

the Internet, the power-law distribution of connections is evident in the distribution of the number of links per node. In other words, most nodes have few links, while a few nodes have many links. This distribution is called scale-free, because it is independent of the size of the system. The Internet is not the only example of a scale-free network; other examples include protein-protein interaction networks, social networks, and biological networks.

Scale-free networks are characterized by a small-world effect, which means that any two nodes in the network are connected by a short path. This is in contrast to a random network, where the average distance between two nodes increases exponentially with the size of the network.

Small-world networks are also characterized by a high clustering coefficient, which means that the neighbors of a node tend to be interconnected.

These properties are shared by many complex systems, such as the brain, the Internet, and social networks. They are also found in many other systems, such as protein-protein interaction networks, social networks, and biological networks.

The discovery of scale-free networks has led to a new understanding of the way complex systems work. It has shown that many systems exhibit self-similarity, which means that they have similar properties at different scales.

For example, the distribution of connections in a scale-free network is similar to the distribution of connections in a smaller part of the network, such as a single node or a cluster of nodes.

This self-similarity is a key feature of scale-free networks, and it has important implications for the way we think about complex systems.

For example, it has been shown that scale-free networks are more robust than random networks against random attacks, such as viruses or worms.

It has also been shown that scale-free networks are more efficient than random networks at spreading information, such as rumors or news.

These properties are shared by many complex systems, such as the brain, the Internet, and social networks. They are also found in many other systems, such as protein-protein interaction networks, social networks, and biological networks.

The discovery of scale-free networks has led to a new understanding of the way complex systems work. It has shown that many systems exhibit self-similarity, which means that they have similar properties at different scales.

For example, the distribution of connections in a scale-free network is similar to the distribution of connections in a smaller part of the network, such as a single node or a cluster of nodes.

This self-similarity is a key feature of scale-free networks, and it has important implications for the way we think about complex systems.

For example, it has been shown that scale-free networks are more robust than random networks against random attacks, such as viruses or worms.

It has also been shown that scale-free networks are more efficient than random networks at spreading information, such as rumors or news.

These properties are shared by many complex systems, such as the brain, the Internet, and social networks. They are also found in many other systems, such as protein-protein interaction networks, social networks, and biological networks.

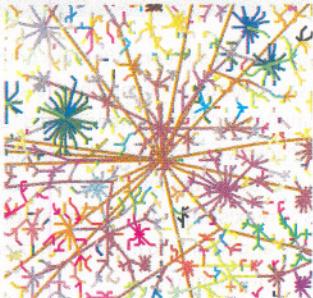
The discovery of scale-free networks has led to a new understanding of the way complex systems work. It has shown that many systems exhibit self-similarity, which means that they have similar properties at different scales.

For example, the distribution of connections in a scale-free network is similar to the distribution of connections in a smaller part of the network, such as a single node or a cluster of nodes.

This self-similarity is a key feature of scale-free networks, and it has important implications for the way we think about complex systems.

For example, it has been shown that scale-free networks are more robust than random networks against random attacks, such as viruses or worms.

It has also been shown that scale-free networks are more efficient than random networks at spreading information, such as rumors or news.



THE INTERNET, mapped on the opposite page, is a scale-free network in that some sites (starbursts and detail above) have a seemingly unlimited number of connections to other sites. This map, made on February 6, 2003, traces the shortest routes from a test Web site to about 100,000 others, using like colors for similar Web addresses.

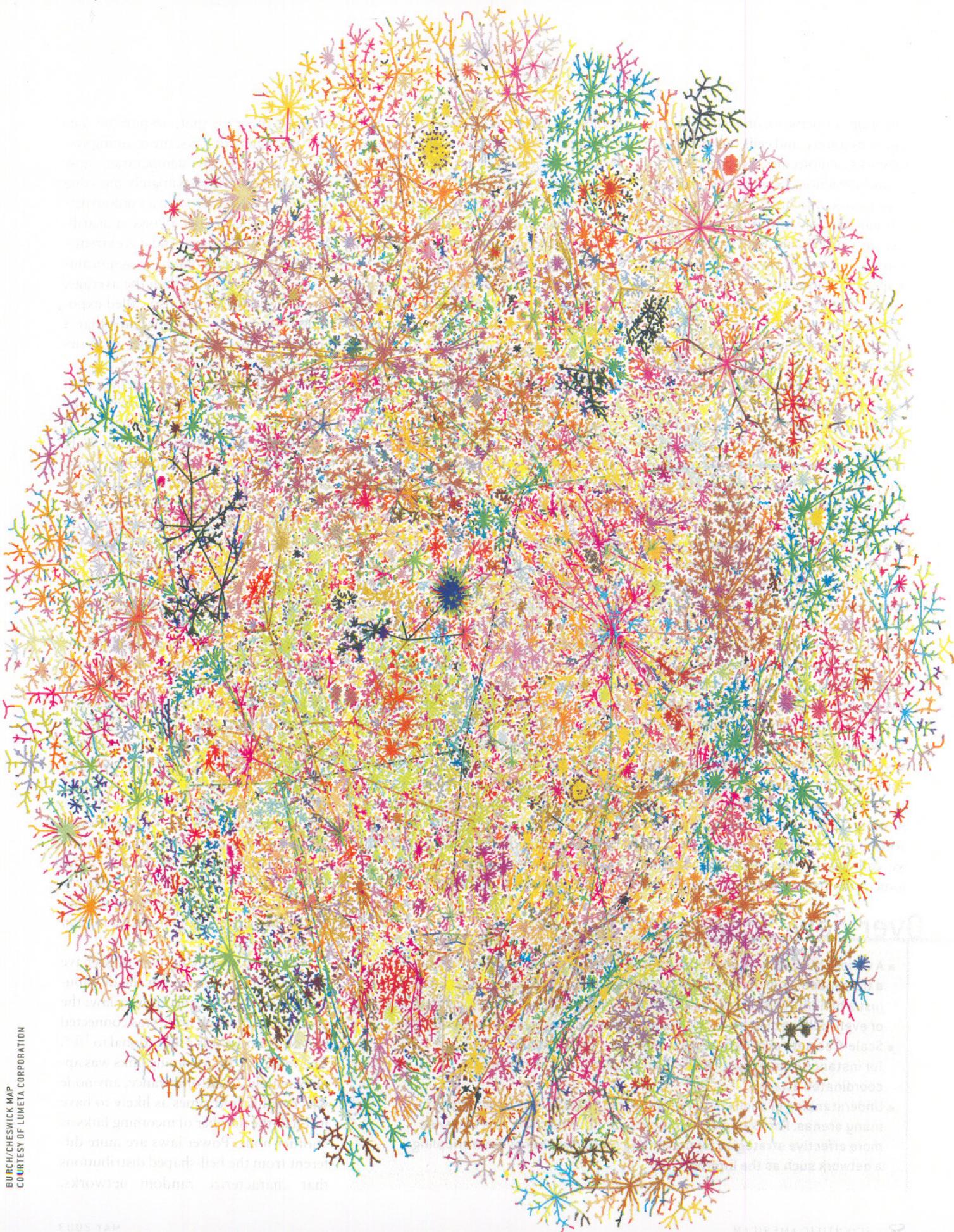
# Scale-Free Networks

**Scientists have recently discovered that various complex systems have an underlying architecture governed by shared organizing principles. This insight has important implications for a host of applications, from drug development to Internet security.**

BY ALBERT-LÁSZLÓ BARABÁSI AND ERIC BONABEAU

THE INTERNET IS A SCALE-FREE NETWORK, WHICH MEANS THAT THE DISTRIBUTION OF CONNECTIONS IS POWER-LAW, INDEPENDENT OF THE SIZE OF THE SYSTEM. THIS MEANS THAT MOST NODES HAVE FEW CONNECTIONS, WHILE A FEW NODES HAVE MANY CONNECTIONS. FOR EXAMPLE, THE INTERNET HAS ABOUT 100,000 NODES WITH MORE THAN 10,000 CONNECTIONS, WHILE THERE ARE MILLIONS OF NODES WITH ONLY A FEW CONNECTIONS. THIS DISTRIBUTION IS KNOWN AS A POWER-LAW DISTRIBUTION, AND IT IS OBSERVED IN MANY OTHER COMPLEX SYSTEMS, SUCH AS PROTEIN-PROTEIN INTERACTION NETWORKS, SOCIAL NETWORKS, AND BIOLOGICAL NETWORKS.

THE INTERNET IS A SCALE-FREE NETWORK, WHICH MEANS THAT THE DISTRIBUTION OF CONNECTIONS IS POWER-LAW, INDEPENDENT OF THE SIZE OF THE SYSTEM. THIS MEANS THAT MOST NODES HAVE FEW CONNECTIONS, WHILE A FEW NODES HAVE MANY CONNECTIONS. FOR EXAMPLE, THE INTERNET HAS ABOUT 100,000 NODES WITH MORE THAN 10,000 CONNECTIONS, WHILE THERE ARE MILLIONS OF NODES WITH ONLY A FEW CONNECTIONS. THIS DISTRIBUTION IS KNOWN AS A POWER-LAW DISTRIBUTION, AND IT IS OBSERVED IN MANY OTHER COMPLEX SYSTEMS, SUCH AS PROTEIN-PROTEIN INTERACTION NETWORKS, SOCIAL NETWORKS, AND BIOLOGICAL NETWORKS.



BURCH/CHESWICK MAP  
COURTESY OF LUMETA CORPORATION

1000 Yards

MADREMA DEEP

# Networks are everywhere.

The brain is a network of nerve cells connected by axons, and cells themselves are networks of molecules connected by biochemical reactions. Societies, too, are networks of people linked by friendships, familial relationships and professional ties. On a larger scale, food webs and ecosystems can be represented as networks of species. And networks pervade technology: the Internet, power grids and transportation systems are but a few examples. Even the language we are using to convey these thoughts to you is a network, made up of words connected by syntactic relationships.

Yet despite the importance and pervasiveness of networks, scientists have had little understanding of their structure and properties. How do the interactions of several malfunctioning nodes in a complex genetic network result in cancer? How does diffusion occur so rapidly in certain social and communications systems, leading to epidemics of diseases and computer viruses? How do some networks continue to function even after the vast majority of their nodes have failed?

Recent research has begun to answer such questions. Over the past few years, investigators from a variety of fields have discovered that many networks—from the World Wide Web to a cell's metabolic system to actors in Hollywood—are dominated by a relatively small number of nodes that are connected to many other sites. Networks containing such important nodes, or hubs, tend to be what we call "scale-free," in the sense that some hubs have a seemingly unlimited

number of links and no node is typical of the others. These networks also behave in certain predictable ways; for example, they are remarkably resistant to accidental failures but extremely vulnerable to coordinated attacks.

Such discoveries have dramatically changed what we thought we knew about the complex interconnected world around us. Unexplained by previous network theories, hubs offer convincing proof that various complex systems have a strict architecture, ruled by fundamental laws—laws that appear to apply equally to cells, computers, languages and society. Furthermore, these organizing principles have significant implications for developing better drugs, defending the Internet from hackers, and halting the spread of deadly epidemics, among other applications.

## Networks without Scale

FOR MORE THAN 40 YEARS, science treated all complex networks as being completely random. This paradigm has its roots in the work of two Hungarian mathematicians, the imitable Paul Erdős and his close collaborator Alfréd Rényi. In 1959, aiming to describe networks seen in communications and the life sciences, Erdős and Rényi suggested that such systems could be effectively modeled by connecting their nodes with randomly placed links. The simplicity of their approach and the elegance of some of their related theorems revitalized graph theory, leading to the emergence of a field in mathematics that focuses on random networks.

An important prediction of random-

network theory is that, despite the random placement of links, the resulting system will be deeply democratic: most nodes will have approximately the same number of links. Indeed, in a random network the nodes follow a Poisson distribution with a bell shape, and it is extremely rare to find nodes that have significantly more or fewer links than the average. Random networks are also called exponential, because the probability that a node is connected to  $k$  other sites decreases exponentially for large  $k$ .

So in 1998, when we, together with Hawoong Jeong and Réka Albert of the University of Notre Dame, embarked on a project to map the World Wide Web, we expected to find a random network. Here's why: people follow their unique interests when deciding what sites to link their Web documents to, and given the diversity of everyone's interests and the tremendous number of pages they can choose from, the resulting pattern of connections should appear fairly random.

The measurements, however, defied that expectation. Software designed for this project hopped from one Web page to another and collected all the links it could. Although this virtual robot reached only a tiny fraction of the entire Web, the map it assembled revealed something quite surprising: a few highly connected pages are essentially holding the World Wide Web together. More than 80 percent of the pages on the map had fewer than four links, but a small minority, less than 0.01 percent of all nodes, had more than 1,000. (A subsequent Web survey would uncover one document that had been referenced by more than two million other pages!)

Counting how many Web pages have exactly  $k$  links showed that the distribution followed a so-called power law: the probability that any node was connected to  $k$  other nodes was proportional to  $1/k^n$ . The value of  $n$  for incoming links was approximately 2, so, for instance, any node was roughly four times as likely to have just half the number of incoming links as another node. Power laws are quite different from the bell-shaped distributions that characterize random networks.

## Overview/Scale-Free Networks

- A variety of complex systems share an important property: some nodes have a tremendous number of connections to other nodes, whereas most nodes have just a handful. The popular nodes, called hubs, can have hundreds, thousands or even millions of links. In this sense, the network appears to have no scale.
- Scale-free networks have certain important characteristics. They are, for instance, robust against accidental failures but vulnerable to coordinated attacks.
- Understanding of such characteristics could lead to new applications in many arenas. For example, computer scientists might be able to devise more effective strategies for preventing computer viruses from crippling a network such as the Internet.

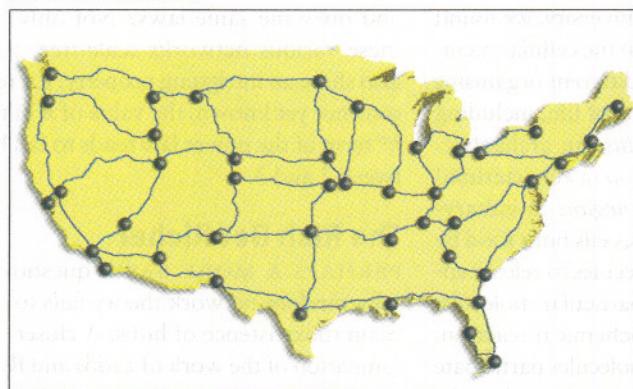
# RANDOM VERSUS SCALE-FREE NETWORKS

RANDOM NETWORKS, which resemble the U.S. highway system (*simplified in left map*), consist of nodes with randomly placed connections. In such systems, a plot of the distribution of node linkages will follow a bell-shaped curve (*left graph*), with most nodes having approximately the same number of links.

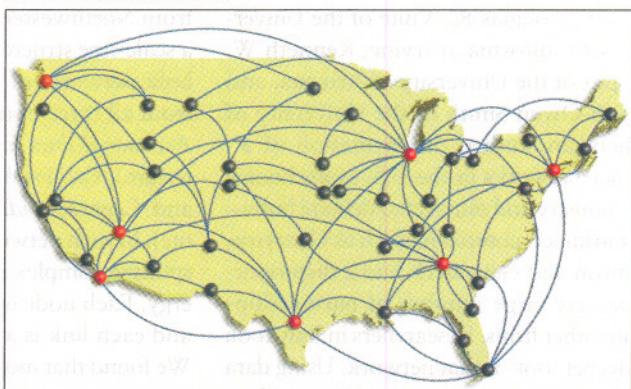
In contrast, scale-free networks, which resemble the U.S. airline system (*simplified in right map*), contain hubs (*red*)—

nodes with a very high number of links. In such networks, the distribution of node linkages follows a power law (*center graph*) in that most nodes have just a few connections and some have a tremendous number of links. In that sense, the system has no "scale." The defining characteristic of such networks is that the distribution of links, if plotted on a double-logarithmic scale (*right graph*), results in a straight line.

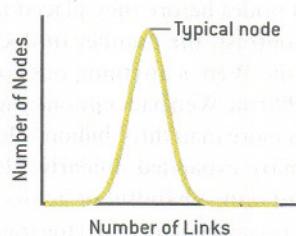
## Random Network



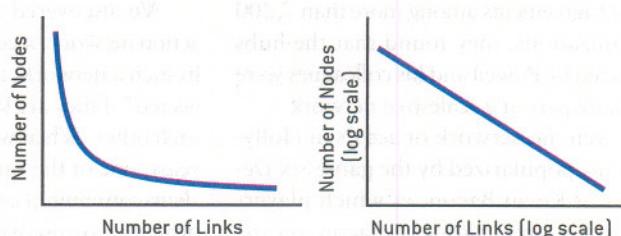
## Scale-Free Network



### Bell Curve Distribution of Node Linkages



### Power Law Distribution of Node Linkages



Specifically, a power law does not have a peak, as a bell curve does, but is instead described by a continuously decreasing function. When plotted on a double-logarithmic scale, a power law is a straight line [*see illustration above*]. In contrast to the democratic distribution of links seen in random networks, power laws describe systems in which a few hubs, such as Yahoo and Google, dominate.

Hubs are simply forbidden in random networks. When we began to map the Web, we expected the nodes to follow a bell-shaped distribution, as do people's heights. Instead we discovered certain nodes that defied explanation, almost as if we had stumbled on a significant number of people who were 100 feet tall, thus prompting us to coin the term "scale-free."

## Scale-Free Networks Abound

OVER THE PAST several years, researchers have uncovered scale-free structures in a stunning range of systems. When we studied the World Wide Web, we looked at the *virtual* network of Web pages connected to one another by hyperlinks. In contrast, Michalis Faloutsos of the University of California at Riverside, Petros Faloutsos of the University of Toronto and Christos Faloutsos of Carnegie Mellon University analyzed the *physical* structure of the Internet. These three computer-scientist brothers investigated the routers connected by optical or other communications lines and found that the topology of that network, too, is scale-free.

Researchers have also discovered that

some social networks are scale-free. A collaboration between scientists from Boston University and Stockholm University, for instance, has shown that a network of sexual relationships among people in Sweden followed a power law: although most individuals had only a few sexual partners during their lifetime, a few (the hubs) had hundreds. A recent study led by Stefan Bornholdt of the University of Kiel in Germany concluded that the network of people connected by e-mail is likewise scale-free. Sidney Redner of Boston University demonstrated that the network of scientific papers, connected by citations, follows a power law as well. And Mark Newman of the University of Michigan at Ann Arbor examined collaborations among scientists in several

disciplines, including physicians and computer scientists, and found that those networks were also scale-free, corroborating a study we conducted focusing on mathematicians and neurologists. (Interestingly, one of the largest hubs in the mathematics community is Erdős himself, who wrote more than 1,400 papers with no fewer than 500 co-authors.)

Scale-free networks can occur in business. Walter W. Powell of Stanford University, Douglas R. White of the University of California at Irvine, Kenneth W. Koput of the University of Arizona, and Jason-Owen Smith of the University of Michigan studied the formation of alliance networks in the U.S. biotechnology industry and discovered definite hubs—for instance, companies such as Genzyme, Chiron and Genentech had a disproportionately large number of partnerships with other firms. Researchers in Italy took a deeper look at that network. Using data collected by the University of Siena's Pharmaceutical Industry Database, which now provides information for around 20,100 R&D agreements among more than 7,200 organizations, they found that the hubs detected by Powell and his colleagues were actually part of a scale-free network.

Even the network of actors in Hollywood—popularized by the game Six Degrees of Kevin Bacon, in which players try to connect actors to Bacon via the movies in which they have appeared together—is scale-free. A quantitative analy-

sis of that network showed that it, too, is dominated by hubs. Specifically, although most actors have only a few links to others, a handful of actors, including Rod Steiger and Donald Pleasence, have thousands of connections. (Incidentally, on a list of most connected actors, Bacon ranked just 876th.)

On a more serious note, scale-free networks are present in the biological realm. With Zoltán Oltvai, a cell biologist from Northwestern University, we found a scale-free structure in the cellular metabolic networks of 43 different organisms from all three domains of life, including *Archaeoglobus fulgidus* (an archaeabacterium), *Escherichia coli* (a eubacterium) and *Caenorhabditis elegans* (a eukaryote). In such networks, cells burn food by splitting complex molecules to release energy. Each node is a particular molecule, and each link is a biochemical reaction. We found that most molecules participate in just one or two reactions, but a few (the hubs), such as water and adenosine triphosphate, play a role in most of them.

We discovered that the protein-interaction network of cells is scale-free as well. In such a network, two proteins are “connected” if they are known to interact with each other. When we investigated Baker’s yeast, one of the simplest eukaryotic (nucleus-containing) cells, with thousands of proteins, we discovered a scale-free topology: although most proteins interact with only one or two others, a few are able to

attach themselves physically to a huge number. We found a similar result in the protein-interaction network of an organism that is very different from yeast, a simple bacterium called *Helicobacter pylori*.

Indeed, the more that scientists studied networks, the more they uncovered scale-free structures. These findings raised an important question: How can systems as fundamentally different as the cell and the Internet have the same architecture and obey the same laws? Not only are these various networks scale-free, they also share an intriguing property: for reasons not yet known, the value of  $n$  in the  $k^n$  term of the power law tends to fall between 2 and 3.

## The Rich Get Richer

PERHAPS A MORE BASIC question is why random-network theory fails to explain the existence of hubs. A closer examination of the work of Erdős and Rényi reveals two reasons.

In developing their model, Erdős and Rényi assumed that they had the full inventory of nodes before they placed the links. In contrast, the number of documents on the Web is anything but constant. In 1990 the Web had only one page. Now it has more than three billion. Most networks have expanded similarly. Hollywood had only a handful of actors in 1890, but as new people joined the trade, the network grew to include more than half a million, with the rookies connecting to veteran actors. The Internet had only a few routers about three decades ago, but it gradually grew to have millions, with the new routers always linking to those that were already part of the network. Thanks to the growing nature of real networks, older nodes had greater opportunities to acquire links.

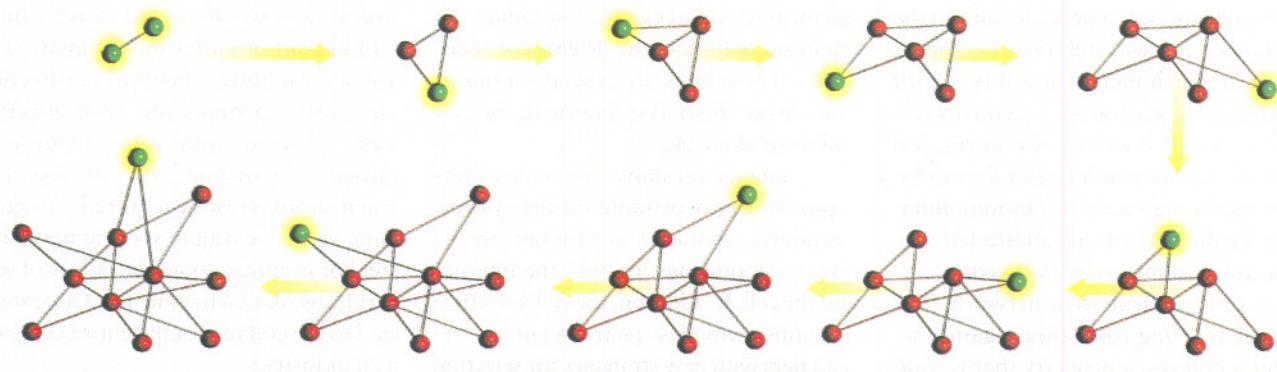
Furthermore, all nodes are not equal. When deciding where to link their Web page, people can choose from a few billion locations. Yet most of us are familiar with only a tiny fraction of the full Web, and that subset tends to include the more connected sites because they are easier to find. By simply linking to those nodes, people exercise and reinforce a bias toward them. This process of “preferential attachment” occurs elsewhere. In Hollywood the more

## Examples of Scale-Free Networks

NETWORK	NODES	LINKS
Cellular metabolism	Molecules involved in burning food for energy	Participation in the same biochemical reaction
Hollywood	Actors	Appearance in the same movie
Internet	Routers	Optical and other physical connections
Protein regulatory network	Proteins that help to regulate a cell’s activities	Interactions among proteins
Research collaborations	Scientists	Co-authorship of papers
Sexual relationships	People	Sexual contact
World Wide Web	Web pages	URLs

## BIRTH OF A SCALE-FREE NETWORK

A SCALE-FREE NETWORK grows incrementally from two to 11 nodes in this example. When deciding where to establish a link, a new node (green) prefers to attach to an existing node (red) that already has many other connections. These two basic mechanisms—growth and preferential attachment—will eventually lead to the system's being dominated by hubs, nodes having an enormous number of links.



connected actors are more likely to be chosen for new roles. On the Internet the more connected routers, which typically have greater bandwidth, are more desirable for new users. In the U.S. biotech industry, well-established companies such as Genzyme tend to attract more alliances, which further increases their desirability for future partnerships. Likewise, the most cited articles in the scientific literature stimulate even more researchers to read and cite them, a phenomenon that noted sociologist Robert K. Merton called the Matthew effect, after a passage in the New Testament: "For unto every one that hath shall be given, and he shall have abundance."

These two mechanisms—growth and preferential attachment—help to explain the existence of hubs: as new nodes appear, they tend to connect to the more connected sites, and these popular locations thus acquire more links over time than their less connected neighbors. And this "rich get richer" process will generally favor the early nodes, which are more likely to eventually become hubs.

Along with Réka Albert, we have used computer simulations and calculations to show that a growing network with preferential attachment will indeed become scale-free, with its distribution of nodes following a power law. Although this theoretical model is simplistic and needs to be adapted to specific situations, it does appear to confirm our explanation for

why scale-free networks are so ubiquitous in the real world.

Growth and preferential attachment can even help explicate the presence of scale-free networks in biological systems. Andreas Wagner of the University of New Mexico and David A. Fell of Oxford Brookes University in England have found, for instance, that the most-connected molecules in the *E. coli* metabolic network tend to have an early evolutionary history: some are believed to be remnants of the so-called RNA world (the evolutionary step before the emergence of DNA), and others are components of the most ancient metabolic pathways.

Interestingly, the mechanism of preferential attachment tends to be linear. In other words, a new node is twice as likely to link to an existing node that has twice as many connections as its neighbor. Redner and his colleagues at Boston University and elsewhere have investigated different types of preferential attachment and have learned that if the mechanism is faster than linear (for example, a new node is four times as likely to link to

an existing node that has twice as many connections), one hub will tend to run away with the lion's share of connections. In such "winner take all" scenarios, the network eventually assumes a star topology with a central hub.

### An Achilles' Heel

AS HUMANITY BECOMES increasingly dependent on power grids and communications webs, a much-voiced concern arises: Exactly how reliable are these types of networks? The good news is that complex systems can be amazingly resilient against accidental failures. In fact, although hundreds of routers routinely malfunction on the Internet at any moment, the network rarely suffers major disruptions. A similar degree of robustness characterizes living systems: people rarely notice the consequences of thousands of errors in their cells, ranging from mutations to misfolded proteins. What is the origin of this robustness?

Intuition tells us that the breakdown of a substantial number of nodes will result in a network's inevitable fragmentation.

### THE AUTHORS

ALBERT-LÁSZLÓ BARABÁSI and ERIC BONABEAU study the behavior and characteristics of myriad complex systems, ranging from the Internet to insect colonies. Barabási is Emil T. Hofman Professor of Physics at the University of Notre Dame, where he directs research on complex networks. He is author of *Linked: The New Science of Networks*. Bonabeau is chief scientist at Icosystem, a consulting firm based in Cambridge, Mass., that applies the tools of complexity science to the discovery of business opportunities. He is co-author of *Swarm Intelligence: From Natural to Artificial Systems*. This is Bonabeau's second article for *Scientific American*.

tion. This is certainly true for random networks: if a critical fraction of nodes is removed, these systems break into tiny, noncommunicating islands. Yet simulations of scale-free networks tell a different story: as many as 80 percent of randomly selected Internet routers can fail and the remaining ones will still form a compact cluster in which there will still be a path between any two nodes. It is equally difficult to disrupt a cell's protein-interaction network: our measurements indicate that even after a high level of random mutations are introduced, the unaffected proteins will continue to work together.

In general, scale-free networks display an amazing robustness against accidental failures, a property that is rooted in their inhomogeneous topology. The random removal of nodes will take out mainly the small ones because they are much more plentiful than hubs. And the elimination of small nodes will not disrupt the network topology significantly, because they contain few links compared with the hubs, which connect to nearly everything. But a reliance on hubs has a serious drawback: vulnerability to attacks.

In a series of simulations, we found

that the removal of just a few key hubs from the Internet splintered the system into tiny groups of hopelessly isolated routers. Similarly, knockout experiments in yeast have shown that the removal of the more highly connected proteins has a significantly greater chance of killing the organism than does the deletion of other nodes. These hubs are crucial—if mutations make them dysfunctional, the cell will most likely die.

A reliance on hubs can be advantageous or not, depending on the system. Certainly, resistance to random breakdown is good news for both the Internet and the cell. In addition, the cell's reliance on hubs provides pharmaceutical researchers with new strategies for selecting drug targets, potentially leading to cures that would kill only harmful cells or bacteria by selectively targeting their hubs, while leaving healthy tissue unaffected. But the ability of a small group of well-informed hackers to crash the entire communications infrastructure by targeting its hubs is a major reason for concern.

The Achilles' heel of scale-free networks raises a compelling question: How many hubs are essential? Recent research

suggests that, generally speaking, the simultaneous elimination of as few as 5 to 15 percent of all hubs can crash a system. For the Internet, our experiments imply that a highly coordinated attack—first removing the largest hub, then the next largest, and so on—could cause significant disruptions after the elimination of just several hubs. Therefore, protecting the hubs is perhaps the most effective way to avoid large-scale disruptions caused by malicious cyber-attacks. But much more work is required to determine just how fragile specific networks are. For instance, could the failure of several hubs like Genzyme and Genentech lead to the collapse of the entire U.S. biotech industry?

## Scale-Free Epidemics

KNOWLEDGE ABOUT scale-free networks has implications for understanding the spread of computer viruses, diseases and fads. Diffusion theories, intensively studied for decades by both epidemiologists and marketing experts, predict a critical threshold for the propagation of a contagion throughout a population. Any virus, disease or fad that is less infectious than that well-defined threshold will inevitably die out, whereas those above the threshold will multiply exponentially, eventually penetrating the entire system.

Recently, though, Romualdo Pastor-Satorras of the Polytechnic University of Catalonia in Barcelona and Alessandro Vespignani of the International Center for Theoretical Physics in Trieste, Italy, reached a disturbing conclusion. They found that in a scale-free network the threshold is zero. That is, all viruses, even those that are weakly contagious, will spread and persist in the system. This result explains why Love Bug, the most damaging computer virus thus far (it shut down the British Parliament in 2000), was still one of the most pervasive viruses a year after its supposed eradication.

Because hubs are connected to many other nodes, at least one hub will tend to be infected by any corrupted node. And once a hub has been infected, it will pass the virus to numerous other sites, eventually compromising other hubs, which will

## The Potential Implications of Scale-Free Networks for...

### Computing

- Computer networks with scale-free architectures, such as the World Wide Web, are highly resistant to accidental failures. But they are very vulnerable to deliberate attacks and sabotage.
- Eradicating viruses, even known ones, from the Internet will be effectively impossible.

### Medicine

- Vaccination campaigns against serious viruses, such as smallpox, might be most effective if they concentrate on treating hubs—people who have many connections to others. But identifying such individuals can be difficult.
- Mapping out the networks within the human cell could aid researchers in uncovering and controlling the side effects of drugs. Furthermore, identifying the hub molecules involved in certain diseases could lead to new drugs that would target those hubs.

### Business

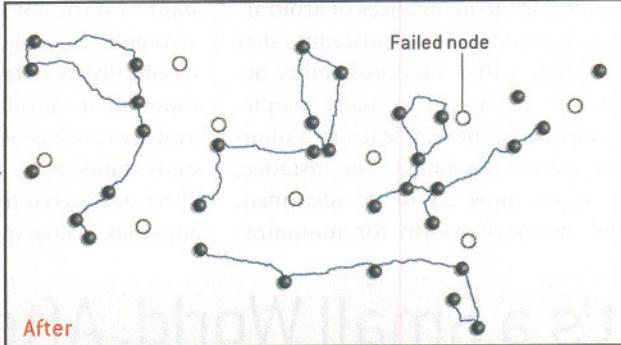
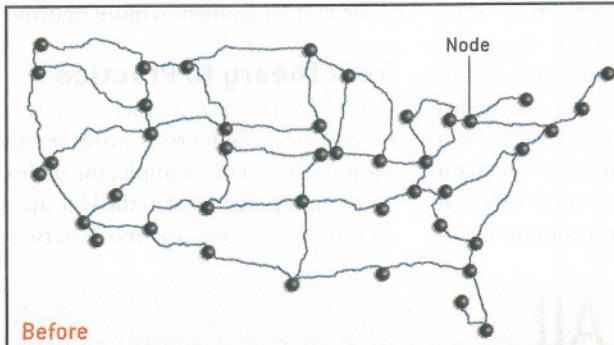
- Understanding how companies, industries and economies are interlinked could help researchers monitor and avoid cascading financial failures.
- Studying the spread of a contagion on a scale-free network could offer new ways for marketers to propagate consumer buzz about their products.

# HOW ROBUST ARE RANDOM AND SCALE-FREE NETWORKS?

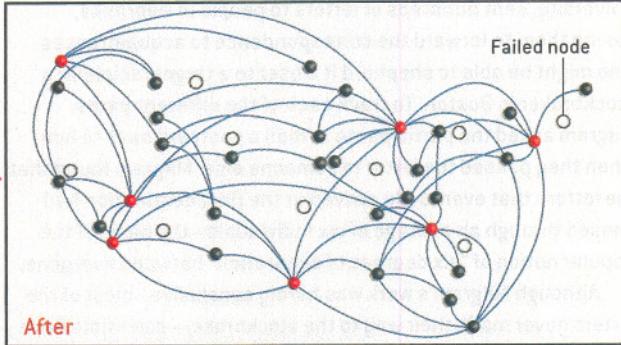
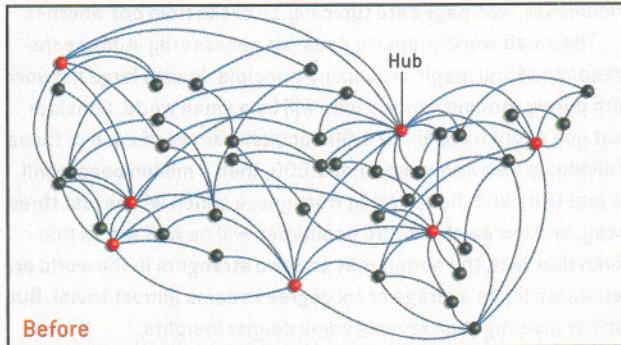
THE ACCIDENTAL FAILURE of a number of nodes in a random network (*top panels*) can fracture the system into non-communicating islands. In contrast, scale-free networks are

more robust in the face of such failures (*middle panels*). But they are highly vulnerable to a coordinated attack against their hubs (*bottom panels*).

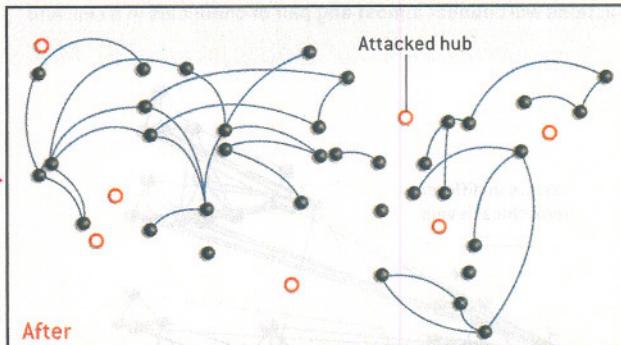
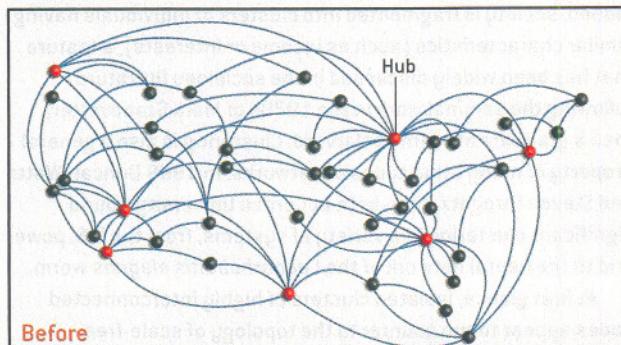
## Random Network, Accidental Node Failure



## Scale-Free Network, Accidental Node Failure



## Scale-Free Network, Attack on Hubs



then spread the virus throughout the entire system.

The fact that biological viruses spread in social networks, which in many cases appear to be scale-free, suggests that scientists should take a second look at the volumes of research written on the interplay of network topology and epidemics. Specifically, in a scale-free network, the

traditional public health approach of random immunization could easily fail because it would very likely neglect a number of the hubs. In fact, nearly everyone would have to be treated to ensure that the hubs were not missed. A vaccination for measles, for instance, must reach 90 percent of the population to be effective.

Instead of random immunizations,

though, what if doctors targeted the hubs, or the most connected individuals? Research in scale-free networks indicates that this alternative approach could be effective even if the immunizations reached only a small fraction of the overall population, provided that the fraction contained the hubs.

But identifying the hubs in a social

network is much more difficult than detecting them in other types of systems. Nevertheless, Reuven Cohen and Shlomo Havlin of Bar-Ilan University in Israel, together with Daniel ben-Avraham of Clarkson University, have proposed a clever solution: immunize a small fraction of the random acquaintances of arbitrarily selected individuals, a procedure that selects hubs with a high probability because they are linked to many people. That approach, though, leads to a number of ethical dilemmas. For instance, even if the hubs could be identified, should they have priority for immuniza-

tions and cures? Such issues notwithstanding, targeting hubs could be the most pragmatic solution for the future distribution of AIDS or smallpox vaccines in countries and regions that do not have the resources to treat everyone.

In many business contexts, people want to start, not stop, epidemics. Viral marketing campaigns, for instance, often specifically try to target hubs to speed the adoption of a product. Obviously, such a strategy is not new. Back in the 1950s, a study funded by pharmaceutical giant Pfizer discovered the important role that hubs play in how quickly a community of

doctors begins using a new drug. Indeed, marketers have intuitively known for some time that certain customers outshine others in spreading promotional buzz about products and fads. But recent work in scale-free networks provides the scientific framework and mathematical tools to probe that phenomenon more rigorously.

### From Theory to Practice

ALTHOUGH SCALE-FREE networks are pervasive, numerous prominent exceptions exist. For example, the highway system and power grid in the U.S. are not scale-free. Neither are most networks

## It's a Small World, After All

IN 1967 STANLEY MILGRAM, a social psychologist at Harvard University, sent hundreds of letters to people in Nebraska, asking them to forward the correspondence to acquaintances who might be able to shepherd it closer to a target recipient: a stockbroker in Boston. To track each of the different paths, Milgram asked the participants to mail a postcard back to him when they passed the letter to someone else. Milgram found that the letters that eventually arrived at the final destination had passed through an average of six individuals—the basis of the popular notion of "six degrees of separation" between everyone.

Although Milgram's work was hardly conclusive—most of the letters never made their way to the stockbroker—scientists have recently learned that other networks exhibit this "small world" property. We have found, for instance, that a path of just three reactions will connect almost any pair of chemicals in a cell. And

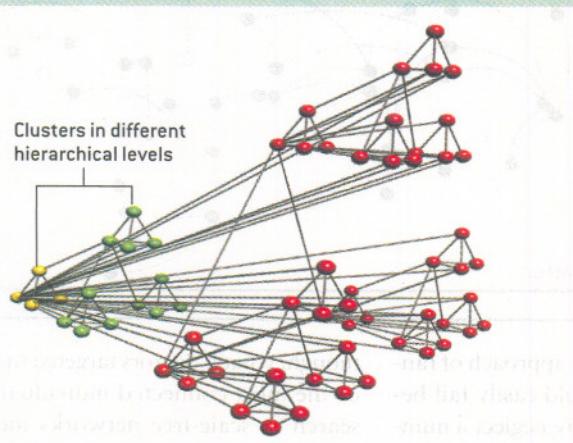
on the World Wide Web, which contains more than three billion documents, Web pages are typically 19 clicks from one another.

The small-world property does not necessarily indicate the presence of any magic organizing principle. Even a large network with purely random connections will be a small world. Consider that you might have about 1,000 acquaintances. If each of those individuals also knows another 1,000, then a million people will be just two handshakes away from you, a billion will be just three away, and the earth's entire population will be well within four. Given that fact, the notion that any two strangers in the world are separated by an average of six degrees seems almost trivial. But further investigation reveals some deeper insights.

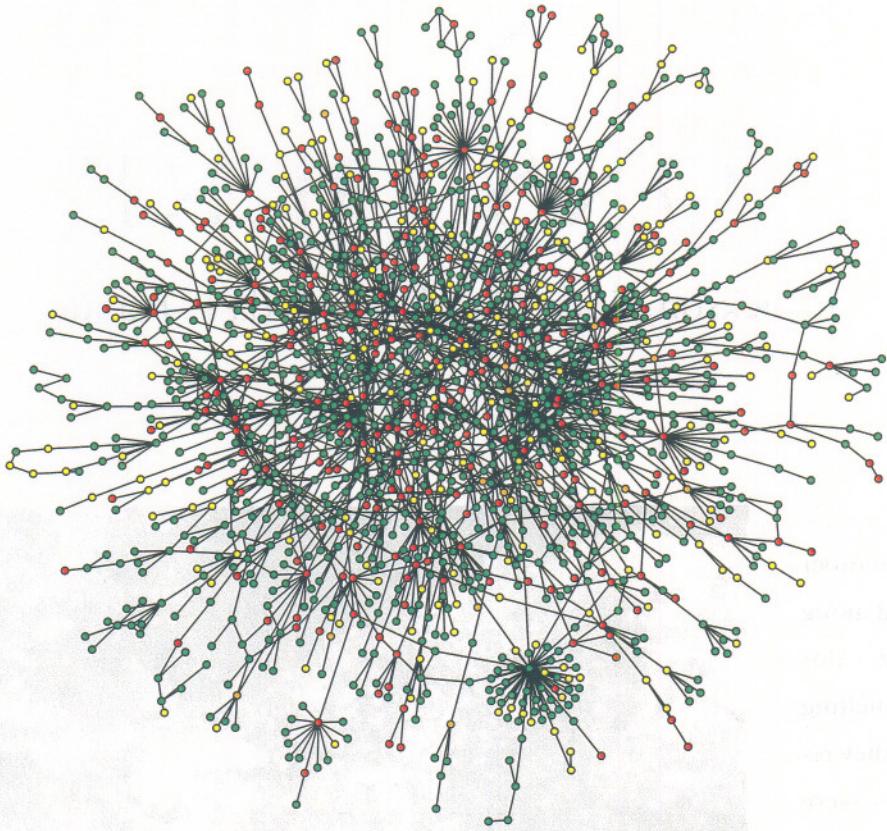
Our simple calculation assumes that the people you know are all strangers to one another. In reality, there is much overlap. Indeed, society is fragmented into clusters of individuals having similar characteristics (such as income or interests), a feature that has been widely discussed in the sociology literature following the seminal work in the 1970s of Mark Granovetter, then a graduate student at Harvard. Clustering is also a general property of many other types of networks. In 1998 Duncan Watts and Steven Strogatz, then both at Cornell University, found significant clustering in a variety of systems, from the U.S. power grid to the neural network of the *Caenorhabditis elegans* worm.

At first glance, isolated clusters of highly interconnected nodes appear to run counter to the topology of scale-free networks, in which a number of hubs radiate throughout the system, linking everything. Recently, however, we have shown that the two properties are compatible: a network can be both highly clustered and scale-free when small, tightly interlinked clusters of nodes are connected into larger, less cohesive groups (left). This type of hierarchy appears to exist in a number of systems, from the World Wide Web (in which clusters are groupings of Web pages devoted to the same topic) to a cell (in which clusters are teams of molecules responsible for a specific function).

A.-L.B. and E.B.



**HIERARCHICAL CLUSTERS**, shown schematically, could include, say, Web pages on the Frank Lloyd Wright home Fallingwater (yellow), which could be linked to other clusters (green) focusing on Wright, famous homes or Pennsylvania's attractions. Those sites, in turn, could be connected to clusters (red) on famous architects or architecture in general.



**MAP OF INTERACTING PROTEINS** in yeast highlights the discovery that highly linked, or hub, proteins tend to be crucial for a cell's survival. Red denotes essential proteins (their removal will cause the cell to die). Orange represents proteins of some importance (their removal will slow cell growth). Green and yellow represent proteins of lesser or unknown significance, respectively.

seen in materials science. In a crystal lattice, for instance, atoms have the same number of links to their neighbors. With other networks, the data are inconclusive. The relatively small size of food webs, which show predator-prey relationships, has prevented scientists from reaching a clear conclusion regarding that network's type. And the absence of large-scale connectivity maps of the brain has kept researchers from knowing the nature of that important network as well.

Determining whether a network is scale-free is important in understanding the system's behavior, but other significant parameters merit attention, too. One such characteristic is the diameter, or path length, of a network: the largest number of hops required to get from one node to another by following the shortest route possible [see box on opposite page].

Finally, knowledge of a network's general topology is just part of the story in understanding the overall characteristics and

behavior of such systems. There might be steep costs, for instance, with the addition of each link to a given node that could prevent certain networks (such as the U.S. highway system) from becoming scale-free. In food chains, some prey are easier to catch than others, and that fact has a profound effect on the overall ecosystem. With social networks, ties among household members are much stronger than connections to casual acquaintances, so diseases (and information) are more likely to spread through such linkages. For transportation, transmission and communications systems (such as the Internet),

congestion along specific links is a major consideration: too much traffic on a particular link can cause it to break down, leading to the potential failure of other links that must then handle the spillover. And the nodes themselves might not be homogeneous—certain Web pages have more interesting content, for instance—which could greatly alter the preferential-attachment mechanism.

Because of these and other factors, scientists have only begun to uncover the behavior of scale-free systems. Immunizing hubs, for instance, might not be sufficient to stop the spread of a disease; a more effective solution might be found by considering not just the number of connections a person has but also the frequency and duration of contact for those links.

In essence, we have studied complex networks first by ignoring the details of their individual links and nodes. By distancing ourselves from those particulars, we have been able to better glimpse some of the organizing principles behind these seemingly incomprehensible systems. At the very least, knowledge from this endeavor has led to the rethinking of many basic assumptions. In the past, for example, researchers modeled the Internet as a random network to test how a new routing protocol might affect system congestion. But we now know that the Internet is a scale-free system with behavior that is dramatically different from a random network's. Consequently, investigators such as John W. Byers and his colleagues at Boston University are revamping the computer models they have been using to simulate the Internet. Similarly, knowledge of the properties of scale-free networks will be valuable in a number of other fields, especially as we move beyond network topologies to probe the intricate and often subtle dynamics taking place within those complex systems.

#### MORE TO EXPLORE

All the World's a Net. David Cohen in *New Scientist*, Vol. 174, No. 2338, pages 24–29; April 13, 2002.

Statistical Mechanics of Complex Networks. Réka Albert and Albert-László Barabási in *Reviews of Modern Physics*, Vol. 74, pages 47–97; January 2002.

Linked: The New Science of Networks. Albert-László Barabási. Perseus Publishing, 2002.

Evolution of Networks: From Biological Nets to the Internet and WWW. J.F.F. Mendes and Sergei N. Dorogovtsev. Oxford University Press, 2003.

Find links to papers on scale-free networks at [www.nd.edu/~networks](http://www.nd.edu/~networks)