasing. Dorogovtsev and Mendes [118, 121] have studied a variation of the Barabási–Albert model that incorporates this process. They assume that the number  $m$  of new edges added per new vertex increases with network size  $n$  as  $n^a$  for some constant  $a$ , and that the probability of attaching to a given vertex goes as  $k + Bn^a$  for constant  $B$ . They show that the resulting degree distribution follows a power law with exponent  $\alpha = 2 + B(1 + a)/(1 - Ba)$ . (Note that when  $a = 0$ , this model reduces to the model studied previously by Dorogovtsev, Mendes, and Samukhin [123], but the expression for  $\alpha$  given here is not valid in this limit.) Thus this process offers another possible mechanism by which the exponent of the degree distribution can be tuned to match that observed in real-world networks.

In Price’s model of citation networks, no new out-going edges are added to a vertex after its first appearance, and edges once added to the graph remain where they are forever. This makes sense for citation networks. But the model of Barabási and Albert is intended to be a model of the World Wide Web, in which new links are often added to preexisting Web sites, and old links are frequently moved or removed. A number of authors have proposed models that incorporate processes like these. In particular, Dorogovtsev and Mendes [116] have proposed a model that adds to the standard Barabási–Albert model an extra mechanism whereby edges appear and disappear between preexisting vertices with stochastically constant but possibly different rates. They find that over a wide range of values of the rates the power-law degree distribution is maintained, although again the exponent varies from the value  $-3$  seen in the original model. Krapivsky and Redner [246] have also proposed a model that allows edges to be added after vertices are created, which we discuss in the next section. Albert and Barabási [12] and Tadić [390, 391] have studied models in which edges can move around the network after they are added. These models can show both power-law and exponential degree distributions depending on the model parameters.

As discussed in section 7.2, Adamic and Huberman [4] have shown that the real World Wide Web does not have the correlations between age and degree of vertices that are found in the model of Barabási and Albert. Adamic and Huberman suggest that this is because the degree of vertices is also a function of their intrinsic worth; some Web sites are useful to more people than others and so gain links at a higher rate. Bianconi and Barabási [52, 53] have proposed an extension of the Barabási–Albert model that mimics this process. In their model each newly appearing vertex  $i$  is given a “fitness”  $\eta_i$  that represents its attractiveness and hence its propensity to accrue new links. Fitnesses are chosen from some distribution  $\rho(\eta)$ , and links attach to vertices with probability proportional now not just to the degree  $k_i$  of vertex  $i$  but to the product  $\eta_i k_i$ .

Depending on the form of the distribution  $\rho(\eta)$  this model shows two regimes of behavior [52, 246]. If the distribution has finite support, then the network shows a power-law degree distribution, as in the original Barabási–Albert model. However, if the distribution has infinite support, then the one vertex with the highest fitness accrues a finite fraction of all the edges in the network—a sort of “winner takes all” phenomenon, which Bianconi and Barabási liken to monopoly dominance of a market.

A number of variations on the fitness theme have been studied by Ergün and Rodgers [145], who looked at a directed version of the Bianconi–Barabási model and at models where instead of multiplying the attachment probability, the fitness  $\eta_i$  contributes additively to the probability of attaching a new edge to vertex  $i$ . Treating the models analytically, they found in each case that for suitable parameter values the power-law degree distribution is preserved, although again the exponent may be affected by the distribution of fitnesses, and in some cases there are also logarithmic corrections to the degree distribution. A model with vertex fitness but no preferential attachment has been studied by Caldarelli et al. [78] and also gives power-law degree distributions under some circumstances.

**7.4. Other Growth Models.** The model of Barabási and Albert [32] is elegant and simple but lacks a number of features that are present in the real World Wide Web:

- The model is a model of an undirected network, whereas the real Web is directed.
- As mentioned previously one can regard the model as a model of a directed network, but in that case attachment is in proportion to the sum of in- and out-degrees of a vertex, which is unrealistic—presumably attachment should be in proportion to in-degree only, as in the model of Price.
- If we regard the model as producing a directed network, then it generates acyclic graphs (see section 1.1), which are a poor representation of the Web.
- All vertices in the model belong to a single connected component (a weakly connected component if the graph is regarded as directed—the graph has no strongly connected components because it is acyclic). In the real Web there are many separate components (and strongly connected components).
- The out-degree distribution of the Web follows a power law, whereas out-degree is a constant in the model.<sup>33</sup>

Many of these criticisms are also true of Price’s model, but Price’s model is intended to be a model of a citation network, and citation networks really are directed, acyclic, and to a good approximation all vertices belong to a single component, unless they cite and are cited by no one else at all. Thus Price’s model is, within its own limited sphere, a reasonable one. For the World Wide Web a number of authors have suggested new growth models that address one or more of the concerns above. Here we describe a number of these models, starting with some very simple ones and working up to the more complex.

Consider first the issue of the component structure of the network. In the models of Price and of Barabási and Albert each vertex joins to at least one other when it first appears. It follows trivially then that, so long as no edges are ever removed, all vertices belong to a single (weakly connected) component. This is not true in the real Web. How can we get around it? To address this question Callaway et al. [80] proposed the following extremely simple model of a growing network. Vertices are added to the network one by one as before, and a mean number  $m$  of undirected

<sup>33</sup>What’s more, although it is rarely pointed out, it is clearly the case that a different mechanism must be responsible for the out-degree distribution from the one responsible for the in-degree distribution. We can justify preferential attachment for in-degree by saying that Web sites are easier to find if they have more links to them, and hence they get more new links because people find them. No such argument applies for out-degree. It is usually assumed that out-degree is subject to preferential attachment nonetheless. One can certainly argue that sites with many out-going links are more likely to add new ones in the future than sites with few, but it’s far from clear that this must be the case.

edges are added with each vertex. As with Price's model, the value of  $m$  is only an average—the actual number of edges added per step can vary—and so  $m$  is not restricted to integer values, and indeed we will see that the interesting behavior of the model takes place at values  $m < 1$ .

The important difference between this model and the previous models is that edges are not, in general, attached to the vertex that has just been added. Instead, both ends of each edge are attached to vertices chosen uniformly at random from the whole graph, without preferential attachment. Vertices therefore normally have degree zero when they are first added to the graph. Because of the lack of preferential attachment this model does not show power-law degree distributions—in fact the degree distribution can be shown to be exponential—but it does have an interesting component structure. A related model has been studied, albeit to somewhat different purpose, by Aldous and Pittel [17]. Their model is equivalent to the model of Callaway et al. in the case  $m = 1$ . Also Bauer and collaborators [44, 100] have investigated a directed-graph version of the model.

Initially, one might imagine that the model of Callaway et al. generated an ordinary Poisson random graph of the Erdős–Rényi type. Further reflection reveals, however, that this is not the case; older vertices in the network will tend to be connected to one another, so the network has a cliquish core of old-timers surrounded by a sea of younger vertices. Nonetheless, like the Poisson random graph, the model does have many separate components, with a phase transition at a finite value of  $m$  at which a giant component appears that occupies a fixed fraction of the volume of the network as  $n \rightarrow \infty$ . To demonstrate this, Callaway et al. used a master-equation approach similar to that used for degree distributions in the preceding sections. One defines  $p_s$  to be the probability that a randomly chosen vertex belongs to a component of  $s$  vertices and writes difference equations that give the change in  $p_s$  when a single vertex and  $m$  edges are added to the graph. Looking for stationary solutions, one then finds in the limit of large graph size that

$$(7.15) \quad p_s = \begin{cases} ms \sum_{j=1}^{s-1} p_j p_{s-j} - 2mp_s & \text{for } s > 1, \\ 1 - 2mp_1 & \text{for } s = 1. \end{cases}$$

Being nonlinear in  $p_s$ , these equations are harder to solve than those for the degree distributions in previous sections, and indeed no exact solution has been found. Nonetheless, we can see that a giant component must form by defining a generating function for the component size distribution similar to that of (4.8):  $H(x) = \sum_{s=0}^{\infty} p_s x^s$ . Then (7.15) implies that

$$(7.16) \quad \frac{dH}{dx} = \frac{1}{2m} \left[ \frac{1 - H(x)/x}{1 - H(x)} \right].$$

If there is no giant component, then  $H(1) = 1$  and the average component size is  $\langle s \rangle = H'(1)$ . Taking the limit  $x \rightarrow 1$  in (7.16), we find that  $\langle s \rangle$  is a solution of the quadratic equation  $2m\langle s \rangle^2 - \langle s \rangle + 1 = 0$ , or

$$(7.17) \quad \langle s \rangle = \frac{1 - \sqrt{1 - 8m}}{4m}.$$

(The other solution to the quadratic gives a nonphysical value.) This solution exists only up to  $m = \frac{1}{8}$ , however, and hence above this point there must be a giant com-



ponent. This doesn't tell us where in the interval  $0 \leq m \leq \frac{1}{8}$  the giant component appears, but a proof that the transition in fact falls precisely at  $m = \frac{1}{8}$  was later given by Durrett [134].

The model of Callaway et al. has been generalized to include preferential attachment by Dorogovtsev, Mendes, and Samukhin [124]. In their version of the model both ends of each edge are attached in proportion to the degrees of vertices plus a constant offset to ensure that vertices of degree zero have a chance of receiving an edge. Again they find many components and a phase transition at nonzero  $m$ , and in addition the power-law degree distribution is now restored.

Taking the process a step further, Krapivsky and Redner [246] studied a full directed-graph model in which both vertices and directed edges are added at stochastically constant rates and the out-going end of each edge is attached to vertices in proportion to their out-degree and the in-going end in proportion to in-degree, plus appropriate constant offsets. This appears to be quite a reasonable model for the growth of the Web. It produces a directed graph, it allows edges to be added after the creation of a vertex, it allows for separate components in the graph, and, as Krapivsky and Redner showed, it gives power laws in both the in- and out-degree distributions, just as observed in the real Web. By varying the offset parameters for the in- and out-degree attachment mechanisms, one can even tune the exponents of the two distributions to agree with those observed in the wild. (Krapivsky and Redner's model is a development of an earlier model that they proposed [249] that had all the same features but gave rise to only a single weakly connected component because each added vertex came with one edge that attached it to the rest of the network from the outset. In their later paper, they abandoned this feature. A similar model has also been studied by Rodgers and Darby-Dowman [354].) A slight variation on the model of Krapivsky and Redner has been proposed independently by Aiello, Chung, and Lu [9], who give rigorous proofs of some of its properties.

**7.5. Vertex Copying Models.** There are some networks that appear to have power-law degree distributions but for which preferential attachment is clearly not an appropriate model. Good examples are biochemical interaction networks of various kinds [153, 211, 213, 375, 382, 404]. A number of studies have been performed, for instance, of the interaction networks of proteins (see section 2.4) in which the vertices are proteins and the edges represent reactions. These networks do change on very long time-scales because of biological evolution, but there is no reason to suppose that protein networks grow according to a simple cumulative advantage or preferential attachment process. Nonetheless, it appears that the degree distribution of these networks obeys a power law, at least roughly.

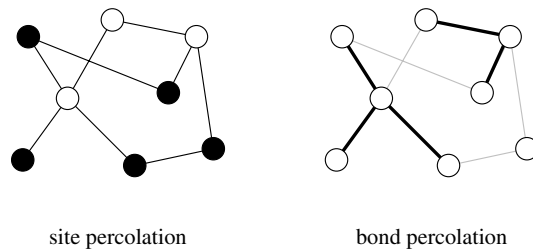
A possible explanation for this observation has been suggested by Kleinberg et al. [240, 253], who proposed that these networks grow, at least in part, by the copying of vertices. Kleinberg et al. were interested in the growth of the Web, for which their model is as follows. The graph grows by stochastically constant addition of vertices and addition of directed edges either randomly or by copying them from another vertex. Specifically, one chooses an existing vertex and a number  $m$  of edges to add to it, and one then decides the targets of those edges by choosing at random another vertex and copying targets from  $m$  of its edges, randomly chosen. If the chosen vertex has less than  $m$  outgoing edges, then its  $m$  edges are copied and one moves on to another vertex and copies its edges, and so forth until  $m$  edges in total have been copied. In its most general form, the model of Kleinberg et al. also incorporates mechanisms for the removal of edges and vertices, which we do not describe here.

It is straightforward to see that the copying mechanism will give rise to power-law distributions. The mean probability that an edge from a randomly chosen vertex will lead to a particular other vertex with in-degree  $k$  is proportional to  $k$  (see section 4.2.1), and hence the rate of increase of a vertex's degree is proportional to its current degree. As with the model of Price, this mechanism will never add new edges to vertices that currently have degree zero, so Kleinberg et al. also include a finite probability that the target of a newly added edge will be chosen at random, so that vertices with degree zero have a chance to gain edges. In their original paper, Kleinberg et al. present only numerical evidence that their model results in a power-law degree distribution, but in a later paper a subset of the same authors [253] proved that the degree distribution is a power law with exponent  $\alpha = (2 - a)/(1 - a)$ , where  $a$  is the ratio of the number of edges added whose targets are chosen at random to the number whose targets are copied from other vertices. For small values of  $a$ , between 0 and  $\frac{1}{2}$ , i.e., for models in which most target selection is made by copying, this produces exponents  $2 \leq \alpha \leq 3$ , which is the range observed in most real-world networks—see Table 3.1. Some further analytic results for copying models have been given by Chung et al. [90].

It is not clear whether the copying mechanism really is at work in the growth of the World Wide Web, but there has been considerable interest in its application as a model of the evolution of protein interaction networks of one sort or another. The argument here is that the genes that code for proteins can and do, in the course of their evolutionary development, duplicate. That is, upon reproduction of an organism, two copies of a gene are erroneously made where only one existed before. Since the proteins coded for by each copy are the same, their interactions are also the same, i.e., the new gene copies its edges in the interaction network from the old. Subsequently, the two genes may develop differences because of evolutionary drift or selection [403]. Models of protein networks that make use of copying mechanisms have been proposed by a number of authors [49, 232, 376, 398].

A variation on the idea of vertex copying appears in the autocatalytic network models of Jain and Krishna [208, 209], in which a network of interacting chemical species evolves by reproduction and mutation, giving rise ultimately to self-sustaining autocatalytic loops reminiscent of the “hypercycles” of Eigen and Schuster [140], which have been proposed as a possible explanation of the origin of life.

**8. Processes Taking Place on Networks.** As discussed in the introduction, the ultimate goal of the study of the structure of networks is to understand and explain the workings of systems built upon those networks. We would like, for instance, to understand how the topology of the World Wide Web affects Web surfing and search engines, how the structure of social networks affects the spread of information, how the structure of a food web affects population dynamics, and so forth. Thus, the next logical step after developing models of network structure, such as those described in the previous sections of this review, is to look at the behavior of models of physical (or biological or social) processes going on on those networks. Progress on this front has been slower than progress on understanding network structure, perhaps because without a thorough understanding of structure an understanding of the effects of that structure is hard to come by. However, there have been some important advances made, particularly in the study of network failure, epidemic processes on networks, and constraint satisfaction problems. In this section we review what has been learned so far.



**Fig. 8.1** *Site and bond percolation on a network. In site percolation, vertices (“sites” in the physics parlance) are either occupied (solid circles) or unoccupied (open circles) and studies focus on the shape and size of the contiguous clusters of occupied sites, of which there are three in this small example. In bond percolation, it is the edges (“bonds” in physics) that are occupied or not (black or gray lines) and the vertices that are connected together by occupied edges that form the clusters of interest.*

**8.1. Percolation Theory and Network Resilience.** One of the first examples to be studied thoroughly of a process taking place on a network has been percolation processes, mostly simple site and bond percolation—see Figure 8.1—although a number of variants have been studied also. A percolation process is one in which vertices or edges on a graph are randomly designated either “occupied” or “unoccupied” and one asks about various properties of the resulting patterns of vertices. One of the main motivations for the percolation model when it was first proposed in the 1950s was the modeling of the spread of disease [73, 186], and it is in this context also that it was first studied in the current wave of interest in real-world networks [324]. We consider epidemiological applications of percolation theory in section 8.2. Here, however, we depart from the order of historical developments to discuss first a simpler application to the question of network resilience.

As discussed in section 3.4, real-world networks are found often to be highly resilient to the random deletion of their vertices. Resilience can be measured in different ways, but perhaps the simplest indicator of resilience in a network is the variation (or lack of variation) in the fraction of vertices in the largest component of the network, which we equate with the giant component in our models (see section 4.1). If one is thinking of a communication network, for example, in which the existence of a connecting path between two vertices means that those two can communicate with one another, then the vertices in the giant component can communicate with an extensive fraction of the entire network, while those in the small components can communicate with only a few others at most. Following the numerical studies of Broder et al. [74] and Albert, Jeong, and Barabási [15] on subsets of the Web graph, it was quickly realized [81, 93] that the problem of resilience to random failure of vertices in a network is equivalent to a site percolation process on the network. Vertices are randomly occupied (working) or unoccupied (failed), and the number of vertices remaining that can successfully communicate is precisely the giant component of the corresponding percolation model.

A number of analytic results have been derived for percolation on networks with the structure of the configuration model of section 4.2.1, i.e., a random graph with a given degree sequence. Cohen et al. [93] made the following simple argument. Suppose we have a configuration model with degree distribution  $p_k$ . That is, a randomly chosen vertex has degree  $k$  with probability  $p_k$  in the limit of large number  $n$  of vertices. Now suppose that only a fraction  $q$  of the vertices are “occupied,” or functional, that

fraction chosen uniformly at random from the entire graph. For a vertex with degree  $k$ , the number  $k'$  of occupied vertices to which it is connected is distributed binomially so that the probability of having a particular value of  $k'$  is  $\binom{k}{k'} q^{k'} (1-q)^{k-k'}$ , and hence the total probability that a randomly chosen vertex is connected to  $k'$  other occupied vertices is

$$(8.1) \quad p_{k'} = \sum_{k=k'}^{\infty} p_k \binom{k}{k'} q^{k'} (1-q)^{k-k'}.$$

Since vertex failure is random and uncorrelated, the subset of all vertices that are occupied forms another configuration model with this degree distribution. Cohen et al. then applied the criterion of Molloy and Reed, (4.11), to determine whether this network has a giant component. (One could also apply (4.12) and (4.13) to determine the size of the giant and nongiant components, although this is not done in [93].)

One of the most interesting conclusions of the work of Cohen et al. is for the case of networks with power-law degree distributions  $p_k \sim k^{-\alpha}$  for some constant  $\alpha$ . When  $\alpha \leq 3$ , they find that the critical value  $q_c$  of  $q$  where the transition takes place at which a giant component forms is zero or negative, indicating that the network always has a giant component, or in the language of physics, the network always percolates. This echoes the numerical results of Albert, Jeong, and Barabási [15], who found that the connectivity of power-law networks was highly robust to the random removal of vertices. In general, the method of Cohen et al. indicates that  $q_c \leq 0$  for any degree distribution with a diverging second moment.

An alternative and more general approach to the percolation problem on the configuration model has been put forward by Callaway et al. [81], using a generalization of the generating function formalism discussed in section 4.2.1. In their method, the probability of occupation of a vertex can be any function of the degree  $k$  of that vertex. Thus the constant  $q$  of the approach of Cohen et al. is generalized to  $q_k$ , the probability that a vertex having degree  $k$  is occupied. One defines generating functions

$$(8.2) \quad F_0(x) = \sum_{k=0}^{\infty} p_k q_k x^k, \quad F_1(x) = \frac{\sum_k k p_k q_k x^{k-1}}{\sum_k k p_k},$$

and it can then be shown that the probability distribution of the size of the component of occupied vertices to which a randomly chosen vertex belongs is generated by  $H_0(x)$ , where

$$(8.3) \quad H_0(x) = 1 - F_0(1) + x F_0(H_1(x)), \quad H_1(x) = 1 - F_1(1) + x F_1(H_1(x)).$$

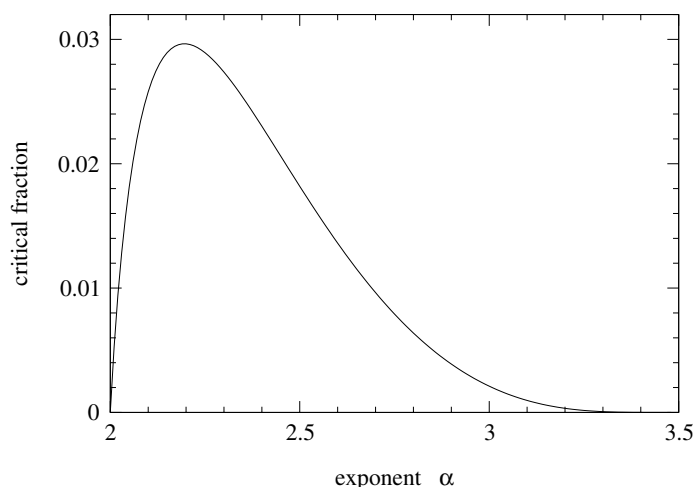
(Note that  $F_0$  is not a properly normalized generating function in the sense that  $F_0(1) \neq 1$ .) From this, one can derive an expression for the mean component size:

$$(8.4) \quad \langle s \rangle = F_0(1) + \frac{F'_0(1) F_1(1)}{1 - F'_1(1)},$$

which immediately tells us that the phase transition at which a giant component forms takes place at  $F'_1(1) = 1$ . The size of the giant component is given by

$$(8.5) \quad S = F_0(1) - F_0(u), \quad u = 1 - F_1(1) + F_1(u).$$

For instance, in the case studied by Cohen et al. [93] of uniform occupation probability  $q_k = q$ , this gives a critical occupation probability of  $q_c = 1/G'_1(1)$ , where



**Fig. 8.2** The fraction of vertices that must be removed from a network to destroy the giant component, if the network has the form of a configuration model with a power-law degree distribution of exponent  $\alpha$ , and vertices are removed in decreasing order of their degrees.

$G_1(x)$  is the generating function for the degree distribution itself, as defined in (4.6). Taking the example of a power-law degree distribution  $p_k = k^{-\alpha}/\zeta(\alpha)$ , (4.15), we find

$$(8.6) \quad q_c = \frac{\zeta(\alpha - 1)}{\zeta(\alpha - 2) - \zeta(\alpha - 1)}.$$

This is negative (and hence unphysical) for  $\alpha < 3$ , confirming the finding that the system always percolates in this regime. Note that  $q_c > 1$  for sufficiently large  $\alpha$ , which is also unphysical. One finds that the system *never* percolates for  $\alpha > \alpha_c$ , where  $\alpha_c$  is the solution of  $\zeta(\alpha - 2) = 2\zeta(\alpha - 1)$ , which gives  $\alpha_c = 3.4788\dots$ . This corresponds to the point at which the underlying network itself ceases to have a giant component, as shown by Aiello, Chung, and Lu [8] and discussed in section 4.2.1.

The main advantage of the approach of Callaway et al. is that it allows us to remove vertices from the network in an order that depends on their degree. If, for instance, we set  $q_k = \theta(k - k_{\max})$ , where  $\theta(x)$  is the Heaviside step function, then we remove all vertices with degrees greater than  $k_{\max}$ . This corresponds precisely to the experiment of Broder et al. [74], who looked at the behavior of the World Wide Web graph as vertices were removed in order of decreasing degree. (Similar but not identical calculations were also performed by Albert, Jeong, and Barabási [15].) In agreement with the numerical calculations (see section 3.4), Callaway et al. find that networks with power-law degree distributions are highly susceptible to this type of targeted attack; one need only remove a small percentage of vertices to destroy the giant component entirely. Similar results were also found independently by Cohen et al. [94], using a closely similar method, and in a later paper [361] some of the same authors extended their calculations to directed networks also, which show a considerably richer component structure, as described in section 4.2.3.

As an example, consider Figure 8.2, which shows the fraction of the highest degree vertices that must be removed from a network with a power-law degree distribution to destroy the giant component, as a function of the exponent  $\alpha$  of the power law [117, 318]. As the figure shows, the maximum fraction is less than three percent, and for

most values of  $\alpha$  the fraction is significantly less than this. This appears to imply that networks like the Internet and the Web that have power-law degree distributions are highly susceptible to such attacks [15, 74, 94].

These results are for the configuration model. Other models offer some further insights. The finding by Cohen et al. [93] that the threshold value  $q_c$  at which percolation sets in for the configuration model is zero for degree distributions with a divergent second moment has attracted particular interest. Vázquez and Moreno [399], for example, have shown that the threshold may be zero even for finite second moment if the degrees of adjacent vertices in the network are positively correlated (see sections 3.6 and 4.2.5). Conversely, if the second moment does diverge, there may still be a nonzero threshold if there are negative degree correlations. Warren, Sander, and Sokolov [407] have shown that there can also be a nonzero threshold for a network incorporating geographical effects, in which each vertex occupies a position in a low-dimensional space (typically two-dimensional), and probability of connection is higher for vertex pairs that are close together in that space. A similar spatial model has been studied by Rozenfeld et al. [358], and both models are closely related to continuum percolation [277].

An issue related to resilience to vertex deletion is the issue of cascading failures. In some networks, such as electrical power networks, that carry load or distribute a resource, the operation of the network is such that the failure of one vertex or edge results in the redistribution of the load on that vertex or edge to other nearby vertices or edges. If vertices or edges fail when the load on them exceeds some maximum capacity, then this mechanism can result in a cascading failure or avalanche in which the redistribution of load pushes a vertex or edge over its threshold and causes it to fail, leading to further redistribution. Such a cascading failure in the western United States in August 1996 resulted in the spread of what was initially a small power outage in El Paso, Texas, through six states as far as Oregon and California, leaving several million electricity customers without power. Watts [412] has given a simple model of this process that can be mapped onto a type of percolation model and hence can be solved using generating function methods similar to those for simple vertex removal processes above.

In Watts's model, a vertex  $i$  fails if a given fraction  $\phi_i$  of its neighbors have failed, where the quantities  $\{\phi_i\}$  are independent identically distributed (iid) variables drawn from a distribution  $f(\phi)$ . The model is seeded by the initial failure of some nonzero density  $\Phi_0$  of vertices, chosen uniformly at random. It is assumed that  $\Phi_0 \ll 1$ , so that the initial seed consists, to leading order, of single isolated vertices. Watts considers networks with the topology of the configuration model (section 4.2.1), for which, because of the vanishing density of short loops making the networks tree-like at small length-scales, each vertex will have at most only a single failed neighboring vertex in the initial stages of the cascade, and hence will fail itself if and only if its threshold for failure satisfies  $\phi < 1/k$ , where  $k$  is its degree. Watts calls vertices satisfying this criterion *vulnerable*. The probability of a vertex being vulnerable is  $q_k = \int_0^{1/k} f(\phi) d\phi$ , and the cascade will spread only if such vertices connect to form a percolating (i.e., extensive) cluster on the network. Thus the problem maps directly onto the generalized percolation process studied by Callaway et al. [81] above, allowing us to find a condition for the spread of the initial seed to give a large-scale cascade. The percolation model applies only to the vulnerable vertices, however, so to calculate the final sizes of cascades Watts performs numerical simulations.

Models of cascading failure have also been studied by Holme and Kim [194, 198], by Moreno et al. [296, 297], and by Motter and Lai [304]. In the model of Holme

and Kim, for instance, load on a vertex is quantified by the betweenness centrality of the vertex (see section 3.9), and vertices fail when the betweenness exceeds a given threshold. Holme and Kim give simulation results for the avalanche size distribution in their model.

**8.2. Epidemiological Processes.** One of the original, and still primary, reasons for studying networks is to understand the mechanisms by which diseases and other things (information, computer viruses, rumors) spread over them. For instance, the main reason for the study of networks of sexual contact [45, 154, 185, 217, 242, 264, 265, 302, 357] (section 2.1) is to help us understand and perhaps control the spread of sexually transmitted diseases. Similarly one studies networks of email contact [136, 320] to learn how computer viruses spread.<sup>34</sup>

**8.2.1. The SIR Model.** The simplest model of the spread of a disease over a network is the SIR model of epidemic disease [23, 26, 191].<sup>35</sup> This model, first formulated, though never published, by Lowell Reed and Wade Hampton Frost in the 1920s, divides the population into three classes: susceptible (S), meaning they don't have the disease of interest but can catch it if exposed to someone who does, infective<sup>36</sup> (I) meaning they have the disease and can pass it on, and recovered (R), meaning they have recovered from the disease and have permanent immunity, so that they can never get it again or pass it on. (Some authors consider the R to stand for "removed," a general term that encompasses also the possibility that people may die of the disease and remove themselves from the infective pool in that fashion. Others consider the R to mean "refractory," which is the common term among those who study the closely related area of reaction diffusion processes [385, 423].)

In traditional mathematical epidemiology [23, 26, 191], one then assumes that any susceptible individual has a uniform probability  $\beta$  per unit time of catching the disease from any infective one and that infective individuals recover and become immune at some stochastically constant rate  $\gamma$ . The fractions  $s$ ,  $i$ , and  $r$  of individuals in the states S, I, and R are then governed by the differential equations

$$(8.7) \quad \frac{ds}{dt} = -\beta is, \quad \frac{di}{dt} = \beta is - \gamma i, \quad \frac{dr}{dt} = \gamma i.$$

Models of this type are called *fully mixed*, and although they have taught us much about the basic dynamics of diseases, they are obviously unrealistic in their assumptions. In reality diseases can only spread between those individuals who have actual physical contact of one sort or another, and the structure of the contact network is important to the pattern of development of the disease.

The SIR model can be generalized in a straightforward manner to an epidemic taking place on a network, although the resulting dynamical system is substantially more complicated than its fully mixed counterpart. The important observation that allows us to make progress, first made by Grassberger [178], is that the model can

<sup>34</sup>Computer viruses are an interesting case in that the networks over which they spread are normally directed, unlike the contact networks for most human diseases [228].

<sup>35</sup>One distinguishes between an *epidemic* disease such as influenza, which sweeps through the population rapidly and infects a significant fraction of individuals in a short outbreak, and an *endemic* disease such as measles, which persists within the population at a level roughly constant over time. The SIR model is a model of the former. The SIS model discussed in section 8.2.2 is a model of the latter.

<sup>36</sup>In everyday parlance the more common word is "infectious," but infective is the standard term among epidemiologists.

be mapped exactly onto bond percolation on the same network. Indeed, as pointed out by Sander et al. [359], significantly more general models can also be mapped to percolation, in which transmission probability between pairs of individuals and the times for which individuals remain infective both vary but are chosen in iid fashion from some appropriate distributions. Let us suppose that the distribution of infection rates  $\beta$ , defined as the probability per unit time that an infective individual will pass the disease onto a particular susceptible network neighbor, is drawn from a distribution  $P_i(\beta)$ . And suppose that the recovery rate  $\gamma$  is drawn from another distribution  $P_r(\gamma)$ . Then the resulting model can be shown [314] to be equivalent to uniform bond percolation on the same network with edge occupation probability

$$(8.8) \quad T = 1 - \int_0^\infty P_i(\beta) P_r(\gamma) e^{-\beta/\gamma} d\beta d\gamma.$$

The extraction of predictions about epidemics from the percolation model is simple: the distribution of percolation clusters (i.e., components connected by occupied edges) corresponds to the distribution of the sizes of disease outbreaks that start with a randomly chosen initial carrier; the percolation transition corresponds to the “epidemic threshold” of epidemiology, above which an epidemic outbreak is possible (i.e., one that infects a nonzero fraction of the population in the limit of large system size); and the size of the giant component above this transition corresponds to the size of the epidemic. What the mapping cannot tell us, but standard epidemiological models can, is the time progression of a disease outbreak. The mapping gives us results only for the ultimate outcome of the disease in the limit of long times, in which all individuals are in either the S or R states, and no new cases of the disease are occurring. Nonetheless, there is much to be learned by studying even the non-time-varying properties of the model.

The solution of bond percolation for the configuration model was given by Callaway et al. [81], who showed that, for uniform edge occupation probability  $T$ , the distribution of the sizes of clusters (i.e., disease outbreaks in epidemiological language) is generated by the function  $H_0(x)$ , where

$$(8.9) \quad H_0(x) = xG_0(H_1(x)), \quad H_1(x) = 1 - T + TxG_1(H_1(x)),$$

where  $G_0(x)$  and  $G_1(x)$  are defined in (4.6). This gives an epidemic transition that takes place at  $T_c = 1/G'_1(1)$ , a mean outbreak size  $\langle s \rangle$  given by

$$(8.10) \quad \langle s \rangle = H'_0(1) = T \left[ 1 + \frac{TG'_0(1)}{1 - TG'_1(1)} \right],$$

and an epidemic outbreak that affects a fraction  $S$  of the network, where

$$(8.11) \quad S = 1 - G_0(u), \quad u = 1 - T + TG_1(u).$$

Similar solutions can be found for a wide variety of other model networks, including networks with correlations of various kinds between the rates of infection or the infectivity times [314], networks with correlations between the degrees of vertices [300], and networks with more complex structure, such as different types of vertices [21, 314].

One of the most important conclusions of this work is for the case of networks with power-law degree distributions, for which, as in the case of site percolation (section 8.1), there is no nonzero epidemic threshold so long as the exponent of the power



law is less than 3. Since most power-law networks satisfy this condition, we expect diseases always to propagate in these networks, regardless of transmission probability between individuals, a point that was first made, in the context of models of computer virus epidemiology, by Pastor-Satorras and Vespignani [332, 335], although, as pointed out by Lloyd and May [266, 276], precursors of the same result can be seen in the earlier work of May and Anderson [275]. May and Anderson studied traditional (fully mixed) differential equation models of epidemics, without network structure, but they divided the population into activity classes with different values of the infection rate  $\beta$ . They showed that the variation of the number of infective individuals over time depends on the variance of this rate over the classes, and in particular that the disease always multiplies exponentially if the variance diverges—precisely the situation in a network with a power-law degree distribution and exponent less than 3.

The conclusion that diseases always spread on scale-free networks has been revised somewhat in the light of later discoveries. In particular, there may be a nonzero percolation threshold for certain types of correlations between vertices [56, 57, 58, 59, 300, 399] if the network is embedded in a low-dimensional (rather than infinite-dimensional) space [358, 407] or if the network has high transitivity [139] (see section 3.2).

An interesting combination of the ideas of epidemiology with those of network resilience explored in the preceding section arises when one considers vaccination of a population against the spread of a disease. Vaccination can be regarded as the removal from a network of some particular set of vertices, and this in turn can be modeled as a site percolation process. Thus one is led to consideration of joint site/bond percolation on networks, which has also been solved, in the simplest uniformly random case, by Callaway et al. [81]. If the site percolation is correlated with vertex degree (as in (8.2) and following), for example, removing the vertices with highest degree, then one has a model for targeted vaccination strategies also. A good discussion has been given by Pastor-Satorras and Vespignani [334]. As with the models of section 8.1, one finds that networks tend to be particularly vulnerable to removal of their highest degree vertices, so this kind of targeted vaccination is expected to be particularly effective. (This of course is not news to the public health community, who have long followed a policy of focusing their most aggressive disease prevention efforts on the “core communities” of high-degree vertices in a network.)

Unfortunately, it is not always easy to find the highest degree vertices in a social network. The number of sexual contacts a person has had, for instance, can normally only be found by asking them, and perhaps not even then. An interesting method that circumvents this problem has been suggested by Cohen, ben-Avraham, and Havlin [92]. They observe that since the probability of reaching a particular vertex by following a randomly chosen edge in a graph is proportional to the vertex’s degree (section 4.2), one is more likely to find high-degree vertices by following edges than by choosing vertices at random. They propose thus that a population can be immunized by choosing a random person from that population and vaccinating a *friend* of that person, and then repeating the process. They show both by analytic calculations and by computer simulation that this strategy is substantially more effective than random vaccination. In a sense, in fact, this strategy is already in use. The “contact tracing” methods [250], used to control sexually transmitted diseases, and the “ring vaccination” method [180, 307], used to control smallpox and foot-and-mouth disease, are both examples of roughly this type of acquaintance vaccination.

**8.2.2. The SIS Model.** Not all diseases confer immunity on their survivors. Diseases that, for instance, are not self-limiting but can be cured by medicine, can usually be caught again immediately by an unlucky patient. Tuberculosis and gonorrhea are two much-studied examples. Computer viruses also fall into this category; they can be “cured” by antivirus software, but without a permanent virus-checking program the computer has no way to fend off subsequent attacks by the same virus.

With diseases of this kind carriers that are cured move from the infective pool not to a recovered pool, but back into the susceptible one. A model with this type of dynamics is called an SIS model, for obvious reasons. In the simplest, fully mixed, single-population case, its dynamics are described by the differential equations

$$(8.12) \quad \frac{ds}{dt} = -\beta is + \gamma i, \quad \frac{di}{dt} = \beta is - \gamma i,$$

where  $\beta$  and  $\gamma$  are, as before, the infection and recovery rates.

The SIS model is a model of endemic disease. Since carriers can be infected many times, it is possible, and does happen in some parameter regimes, that the disease will persist indefinitely, circulating around the population and never dying out. The equivalent of the SIR epidemic transition is the phase boundary between the parameter regimes in which the disease persists and those in which it does not.

The SIS model cannot be solved exactly on a network as the SIR model can, but a detailed mean-field treatment has been given by Pastor-Satorras and Vespignani [331, 332] for SIS epidemics on the configuration model. Their approach is based on the differential equations, (8.12), but they allow the rate of infection  $\beta$  to vary between members of the population rather than holding it constant. (This is similar to the approach of May and Anderson [275] for the SIR model, discussed in section 8.2.1, but is more general, since it does not involve the division of the population into a binned set of activity classes, as the May–Anderson approach does.) The calculation proceeds as follows.

The quantity  $\beta i$  appearing in (8.12) represents the average rate at which susceptible individuals become infected by their neighbors. For a vertex of degree  $k$ , Pastor-Satorras and Vespignani make the replacement  $\beta i \rightarrow k\lambda\Theta(\lambda)$ , where  $\lambda$  is the rate of infection via contact with a single infective individual and  $\Theta(\lambda)$  is the probability that the neighbor at the other end of an edge will in fact be infective. Note that  $\Theta$  is a function of  $\lambda$  since presumably the probability of being infective will increase as the probability of passing on the disease increases. The remaining occurrences of the variables  $s$  and  $i$  Pastor-Satorras and Vespignani replace by  $s_k$  and  $i_k$ , which are degree-dependent generalizations representing the fraction of vertices of degree  $k$  that are susceptible or infective. Then, noticing that  $i_k$  and  $s_k$  obey  $i_k + s_k = 1$ , we can rewrite (8.12) as the single differential equation

$$(8.13) \quad \frac{di_k}{dt} = k\lambda\Theta(\lambda)(1 - i_k) - i_k,$$

where we have, without loss of generality, set the recovery rate  $\gamma$  equal to 1. There is an approximation inherent in this formulation, since we have assumed that  $\Theta(\lambda)$  is the same for all vertices, when in general it too will be dependent on vertex degree. This is in the nature of a mean-field approximation and can be expected to give a reasonable guide to the qualitative behavior of the system, although certain properties (particularly close to the phase transition) may be quantitatively mispredicted.

Looking for stationary solutions, we find

$$(8.14) \quad i_k = \frac{k\lambda\Theta(\lambda)}{1 + k\lambda\Theta(\lambda)}.$$

To calculate the value of  $\Theta(\lambda)$ , one averages the probability  $i_k$  of being infected over all vertices. Since  $\Theta(\lambda)$  is defined as the probability that the vertex at the end of an edge is infective,  $i_k$  should be averaged over the distribution  $kp_k/z$  of the degrees of such vertices (see section 4.2.1), where  $z = \sum_k kp_k$  is, as usual, the mean degree. Thus

$$(8.15) \quad \Theta(\lambda) = \frac{1}{z} \sum_k kp_k i_k.$$

Eliminating  $i_k$  from (8.13) and (8.15) we then obtain an implicit expression for  $\Theta(\lambda)$ :

$$(8.16) \quad \frac{\lambda}{z} \sum_k \frac{k^2 p_k}{1 + k\lambda\Theta(\lambda)} = 1.$$

For particular choices of  $p_k$  this equation can be solved for  $\Theta(\lambda)$  either exactly or approximately. For instance, for a power-law degree distribution of the form (4.15), Pastor-Satorras and Vespignani solve it by making an integral approximation, and hence show that there is no nonzero epidemic threshold for the SIS model in the power-law case—the disease will always persist, regardless of the value of the infection rate parameter  $\lambda$  [332]. They have also generalized the solution to a number of other cases, including other degree distributions [331], finite-sized networks [333], and models that include vaccination of some fraction of individuals [334, 335]. In the latter case, they tackle both random vaccination and vaccination targeted at the vertices with highest degree using a method similar to that of Cohen et al. [93] in which they calculate the effective degree distribution of the network after the removal of a given set of vertices and then apply their mean-field method to the resulting network. As we would expect from the results of Cohen et al., propagation of the disease turns out to be relatively robust against random vaccination, at least in networks with right-skewed degree distributions, but highly susceptible to vaccination of the highest-degree individuals. The mean-field method has also been applied to networks with degree correlations of the type discussed in section 3.6; see Boguñá, Pastor-Satorras, and Vespignani [58]. Of particular note is their finding that for the case of power-law degree distributions neither assortative nor disassortative mixing by degree can produce a nonzero epidemic threshold in the SIS model, at least within the mean-field approximation. This contrasts with the case for the SIR model, where it was found that disassortative mixing can produce a nonzero threshold [399].

The mean-field method can also be applied to the SIR model [24, 298]. Although we have an exact solution for the SIR model as described in section 8.2.1, that solution can only tell us about the long-time behavior of an outbreak—its expected final size and so forth. The mean-field method, although approximate, can tell us about the time evolution of an outbreak, so the two methods are complementary. The mean-field method for the SIR model can also be used to treat approximately the effects of network transitivity [24, 154, 227, 234].

**8.3. Search on Networks.** Another example of a process taking place on a network that has important practical applications is network search. Suppose some

resource of interest is stored at the vertices of a network, such as information on Web pages, or computer files on a distributed database or file-sharing network. One would like to determine rapidly where on the network a particular item of interest can be found (or determine that it is not on the network at all). One way of doing this, which is used by Web search engines, is simply to catalog exhaustively (or “crawl”) the entire network, creating a distilled local map of the data found. Such a strategy is favored in cases where there is a heavy communication cost to searching the network in real time, so that it makes sense to create a local index. While performing a network crawl is, in principle, straightforward (although in practice it may be technically very challenging [72]), there are nonetheless some interesting theoretical questions arising.

**8.3.1. Exhaustive Network Search.** One of the triumphs of recent work on networks has been the development of effective algorithms for mining network crawl data for information of interest, particularly in the context of the World Wide Web. The important trick here turns out to be to use the information contained in the edges of the network as well as in the vertices. Since the edges, or hyperlinks, in the World Wide Web are created by people in order to highlight connections between the contents of pairs of pages, their structure contains information about page content and relevance which can help us to improve search performance. The good search engines therefore make a local catalog not only of the contents of web pages, but also of which ones link to which others. Then when a query is made of the database, usually in the form of a textual string of interest, the typical strategy would be to select a subset of pages from the database by searching for that string, and then to rank the results using the edge information. The classic algorithm, due to Brin and Page [72] and Page et al. [327], is essentially identical in its simplest form to the eigenvector centrality long used in social network analysis [66, 67, 362, 408]. Each vertex  $i$  is assigned a weight  $x_i > 0$ , which is defined to be proportional to the sum of the weights of all vertices that point to  $i$ :  $x_i = \lambda^{-1} \sum_j A_{ij} x_j$  for some  $\lambda > 0$ , or in matrix form

$$(8.17) \quad \mathbf{A}\mathbf{x} = \lambda\mathbf{x},$$

where  $\mathbf{A}$  is the (asymmetric) adjacency matrix of the graph, whose elements are  $A_{ij}$ , and  $\mathbf{x}$  is the vector whose elements are the  $x_i$ . This of course means that the weights we want are an eigenvector of the adjacency matrix with eigenvalue  $\lambda$  and, provided the network is connected (there are no separate components), the Perron–Frobenius theorem then tells us that there is only one eigenvector with all weights nonnegative, which is the unique eigenvector corresponding to the largest eigenvalue. This eigenvector can be found trivially by repeated multiplication of the adjacency matrix into any initial nonzero vector which is not itself an eigenvector.

This algorithm, which is implemented (along with many additional tricks) in the widely used search engine *Google*, appears to be highly effective. In essence the algorithm makes the assumption that a page is important if it is pointed to by other important pages. A more sophisticated version of the same idea has been put forward by Kleinberg [235, 236], who notes that, since the Web is a directed network, one can ask not only about which vertices point to a vertex of interest, but also about which vertices are pointed to by that vertex. This then leads to two different weights  $x_i$  and  $y_i$  for each vertex. Kleinberg refers to a vertex that is pointed to by highly ranked vertices as an authority—it is likely to contain relevant information. Such a vertex gets a weight  $x_i$  that is large. A vertex that points to highly ranked vertices is referred to as a hub; while it may not contain directly relevant information, it can tell you where to find such information. It gets a weight  $y_i$  that is large. (Certainly it is

possible for a vertex to have both weights large; there is no reason why the same page cannot be both a hub and an authority.) The appropriate generalization of (8.17) for the two weights is then

$$(8.18) \quad \mathbf{A}\mathbf{y} = \lambda\mathbf{x}, \quad \mathbf{A}^T\mathbf{x} = \mu\mathbf{y},$$

where  $\mathbf{A}^T$  is the transpose of  $\mathbf{A}$ . Most often we are interested in the authority weights which, eliminating  $\mathbf{y}$ , obey  $\mathbf{A}\mathbf{A}^T\mathbf{x} = \lambda\mu\mathbf{x}$ , so that the primary difference between the method of Brin and Page [72] and the method of Kleinberg is the replacement of the adjacency matrix with the symmetric product  $\mathbf{A}\mathbf{A}^T$ . More general forms than (8.18) are also possible. One could, for example, allow the authority weight of a vertex to depend on the authority weights of the vertices that point to it (and not just their hub weights, as in (8.18)). This leads to a model that interpolates smoothly between the Brin–Page and Kleinberg methods. As far as we are aware, however, this has not been tried. Neither has Kleinberg’s method been implemented yet in a commercial web search engine, to the best of our knowledge.

The methods described here can also be used for search on other directed information networks. Kleinberg’s method is particularly suitable for ranking publications in citation networks, for example. The *Citeseer* literature search engine implements a form of article ranking of this type.

**8.3.2. Guided Network Search.** An alternative approach to searching a network is to perform a guided search. Guided search strategies may be appropriate for certain kinds of Web search, particularly searches for specialized content that could be missed by generic search engines (whose coverage tends to be quite poor), and also for searching on other types of networks such as distributed databases. Exhaustive search of the type discussed in the preceding section crawls a network once to create an index of the data found, which is then stored and searched locally. Guided searches perform small special-purpose crawls for every search query, crawling only a small fraction of the network, but doing so in an intelligent fashion that deliberately seeks out the network vertices most likely to contain relevant information.

One practical example of a guided search is the specialized Web crawler or “spider” of Menczer et al. [279, 280]. This is a program that performs a Web crawl to find results for a particular query. The method used is a type of genetic algorithm [284] or enrichment method [179] that in its simplest form has a number of “agents” that start crawling the Web at random, looking for pages that contain, for example, particular words or sets of words given by the user. Agents are ranked according to their success at finding matches to the words of interest and those that are least successful are killed off. Those that are most successful are duplicated so that the density of agents will be high in regions of the Web graph that contain many pages that look promising. After some specified amount of time has passed, the search is halted and a list of the most promising pages found so far is presented to the user. The method relies for its success on the assumption that pages that contain information on a particular topic tend to be clustered together in local regions of the graph. Other than this, however, the algorithm makes little use of statistical properties of the structure of the graph.

Adamic et al. [5, 6] have given a completely different algorithm that directly exploits network structure and is designed for use on peer-to-peer networks. Their algorithm makes use of the skewed degree distribution of most networks to find the desired results quickly. It works as follows.

Simple breadth-first search can be thought of as a query that starts from a single source vertex on a network. The query goes out to all neighbors of the source vertex

and says, “Have you got the information I am looking for?” Each neighbor either replies “Yes, I have it,” in which case the search is over, or “No, I don’t, but I have forwarded your request to all of my neighbors.” Each of their neighbors, when they receive the request, either recognizes it as one they have seen before, in which case they discard it, or they repeat the process as above. A query of this kind takes aggregate effort  $O(n)$  in the network size. Adamic et al. propose to modify this algorithm as follows. The initial source vertex again queries each of its neighbors for the desired information. But now the reply is either “Yes, I have it” or “No, I don’t, and I have  $k$  neighbors,” where  $k$  is the degree of the vertex in question. Upon receiving replies of the latter type from each of its neighbors, the source vertex finds which of its neighbors has the highest value of  $k$  and passes the responsibility for the query like a runner’s baton to that neighbor, who then repeats the entire process with their neighbors. (If the highest-degree vertex has already handled the query in the past, then the second highest is chosen, and so forth; complete recursive back-tracking is used to make sure the algorithm never gets stuck in a dead end.)

The upshot of this strategy is that the baton gets passed rapidly up a chain of increasing vertex degree until it reaches the highest degree vertices in the network. On networks with highly skewed degree distributions, particularly scale-free (i.e., power-law) networks, the neighbors of the high-degree vertices account for a significant fraction of all the vertices in the network. On average, therefore, we need only go a few steps along the chain before we find a vertex with a neighbor that has the information we are looking for. The maximum degree on a scale-free network scales with network size as  $n^{1/(\alpha-1)}$  (see section 3.3.2), and hence the number of steps required to search  $O(n)$  vertices is of order  $n/n^{1/(\alpha-1)} = n^{(\alpha-2)/(\alpha-1)}$ , which lies between  $O(n^{1/2})$  and  $O(\log n)$  for  $2 \leq \alpha \leq 3$ , which is the range generally observed in power-law networks (see Table 3.1). This is a significant improvement over the  $O(n)$  of the simple breadth-first search, especially for the smaller values of  $\alpha$ .

This result differs from that given by Adamic et al. [5, 6], who adopted the more conservative assumption that the maximum degree goes as  $n^{1/\alpha}$  [8], which gives significantly poorer search times between  $O(n^{2/3})$  and  $O(n^{1/2})$ . They point out, however, that if each vertex to which the baton passes is allowed to query not only its immediate network neighbors but also its second neighbors, then the performance improves markedly to  $O(n^{2(1-2/\alpha)})$ .

The algorithm of Adamic et al. has been tested numerically on graphs with the structure of the configuration model [5] (section 4.2.1) and the Barabási–Albert preferential attachment model [5, 231] (section 7.2) and shows behavior in reasonable agreement with the expected scaling forms.

The reader might be forgiven for feeling that these algorithms are cheating a little, since the running time of the algorithm is measured by the number of hands the baton passes through. If one measures it in terms of the number of queries that must be responded to by network vertices, then the algorithm is still  $O(n)$ , just as the simple breadth-first search is. Adamic et al. suggest that each vertex therefore keep a local directory or index of the information (such as data files) stored at neighboring vertices, so that queries concerning those vertices can be resolved locally. For distributed databases and file sharing networks, where bandwidth, in terms of communication overhead between vertices, is the costly resource, this strategy really does improve scaling with network size, reducing overhead per query to  $O(\log n)$  in the best case.

**8.3.3. Network Navigation.** The work of Adamic et al. [5, 6] discussed in the preceding section considers how one can design a network search algorithm to exploit

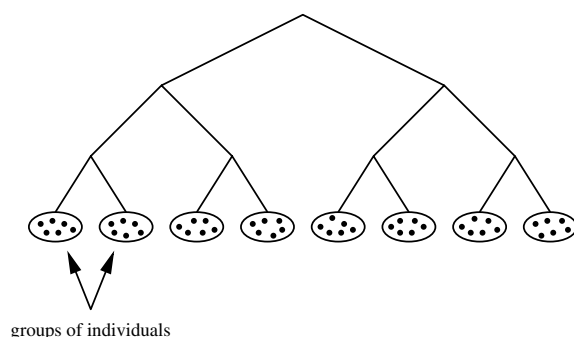
statistical features of network structure to improve performance. A complementary question has been considered by Kleinberg [237, 238]: Can one design network structures to make a particular search algorithm perform well? Kleinberg's work is motivated by the observation, discussed in section 3.8, that people are able to navigate social networks efficiently with only local information about network structure. Furthermore, this ability does not appear to depend on any particularly sophisticated behavior on the part of the people. When performing the letter-passing task of Milgram [282, 392], for instance, in which participants are asked to communicate a letter or message to a designated target person by passing it through their acquaintance network (section 2.1), the search for the target is performed, roughly speaking, using a simple "greedy algorithm." That is, at each step along the way the letter is passed to the person that the current holder believes to be closest to the target. (This in fact is precisely how participants were instructed to act in Milgram's experiments.) The fact that the letter often reaches the target in only a short time then indicates that the network itself must have some special properties, since the search algorithm clearly doesn't.

Kleinberg suggested a simple model that illustrates this behavior. His model is a variant of the small-world model of Watts [411] and Watts and Strogatz [415] (section 6) in which shortcuts are added between pairs of sites on a regular lattice (a square lattice in Kleinberg's studies). Rather than adding these shortcuts uniformly at random, as Watts and Strogatz proposed, Kleinberg adds them in a biased fashion, with shortcuts more likely to fall between lattice sites that are close together in the Euclidean space defined by the lattice. The probability of a shortcut falling between two sites goes as  $r^{-\alpha}$ , where  $r$  is the distance between the sites and  $\alpha$  is a constant. Kleinberg proves a lower bound on the mean time  $t$  (i.e., number of steps) taken by the greedy algorithm to find a randomly chosen target on such a network. His bound is  $t \geq cn^\beta$  where  $c$  is independent of  $n$  and

$$(8.19) \quad \beta = \begin{cases} (2 - \alpha)/3 & \text{for } 0 \leq \alpha < 2, \\ (\alpha - 2)/(\alpha - 1) & \text{for } \alpha > 2. \end{cases}$$

Thus the best performance of the algorithm is when  $\alpha$  is close to 2, and precisely at  $\alpha = 2$  the greedy algorithm should be capable of finding the target in  $O(\log n)$  steps. Kleinberg also gave computer simulation results confirming this result. More generally, for networks built on an underlying lattice in  $d$  dimensions, the optimal performance of the greedy algorithm occurs at  $\alpha = d$  [237, 238]. (See also [192] for some rigorous results on the performance of greedy algorithms on Watts-Strogatz-type networks.)

Kleinberg's work shows that many networks do not allow fast search using a simple algorithm such as a greedy algorithm but that it is possible to design networks that do allow such fast search. The particular model he studies, however, is quite specialized and certainly not a good representation of the real social networks that inspired his investigations. An alternative model that shows similar behavior to Kleinberg's, but which may shed more light on the true structure of social networks, has been proposed by Watts, Dodds, and Newman [414] and independently by Kleinberg [239]. The "index" experiments of Killworth and Bernard [229] and Bernard et al. [50] indicate that people in fact navigate social networks by looking for common features between their acquaintances and the target, such as geographical location or occupation. This suggests a model in which individuals are grouped (at least in the participants minds) into categories according, for instance, to their jobs.



**Fig. 8.3** The hierarchical “social distance” tree proposed by Watts, Dodds, and Newman [414] and by Kleinberg [239]. Individuals are grouped together by occupation, location, interest, etc., and then those groups are grouped together into bigger groups and so forth. The social distance between two individuals is measured by how far one must go up the tree to find the lowest “common ancestor” of the pair.

These categories may then themselves be grouped in to supercategories, and so forth, creating a tree-like hierarchy of organization that defines a “social distance” between any two people: the social distance between two individuals is measured by the height of lowest level in tree at which the two are connected—see Figure 8.3.

The tree, however, is not the network; it is merely a mental construct that affects the way the network grows. It is assumed that the probability of their being an edge between two vertices is greater the shorter the social distance between those vertices, and both Watts, Dodds, and Newman [414] and Kleinberg [239] assumed that this probability falls off exponentially with social distance. The greedy algorithm for communicating a message to a target person then specifies that the message should at each step be passed to that network neighbor of the current holder who has the shortest social distance to the target. Watts et al. showed by computer simulation that such an algorithm performs well over a broad range of parameters of the model, and Kleinberg showed that for appropriate parameter choices the search can be completed in time again  $O(\log n)$ .

While this model is primarily a model of search on social networks (or possibly the Web [239]), Watts et al. also suggested that it could be used as a model for designed networks. If one could arrange for items in a distributed database to be grouped hierarchically according to some identifiable characteristics, then a greedy algorithm that is aware of those characteristics should be able to find a desired element in the database quickly, possibly in time only logarithmic in the size of the database. This idea has been studied in more detail by Iamnitchi, Ripeanu, and Foster [204] and Arenas et al. [25].

One disadvantage of the hierarchical organizational model is that in reality the categories into which network vertices fall almost certainly overlap, whereas in the hierarchical model they are disjoint. Kleinberg has proposed a generalization of the model that allows for overlapping categories and shows search behavior qualitatively similar to the hierarchical model [239].

**8.4. Phase Transitions on Networks.** Another group of papers has dealt with the behavior on networks of traditional statistical mechanical models that show phase transitions. For example, several authors have studied spin models such as the Ising



model on networks of various kinds. Barrat and Weigt [40] studied the Ising model on networks with the topology of the small-world model [415] (see section 6) using replica methods. They found, unsurprisingly, that in the limit  $n \rightarrow \infty$  the model has a finite-temperature transition for all values of the shortcut density  $p > 0$ . Further results for Ising models on small-world networks can be found in [190, 201, 255, 336, 429], and the model has also been studied on random graphs [112, 263] and on networks with the topology of the Barabási–Albert growing network model [18, 51] (section 7.2).

The motivation behind studies of spin models on networks is usually either that they can be regarded as simple models of opinion formation in social networks [426] or that they provide general insight into the effects of network topology on phase transition processes. There are, however, other more direct approaches to both of these issues. Opinion formation can be studied more directly using actual opinion formation models [84, 108, 163, 380, 389, 402]. And Goltsev, Dorogovtsev, and Mendes [177] have examined phase transition behavior on networks using the general framework known as Landau theory. They find that the critical behavior of models on a network depends in general on the degree distribution and is in particular strongly affected by power-law degree distributions.

One class of networked systems showing a phase transition that is of real interest is the class of NP-hard computational problems such as satisfiability and colorability that show solvability transitions. The simplest example of such a system is the colorability problem, which is related to problems in operations research such as scheduling problems and also to the Potts model of statistical mechanics. In this problem a number of items (vertices) are divided into a number of groups (colors). Some pairs of vertices cannot be in the same group. Such a constraint is represented by placing an edge between those vertices, so that the set of all constraints forms a graph. A solution to the problem of satisfying all constraints simultaneously (if a solution exists) is then equivalent to finding a coloring of the graph such that no two adjacent vertices have the same color. Problems of this type are found to show a phase transition between a region of low graph density (low ratio of edges to vertices) in which most graphs are colorable, to one of high density in which most are not. A considerable amount of work has been carried out on this and similar problems in the computer science community [131]. However, this work has primarily been restricted to Poisson random graphs; it is largely an open question how the results will change when we look at more realistic network topologies. Walsh [405] has looked at colorability in the Watts–Strogatz small-world model (section 6) and found that these networks are easily colorable for both small and large values of the shortcut density parameter  $p$  but harder to color in intermediate regimes. Vázquez and Weigt [401] examined the related problem of vertex covers and found that on generalized random graphs solutions are harder to find for networks with strong degree correlations of the type discussed in section 3.6.

**8.5. Other Processes on Networks.** Preliminary investigations, primarily numerical in nature, have been carried out of the behavior of various other processes on networks. A number of authors have looked at diffusion processes. Random walks, for example, have been examined by Jespersen, Sokolov, and Blumen [215], Pandit and Amritkar [328], and Lahtinen, Kertész, and Kaski [257, 258]. Solutions of the diffusion equation can be expressed as linear combinations of eigenvectors of the graph Laplacian, which has led a number of authors to investigate the Laplacian and its eigenvalue spectrum [150, 173, 288]. Discrete dynamical processes have also attracted some attention. One of the earliest examples of a statistical model

of a networked system falls in this category, the random Boolean net of Kauffman [11, 16, 97, 98, 159, 223, 224, 225, 372], which is a model of a genetic regulatory network (see section 2.4). Cellular automata on networks have been investigated by Watts and Strogatz [411, 415], and voter models and models of opinion formation can also be regarded as cellular automata [84, 255, 402]. Iterated games on networks have been investigated by several authors [1, 135, 230, 415], and some interesting differences are seen between behavior on networks and on regular lattices. Other topics of investigation have included weakly coupled oscillators [37, 200, 415], neural networks [256, 381], and self-organized critical models [106, 251, 299]. A useful discussion of the behavior of dynamical systems on networks has been given by Strogatz [386].

**9. Summary and Directions for Future Research.** In this article we have reviewed some recent work on the structure and function of networked systems. Work in this area has been motivated to a high degree by empirical studies of real-world networks such as the Internet, the World Wide Web, social networks, collaboration networks, citation networks, and a variety of biological networks. We have reviewed these empirical studies in sections 2 and 3, focusing on a number of statistical properties of networks that have received particular attention, including path lengths, degree distributions, clustering, and resilience. Quantitative measurements for a variety of networks are summarized in Table 3.1. The most important observation to come out of studies such as these is that networks are generally very far from random. They have highly distinctive statistical signatures, some of which, such as high clustering coefficients and highly skewed degree distributions, are common to networks of a wide variety of types.

Inspired by these observations many researchers have proposed models of networks that typically seek to explain either how networks come to have the observed structure or what the expected effects of that structure will be. The largest portion of this review has been taken up with discussion of these models, covering random graph models and their generalizations (section 4), Markov graphs (section 5), the small-world model (section 6), and models of network growth, particularly the preferential attachment models (section 7).

In the last part of this review (section 8) we have discussed work on the behavior of processes that take place *on* networks. The notable successes in this area so far have been studies of the spread of infection over networks such as social networks or computer networks, and studies of the effect of the failure of network nodes on performance of communications networks. Some progress has also been made on phase transitions on networks and on dynamical systems on networks, particularly discrete dynamical systems.

In looking forward to future developments in this area it is clear that there is much to be done. The study of complex networks is still in its infancy. Several general areas stand out as promising for future research. First, while we are beginning to understand some of the patterns and statistical regularities in the structure of real-world networks, our techniques for analyzing networks are at present no more than a grab-bag of miscellaneous and largely unrelated tools. We do not yet, as we do in some other fields, have a systematic program for characterizing network structure. We count triangles on networks or measure degree sequences, but we have no idea if these are the only important quantities to measure (almost certainly they are not) or even if they are the most important. We have as yet no theoretical framework to tell us if we are even looking in the right place. Perhaps there are other measures, so far unthought-of, that are more important than those we have at present. A true

understanding of which properties of networks are the important ones to focus on will almost certainly require us to state first what questions we are interested in answering about a particular network. And knowing how to tie the answers to these questions to structural properties of the network is therefore also an important goal.

Second, there is much to be done in developing more sophisticated models of networks, both to help us understand network topology and to act as a substrate for the study of processes taking place on networks. While some network properties, such as degree distributions, have been thoroughly modeled and their causes and effects well understood, others such as correlations, transitivity, and community structure have not. It seems certain that these properties will affect the behavior of networked systems substantially, so our current lack of suitable techniques to handle them leaves a large gap in our understanding.

Which leads us to our third and perhaps most important direction for future study, the behavior of processes taking place on networks. The work described in section 8 represents only a few first attempts at answering questions about such processes, and yet this, in a sense, is our ultimate goal in this field: to understand the behavior and function of the networked systems we see around us. If we can gain such understanding, it will give us new insight into a vast array of complex and previously poorly understood phenomena.

**Acknowledgments.** For useful feedback on early versions of this article, the author would particularly like to thank Lada Adamic, Michelle Girvan, Petter Holme, Randy LeVeque, Sidney Redner, Ricard Solé, Steve Strogatz, Alexei Vázquez, and an anonymous referee. For other helpful conversations and comments about networks thanks go to Lada Adamic, László Barabási, Stefan Bornholdt, Duncan Callaway, Peter Dodds, Jennifer Dunne, Rick Durrett, Stephanie Forrest, Michelle Girvan, Jon Kleinberg, Jim Moody, Cris Moore, Martina Morris, Juyong Park, Richard Rothenberg, Larry Ruzzo, Matthew Salganik, Len Sander, Steve Strogatz, Alessandro Vespignani, Chris Warren, Duncan Watts, and Barry Wellman. For providing data used in calculations and figures, thanks go to Lada Adamic, László Barabási, Jerry Davis, Jennifer Dunne, Ramón Ferrer i Cancho, Paul Ginsparg, Jerry Grossman, Oleg Khovayko, Hawoong Jeong, David Lipman, Neo Martinez, Stephen Muth, Richard Rothenberg, Ricard Solé, Grigoriy Starchenko, Duncan Watts, Geoffrey West, Janet Wiener, and Clip2 Distributed Search Solutions. Figures 1.2a, 1.2b, and 3.4 were kindly provided by Bill Cheswick, Richard Williams, and Jim Moody, respectively.

#### REFERENCES

- [1] G. ABRAMSON AND M. KUPERMAN, *Social games in a social network*, Phys. Rev. E, 63 (2001), art. no. 030901.
- [2] L. A. ADAMIC, *The small world web*, in Research and Advanced Technology for Digital Libraries, Lecture Notes in Comput. Sci. 1696, S. Abiteboul and A.-M. Vercoustre, eds., Springer-Verlag, New York, 1999, pp. 443–452.
- [3] L. A. ADAMIC AND E. ADAR, *Friends and neighbors on the Web*, Social Networks (to appear).
- [4] L. A. ADAMIC AND B. A. HUBERMAN, *Power-law distribution of the world wide web*, Science, 287 (2000), p. 2115a.
- [5] 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.
- [6] 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.
- [7] R. K. AHUJA, T. L. MAGNANTI, AND J. B. ORLIN, *Network Flows: Theory, Algorithms, and Applications*, Prentice-Hall, Upper Saddle River, NJ, 1993.

- [8] 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, Association of Computing Machinery, New York, 2000, pp. 171–180.
- [9] W. AIELLO, F. CHUNG, AND L. LU, *Random evolution of massive graphs*, in Handbook of Massive Data Sets, J. Abello, P. M. Pardalos, and M. G. C. Resende, eds., Kluwer Academic, Dordrecht, 2002, pp. 97–122.
- [10] R. ALBERICH, J. MIRO-JULIA, AND F. ROSSELLO, *Marvel Universe Looks Almost Like a Real Social Network*, Preprint 0202174 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [11] R. ALBERT AND A.-L. BARABÁSI, *Dynamics of complex systems: Scaling laws for the period of Boolean networks*, Phys. Rev. Lett., 84 (2000), pp. 5660–5663.
- [12] R. ALBERT AND A.-L. BARABÁSI, *Topology of evolving networks: Local events and universality*, Phys. Rev. Lett., 85 (2000), pp. 5234–5237.
- [13] R. ALBERT AND A.-L. BARABÁSI, *Statistical mechanics of complex networks*, Rev. Modern Phys., 74 (2002), pp. 47–97.
- [14] R. ALBERT, H. JEONG, AND A.-L. BARABÁSI, *Diameter of the world-wide web*, Nature, 401 (1999), pp. 130–131.
- [15] R. ALBERT, H. JEONG, AND A.-L. BARABÁSI, *Attack and error tolerance of complex networks*, Nature, 406 (2000), pp. 378–382.
- [16] M. ALDANA, *Dynamics of Boolean Networks with Scale-Free Topology*, Preprint 0209571 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [17] D. ALDOUS AND B. PITTEL, *On a random graph with immigrating vertices: Emergence of the giant component*, Random Structures Algorithms, 17 (2000), pp. 79–102.
- [18] A. ALEKSIEJUK, J. A. HOLYST, AND D. STAUFFER, *Ferromagnetic phase transition in Barabási–Albert networks*, Phys. A, 310 (2002), pp. 260–266.
- [19] E. ALMAAS, R. V. KULKARNI, AND D. STROUD, *Characterizing the structure of small-world networks*, Phys. Rev. Lett., 88 (2002), art. no. 098101.
- [20] L. A. N. AMARAL, A. SCALA, M. BARTHÉLÉMY, AND H. E. STANLEY, *Classes of small-world networks*, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 11149–11152.
- [21] L. ANCEL MEYERS, M. E. J. NEWMAN, M. MARTIN, AND S. SCHRAG, *Applying network theory to epidemics: Control measures for outbreaks of Mycoplasma pneumoniae*, Emerging Infectious Diseases, 9 (2001), pp. 204–210.
- [22] C. ANDERSON, S. WASSERMAN, AND B. CROUCH, *A  $p^*$  primer: Logit models for social networks*, Social Networks, 21 (1999), pp. 37–66.
- [23] R. M. ANDERSON AND R. M. MAY, *Infectious Diseases of Humans*, Oxford University Press, Oxford, 1991.
- [24] H. ANDERSSON, *Epidemic models and social networks*, Math. Sci., 24 (1999), pp. 128–147.
- [25] A. ARENAS, A. CABRALES, A. DÍAZ-GUILERA, R. GUIMERÀ, AND F. VEGA-REDONDO, *Search and congestion in complex networks*, in Proceedings of the 18th Sitges Conference on Statistical Mechanics, R. Pastor-Satorras and J. Rubi, eds., Lecture Notes in Phys., Springer-Verlag, Berlin, 2003.
- [26] N. T. J. BAILEY, *The Mathematical Theory of Infectious Diseases and Its Applications*, Hafner Press, New York, 1975.
- [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. SCALIA-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. Sociology, 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, AND H. JEONG, *Mean-field theory for scale-free random networks*, Phys. A, 272 (1999), pp. 173–187.
- [34] A.-L. BARABÁSI, R. ALBERT, AND H. JEONG, *Scale-free characteristics of random networks: The topology of the World Wide Web*, Phys. A, 281 (2000), pp. 69–77.
- [35] A.-L. BARABÁSI, R. ALBERT, H. JEONG, AND G. BIANCONI, *Power-law distribution of the World Wide Web*, Science, 287 (2000), p. 2115a.
- [36] A.-L. BARABÁSI, H. JEONG, E. RAVASZ, Z. NÉDA, A. SCHUBERTS, AND T. VICSEK, *Evolution of the social network of scientific collaborations*, Phys. A, 311 (2002), pp. 590–614.

- [37] M. BARAHONA AND L. M. PECORA, *Synchronization in small-world systems*, Phys. Rev. Lett., 89 (2002), art. no. 054101.
- [38] A. D. BARBOUR AND G. REINERT, *Small worlds*, Random Structures Algorithms, 19 (2001), pp. 54–74.
- [39] A. BARRAT, *Comment on “Small-World Networks: Evidence for Crossover Picture,”* Preprint 9903323 (1999); available from <http://arxiv.org/abs/cond-mat/>.
- [40] A. BARRAT AND M. WEIGT, *On the properties of small-world networks*, Eur. Phys. J. B, 13 (2000), pp. 547–560.
- [41] M. BARTHÉLÉMY AND L. A. N. AMARAL, *Erratum: Small-world networks: Evidence for a crossover picture*, Phys. Rev. Lett., 82 (1999), p. 5180.
- [42] M. BARTHÉLÉMY AND L. A. N. AMARAL, *Small-world networks: Evidence for a crossover picture*, Phys. Rev. Lett., 82 (1999), pp. 3180–3183.
- [43] V. BATAGELJ AND A. MRVAR, *Some analyses of Erdős collaboration graph*, Social Networks, 22 (2000), pp. 173–186.
- [44] M. BAUER AND D. BERNARD, *A Simple Asymmetric Evolving Random Network*, Preprint 0203232 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [45] P. S. BEARMAN, J. MOODY, AND K. STOVEL, *Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks*, preprint, Department of Sociology, Columbia University, New York, 2002.
- [46] A. BEKESSY, P. BEKESSY, AND J. KOMLOS, *Asymptotic enumeration of regular matrices*, Stud. Sci. Math. Hungar., 7 (1972), pp. 343–353.
- [47] E. A. BENDER AND E. R. CANFIELD, *The asymptotic number of labeled graphs with given degree sequences*, J. Combin. Theory Ser. A, 24 (1978), pp. 296–307.
- [48] J. BERG AND M. LÄSSIG, *Correlated random networks*, Phys. Rev. Lett., 89 (2002), art. no. 228701.
- [49] J. BERG, M. LÄSSIG, AND A. WAGNER, *Evolution Dynamics of Protein Networks*, Preprint 0207711 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [50] H. R. BERNARD, P. D. KILLWORTH, M. J. EVANS, C. MCCARTY, AND G. A. SHELLEY, *Studying social relations cross-culturally*, Ethnology, 2 (1988), pp. 155–179.
- [51] G. BIANCONI, *Mean Field Solution of the Ising Model on a Barabási–Albert Network*, Preprint 0204455 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [52] G. BIANCONI AND A.-L. BARABÁSI, *Bose–Einstein condensation in complex networks*, Phys. Rev. Lett., 86 (2001), pp. 5632–5635.
- [53] G. BIANCONI AND A.-L. BARABÁSI, *Competition and multiscaling in evolving networks*, Europhys. Lett., 54 (2001), pp. 436–442.
- [54] G. BIANCONI AND A. CAPOCCI, *Number of loops of size  $h$  in growing scale-free networks*, Phys. Rev. Lett., 90 (2003), art. no. 078701.
- [55] S. BILKE AND C. PETERSON, *Topological properties of citation and metabolic networks*, Phys. Rev. E, 64 (2001), art. no. 036106.
- [56] P. BLANCHARD, C.-H. CHANG, AND T. KRÜGER, *Epidemic Thresholds on Scale-Free Graphs: The Interplay between Exponent and Preferential Choice*, Preprint 0207319 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [57] M. BOGUÑÁ AND R. PASTOR-SATORRAS, *Epidemic spreading in correlated complex networks*, Phys. Rev. E, 66 (2002), art. no. 047104.
- [58] M. BOGUÑÁ, R. PASTOR-SATORRAS, AND A. VESPIGNANI, *Absence of Epidemic Threshold in Scale-Free Networks with Connectivity Correlations*, Preprint 0208163 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [59] M. BOGUÑÁ, R. PASTOR-SATORRAS, AND A. VESPIGNANI, *Epidemic spreading in complex networks with degree correlations*, in Proceedings of the 18th Sitges Conference on Statistical Mechanics, R. Pastor-Satorras and J. Rubi, eds., Lecture Notes in Phys., Springer-Verlag, Berlin, 2003.
- [60] B. BOLLOBÁS, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin., 1 (1980), pp. 311–316.
- [61] B. BOLLOBÁS, *The diameter of random graphs*, Trans. Amer. Math. Soc., 267 (1981), pp. 41–52.
- [62] B. BOLLOBÁS, *Modern Graph Theory*, Springer-Verlag, New York, 1998.
- [63] B. BOLLOBÁS, *Random Graphs*, 2nd ed., Academic Press, New York, 2001.
- [64] B. BOLLOBÁS AND O. RIORDAN, *The Diameter of a Scale-Free Random Graph*, preprint, Department of Mathematical Sciences, University of Memphis, 2002.
- [65] B. BOLLOBÁS, O. RIORDAN, J. SPENCER, AND G. TUSNÁDY, *The degree sequence of a scale-free random graph process*, Random Structures Algorithms, 18 (2001), pp. 279–290.

- [66] P. F. BONACICH, *A technique for analyzing overlapping memberships*, in Sociological Methodology, H. Costner, ed., Jossey-Bass, San Francisco, CA, 1972.
- [67] P. F. BONACICH, *Power and centrality: A family of measures*, Amer. J. Sociol., 92 (1987), pp. 1170–1182.
- [68] M. BORDENS AND I. GÓMEZ, *Collaboration networks in science*, in The Web of Knowledge: A Festschrift in Honor of Eugene Garfield, B. Cronin and H. B. Atkins, eds., Information Today, Medford, NJ, 2000.
- [69] S. BORNHOLDT AND H. EBEL, *World Wide Web scaling exponent from Simon's 1955 model*, Phys. Rev. E, 64 (2001), art. no. 035104.
- [70] S. BORNHOLDT AND H. G. SCHUSTER, EDS., *Handbook of Graphs and Networks*, Wiley-VCH, Berlin, 2003.
- [71] R. L. BREIGER, S. A. BOORMAN, AND P. ARABIE, *An algorithm for clustering relations data with applications to social network analysis and comparison with multidimensional scaling*, J. Math. Psych., 12 (1975), pp. 328–383.
- [72] S. BRIN AND L. PAGE, *The anatomy of a large-scale hypertextual Web search engine*, Computer Networks, 30 (1998), pp. 107–117.
- [73] S. R. BROADBENT AND J. M. HAMMERSLEY, *Percolation processes: I. Crystals and mazes*, Proc. Cambridge Philos. Soc., 53 (1957), pp. 629–641.
- [74] A. BRODER, R. KUMAR, F. MAGHOUL, P. RAGHAVAN, S. RAJAGOPALAN, R. STATA, A. TOMKINS, AND J. WIENER, *Graph structure in the web*, Computer Networks, 33 (2000), pp. 309–320.
- [75] 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., Proc. SPIE 4526, International Society for Optical Engineering, Bellingham, WA, 2001, pp. 172–187.
- [76] M. BUCHANAN, *Nexus: Small Worlds and the Groundbreaking Science of Networks*, Norton, New York, 2002.
- [77] Z. BURDA, J. D. CORREIA, AND A. KRZYWICKI, *Statistical ensemble of scale-free random graphs*, Phys. Rev. E, 64 (2001), art. no. 046118.
- [78] 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.
- [79] G. CALDARELLI, R. PASTOR-SATORRAS, AND A. VESPIGNANI, *Cycles Structure and Local Ordering in Complex Networks*, Preprint 0212026 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [80] 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.
- [81] 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.
- [82] J. CAMACHO, R. GUIMERÀ, AND L. A. N. AMARAL, *Robust patterns in food web structure*, Phys. Rev. Lett., 88 (2002), art. no. 228102.
- [83] A. CAPOCCI, G. CALDARELLI, AND P. DE LOS RIOS, *Quantitative Description and Modeling of Real Networks*, Preprint 0206336 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [84] C. CASTELLANO, D. VILONE, AND A. VESPIGNANI, *Incomplete Ordering of the Voter Model on Small-World Networks*, Preprint 0210465 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [85] J. A. CATANIA, T. J. COATES, S. KEGELS, AND M. T. FULLILOVE, *The population-based AMEN (AIDS in Multi-Ethnic Neighborhoods) study*, Amer. J. Public Health, 82 (1992), pp. 284–287.
- [86] Q. CHEN, H. CHANG, R. GOVINDAN, S. JAMIN, S. J. SHENKER, AND W. WILLINGER, *The origin of power laws in Internet topologies revisited*, in Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE Computer Society, Los Alamitos, CA, 2002.
- [87] G. CHOWELL, J. M. HYMAN, AND S. EUBANK, *Analysis of a Real World Network: The City of Portland*, Technical Report BU-1604-M, Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, 2002.
- [88] F. CHUNG AND L. LU, *The average distances in random graphs with given expected degrees*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 15879–15882.
- [89] F. CHUNG AND L. LU, *Connected components in random graphs with given degree sequences*, Ann. Comb., 6 (2002), pp. 125–145.
- [90] F. CHUNG, L. LU, T. G. DEWEY, AND D. J. GALAS, *Duplication models for biological networks*, J. Comput. Biology (to appear).
- [91] J. E. COHEN, F. BRIAND, AND C. M. NEWMAN, *Community Food Webs: Data and Theory*, Springer-Verlag, New York, 1990.

- [92] R. COHEN, D. BEN-AVRAHAM, AND S. HAVLIN, *Efficient Immunization of Populations and Computers*, Preprint 0207387 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [93] 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.
- [94] R. COHEN, K. EREZ, D. BEN-AVRAHAM, AND S. HAVLIN, *Breakdown of the Internet under intentional attack*, Phys. Rev. Lett., 86 (2001), pp. 3682–3685.
- [95] R. COHEN AND S. HAVLIN, *Scale-free networks are ultrasmall*, Phys. Rev. Lett., 90 (2003), art. no. 058701.
- [96] R. C. CONNOR, M. R. HEITHAUS, AND L. M. BARRE, *Superalliance of bottlenose dolphins*, Nature, 397 (1999), pp. 571–572.
- [97] S. N. COPPERSMITH, L. P. KADANOFF, AND Z. ZHANG, *Reversible Boolean networks: I. Distribution of cycle lengths*, Phys. D, 149 (2000), pp. 11–29.
- [98] S. N. COPPERSMITH, L. P. KADANOFF, AND Z. ZHANG, *Reversible Boolean networks: II. Phase transition, oscillation, and local structures*, Phys. D, 157 (2001), pp. 54–74.
- [99] S. R. CORMAN, T. KUHN, R. D. MCPHEE, AND K. J. DOOLEY, *Studying complex discursive systems: Centering resonance analysis of organizational communication*, Human Communication Res., 28 (2002), pp. 157–206.
- [100] S. COULUMB AND M. BAUER, *Asymmetric Evolving Random Networks*, Preprint 0212371 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [101] D. CRANE, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*, University of Chicago Press, Chicago, 1972.
- [102] J. DAVIDSEN, H. EBEL, AND S. BORNHOLDT, *Emergence of a small world from local interactions: Modeling acquaintance networks*, Phys. Rev. Lett., 88 (2002), art. no. 128701.
- [103] A. DAVIS, B. B. GARDNER, AND M. R. GARDNER, *Deep South*, University of Chicago Press, Chicago, 1941.
- [104] G. F. DAVIS AND H. R. GREVE, *Corporate elite networks and governance changes in the 1980s*, Amer. J. Sociol., 103 (1997), pp. 1–37.
- [105] G. F. DAVIS, M. YOO, AND W. E. BAKER, *The Small World of the Corporate Elite*, preprint, University of Michigan Business School, Ann Arbor, MI, 2001.
- [106] L. DE ARCANGELIS AND H. J. HERRMANN, *Self-organized criticality on small world networks*, Phys. A, 308 (2002), pp. 545–549.
- [107] R. DE CASTRO AND J. W. GROSSMAN, *Famous trails to Paul Erdős*, Math. Intelligencer, 21 (1999), pp. 51–63.
- [108] M. H. DE GROOT, *Reaching a consensus*, J. Amer. Statist. Assoc., 69 (1974), pp. 118–121.
- [109] M. A. DE MENEZES, C. MOUKARZEL, AND T. J. P. PENNA, *First-order transition in small-world networks*, Europhys. Lett., 50 (2000), pp. 574–579.
- [110] P. S. DODDS, R. MUHAMAD, AND D. J. WATTS, *An Experiment Study of Social Search and the Small World Problem*, preprint, Department of Sociology, Columbia University, New York, 2002.
- [111] P. S. DODDS AND D. H. ROTHMAN, *Geometry of river networks*, Phys. Rev. E, 63 (2001), art. no. 016115, 016116, 016117.
- [112] S. N. DOROGOVTSSEV, A. 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.
- [113] S. N. DOROGOVTSSEV, A. V. GOLTSEV, AND J. F. F. MENDES, *Pseudofractal scale-free web*, Phys. Rev. E, 65 (2002), art. no. 066122.
- [114] S. N. DOROGOVTSSEV AND J. F. F. MENDES, *Evolution of networks with aging of sites*, Phys. Rev. E, 62 (2000), pp. 1842–1845.
- [115] S. N. DOROGOVTSSEV AND J. F. F. MENDES, *Exactly solvable small-world network*, Europhys. Lett., 50 (2000), pp. 1–7.
- [116] S. N. DOROGOVTSSEV AND J. F. F. MENDES, *Scaling behaviour of developing and decaying networks*, Europhys. Lett., 52 (2000), pp. 33–39.
- [117] S. N. DOROGOVTSSEV AND J. F. F. MENDES, *Comment on “Breakdown of the Internet under intentional attack,”* Phys. Rev. Lett., 87 (2001), art. no. 219801.
- [118] S. N. DOROGOVTSSEV AND J. F. F. MENDES, *Effect of the accelerating growth of communications networks on their structure*, Phys. Rev. E, 63 (2001), art. no. 025101.
- [119] S. N. DOROGOVTSSEV AND J. F. F. MENDES, *Language as an evolving word web*, Proc. Roy. Soc. London Ser. B, 268 (2001), pp. 2603–2606.
- [120] S. N. DOROGOVTSSEV AND J. F. F. MENDES, *Evolution of networks*, Adv. in Phys., 51 (2002), pp. 1079–1187.
- [121] S. N. DOROGOVTSSEV AND J. F. F. MENDES, *Accelerated growth of networks*, in Handbook of Graphs and Networks, S. Bornholdt and H. G. Schuster, eds., Wiley-VCH, Berlin, 2003, pp. 318–341.

- [122] S. N. DOROGOVTSSEV AND J. F. F. MENDES, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press, Oxford, 2003.
- [123] S. N. DOROGOVTSSEV, J. F. F. MENDES, AND A. N. SAMUKHIN, *Structure of growing networks with preferential linking*, Phys. Rev. Lett., 85 (2000), pp. 4633–4636.
- [124] S. N. DOROGOVTSSEV, J. F. F. MENDES, AND A. N. SAMUKHIN, *Anomalous percolation properties of growing networks*, Phys. Rev. E, 64 (2001), art. no. 066110.
- [125] S. N. DOROGOVTSSEV, J. F. F. MENDES, AND A. N. SAMUKHIN, *Giant strongly connected component of directed networks*, Phys. Rev. E, 64 (2001), art. no. 025101.
- [126] S. N. DOROGOVTSSEV, J. F. F. MENDES, AND A. N. SAMUKHIN, *Size-dependent degree distribution of a scale-free growing network*, Phys. Rev. E, 63 (2001), art. no. 062101.
- [127] S. N. DOROGOVTSSEV, J. F. F. MENDES, AND A. N. SAMUKHIN, *Metric Structure of Random Networks*, Preprint 0210085 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [128] S. N. DOROGOVTSSEV, J. F. F. MENDES, AND A. N. SAMUKHIN, *Principles of Statistical Mechanics of Random Networks*, Preprint 0204111 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [129] S. N. DOROGOVTSSEV AND A. N. SAMUKHIN, *Mesoscopics and Fluctuations in Networks*, Preprint 0211518 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [130] J. P. K. DOYE, *Network topology of a potential energy landscape: A static scale-free network*, Phys. Rev. Lett., 88 (2002), art. no. 238701.
- [131] D. DU, J. GU, AND P. M. PARDALOS, EDS., *Satisfiability Problem: Theory and Applications*, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 35, AMS, Providence, RI, 1997.
- [132] 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.
- [133] J. A. DUNNE, R. J. WILLIAMS, AND N. D. MARTINEZ, *Network structure and biodiversity loss in food webs: Robustness increases with connectance*, Ecology Lett., 5 (2002), pp. 558–567.
- [134] R. T. DURRETT, *Rigorous Results for the Callaway–Hopcroft–Kleinberg–Newman–Strogatz Model*, preprint, Department of Mathematics, Cornell University, Ithaca, NY, 2003.
- [135] H. EBEL AND S. BORNHOLDT, *Co-evolutionary games on networks*, Phys. Rev. E, 66 (2002), art. no. 056118.
- [136] H. EBEL, L.-I. MIELSCH, AND S. BORNHOLDT, *Scale-free topology of e-mail networks*, Phys. Rev. E, 66 (2002), art. no. 035103.
- [137] J.-P. ECKMANN AND E. MOSES, *Curvature of co-links uncovers hidden thematic layers in the world wide web*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 5825–5829.
- [138] L. EGGHE AND R. ROUSSEAU, *Introduction to Informetrics*, Elsevier, Amsterdam, 1990.
- [139] V. M. EGUILUZ AND K. KLEMM, *Epidemic threshold in structured scale-free networks*, Phys. Rev. Lett., 89 (2002), art. no. 108701.
- [140] M. EIGEN AND P. SCHUSTER, *The Hypercycle: A Principle of Natural Self-Organization*, Springer-Verlag, New York, 1979.
- [141] P. ERDŐS AND A. RÉNYI, *On random graphs*, Publ. Math. Debrecen, 6 (1959), pp. 290–297.
- [142] P. ERDŐS AND A. RÉNYI, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl., 5 (1960), pp. 17–61.
- [143] P. ERDŐS AND A. RÉNYI, *On the strength of connectedness of a random graph*, Acta Math. Acad. Sci. Hungar., 12 (1961), pp. 261–267.
- [144] G. ERGÜN, *Human sexual contact network as a bipartite graph*, Phys. A, 308 (2002), pp. 483–488.
- [145] G. ERGÜN AND G. J. RODGERS, *Growing random networks with fitness*, Phys. A, 303 (2002), pp. 261–272.
- [146] K. A. ERIKSEN, I. SIMONSEN, S. MASLOV, AND K. SNEPPEN, *Modularity and Extreme Edges of the Internet*, Preprint 0212001 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [147] B. EVERITT, *Cluster Analysis*, John Wiley, New York, 1974.
- [148] M. FALOUTSOS, P. FALOUTSOS, AND C. FALOUTSOS, *On power-law relationships of the internet topology*, Computer Communications Rev., 29 (1999), pp. 251–262.
- [149] T. J. FARARO AND M. SUNSHINE, *A Study of a Biased Friendship Network*, Syracuse University Press, Syracuse, NY, 1964.
- [150] I. J. FARKAS, I. DERÉNYI, A.-L. BARABÁSI, AND T. VICSEK, *Spectra of “real-world” graphs: Beyond the semicircle law*, Phys. Rev. E, 64 (2001), art. no. 026704.
- [151] I. J. FARKAS, I. DERÉNYI, H. JEONG, Z. NEDA, Z. N. OLTVAI, E. RAVASZ, A. SCHURBERT, A.-L. BARABÁSI, AND T. VICSEK, *Networks in life: Scaling properties and eigenvalue spectra*, Phys. A, 314 (2002), pp. 25–34.
- [152] I. J. FARKAS, H. JEONG, T. VICSEK, A.-L. BARABÁSI, AND Z. N. OLTVAI, *The topology of the transcription regulatory network in the yeast, Saccharomyces cerevisiae*, Phys. A, 381 (2003), pp. 601–612.



- [153] D. A. FELL AND A. WAGNER, *The small world of metabolism*, Nature Biotechnology, 18 (2000), pp. 1121–1122.
- [154] 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.
- [155] 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.
- [156] R. FERRER I CANCHO AND R. V. SOLÉ, *Optimization in Complex Networks*, Preprint 0111222 (2001); available from <http://arxiv.org/abs/cond-mat/>.
- [157] R. FERRER I CANCHO AND R. V. SOLÉ, *The small world of human language*, Proc. Roy. Soc. London Ser. B, 268 (2001), pp. 2261–2265.
- [158] G. W. FLAKE, S. R. LAWRENCE, C. L. GILES, AND F. M. COETZEE, *Self-organization and identification of Web communities*, IEEE Computer, 35 (2002), pp. 66–71.
- [159] J. J. FOX AND C. C. HILL, *From topology to dynamics in biochemical networks*, Chaos, 11 (2001), pp. 809–815.
- [160] O. FRANK AND D. STRAUSS, *Markov graphs*, J. Amer. Statist. Assoc., 81 (1986), pp. 832–842.
- [161] L. FREEMAN, *A set of measures of centrality based upon betweenness*, Sociometry, 40 (1977), pp. 35–41.
- [162] L. C. FREEMAN, *Some antecedents of social network analysis*, Connections, 19 (1996), pp. 39–42.
- [163] J. R. P. FRENCH, *A formal theory of social power*, Psych. Rev., 63 (1956), pp. 181–194.
- [164] A. FRONCZAK, P. FRONCZAK, AND J. A. HOLYST, *Exact Solution for Average Path Length in Random Graphs*, Preprint 0212230 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [165] A. FRONCZAK, J. A. HOLYST, M. JEDYNIAK, AND J. SIENKIEWICZ, *Higher order clustering coefficients in Barabasi-Albert networks*, Phys. A, 316 (2002), pp. 688–694.
- [166] V. GAFYCHUK, I. LUBASHEVSKY, AND A. STOSYK, *Remarks on Scaling Properties Inherent to the Systems with Hierarchically Organized Supplying Network*, Preprint 0004033 (2000); available from <http://arxiv.org/abs/nlin/>.
- [167] J. GALASKIEWICZ, *Social Organization of an Urban Grants Economy*, Academic Press, New York, 1985.
- [168] J. GALASKIEWICZ AND P. V. MARSDEN, *Interorganizational resource networks: Formal patterns of overlap*, Social Sci. Res., 7 (1978), pp. 89–107.
- [169] E. GARFIELD, *It's a small world after all*, Current Contents, 43 (1979), pp. 5–10.
- [170] I. GARFINKEL, D. A. GLEI, AND S. S. McLANAHAN, *Assortative mating among unmarried parents*, J. Population Economics, 15 (2002), pp. 417–432.
- [171] M. GIRVAN AND M. E. J. NEWMAN, *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 8271–8276.
- [172] P. M. GLEISS, P. F. STADLER, A. WAGNER, AND D. A. FELL, *Relevant cycles in chemical reaction networks*, Adv. in Complex Systems, 4 (2001), pp. 207–226.
- [173] K.-I. GOH, B. KAHNG, AND D. KIM, *Spectra and eigenvectors of scale-free networks*, Phys. Rev. E, 64 (2001), art. no. 051903.
- [174] 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.
- [175] 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.
- [176] L. GOLDWASSER AND J. ROUGHGARDEN, *Construction and analysis of a large Caribbean food web*, Ecology, 74 (1993), pp. 1216–1233.
- [177] A. V. GOLTSEV, S. N. DOROGVTSEV, AND J. F. F. MENDES, *Critical phenomena in networks*, Phys. Rev. E, 67 (2003), art. no. 026123.
- [178] P. GRASSBERGER, *On the critical behavior of the general epidemic process and dynamical percolation*, Math. Biosci., 63 (1983), pp. 157–172.
- [179] P. GRASSBERGER, *Go with the winners: A general Monte Carlo strategy*, Comput. Phys. Comm., 147 (2002), pp. 64–70.
- [180] D. GREENHALGH, *Optimal control of an epidemic by ring vaccination*, Comm. Statist. Stochastic Models, 2 (1986), pp. 339–363.
- [181] J. W. GROSSMAN AND P. D. F. ION, *On a portion of the well-known collaboration graph*, Congr. Numer., 108 (1995), pp. 129–131.
- [182] J. GUARE, *Six Degrees of Separation: A Play*, Vintage, New York, 1990.
- [183] N. GUELZIM, S. BOTTANI, P. BOURGINE, AND F. KEPES, *Topological and causal structure of the yeast transcriptional regulatory network*, Nature Genetics, 31 (2002), pp. 60–63.

- [184] R. GUIMERÀ, L. DANON, A. DÍAZ-GUILERA, F. GIRALT, AND A. ARENAS, *Self-Similar Community Structure in Organisations*, Preprint 0211498 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [185] S. GUPTA, R. M. ANDERSON, AND R. M. MAY, *Networks of sexual contacts: Implications for the pattern of spread of HIV*, AIDS, 3 (1989), pp. 807–817.
- [186] J. M. HAMMERSLEY, *Percolation processes: II. The connective constant*, Proc. Cambridge Philos. Soc., 53 (1957), pp. 642–645.
- [187] F. HARARY, *Graph Theory*, Perseus, Cambridge, MA, 1995.
- [188] B. HAYES, *Graph theory in practice: Part I*, Amer. Sci., 88 (2000), pp. 9–13.
- [189] B. HAYES, *Graph theory in practice: Part II*, Amer. Sci., 88 (2000), pp. 104–109.
- [190] C. P. HERRERO, *Ising model in small-world networks*, Phys. Rev. E, 65 (2002), art. no. 066110.
- [191] H. W. HETHCOTE, *The mathematics of infectious diseases*, SIAM Rev., 42 (2000), pp. 599–653.
- [192] D. J. HIGHAM, *Greedy Pathlengths and Small World Graphs*, Mathematics Research Report 8, University of Strathclyde, Glasgow, UK, 2002.
- [193] P. W. HOLLAND AND S. LEINHARDT, *An exponential family of probability distributions for directed graphs*, J. Amer. Statist. Assoc., 76 (1981), pp. 33–65.
- [194] P. HOLME, *Edge overload breakdown in evolving networks*, Phys. Rev. E, 66 (2002), art. no. 036119.
- [195] P. HOLME, C. R. EDLING, AND F. LILJEROS, *Structure and Time-Evolution of the Internet Community pussokram.com*, Preprint 0210514 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [196] P. HOLME, M. HUSS, AND H. JEONG, *Subnetwork Hierarchies of Biochemical Pathways*, Preprint 0206292 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [197] P. HOLME AND B. J. KIM, *Growing scale-free networks with tunable clustering*, Phys. Rev. E, 65 (2002), art. no. 026107.
- [198] P. HOLME AND B. J. KIM, *Vertex overload breakdown in evolving networks*, Phys. Rev. E, 65 (2002), art. no. 066109.
- [199] 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.
- [200] H. HONG, M. Y. CHOI, AND B. J. KIM, *Synchronization on small-world networks*, Phys. Rev. E, 65 (2002), art. no. 026139.
- [201] 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.
- [202] B. A. HUBERMAN, *The Laws of the Web*, MIT Press, Cambridge, MA, 2001.
- [203] M. HUXHAM, S. BEANEY, AND D. RAFFAELLI, *Do parasites reduce the chances of triangulation in a real food web?*, Oikos, 76 (1996), pp. 284–300.
- [204] A. IAMNITCHI, M. RIPEANU, AND I. FOSTER, *Locating data in (small-world?) peer-to-peer scientific collaborations*, in Proceedings of the First International Workshop on Peer-to-Peer Systems, P. Druschel, F. Kaashoek, and A. Rowstron, eds., Lecture Notes in Comput. Sci. 2429, Springer-Verlag, Berlin, 2002, pp. 232–241.
- [205] T. ITO, T. CHIBA, R. OZAWA, M. YOSHIDA, M. HATTORI, AND Y. SAKAKI, *A comprehensive two-hybrid analysis to explore the yeast protein interactome*, Proc. Natl. Acad. Sci. USA, 98 (2001), pp. 4569–4574.
- [206] A. JAFFE AND M. TRAJTENBERG, *Patents, Citations and Innovations: A Window on the Knowledge Economy*, MIT Press, Cambridge, MA, 2002.
- [207] A. K. JAIN, M. N. MURTY, AND P. J. FLYNN, *Data clustering: A review*, ACM Comput. Surveys, 31 (1999), pp. 264–323.
- [208] S. JAIN AND S. KRISHNA, *Autocatalytic sets and the growth of complexity in an evolutionary model*, Phys. Rev. Lett., 81 (1998), pp. 5684–5687.
- [209] S. JAIN AND S. KRISHNA, *A model for the emergence of cooperation, interdependence, and structure in evolving networks*, Proc. Natl. Acad. Sci. USA, 98 (2001), pp. 543–547.
- [210] S. JANSON, T. ŁUCZAK, AND A. RUCINSKI, *Random Graphs*, John Wiley, New York, 1999.
- [211] H. JEONG, S. MASON, A.-L. BARABÁSI, AND Z. N. OLTVAI, *Lethality and centrality in protein networks*, Nature, 411 (2001), pp. 41–42.
- [212] H. JEONG, Z. NÉDA, AND A.-L. BARABÁSI, *Measuring preferential attachment in evolving networks*, Europhys. Lett., 61 (2003), pp. 567–572.
- [213] H. JEONG, B. TOMBOR, R. ALBERT, Z. N. OLTVAI, AND A.-L. BARABÁSI, *The large-scale organization of metabolic networks*, Nature, 407 (2000), pp. 651–654.
- [214] S. JESPERSEN AND A. BLUMEN, *Small-world networks: Links with long-tailed distributions*, Phys. Rev. E, 62 (2000), pp. 6270–6274.
- [215] S. JESPERSEN, I. M. SOKOLOV, AND A. BLUMEN, *Relaxation properties of small-world networks*, Phys. Rev. E, 62 (2000), pp. 4405–4408.

- [216] E. M. JIN, M. GIRVAN, AND M. E. J. NEWMAN, *The structure of growing social networks*, Phys. Rev. E, 64 (2001), art. no. 046132.
- [217] J. H. JONES AND M. S. HANDCOCK, *An Assessment of Preferential Attachment as a Mechanism for Human Sexual Network Formation*, preprint, University of Washington, Seattle, 2003.
- [218] P. JORDANO, J. BASCOMPTE, AND J. M. OLESEN, *Invariant properties in coevolutionary networks of plant-animal interactions*, Ecology Lett., 6 (2003), pp. 69–81.
- [219] J. JOST AND M. P. JOY, *Evolving networks with distance preferences*, Phys. Rev. E, 66 (2002), art. no. 036126.
- [220] V. K. KALAPALA, V. SANWALANI, AND C. MOORE, *The Structure of the United States Road Network*, preprint, University of New Mexico, Albuquerque, 2003.
- [221] F. KARINTHY, *Chains*, in Everything is Different, Budapest, 1929.
- [222] M. KAROŃSKI, *A review of random graphs*, J. Graph Theory, 6 (1982), pp. 349–389.
- [223] S. A. KAUFFMAN, *Metabolic stability and epigenesis in randomly connected nets*, J. Theor. Bio., 22 (1969), pp. 437–467.
- [224] S. A. KAUFFMAN, *Gene regulation networks: A theory for their structure and global behavior*, in Current Topics in Developmental Biology 6, A. Moscana and A. Monroy, eds., Academic Press, New York, 1971, pp. 145–182.
- [225] S. A. KAUFFMAN, *The Origins of Order*, Oxford University Press, Oxford, 1993.
- [226] H. KAUTZ, B. SELMAN, AND M. SHAH, *ReferralWeb: Combining social networks and collaborative filtering*, Comm. ACM, 40 (1997), pp. 63–65.
- [227] M. J. KEELING, *The effects of local spatial structure on epidemiological invasion*, Proc. Roy. Soc. London Ser. B, 266 (1999), pp. 859–867.
- [228] J. O. KEPHART AND S. R. WHITE, *Directed-graph epidemiological models of computer viruses*, in Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy, IEEE Computer Society, Los Alamitos, CA, 1991, pp. 343–359.
- [229] P. D. KILLWORTH AND H. R. BERNARD, *The reverse small world experiment*, Social Networks, 1 (1978), pp. 159–192.
- [230] 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 on small-world networks*, Phys. Rev. E, 66 (2002), art. no. 021907.
- [231] 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.
- [232] 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.
- [233] O. KINOCHI, A. S. MARTINEZ, G. F. LIMA, G. M. LOURENÇO, AND S. RISAU-GUSMAN, *Deterministic walks in random networks: An application to thesaurus graphs*, Phys. A, 315 (2002), pp. 665–676.
- [234] A. KLECZKOWSKI AND B. T. GRENFELL, *Mean-field-type equations for spread of epidemics: The “small world” model*, Phys. A, 274 (1999), pp. 355–360.
- [235] J. KLEINBERG AND S. LAWRENCE, *The structure of the Web*, Science, 294 (2001), pp. 1849–1850.
- [236] J. M. KLEINBERG, *Authoritative sources in a hyperlinked environment*, J. ACM, 46 (1999), pp. 604–632.
- [237] J. M. KLEINBERG, *Navigation in a small world*, Nature, 406 (2000), p. 845.
- [238] J. M. KLEINBERG, *The small-world phenomenon: An algorithmic perspective*, in Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, Association of Computing Machinery, New York, 2000, pp. 163–170.
- [239] J. M. KLEINBERG, *Small world phenomena and the dynamics of information*, in Proceedings of the 2001 Neural Information Processing Systems Conference, T. G. Dietterich, S. Becker, and Z. Ghahramani, eds., MIT Press, Cambridge, MA, 2002.
- [240] 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, Lecture Notes in Comput. Sci. 1627, Springer-Verlag, Berlin, 1999, pp. 1–18.
- [241] K. KLEMM AND V. M. EGUILUZ, *Highly clustered scale-free networks*, Phys. Rev. E, 65 (2002), art. no. 036123.
- [242] 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.
- [243] D. E. KNUTH, *The Stanford GraphBase: A Platform for Combinatorial Computing*, Addison-Wesley, Reading, MA, 1993.

- [244] P. L. KRAPIVSKY AND S. REDNER, *Organization of growing random networks*, Phys. Rev. E, 63 (2001), art. no. 066123.
- [245] P. L. KRAPIVSKY AND S. REDNER, *Finiteness and fluctuations in growing networks*, J. Phys. A, 35 (2002), pp. 9517–9534.
- [246] P. L. KRAPIVSKY AND S. REDNER, *A statistical physics perspective on Web growth*, Computer Networks, 39 (2002), pp. 261–276.
- [247] P. L. KRAPIVSKY AND S. REDNER, *Rate equation approach for growing networks*, in Proceedings of the 18th Sitges Conference on Statistical Mechanics, R. Pastor-Satorras and J. Rubi, eds., Lecture Notes in Phys., Springer-Verlag, Berlin, 2003.
- [248] P. L. KRAPIVSKY, S. REDNER, AND F. LEYVRAZ, *Connectivity of growing random networks*, Phys. Rev. Lett., 85 (2000), pp. 4629–4632.
- [249] P. L. KRAPIVSKY, G. J. RODGERS, AND S. REDNER, *Degree distributions of growing networks*, Phys. Rev. Lett., 86 (2001), pp. 5401–5404.
- [250] M. KRETZSCHMAR, Y. T. H. P. VAN DUYNHOVEN, AND A. J. SEVERIJNEN, *Modeling prevention strategies for gonorrhea and chlamydia using stochastic network simulations*, Amer. J. Epidemiol., 114 (1996), pp. 306–317.
- [251] R. V. KULKARNI, E. ALMAAS, AND D. STROUD, *Evolutionary Dynamics in the Bak-Sneppen Model on Small-World Networks*, Preprint 9908216 (1999); available from <http://arxiv.org/abs/cond-mat/>.
- [252] R. V. KULKARNI, E. ALMAAS, AND D. STROUD, *Exact results and scaling properties of small-world networks*, Phys. Rev. E, 61 (2000), pp. 4268–4271.
- [253] 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, Institute of Electrical and Electronics Engineers, New York, 2000, pp. 57–65.
- [254] M. KUPERMAN AND G. ABRAMSON, *Small world effect in an epidemiological model*, Phys. Rev. Lett., 86 (2001), pp. 2909–2912.
- [255] M. KUPERMAN AND D. H. ZANETTE, *Stochastic resonance in a model of opinion formation on small world networks*, Eur. Phys. J. B, 26 (2002), pp. 387–391.
- [256] L. F. LAGO-FERNÁNDEZ, R. HUERTA, F. CORBACHO, AND J. A. SIGÜENZA, *Fast response and temporal coherent oscillations in small-world networks*, Phys. Rev. Lett., 84 (2000), pp. 2758–2761.
- [257] J. LAHTINEN, J. KERTÉSZ, AND K. KASKI, *Scaling of random spreading in small world networks*, Phys. Rev. E, 64 (2001), art. no. 057105.
- [258] J. LAHTINEN, J. KERTÉSZ, AND K. KASKI, *Random spreading phenomena in annealed small world networks*, Phys. A, 311 (2002), pp. 571–580.
- [259] V. LATORA AND M. MARCHIORI, *Efficient behavior of small-world networks*, Phys. Rev. Lett., 87 (2001), art. no. 198701.
- [260] V. LATORA AND M. MARCHIORI, *Economic Small-World Behavior in Weighted Networks*, Preprint 0204089 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [261] V. LATORA AND M. MARCHIORI, *Is the Boston subway a small-world network?*, Phys. A, 314 (2002), pp. 109–113.
- [262] S. LAWRENCE AND C. L. GILES, *Accessibility of information on the web*, Nature, 400 (1999), pp. 107–109.
- [263] 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.
- [264] F. LILJEROS, C. R. EDLING, AND L. A. N. AMARAL, *Sexual networks: Implication for the transmission of sexually transmitted infection*, Microbes and Infections (to appear).
- [265] F. LILJEROS, C. R. EDLING, L. A. N. AMARAL, H. E. STANLEY, AND Y. ÅBERG, *The web of human sexual contacts*, Nature, 411 (2001), pp. 907–908.
- [266] A. L. LLOYD AND R. M. MAY, *How viruses spread among computers and people*, Science, 292 (2001), pp. 1316–1317.
- [267] T. ŁUCZAK, *Sparse random graphs with a given degree sequence*, in Proceedings of the Symposium on Random Graphs, Poznań 1989, A. M. Frieze and T. Łuczak, eds., John Wiley, New York, 1992, pp. 165–182.
- [268] P. MARIOLIS, *Interlocking directorates and control of corporations: The theory of bank control*, Social Sci. Quart., 56 (1975), pp. 425–439.
- [269] A. MARITAN, A. RINALDO, R. RIGON, A. GIACOMETTI, AND I. RODRÍGUEZ-ITURBE, *Scaling laws for river networks*, Phys. Rev. E, 53 (1996), pp. 1510–1515.
- [270] P. V. MARSDEN, *Network data and measurement*, Ann. Rev. Sociology, 16 (1990), pp. 435–463.
- [271] N. D. MARTINEZ, *Artifacts or attributes? Effects of resolution on the Little Rock Lake food web*, Ecological Monographs, 61 (1991), pp. 367–392.

- [272] N. D. MARTINEZ, *Constant connectance in community food webs*, Amer. Naturalist, 139 (1992), pp. 1208–1218.
- [273] S. MASLOV AND K. SNEPPEN, *Specificity and stability in topology of protein networks*, Science, 296 (2002), pp. 910–913.
- [274] S. MASLOV, K. SNEPPEN, AND A. ZALIZNYAK, *Pattern Detection in Complex Networks: Correlation Profile of the Internet*, Preprint 0205379 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [275] R. M. MAY AND R. M. ANDERSON, *The transmission dynamics of human immunodeficiency virus (HIV)*, Philos. Trans. Roy. Soc. London Ser. B, 321 (1988), pp. 565–607.
- [276] R. M. MAY AND A. L. LLOYD, *Infection dynamics on scale-free networks*, Phys. Rev. E, 64 (2001), art. no. 066112.
- [277] R. MEESTER AND R. ROY, *Continuum Percolation*, Cambridge University Press, Cambridge, UK, 1996.
- [278] G. MELIN AND O. PERSSON, *Studying research collaboration using co-authorships*, Scientometrics, 36 (1996), pp. 363–377.
- [279] F. MENCZER AND R. K. BELEW, *Adaptive retrieval agents: Internalizing local context and scaling up to the Web*, Machine Learning, 39 (2000), pp. 203–242.
- [280] F. MENCZER, G. PANT, M. RUIZ, AND P. SRINIVASAN, *Evaluating topic-driven Web crawlers*, in Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, D. H. Kraft, W. B. Croft, D. J. Harper, and J. Zobel, eds., Association of Computing Machinery, New York, 2001, pp. 241–249.
- [281] R. K. MERTON, *The Matthew effect in science*, Science, 159 (1968), pp. 56–63.
- [282] S. MILGRAM, *The small world problem*, Psych. Today, 2 (1967), pp. 60–67.
- [283] R. MILO, S. SHEN-ORR, S. ITZKOVITZ, N. KASHTAN, D. CHKLOVSKII, AND U. ALON, *Network motifs: Simple building blocks of complex networks*, Science, 298 (2002), pp. 824–827.
- [284] M. MITCHELL, *Introduction to Genetic Algorithms*, MIT Press, Cambridge, MA, 1996.
- [285] M. S. MIZRUCHI, *The American Corporate Network, 1904–1974*, Sage, Beverly Hills, CA, 1982.
- [286] M. MOLLOY AND B. REED, *A critical point for random graphs with a given degree sequence*, Random Structures Algorithms, 6 (1995), pp. 161–179.
- [287] M. MOLLOY AND B. REED, *The size of the giant component of a random graph with a given degree sequence*, Combin. Probab. Comput., 7 (1998), pp. 295–305.
- [288] R. MONASSON, *Diffusion, localization and dispersion relations on “small-world” lattices*, Eur. Phys. J. B, 12 (1999), pp. 555–567.
- [289] J. M. MONTÓYA AND R. V. SOLÉ, *Small world patterns in food webs*, J. Theor. Bio., 214 (2002), pp. 405–412.
- [290] J. MOODY, *Race, school integration, and friendship segregation in America*, Amer. J. Sociol., 107 (2001), pp. 679–716.
- [291] J. MOODY, *The structure of a social science collaboration network*, preprint, Department of Sociology, Ohio State University, Columbus, 2003.
- [292] C. MOORE AND M. E. J. NEWMAN, *Epidemics and percolation in small-world networks*, Phys. Rev. E, 61 (2000), pp. 5678–5682.
- [293] C. MOORE AND M. E. J. NEWMAN, *Exact solution of site and bond percolation on small-world networks*, Phys. Rev. E, 62 (2000), pp. 7059–7064.
- [294] A. A. MOREIRA, J. S. ANDRADE, JR., AND L. A. N. AMARAL, *Extremum Statistics in Scale-Free Network Models*, Preprint 0205411 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [295] J. L. MORENO, *Who Shall Survive?*, Beacon House, Beacon, NY, 1934.
- [296] 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.
- [297] Y. MORENO, R. PASTOR-SATORRAS, A. VÁZQUEZ, AND A. VESPIGNANI, *Critical Load and Congestion Instabilities in Scale-Free Networks*, Preprint 0209474 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [298] Y. MORENO, R. PASTOR-SATORRAS, AND A. VESPIGNANI, *Epidemic outbreaks in complex heterogeneous networks*, Eur. Phys. J. B, 26 (2002), pp. 521–529.
- [299] Y. MORENO AND A. VÁZQUEZ, *The Bak-Sneppen model on scale-free networks*, Europhys. Lett., 57 (2002), pp. 765–771.
- [300] Y. MORENO AND A. VÁZQUEZ, *Disease Spreading in Structured Scale-Free Networks*, Preprint 0210362 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [301] M. MORRIS, *Data driven network models for the spread of infectious disease*, in Epidemic Models: Their Structure and Relation to Data, D. Mollison, ed., Cambridge University Press, Cambridge, UK, 1995, pp. 302–322.

- [302] M. MORRIS, *Sexual networks and HIV*, AIDS 97: Year in Review, 11 (1997), pp. 209–216.
- [303] 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.
- [304] A. E. MOTTER AND Y.-C. LAI, *Cascade-based attacks on complex networks*, Phys. Rev. E, 66 (2002), art. no. 065102.
- [305] C. F. MOUKARZEL, *Spreading and shortest paths in systems with sparse long-range connections*, Phys. Rev. E, 60 (1999), pp. 6263–6266.
- [306] C. F. MOUKARZEL AND M. A. DE MENEZES, *Shortest paths on systems with power-law distributed long-range connections*, Phys. Rev. E, 65 (2002), art. no. 056709.
- [307] J. MÜLLER, B. SCHÖNFISCH, AND M. KIRKILIONIS, *Ring vaccination*, J. Math. Biol., 41 (2000), pp. 143–171.
- [308] M. E. J. NEWMAN, *Models of the small world*, J. Statist. Phys., 101 (2000), pp. 819–841.
- [309] M. E. J. NEWMAN, *Clustering and preferential attachment in growing networks*, Phys. Rev. E, 64 (2001), art. no. 025102.
- [310] M. E. J. NEWMAN, *Scientific collaboration networks: I. Network construction and fundamental results*, Phys. Rev. E, 64 (2001), art. no. 016131.
- [311] M. E. J. NEWMAN, *Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality*, Phys. Rev. E, 64 (2001), art. no. 016132.
- [312] M. E. J. NEWMAN, *The structure of scientific collaboration networks*, Proc. Natl. Acad. Sci. USA, 98 (2001), pp. 404–409.
- [313] M. E. J. NEWMAN, *Assortative mixing in networks*, Phys. Rev. Lett., 89 (2002), art. no. 208701.
- [314] M. E. J. NEWMAN, *Spread of epidemic disease on networks*, Phys. Rev. E, 66 (2002), art. no. 016128.
- [315] M. E. J. NEWMAN, *The structure and function of networks*, Comput. Phys. Comm., 147 (2002), pp. 40–45.
- [316] M. E. J. NEWMAN, *Ego-centered networks and the ripple effect*, Social Networks, 25 (2003), pp. 83–95.
- [317] M. E. J. NEWMAN, *Mixing patterns in networks*, Phys. Rev. E, 67 (2003), art. no. 026126.
- [318] M. E. J. NEWMAN, *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.
- [319] M. E. J. NEWMAN, A.-L. BARABÁSI, AND D. J. WATTS, *The Structure and Dynamics of Networks*, Princeton University Press, Princeton, NJ, 2003.
- [320] 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.
- [321] M. E. J. NEWMAN, C. MOORE, AND D. J. WATTS, *Mean-field solution of the small-world network model*, Phys. Rev. Lett., 84 (2000), pp. 3201–3204.
- [322] 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.
- [323] M. E. J. NEWMAN AND D. J. WATTS, *Renormalization group analysis of the small-world network model*, Phys. Lett. A, 263 (1999), pp. 341–346.
- [324] M. E. J. NEWMAN AND D. J. WATTS, *Scaling and percolation in the small-world network model*, Phys. Rev. E, 60 (1999), pp. 7332–7342.
- [325] M. OZANA, *Incipient spanning cluster on small-world networks*, Europhys. Lett., 55 (2001), pp. 762–766.
- [326] 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.
- [327] L. PAGE, S. BRIN, R. MOTWANI, AND T. WINOGRAD, *The Pagerank Citation Ranking: Bringing Order to the web*, technical report, Stanford University, Stanford, CA, 1998.
- [328] S. A. PANDIT AND R. E. AMRITKAR, *Random spread on the family of small-world networks*, Phys. Rev. E, 63 (2001), art. no. 041104.
- [329] R. PASTOR-SATORRAS AND J. RUBI, EDS., *Proceedings of the 18th Sitges Conference on Statistical Mechanics*, Lecture Notes in Phys., Springer-Verlag, Berlin, 2003.
- [330] R. PASTOR-SATORRAS, A. VÁZQUEZ, AND A. VESPIGNANI, *Dynamical and correlation properties of the Internet*, Phys. Rev. Lett., 87 (2001), art. no. 258701.
- [331] R. PASTOR-SATORRAS AND A. VESPIGNANI, *Epidemic dynamics and endemic states in complex networks*, Phys. Rev. E, 63 (2001), art. no. 066117.
- [332] R. PASTOR-SATORRAS AND A. VESPIGNANI, *Epidemic spreading in scale-free networks*, Phys. Rev. Lett., 86 (2001), pp. 3200–3203.
- [333] R. PASTOR-SATORRAS AND A. VESPIGNANI, *Epidemic dynamics in finite size scale-free networks*, Phys. Rev. E, 65 (2002), art. no. 035108.

- [334] R. PASTOR-SATORRAS AND A. VESPIGNANI, *Immunization of complex networks*, Phys. Rev. E, 65 (2002), art. no. 036104.
- [335] R. PASTOR-SATORRAS AND A. VESPIGNANI, *Epidemics and immunization in scale-free networks*, in Handbook of Graphs and Networks, S. Bornholdt and H. G. Schuster, eds., Wiley-VCH, Berlin, 2003.
- [336] A. PEKALSKI, *Ising model on a small world network*, Phys. Rev. E, 64 (2001), art. no. 057104.
- [337] D. M. PENNOCK, G. W. FLAKE, S. LAWRENCE, E. J. GLOVER, AND C. L. GILES, *Winners don't take all: Characterizing the competition for links on the web*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 5207–5211.
- [338] S. L. PIMM, *Food Webs*, 2nd ed., University of Chicago Press, Chicago, 2002.
- [339] J. PODANI, Z. N. OLTVAI, H. JEONG, B. TOMBOR, A.-L. BARABÁSI, AND E. SZATHMARY, *Comparable system-level organization of Archaea and Eukaryotes*, Nature Genetics, 29 (2001), pp. 54–56.
- [340] I. DE S. POOL AND M. KOCHEN, *Contacts and influence*, Social Networks, 1 (1978), pp. 1–48.
- [341] 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. i159–i163.
- [342] D. J. DE S. PRICE, *Networks of scientific papers*, Science, 149 (1965), pp. 510–515.
- [343] D. J. DE S. PRICE, *A general theory of bibliometric and other cumulative advantage processes*, J. Amer. Soc. Inform. Sci., 27 (1976), pp. 292–306.
- [344] A. RAMEZANPOUR, V. KARIMPOUR, AND A. MASHAGHI, *Generating Correlated Networks from Uncorrelated Ones*, Preprint 0212469 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [345] A. RAPOPORT, *Contribution to the theory of random and biased nets*, Bull. Math. Biophys., 19 (1957), pp. 257–277.
- [346] A. RAPOPORT, *Cycle distribution in random nets*, Bull. Math. Biophys., 10 (1968), pp. 145–157.
- [347] A. RAPOPORT AND W. J. HORVATH, *A study of a large sociogram*, Behavioral Sci., 6 (1961), pp. 279–291.
- [348] E. RAVASZ AND A.-L. BARABÁSI, *Hierarchical organization in complex networks*, Phys. Rev. E, 67 (2003), art. no. 026112.
- [349] E. RAVASZ, A. L. SOMERA, D. A. MONGRU, Z. OLTVAI, AND A.-L. BARABÁSI, *Hierarchical organization of modularity in metabolic networks*, Science, 297 (2002), pp. 1551–1555.
- [350] S. REDNER, *How popular is your paper? An empirical study of the citation distribution*, Eur. Phys. J. B, 4 (1998), pp. 131–134.
- [351] P. RESNICK AND H. R. VARIAN, *Recommender systems*, Comm. ACM, 40 (1997), pp. 56–58.
- [352] A. RINALDO, I. RODRÍGUEZ-ITURBE, AND R. RIGON, *Channel networks*, Ann. Rev. Earth and Planetary Sci., 26 (1998), pp. 289–327.
- [353] M. RIPEANU, I. FOSTER, AND A. IAMNITCHI, *Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design*, IEEE Internet Comput., 6 (2002), pp. 50–57.
- [354] G. J. RODGERS AND K. DARBY-DOWMAN, *Properties of a growing random directed network*, Eur. Phys. J. B, 23 (2001), pp. 267–271.
- [355] I. RODRÍGUEZ-ITURBE AND A. RINALDO, *Fractal River Basins: Chance and Self-Organization*, Cambridge University Press, Cambridge, UK, 1997.
- [356] F. J. ROETHLISBERGER AND W. J. DICKSON, *Management and the Worker*, Harvard University Press, Cambridge, MA, 1939.
- [357] R. ROTHENBERG, J. BALDWIN, R. TROTTER, AND S. MUTH, *The risk environment for HIV transmission: Results from the Atlanta and Flagstaff network studies*, J. Urban Health, 78 (2001), pp. 419–431.
- [358] A. F. ROZENFELD, R. COHEN, D. BEN-AVRAHAM, AND S. HAVLIN, *Scale-free networks on lattices*, Phys. Rev. Lett., 89 (2002), art. no. 218701.
- [359] 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.
- [360] A. SCALA, L. A. N. AMARAL, AND M. BARTHÉLÉMY, *Small-world networks and the conformation space of a short lattice polymer chain*, Europhys. Lett., 55 (2001), pp. 594–600.
- [361] N. SCHWARTZ, R. COHEN, D. BEN-AVRAHAM, A.-L. BARABÁSI, AND S. HAVLIN, *Percolation in directed scale-free networks*, Phys. Rev. E, 66 (2002), art. no. 015104.
- [362] J. SCOTT, *Social Network Analysis: A Handbook*, 2nd ed., Sage, London, 2000.
- [363] P. O. SEGLEN, *The skewness of science*, J. Amer. Soc. Inform. Sci., 43 (1992), pp. 628–638.

- [364] P. SEN AND B. K. CHAKRABARTI, *Small-world phenomena and the statistics of linear polymers*, J. Phys. A, 34 (2001), pp. 7749–7755.
- [365] P. SEN, S. DASGUPTA, A. CHATTERJEE, P. A. SREERAM, G. MUKHERJEE, AND S. S. MANNA, *Small-World Properties of the Indian Railway Network*, Preprint 0208535 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [366] 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, Association of Computing Machinery, New York, 1995, pp. 210–217.
- [367] 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.
- [368] M. SIGMAN AND G. A. CECCHI, *Global organization of the Wordnet lexicon*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 1742–1747.
- [369] H. A. SIMON, *On a class of skew distribution functions*, Biometrika, 42 (1955), pp. 425–440.
- [370] R. D. SMITH, *Instant Messaging as a Scale-Free Network*, Preprint 0206378 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [371] T. A. B. SNIJDERS, *Markov chain Monte Carlo estimation of exponential random graph models*, J. Social Structure, 2 (2002).
- [372] J. E. S. SOCOLAR AND S. A. KAUFFMAN, *Scaling in ordered and critical random Boolean networks*, Phys. Rev. Lett., 90 (2003), art. no. 068702.
- [373] B. SÖDERBERG, *General formalism for inhomogeneous random graphs*, Phys. Rev. E, 66 (2002), art. no. 066121.
- [374] R. V. SOLÉ AND J. M. MONTÓYA, *Complexity and fragility in ecological networks*, Proc. Roy. Soc. London Ser. B, 268 (2001), pp. 2039–2045.
- [375] R. V. SOLÉ AND R. PASTOR-SATORRAS, *Complex networks in genomics and proteomics*, in Handbook of Graphs and Networks, S. Bornholdt and H. G. Schuster, eds., Wiley-VCH, Berlin, 2003, pp. 145–167.
- [376] R. V. SOLÉ, R. PASTOR-SATORRAS, E. SMITH, AND T. B. KEPLER, *A model of large-scale proteome evolution*, Adv. in Complex Systems, 5 (2002), pp. 43–54.
- [377] R. SOLOMONOFF AND A. RAPOPORT, *Connectivity of random nets*, Bull. Math. Biophys., 13 (1951), pp. 107–117.
- [378] O. SPORNS, *Network analysis, complexity, and brain function*, Complexity, 8 (2002), pp. 56–60.
- [379] 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.
- [380] D. STAUFFER, *Monte Carlo simulations of Sznajd models*, J. Artificial Societies and Social Simulation, 5 (1) (2002); available online from <http://jasss.soc.surrey.ac.uk/5/1/4.html>.
- [381] D. STAUFFER, A. AHARONY, L. DA FONTOURA COSTA, AND J. ADLER, *Efficient Hopfield Pattern Recognition on a Scale-Free Neural Network*, Preprint 0212601 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [382] J. STELLING, S. KLAMT, K. BETTENBROCK, S. SCHUSTER, AND E. D. GILLES, *Metabolic network structure determines key aspects of functionality and regulation*, Nature, 420 (2002), pp. 190–193.
- [383] M. STEYVERS AND J. B. TENENBAUM, *The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model for Semantic Growth*, Preprint 0110012 (2001); available from <http://arxiv.org/abs/cond-mat/>.
- [384] D. STRAUSS, *On a general class of models for interaction*, SIAM Rev., 28 (1986), pp. 513–527.
- [385] S. H. STROGATZ, *Nonlinear Dynamics and Chaos*, Addison-Wesley, Reading, MA, 1994.
- [386] S. H. STROGATZ, *Exploring complex networks*, Nature, 410 (2001), pp. 268–276.
- [387] P. SVENSON, *From Néel to NPC: Colouring Small Worlds*, Preprint 0107015 (2001); available from <http://arxiv.org/abs/cs/>.
- [388] G. SZABÓ, M. ALAVA, AND J. KERTÉSZ, *Structural Transitions in Scale-Free Networks*, Preprint 0208551 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [389] K. SZNAJD-WERON AND J. SZNAJD, *Opinion evolution in closed community*, Internat. J. Modern Phys. C, 11 (2000), pp. 1157–1165.
- [390] B. TADIĆ, *Dynamics of directed graphs: The World-Wide Web*, Phys. A, 293 (2001), pp. 273–284.
- [391] B. TADIĆ, *Temporal fractal structures: Origin of power laws in the World-Wide Web*, Phys. A, 314 (2002), pp. 278–283.
- [392] J. TRAVERS AND S. MILGRAM, *An experimental study of the small world problem*, Sociometry, 32 (1969), pp. 425–443.



- [393] P. UETZ, L. GIOT, G. CAGNEY, T. A. MANSFIELD, R. S. JUDSON, J. R. KNIGHT, D. LOCKSHON, V. NARAYAN, M. SRINIVASAN, P. POCHART, A. QURESHI-EMILI, Y. LI, B. GODWIN, D. CONOVER, T. KALBFLEISCH, G. VIJAYADAMODAR, M. YANG, M. JOHNSTON, S. FIELDS, AND J. M. ROTHBERG, *A comprehensive analysis of protein-protein interactions in saccharomyces cerevisiae*, *Nature*, 403 (2000), pp. 623–627.
- [394] S. VALVERDE, R. F. CANCHO, AND R. V. SOLÉ, *Scale-free networks from optimal design*, *Europhys. Lett.*, 60 (2002), pp. 512–517.
- [395] A. VÁZQUEZ, *Statistics of Citation Networks*, Preprint 0105031 (2001); available from <http://arxiv.org/abs/cond-mat/>.
- [396] A. VÁZQUEZ, *Growing Networks with Local Rules: Preferential Attachment, Clustering Hierarchy and Degree Correlations*, Preprint 0211528 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [397] A. VÁZQUEZ, M. BOGUÑA, Y. MORENO, R. PASTOR-SATORRAS, AND A. VESPIGNANI, *Topology and Correlations in Structured Scale-Free Networks*, Preprint 0209183 (2002); available from <http://arxiv.org/abs/cond-mat/>.
- [398] A. VÁZQUEZ, A. FLAMMINI, A. MARITAN, AND A. VESPIGNANI, *Modeling of protein interaction networks*, *Complexus*, 1 (2003), pp. 38–44.
- [399] A. VÁZQUEZ AND Y. MORENO, *Resilience to damage of graphs with degree correlations*, *Phys. Rev. E*, 67 (2003), art. no. 015101.
- [400] 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.
- [401] A. VÁZQUEZ AND M. WEIGT, *Computational complexity arising from degree correlations in networks*, *Phys. Rev. E*, 67 (2003), art. no. 027101.
- [402] F. VAZQUEZ, P. L. KRAPIVSKY, AND S. REDNER, *Constrained opinion dynamics: Freezing and slow evolution*, *J. Phys. A*, 36 (2003), pp. L61–L68.
- [403] A. WAGNER, *The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes*, *Mol. Biol. Evol.*, 18 (2001), pp. 1283–1292.
- [404] A. WAGNER AND D. FELL, *The small world inside large metabolic networks*, *Proc. Roy. Soc. London Ser. B*, 268 (2001), pp. 1803–1810.
- [405] T. WALSH, *Search in a small world*, in *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, T. Dean, ed., Morgan-Kaufmann, San Francisco, CA, 1999.
- [406] B.-Y. WANG AND F. ZHANG, *Exact counting of  $(0,1)$  matrices with given row and column sums*, *Discrete Math.*, 187 (1998), pp. 211–220.
- [407] C. P. WARREN, L. M. SANDER, AND I. SOKOLOV, *Geography in a scale-free network model*, *Phys. Rev. E*, 66 (2002), art. no. 056105.
- [408] S. WASSERMAN AND K. FAUST, *Social Network Analysis*, Cambridge University Press, Cambridge, UK, 1994.
- [409] S. WASSERMAN AND P. PATTISON, *Logit models and logistic regressions for social networks: I. An introduction to Markov random graphs and  $p^*$* , *Psychometrika*, 61 (1996), pp. 401–426.
- [410] D. J. WATTS, *Networks, dynamics, and the small world phenomenon*, *Amer. J. Sociol.*, 105 (1999), pp. 493–592.
- [411] D. J. WATTS, *Small Worlds*, Princeton University Press, Princeton, NJ, 1999.
- [412] D. J. WATTS, *A simple model of global cascades on random networks*, *Proc. Natl. Acad. Sci. USA*, 99 (2002), pp. 5766–5771.
- [413] D. J. WATTS, *Six Degrees: The Science of a Connected Age*, Norton, New York, 2003.
- [414] D. J. WATTS, P. S. DODDS, AND M. E. J. NEWMAN, *Identity and search in social networks*, *Science*, 296 (2002), pp. 1302–1305.
- [415] D. J. WATTS AND S. H. STROGATZ, *Collective dynamics of “small-world” networks*, *Nature*, 393 (1998), pp. 440–442.
- [416] G. B. WEST, J. H. BROWN, AND B. J. ENQUIST, *A general model for the origin of allometric scaling laws in biology*, *Science*, 276 (1997), pp. 122–126.
- [417] 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.
- [418] H. C. WHITE, S. A. BOORMAN, AND R. L. BREIGER, *Social structure from multiple networks: I. Blockmodels of roles and positions*, *Amer. J. Sociol.*, 81 (1976), pp. 730–779.
- [419] H. D. WHITE, B. WELLMAN, AND N. NAZER, *Does citation reflect social structure? Longitudinal evidence from the “GloboNet” interdisciplinary research group*, preprint, University of Toronto, Toronto, ON, Canada, 2003.
- [420] 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. London*, 314 (1986), pp. 1–340.

- [421] D. WILKINSON AND B. A. HUBERMAN, *A Method for Finding Communities of Related Genes*, preprint, Stanford University, Stanford, CA, 2002.
- [422] 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.
- [423] A. T. WINFREE, *The Geometry of Biological Time*, 2nd ed., Springer-Verlag, New York, 2000.
- [424] N. C. WORMALD, *The asymptotic connectivity of labelled regular graphs*, J. Combin. Theory Ser. B, 31 (1981), pp. 156–167.
- [425] S. H. YOOK, H. JEONG, AND A.-L. BARABASI, *Modeling the internet's large-scale topology*, Proc. Natl. Acad. Sci. USA, 99 (2001), pp. 13382–13386.
- [426] H. P. YOUNG, *The diffusion of innovations in social networks*, in The Economy as an Evolving Complex System, Vol. 3, L. E. Blume and S. N. Durlauf, eds., Oxford University Press, Oxford, 2003.
- [427] D. H. ZANETTE, *Critical behavior of propagation on small-world networks*, Phys. Rev. E, 64 (2001), art. no. 050901.
- [428] N. ZEKRI AND J. P. CLERC, *Statistical and dynamical study of disease propagation in a small world network*, Phys. Rev. E, 64 (2001), art. no. 056115.
- [429] J.-Y. ZHU AND H. ZHU, *Introducing Small-World Network Effect to Critical Dynamics*, Preprint 0212542 (2002); available from <http://arxiv.org/abs/cond-mat/>.