from *SIAM News*, Volume 34, Number 9

Community Lecture 2001

The Continuing Appeal of Small-world Networks

By James Case

The I.E. Block Community Lecture has become a highlight of the SIAM Annual Meeting; this year's lecture, given in San Diego on July 11 by Steve Strogatz of Cornell University, was no exception. Titled "Collective Dynamics of Small-world Networks," the lecture concerned a loosely defined class of networks that seems to include many if not most of the ones science and industry would like to learn more about. Many of the results presented were first obtained by Strogatz in collaboration with Duncan Watts, then a graduate student at Cornell and currently at Columbia University (Department of Sociology) and the Santa Fe Institute.

Large sparse networks are all around us. Of the six billion people in the world, few are acquainted with more than a thousand others. The "global acquaintance graph" thus contains fewer than 3×10^{12} edges and is sparse in comparison with the complete graph on six billion nodes, which has about six million times as many edges. A brain, with about ten billion neurons, each connected to perhaps ten thousand others, has a comparably sparse wiring diagram. The Internet links millions of computers, relatively few of which ever communicate directly with one another, meaning that the "Internet connection graph" is also large and sparse.

The foregoing networks are partially random, in that people make unlikely acquaintances, and neurons and computers make chance long-range connections. Yet people know their neighbors, neurons connect to nearby neurons, and computers are linked in local networks far more often than if such connections were completely random. Practically important networks, then, seldom resemble either totally random graphs or the completely regular graphs (such as chains, rings, grids, lattices, trivalent maps, and complete graphs) of traditional graph theory.

A random graph of order n is nothing more than a set of n vertices, together with an edge set that is generated in some random fashion. But random graph theory as it currently exists is mainly concerned with edge sets generated in either of two specific ways. After $n(n - 1)/2$ numbered balls—representing the edges of the complete graph on n nodes—have been placed in an urn, either m of them are withdrawn at random, or all the balls are withdrawn in succession, each one retained or discarded with a probability of p or $1 - p$, respectively. Many of the fundamental properties of the resulting classes of random graphs, along with the techniques for analyzing them, were developed during the late 1950s and early 1960s by Alfred Rényi and Paul Erdös. Statistically speaking, there is little difference between the resulting classes of graphs as long as n is large and $m = np$. The results can be compared with graphs of commercial or scientific importance, large numbers of which can now be constructed from Internet data.

Many such graphs—representing road maps, food chains, electric power grids,cellular, metabolic, or neural networks, telephone call graphs, influence networks, and the like—are large and sparse. More than a few resemble those discussed by Watts and Strogatz in that they share some traits with random graphs and others with regularly structured graphs.

The large sparse graphs discussed in San Diego by I.E. Block Community Lecturer Steve Strogatz (above) represent real-world networks ranging from electric power grids to "influence networks"; these "small-world" graphs, as discussed in the accompanying article, "occupy an intermediate position in the graph-theoretic firmament, between the perfectly regular and the totally random extremes."
Another San Diego highlight (right): Vincent Blondel of l'Université Catholique de Louvain received the SIAM Activity Group on Control and Systems Theory Prize, which is awarded every three years in recognition of outstanding work by young researchers. Blondel, shown here with SIAG/CST vice-chair Mary Ann Horn, has studied the computational complexity of a variety of control problems. He was cited for addressing "fundamental problems in systems and control theory from a novel point of view, involving a creative combination of disparate tools." In certain contexts, according to the prize citation, Blondel's research "delineates the limitations of mathematical analysis and computation."

Three statistics have proven particularly fruitful in the study of large sparse networks: (1) the average k over all vertices v of the number of edges incident on v; (2) the average over all connected pairs of distinct vertices of the length L (measured in edges) of the shortest connecting path; and (3) the (dimensionless) frequency C with which three connected vertices are completely connected, meaning that each pair of vertices is connected by an edge. Obviously, $0 \le k \le n - 1$, $1 \le L \le n - 1$, and $0 \le C \le 1$. L is known as the "characteristic path length" of the graph in question; k is called the "average vertex degree," and C the "clustering coefficient." Graphs in which L is about as small as in a random graph with the same n,k, but in which C is much larger, are called "small-world graphs," in homage to the cliché traditionally invoked when people from distant parts of the world turn out to have friends in common.

The ultimate small-world graph is the so-called Kevin Bacon graph (KBG), in which the nodes represent actors who have appeared (any time, anywhere) in one or more feature films in which pairs of vertices are connected by an edge whenever the corresponding actors have appeared together in at least one feature film. Watts lists a number of factors that make the KBG worthy of serious study [3]. For one thing, the data are reliable. The entire history of motion pictures has resulted in the creation of only about 150,000 feature films, with a combined cast of about 300,000 actors, all of which are listed in a single searchable database (at www.us.imdb.com). Moreover, about 90% of the actors listed are part of a single *connected component* KBG* of the graph that includes about 225,000 actors in about 110,000 films. Strictly speaking, Watts and Strogatz's analysis of the KBG applies only to KBG*, since only connected graphs are of interest in the present line of inquiry.

KBG* is sufficiently large ($n = 225{,}226$) and sparse ($k \approx 61$) that L could conceivably vary over several orders of magnitude, while C might lie almost anywhere in the unit interval. Yet the graph is not so large that it cannot be stored and manipulated by a (suitably powerful) computer. Specifically, if n were even a single order of magnitude larger, it would be difficult to find *any* computer that could hold the connected component in memory at one time, as must be done to compute statistics like L and C with reasonable dispatch. The KBG is so named because (a) every actor who has ever appeared in an American-made film is connected to Bacon by a path of length four or less, and (b) every actor in the entire graph—whatever his or her nationality—is connected to him by a path of length at most eight. It is almost anticlimactic to learn that the decidedly more accomplished Rod Steiger has an even shorter average path length than Bacon to other members of KBG*.

Two other small-world networks investigated by Watts and Strogatz are the moderately large ($n = 282$) and sparse ($k = 14$) neural wiring diagram of the tiny worm known as *Caenorhabditis elegans*—*C. elegans* for short—and the electric power transmission grid for the several states (as well as portions of the Canadian provinces) located west of the Rocky Mountains, wherein $n = 4971$ and $k = 2.67$. The former graph is scientifically important, while the latter is commercially so, especially in a year of rolling brownouts. Like the KBG, both offer the advantages of tractable size and accurate documentation.

Empirical evidence suggests that even remarkably sparse random graphs have low values of L/n and C, while comparably sparse regularly structured graphs combine significantly higher values of both quantities. It might therefore be conjectured that the two statistics are correlated, causing the pairs $(L/n,C)$ obtained from a variety of large sparse graphs to cluster about a straight line of positive slope in the unit square. To demonstrate that this is not the case, Watts and Strogatz devised a scheme for redirecting, in turn—with probability p—each edge of a ring graph in which n vertices are arranged in a circle and each vertex is originally connected to each of its k nearest neighbors. A small example (for $n = 20$, $k = 4$) appears in Figure 1.

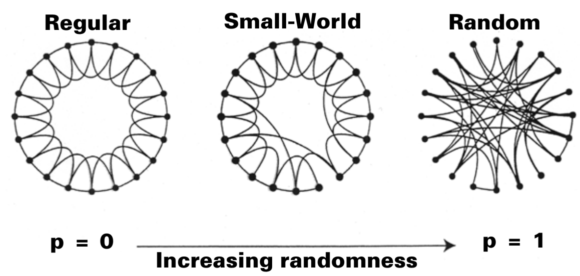

Figure 1. *Progression from regularity to randomness, beginning with a small ring graph in which n = 20, k = 4. (Figures 1 and 2 adapted from [3].)*

The ring graphs actually employed in the analysis were significantly larger. The results for small positive values of p are somewhat disordered versions of the originals in which approximately $nkp/2$ of the $nk/2$ original edges have been assigned new destinations. The degree of disorder obviously increases with p, becoming nearly total as p nears unity. The surprising results are tabulated in Figure 2, in which the ratios $L(p)/L(0)$ and $C(p)/C(0)$ are plotted against the parameter p, which is scaled exponentially for ease of interpretation. Whereas $L(p)$ falls off rapidly to something barely exceeding $L(1)$ as p increases from 0, $C(p)$ is almost constant for small values of p, indicating the existence of an entire interval of p-values in which $C(p)$ approximates $C(0)$ while $L(p)$ approximates $L(1)$. For such values of p, the corresponding graphs combine short characteristic lengths with nearly maximal clustering. By this construction, Watts and Strogatz give precise meaning—in at least one special case—to the notion that small-world graphs occupy an intermediate position in the graph-theoretic firmament, between the perfectly regular and the totally random extremes.

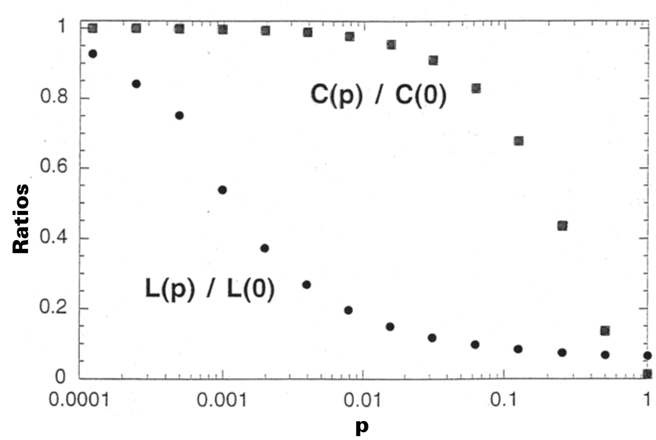

Figure 2. *Decline of the ratios C(p)/C(0) and L(p)/L(0) for the case n = 1000, k = 10.*

The combination of a short characteristic length and a high degree of clustering—which roughly distinguishes small-world networks from other large sparse networks—can be deadly, as in the case of certain diseases likely to be transmitted from one acquaintance to another. The short characteristic lengths facilitate transmission between remote clusters, while the high degree of clustering assures that few members of an infected cluster will remain uninfected. On the other hand, the same two characteristics are ideal for communication networks—either human or robotic—since they decrease the number of times a message must be "handled" en route to its destination.

A somewhat similar and particularly striking result was obtained by Newman, Moore, and Watts [1] when they considered a slightly different procedure for turning the original (completely regular) ring into a small-world graph. Instead of redirecting each edge in turn—with probability p—to a new destination, they simply inserted nkp shortcuts between randomly selected pairs of nodes to conclude that $L(p) \approx n*f(nkp)/k$, where

$$f(x) = \left[\tanh^{-1}\left(x/\sqrt{x^2+2x}\right)\right] / 2\sqrt{x^2+2x}.$$

The estimate is asymptotically correct in the limits $n \to \infty$ (corresponding to large system size) and either $nkp \to \infty$ or $nkp \to 0$ (corresponding to many or few shortcuts, respectively). Figure 3 indicates that the formula is also qualitatively correct for intermediate numbers of shortcuts, although the actual average shortest distances are slightly longer than the estimated ones.

The "old-boy" network to which the directors of the nation's largest companies belong can be described as a graph in which vertices represent directors, and pairs of vertices are joined by edges if and only if the corresponding directors sit together on one or more boards. Vernon Jordan turns out (or did as of 1999) to play the Rod Steiger role among the directors of companies included in the Fortune 1000—a list that ranks the largest U.S. firms according to revenue—by sitting on no fewer than nine distinct boards. Jordan is one of an elite group who sit on many boards, creating an "overlap" that unites virtually all large U.S. firms into a giant "web of corporate governance." Reformers once portrayed this web of "interlocking directorates" as a potent menace to society, although the public quickly lost interest.

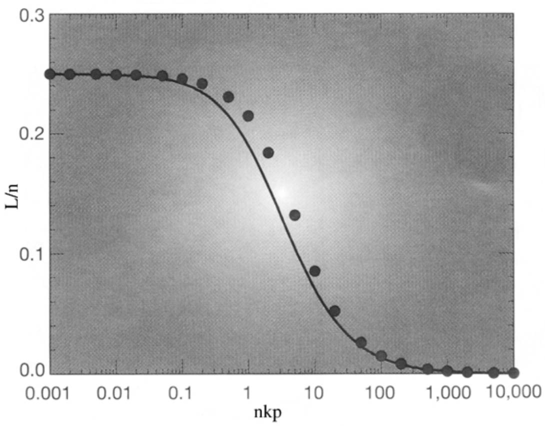

Figure 3. *Average path length, normalized by system size, plotted as a function of the average number of shortcuts (small circles) as compared with the theoretical formula (solid curve). (Figures 3–5 adapted from [2].)*

The data on interlocking directorates are more accurately recorded by a bipartite than a unipartite graph, as illustrated in Figure 4 for a situation involving four boards with a total of 16 members, 11 of whom are distinct. The unipartite display (Figure 4b) omits some of the information contained in Figure 4a, in that triangle FHI, which corresponds to board 3, looks for all the world like triangle FGI, which does not correspond to any of the boards.

Let p_j denote the probability that a director sits on exactly j boards, and let q_k denote the probability that a board includes exactly k members. The sequence $\{p_j\}$ is a rapidly decreasing one, with most directors sitting on only one Fortune 1000 board, while $\{q_k\}$ has a distinct peak at around 10 members. Assume, by way of a null hypothesis, that the Fortune 1000 network is a random member of the ensemble of all bipartite graphs with the same sequences $\{p_j\}$ and $\{q_k\}$. It seems natural to measure the degree of clustering in the "web of corporate governance" by determining the sequence $\{r_z\}$ of probabilities r_z that a randomly chosen board member sits with a total of z others on the several Fortune 1000 boards to which he or she belongs. To that end, let

$$f_0(x) = \Sigma_j \, p_j x^j$$

and

$$g_0(x) = \Sigma_k \, q_k x^k$$

be the generating functions associated with the empirical distributions $\{p_j\}$ and $\{q_k\}$. If we now choose a random edge, and follow it to its board (rather than its member) end, the distribution of the number of other edges emanating from that board can be shown to be generated by $g_1(x) = g_0'(x)/g_0'(1)$. For a randomly chosen director, the generating function for z is then

$$G_0(x) = f_0(g_1(x)) = \Sigma_z \, r_z x^z,$$

the coefficients of which can be determined by repeated differentiation: $r_z = (1/z!)(d^zG/dx^z)|_{x=0}$.

Figure 5 shows that the predicted sequence $\{r_z\}$ agrees closely with the actual distribution. Similarly, the predicted clustering coefficient agrees, to within 1%, with the observed value, suggesting that the proposed random model captures a good deal of the structure of the actual network. Yet it must be confessed that in two other bipartite networks—pairing film actors with the films they appeared in, and biomedical scientists with the papers they co-authored—the model underestimates the degree of clus-

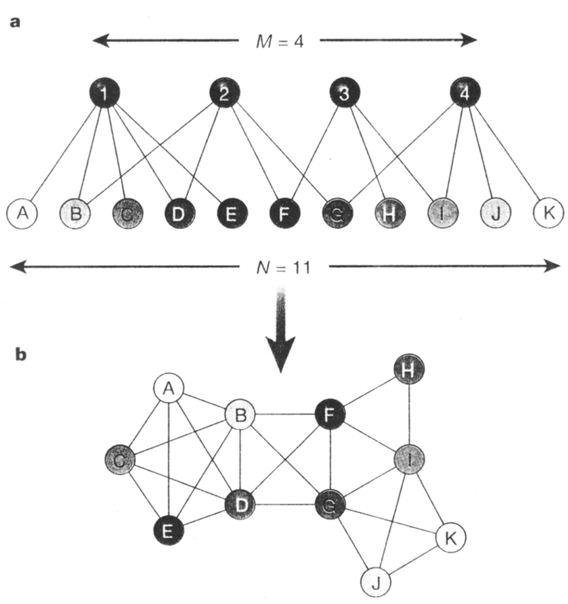

Figure 4. *Information is lost in the passage from a bipartite (a) to a unipartite (b) representation.*

tering by half.

Useful generalizations have already been made regarding large sparse real-world networks, and still more useful ones appear to await discovery. It should therefore become increasingly possible to distinguish between surprising properties of a particular large sparse network, and properties to be expected in any such graph. Similar networks are under investigation in numerous branches of science and technology, and closely related discoveries are being made with surprising regularity. Although it is probably too soon to unify so diverse a field, the first to try have much to show for their efforts.

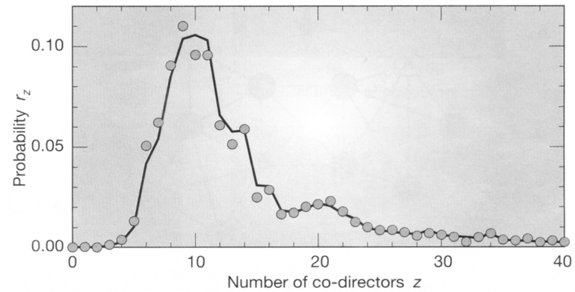

Figure 5. *Agreement between predicted and observed values of $\{r_z\}$.*

References

[1] M.E.J. Newman, C. Moore, and D.J. Watts, *Mean-field solution of the small-world model*, Phys. Rev. Lett., 84 (2000), 3201–3204.

[2] S.H. Strogatz, *Exploring complex networks*, Nature, 410 (2001), 268–276.

[3] D.J. Watts, *Small Worlds: The Dynamics of Networks Between Order and Randomness*, Princeton University Press, Princeton, NJ, 1999.

A review of [3], "Random Shortcuts Make It a Small World Indeed," by Gilbert Strang, appeared in SIAM News *in December 1999; http://www.siam.org/siamnews/12-99/shortcuts.pdf.*

James Case writes from Baltimore, Maryland.