Exploring complex networks

Steven H. Strogatz

Department of Theoretical and Applied Mechanics and Center for Applied Mathematics, 212 Kimball Hall, Cornell University, Ithaca, New York 14853-1503, USA (e-mail: strogatz@cornell.edu)

The study of networks pervades all of science, from neurobiology to statistical physics. The most basic
issues are structural: how does one characterize the wiring diagram of a food web or the Internet or
the metabolic network of the bacterium Escherichia coli? Are there any unifying principles underlying their
topology? From the perspective of nonlinear dynamics, we would also like to understand how an enormous
network of interacting dynamical systems — be they neurons, power stations or lasers — will behave
collectively, given their individual dynamics and coupling architecture. Researchers are now only beginning
to unravel the structure and dynamics of complex networks.

Networks are on our minds nowadays. Sometimes we fear their power — and with
good reason. On 10 August 1996, a fault in
two power lines in Oregon led, through a
cascading series of failures, to blackouts in 11
US states and two Canadian provinces, leaving about 7
million customers without power for up to 16 hours1. The
Love Bug worm, the worst computer attack to date, spread
over the Internet on 4 May 2000 and inflicted billions of
dollars of damage worldwide.

In our lighter moments we play parlour games about
connectivity. 'Six degrees of Marlon Brando' broke out as a
nationwide fad in Germany, as readers of Die Zeit tried to
connect a falafel vendor in Berlin with his favourite actor
through the shortest possible chain of acquaintances2. And
during the height of the Lewinsky scandal, the New York
Times printed a diagram3 of the famous people within 'six
degrees of Monica'.

Meanwhile scientists have been thinking about
networks too. Empirical studies have shed light on the
topology of food webs4-9, electrical power grids, cellular and
metabolic networks10-12, the World-Wide Web13, the Internet
backbone14, the neural network of the nematode worm
Caenorhabditis elegans15, telephone call graphs16, coauthor-
ship and citation networks of scientists17-19, and the
quintessential 'old-boy' network, the overlapping boards of
directors of the largest companies in the United States20
(Fig. 1). These databases are now easily accessible, courtesy
of the Internet. Moreover, the availability of powerful
computers has made it feasible to probe their structure;
until recently, computations involving million-node
networks would have been impossible without specialized
facilities.

Why is network anatomy so important to characterize?
Because structure always affects function. For instance, the
topology of social networks affects the spread of informa-
tion and disease, and the topology of the power grid affects
the robustness and stability of power transmission.

From this perspective, the current interest in networks is
part of a broader movement towards research on complex
systems. In the words of E. O. Wilson21, "The greatest
challenge today, not just in cell biology and ecology but in all
of science, is the accurate and complete description of
complex systems. Scientists have broken down many kinds
of systems. They think they know most of the elements
and forces. The next task is to reassemble them, at least
in mathematical models that capture the key properties of
the entire ensembles."
the nervous system can be strong or weak, inhibitory or excitatory.

4. Dynamical complexity: the nodes could be nonlinear dynamical systems. In a gene network or a Josephson junction array, the state of each node can vary in time in complicated ways.

5. Node diversity: there could be many different kinds of nodes. The biochemical network that controls cell division in mammals consists of a bewildering variety of substrates and enzymes, only a few of which are shown in Fig. 1c.

6. Meta-complication: the various complications can influence each other. For example, the present layout of a power grid depends on how it has grown over the years—a case where network evolution affects topology. When coupled neurons fire together repeatedly, the connection between them is strengthened; this is the basis of memory and learning. Here nodal dynamics affect connection weights.

To make progress, different fields have suppressed certain complications while highlighting others. For instance, in nonlinear dynamical systems coupled together in simple, geometrically regular ways. Furthermore we usually assume that the network architecture is static. These simplifications allow us to sidestep any issues of structural complexity and to concentrate instead on the system’s potentially formidable dynamics.

Laser arrays provide a concrete example. In the single-mode approximation, each laser is characterized by its time-dependent gain, polarization, and the phase and amplitude of its electric field. These evolve according to four coupled, nonlinear differential equations. We usually hope the laser will settle down to a stable state, corresponding to steady emission of light, but periodic pulsations and even chaotic intensity fluctuations can occur in some cases.

Now suppose that many identical lasers are arranged side by side in a regular chain or ring, interacting with their neighbours by evanescent coupling or by overlap of their electric fields. Will the lasers lock their phases together spontaneously, or break up into a standing wave pattern, or beat each other into incoherence? From a technological standpoint, self-synchronization would be the most desirable outcome, because a perfectly coherent array of \( N \) lasers would allow us to exploit the potentialities of each laser in isolation.

**Figure 1** Wiring diagrams for complex networks. **a.** Food web of Little Rock Lake, Wisconsin, currently the largest food web in the primary literature. Nodes are functionally distinct ‘trophic species’ containing all taxa that share the same set of predators and prey. Height indicates trophic level with mostly phytoplankton at the bottom and fishes at the top. Cannibalism is shown with self-loops, and omnivory (feeding on more than one trophic level) is shown by different coloured links to consumers. (Figure provided by N. D. Martinez.) **b.** New York State electric power grid. Generators and substations are shown as small blue bars. The lines connecting them are transmission lines and transformers. Line thickness and colour indicate the voltage level: red, 765 kV and 500 kV; brown, 345 kV; green, 230 kV; grey, 138 kV and below. Pink dashed lines are transformers. (Figure provided by J. Thorp and H. Wang.) **c.** A portion of the molecular interaction map for the regulatory network that controls the mammalian cell cycle. Colours indicate different types of interactions: black, binding interactions and stoichiometric conversions; red, covalent modifications and gene expression; green, enzyme actions; blue, stimulations and inhibitions. (Reproduced from Fig. 6a in ref. 6, with permission. Figure provided by K. Kohn.)
The state of each oscillator is represented geometrically as a dot in the complex plane. The amplitude and phase of the oscillation correspond to the radius and angle of the dot in polar coordinates. Colours code the oscillators’ natural frequencies, running from slowest (red) to fastest (violet).

In the absence of coupling, each oscillator would settle onto its limit cycle (circle) and rotate at its natural frequency. However, here all the oscillators are also pulled towards the mean field that they generate collectively (shown as an asterisk at the centre of the population). Time increases from left to right, and from top to bottom. Starting from a random initial condition, the oscillators self-organize by collapsing their amplitudes; then they sort their phases so that the fastest oscillators are in the lead. Ultimately they all rotate as a synchronized pack, with locked amplitudes and phases. The governing equations describe a mean-field model of a laser array. (Simulation provided by R. Oliva.)

Figure 2  Spontaneous synchronization in a network of limit-cycle oscillators with distributed natural frequencies. The state of each oscillator is represented geometrically as a dot in the complex plane. The amplitude and phase of the oscillation correspond to the radius and angle of the dot in polar coordinates. Colours code the oscillators’ natural frequencies, running from slowest (red) to fastest (violet).

In the absence of coupling, each oscillator would settle onto its limit cycle (circle) and rotate at its natural frequency. However, here all the oscillators are also pulled towards the mean field that they generate collectively (shown as an asterisk at the centre of the population). Time increases from left to right, and from top to bottom. Starting from a random initial condition, the oscillators self-organize by collapsing their amplitudes; then they sort their phases so that the fastest oscillators are in the lead. Ultimately they all rotate as a synchronized pack, with locked amplitudes and phases. The governing equations describe a mean-field model of a laser array. (Simulation provided by R. Oliva.)
analogue of a phase transition. He proposed a mean-field model of nearly identical, weakly coupled limit-cycle oscillators and showed that when the coupling is small compared to the spread of natural frequencies, the system behaves incoherently, with each oscillator running at its natural frequency. As the coupling is increased, the incoherence persists until a certain threshold is crossed — then a small cluster of oscillators suddenly ‘freezes’ into synchrony. For still greater coupling, all the oscillators become locked in phase and amplitude (Fig. 2).

Kuramoto refined this connection between nonlinear dynamics and statistical physics. He proposed an exactly solvable model of collective synchronization, given by

$$\frac{d\theta_i}{dt} = \omega_i + \frac{K}{N} \sum_{j=1}^{N} \sin(\theta_j - \theta_i), \; i = 1, \ldots, N$$

where $$\theta_i(t)$$ is the phase of the $$i$$th oscillator and $$\omega_i$$ is its natural frequency, chosen at random from a lorentzian probability density

$$g(\omega) = \frac{\gamma}{\pi[\gamma^2 + (\omega - \omega_0)^2]}$$

of width $$\gamma$$ and mean $$\omega_0$$. Using an ingenious self-consistency argument, Kuramoto solved for the order parameter

$$r(t) = \left| \frac{1}{N} \sum_{i=1}^{N} e^{i\theta_i(t)} \right|$$

(a convenient measure of the extent of synchronization) in the limit $$N \to \infty$$ and $$t \to \infty$$. He found that

$$r = \begin{cases} 0, & K < K_c \\ \sqrt{1 - (K/K_c)}, & K \geq K_c \end{cases}$$

where $$K_c = 2\gamma$$. In other words, the oscillators are desynchronized completely until the coupling strength $$K$$ exceeds a critical value $$K_c$$. After that, the population splits into a partially synchronized state.
consisting of two groups of oscillators: a synchronized group that contributes to the order parameter \( r \), and a desynchronized group whose natural frequencies lie in the tails of the distribution \( g(\omega) \) and are too extreme to be entrained. With further increases in \( K \), more and more oscillators are recruited into the synchronized group, and \( r \) grows accordingly.

Twenty-five years later, the Kuramoto model continues to surprise us (see ref. 45 for a review). First, the incoherent state with \( r = 0 \) was found to be neutrally stable below threshold, despite its apparent stability in simulations; the analysis reveals a connection to Landau damping in plasmas. Second, the square-root critical behaviour of \( r \), almost a cliché for mean-field models in statistical mechanics, turns out to be non-generic; if the sinusoidal coupling is replaced by a periodic function with second harmonics, the scaling changes to \( r \sim K - K_0 \). Third, although the model was motivated originally by biological oscillators, it has appeared in such far-flung settings as the flavour evolution of neutrinos\(^{32}\), and arrays of Josephson junctions\(^{27}\) and semiconductor lasers\(^{24}\).

The main unsolved problem is the stability of the partially synchronized state for \( K > K_0 \). Numerical simulations indicate that it is globally stable, in the sense that it attracts almost all solutions, but even the linear stability problem has yet to be solved. Another issue concerns the extension of the model to nearest-neighbour coupling on a d-dimensional cubic lattice. Simulations\(^{46}\) and renormalization arguments\(^{47}\) indicate that the synchronization phase transition persists for \( d > 3 \) and vanishes for \( d = 1 \); the results are ambiguous for \( d = 2 \). All of this awaits a mathematical resolution.

In contrast to the mean-field models of Winfree and Kuramoto, Ermentrout and Kopell’s classic work deals with one-dimensional chains of oscillators, first in connection with neuromuscular rhythms in the mammalian intestine\(^{48}\), and later in their model of the central pattern generator for the lamprey\(^{49,50}\). The main phenomena here involve travelling waves, rather than the synchrony found in mean-field models. This is not accidental, as wave propagation is essential for the generation of peristalsis in the intestine, and for the creation of the swimming rhythm in lamprey.

Ermentrout and Kopell introduced several deep mathematical innovations, but perhaps their most impressive result is a counterintuitive biological prediction. Their lamprey model suggested that the tail-to-head neural connections along the spinal cord would be stronger than those running from head to tail, despite the fact that the wave associated with swimming travels from head to tail. To everyone’s delight, that prediction was later confirmed by their experimental collaborators\(^{51}\).

**Complex network architectures**

All the network topologies discussed so far — chains, grids, lattices and fully-connected graphs — have been completely regular (Fig. 3a, b). Those simple architectures allowed us to focus on the complexity caused by the nonlinear dynamics of the nodes, without being burdened by any additional complexity in the network structure itself. Now I take the complementary approach, setting dynamics aside and turning to more complex architectures. A natural place to start is at the opposite end of the spectrum from regular networks, with graphs that are completely random.

**Random graphs**

Imagine \( m \) buttons strewn across the floor\(^{51}\). Pick two buttons at random and tie them together with thread. Repeat this process \( m \) times, always choosing pairs of buttons at random. (If \( m \) is large, you might eventually select buttons that already have threads attached. That is certainly allowed; it merely creates clusters of connected buttons.) The result is a physical example of a random graph with \( n \) nodes and \( m \) links (Fig. 3c). Now slowly lift a random button off the floor. If it is tied to other buttons, either directly or indirectly, those are dragged up too. So what happens? Are you likely to pull up an isolated button, a small cluster or a vast meshwork?

Erdős and Rényi\(^{28}\) studied how the expected topology of this random graph changes as a function of \( m \). When \( m \) is small, the graph is likely to be fragmented into many small clusters of nodes, called components. As \( m \) increases, the components grow, at first by linking to isolated nodes and later by coalescing with other components. A phase transition occurs at \( m = n/2 \), where many clusters crosslink spontaneously to form a single giant component. For \( m > n/2 \), this giant component contains on the order of \( n \) nodes and \( m \) links, roughly the maximum number of ‘degrees of separation’ between any two nodes grows slowly, like \( \log n \).

In the decades since this pioneering work, random graphs have been studied deeply within pure mathematics\(^\text{33} \). They have also served as idealized coupling architectures for dynamical models of gene networks, ecosystems and the spread of infectious diseases and computer viruses\(^{29,35,34,35}\).

**Small-world networks**

Although regular networks and random graphs are both useful idealizations, many real networks lie somewhere between the extremes of order and randomness. Watts and Strogatz\(^{26,32}\) studied a simple model that can be tuned through this middle ground: a regular lattice

---

**Table 1 Clustering for three affiliation networks**

<table>
<thead>
<tr>
<th>Network</th>
<th>Clustering C</th>
<th>Actual</th>
<th>Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company directors</td>
<td>0.590</td>
<td>0.588</td>
<td></td>
</tr>
<tr>
<td>Movie authors</td>
<td>0.084</td>
<td>0.199</td>
<td></td>
</tr>
<tr>
<td>Biomedical authors</td>
<td>0.042</td>
<td>0.088</td>
<td></td>
</tr>
</tbody>
</table>

US corporate directors: 7,673 company directors linked by joint membership on 914 boards of the Fortune 1,000 companies for 1999. Movie actors: 449,913 actors linked by mutual appearances in 151,261 feature films, as specified by the Internet Movie Database (www.imdb.com) as of 1 May 2000. Biomedical collaborations: 1,388,989 scientists linked by coauthorship of at least one of 2,156,769 biomedical journal articles published between 1995 and 1999 inclusive, as listed in the MEDLINE database. The clustering coefficient C is defined as the probability that a connected triple of nodes is actually a triangle; here nodes correspond to people, as in the unipartite representation shown in Fig. 7b. Intuitively, C measures the likelihood that two people who have a mutual collaborator are also collaborators of each other. The results show that the random model accurately predicts C for the corporate director network, given the network’s bipartite structure and its degree distributions; no additional social forces need to be invoked. For the networks of actors and scientists, the model accounts for about half of the observed clustering. The remaining portion depends on social mechanisms at work in those communities (see text). (Adapted from ref. 91.)
where the original links are replaced by random ones with some probability $0 < \phi < 1$. They found that the slightest bit of rewiring transforms the network into a ‘small world’, with short paths between any two nodes, just as in the giant component of a random graph. Yet the network is much more highly clustered than a random graph, in the sense that if $A$ is linked to $B$ and $B$ is linked to $C$, there is a greatly increased probability that $A$ will also be linked to $C$ (a property that sociologists call ‘transitivity’).

Watts and Strogatz conjectured that the same two properties — short paths and high clustering — would hold also for many natural and technological networks. Furthermore, they conjectured that dynamical systems coupled in this way would display enhanced signal propagation speed, synchronizability and computational power, as compared with regular lattices of the same size. The intuition is that the short paths could provide high-speed communication channels between distant parts of the system, thereby facilitating any dynamical process (like synchronization or computation) that requires global coordination and information flow.

Research has proceeded along several fronts. Many empirical examples of small-world networks have been documented, in fields ranging from cell biology to business11,14,10–64. On the theoretical side, small-world networks are turning out to be a Rorschach test — how to actually find the shortest path? If $\ell$ denotes that average separation, we find that $\ell$ drops sharply near $\phi = 0$, confirming that a few shortcuts do indeed shrink the world dramatically. One of the most striking results is the following formula derived by Newman, Moore and Watts56:

$$\ell = \frac{N}{k} f(n k \phi)$$

where

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

This solution is asymptotically exact in the limits $n \to \infty$ (large system size) and either $n k \phi \to \infty$ or $n k \phi \to 0$ (large or small number of
The simplest random graph models predict a bell-shaped Poisson distribution for $p_k$. But for many real networks, $p_k$ is highly skewed and decays much more slowly than a Poisson. For instance, the distribution decays as a power law $\propto k^{-3}$. More sophisticated models include the effects of adding or rewiring links, allowing nodes to age so that they can no longer accept new links, or varying the form of preferential attachment. These generalized models predict exponential and truncated power-law $p_k$ in some parameter regimes, as well as scale-free distributions.

Could there be a functional advantage to scale-free architecture? Albert, Jeong and Barabási suggested that scale-free networks are resistant to random failures because a few hubs dominate their topology (Fig. 3d). Any node that fails probably has small degree (like most nodes) and so is expendable. The flip side is that such networks are vulnerable to deliberate attacks on the hubs. These intuitive ideas have been confirmed numerically and analytically by examining how the average path length and size of the giant component depend on the number and degree of the nodes removed. Some possible implications for the resilience of the Internet, the design of therapeutic drugs, and the evolution of metabolic networks have been discussed.

Generalized random graphs

As mentioned above, the simplest random graph predicts a Poisson degree distribution, and so cannot accommodate the other types of distribution found in real networks. Molloy and Reed introduced a more flexible class of random graphs in which any degree distribution is permitted. Given a sequence of non-negative integers $\{d_k\}$, where $d_k$ denotes the number of nodes with degree $k$, consider the ensemble of all graphs with that prescribed degree sequence, and weight them all equally when computing statistical averages of interest. For this class of graphs, Molloy and Reed derived a simple condition for the birth of the giant component, and they also found an implicit formula for its size as a fraction of $n$, the total number of nodes. Specifically, let $n \gg 1$ and define

$$Q = \sum_{k=1}^{\infty} p_k (k - 2)$$

where $p_k = d_k/n$. If $Q < 0$, the graph consists of many small components. The average component size diverges as $Q \to 0$ from below, and a giant component exists for $Q > 0$. (In technical terms, these results hold 'almost surely'; that is, with probability tending to 1 as $n \to \infty$.)

Aiello, Chung and Lu applied these results to a random graph model for scale-free networks. For $p_k$ of power-law form, the condition on $Q$ implies that a giant component exists if and only if $\gamma < 3.47$, which holds for most scale-free networks measured so far. If $\gamma < 1$, there are so many high-degree hubs that the network forms one huge, connected piece. They also proved theorems about the number and size of small components outside the giant component, and compared these to a real graph of about 47 million telephone numbers correctly and the calls between them in one day. They found that the data are best fit by an exponent $\gamma \approx 2.1$, which predicts correctly that the call graph is not connected but has a giant component.

Nevertheless, the scale-free case has stimulated a great deal of theorizing. The earliest work is due to Simon in 1955, now independently rediscovered by Barabási, Albert and Jeong. They showed that a heavy-tailed degree distribution emerges automatically from a stochastic growth model in which new nodes are added continuously and attach themselves preferentially to existing nodes, with probability proportional to the degree of the target node. Richly connected nodes get richer, and the result is $p_k \propto k^{-3}$. More sophisticated models include the effects of adding or rewiring links, allowing nodes to age so that they can no longer accept new links, or varying the form of preferential attachment. These generalized models predict exponential and truncated power-law $p_k$ in some parameter regimes, as well as scale-free distributions.

Could there be a functional advantage to scale-free architecture? Albert, Jeong and Barabási suggested that scale-free networks are resistant to random failures because a few hubs dominate their topology (Fig. 3d). Any node that fails probably has small degree (like most nodes) and so is expendable. The flip side is that such networks are vulnerable to deliberate attacks on the hubs. These intuitive ideas have been confirmed numerically and analytically by examining how the average path length and size of the giant component depend on the number and degree of the nodes removed. Some possible implications for the resilience of the Internet, the design of therapeutic drugs, and the evolution of metabolic networks have been discussed.

Generalized random graphs

As mentioned above, the simplest random graph predicts a Poisson degree distribution, and so cannot accommodate the other types of distribution found in real networks. Molloy and Reed introduced a more flexible class of random graphs in which any degree distribution is permitted. Given a sequence of non-negative integers $\{d_k\}$, where $d_k$ denotes the number of nodes with degree $k$, consider the ensemble of all graphs with that prescribed degree sequence, and weight them all equally when computing statistical averages of interest. For this class of graphs, Molloy and Reed derived a simple condition for the birth of the giant component, and they also found an implicit formula for its size as a fraction of $n$, the total number of nodes. Specifically, let $n \gg 1$ and define

$$Q = \sum_{k=1}^{\infty} p_k (k - 2)$$

where $p_k = d_k/n$. If $Q < 0$, the graph consists of many small components. The average component size diverges as $Q \to 0$ from below, and a giant component exists for $Q > 0$. (In technical terms, these results hold 'almost surely'; that is, with probability tending to 1 as $n \to \infty$.)

Aiello, Chung and Lu applied these results to a random graph model for scale-free networks. For $p_k$ of power-law form, the condition on $Q$ implies that a giant component exists if and only if $\gamma < 3.47$, which holds for most scale-free networks measured so far. If $\gamma < 1$, there are so many high-degree hubs that the network forms one huge, connected piece. They also proved theorems about the number and size of small components outside the giant component, and compared these to a real graph of about 47 million telephone numbers correctly and the calls between them in one day. They found that the data are best fit by an exponent $\gamma \approx 2.1$, which predicts correctly that the call graph is not connected but has a giant component.

Nevertheless, the scale-free case has stimulated a great deal of theorizing. The earliest work is due to Simon in 1955, now independently rediscovered by Barabási, Albert and Jeong. They showed that a heavy-tailed degree distribution emerges automatically from a stochastic growth model in which new nodes are added continuously and attach themselves preferentially to existing nodes, with probability proportional to the degree of the target node. Richly connected nodes get richer, and the result is $p_k \propto k^{-3}$. More sophisticated models include the effects of adding or rewiring links, allowing nodes to age so that they can no longer accept new links, or varying the form of preferential attachment. These generalized models predict exponential and truncated power-law $p_k$ in some parameter regimes, as well as scale-free distributions.
The bipartite case is especially interesting for applications. By definition, in a bipartite graph there are two types of nodes, with links running only between different kinds (Fig. 7). For example, consider the network of the boards of directors of the Fortune 1,000 companies, the largest US corporations ranked according to revenues. This network is fascinating because the boards ‘interlock’ — some important people sit on several of them — and this overlap knits virtually all large US firms together into a giant web of corporate governance.

Let \( p_j \) denote the probability that a director sits on exactly \( j \) boards, and let \( q_k \) denote the probability that a board consists of \( k \) directors. Figure 8 shows that \( p_j \) is approximately exponential, with most directors sitting on only one board, whereas \( q_k \) is strongly peaked around \( k = 10 \), indicating that most boards have about ten members. As a null hypothesis, assume that the Fortune 1,000 network is a random member of the ensemble of all bipartite graphs with the same \( p_j \) and \( q_k \). Then generating functions yield predictions for various quantities of interest. For example, suppose we want to calculate \( r_z \), the probability that a random director works with a total of \( z \) other co-directors, summed over all the boards on which he or she serves. Let

\[
f_j(x) = \sum_{j=0}^{\infty} p_j x^j
\]

\[
g_k(x) = \sum_{k=0}^{\infty} q_k x^k
\]

be the generating functions associated with the empirical degree distributions \( p_j \) and \( q_k \). If we now choose a random edge on the bipartite graph and follow it to the board at one of its ends, the distribution of the number of other edges leaving that board can be shown to be generated by \( g_k(x) = g_k^r(x) n \), where \( n = g_k^r(1) \). Then for a randomly chosen director, the generating function for \( z \) is given by \( G_z(x) = f_z(g_k(x)) \). If we expand \( G_z \) in a series as

\[
G_z(x) = \sum_{z=0}^{\infty} r_z x^z,
\]

the coefficients \( r_z \) are exactly the quantities we seek. They can be extracted by repeated differentiation:

\[
r_j = \frac{1}{j!} \frac{d^j G_z}{dx^j} \bigg|_{x=0}
\]

Figure 8c shows that the predicted \( r_z \) agrees almost perfectly with the actual distribution. Similarly, the clustering coefficient \( c_z \) predicted for the directors lies within 1% of the observed value (Table 1). Clearly the random model captures much of the structure of the real network.

However, for two other bipartite graphs — film actors and the movies they appeared in, and biomedical scientists and the papers they coauthored — the model underestimates the clustering coefficients by half (Table 1). The reason is that the random model quantifies only the generic portion of the clustering; it reflects the cliques that are formed automatically whenever a bipartite collaboration graph is projected onto the space of people, as in Fig. 7b. For the corporate board data, those cliques account for essentially all the clustering (simply because most directors sit on only one board, thus preventing clustering across boards). But for the scientists and actors, some further mechanisms must be at work. One possible explanation is that scientists tend to introduce pairs of their collaborators to each other, engendering new collaborations.

In this way the random model allows us to disentangle the generic features of bipartite graphs from those that could reflect sociological effects. Beyond their benchmarking role, generalized random graphs provide a promising new class of substrates on which dynamical processes can be simulated and even approached analytically. Using this approach, Watts has given an intriguing explanation of fads and normal accidents as the natural consequence of cascade dynamics on sparse interaction networks.

**Outlook**

In the short run there are plenty of good problems about the nonlinear dynamics of systems coupled according to small-world, scale-free or generalized random connectivity. The speculations that these architectures are dynamically advantageous (for example, more synchronization or error-tolerant) need to be sharpened, then confirmed or refuted mathematically for specific examples. Other ripe topics include the design of self-healing networks, and the relationships among optimization principles, network growth rules and network topology.

In the longer run, network thinking will become essential to all branches of science as we struggle to interpret the data pouring in from neurobiology, genomics, ecology, finance and the World-Wide Web. Will theory be able to keep up? Time to log back on to the Internet...

---
