The New Science of (Random) Networks

Two Books Reviewed by Rick Durrett

Linked: The New Science of Networks
Albert-László Barabási
Perseus Publishing, May 2002
256 pages, $26.00
ISBN 0-738-20667-9

Six Degrees: The Science of a Connected Age
Duncan J. Watts
W. W. Norton & Company, February 2003
368 pages, $27.95
ISBN 0-393-04142-5

At the beginning of the 21st century, a maverick group of scientists is discovering that all networks—from a cocktail party to a terrorist cell to an international conglomerate—have a deep underlying order and operate according to simple but powerful rules. Knowledge of the structure and behavior of these networks illuminates everything from the vulnerability of economies to the way diseases are spread, allowing us to design the "perfect" business or stop an outbreak before it goes global.

This paragraph is from the press release for the book whose full title is: *Linked: The New Science of Networks. How Everything is Connected to Everything Else and What it Means for Science, Business, and Everyday Life.* The content of the book is considerably less hyperbolic than the press release, but Barabási, who publishes frequently in *Nature* and *Science*, speaks with great confidence about the far-reaching implications of his results. The second book under review, *Six Degrees: The Science of a Connected Age* by Duncan J. Watts, has a more modest tone that mathematicians will be more comfortable with.

Having started at the end of the story, we will now go back more than forty years in time and across the Atlantic to Hungary to explain the developments that led up to this point. Along the way we will give references to let you know who did what and where you can learn more. Due to the laziness of the reviewer, there is no list of references here. However, most of them are contained in the 429 works cited in Mark Newman's recent article [1] in *SIAM Review*.

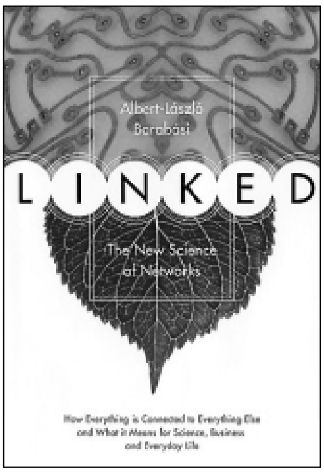
In a series of papers in the late 1950s, Erdős and Renyi introduced the simplest random graph model: $G(n, p)$. There are n vertices, and each of the $n(n-1)/2$ possible edges between vertices are independently present with probability p. When $p = c/n$ and n is large, there is a dramatic change in the connectedness of the graph at $c = 1$. If $c < 1$, then the connected component containing a given vertex will have an average of $1/(1-c)$ members; while if $c > 1$, it will with positive probability belong to a giant component of size $g(c)n$ where $g(c)$ is a constant that depends on c. To foreshadow later developments, we note that when $c > 1$ the world is small. One can go between two vertices in the giant component in $(1 + o(1))ln(n)/ln(c)$ steps.

The next scene in our story takes place in 1967. Stanley Milgram, a Harvard social psychologist, was interested in the social distance between any two people in the United States. He gave letters to a few hundred randomly selected people in Omaha, Nebraska. The letters were to be sent toward a target person, a stockbroker in Boston, but recipients could send the letters only to someone they knew

Rick Durrett is professor of mathematics at Cornell University. His email address is `rtd1@cornell.edu`.

on a first-name basis. The median number of steps the letters took to reach their destination was 5.5. Rounding up gives rise to the now famous phrase "six degrees of separation", which comes, not from Milgram's work, but from the title of John Guare's 1991 play.

The neat story in the last paragraph becomes a little more complex when one looks at the details. One third of the test subjects were from Boston, not Omaha, and one half of those in Omaha were stockbrokers. A large fraction of the letters never reached their destination, and those that did provide only an upper bound on the distance between the two individuals. Indeed, computing distances in real networks is difficult. As Watts mentions in his book, it took Steve Strogatz (Watts's Ph.D. advisor) two days to find that his Erdős number is four. Upper bounds of course are easy, but lower bounds require exhaustive search.



Jon Kleinberg (2000) has done an interesting study of navigation in small worlds, which shows that in most cases short paths are hard to find. A recent email version of Milgram's experiment carried out by Watts and coworkers suggests that people succeed in finding short paths by using features of the target person. For example, if I wanted to reach a policeman in a small town in France in a few steps, I would first send a message to a professor I know in Paris.

The next development in our chain of events is due to three college students scheming to get on Jon Stewart's radio talk show. They wrote to him: "We are three men on a mission, to prove to the world that Kevin Bacon is God." As they later explained on his show, if one says that two actors are "linked" if they appeared in a movie together, then almost every actor could be linked to Kevin Bacon with two or three links, another example of "six degrees". While this observation has made Kevin Bacon's name a household word, later analysis showed that he was far from the best choice. His average separation from everyone else in Hollywood is 2.79, the 876th best score among all actors. Rod Steiger is the winner with 2.51.

At about the time of these developments, Duncan Watts was a first-year graduate student at Cornell. "Like many grad students, I had harbored dreamy visions of life in a research university and was pretty disillusioned by the difficult and often dull reality. This guy Strogatz (then at MIT) had recently given a talk in my department—the first such talk that I felt I had actually understood—so I called him to see if he could take on a new research assistant."

At this point, if our story was a Hollywood movie, Watts and Strogatz would meet for coffee and emerge with their small-world model scribbled on a napkin. However, in the real world, Watts began first to work on a problem of synchronization of cricket songs and was only later led to the random graphs problem, which would become his Ph.D. thesis topic, by a remark his father made about six degrees of separation.

Watts and Strogatz took an interesting approach to their research. To quote several excerpts from Watts's description:

We were reasonably certain that someone must have thought about this problem before…but we also thought that if we went out looking for it, we might get discouraged by how much had already been done or else trapped into thinking about the problem from the same perspective. We met in Steve's office in January 1996 and made up our minds: we would go it alone. Steve insisted we only give it four months—a single semester—after which if we hadn't made some significant progress, we would concede defeat and return to crickets.

Things must have gone well, since their article "Collective dynamics of small-world networks" appeared in the June 4, 1998, issue of *Nature*. This article concerns what Watts calls their second or beta model. One starts with a one-dimensional ring of individuals, each of which is connected to a fixed number of neighbors in each direction, and then one rewires a fraction β of the connection to new randomly chosen individuals. The average path length between two individuals drops dramatically as β increases from 0. However, the clustering coefficient, roughly the fraction of your neighbors who are neighbors of each other, remains close to its original value. This feature (a) distinguishes the Watts-Strogatz model from the Erdős-Renyi one, in which almost all nongiant connected components are trees; and (b) matches the characteristics of the networks of film actors, the power grid of the western U.S., and the 282-neuron nervous system of the nematode *C. elegans*.

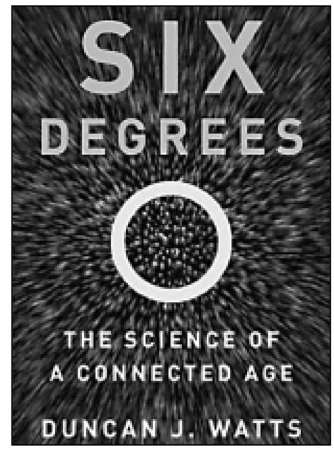
The pioneering article of Watts and Strogatz led to an explosion of results on the beta model and a variant due to Newman and Watts in which edges are added but none are taken away. This work also led to a number of other systems for which data were publicly available, such as boards of directors of Fortune 1000 companies and the World Wide Web. In a letter in the September 9, 1999, issue of *Nature*, Albert, Jeong, and Barabási declared that "despite its huge size, the Web is a highly connected graph with an average diameter of only 19 links." To find this number, they sent a software robot crawling the Web; it examined 325,729 nodes and found an average distance of 11.2. The three humans then extrapolated, assuming a logarithmic growth of maximum distance with number of nodes, to conclude that for 800 million Web pages, the diameter would be 19.

The next big idea in this area came from Barabási and Albert (1999). They noted that a common property of many large networks such as the World Wide Web or metabolic networks in genetics is that degrees of vertices follow a power law, and they introduced a simple model, called the preferential attachment model, that has this behavior. At each stage we add a new vertex with a fixed number of connections m to the rest of the network. The new vertex chooses the vertices to connect to with probabilities proportional to their degrees. In the long run the probability a vertex has degree $k \geq m$ is a power law $p(k) = 2m^2/k^3$. Some readers may have noticed that this is not a probability distribution. This comes from the fact that for a physicist there is no difference between sums and integrals. The exact result is $p(k) = 2m(m+1)/(k(k+1)(k+2))$ for $k \geq m$.

Like the Watts and Strogatz model, the preferential attachment model with its "scale-free" power laws, soon saw many applications. To quote the cover of the April 13, 2002, issue of *New Scientist*: "How can a single law govern our sex lives, the proteins in our bodies, movie stars, and supercool atoms? Nature is telling us something.…" The common feature alluded to on the cover is the fact that (in all cases but supercool atoms) the degree distributions of vertices in all of these systems have power law $p(k) \approx Ck^{-\gamma}$. A study of sexual relations of 2,810 people in Sweden found that the number of sexual partners for females and males followed power laws with exponents $\gamma_f = 3.5$ and $\gamma_m = 3.3$, respectively. A study of metabolic networks in forty-three organisms yielded $\gamma \approx 2.1$. For movie actors several studies have found $\gamma \approx 2.3$. A fourth example, which may be of interest to readers, is the collaboration graph of mathematicians, which has average path length 9.5 and $\gamma = 2.1$.

These power laws are certainly not the first to have been found. Alfred Lotka noticed them in his law of scientific productivity in 1926. Zipf found them in word usage in the English language; Pareto, in income distributions. It is not clear to me why *New Scientist* concludes that the examples we have cited "grow in the same way and have the same strengths and weaknesses: understand one and you understand them all." Albert and Barabási's calculations show that preferential attachment implies power laws, but the converse is not true. The Swedish researchers who did the sex study suffered from this confusion: "Maybe people become more attractive the more partners they get." In an age with a large number of serious sexually transmitted diseases, this is a curious conclusion. A simpler, more logical explanation is that people vary in the rate at which they seek partners and hence also in the number that they have had.

A second problem with the application of the model of Albert and Barabási that is easier to resolve is the fact that the preferential attachment model has $\gamma = 3$, but the "applications" show a variety of powers. If one postulates preferential attachment with probability proportional to k^a, then for $a > 1$ we end up with one man having sex with all of the women in Sweden, while for $a < 1$ the degree distribution $p(k)$ is a stretched exponential $exp(-ck^{1-a})$. On the borderline, if one uses an affine function $A + k$, then one can generate powers between 2 and infinity. See Krapivsky, Redner, and Leyvraz (2000), or Dorogovtsev, Mendes, and Samukhin (2000).



At this point one may ask, Why do we care that the Albert and Barabási model (and many real networks) have a power law distribution of vertex degrees, while the Erdős and Renyi model has a Poisson distribution ($p(k) = e^{-\lambda}\lambda^k/k!$), which is concentrated near its mean? One answer is that this difference in the degree distribution drastically changes properties of processes that take place on the network. Consider, for example, epidemics (of sexually transmitted diseases or computer viruses). On a regular lattice or on the giant component of an Erdős-Renyi random graph, only sufficiently virulent diseases will cause epidemics. On a scale-free network, if the power is < 3, then the threshold for transmissibility of the disease needed to cause a widespread epidemic is 0. In other words, even the most inefficient email virus will spread widely through the system. The last result is easy to understand: if the power is < 3 and we pick a vertex with a probability proportional to its degree, then the average number of neighbors it has is infinite, so the average number of neighbors it will infect is > 1. However, it is not clear if this result has anything to say about the current plague of computer viruses or strategies for treating sexually transmitted diseases, as some have claimed.

A second answer concerns network reliability. In an Erdős-Renyi random graph with $p = c/n$, if one independently destroys a fraction q of the edges and if $c(1-q) < 1$, then the network falls apart. However, power law networks are highly resilient to this type of random destruction. Even if a large percentage of the edges are destroyed, the high-degree hubs cause the existence of a giant component. Conversely, the failure of a small number of carefully chosen nodes can have disastrous consequences for a power law network but has little impact on an Erdős-Renyi one.

Up to this point we have taken the approach Halmos did as book review editor for the *Bulletin of the AMS*: a book review is an excuse for a miniessay on a given topic. Turning our attention now to the books themselves, Barabási's is one that mathematicians love to hate. Simple results, arrived at mostly by simulation and computer processing of database information, are used to make far-reaching conclusions about a wide variety of systems.

As I was reading this book, I plotted a review filled with pointed remarks: "Barabási's book, like catastrophe theory, says nothing about everything, which will not please mathematicians, who are more inclined to knowing everything about nothing." "Mathematicians have much to learn from Barabási. We would generate many more sales with titles like *Calculus: The Science of Motion. A theory that explains everything from the workings of nanotechnology to the motion of galaxies, from the flowing of blood in our hearts to the emotions that govern it.*"

However, as time went on, I found myself frequently quoting in conversations things I had learned from the book. For all that I may be irritated by Barabási's bluster, he knows the story behind the results very well. In addition to many of the tidbits in this article, he has a number of interesting anecdotes. To mention four from a much larger list, he has stories about Erdős and Renyi, about the Hungarian author Frigyes Karinthy, who made a claim about six degrees of separation in 1929; about MafiaBoy, who crashed a number of ecommerce sites by forcing them to serve billions of ghost customers; and about how one overloaded power line led to a cascading failure that put the West Coast in the dark.

Watts's book has a tone more in keeping with a mathematician's temperament. His approach to telling the story is: here are some interesting questions that we are trying to figure out, and this is how far we have gotten on them. His book also has a wealth of background information but with a slightly different emphasis. Having moved from a Ph.D. program in theoretical and applied mechanics to a postdoc at the Santa Fe Institute and now to the sociology department at Columbia University, Watts spends a lot of effort relating the random graph results to contemporary social science.

Both books are entertaining bedtime reading, though Barabási's should carry a surgeon general's warning for pure mathematicians with high blood pressure. However, neither book gives much of a hint about the mathematics behind the headlines. For this reason I think that most readers of the *Notices* would be better off keeping their credit cards in their wallets and instead getting Newman's article [1]. It covers the same ground in a clear and enlightening manner but with more details. From his article it is clear that there are a number of interesting features of these models that have been demonstrated by computer simulation or approximate computations whose proofs will lead to interesting mathematical developments.

References

[1] M. Newman, The structure and function of complex networks, *SIAM Review* **45** (2003), 167–256.