From *SIAM News*, Volume 37, Number 3, April 2004

***The Mathematics of Networks***

# Large Dynamic Graphs: What Can Researchers Learn from Them?

*By Fan Chung Graham*

Researchers who study massive real-world networks like the Internet and the Web are facing new and challenging mathematical problems. These problems arise in part because the usual assumptions mathematicians make in problems of this type may no longer hold. In problems involving massive data sets, for instance, the networks or graphs we use to study the systems are prohibitively large—so much so that the (exact) number of nodes is no longer a useful parameter. Instead, we can obtain only partial information, and even that is changing dynamically. In terms of the Internet: The number of Internet hosts topped 109 million as of March 2001, and is estimated to be growing at more than 51% per year. And the number of Web pages indexed by large search engines now exceeds 2 billion, and more than 4000 Web sites are thought to be created every day.

*Power law distributions are found in widely different settings, including linguistics, physics, the rates at which academics cite one another in journals, and even in nature and the economy.*

Given the massive and dynamic graphs arising from the Internet and the Web, is it even possible for researchers to determine simple structural properties? Here are some examples of things we would like to know about these graphs: How connected are they? In particular, what are the size and diameter of the largest component? What about the second-largest component? Are there interesting structural properties that govern or influence the development and use of these physical and virtual networks?

Despite the complexity and elusiveness of large dynamic graphs, some useful structural characteristics have recently come to light. In particular, several research groups have observed that massive graphs, such as graphs of the Web or graphs of phone calls (arising from telecommunication networks), have degree distributions that obey a *power law*. In a power law degree distribution, the fraction of nodes with degree $d$ is proportional to $1/d^{\beta}$ for some constant $\beta > 0$.

Power law distributions are found in widely different settings, including linguistics, physics, the rates at which academics cite one another in journals, and even in nature and the economy. As it turns out, the history of power laws can be traced back to various statistical studies performed as far back as the 1920s. Still, with power law graphs intrinsically embedded in so many of the data bases that surround us today, a whole new dimension of problems and research directions is emerging.

Researchers who study power law graph models take one of two basic approaches. In the first, they try to model power law graphs and the manner in which the power law degree distribution arises, with the aim of approximating the statistical behavior of some targeted massive graphs. The second approach is to focus on classes of graphs with a given degree distribution, such as the power law. In this case the goal is to derive, in a rigorous way, structures and properties, such as connected components, diameters, and the reachability of the graphs, that will produce the given degree distribution.

The starting point for these studies is random graph theory. The classic random graph model, introduced in 1959 by the celebrated mathematicians Paul Erdős and Alfred Rényi, turns out to be useful for investigating some of the currently interesting massive graphs. The process by which links are added to or subtracted from a massive graph, one link at a time, has a flavor very similar to the traditional study of the evolution of random graphs. Furthermore, a subgraph of a random graph still has certain quantitative random-like behavior, and the same is true for most realistic graphs.

But there are also differences: The classic model of random graphs focuses on graphs that are for the most part regular (almost all the vertices, that is, have the same expected degree), while in a real-world graph, a small number of the vertices often have very large degrees. Nevertheless, we have learned that we can extend the Erdős–Rényi model to consider random graphs of any expected degrees.

This approach has already led to new directions in random graph theory. For example, many real-world networks show the "small-world" phenomenon: In a series of experiments conducted in 1967, the social psychologist Stanley Milgram found that any two strangers are connected by a chain of intermediate acquaintances of length at most six. (Readers who thought the playwright John Guare invented this concept with his play *Six Degrees of Separation* will see that Guare was basing his art on a thirty-year-old idea.) In 1999, Barabási et al. observed that for certain portions of the Internet, any two Web pages are at most 19 clicks away from one another. It turns out that we can use random graph theory to articulate this interesting theory in a rigorous way. Many Internet, social, and citation networks exhibit power law degree distributions with exponents in the range $2 < \beta < 3$; that is, the number of vertices of degree $k$ is proportional to $1/k^{\beta}$. It can be shown that the average distance of such random power law graphs is almost surely of order log log $n$.

One of the main issues in the age of information technology is the management of massive data sets. One way to deal with massive amounts of data is to take advantage of the associated network structure, which represents the interrelations of the data. The developers of the Google search engine have done just that: Using spectral methods, Google assigns rankings to all pages in the

Web, determining the order in which responses to a user's search appear. In similar ways, many of the information networks that surround us today provide interesting motivation and suggest new and challenging research directions that will engage researchers for years to come.

*Fan Chung Graham is a professor of mathematics at the University of California, San Diego.*