

# THE "NEW" SCIENCE OF NETWORKS

---

Duncan J. Watts

*Department of Sociology, Columbia University, New York, NY 10027; Santa Fe Institute, Santa Fe, New Mexico 97501; email: djw24@columbia.edu*

**Key Words** graph theory, mathematical models, network data, dynamical systems

■ **Abstract** In recent years, the analysis and modeling of networks, and also networked dynamical systems, have been the subject of considerable interdisciplinary interest, yielding several hundred papers in physics, mathematics, computer science, biology, economics, and sociology journals (Newman 2003c), as well as a number of books (Barabasi 2002, Buchanan 2002, Watts 2003). Here I review the major findings of this emerging field and discuss briefly their relationship with previous work in the social and mathematical sciences.

## INTRODUCTION

Building on a long tradition of network analysis in sociology and anthropology (Degenne & Forse 1994, Scott 2000, Wasserman & Faust 1994) and an even longer history of graph theory in discrete mathematics (Ahuja et al. 1993, Bollobas 1998, West 1996), the study of networks and networked systems has exploded across the academic spectrum in the past five years. Spurred by the rapidly growing availability of cheap yet powerful computers and large-scale electronic datasets, researchers from the mathematical, biological, and social sciences have made substantial progress on a number of previously intractable problems, reformulating old ideas, introducing new techniques, and uncovering connections between what had seemed to be quite different problems. The result has been called the "new science of networks" (Barabasi 2002, Buchanan 2002, Watts 2003)—a label that may strike many sociologists as misleading, given the familiarity (to social network analysts) of many of its central ideas. Nevertheless, the label does capture the sense of excitement surrounding what is unquestionably a fast developing field—new papers are appearing almost daily—and also the unprecedented degree of synthesis that this excitement has generated across the many disciplines in which network-related problems arise.

In reviewing the main results of this field, I have tried to strike a balance between the historical progression of the ideas and their logical order. To this end, the first section describes the main network modeling approaches, regarding local structure, global connectivity, searchability, and highly skewed degree distributions.

I then summarize related empirical studies and discuss some of the issues associated with interpreting empirical data. Finally, I review preliminary progress on applications of network models to dynamical processes like disease spreading and information exchange. Throughout the discussion, I focus on the motivation and interpretation of ideas and results, referring the reader elsewhere (Newman 2003c) for the mathematical details. And although I concentrate on developments that seem most relevant to social networks, it is worth bearing in mind that many of the same results have been applied to, or motivated by, problems in other disciplines.

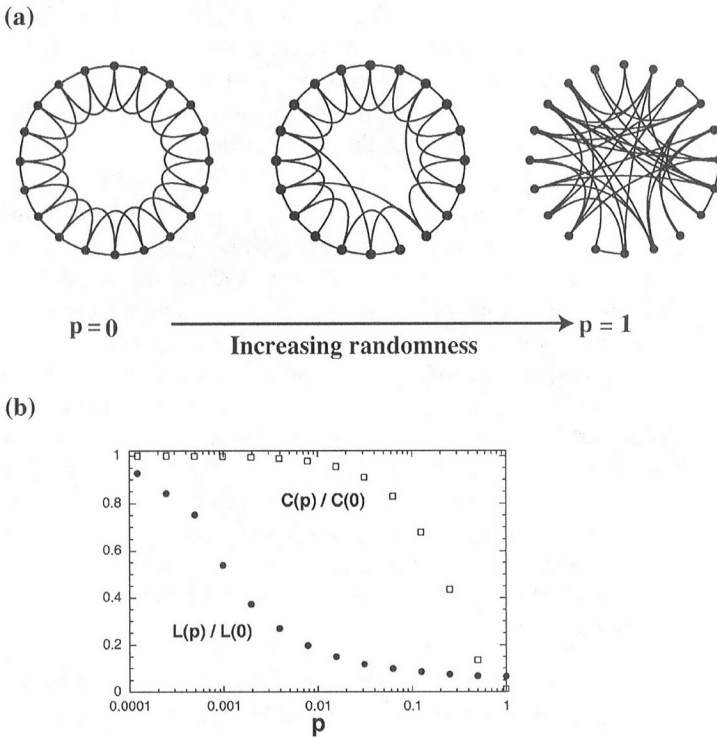
## MODELS OF NETWORK STRUCTURE

### Small-World Networks

An early contribution to the new science of networks, and one that captures some of its major themes, was that by Watts & Strogatz (1998), in which the authors made several related but distinct points:

1. Real-world networks are neither completely ordered nor completely random, but rather exhibit important properties of both.
2. Some properties of these networks can be embodied by simple mathematical models that interpolate between order and randomness. In Watts & Strogatz's specific example, "order" was represented by a uniform one-dimensional lattice,<sup>1</sup> where each node was connected to its  $k$  nearest neighbors on the lattice, and "randomness" was characterized by a tunable parameter  $p$  that specified the fraction of randomly rewired links (see Figure 1a).
3. These properties can be quantified with simple statistics: for example, the clustering coefficient  $C$  of a network is a measure of local density, and the average shortest path length  $L$  is a global measure of separation. In the case of the proposed model,  $L$  and  $C$  can be measured as a function of  $p$ .
4. When  $p = 0$  (completely ordered), the network is "large" [ $L(0) \sim N/2k$ ] and "highly clustered" [ $C(0) \sim 3/4$ ], and when  $p = 1$  (completely random), it is "small" [ $L(1) \sim \ln(N)/\ln(k)$ ] and "poorly clustered" [ $C(1) \sim k/N$ ], suggesting that path lengths are short only when clustering is low.
5. However, the model exhibits a broad region of  $p$  values in which  $C(p)$  is high relative to its random limit  $C(1)$ , yet  $L(p)$  is, roughly speaking, as "small" as possible (see Figure 1b). Watts & Strogatz (1998) coined the term

<sup>1</sup>Throughout this paper, the term "lattice" is used in the physicist's sense of a *crystal lattice* (roughly, a low-dimensional grid), not the mathematician's sense of a partial ordering on a set of elements [as in, for example, *Galois Lattices* that have been used to represent affiliation networks (Freeman & White 1993)].



**Figure 1** (a) Schematic of the Watts-Strogatz model. (b) Normalized average shortest path length  $L$  and clustering coefficient  $C$  as a function of the random rewiring parameter  $p$  for the Watts-Strogatz model with  $N = 1000$ , and  $\langle k \rangle = 10$ .

*small-world networks*<sup>2</sup> to refer to networks in this class, in reference to the early work of Pool & Kochen (1978), and subsequent experiments of Milgram and colleagues (Korte & Milgram 1970; Milgram 1967; Travers & Milgram 1969).

6. Because the conditions required for any network to belong to the small-world class (some nontrivial local order, combined with just a small fraction of long-range, random shortcuts) were relatively weak, Watts & Strogatz (1998) predicted that many real-world networks—whether social networks

<sup>2</sup>Watts & Strogatz were not, in fact, the first to use the term small-world networks: that distinction belongs to Eugene Garfield (1979), who used it in a review article on the topic of small-world research. The term does not appear to have caught on, however, until Watts & Strogatz defined it to refer to networks with high local clustering and short global path lengths.

or otherwise—ought to be small-world networks. They then checked this prediction by considering three network datasets—the affiliation network of movie actors, the power transmission grid of the western United States, and the neural network of the nematode *Caenorhabditis elegans*—and found that all three examples satisfied the small-world criteria.

7. Finally, the authors claimed that the structure of a network can have dramatic implications for the collective dynamics of a system, whose connectivity the network represents. In particular, they claimed that large changes in dynamical behavior could be driven by even subtle modifications to the network structure—modifications that may be imperceptible to actors with only local knowledge of the network. They supported this claim with the example of a simple disease-spreading model, demonstrating that both the extent of an epidemic, and also the time taken for it to spread, were sensitively dependent on the level of randomness in the network—only a small amount of randomness was required for large, rapidly spreading epidemics to occur. In related work, Watts (1999) found equally dramatic effects of network structure on dynamics for the emergence of cooperation in a multiplayer game of Prisoner's Dilemma, the synchronization of coupled oscillators, and the ability of locally connected cellular automata to “solve” certain global computation problems.

Watts & Strogatz's observations, although new to the mathematical literature of coupled dynamical systems, and also to the condensed matter physics community, were not, in spirit at least, new to mathematical sociologists. Parameterized models of locally ordered but otherwise random networks, for example, had been the object of social network analysts' attention for decades, at first according to Rapoport's notion of random-biased nets (Rapoport 1951, 1957), and more recently according to what are generically referred to as exponential random graph models, developed by Holland & Leinhardt (1981), Frank & Strauss (1986), and others (Anderson et al. 1999, Pattison & Wasserman 1999, Robins et al. 1999, Wasserman & Pattison 1996). Furthermore, Granovetter (1973), following Rapoport, had introduced the distinction between “strong” and “weak” ties, where the former could be construed as arising from local ordering principles like homophily (Lazarsfeld & Merton 1954) and triadic closure (Rapoport 1953a), and the latter from occasional random contacts, which Granovetter called local bridges, but which are clearly analogous to Watts & Strogatz's shortcuts. And finally, Rapoport (1953a,b) and others (Kretschmar & Morris 1996) had long been concerned with the effect of socio-structural biases in otherwise random networks on the spread of information or infectious disease.

However, a distinguishing feature of Watts & Strogatz (1998), and the one that arguably generated much of the subsequent interest in the physics community, was their identification of a universal class of networks; that is, a family of networks that share certain aggregate properties (in this case small  $L$  and large  $C$ ) regardless of

many of their individual details. This finding had two implications, both of which fit naturally into a physicist's worldview: (a) that at least some interesting features of even very complex networks could be captured by extremely simple models, and (b) that metrics and models devised to address social network problems might be usefully applied in other disciplines as well. Condensed matter physicists have subsequently analyzed what is now called the Watts-Strogatz model, or variants of it, in great detail (Barrat & Weigt 2000; Barthelemy & Amaral 1999; Dorogovtsev & Mendes 2000; Kulkarni et al. 2000; Newman & Watts 1999a,b; Newman et al. 2000). Specifically, it has been shown that whereas clustering  $C$  is a function of the fraction of random shortcuts  $p$  (Barrat & Weigt 2000), the average shortest path length  $L$  is a function of their absolute number  $Nkp$  (Newman et al. 2000). Although this difference may appear subtle, its implication is that in a sufficiently large network (i.e., as  $N$  is increased without bound) small-world behavior will result whenever the fraction of random shortcuts is greater than zero—that is, for any randomness at all. Furthermore, this result does not depend on the presence of a one-dimensional lattice—it is true for any connected substrate—hence the conditions required for small-world behavior to pertain are even weaker than at first thought. A particularly striking result in this regard is that for a slightly modified version of the ring lattice model (in which random shortcuts are added rather than rewired), Newman et al. (2000) showed that the addition of only five random shortcuts would halve the average path length  $L$  of the network, regardless of its size.<sup>3</sup>

## The Algorithmic Small-World Problem

Unfortunately, the Watts-Strogatz model suffers from some serious problems that render it unsuitable as a model of social networks. One such problem was pointed out by Kleinberg (2000a,b), who observed that the small-world experiments of Milgram (Korte & Milgram 1970, Milgram 1967, Travers & Milgram 1969) demonstrated not only that short paths existed between randomly chosen individuals in a large population, but also that the individuals in question could locate these paths using only their local information about the network. Social networks, in other words, are not only small; they are also searchable. In a remarkable piece of analysis, Kleinberg proved that while models like that of Watts & Strogatz, in which links are rewired uniformly at random, exhibited short global path lengths, they could not exhibit searchability. He then proposed a class of *generalized small-world networks* comprising (a) an underlying  $d$ -dimensional lattice, and (b) random links superposed on the lattice, where the probability  $p_{ij}$  of two nodes being connected randomly is

$$p_{ij} \propto r_{ij}^{-\gamma}, \quad 1.$$

<sup>3</sup>Strictly speaking, this result applies only in the limit of large network size  $N$ .

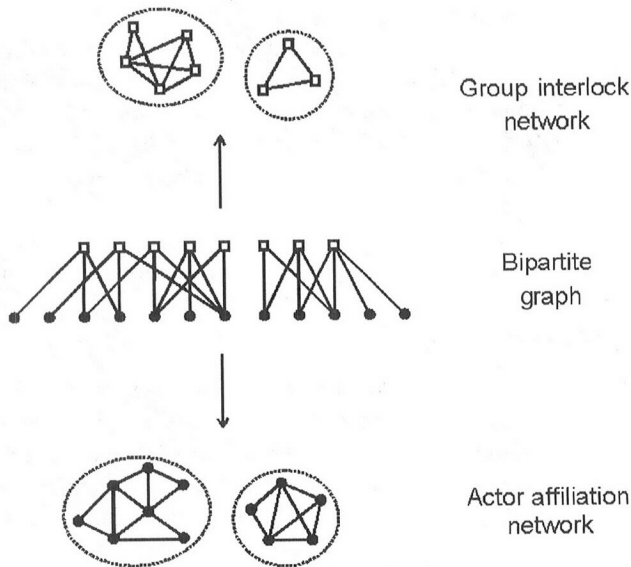
and  $r_{ij}$  is the lattice distance from  $i$  to  $j$ . When  $\gamma = 0$ , random edges are formed with uniform probability (equivalent to the original Watts-Strogatz model), and when  $\gamma$  is sufficiently large, only local ties are possible. Clearly in neither limit is the network searchable—in the former case because short paths cannot be found (Kleinberg's earlier result), and in the latter case because no short paths exist. In fact, Kleinberg proved that only when  $\gamma = d$ , the dimension of the underlying lattice, would the network be searchable in the sense that the number of steps required to forward a message from a randomly chosen starter to a randomly chosen target is "short"<sup>4</sup> (Kleinberg 2000a). More generally, what Kleinberg showed is that network structure is important not only locally, i.e., because an individual's neighborhood provides him with information and resources, but also globally in that it enables him to navigate when searching for information or resources outside his neighborhood.

### Affiliation Networks

Another problem with the Watts-Strogatz model, and also with Kleinberg's, is that both models rely on the presence of some underlying lattice substrate to (a) guarantee global connectivity and (b) provide a distance metric that is independent of the network distance. Clearly, social networks are not built on a lattice substrate, so the question arises whether one can account for clustering, short path lengths, and searchability in a less artificial manner. One alternative method of generating networks with small-world properties that uses more plausible assumptions, but is still parsimonious, is based on a bipartite (or "two mode") affiliation network (Wasserman & Faust 1994) comprising two classes of nodes:  $N$  actors each of whom belongs to one or more of  $M$  groups. As shown schematically in Figure 2, "single mode" (unipartite) networks can then be generated by projecting the bipartite network onto either the set of actors (the "actor affiliation network") or the set of groups (the "group interlock network"), where the standard projection assumes that actors/groups will be connected if they share at least one group/actor.

Following this construction, a simple affiliation-based network model was proposed by Newman et al. (2001, 2002), who assumed that actors choose groups uniformly at random. Using the method of generating functions (Wilf 1994), they derived a number of results for random bipartite graphs, including the conditions required for the formation of a giant connected component, as well as the degree distribution, approximate average shortest path length, and clustering coefficient of the actor affiliation and group interlock networks. Generalizing their approach to allow for arbitrary degree distributions, Newman et al. were able to replicate a number of features of real-world affiliation networks, such as Davis et al.'s (2003) boards of directors data, and the collaboration networks of movie actors (Watts 1999) and scientists (Newman 2001). One result of this work was that the

<sup>4</sup>Specifically, Kleinberg showed that whenever  $\gamma \neq d$  the expected number of steps required to complete a search increased as a polynomial function of the system size, but when  $\gamma = d$ , it increased as a polynomial of the logarithm of the system size.



**Figure 2** Bipartite (two mode) affiliation network (*center*), defined in terms of  $N$  actors belonging to  $M$  groups, along with unipartite (single mode) projections onto the set of groups (*top*) and actors (*bottom*). A connection in the single mode network occurs whenever two actors/groups share at least one group/actor respectively.

unipartite projection of a random bipartite network can account for much of the observed clustering in empirical affiliation network data. Because two actors are by definition connected whenever they share a group, the single mode projections exhibit many overlapping cliques that contribute to high local clustering, even when individuals choose groups at random. Furthermore, random bipartite networks almost always have short global path lengths; thus both actor affiliation and group interlock networks will inevitably fall in the small-world class.

Subsequently, the basic affiliation model has been generalized in two ways. First, Newman & Park (2003) have derived a number of results for random affiliation networks in which ties between actors in the single mode projection exist only with a probability that is proportional to the number of shared groups and inversely proportional to the size of each group. And second, Watts et al. (2002) have proposed a model of *generalized affiliation networks* in which distance between groups is defined according to some number of social dimensions (e.g., geography and occupation), and individuals are characterized by the coordinates of the groups to which they belong. Ties between individuals are then allowed to form with a probability that depends on the distance between the corresponding groups and a tunable homophily (Lazarsfeld & Merton 1954) parameter that biases interactions toward or away from similar nodes. They find that when groups are classified according to more than one social dimension and when interactions are homophilous

within any given dimension, then not only are the resulting networks highly clustered with short global path lengths, but they are also searchable in a sense similar to that proposed by Kleinberg (Watts et al. 2002). A recent Web-based experiment (Dodds et al. 2003a) replicating the small-world method of Milgram (Korte & Milgram 1970, Milgram 1967, Travers & Milgram 1969), but using e-mail instead of paper mail, appears to support the claim of the generalized affiliation model—and also earlier work (Bernard et al. 1988, Killworth & Bernard 1978)—that chains of individuals can locate remote targets given only a few pieces of identifying information such as geographical location and occupation.

## Scale-Free Networks

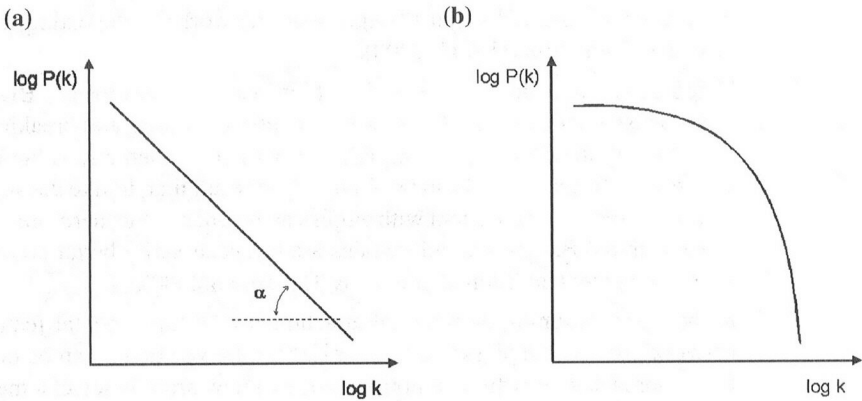
A separate development in the recent literature on networks has been the growing realization that in many real-world networks, the distribution of the number of network neighbors—the degree distribution—is typically right-skewed with a “heavy tail,” meaning that a majority of nodes have less-than-average degree and that a small fraction of hubs are many times better connected than average. This qualitative description can be satisfied by several mathematical functions, but a particularly popular one in the current literature is a power law (Barabasi & Albert 1999), which has the asymptotic form

$$p(k) \sim k^{-\alpha}. \quad 2.$$

In other words, the probability of a randomly chosen node having degree  $k$  decays like a power of  $k$ , where the exponent  $\alpha$ , typically measured in the range  $2 < \alpha < 3$ , determines the rate of decay (smaller  $\alpha$  implies slower decay, hence a more skewed distribution). A distinguishing feature of power-law distributions is that when plotted on a double logarithmic scale, a power law appears as a straight line with negative slope  $\alpha$ . As displayed schematically in Figure 3a, the straight line form of a power law contrasts with a normal distribution (Figure 3b), which curves sharply on a log-log plot, such that the probability of a node having a degree greater than a certain “cutoff” value is effectively zero. A cutoff therefore implies a characteristic scale for the degree distribution of the network, and because a power-law degree distribution lacks any such cutoff value, it is often called a scale-free distribution.

Distributions that are approximated by power laws in their tails have been studied for over a century, initially with respect to distributions of wealth (Mandelbrot 1960), but subsequently in a wide range of applications, including distributions of species size (Yule 1925), word frequencies (Zipf 1949), scientific citation counts (Price 1965, Redner 1998), and the sizes of various natural events [e.g., forest fires and earthquakes (Sornette 2000)] and social entities [e.g., business firms (Axtell 2001, Simon & Bonini 1958) and cities (Zipf 1949)]. After examining the degree distributions of three real-world networks—the collaboration network of movie actors, the link structure of a portion of the World Wide Web, and the power transmission grid of the western United States—Barabasi & Albert (1999) concluded





**Figure 3** (a) Power law with exponent  $\alpha$ , when plotted on a double-logarithmic scale, appears as a straight line with negative slope  $\alpha$ . (b) A normal-type distribution (i.e., with finite mean and variance) plotted on a double-logarithmic scale displays a well-defined cutoff, above which probabilities are effectively zero.

that all three exhibited power-law tails [a claim later shown to be incorrect in the case of the power grid (Amaral et al. 2000)], and coined the term *scale-free networks* to describe them.

Barabasi & Albert (1999) also proposed a simple mathematical model (known as the BA model) to account for their empirical observations. Their model—a special case of a much earlier model developed by Simon (1955) and later applied to citation networks by Price (1980)—embodied two mechanisms: population growth and preferential attachment. The intuition behind the first mechanism is straightforward: Real networks grow in time as new members join the population. The mechanism of preferential attachment then expressed the notion, analogous to Simon’s “Gibrat principle” (Simon 1955), Merton’s “Matthew Effect” (Merton 1968), and Price’s “cumulative advantage” (Price 1980), that newly arriving nodes will tend to connect to already well-connected nodes rather than poorly connected nodes. Specifically, Barabasi & Albert defined the probability  $p(k_i)$  of an existing node  $i$  with  $k_i$  links receiving a new link as  $p(k_i) = ck_i$  (where  $c$  is a normalizing constant). Barabasi & Albert showed that over a sufficiently long time horizon, the degree distribution of a growing network exhibiting linear preferential attachment would converge to something in the form of Equation 2 with an exponent  $\alpha = 3$ .

A number of subsequent models have been proposed that expand on the basic theme of growth with preferential attachment.

1. When either the creation of additional ties (Amaral et al. 2000) or the process of searching for additional neighbors (Mossa et al. 2002) is costly, the power-law distribution of degree becomes truncated, yielding a scaling region with a characteristic cutoff (“broad-scale networks”), or even a narrow distribution

- around a well-defined mean (“single-scale networks”), depending on the costs involved (Amaral et al. 2000).
2. When nodes are endowed with different intrinsic “attractiveness” (Bianconi & Barabasi 2001), or else are allowed to continue making and breaking ties after joining the network (Bornholdt & Ebel 2001), then any value of the exponent between 2 and 3 can be obtained—a result that, unlike the singular value of 3 above, is consistent with empirical exponents. Furthermore, unlike in the original BA model, older nodes are not necessarily better connected than more recent additions (Adamic & Huberman 2000).
  3. Rather than assuming preferential attachment as a fundamental feature of network growth, Krapivsky & Redner (2001) showed that it can be derived from more basic growth assumptions: (a) that new arrivals select a member of the existing population uniformly at random; and (b) that with some probability  $q$  they are redirected by their initial choice to one of its neighbors. Because neighbors of a randomly selected node in a network have a degree distribution that is proportional to  $kp(k)$  (Feld 1991, Newman 2003a), linear preferential attachment follows automatically, except that any exponent between 2 and 3 can be generated, depending on the choice of  $q$ .<sup>5</sup>
  4. Krapivsky et al. (2000) also showed that although linear preferential attachment is only one of many possible preferential attachment mechanisms, it is the only choice that leads to scale-free networks. If, for example, preferential attachment is sublinear [e.g.,  $p(k_i) = ck_i^\gamma$ , with  $\gamma < 1$ ] then the resulting degree distribution exhibits a characteristic cutoff; and if it is superlinear ( $\gamma > 1$ ) then a single “winner” will inevitably capture a finite fraction of all available edges—an inequality even more extreme than is implied by a power-law distribution.
  5. Finally, a more general approach to growing random networks that does not force new arrivals to connect to the existing network has been proposed by Callaway et al. (2001), who identify conditions on the relative rates of node and edge creation for the network to exhibit a giant connected component, rather than assuming (as the original BA model does) that everyone is part of the same growing cluster.

## EMPIRICAL NETWORK ANALYSIS

### Small-World and Scale-Free Networks

Stimulated by the predictions of universal features of networks, and encouraged by the rapidly increasing availability of large data sets, numerous authors have identified short global path lengths, high local clustering, and skewed degree

<sup>5</sup>Surprisingly, the model of Krapivsky & Redner (2001) turns out to be mathematically identical to that of Simon (1955), even though it is motivated differently.

distributions (often with power-law tails) in a variety of real-world networks, including metabolic reaction networks (Ebenhoh & Heinrich 2003, Gomez et al. 2001, Jeong et al. 2000, Schuster et al. 2002, Wagner & Fell 2001), genetic regulatory networks (Bhan et al. 2002, Guelzim et al. 2002), protein-protein interaction networks (Goldberg & Roth 2003, Vendruscolo et al. 2002, Wagner 2001, Wuchty 2002), biological neural networks (Morita et al. 2001, Shefi et al. 2002), food webs (Dunne et al. 2002, Montoya & Sole 2002), transportation networks (Sen et al. 2003), firm co-ownership networks (Kogut & Walker 2001), collaboration networks of boards of directors (Davis et al. 2003) and scientists (Newman 2001), the Internet (Faloutsos et al. 1999), and the link structure of the World Wide Web (Adamic 1999, Albert et al. 1999, Barabasi et al. 2000, Broder et al. 2000). The relative paucity of social network data in this list can be attributed to the difficulty inherent in recording social interactions on a large scale—a practical problem that the increasing extent of online activity (as a means of communicating, conducting business, recording activities, etc.) may help at least partially overcome.

## Network Motifs

In recent work, Shen-Orr et al. (2002) and Milo et al. (2002) have extended the exploration of network structure beyond small-world and scale-free categorizations with their approach of detecting what they call network motifs—topologically distinct subgraphs whose frequencies in a network can be used to characterize its local structure. By systematically enumerating all the motifs comprising three and four nodes in a number of networks, and then comparing the resulting counts with those for random networks—a technique that, for the three-node motif case, is identical in spirit to the triad census of Holland & Leinhardt (1981)—Milo et al. identified three classes of networks: (a) networks that process information (e.g., gene networks, the neural network of *C. elegans*, and forward logic chips); (b) networks that process energy (e.g., food webs); and (c) networks that symbolize communities of knowledge (e.g., the World Wide Web). Whether social networks fall into any of these classes or constitute one or more additional classes remains to be seen.

## Community Structure

Another area of recent progress has been in the identification of community structure—an intermediate scale of analysis between local (e.g., clustering, network motifs) and global (e.g., connectivity, path lengths) structure. Standard approaches to identifying community structure have tended to rely on some version of hierarchical clustering (Wasserman & Faust 1994)—a method for partitioning the network into increasingly similar subsets of nodes, where pairwise similarity can be measured in terms of, for example, cohesion (Moody & White 2003, White & Harary 2001) or equivalence (Lorrain & White 1971). Hierarchical clustering based on similarity measures, however, has a tendency to split off

poorly connected nodes, representing them as isolates rather than as members of the groups with which they would naturally be associated. Recently, Girvan & Newman (2002) have introduced a new method for identifying community structure that departs from the standard approach in two ways: (a) instead of computing pairwise node similarity, they compute edge betweenness centrality [roughly speaking, Freeman's (1977) notion of betweenness centrality, but applied to edges instead of nodes]; and (b) instead of adding edges to an empty graph, they delete them from the original graph in order of highest betweenness first, recomputing the betweenness of all remaining edges after each deletion. Their results are a striking improvement over hierarchical clustering methods, which they compare directly with their own for a number of test networks. In addition, they identify plausible communities in previously untested networks, such as the NCAA football competition and the research community of the Santa Fe Institute. Subsequently, Newman & Girvan (2003) have extended their method to include different measures of edge centrality and also to optimize the edge deletion algorithm such that the ratio of in-group to out-of-group ties in the partitioned network is maximized.

## Interpreting Empirical Data

When interpreting the measured structural features—whether local, global, or community—present in empirical network data, an important but typically overlooked distinction is that between what might be called “symbolic” networks, which can be thought of as network representations of abstract relations between discrete entities, and “interactive” networks, whose links describe tangible interactions that are capable of transmitting information, influence, or material. Metabolic reaction networks, word networks, and the World Wide Web, for example, are all symbolic networks, whereas advice networks, the Internet, and the power transmission grid are clearly interaction networks. The distinction is relevant to the interpretation of structural features because in a symbolic relationship it is frequently unclear how network metrics such as degree, path length, or centrality should be interpreted with respect to their consequences for some particular social, physical, or biological process. What does it mean, for example, that any two words in a large body of text can be “connected” through a short series of links, defined as two words co-occurring in the same sentence (Cancho & Sole 2001)? And how is this relationship different from the same two words being connected through a short chain of semantic relationships (Motter et al. 2002)? Or in the case of collaboration networks of movie actors, corporate directors, and scientists—often cited as examples of social networks—how should one interpret costarring in a movie versus sitting together on a board or coauthoring a paper? In all likelihood, these relationships involve different kinds of social interactions, but because the interactions themselves are underspecified, the network alone cannot be said to reveal much about actual social processes.

In the case of interaction networks, another frequently overlooked issue that bears on the interpretation of empirical findings is how much the nodes themselves vary in their intrinsic capacity to support connections. A widely cited paper by Faloutsos et al. (1999), for example, examines the connectivity of the Internet at two distinct scales of analysis: at the level of routers (the physical devices that actually perform packet switching), and at the level of autonomous systems (collections of routers that are administered by the same organization and typically correspond to domain names;<sup>6</sup> for example, columbia.edu can be considered a single autonomous system). In network terms, at the router scale, routers are the nodes, and links represent physical connections between routers; and at the autonomous system scale, domains are nodes, and two domains are linked if any router in one domain is connected to any router in the other. Faloutsos et al. (1999) claim that both networks exhibit power-law degree distributions, and this claim has been widely interpreted to imply that the Internet is scale-free regardless of the choice of representation (Yook et al. 2002). But a careful look at their data reveals that whereas nodes at the domain scale do indeed exhibit a large variance in degree, ranging between 1 and 1000 ties, individual routers never have fewer than 2 ties and never more than 40.<sup>7</sup> Furthermore, the exponent of the rank-ordered distribution for router degree corresponds to one with finite variance; hence, it is not scale-free.<sup>8</sup>

Based on their own results, therefore, Faloutsos et al.'s conclusion should have been that the Internet is scale-free at the autonomous system level, but not at the router level, where the likely explanation is that the autonomous systems themselves exhibit a wide range of sizes—large autonomous systems correspond to large organizations that use many routers, whereas small organizations require only a few. Naturally, the more routers that lie within a single domain, the more connections it can support to other domains; hence, as recent work (Tangmunarunkit et al. 2001) suggests, much of the observed variance in the degree distribution can be explained by the variance in the “size” of the nodes themselves. What the network analysis seems to reveal, in other words, may not be a property of network growth processes at all but merely a reflection of whatever processes generate the size distribution of the organizations that host Internet routers (Simon 1955). More generally, the artifactual or epiphenomenal nature of at least some

---

<sup>6</sup>Although there is some disagreement over whether autonomous systems and domains are in fact equivalent, they are generally agreed to be approximately so, and the distinction is typically ignored (Faloutsos et al. 1999).

<sup>7</sup>These numbers apply to the Internet as of 1995, the only router-level data that Faloutsos et al. (1999) present, and may well be different in today's Internet.

<sup>8</sup>If  $\gamma$  is the exponent of the tail of a rank-ordered distribution and  $\alpha$  is the exponent for the tail of the corresponding density function, then necessarily  $\alpha = 1 + 1/\gamma$  (this result follows automatically from the definitions of the two distributions). Faloutsos et al. (1999) measure  $\gamma < 0.5$ , corresponding to  $\alpha > 3$ , which implies a finite variance and hence a finite scale. Oddly, Faloutsos et al. do not make the connection between  $\alpha$  and  $\gamma$ , instead measuring  $\alpha$  independently, with the (inconsistent) result that  $\alpha < 3$ .

recent empirical network findings suggest that care is required when claiming universal features of networks, particularly when the claims in question extend to the collective dynamics of networked systems—the topic to which we now turn.

## NETWORKS AND COLLECTIVE DYNAMICS

Without doubt, the structure of real-world networks, and the evolution thereof, are legitimate scientific problems the general resolution of which presents us with substantial modeling and empirical challenges. It remains the case, however, that networks are interesting to sociologists (and to researchers in other disciplines as well) primarily because they are thought to influence individual (micro) and collective (macro) behavior, as well as the relationship between the two. At the individual level, network exchange theory (Cook et al. 1983, 1993; Walker et al. 2000; Willer 1999), for example, has demonstrated that the outcomes experienced by actors who engage in bilateral exchanges depend not only on their intrinsic preferences and perceived status differences, but also on the larger patterns of relations (i.e., the network) within which both actors are embedded. And at the collective level, sociologists since Solomonoff & Rapoport (1951) have been aware that the nature and arrangement of network ties may play an important role in fostering the global spread of information and disease (Boorman & Levitt 1980, Rapoport 1963), or facilitating collective action (Granovetter 1973). Unfortunately, although a rigorous and comprehensive vocabulary of network structure is clearly an important part of this larger problem—and thus recent progress in that respect is certainly helpful—the relationship between network structure and dynamical consequences is anything but straightforward. For example, it is almost certainly the case that the details of network structure (as in statistical measures like triad densities or degree distributions) that are relevant to individual and collective behavior will depend on the nature of the particular dynamical process under investigation (Watts 1999). The interaction between network structure and collective dynamics is therefore a daunting area of study, into which efforts to date have made only limited headway. Nevertheless, some insights have been forthcoming for very simple dynamical processes like the spread of disease and information.

### Disease Spreading

Many infectious diseases of humans spread either through human-to-human or human-to-animal interactions; hence, epidemiological models of infection seem like an obvious area of application for the science of networks, and for social networks in particular. It may come as a surprise, therefore, that the literature on mathematical epidemiology has paid relatively little attention to the structure of the networks—whether sexual, friendship, community, or transportation networks—through which infectious diseases manifestly spread. Whereas clinical epidemiologists routinely apply methods like contact tracing that are inherently network-oriented (Klondahl et al. 1994, Potterat et al. 1999), mathematical

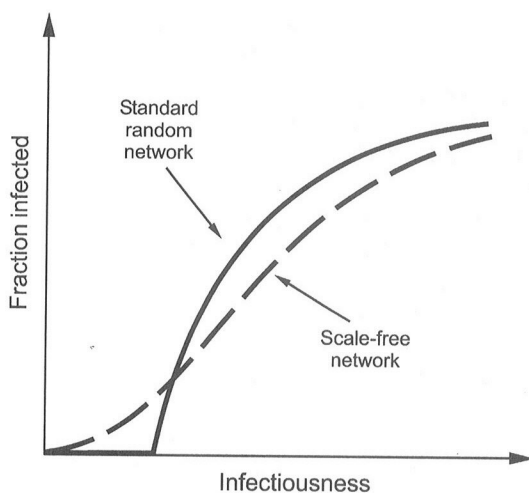
epidemiologists have preferred to develop modeling approaches that assume individuals within any given population mix randomly (Anderson & May 1991, Bailey 1975, Murray 1993).

The basic model of disease spread, introduced by Kermack & McKendrick (1927), is known as the SIR model representing the three basic classes—susceptible (S), infected (I), and removed (R)—that any given individual can inhabit at a particular time. The corresponding dynamics are as follows: (a) for each unit of time during which a susceptible is in contact with an infective, he or she may become infected with some probability  $\beta$  (the infectiousness of the disease); (b) infectives are removed from the population at some rate  $r$  (either by achieving immunity or by death); and (c) removed individuals are reinjected into the susceptible pool (through loss of immunity) at some other rate  $\rho$ . The populations  $S(t)$ ,  $I(t)$ , and  $R(t)$  [which necessarily sum to  $N(t)$ , the population size] therefore contain all current information about the state of the system. But because the infection stage requires contact between a susceptible and an infective, an additional assumption (or set of assumptions) is required to specify how individuals in different classes interact. In the basic model, the simplest assumption is made—that individuals interact randomly, in the manner of chemicals being stirred in a beaker. Thus the probability of a susceptible encountering an infective is proportional to the product of the S and I populations, where an interaction coefficient  $\gamma$  is usually introduced to specify the average rate of pairwise interactions. This assumption dramatically simplifies the subsequent analysis, yielding in the limit of large  $N$  a set of nonlinear differential equations, which can either be solved explicitly (in trivial cases) or else be used to determine the equilibrium states of the disease (typically the fraction of the population that becomes infected) for different parameter settings and initial conditions.

## SIR Models on Networks

Although the bulk of the mathematical epidemiology literature comprises increasingly elaborate versions of the SIR model with random mixing, some progress has been made in analyzing the steady-state behavior of some simple SIR-type models in which interactions are described by a network. Most of this work to date has focused on a generalized class of random network models in which properties like the shape of the degree distribution and the tendency of high-degree nodes to associate with each other (assortativity) can be varied arbitrarily (as opposed to Bernoulli random graphs, in which they are fixed), but that do not exhibit any local or community structure.

Early work by Kretschmar & Morris (1996) focused on the effects of concurrent relations on the spread of sexually transmitted diseases. Individuals were allowed to choose partners at random but were constrained in how many partnerships could be maintained simultaneously. By fixing the average degree of the network at  $\langle k \rangle = 1$ , Kretschmar & Morris showed that when concurrent relations were forbidden, the network would necessarily disintegrate into  $N/2$  isolated dyads, thus limiting any outbreak of disease. But when concurrency constraints were



**Figure 4** Comparison of disease spreading on Bernoulli (standard) and scale-free random networks. The steady-state fraction of infectives displays a distinct phase transition (i.e., an epidemic threshold) in the case of a Bernoulli random graph, but not for the scale-free network.

lifted, the resulting variance in degree caused the network to exhibit a giant connected component, thereby enabling most outbreaks to reach a finite fraction of the population. Pastor-Satorris & Vespignani (2001) have extended Kretschmar & Morris's notion of concurrency by investigating disease spreading on random networks with power-law distributions of degree. Their main finding has been that infinite random networks exhibiting power-law degree distributions with an exponent  $\alpha < 3$  do not exhibit the epidemic threshold predicted by uniform mixing models. Rather, as indicated schematically in Figure 4, the fraction of the population infected in a standard SIR model is finite for all non-zero values of the infectiousness—a result that Lloyd & May (2001) subsequently pointed out had been derived earlier using different methods (Anderson & May 1988). The explanation is that in a scale-free network the presence of highly connected hubs (roughly equivalent to very high concurrency in the language of Kretschmar & Morris) ensures that even relatively benign diseases will spread widely if they infect even a single hub. The claim that networks of sexual contacts are, in fact, scale-free (Liljeros et al. 2001), however, has subsequently been questioned (Jones & Handcock 2003), as has the absence of the epidemic threshold when factors such as the geography (Warren et al. 2002), transitivity (Eguiluz & Klemm 2002), attenuation of infectiousness over time (Wu et al. 2004), and the finite size of real networks (May & Lloyd 2001) are taken into account.

Another area of active investigation concerns assortative mixing in random networks, where assortativity implies that high-degree nodes tend to be connected



disproportionately to other high-degree nodes (in disassortative networks, high-degree nodes tend to be connected to low-degree nodes). A number of measures of assortativity have been proposed (Gupta et al. 1989, Maslov & Sneppen 2002, Pastor-Satorras et al. 2001, Vazquez et al. 2002), but the simplest and most flexible appears to be that of Newman (2002, 2003b), who defines it in terms of the Pearson correlation coefficient of the degrees of adjacent nodes<sup>9</sup>—a definition that can be applied somewhat more generally to any scalar attribute such as age, race, etc. (Newman 2003b). Surveying a variety of real-world networks, Newman concluded that social networks tend to exhibit positive degree correlations—hence assortative mixing—whereas technological and biological networks (e.g., power grid, Internet, food webs) are disassortatively mixed. Applying this finding to the spread of disease, Newman showed that in networks with assortative mixing, the giant connected component tends to be smaller than in purely random networks but has an exceptionally small diameter and is also extremely resilient to disconnection. When mixing is disassortative, the opposite result applies, where both findings are consistent with earlier claims concerning the effects of assortativity, based on traditional random mixing models (Gupta et al. 1993).

In addition to studying random networks, scholars have made some efforts to study the effects of local network structure on disease propagation. For a standard SIR-type model adapted to spread on a small-world network, Watts & Strogatz (1998) used computer simulations to suggest that even a small probability of random rewiring would dramatically reduce the infectiousness required of a disease in order for it to infect a finite fraction of the population—a result that was similar to earlier work by Ball et al. (1997). Furthermore, Watts & Strogatz showed that for a highly infectious disease—that is, one for which a global epidemic is a certainty—the time required for the entire population to become infected is approximately proportional to the average path length of the network and hence highly susceptible to the fraction of random shortcuts. These early results have been refined subsequently by Newman & Watts (1999b) and Moore & Newman (2000a,b), who used an equivalent percolation model (Stauffer & Aharony 1992) to study the impact of random shortcuts on the steady-state size of an epidemic in one- and two-dimensional lattices. Similar models have also been proposed by Keeling (1999) and Eames & Keeling (2002), with similar results. And Kuperman & Abramson (2001) found that when recovered individuals in an SIR model instantly become resusceptible (the so-called SIS model), low levels of randomness lead to low-level endemicity, high randomness leads to global infectivity, and intermediate levels generate self-sustained oscillations of the infected population—reminiscent of the seasonality of diseases like influenza.

A somewhat more general outcome of these models is the notion of a correlation length of a network, defined as the distance between ends of shortcuts (Newman &

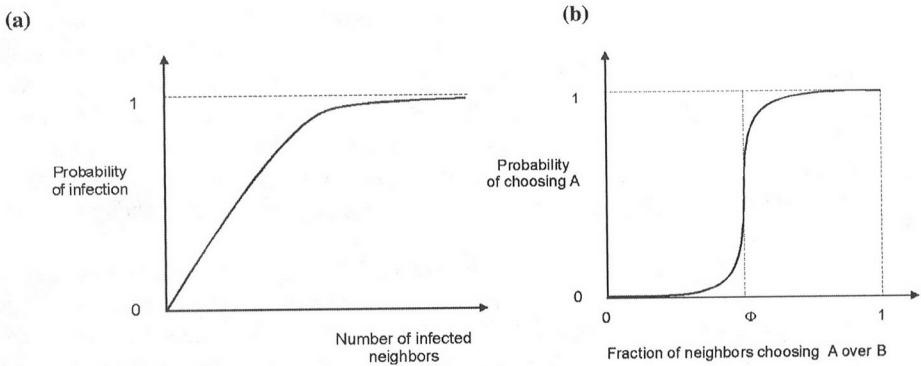
---

<sup>9</sup>In a Bernoulli random graph, degree assortativity is by construction zero; hence, real networks can be characterized as assortative or disassortative to the extent that they exhibit a positive or negative degree correlation, respectively.

Watts 1999b). The correlation length regulates how far from its source a contagious influence must spread before it “feels” the effect of the shortcuts. Hence an outbreak of disease can be divided naturally into two stages: (a) a local growth phase in which its spread is slowed by the redundancy present in the local topology of the network (e.g., a lattice); and (b) a global growth phase, in which it can spread exponentially throughout the network via the random shortcuts. Because the correlation length governs the transition from local to global growth, these models suggest that to be effective, strategies for preventing epidemics should focus on eliminating even very rare shortcuts, thus prolonging local growth for as long as possible. The correlation length of networks may also be important from an evolutionary perspective. Boots & Sasaki (1999) have observed that vector-borne diseases such as malaria and cholera tend to be more virulent (that is, they kill off their hosts rapidly) than infectious diseases spread by direct contact. Their explanation was that vector-borne diseases have historically been able to spread via long-range connections—transported by mosquitoes or flowing water, for example—whereas human-to-human infections have been more constrained by geographical proximity and hence must survive a local growth phase before they can reach epidemic status. Diseases that rely on direct contact will therefore evolve lower virulence, they argued, in order not to burn themselves out before they reach the correlation length, whereas diseases that are not constrained by an initial local growth phase will evolve to be highly virulent. Boots & Sasaki then suggested that as the correlation length of a network shrinks on account of the introduction of random shortcuts—as the world becomes smaller—one might expect the evolution of more virulent human-to-human diseases.

## Social Contagion

An important but less-studied variant of contagion applies to a class of models that deal with collective decision making. Whether in the context of innovations (Rogers 1995), competing technologies (Arthur 1989), cultural fads (Banerjee 1992, Bikhchandani et al. 1992), social norms and conventions (Young 1998), cooperation (Axelrod 1984, Glance & Huberman 1993), social disorder (Lohmann 1994), or financial markets (Mackay 1932, Shiller 2000, Sornette 2003), individuals making decisions frequently pay attention to the actions or decisions of other individuals (Katz & Lazarsfeld 1955, Watts 2003). Hence, the aggregation of individual to collective decision making can be understood in terms of social contagion, where decisions are “transmitted” from one individual to another in a manner reminiscent of disease. Unlike SIR-type models, however, which assume contagion to be a memory-free process, individuals making decisions are affected simultaneously by past as well as current interactions. In an SIR model, for example, the probability of a susceptible becoming infected upon contact with an infective is the same regardless of how many infectives it has previously encountered. Hence, the cumulative probability of infection over a succession of contacts looks roughly like Figure 5a. By contrast, the probability of believing an apocryphal rumor or adopting a new innovation may change dramatically upon a



**Figure 5** (a) The cumulative probability of infection for an SIR-type model increases approximately linearly with the number of exposures to infectives—a consequence of the assumed memory-free nature of biological infection. (b) In social contagion, decisions are costly and are affected by the memory of past interactions; hence, the cumulative probability of “infection” exhibits threshold-like behavior.

second or third (or subsequent) exposure, thus resembling Figure 5b, where the sudden jump occurs when the individual’s “threshold” of belief is exceeded.

Threshold models of this kind arise widely (Boorman & Levitt 1980, Granovetter & Soong 1988, Schelling 1978, Watts 2003) and can display very different collective dynamics than the SIR-type models described above. When individuals are assumed to mix randomly, for example, SIR models admit the possibility that an epidemic can be triggered by an infinitesimal initial seed (even a single infective may suffice), whereas models that embody nontrivial thresholds require an initial critical mass—that is, a finite fraction of the population—before global infection is possible (Dodds & Watts 2004). And when social contagion takes place via a sparse network, epidemics—sometimes called global cascades—can only occur for social contagion when the connectivity of the network is neither too sparse nor too dense (Watts 2002). It is the latter condition that is surprising because in biological contagion greater connectivity always implies a greater possibility that a disease will spread. In social contagion, however, dense connectivity renders many individuals impervious to novel signals, thereby quashing cascades before they can even begin. Another consequence of increasing the connectivity of sparse random networks is that cascades become larger but increasingly rare. As a result, it is possible for a universal cascade (i.e., one affecting the entire population) to be triggered by a random shock that, a priori, is indistinguishable from hundreds of innocuous shocks (Watts 2002).

## Network Robustness

Another class of network-related processes that has been the subject of recent attention concerns the origins and effects of node failure. Whether applied to

routers in the Internet, airports in the air-transportation system, or individuals in a large organization, nodes in interactive networks are generally susceptible both to endogenous failure—the risk of which rises dramatically whenever a node's limited capacity to process information or material is exceeded by endogenously generated demand—and also to exogenous failure, driven for example by random component breakdown, sickness, accidents, environmental catastrophes, or even intentional attack. A robust network is one that resists one or both kinds of failure, ideally under a wide range of environmental conditions.

Assuming exogenous failures, Albert et al. (2000) demonstrated with computer simulations that scale-free random networks of the kind generated through growth and preferential attachment exhibit much greater resilience to random failures than Bernoulli random graphs, where they defined resilience as the size of the largest connected component remaining after a certain fraction of nodes in the network had been removed. When nodes were removed preferentially, however, in order from highest to lowest degree, the opposite result pertained. The explanation for both results is that the connectivity of a scale-free network is determined almost entirely by the small fraction of highly connected hubs that occupy the tail of the degree distribution. Because random failures are relatively unlikely to affect the hubs, the only nodes that are likely to become disconnected are those that fail. But when the hubs are deliberately targeted, each individual failure has the potential to impact the connectivity of the entire network. Subsequent work (Callaway et al. 2000; Cohen et al. 2000, 2001) has confirmed this basic feature of scale-free networks and extended the findings to more general classes of degree distributions (for example, power-law distributions with finite cutoffs), and also for link as well as node failure (Callaway et al. 2000).

More recently, endogenous failure—due, for example, to network congestion—has begun to be incorporated into models of network robustness (Arenas et al. 2001; Dodds et al. 2003b; Goh et al. 2001; Guimera et al. 2001, 2002). For example, Arenas et al. (2001) have shown that when randomly chosen pairs in a pure hierarchy pass messages to each other, where intermediate nodes incur a small cost for each message forwarded, then at a critical rate of message passing network congestion undergoes a phase transition (Newman & Barkema 1999), above which almost all messages are inhibited by the limited capacity of the small fraction of nodes at the top of the hierarchy. Thus both traditional (Williamson 1975) and also more recent (Bolton & Dewatripont 1994, Radner 1993) economic models of the internal architecture of firms, which typically propose pure hierarchies as the most efficient form of organization, can be seen as ignoring a critical aspect of distributed problem solving—namely the capacity constraints inherent in distributed communication.

Recently, Dodds et al. (2003b) have proposed a model of information exchange in organizational networks that combines the notion of efficiency (understood roughly as maximizing performance at minimum cost of adding network ties) and robustness, finding that a simple algorithm for adding horizontal and vertical ties to a pure hierarchical backbone yields what they call *ultrarobust networks*—networks

that simultaneously minimize the likelihood of individual node failure (due to endogenously generated congestion), and also the impact on global connectivity that results from failures arising out of either endogenous or exogenous causes. The key to ultrarobustness, they find, is that organizational networks must exhibit nonhierarchical ties that extend across all scales of the underlying hierarchy—a quantitative result that stands in qualitative agreement with a longstanding body of work in organizational sociology (Burns & Stalker 1961, Granovetter 1985, Lawrence & Lorsch 1967), and is also consistent with recent accounts of firms surviving major disasters (Kelly & Stark 2002, Nishiguchi & Beaudet 2000).

## CONCLUSION

As I hope this review has made clear, many of the ideas and metrics of the “new” science of networks have either been borrowed from, or else rediscovered independently of, a distinguished lineage of work in mathematics, economics, and sociology. For example, the intuition that a few random “shortcuts” in the midst of locally dense neighborhoods can generate the structure of small-world networks (Watts & Strogatz 1998) bears a remarkable similarity to Rapoport’s (1963) analysis of random-biased nets, which inspired Granovetter’s (1973) investigation of the strength of weak ties. Recent Internet-based social search experiments (Dodds et al. 2003) explicitly acknowledge their debt to Milgram’s seminal studies (Korte & Milgram 1970, Milgram 1967, Travers & Milgram 1969). The finding that preferential attachment mechanisms in growing networks can lead to what are now called scale-free networks (Barabasi & Albert 1999) turns out to be a special case of Simon’s (1955) implementation of the Gibrat principle, better known in sociology as the Matthew Effect (Merton 1968), and first applied to networks by Price (1980). Related claims that highly connected hubs in a sexual network can enhance the spread of infectious diseases (Pastor-Satorras & Vespignani 2001) are equivalent to findings by Anderson & May (1988) that are well known in the mathematical epidemiology community. The topic of network motifs, which is gaining attention in the biological sciences (Milo et al. 2002, Shen-Orr et al. 2002), is, in principle, identical to the triad census approach of Holland & Leinhardt (1981). And recent work on the structure of ultrarobust networks (Dodds et al. 2003b) fits into a long line of work in organizational sociology starting with Burns & Stalker’s (1961) descriptions of organic organizations.

As much as it bears similarities to previous work, however, what is new about the current generation of network-related research is a rapidly emerging and highly interdisciplinary synthesis of new analytical techniques, enormously greater computing power, and an unprecedented volume of empirical data. Sociologists have much to gain from this progress, and also much to contribute. Most of the work discussed in this review is taking place in the mathematical sciences, particularly in physics. That is hardly surprising, as physics and mathematics typically lead

the way methodologically, whereas biological and social sciences follow with applications. But in this particular case, the flow of ideas has been bidirectional, with many of the core ideas—not just applications—having come from sociology. Physicists may be marvelous technicians, but they are mediocre sociologists. Thus, if the science of networks is to live up to its early promise, then the other disciplines—sociology in particular—must offer guidance in, for example, the interpretation of empirical and theoretical findings, particularly in the context of policy applications, and also in suggesting measures and models that are increasingly relevant to the important problems at hand.

## ACKNOWLEDGMENTS

The author acknowledges the support of the National Science Foundation (SES-0094162), Office of Naval Research, James S. McDonnell Foundation, Legg-Mason Funds, and Columbia Office of Strategic Initiatives.

*The Annual Review of Sociology* is online at <http://soc.annualreviews.org>

## LITERATURE CITED

- Adamic LA. 1999. The small world web. In *Lecture Notes in Computer Science*, pp. 443–54. New York: Springer
- Adamic LA, Huberman BA. 2000. Power-law distribution of the World Wide Web. *Science* 287:A2115
- Ahuja RK, Magnanti TL, Orlin JB. 1993. *Network Flows: Theory, Algorithms, and Applications*. Upper Saddle River, NJ: Prentice Hall
- Albert R, Jeong H, Barabasi A-L. 1999. Diameter of the World-Wide Web. *Nature* 401:130–31
- Albert R, Jeong H, Barabasi A-L. 2000. Error and attack tolerance of complex networks. *Nature* 406:378–82
- Amaral LAN, Scala A, Barthelemy M, Stanley HE. 2000. Classes of small-world networks. *Proc. Natl. Acad. Sci. USA* 97:11149–52
- Anderson C, Wasserman S, Crouch B. 1999. A p\* primer: logit models for social networks. *Soc. Netw.* 21:37–66
- Anderson RM, May RM. 1988. Epidemiological parameters of HIV transmission. *Nature* 333:514–19
- Anderson RM, May RM. 1991. *Infectious Diseases of Humans*. Oxford: Oxford Univ. Press
- Arenas A, Diaz-Guilera A, Guimera R. 2001. Communication in networks with hierarchical branching. *Phys. Rev. Lett.* 86:3196–99
- Arthur WB. 1989. Competing technologies, increasing returns, and lock-in by historical events. *Econ. J.* 99:116–31
- Axelrod RM. 1984. *The Evolution of Cooperation*. New York: Basic Books. 241 pp.
- Axtell R. 2001. Zipf distribution of U.S. firm sizes. *Science* 293:1818–20
- Bailey NTJ. 1975. *The Mathematical Theory of Infectious Diseases and Its Applications*. New York: Hafner
- Ball F, Mollison D, Scalia-Tomba G. 1997. Epidemics with two levels of mixing. *Ann. Appl. Probab.* 7:46–89
- Banerjee AV. 1992. A simple model of herd behavior. *Q. J. Econ.* 107:797–817
- Barabasi A-L. 2002. *Linked: The New Science of Networks*. Cambridge, MA: Perseus. 280 pp.
- Barabasi A-L, Albert R. 1999. Emergence of scaling in random networks. *Science* 286:509–12

- Barabasi A-L, Albert R, Jeong H. 2000. Scale-free characteristics of random networks: the topology of the World Wide Web. *Physica A* 281:69–77
- Barrat A, Weigt M. 2000. On the properties of small-world network models. *Eur. Phys. J. B* 13:547–60
- Barthelemy M, Amaral LAN. 1999. Small-world networks: evidence for a crossover picture. *Phys. Rev. Lett.* 82:3180–83. Erratum. 1999. *Phys. Rev. Lett.* 82:5180
- Bernard HR, Killworth PD, Evans MJ, McCarty C, Shelley GA. 1988. Studying social relations cross-culturally. *Ethnology* 27(2):155–79
- Bhan A, Galas DJ, Dewey TG. 2002. A duplication growth model of gene expression networks. *Bioinformatics* 18:1486–93
- Bianconi G, Barabasi A-L. 2001. Competition and multiscaling in evolving networks. *Europhys. Lett.* 54:436–42
- Bikhchandani S, Hirshleifer D, Welch I. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Polit. Econ.* 100:992–1026
- Bollobas B. 1998. *Modern Graph Theory*. New York: Springer
- Bolton P, Dewatripont M. 1994. The firm as a communication network. *Q. J. Econ.* 109:809–39
- Boorman SA, Levitt PR. 1980. *The Genetics of Altruism*. New York: Academic. 459 pp.
- Boots M, Sasaki A. 1999. ‘Small worlds’ and the evolution of virulence: Infection occurs locally and at a distance. *Proc. R. Soc. London Ser. B* 266:1933–38
- Bornholdt S, Ebel H. 2001. World Wide Web scaling exponent from Simon’s 1955 model. *Phys. Rev. E* 6403:035104
- Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, et al. 2000. Graph structure in the web. *Comput. Netw.* 33:309–20
- Buchanan M. 2002. *Nexus: Small Worlds and the Groundbreaking Science of Networks*. New York: Norton. 235 pp.
- Burns T, Stalker GM. 1961. *The Management of Innovation*. London: Tavistock. 269 pp.
- Callaway DS, Hopcroft JE, Kleinberg JM, Newman MEJ, Strogatz SH. 2001. Are randomly grown graphs really random? *Phys. Rev. E* 6404:041902
- Callaway DS, Newman MEJ, Strogatz SH, Watts DJ. 2000. Network robustness and fragility: percolation on random graphs. *Phys. Rev. Lett.* 85:5468–71
- Cancho RFI, Sole RV. 2001. The small world of human language. *Proc. R. Soc. London Ser. B* 268:2261–65
- Cohen R, Erez K, Ben-Avraham D, Havlin S. 2000. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* 85:4626–28
- Cohen R, Erez K, Ben-Avraham D, Havlin S. 2001. Breakdown of the Internet under intentional attack. *Phys. Rev. Lett.* 86:3682–85
- Cook KS, Emerson RM, Gillmore MR, Yamagishi T. 1983. The distribution of power in exchange networks: theory and experimental results. *Am. J. Sociol.* 89:275–305
- Cook KS, Molm LD, Yamagishi T. 1993. Exchange relations and exchange networks: recent developments in social exchange theory. In *Theoretical Research Programs: Studies in Theory Growth*, ed. J Berger, M Zelditch, pp. 296–322. Stanford, CA: Stanford Univ. Press
- Davis GF, Yoo M, Baker WE. 2003. The small world of the American corporate elite, 1982–2001. *Strateg. Organ.* 1:301–26
- Degenne A, Forse M. 1994. *Introducing Social Networks*. London: Sage
- Dodds PS, Muhamad R, Watts DJ. 2003a. An experimental study of search in global social networks. *Science* 301:827–29
- Dodds PS, Watts DJ. 2004. Universal behavior in a generalized model of contagion. *Phys. Rev. Lett.* In press
- Dodds PS, Watts DJ, Sabel CF. 2003b. Information exchange and the robustness of organizational networks. *Proc. Natl. Acad. Sci. USA* 100:12516–21
- Dorogovtsev SN, Mendes JFF. 2000. Exactly solvable small-world network. *Europhys. Lett.* 50:1–7
- Dunne JA, Williams RJ, Martinez ND. 2002. Food-web structure and network theory: the

- role of connectance and size. *Proc. Natl. Acad. Sci. USA* 99:12917–22
- Eames KTD, Keeling MJ. 2002. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc. Natl. Acad. Sci. USA* 99:13330–35
- Ebenhoh O, Heinrich R. 2003. Stoichiometric design of metabolic networks: multifunctionality, clusters, optimization, weak and strong robustness. *Bull. Math. Biol.* 65:323–57
- Eguiluz VM, Klemm K. 2002. Epidemic threshold in structured scale-free networks. *Phys. Rev. Lett.* 89:108701
- Faloutsos M, Faloutsos P, Faloutsos C. 1999. On power-law relationships of the Internet topology. *Comput. Commun. Rev.* 29:251–62
- Feld S. 1991. Why your friends have more friends than you do. *Am. J. Sociol.* 96:1464–77
- Frank O, Strauss D. 1986. Markov graphs. *J. Am. Stat. Assoc.* 81:832–42
- Freeman L. 1977. A set of measures of centrality based upon betweenness. *Sociometry* 40:35–41
- Freeman LC, White DR. 1993. Using Galois lattices to represent network data. *Sociol. Methodol.* 23:127–46
- Garfield E. 1979. It's a small world after all. In *Essays of an Information Scientist*, ed. E Garfield, pp. 299–304. Philadelphia, PA: ISI
- Girvan M, Newman MEJ. 2002. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* 99:7821–26
- Glance NS, Huberman BA. 1993. The outbreak of cooperation. *J. Math. Sociol.* 17:281–302
- Goh KI, Kahng B, Kim D. 2001. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.* 87:278701
- Goldberg DS, Roth FP. 2003. Assessing experimentally derived interactions in a small world. *Proc. Natl. Acad. Sci. USA* 100:4372–76
- Gomez SM, Lo SH, Rzhetsky A. 2001. Probabilistic prediction of unknown metabolic and signal-transduction networks. *Genetics* 159:1291–98
- Granovetter M. 1985. Economic action and social structure: the problem of embeddedness. *Am. J. Sociol.* 91:481–510
- Granovetter M, Soong R. 1988. Threshold models of diversity: Chinese restaurants, residential segregation, and the spiral of silence. *Sociol. Methodol.* 18:69–104
- Granovetter MS. 1973. The strength of weak ties. *Am. J. Sociol.* 78:1360–80
- Guelzim N, Bottani S, Bourguin P, Kepes F. 2002. Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* 31:60–63
- Guimera R, Arenas A, Diaz-Guilera A. 2001. Communication and optimal hierarchical networks. *Physica A* 299:247–52
- Guimera R, Diaz-Guilera A, Vega-Redondo F, Cabrales A, Arenas A. 2002. Optimal network topologies for local search with congestion. *Phys. Rev. Lett.* 89:248701
- Gupta S, Anderson RM, May RM. 1989. Networks of sexual contacts: implications for the pattern of spread of HIV. *AIDS* 3:807–17
- Gupta S, Anderson RM, May RM. 1993. Mathematical models and the design of public health policy: HIV and antiviral therapy. *SIAM Rev.* 35:1–16
- Holland PW, Leinhardt S. 1981. An exponential family of probability distributions for directed graphs. *J. Am. Stat. Assoc.* 76:33–65
- Jeong H, Tombor B, Albert R, Oltval ZN, Barabasi AL. 2000. The large-scale organization of metabolic networks. *Nature* 407:651–54
- Jones JH, Handcock MS. 2003. An assessment of preferential attachment as a mechanism for human sexual network formation. *Proc. R. Soc. London Ser. B* 270:1123–28
- Katz E, Lazarsfeld PF. 1955. *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Glencoe, IL: Free Press. 400 pp.
- Keeling MJ. 1999. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. London Ser. B* 266:859–67
- Kelly J, Stark D. 2002. Crisis, recovery, innovation: learning from 9/11. *Environ. Plan. A* 34:1523–33



- Kermack WO, McKendrick AG. 1927. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. London Ser. A* 115:700–21
- Killworth PD, Bernard HR. 1978. The reverse small world experiment. *Soc. Netw.* 1:159–92
- Kleinberg J. 2000a. The small-world phenomenon: an algorithmic perspective. *Proc. 32nd ACM Symp. Theory Comput.*, pp. 163–70
- Kleinberg JM. 2000b. Navigation in a small world—it is easier to find short chains between points in some networks than others. *Nature* 406:845
- Klov Dahl AS, Poterat JJ, Woodhouse DE, Muth JB, Muth SQ, Darrow WW. 1994. Social networks and infectious disease: the Colorado Springs study. *Soc. Sci. Med.* 38:79–88
- Kogut B, Walker G. 2001. The small world of Germany and the durability of national networks. *Am. Sociol. Rev.* 66:317–35
- Korte C, Milgram S. 1970. Acquaintance linking between white and Negro populations: application of the small world problem. *J. Personal. Soc. Psychol.* 15:101–18
- Krapivsky PL, Redner S. 2001. Organization of growing random networks. *Phys. Rev. E* 6306:066123
- Krapivsky PL, Redner S, Leyvraz F. 2000. Connectivity of growing random networks. *Phys. Rev. Lett.* 85:4629–32
- Kretschmar M, Morris M. 1996. Measures of concurrency in networks and the spread of infectious disease. *Math. Biosci.* 133:165–95
- Kulkarni RV, Almaas E, Stroud D. 2000. Exact results and scaling properties of small-world networks. *Phys. Rev. E* 61:4268–71
- Kuperman M, Abramson G. 2001. Small world effect in an epidemiological model. *Phys. Rev. Lett.* 86:2909–12
- Lawrence PR, Lorsch JW. 1967. *Organization and Environment: Managing Differentiation and Integration*. Boston: Div. Res. Grad. Sch. Bus. Adm. Harvard Univ. 279 pp.
- Lazarsfeld P, Merton R. 1954. Friendship as social process: a substantive and methodological analysis. In *Freedom and Control in Modern Society*, ed. M Berger, T Abel, C Page, pp. 18–66. New York: Van Nostrand
- Liljeros F, Edling CR, Amaral LAN, Stanley HE, Aberg Y. 2001. The web of human sexual contacts. *Nature* 411:907–8
- Lloyd AL, May RM. 2001. Epidemiology—how viruses spread among computers and people. *Science* 292:1316–17
- Lohmann S. 1994. The dynamics of informational cascades: the Monday demonstrations in Leipzig, East Germany, 1989–91. *World Polit.* 47:42–101
- Lorrain FP, White HC. 1971. Structural equivalence of individuals in social networks. *J. Math. Sociol.* 1:49–80
- Mackay C. 1932. *Extraordinary Popular Delusions and the Madness of Crowds*. Boston: Page. 724 pp.
- Mandelbrot B. 1960. The Pareto-Levy Law and the distribution of income. *Int. Econ. Rev.* 1:79–106
- Maslov S, Sneppen K. 2002. Specificity and stability in topology of protein networks. *Science* 296:910–13
- May RM, Lloyd AL. 2001. Infection dynamics on scale-free networks. *Phys. Rev. E* 6406:066112
- Merton RK. 1968. The Matthew effect in science. *Science* 159:56–63
- Milgram S. 1967. The small world problem. *Psychol. Today* 2:60–67
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. 2002. Network motifs: simple building blocks of complex networks. *Science* 298:824–27
- Montoya JM, Sole RV. 2002. Small world patterns in food webs. *J. Theor. Biol.* 214:405–12
- Moody J, White DR. 2003. Social cohesion and embeddedness: a hierarchical concept of social groups. *Am. Sociol. Rev.* 68:1–25
- Moore C, Newman MEJ. 2000a. Epidemics and percolation in small-world networks. *Phys. Rev. E* 61:5678–82
- Moore C, Newman MEJ. 2000b. Exact solution of site and bond percolation on small-world networks. *Phys. Rev. E* 62:7059–64

- Morita S, Oshio K, Osana Y, Funabashi Y, Oka K, Kawamura K. 2001. Geometrical structure of the neuronal network of *Caenorhabditis elegans*. *Physica A* 298:553–61
- Mossa S, Barthelemy M, Stanley HE, Amaral LAN. 2002. Truncation of power law behavior in “scale-free” network models due to information filtering. *Phys. Rev. Lett.* 88: 138701
- Motter AE, de Moura APS, Lai YC, Dasgupta P. 2002. Topology of the conceptual network of language. *Phys. Rev. E* 65:065102
- Murray JD. 1993. *Mathematical Biology*. Berlin: Springer-Verlag. 2nd ed.
- Newman MEJ. 2001. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* 98:404–9
- Newman MEJ. 2002. Assortative mixing in networks. *Phys. Rev. Lett.* 89:208701
- Newman MEJ. 2003a. Ego-centered networks and the ripple effect. *Soc. Netw.* 25:83–95
- Newman MEJ. 2003b. Mixing patterns in networks. *Phys. Rev. E* 67:026126
- Newman MEJ. 2003c. The structure and function of complex networks. *SIAM Rev.* 45:167–256
- Newman MEJ, Barkema GT. 1999. *Monte Carlo Methods in Statistical Physics*. Oxford/New York: Clarendon/Oxford Univ. Press. 475 pp.
- Newman MEJ, Girvan M. 2003. *Finding and evaluating community structure in networks*. <http://xxx.lanl.gov/abs/cond-mat/0308217>
- Newman MEJ, Moore C, Watts DJ. 2000. Mean-field solution of the small-world network model. *Phys. Rev. Lett.* 84:3201–4
- Newman MEJ, Park J. 2003. Why social networks are different from other types of networks. *Phys. Rev. E* 68:036122
- Newman MEJ, Strogatz SH, Watts DJ. 2001. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* 64:02:026118
- Newman MEJ, Watts DJ. 1999a. Renormalization group analysis of the small-world network model. *Phys. Lett. A* 263:341–46
- Newman MEJ, Watts DJ. 1999b. Scaling and percolation in the small-world network model. *Phys. Rev. E* 60:7332–42
- Newman MEJ, Watts DJ, Strogatz SH. 2002. Random graph models of social networks. *Proc. Natl. Acad. Sci. USA* 99:2566–72
- Nishiguchi T, Beaudet A. 2000. Fractal design: self-organizing links in supply chain. In *Knowledge Creation: A New Source of Value*, ed. G Von Krogh, I Nonaka, T Nishiguchi, pp. 199–230. London: Macmillan
- Pastor-Satorras R, Vazquez A, Vespignani A. 2001. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.* 87:258701
- Pastor-Satorras R, Vespignani A. 2001. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86:3200–3
- Pattison P, Wasserman S. 1999. Logit models and logistic regressions for social networks. II. Multivariate relations. *Br. J. Math. Stat. Psychol.* 52:169–93
- Pool IdeS, Kochen M. 1978. Contacts and influence. *Soc. Netw.* 1:1–48
- Potterat JJ, Rothenberg RB, Muth SQ. 1999. Network structural dynamics acid infectious disease propagation. *Int. J. STD AIDS* 10:182–85
- Price DJdeS. 1965. Networks of scientific papers. *Science* 149:510–15
- Price DJdeS. 1980. A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inf. Sci.* 27:292–306
- Radner R. 1993. The organization of decentralized information processing. *Econometrica* 61:1109–46
- Rapoport A. 1951. Nets with distance bias. *Bull. Math. Biophys.* 13:85–91
- Rapoport A. 1953a. Spread of information through a population with socio-structural bias: I. Assumption of transitivity. *Bull. Math. Biophys.* 15:523–33
- Rapoport A. 1953b. Spread of information through a population with socio-structural bias: II. Various models with partial transitivity. *Bull. Math. Biophys.* 15:535–46
- Rapoport A. 1957. Contribution to the theory of random and biased nets. *Bull. Math. Biophys.* 19:257–77
- Rapoport A. 1963. Mathematical models of

- social interaction. In *Handbook of Mathematical Psychology*, ed. RD Luce, RR Bush, E Galanter, 2:493–579. New York: Wiley
- Redner S. 1998. How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B* 4:131–34
- Robins G, Pattison P, Wasserman S. 1999. Logit models and logistic regressions for social networks: III. Valued relations. *Psychometrika* 64:371–94
- Rogers EM. 1995. *Diffusion of Innovations*. New York: Free Press. 519 pp.
- Schelling TC. 1978. *Micromotives and Macrobehavior*. New York: Norton. 252 pp.
- Schuster S, Pfeiffer T, Moldenhauer F, Koch I, Dandekar T. 2002. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to mycoplasma pneumoniae. *Bioinformatics* 18:351–61
- Scott J. 2000. *Social Network Analysis: A Handbook*. London: Sage
- Sen P, Dasgupta S, Chatterjee A, Sreeram PA, Mukherjee G, Manna SS. 2003. Small-world properties of the Indian railway network. *Phys. Rev. E* 67:036106
- Shefi O, Golding I, Segev R, Ben-Jacob E, Ayali A. 2002. Morphological characterization of in vitro neuronal networks. *Phys. Rev. E* 66:021905
- Shen-Orr SS, Milo R, Mangan S, Alon U. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* 31:64–68
- Shiller RJ. 2000. *Irrational Exuberance*. Princeton, NJ: Princeton Univ. Press. 296 pp.
- Simon HA. 1955. On a class of skew distribution functions. *Biometrika* 42:425–40
- Simon HA, Bonini CP. 1958. The size distribution of business firms. *Am. Econ. Rev.* 48:607–17
- Solomonoff R, Rapoport A. 1951. Connectivity of random nets. *Bull. Math. Biophys.* 13:107–17
- Sornette D. 2000. *Critical Phenomena in Natural Sciences*. Berlin: Springer-Verlag
- Sornette D. 2003. *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton, NJ: Princeton Univ. Press. 421 pp.
- Stauffer D, Aharony A. 1992. *Introduction to Percolation Theory*. London: Taylor & Francis
- Tangmunarunkit H, Doyle J, Govindan R, Jamin S, Shenker S, Willinger W. 2001. Does AS size determine degree in AS topology? *Comput. Commun. Rev.* 31:7–10
- Travers J, Milgram S. 1969. An experimental study of the small world problem. *Sociometry* 32:425–43
- Vazquez A, Pastor-Satorras R, Vespignani A. 2002. Large-scale topological and dynamical properties of the Internet. *Phys. Rev. E* 65:066130
- Vendruscolo M, Dokholyan NV, Paci E, Karplus M. 2002. Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E* 65:061910
- Wagner A. 2001. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol. Biol. Evol.* 18:1283–92
- Wagner A, Fell DA. 2001. The small world inside large metabolic networks. *Proc. R. Soc. London Ser. B* 268:1803–10
- Walker HA, Thye SR, Simpson B, Lovaglia MJ, Willer D, Markovsky B. 2000. Network exchange theory: recent developments and new directions. *Soc. Psychol. Q.* 63:324–36
- Warren CP, Sander LM, Sokolov IM. 2002. Geography in a scale-free network model. *Phys. Rev. E* 66:056105
- Wasserman S, Faust K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge Univ. Press
- Wasserman S, Pattison P. 1996. Logit models and logistic regressions for social networks. 1. An introduction to Markov graphs and p. *Psychometrika* 61:401–25
- Watts DJ. 1999. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton, NJ: Princeton Univ. Press
- Watts DJ. 2002. A simple model of information cascades on random networks. *Proc. Natl. Acad. Sci. USA* 99:5766–71

- Watts DJ. 2003. *Six Degrees: The Science of a Connected Age*. New York: Norton
- Watts DJ, Dodds PS, Newman MEJ. 2002. Identity and search in social networks. *Science* 296:1302–5
- Watts DJ, Strogatz SH. 1998. Collective dynamics of 'small-world' networks. *Nature* 393:440–42
- West DB. 1996. *Introduction to Graph Theory*. Upper Saddle River, NJ: Prentice Hall
- White DR, Harary F. 2001. The cohesiveness of blocks in social networks: connectivity and conditional density. *Sociol. Methodol.* 31:305–59
- Wilf H. 1994. *Generating Functionology*. London: Academic
- Willer D. 1999. *Network Exchange Theory*. Westport, CT: Praeger
- Williamson OE. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press
- Wu F, Huberman BA, Adamic IA, Tyler JR. 2004. Information flow in social groups. *Physica A: Stat. Mech. Appl.* In press
- Wuchty S. 2002. Interaction and domain networks of yeast. *Proteomics* 2:1715–23
- Yook SH, Jeong HW, Barabasi AL. 2002. Modeling the Internet's large-scale topology. *Proc. Natl. Acad. Sci. USA* 99:13382–86
- Young HP. 1998. Conventional contracts. *Rev. Econ. Stud.* 65:773–92
- Yule GU. 1925. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. *Philos. Trans. R. Soc. London Ser. B* 213:21–87
- Zipf GK. 1949. *Human Behaviour and the Principle of Least Effort*. Reading, MA: Addison-Wesley