THE THIRTEENTH LINK

# *The Map of Life*

IN FEBRUARY 1987 the journal *Nature* reported a landmark discovery: the gene for manic depression, or by its more recent name, bipolar disorder. Manic depression affects 1 to 5 percent of adults in the United States, and as many as 25 to 50 percent of those attempt suicide at least once. Because the risk of developing manic depression is five to ten times higher if first-degree relatives have the disease, the prevailing view is that manic depression is a genetic disorder. So as soon as methods for linking illnesses to specific genes emerged, the race was on to find the manic depression gene. The much coveted "first" seemed to have gone to the authors of the 1987 *Nature* paper, who located the gene on chromosome 11 while studying a large Amish family in Lancaster, Pennsylvania. Yet two years later the research group recanted the results. The blunder did not discourage other gene hunters, however. If anything, it gave them extra motivation to find the real gene. In 1996, almost a decade after the first published study, three independent research groups reported links to genes on other chromosomes. Another Amish study implicated chromosomes 6, 13, and 15; a study focusing on the isolated population of Costa Rica's Central Valley documented links to chromosome 18; and results derived from a large Scottish family indicated the involvement of chromosome 4. Research on another prominent mental disorder, schizophrenia, followed a similar pattern, linking the disease to two different

regions of chromosome 1, with a different research group implicating chromosome 5 a few years later.

Absentminded scientists? Bad research? Far from it. These are not conflicting results. They simply demonstrate that most illnesses, ranging from manic depression to cancer, are not caused by a single malfunctioning gene. Rather, several genes interacting through a complex network hidden within our cells are simultaneously responsible. Faced with the gigantic task of figuring out the building blocks of the cell, from genes to proteins, scientists until recently focused on biology rather than networks. But with the pieces now in hand, postgenomic biology is taking a step back to grasp the big picture. New and exciting discoveries that are revolutionizing biology and medicine tell us loud and clear: If we want to understand life—and ultimately cure disease—we must think networks.

## *1.*

"Today we are learning the language in which God created life," said President Bill Clinton on June 26, 2000, at the White House ceremony announcing the decoding of the 3 billion chemical "letters" of the human genome. Is it true? Has humanity been handed the "book of life"? Are Francis Collins and Craig Venter, the two gentlemen who stood on either side of the president, the prophets of the twenty-first century? After all, Collins and Venter, representing the publicly funded Human Genome Project and the private Celera Genomics, which each decoded the human genome, brought the book to us.

Open the "book of life" and you will see a "text" of about 3 billion letters, filling about 10,000 copies of the *New York Times* Sunday edition. Each line looks something like this:

TCTAGAAACA ATTGCCATTG TTTCTTCTCA TTTTCTTTTC ACGGGCAGCC

These letters, abbreviations of the molecules making up the DNA, could easily mean that the anonymous donor whose genome has been sequenced will be bald by the age of fifty. Or they could reveal that he

will develop Alzheimer's disease by seventy. We are repeatedly told that everything from our personality to future medical history is encoded in this book. Can you read it? I doubt it. Let me share a secret with you: Neither can biologists or doctors.

To be sure, the sequencing of the human genome is a triumph, the result of modern molecular biology's ability to reduce complex living systems to their smallest parts. It is undoubtedly a catalyst of a new era in both medicine and biology. But the genome project has brought along a new realization: The behavior of living systems can seldom be reduced to their molecular components.

Our inability to find a single gene responsible for manic depression is the best illustration. A list of suspected genes is not sufficient. To cure most illnesses, we need to understand living systems in their integrity. We need to decipher how and when different genes work together, how messages travel within the cell, which reactions are taking place or not in any given moment, and how the effects of a reaction spread along this complex cellular network. To achieve this we must map out the network within the cell. This web of life determines whether a cell develops into skin or labors constantly in the heart, decides the cell's response to external disturbances, holds the key to survival in constantly changing environments, tells the cell when to divide or die, and is responsible for illnesses ranging from cancer to psychiatric disorders. As the historic *Science* article that reported the decoding of the human genome concluded, "there are no 'good' genes or 'bad' genes, but only networks that exist at various levels."

## 2.

The decoding of the human genome offered us an inventory of the cell's parts. To return to our car analogy, it is like having thousands of car parts in your backyard. If you ever want to see that car running again, you must find the blueprint, a map telling you how to assemble it. For most cells this map is almost as elusive now as it was fifteen years ago at the beginning of the Human Genome Project. The absence of a cellular search engine is only part of the problem. The biggest difficulty

is that within each cell there are many layers of organization that can each be viewed as a complex network. To understand the web of life, we need to acquaint ourselves with some of these.

In today's weight-conscious society, it is common knowledge that cells burn food by splitting complex molecules to create the cells' building blocks and the energy they require to stay alive. This is achieved through a web of hundreds of multistep intracellular biochemical reactions, together referred to as the *metabolic network*. The nodes of this network can be simple chemicals, such as water or carbon dioxide, or more complex molecules made of dozens of atoms, such as ATP. The links are the biochemical reactions that take place between these molecules. If two molecules, A and B, react with each other to create C and D, then all four of them are connected in the cell's complex metabolism.

Think of the cellular metabolism as the engine in your car. Having an engine in and of itself will not get you very far. You need wheels, suspension, brakes, lights, and many other components, each ensuring that the car will run safely on the road. In a similar vein, the cell has an intricate *regulatory network* that controls everything from metabolism to cell death. The nodes of this network are the genes and the proteins encoded by the gigantic DNA molecule. The links are the various biochemical interactions between these components. The genes are first copied into unique messenger RNA molecules, which are then translated into proteins. Some proteins interact with the DNA, initiating or suppressing the translation of new genes, repairing accidental DNA damage, copying the two strands of DNA when the cell replicates, and so on. Other proteins interact with each other, forming large protein complexes. A prominent example is hemoglobin, a protein complex made of four proteins that bind together to transport oxygen in our bloodstream. Therefore, proteins can be viewed as nodes of a complex protein-protein interaction network in which two proteins are connected if they can physically attach to each other. The full weblike molecular architecture of a cell is encoded in the *cellular network*, a sum of all cellular components (genes, proteins, and other molecules), connected by all physiologically relevant interactions, ranging from biochemical reac-

tions to physical links. This web of life contains all metabolic, protein-protein, and protein-DNA interactions present in the cell.

Not too long ago it was widely believed that everything that matters for an organism's biological history is encoded in the genes. Postgenomic biology, though still in its infancy, is already fighting an important battle. It aims to diminish the all-encompassing role historically attributed to individual genes. Genes are known to play a *structural* role, determining the scope and make of proteins and passing this information in a hereditary manner to subsequent generations. Recently, however, scientists have discovered that genes also play an important *functional* role as members of a complex cellular network. This functional role is apparent only in the dynamic context in which an individual gene interacts with many other cellular components. The gene's structural role can be unearthed from its sequence. We now have the complete sequence for several key organisms, ranging from *Esherichia coli* bacteria to humans. We are only at the beginning, however, of the second, equally revolutionary scientific endeavor: uncovering the gene's functional role. To achieve this we need a second genome project, this time mapping the web within the cell. We have the "book of life." Now we need the *map of life*.

## 3.

Zoltán Oltvai, a cell biologist at Northwestern University Medical School in Chicago, had several significant and much cited discoveries under his belt when we met in 1998. At that time we both lived in Oak Park, a Chicago suburb styled by the towering architectural presence of Frank Lloyd Wright. With small children of similar age, we started to visit each other regularly. After exhausting all topics related to culture and politics, our conversation turned to science and biology. By then my group was pressing ahead with research on the Web and Internet. Inevitably, our weekend chats drifted toward the similarities and differences between the web of life and other complex networks. Soon an ongoing argument developed. The Web and the actor network are scale-free because they emerged thanks to growth and preferential attachment,

processes that are easily identifiable in both networks. The cell, on the other hand, is different. To be sure, the original assembly of the first protocells from a primordial soup of organic molecules might have resembled a growing network. But during the past three billion years evolution and natural selection took their course. During this time there was significantly less growth, just a lot of tinkering with the cellular network, streamlining and optimizing it. Thus, on the one hand, even if a scale-free topology had developed when lifeless molecules took their first steps towards life, it might have been lost because of the all-encompassing effects of evolution. On the other hand, it is hard to fathom that the complex biochemical web within the cell would be completely random. So is the map of life, like the Erdős-Rényi network, random, or is it scale-free, like the Web? How do we characterize the cell's complex topology?

After we ran out of arguments to convince ourselves one way or another, Oltvai and I decided to move our discussions off the playground and look for real data on the web of life. Fortunately, for most of the twentieth century, biology and biochemistry were devoted to identifying and interrelating the various molecules within the cell. James Watson, the codiscoverer of the double helix structure of DNA, wrote in 1970 in the now classic *Molecular Biology*, "We already know at least one-fifth, and maybe more than one-third of all metabolic reactions that will ever be described [in *E. coli* bacteria]," suggesting that "within the next ten to twenty years we shall approach a state in which it will be possible to describe essentially all metabolic reactions." Watson's vision has been fulfilled. Today bacteriologists believe that the complex network of more than seven hundred nodes and close to a thousand links represents pretty much the full list of reactions fueling the *E. coli* metabolism. What Watson could not have imagined in 1970 is that thirty years later online databases would be compiling the network of metabolic reactions for hundreds of organisms. While we are still missing a detailed metabolic map of the highly complex human cell, our knowledge of several simpler organisms is close to being complete.

So my discussions with Oltvai could not have been better timed. A few years earlier my lab's quest to study the cell's topology would

have been brought to a halt by an absence of data. In late 1999, however, several Websites had the maps we were looking for. After researching the available databases, we settled on a new one, run by the Argonne National Laboratory outside Chicago, nicknamed "What Is There?" which compiled the metabolic network of forty-three diverse organisms. Hawoong Jeong, once again displaying his computer wizardry, wrote a program that downloaded each reaction individually. Oltvai and I watched over his shoulder as he made sense of this extremely complex web, assembling one by one the full metabolic map for these forty-three organisms. Having finished that, he moved on to characterize these networks, calculating how many reactions each molecule participates in. The robustness of the results was shocking. No matter which organism we examined, a clear scale-free topology greeted us. Each cell looked like a tiny web, extremely uneven, with a few molecules involved in the majority of reactions—the hubs of the metabolism—while most molecules participated in only one or two.

## *4.*

To harken back to our social networks, if two molecules participate in the same reaction, their separation is one. If, however, two subsequent reactions are needed to connect them, their separation is two. Putting all nodes and links together, will this complex network within the cell have small-world properties?

Measuring the separation between molecules is not an outgrowth of our obsession with six degrees of separation. The diameter of the network—or degree of separation between nodes—has biological significance. For instance, if we should find that the shortest chemical path between two molecules is one hundred, then any change in the concentration of the first molecule will have to go through one hundred intermediate reactions before reaching the second molecule. Any perturbation will decay and die along such a long path.

To our great surprise, the measurements indicated that the typical path lengths are much shorter than one hundred. In fact, cells are small

worlds with *three degrees of separation*. That is, most pairs of molecules can be linked by a path of three reactions. Perturbations, therefore, are never localized: Any change in the concentration of a molecule will shortly reach most other molecules. This finding was supported by the study of Andreas Wagner, from the University of New Mexico, and by David A. Fell, from Oxford Brooks University, who independently concluded that the *E. coli* metabolic network is scale-free and has small-world properties.

Though unexpectedly short, the *three* degrees was not the most interesting aspect of our finding. Because the forty-three organisms all had different sizes, we expected that the separation would increase with the organism's size, just as the Web's diameter increases with the number of documents. Surprisingly, the measurements indicated that whether we are navigating the tiny network of a small parasite bacterium or the highly developed highway system of a multicellular organism, such as a flower, the separation is the same. Although the difference in the cellular architecture between a primitive bacterium and a cell from a multicellular organism could be as large as the difference between a tiny village and New York City, stripped to their dynamically relevant networks, all cells feel like a small town. Digging deeper, we learned that most cells share the same hubs as well. That is, for the vast majority of organisms the ten most-connected molecules are the same. Adenosine triphosphate (ATP) is almost always the biggest hub, followed closely by adenosine diphosphate (ADP) and water.

To be sure, the role of ATP, ADP, and water as prominent hubs was by no means surprising. In cells, ATP serves as a convenient and versatile store of energy, driving hundreds of biochemical reactions. By supplying energy to these reactions, ATP turns into ADP by giving up a phosphate group; thus, within the metabolic web, both ATP and ADP are linked to a huge number of molecules participating in energy-hungry reactions. Yet, taken together, the top-ten list of highly connected molecules was rather revealing. A key prediction of the scale-free model is that nodes with a large number of links are those that have been added early to the network. In terms of metabolism this would imply that the most connected molecules should be the oldest ones within

the cell. And indeed, the analysis of Wagner and Fell has shown that the most-connected molecules have an early evolutionary history as well. Some of these molecules are believed to be the remnants of the so-called RNA world, the evolutionary step before the emergence of DNA, while others are known to be the components of the most ancient metabolic pathways. Therefore, the first mover advantage seems to pervade the emergence of life as well.

If all organisms have the same scale-free topology and the same node separation and share the same hubs, how do cells of different organisms differ from one another? Is there any difference between the chemical architecture of a bacterium and that of a human cell? It turns out that there are significant differences. Comparing the metabolic network of all forty-three organisms, we found that only 4 percent of the molecules appear in all of them. Though the hubs are identical, when it comes to the less connected molecules, all organisms have their own distinct varieties. Life looks like a suburb in which each house was designed by the same architect, but different builders and interior designers were commissioned to offer the finishing touches, from the material of the floor to the size and make of the windows. In an aerial photograph all houses appear to be alike. The closer you get to them, however, the more you start noticing the differences.

Metabolism represents only one component, albeit an important one, of the cellular network. Will the same scale-free architecture also be present in the regulatory network—the web responsible for running the cell? Indeed, we are ultimately interested in the full weblike molecular architecture of living organisms. The question is, do the different components of this web of life follow the same laws and architectural features, or has evolution discovered different solutions for the various components? Beyond our desire to comprehend the fundamental features of the cell's architecture, understanding the regulatory network has important practical implications as well. Indeed, genetic disorders result from malfunctions of the nodes of the regulatory network. Therefore, the robustness of this network to node failures determines our ability to survive various diseases, as well as researchers' ability to design drugs that can cure those disorders that we cannot easily tolerate.

## 5.

Baker's yeast, one of the simplest eukaryotic cells, has about 6,300 genes, encoding about the same number of proteins. Though this is only a fifth of the estimated 30,000 different genes a human cell contains, it is already an enormous number. In general, when proteins interact by sticking to one another, they have a good reason for doing so. Most interactions play some important functional role in the cell's life. Therefore, to understand how cells work we must identify all pairs of proteins that can interact. For baker's yeast, that requires checking 6,300 times 6,300 pairs—close to forty million potential interactions. With standard molecular biology tools this would take decades and hundreds of people. Yet, despite the magnitude of the job, two research groups have independently obtained a detailed map of the yeast protein network. They succeeded thanks to an important technological breakthrough, the so-called two-hybrid method. Developed by Stanley Fields in 1989, the two-hybrid method offers a relatively rapid semiautomated technique for detecting protein-protein interactions. Though the method is known to provide numerous false negatives and positives, the map it generated offers an unprecedented opportunity to peek into the cell's regulatory organization.

Electrified by the insights offered by the topological analysis of cellular metabolism, in the fall of 2000, Oltvai, Jeong, and I, together with a young student, Sean Mason, became interested in the structure of the protein interaction network. The two-hybrid data, published a few months earlier, offered an excellent opportunity for such a study. After downloading all known protein-protein interactions, we reconstructed the protein network of yeast with the aim of studying its large-scale features. Once again, the results left little room for ambiguity: They demonstrated that the protein interaction network has a scale-free topology. That is, most proteins in the cell play a very specific role, interacting with only one or two other proteins. A few proteins, however, are able to physically attach to a huge number of other proteins. These hubs are crucial for the cell's proper functioning and survival. Indeed, we were able to

show that removing a gene responsible for a hub protein kills the cell 60 to 70 percent of the time. Mutations affecting a weakly connected protein, in contrast, have a less than 20 percent likelihood of proving lethal.

A series of parallel results supported these findings. Andreas Wagner independently confirmed that the yeast protein network has a scale-free topology. Stefan Wuchty, a young researcher working at the European Media Laboratories, found a similar architecture in a markedly different network within the cell. In his so-called protein domain network, the nodes are different facets through which proteins link to each other, two facets being considered connected if they are simultaneously present on the same protein. Jong Park and collaborators from the European Bioinformatics Institute in the United Kingdom spotted a scale-free topology when they reconstructed the yeast network from protein interaction data collected by the Protein Data Bank. Our research group has found the same structure in an organism very different from yeast, a simple bacterium called *Helicobacter pylori*, suggesting that the scale-free nature of the protein interaction network is a generic feature of all organisms.

Taken together, the similar large-scale topology of the metabolic and the protein interaction networks indicate the existence of a high degree of harmony in the cell's architecture: Whichever organizational level we examine, a scale-free topology greets us. These journeys within the cell indicate that Hollywood and the Web have only rediscovered the topology that life had already developed 3 billion years earlier. Cells are really small worlds that share the topology of many other nonbiological networks, as if the architect of life could design only these.

How did life arrive at this architecture? Almost as soon as we asked the question, we had an answer. Approximately a half year after the publication of our findings on the topology of the protein interaction network, I received three e-mails within about a month. Each of them contained a manuscript by a different research group. Amazingly, each of the three research groups independently offered the same simple and elegant explanation, claiming that the cell's scale-free topology is a result of a common mistake cells make while reproducing.

## 6.

Cells reproduce by duplicating their content and dividing into two. The details of these processes may vary for simple bacteria and more complex human cells. Certain steps are universal, however. First, in order to produce a genetically identical offspring cell, the DNA must be faithfully replicated. This process is not free of errors, however. Although the cell's intricate copying mechanism insures that DNA sequences are inherited with extraordinary fidelity, about one letter in a thousand is randomly changed every 200,000 years. Another common error is gene duplication. Through a rare accident in the copying process, gene duplication can occur when the ends of broken DNA molecules join together. As a result, segments of varying length of the parent DNA will appear twice in the offspring's genome. Such copying mistakes sometimes kill the cell. In other cases, multiple copies of the same gene have evolutionary advantages and are passed on to future generations. Hemoglobin is a well-known example.

Originally cells had only one hemoglobin gene. About 500 million years ago, during the evolution of the higher fish species, a series of gene duplications occurred, resulting in four copies of the hemoglobin gene scattered along the genome. Today each of these genes encodes one of the four components of the hemoglobin protein complex.

Gene duplication has a significant impact on the cellular network. It results in two identical genes, which produce identical proteins, that in turn interact with the same proteins. A new node thus has been created, the protein generated by the duplicated gene. Its neighbors, the proteins with which the duplicated protein interacts, will each now interact with both the parent and the identical offspring protein. Therefore, each protein in contact with the duplicated protein gains an extra link. In this game highly connected proteins have a natural advantage: They are more likely to have a link to the duplicating protein than their weakly connected cousins. It's not that hubs duplicate more often. Rather, since the hubs are in contact with more proteins, they are more likely to have a link to a

duplicating node, which offers them an extra link, a subtle version of preferential attachment.

The most important feature of this explanation is that it traces the origin of the scale-free topology back to a well-known biological mechanism, gene duplication. It does so by showing that gene duplication can simultaneously lead to both the growth of the protein network by adding an extra protein and to preferential attachment by adding new links at a higher rate to the more connected proteins. It is too early to determine if this is the only explanation, since it is conceivable that different mechanisms, yet unexplored, could generate the same topology. It is unclear if it explains the scale-free structure seen in the metabolism, as well. Nevertheless, it demonstrates that mechanisms present in the cell can generate the scale-free topology. Therefore, at this point we are ready to turn to the next important question: Will the map of life help us better understand diseases and enhance our ability to eventually cure them?

## 7.

Cancer is the most researched human illness ever. The extraordinary attention the medical community has devoted to it has resulted in several significant breakthroughs. Probably the most important is the discovery of the p53 gene. Though reported as early as 1979 by David Lane and Arnold J. Levine, it was not until the late 1980s following the work of Bert Vogelstein that its role in cancer was fully appreciated. Vogelstein recognized that the p53 protein, created by the p53 gene, is a tumor suppressor. Just as your brakes allow you to stop your car, tumor suppressor genes act to slow and halt DNA replication and division into new cells. Healthy cells keep a small number of p53 molecules around. If radiation or some other injury damages the cell, more p53 is produced, preventing the progression of the cell through cell division. This gives the cell time to repair the damage before further copies of the malfunctioning cell can be produced. However, if the damage is irreparable, the p53 protein will activate a group of genes to kill the cell.

If the cell's brake—the p53 protein—malfunctions, the cell can run amok. Cancerous cells differ from healthy cells in their ability to multiply at a very high rate. Indeed, about 50 percent of human cancers contain mutations in the p53 gene. This observation has stimulated an avalanche of research, resulting in over 17,000 publications since 1989. In recognition of its central role in cancer, in 1993 the p53 molecule was named "Molecule of the Year" by *Science*. Considering the attention the p53 molecule has received, one might have expected that a cure for cancer would have been found by now. After all, all we need to do is to develop drugs that make sure the p53 molecule always does its job. Why, then, has this huge amount of research not yet translated into a universal cancer drug?

Despite its important role in human cancer, fixing the p53 gene alone will not lead to a cure for this deadly disease. The reason was recently articulated by the very people responsible for placing p53 at the center of cancer research. Vogelstein, Lane, and Levine in November 2000 coauthored a *Nature* paper that made networks the crux of their argument. The reason why we do not fully understand cancer, the three suggested, is that the cell is like the Internet.

The three researchers argued that we must stop our obsession with the omnipresent p53 molecule and focus instead on what they called the *p53 network*, a sum of all molecules and genes interacting with the p53 molecule. As they put it, "One way to understand the p53 network is to compare it to the Internet. The cell, like the Internet, appears to be a 'scale-free network': a small subset of proteins is highly connected (linked) and controls the activity of a large number of other proteins, whereas most proteins interact with only a few others. The proteins in this network serve as the nodes, and the most highly connected nodes are the hubs. In such a network, performance is almost unchanged by random removal of nodes. But such systems contain an Achilles' heel."

The "Achilles' heel" of a network, you'll recall, refers to the vulnerability of its hubs. The inactivation of less connected molecules does not have draconian effects on the cell, whereas a mutation in the p53 molecule, one of the clear hubs of the cellular network, turns the cell cancerous and eventually kills the organism. This explains why com-

bined pharmaceutical attacks on molecules that interact with the p53 molecule have progressively more severe effects on the cell, resembling an attack on the p53 molecule itself.

Vogelstein, Lane, and Levine's *Nature* paper demonstrated the strength and ubiquity of network thinking. Ideas developed to better defend the Internet and quantify the effects of hacker attacks have fallen on fertile ground in cell biology, which is concerned with the defense of healthy human cells against all threatening organisms. At the heart of Internet research and cell biology, the questions are similar. The first step is to map out the network behind these systems. Then from these maps we need to infer the laws that govern the networks. At that point the Internet topographer, the Web mapper, and the cancer researcher will be in the same camp.

Yet the most important implication of the p53 network goes beyond the fundamental analogies it illuminates between cells and the Internet. It points to a new approach to drug therapies and drug development. The ultimate goal of studying the p53 network is to find a cure for cancer. As we discuss next, this is largely a trial-and-error process. In most cases cancer therapies aim for destruction: They kill the cancerous cells by disrupting their cellular network with either drugs or radiation. The increasing understanding of the p53 network suggests another avenue: We must first decipher the precise topology of this network, fully understanding all interactions. With such a map in hand, we can start a frontal attack, finding drugs that restore the functions of the p53 molecule without dismantling the network around it.

## 8.

Until recently we could only treat the symptoms of illnesses like cancer, heart disease, and psychiatric disorders. We searched for rare chemicals everywhere, from chemistry labs to rain forests, hoping that they would offer miracle drugs for some diseases. According to some estimates, the drugs available on the market target only about 500 of the 30,000 proteins in the human body. And though we have multiple drugs for many

diseases, it is often a trial-and-error process to figure out which works for a given patient.

A detailed understanding of the full biochemical network within the cell promises to eliminate this guesswork. With knowledge of the precise wiring diagram of a cell and diagnostic tools capable of capturing the strength of the various cellular interactions, doctors in the future could test the response of your cells to a drug before you even take it. Thanks to the map of life, which implies a detailed understanding of how genes work together, we will someday be able to diagnose diseases like manic depression or cancer before any of the symptoms have occurred. This knowledge will help us develop drugs that are so fine-tuned and highly precise that they affect only the malfunctioning cells, leaving the healthy cells alone. In other words, they will provide real cures.

Changing the concentration of a chemical in your body via a drug could reduce the symptoms of a particular disease. However, since the cell is controlled by a complex network with small-world properties, a drug-induced perturbation inevitably affects many other chemicals, possibly creating undesired side effects. Patients treated for manic depression might die of heart disease, a condition they had never experienced before. Furthermore, the drug that causes heart disease for you could have no side effects on another individual. We all have different eye and hair colors and facial features, after all, so it is not surprising that we metabolize drugs differently as well. With the map of life in hand and with tools such as the recently developed DNA chips that monitor the links between the genes, doctors will be able to obtain a detailed list of all molecules and genes affected by a given drug. Exploring side effects will no longer be guesswork. We will have personalized medicine, allowing the marketing and approval of drugs that are effective for only 10 percent of the population and potentially lethal for everybody else.

## 9.

If you suffered from manic depression in recent years, your first visit to the doctor probably started with an hour-long discussion to carefully

examine your thoughts and feelings. Eventually you walked away with a drug prescription. If you had never appreciated how much of your brain's activity and well-being was a matter of brain chemistry, now, after taking the drug, you did. A layer of chemicals injected into your body rapidly took over your behavior and impulses. You discovered yourself doing things and having feelings that you never experienced before. In most cases the first drug didn't work. It perhaps made you hyperactive or even more depressed. A few weeks later the drug had to be switched for another, in hope of better results. Patients would routinely try five or six drugs over a period of several months before finding the one that worked best. While they made you feel better, these drugs didn't cure your illness. They temporarily altered your brain's chemistry, offsetting the changes caused by the malfunctioning of your genetic network. If you stopped taking them, the chemical imbalance would return, along with the symptoms of manic depression.

Twenty years from now things could look quite different. Facing the same doctor, you will have a five-minute discussion, just as you do in cases of simple influenza. An assistant will take a few drops of blood, and you will walk home empty-handed. In the evening you will pick up the medicine from the nearest pharmacy. The next day you will wake up fresh and happy, as you did before your symptoms appeared. Both the manic and the depressive you will have been washed away.

How will this breakthrough come about? First, the full biochemical network of the human cell will have been mapped by then, allowing us to understand in detail how different genes and molecules work together. Second, DNA and protein chips, new technologies now under development, will be in each doctor's office, allowing her or him to monitor which genes and proteins malfunction in your cells. While mapping the human cellular network will probably take over a decade, the instant monitoring of gene activity is already possible in some research labs.

By 2020 these advances will change medicine across the board. Kids will not be taken to the doctor with a sore throat—Mom will have a handheld device, with a replaceable chip, that will reveal that Tommy's sore throat is a streptococci infection, identifying the strain as well. She will be able to link the device to the computer and e-mail the

profile to the doctor's office, so when Tommy shows up for school the drug is ready for him in the nurse's office. Most important, Tommy's drug will not be a strafing antibiotic that kills all bacteria, harmful or not, in his body. It will be designed and mixed on the spot to take out only the organism that made Tommy's throat hurt. It will be ineffective against any other bacteria, minimizing the chance that Tommy will develop an antibiotic resistance.

I don't believe that this vision is far-fetched. In fact, it is rather modest, perhaps even shortsighted. It is only a simple interpolation of the tools already present in most research laboratories around the world. These advances are rooted in a fundamental shift in how we look at everything from life to disease. They are the result of seeing the cell as a whole—as a network—rather than a bag of independent chemicals.

## *10.*

The genome project is the ultimate celebration of the *gene*. Until recently we believed that the complete biological history of a human being was encoded in the 3 billion letters of the helical DNA. To be sure, the mapping of the human genome revolutionized biological research. But it also showed us what a small fraction of the vast world is really known to us and how much more is left to be explored.

In 1996 the decoding of the yeast genome gave the scientific community a shock: It contained as many as 6,300 genes. Only about a quarter of these were expected and could be assigned vague functions. To be on the safe side, and boosted by humans' perceived importance as the pinnacle of evolution, biologists estimated that the human genome would have at least 100,000 genes. This number was believed to be sufficient to account for the high complexity of *Homo sapiens*. Then came February 2001 and the publication of the human genome. It turned out that we have less than a third of the anticipated genes—only about 30,000. Therefore, a mere one-third increase in genes must explain the difference between us and the unsophisticated *Caenorhabditis elegans* worm—quite a provocative idea when we consider that the 20,000

genes of *C. elegans* need to encode only three hundred neurons, whereas our extra 10,000 genes have to account for the billion nerve cells present in our brain.

In short, it is now clear that the number of genes is not proportional to our perceived complexity. Then what does complexity mean? Networks point to the answer. Framed in terms of networks, our question becomes: How many different potentially distinct behaviors can a genetic network display with the same number of genes? In principle, two cells that are identical except that a specific gene is *on* in the first cell and *off* in the second could behave differently. Assuming that each gene can be turned on or off independently, a cell with $N$ genes could display $2^N$ distinct states. If we adopt as a measure of complexity the potential number of distinct behaviors displayed by a typical cell, the difference between the worm and humans is staggering: Humans could be viewed as $10^{3,000}$ times more complex than our wormy relatives!

Whereas the twentieth century was seen as the century of physics, the twenty-first is often predicted to be the century of biology. A decade ago it would have been tempting to call it the century of the gene. Few people would dare say that any longer about the century we have just entered. It will most likely be a century of complexity. It must be a century of biological networks as well. If there is any area in which network thinking could trigger a revolution, I believe that biology is it.

THE FOURTEENTH LINK

# *Network Economy*

TEN YEARS AGO AN EARLY and largely unknown Internet startup was desperately short of cash. As a manager for Time Warner, a member of the startup's directorial board saw these problems as an opportunity for the entertainment giant for which he worked. He therefore suggested to a Time Warner senior executive that they bail out the startup. For a mere $5 million the media conglomerate could have owned 11 percent of the company. This would have been petty cash for Time Warner and would have offered access to the Internet, at that time a brand new distribution channel. "If we did that," the senior executive replied, meaning that if he accepted the Internet as a viable distribution channel for Time Warner, "then everything we have done since 1923 would be thrown out the window."

He certainly was a terrible stock picker: Ten years later the $5 million investment would have been worth over $15 billion. The purchase would have altered history too. Indeed, a decade later Steve Case, the CEO of America Online (AOL), the once unknown Internet startup, and Jerry Levin, the chairman of Time Warner, announced the merger of the two companies at a Manhattan press conference. A few years earlier Time Warner could have easily digested the Internet startup. In 2000, however, it was AOL, a company that few had heard of a decade earlier, that swallowed the media giant.

Time Warner had content, and AOL had the means of delivering it to the consumer. Just before the collapse of the NASDAQ bubble in

spring 2000, Jerry Levin was under pressure to go dot.com to regain Wall Street's attention, and Steve Case needed access to Time Warner's cable to get into your living room. Despite the very different cultures of the two companies, business analysts were eager to convince us that it was a match made in heaven. The same analysts had told us that the 1998 Daimler-Benz takeover of Chrysler also was a sound step for both companies. So was the fusion of the oil industry titans Exxon and Mobil in 1998, four months after another major acquisition in which Amoco was bought by British Petroleum. The list of attention-grabbing mergers and acquisitions does not end here, however. In 1998 alone Bell Atlantic paired up with GTE, SBC Communications bought Ameritech, BankAmerica joined up with NationsBank, Citicorp merged with Travelers Group.

Do these mergers make sense? Not if you listen to antiglobalization activists, who accuse big corporations of dictating everything from policy to fashion. They are unavoidable, however, if we view the economy as a complex network, whose nodes are companies and whose links represent the various economic and financial ties connecting them. Indeed, in a network economy the hubs must get bigger as the network grows. To satisfy their hunger for links, nodes of the business web learn to swallow the smaller nodes, a novel method unseen in other networks. As globalization pressures the nodes to grow bigger, mergers and acquisitions are a natural consequence of an expanding economy.

Motivated by the renaissance of networks in physics and mathematics, recently a number of new findings has documented the power of networks in everything from company structure to the marketplace. We have learned that a sparse network of a few powerful directors controls all major appointments in Fortune 1000 companies; a network of alliances determines the success in the biotech industry; the structure of the network within the firm is responsible for the organization's ability to adapt to rapidly changing market conditions; and strategies taking advantage of the network nature of the consumer base lead to phenomenal successes in marketing. As links and connections take over, understanding network effects become the key to survival in a rapidly evolving new economy.

## *1.*

Regardless of industry and scope, the network behind all twentieth century corporations has the same structure: It is a *tree*, where the CEO occupies the root and the bifurcating branches represent the increasingly specialized and nonoverlapping tasks of lower-level managers and workers. Responsibility decays as you move down the branches, ending with the drone executors of orders conceived at the roots.

Despite its pervasiveness, there are many problems with the corporate tree. First, information must be carefully filtered as it rises in the hierarchy. If filtering is less than ideal, the overload at the top level, where all branches meet, could be huge. As a company expands and the tree grows, information at the top level inevitably explodes. Second, integration leads to unexpected organizational rigidity. A typical example comes from Ford's car factories, one of the first manufacturing plants to fully implement the hierarchical organization. The problem was that they got too good at it. Ford's assembly lines became so tightly integrated and optimized that even small modifications in automobile design required shutting down factories for weeks or months. Optimization leads to what some call *Byzantine monoliths*, organizations so overorganized that they are completely inflexible, unable to respond to changes in the business environment.

The tree model is best suited for mass production, which was the way of economic success until recently. These days, however, the value is in ideas and information. We have gotten to the point that we can produce anything that we can dream of. The expensive question now is, what should that be?

As companies face an information explosion and an unprecedented need for flexibility in a rapidly changing marketplace, the corporate model is in the midst of a complete makeover. This does not mean a superficial shift in the job description of a few individuals. It is a fundamental rethinking of how to respond to the new business environment in the postindustrial era, dubbed the information economy.

The most visible element of this remaking is a shift from a tree to a web or a network organization, flat and with lots of cross-links between the nodes. As valuable resources shift from physical assets to bits and information, operations move from vertical to virtual integration, the reach of businesses increasingly expands from domestic to global, the lifetime of inventories decreases from months to hours, business strategy changes from top-down to bottom-up, and workers transform into employees or free agents.

New products require new alliances both within and outside the company, demanding a new topology. To achieve this, layers of middle managers have been scrapped. Employees who previously played secondary roles are in charge of major products from one day to the next. Project teams, alliances within and outside the organization, and outsourcing proliferate. Therefore, companies aiming to compete in a fast-moving marketplace are shifting from a static and optimized tree into a dynamic and evolving web, offering a more malleable, flexible command structure. Those that resist this change could easily be forced to the periphery.

The internal remaking of the web within the firm is only one consequence of a network economy. Another is the realization that companies never work alone. They collaborate with other institutions, adapting business practices proved successful in other organizations. The crucial high-level connection to the rest of the corporate world is often maintained by the CEO and the board of directors. As we will see next, network effects play a fundamental role in these interactions.

## *2.*

"I want to say to you absolutely and unequivocally that Ms. Lewinsky told me in no uncertain terms that she did not have a sexual relationship with the President," read Vernon Jordan at a hastily convened press conference in the midst of the Clinton-Lewinsky scandal. But he soon was to "pull off some of the fanciest footwork of his career—dancing out of the box that he put himself in," according to *Time* magazine's Eric Pooley, as everyone pressed him for a satisfactory explanation for the four meetings

and seven phone calls Jordan had with the former White House intern, trying to arrange a job for her at one of several major companies.

Jordan's role in finding Monica Lewinsky a corporate job was no surprise to Washington insiders. His inability to steer the attention away from himself was something new, however. An effective civil rights leader in the 1970s, Jordan was shot in the back in 1980 by a white supremacist, who settled on him after learning that Jesse Jackson, whom he really wanted to kill, was out of town. Jordan carefully had avoided the spotlight ever since, becoming the most powerful unknown in D.C., a rarely heard or seen top deal maker and superlawyer in Washington's media-fixated crowd. As Pooley wrote in *Time*, Jordan "earns $1 million a year from a law practice that requires him to file no brief and visit no courtroom, because his billable hours tend to be logged in posh restaurants, on cellular telephones, in the tufted-leather backseats of limousines—making a deft introduction here, nudging a legislative position there, ironing out an indelicate situation before it makes the papers."

Uncharacteristically, Jordan found *himself* in the papers all over the nation in 1998, his meetings and phone calls being scrutinized by everyone from the media to independent counsel Kenneth Starr. He emerged as a prominent node in the entangled web of the Clinton-Lewinsky scandal, often dubbed the Six Degrees of Monica.

Jordan was not a newcomer to small worlds. He acquired his unique status as a consummate Washington insider by successfully surfing one of the most influential small-world networks in the American economy, the corporate web. During the years preceding the Clinton-Lewinsky scandal and the Clinton presidency, Jordan became the most central director of the small corporate elite running the Fortune 1000 corporate world.

The board of directors, a group of about a dozen individuals, holds unusual power in overseeing a company's future. It is responsible for all major decisions, from ousting poorly performing CEOs to approving major mergers and acquisitions. Therefore, corporations make all efforts to recruit well-connected and experienced directors. Successful CEOs, lawyers, and politicians are frequently sought after, being courted for directorship on several boards at the same time.

Despite concerns that directors serving on a large number of boards cannot possibly find the time to do justice to all of them, most companies want their directors to have experience on other boards. As directors apply the knowledge and experience they acquired on one board to bear on questions faced by another, this interlocked network of board members plays a crucial role in spreading corporate practices and maintaining the political and economic clout of big corporations.

Thanks to the important role boards play in shaping the landscape of American corporate life, the web of directors has often been scrutinized in business literature. But only recently, with the advent of methods to analyze complex networks, have we started to understand to what degree the power of this web is rooted in its interlocked topology.

In the director network each node is a board member linked to directors serving on the same board. With thousands of companies, each with about a dozen or so directors, this is a rather large web. Gerald F. Davis, Mina Yoo, and Wayne E. Baker, from the University of Michigan Business School, recently studied the most influential component of this web, focusing on the network of Fortune 1000 companies, made up of 10,100 directorships held by 7,682 directors. If each director were to serve on one board only, the network would be broken into tiny, fully connected circles, each the size of a single board. This is not the case, however. While 79 percent of directors serve on only one board, 14 percent serve on two, and about 7 percent serve on three or more. The measurements indicated that these few overlapping directors create a small-world network with five degrees of separation. Indeed, the distance between any two directors belonging to the major cluster, which contains 6,724 directors, was 4.6 handshakes on average.

The small-world nature of the director web is due to the 21 percent of directors who serve on more than one board, since they are the ones who hold this complex network together. Of these, Vernon Jordan plays a very special role. With membership on ten boards, in which he regularly meets 106 other Fortune 1000 directors, Jordan is the most central director of the corporate elite, within three handshakes from most other directors.

## 3.

Jordan's career offers a vivid demonstration of how the interlocked, small-world nature of corporate directorships determines most major appointments in corporate life. Indeed, in most cases when Jordan joined a board, he already knew at least one director from his service on other boards. In the early seventies, as president of the National Urban League, the influential civil rights organization, Jordan repeatedly called for the inclusion of blacks in the powerful corporate elite. In 1972 John Brooks, the chairman of Celanese Corporation, a diversified manufacturer of chemicals, told him, "I think you ought to put your money where your mouth is. . . . You're talking about blacks on the board of directors. Why don't you come on the board at Celanese?"

Soon after joining the board of Celanese, Jordan received two calls inviting him to join the boards of both Marine Midland Bank and Bankers Trust. Undecided as to which he should accept, Jordan called John Brooks for advice. "You don't have a choice. It's Bankers Trust," came the short reply. When Jordan asked why, Brooks answered simply, "How do you think you got nominated to be on the Bankers Trust board? I am on the board. I nominated you." At Bankers Trust Jordan served together with William M. Ellinghaus, who held a directorship at JC Penney as well. A year later Jordan was invited to serve on the board of JC Penney.

Three years later Jordan asked Peter McCullough, the CEO of Xerox, to be the corporate chairman of the National Urban League. He accepted with a condition: "I'll be your corporate chairman if you come on the Xerox board." Jordan agreed. Three years after becoming a Xerox director, Jordan was invited to the board of American Express, where two other Xerox directors already served. It comes as no surprise that in 1980 Jordan joined the board of RJ Reynolds. Indeed, the CEO of Celanese and another JC Penney board member both served on the RJ Reynolds board, and Jordan had close links to the RJ Reynolds CEO as well, who was a fellow director on the Celanese board.

Prior acquaintanceship allows directors to vouch for prospective recruits. Therefore, the small-world dynamics help the creation of a powerful "old boy network," or corporate elite, that has unparalleled influence in economic and political life. Jordan's current job at Akin, Gump, Strauss, Hauer & Feld, one of the biggest law practices in Washington, can be also traced back to this old boy network: Robert S. Strauss, the partner responsible for recruiting Jordan, was a fellow director on the Xerox board.

Jordan's path is by no means unique. Network effects are known to be present in all industries. For example, in Silicon Valley the extensive movements of labor between companies create dense personal intercompany links. These subtle social networks are extensively utilized for hiring new employees and attracting managers. Since current employees can vouch for their social links, just as directors do for fellow board members, employees hired through social networks quit less frequently and perform better than those recruited otherwise.

The intricate and interlocked nature of board directorships and Silicon Valley employees provides just two examples of the complex social and power networks behind the U.S. economy. But to comprehend how an economy truly works, we need to understand how corporations and other economic institutions run by these highly connected directors interact with each other.

## *4.*

Although universities and their spin-offs, small biotech companies, have been recently the driving force behind the development of new drugs, the cash and experience needed to launch large-scale clinical trials and the worldwide marketing channels continue to be located in large chemical and pharmaceutical companies. Because the development and marketing of a new drug can cost anywhere from $150 million to $500 million, the different players of this field, ranging from universities and research labs to government agencies, chemical and pharmaceutical companies and venture capital firms, have been forced to form strategic partnerships. These alliances, together with the relatively

young age of the biotech industry, offer an unusually well documented case of network formation, allowing us to follow and understand the emergence of networks in economic systems.

From its early days the biotech industry displayed the essential attributes of a growing network. This growth was captured in a dynamic graph developed by Walter W. Powell, Douglas White, and Kenneth W. Koput, depicting the biotech network at different stages of its evolution between 1988 and 1999. In 1988, representing the early days of the industry, there were far fewer links than nodes: Seventy-nine organizations connected by only thirty-one links. According to the famous Erdős-Rényi prediction, the network should have been broken into many tiny clusters. In reality, however, the nodes formed two major components, one with twenty-seven and the other with four organizations. That is, none of the thirty-one links was wasted—each of them contributed to a major component developing around a few biotech companies, leading to a level of connectedness that could not emerge in a random network. A few hubs visible already at this early stage were the first-mover biotech companies, such as Centocor, Genzyme, Chiron, Alza, and Genentech. Without them the biotech network would have broken into many tiny disconnected nodes.

But the existence of a few companies with a large number of partnerships, resembling hubs, is not enough for us to identify the nature of the network. For this we have to analyze the degree distribution, a study recently performed by two economists, Massimo Riccaboni and Fabio Pammolli, both from the University of Siena, working with physicist Guido Cardarelli from La Sapienza University in Rome, Italy. Their study was based on data collected by the Pharmaceutical Industry Database, hosted by the University of Siena, which provides information for 3,973 research and development agreements between 1,709 firms and institutions. The analysis indicates that the hubs noticed by Powell, White, and Koput are not accidental but are rooted in the scale-free nature of the network behind the pharmaceutical industry. Indeed, the number of companies that entered in partnership with exactly $k$ other institutions, representing the number of links they have within the network, followed a power law, the signature of a scale-free

topology. A hierarchy of well-connected large corporations brought together a large number of small companies, seamlessly integrating all players into an evolving scale-free economy.

As research, innovation, product development, and marketing become more and more specialized and divorced from each other, we are converging to a network economy in which strategic alliances and partnerships are the means for survival in all industries. The interfirm linkages of suppliers and subcontractors are well documented in southwestern Germany and north central Italy; Japanese business has long relied on interfirm collaborations to diffuse responsibilities for technological innovations; the Korean business model marries a whole array of diverse companies under the umbrella of large conglomerates; Silicon Valley regularly takes advantage of technology transfers by pairing up startups with established companies. These fluid alliances, which are periodically renegotiated as the marketplace shifts or the focus of the participants changes, offer a glimpse of the future of the world's business environment.

## 5.

Despite the important role these interfirm alliances play in the economy, economic theory pays surprisingly little attention to networks. Until recently economists viewed the economy as a set of autonomous and anonymous individuals interacting through the price system only, a model often called the *standard formal model* of economics. The individual actions of companies and consumers were assumed to have little consequence on the state of the market. Instead, the state of the economy was best captured by such aggregate quantities as employment, output, or inflation, ignoring the interrelated microbehavior responsible for these aggregate measures. Companies and corporations were seen as interacting not with each other but rather with "the market," a mythical entity that mediates all economic interactions.

In reality, the market is nothing but a directed network. Companies, firms, corporations, financial institutions, governments, and all potential economic players are the nodes. Links quantify various interac-

tions between these institutions, involving purchases and sales, joint research and marketing projects, and so forth. The *weight* of the links captures the value of the transaction, and the direction points from the provider to the receiver. The structure and evolution of this weighted and directed network determine the outcome of all macroeconomic processes.

As Walter W. Powell writes in *Neither Market nor Hierarchy: Network Forms of Organization*, "in markets the standard strategy is to drive the hardest possible bargain on the immediate exchange. In networks, the preferred option is often creating indebtedness and reliance over the long haul." Therefore, in a network economy, buyers and suppliers are not competitors but partners. The relationship between them is often very long lasting and stable.

The stability of these links allows companies to concentrate on their core business. If these partnerships break down, the effects can be severe. Most of the time failures handicap only the partners of the broken link. Occasionally, however, they send ripples through the whole economy. As we will see next, macroeconomic failures can throw entire nations into deep financial disarray, while failures in corporate partnerships can severely damage the jewels of the new economy.

## 6.

On February 5, 1997, Somprasong Land, a Thai property development company, failed to pay interest of $3.1 million on Euro-convertible debt. In a globalized economy where trillions of dollars change hands daily, this is petty cash. Not surprisingly, the event easily evaded the attention of the average investor. Unnoticed by most, this single failure was nevertheless the spark that led to the melting of the world's financial architecture.

A month later the Thai government made the first in a series of desperate attempts to save the country's economy from imminent collapse, announcing that it would buy $3.9 billion in bad property debt from financial institutions. A few days later it reneged on its promise, a move that some financial experts took as a sign of stability. The

International Monetary Fund's managing director, Michel Camdessus, who was later criticized for his organization's role in the Asian financial meltdown, said, "I don't see any reason for this crisis to develop further."

Subsequent events proved him wrong. Two weeks later the financial sector was trembling in Malaysia, prompting its central bank to restrict loans. At the same time, Sammi Steel, the main firm of Korea's twenty-sixth largest conglomerate, sought court receivership, the first step toward bankruptcy. In May, Japan hinted that it would raise interest rates to stop the decline of the yen (which never happened), triggering a global sell-off of Southwest Asian currencies and shaking the local stock markets. A week later Thailand failed to save its largest finance company, Finance One, which effectively went bankrupt. The event triggered a strong speculative attack on Thailand's currency, the baht, which, despite repeated promises to the contrary by the government, was abandoned on July 2.

The cascading failures of companies and financial institutions in Thailand, Indonesia, Malaysia, Korea, and the Philippines would take hundreds of pages to fully document. So would the chronicle of finger-pointing, including such highlights as Malaysian Prime Minister Mahathir Mohamad's bitter attack on "rogue speculators," which culminated in a talk given to the IMF/World Bank annual conference in which he called currency trading immoral. George Soros, the prominent international financier, responded a day later, "Dr. Mahathir is a menace to his own country."

Some economists blamed the "structural and policy distortions in the countries of the region" for the financial meltdown. Yet President Clinton and his economic team in the economic report of the president to the Congress in 1999 maintained that the crisis "was not due to problems with the economic fundamentals." Less than a year after the events, Paul Krugman, professor of economics and international affairs at Princeton, summarized the overall feeling: "It seems safe to say that nobody anticipated anything like the current crisis in Asia." A few small, localized financial difficulties had set off a chain reaction of failures that swept across national boundaries, creating a huge currency de-

valuation and stock market crashes from Asia to South America. It eventually caused the single biggest point loss ever of the Dow Jones industrial average, which tumbled 554.26 points on October 27, 1997.

How could the failure of a large but far from dominant property development company shake the world's largest stock market and keep the president of the "world's strongest nation" explaining even two years after? If we view the economy as a highly interconnected network of companies and financial institutions, we can begin to make sense of these events. In such networks the failure of a node has little effect on the system's integrity. Occasionally, however, the breakdown of some well-selected nodes sets off a cascade of failures that can shake the whole system.

The Asian crisis was a large-scale example of a cascading financial failure similar to those we discussed in Chapter 9, a natural consequence of connectedness and interdependency. It was not the first, however: South America and Mexico had experienced similar cascading failures two years earlier. It is surely not the last either, despite all the measures banks and governments seem to have taken to avoid it.

These events cannot be explained within a framework in which all organizations interact with a mythical market only. Cascading failures are a direct consequence of a network economy, of interdependencies induced by the fact that in a global economy no institution can work alone. Understanding macroeconomic interdependencies in terms of networks can help us to foresee and limit future crises. Thinking networks can teach us to monitor the path of the damage and to set firewalls by identifying and strengthening the nodes that can stop the spread of macroeconomic fires.

We should not let ourselves believe that such cascading failures as the Asian crisis and its Latin American counterparts are the side effects of the unstable financial systems of rapidly developing nations. Established economies, such as the United States', that have the cash and the expertise to root out such failures before they turn global aren't immune to cascading failures. Vulnerabilities related to interconnectivity exist in stable economies as well, as the burst of the dot.com bubble illustrates.

## 7.

In late 1999, Compaq's Pocket PC became the company's biggest hit. As discussed by a recent *Strategy & Business* study, demand for the device outpaced supply twenty-five times, making some Compaq executives dream that, with support and accessories, the handheld devices could soon offer a bigger market than traditional PCs. Then problems started surfacing.

Compaq, Cisco Systems, and several other companies are leaders of a new business strategy: outsourcing. Cisco, which not long ago was poised to become the first trillion-dollar company, is the driving force behind this trend. It reached a 30 to 40 percent annual revenue growth with a novel and aggressive approach to manufacturing: It didn't build anything that it sold. Rather, it established strong ties to a large number of manufacturers who built and assembled the pieces sold under Cisco's logo. Compaq and many others followed suit.

Outsourcing requires a tight integration of suppliers, making sure that all pieces arrive just in time. Therefore, when some suppliers were unable to deliver certain basic components like capacitors and flash memory, Compaq's network was paralyzed. The company was looking at 600,000 to 700,000 unfilled orders in handheld devices. The $499 Pocket PCs were selling for $700 to $800 at auctions on eBay and Amazon.com. Cisco experienced a different but equally damaging problem: When orders dried up, Cisco neglected to turn off its supply chain, resulting in a 300 percent ballooning of its raw materials inventory.

The final numbers are frightening: The aggregate market value loss between March 2000 and March 2001 of the twelve major companies that adopted outsourcing—Cisco, Dell, Compaq, Gateway, Apple, IBM, Lucent, Hewlett-Packard, Motorola, Ericcson, Nokia, and Nortel—exceeded $1.2 trillion. The painful experience of these companies and their investors is a vivid demonstration of the consequences of ignoring network effects. A *me* attitude, where the company's immediate financial balance is the only factor, limits network thinking. Not understanding how the actions of one node affect other nodes easily cripples whole segments of the network.

Experts agree that such rippling losses are not an inevitable downside of the network economy. Rather, these companies failed because they outsourced their manufacturing without fully understanding the changes required in their business models. Hierarchical thinking does not fit a network economy. In traditional organizations, rapid shifts can be made within the organization, with any resulting losses being offset by gains in other parts of the hierarchy. In a network economy each node must be profitable. Failing to understand this, the big players of the network game exposed themselves to the risks of connectedness without benefiting from its advantages. When problems arose, they failed to make the right, tough decisions, such as shutting down the supply line in Cisco's case, and got into even bigger trouble.

At both the macro- and the microeconomic level, the network economy is here to stay. Despite some high-profile losses, outsourcing will be increasingly common. Financial interdependencies, ignoring national and continental boundaries, will only be strengthened with globalization. A revolution in management is in the making. It will take a new, network-oriented view of the economy and an understanding of the consequences of interconnectedness to smooth the way.

## 8.

Sabeer Bhatia did not know how to sell a company. But having been born and raised in India, he did know how to buy onions. You have to negotiate. Now he had a very hot onion to sell. He and his partner, Jack Smith, on July 4, 1996, launched a service offering nothing but e-mail—free to anybody in the world. They named it Hotmail. By year's end they had signed up a million customers, each of whom view daily the banner ads displayed on their e-mail account, Hotmail's main source of revenue. When Microsoft came knocking a year later, nearly 10 million users had Hotmail accounts. Bhatia was only twenty-eight when, after touring all twenty-six buildings at Microsoft's Redmond, Washington, empire and shaking hands with Bill Gates, he was ushered into a room packed with twelve Microsoft negotiators. They offered him $160 million. "I'll get back to you," he said, and walked away.

Currently Hotmail has about a quarter of all e-mail accounts. It is the biggest e-mail service provider in Sweden and India, countries in which it has never advertised. Microsoft eventually paid $400 million for the company, which a year later, before the burst of the dot.com bubble, was worth $6 billion.

How did an underfunded startup sign up a quarter of all e-mail users? The answer is simple: They exploited the power of networks, using a hot new marketing technique called viral marketing. Viral marketing works on the same principle that allowed Love Bug to circle the globe in a few hours. The computer virus reached everybody by looking up the e-mail list you store in your Microsoft Outlook program, sending a copy of itself to each address. Thanks to a similar innovation, Hotmail users voluntarily offer the same service.

Tim Draper, from the Draper, Fisher and Jurvetson venture capital firm, after providing $300,000 seed money to launch Hotmail, persuaded Bhatia and Smith to add an extra line at the end of each email: "Get Your Private, Free Email at http://www.hotmail.com." Therefore, whenever Hotmail users send e-mails to their friends, they advertise and endorse the company. The news about Hotmail travels on a scale-free network, utilizing exactly the same routes that helped the spread of Love Bug. Because the critical threshold for innovation spreading vanishes on such networks, it was likely that Hotmail would succeed. It was unexpected and surprising, however, how fast and to what degree it did.

What is the source of Hotmail's phenomenal success? The answer is partially contained in the Trieste study discussed in Chapter 10. Innovations and products with a higher spreading rate have a higher chance of reaching a large fraction of the network. Hotmail enhanced its spreading rate by eliminating the adoption threshold individuals experience. First, it is free; thus you do not have to think about whether you are making a wise investment. Second, the Hotmail interface makes it very easy to sign up. In two minutes you have an account; thus there is no time investment. Third, once you sign up, every time you send an e-mail, you offer free advertisement for Hotmail. Combine these three features, and you get a service that has a

very high infection rate, a built-in mechanism to spread. Traditional marketing theories will tell you that the combination of free service, low learning path, and rapid reach through consumer marketing has put the product above the threshold, and that is why it reached everybody. Based on our new understanding of diffusion in complex networks, we now know that this is only partially correct. It is true that you have a very high rate of spread. But you have no threshold either. Products and ideas spread by being adapted by hubs, the highly connected nodes of the consumer network.

Can Hotmail be replicated? Don't bet on it. Take for example EpidemicMarketing.com, a company that spent $2.1 million on a thirty-second Super Bowl advertisement in 2000, dreaming big to exploit the power of networks. In the Super Bowl ad a man visits a public restroom and receives a tip from the washroom attendant, instead of tipping the attendant as is customary. As was so cleverly expressed in their commercial, Epidemic planned to reward people for doing things they do every day. Their business model was to pay consumers to attach links to Internet businesses on their outgoing e-mail. Therefore, information about a company or promotion was expected to spread largely through word of mouth, replicating the phenomenal success of Hotmail. The model was missing a crucial element of viral marketing, however: Your friend had little interest in passing on the link to his or her acquaintances. It comes as no surprise, therefore, that Epidemic closed its doors and laid off its sixty-person staff in June 2000 after burning through the $7.6 million it raised.

Hotmail demonstrates the power of consumer networks. Some products do not need expensive telemarketing or TV and newspaper ads to prevail. They simply spread by word of mouth like a virus. Though it may not work for all products, throwing in elements of viral marketing could enhance just about all sales. Yet Epidemic's failure indicates that Hotmail cannot be easily copied. Instead, Hotmail's experience should be the starting point for new marketing approaches, combining traditional strategies with a better understanding of network effects.

## 9.

Network effects proliferate in the business world. We saw Vernon Jordan successfully surf the complex corporate network, becoming an influential member of the corporate elite. We saw Hotmail take advantage of the scale-free nature of the consumer network to become the biggest e-mail provider worldwide. The list does not stop here. Motivated by the evolving marketplace, an array of new companies have lately vowed to put network thinking at the core of their business models. Their record is mixed at best.

Take for example SixDegrees.com, a New York–based startup that asked its members to submit the names of their friends, inviting them to join too. If they enrolled, they also submitted the names of *their* friends. Step by step SixDegrees acquired a detailed map of the social network around each of its members, allowing them to reach everybody two links away from them. This consumer-driven viral marketing allowed SixDegrees to sign up over 3 million consumers. Yet the startup closed its doors on December 30, 2000, failing to turn six degrees into a viable business plan.

The burst of the dot.com bubble is often attributed to the one-dimensional thinking of many Internet enthusiasts. Most startups were based on the simple philosophy that offering things online was sufficient to replicate the success stories of the new economy. Yet, apart from a few early starts, such as Amazon.com, AOL, or eBay, most failed. The real legacy of the Internet is not the birth of thousands of new online companies but the transformation of existing businesses. We can see its signature on everything from mom-and-pop stores to large multinational agglomerates.

Networks do not offer a miracle drug, a strategy that makes you invincible in any business environment. The truly important role networks play is in helping existing organizations adapt to rapidly changing market conditions. The very concept of network implies a multidimensional approach.

The diversity of networks in business and the economy is mind-boggling. There are policy networks, ownership networks, collaboration networks, organizational networks, network marketing—you name it. It would be impossible to integrate these diverse interactions into a single all-encompassing web. Yet no matter what organizational level we look at, the same robust and universal laws that govern nature's webs seem to greet us. The challenge is for economic and network research alike to put these laws into practice.

THE LAST LINK

# *Web Without a Spider*

BY MARCH 1998, when in an unusual move I invited Réka Albert to lunch, she was only a year and a half into her graduate studies but had enough publications to receive a Ph.D. One of her papers, on granular media and sand castles, was featured on the cover of *Nature* and *Science News*, and the preliminary results of her current projects were promising, as well. So the purpose of this lunch defied all wisdom: I wanted to persuade her to give up the research she had been so good at and start something entirely different. I told her about my dream to study networks.

Four years earlier, in the fall of 1994, with a fresh doctorate in theoretical physics, I had started as a postdoc in the legendary corporate ivory tower of IBM, the T. J. Watson Research Center in Yorktown Heights, New York. Four months into my job there, perhaps touched by the spirit of the place, I checked out from the library a general-audience book on computer science to read over the Christmas break. As I immersed myself in algorithms, graphs, and Boolean logic, I started to sense how little was known about networks in general. All my readings told me that the millions of electric, telephone, and Internet cables cramped under the pavement in Manhattan formed a fundamentally random network. The more I thought about it, the more I was convinced that there must be some organizing principles governing the complex webs around us. Dreaming of

identifying some signature of order, I started to study network theory, beginning with the classical works of Erdős and Rényi. Before I left IBM in the fall of 1995 for a faculty position in physics at University of Notre Dame, I had submitted my first research paper about complex networks.

At Notre Dame, I tried with little success to contact search engines for data on the Web's topology. Under pressure to publish and obtain grants, I gradually replaced networks with safer and more conventional research. By the beginning of 1998, however, I was ready to return to thinking about nodes and links. Now I was asking one of my best students to drop everything she was doing and join me on that risky journey. I could offer her little encouragement at that time. I had to tell her that my only paper about networks had been rejected by four journals and never been published. I told her she was risking a sudden end to the success story she was part of so far. But I also told her that sometimes we should be ready to take risks. In my view, networks were worth the try.

In 1994, or even in early 1998, nobody could have anticipated the flood of discoveries that would completely reshape our understanding of our interconnected world in the following years. At that lunch with Albert when I made my pitch for networks, I could not tell her about small worlds. Not even in my wildest dreams could I conjure power laws or scale-free networks. I could not talk about error-and-attack tolerance either, since these were nonissues in network research at that time. In fact, every question worth studying that I could tell her about has since been proven ill-founded or simply irrelevant.

It was Hawoong Jeong's robot that forced us to think outside the box. Jeong joined my research group as a postdoctoral researcher in August 1998, five months after Albert and I took up networks as a research topic. Recently graduated from Korea's prestigious Seoul National University, his fascination with and knowledge of computers were prodigious. One day, after a late night discussion, I casually asked him if he would be able to build a robot to map out the World Wide Web. He made no promises. But a month later his robot was

busy carrying home the nodes and links. By that time we were somewhat familiar with the classical literature dealing with random graphs and networks. Thus it was immediately clear that the power laws seen by the robot represented a serious deviation from everything then known about networks. It was only after the construction of the scale-free model, however, that we fully understood how different real webs are from the random universe Erdős and Rényi depicted.

Today we know that, though real networks are not as random as Erdős and Rényi envisioned, chance and randomness do play an important role in their construction. Real networks are not static, as all graph theoretical models were until recently. Instead, growth plays a key role in shaping their topology. They are not as centralized as a star network is. Rather, there is a hierarchy of hubs that keep these networks together, a heavily connected node closely followed by several less connected ones, trailed by dozens of even smaller nodes. No central node sits in the middle of the spider web, controlling and monitoring every link and node. There is no single node whose removal could break the web. A scale-free network is a web without a spider.

In the absence of a spider, there is no meticulous design behind these networks either. Real networks are self-organized. They offer a vivid example of how the independent actions of millions of nodes and links lead to spectacular emergent behavior. Their spiderless scale-free topology is an unavoidable consequence of their evolution. Each time nature is ready to spin a new web, unable to escape its own laws, it creates a network whose fundamental structural features are those of dozens of other webs spun before. The robustness of the laws governing the emergence of complex networks is the explanation for the ubiquity of the scale-free topology, describing such diverse systems as the network behind language, the links between the proteins in the cell, sexual relationships between people, the wiring diagram of a computer chip, the metabolism of the cell, the Internet, Hollywood, the World Wide Web, the web of scientists linked by coauthorships, and the intricate collaborative web behind the economy, to name only a few.

## *1.*

One of the most fascinating aspects of the birth of a new science is the new language it creates, allowing us to casually converse about ideas and issues that we were struggling to describe before. The renaissance of network theory has done this for our interconnected world. The connectors of society, the stars of Hollywood, and the keystone species of an ecosystem are suddenly only manifestations of a single reality, their perceived importance within their environment attributable to their status as hubs within their respective networks. Network thinking is poised to invade all domains of human activity and most fields of human inquiry. It is more than another helpful perspective or tool. Networks are by their very nature the fabric of most complex systems, and nodes and links deeply infuse all strategies aimed at approaching our interlocked universe.

A dramatic example of the pervasiveness of this new language came after September 11, 2001, when networks acquired a meaning previously unfamiliar to most of us. Most of what led to the tragedy make perfect sense from a network perspective. Al Qaeda, the terrorist network held responsible for the attacks, was not created in seven days. Driven by religious beliefs and impatience with the existing social and political order, thousands were drawn to the radical organization over several years. The network expanded one node at a time, taking on all the characteristics of a web without a spider. Indeed, al Qaeda failed to turn into the hub-and-spoke network that offers a central leader control over all details. It avoided the tree structure as well, the chain of command characterizing the military and twentieth-century corporations. Rather, it evolved into a self-organized spiderless web in which a hierarchy of hubs kept the organization together.

After September 11, Valdis Krebs, a management consultant who normally uses network theory to analyze corporate communications, assembled a map of the nineteen hijackers aboard the four planes involved in the attacks and the fifteen people whom authorities claimed to have been connected to them. Krebs carefully entered all publicly disclosed contacts between these thirty-four individuals, weighting the

links based on the known closeness of the relationship. The obtained web is extremely revealing for anybody who wants to understand the inner workings of the deadly cell that carried out the attacks. It offers few surprises to those familiar with the shape of real networks. Mohamed Atta, the purported mastermind of the attacks, is the most connected node indeed. Yet, he had direct contact with only sixteen of the twenty-three nodes. He is closely trailed by Marwan Al-Shehhi, the second most connected node, with links to fourteen nodes. As we go down the list, we encounter numerous nodes poor in links, the peripheral soldiers of the deadly organization.

The map also shows that, despite his central role, taking out Atta would not have crippled the cell. The rest of the hubs would have kept the web together, possibly carrying out the attack without his help. Many suspect that the structure of the cell involved in the September 11 attack characterizes the whole terrorist organization. Because of its distributed self-organized topology, Al Qaeda is so scattered and self-sustaining that even the elimination of Osama bin Laden and his closest deputies might not eradicate the threat they created. It is a web without a true spider.

Today the world's most dangerous aggressors, ranging from al Qaeda to the Columbian drug cartels, are not military organizations with divisions but self-organized networks of terror. In the absence of familiar signs of organization and order, we often call them "irregular armies." Yet by doing so we again equate complexity with randomness. In reality, terrorist networks obey rigid laws that determine their topology, structure, and therefore their ability to function. They exploit all the natural advantages of self-organized networks, including flexibility and tolerance to internal failures. Unfamiliarity with this new order and a lack of language for formalizing our experience are perhaps our most deadly enemies.

To be sure, the battle against al Qaeda can and will be won by crippling the network, either by removing enough of its hubs to reach the critical point for fragmentation or by draining its resources, preparing the groundwork for cascading internal failures. Yet, collapsing al Qaeda

will not end the war. Other networks with similar scope and ideology will no doubt take its place. Bin Laden and his lieutenants did not invent terrorist networks. They only rode the rage of Islamic militants, exploiting the laws of self-organization along their journey. If we ever want to win the war, our only hope is to tackle the underlying social, economic, and political roots that fuel the network's growth. We must help eliminate the need and desire of the nodes to form links to terrorist organizations by offering them a chance to belong to more constructive and meaningful webs. No matter how good we become at winning each net battle, if we are unable to inhibit the desire for links, the prerequisite for the formation of these deadly self-organized webs, the net war will never end.

## *2.*

On June 23, 1995, the *New York Times* carried a large photograph of the German parliament, the century-old Reichstag, on its cover. This was five years after German reunification and almost exactly four years after the Bundestag, sitting in Bonn, voted to make Berlin the capital of the united Germany once again. Yet, politics and the collapse of communism had little to do with the renewed worldwide attention to the Reichstag. The real attraction for the 5 million visitors who flooded to Berlin during the coming two weeks was the fact that none of them could actually spot even a square inch of the building. The Reichstag's signature sober gray walls, the dark and quiet witnesses of a century of tumultuous German history, were all invisible. This ultimate symbol of power was wrapped in an aluminum-colored fabric, from its stairs to its flag post, transforming it into a monumental piece of public art. Over a million square feet of thickly woven polypropylene fabric held together by 5,000 feet of blue rope covered every square inch of the structure, offering one of the most magnificent artistic spectacles of our time.

The portfolio of the Bulgarian-born artist Christo and his partner, the French artist Jeanne-Claude, includes such monumental works as the Wrapped Pont Neuf, which covered the famous Parisian bridge with a yellow drapery, and the magnificent Surrounded Islands, for

which they placed six million square feet of pink fabric around eleven islands in Biscayne Bay, Miami, Florida. In many ways the Wrapped Reichstag was the culmination of their decades-long wrapping activity. Yet, it would be simplistic to perceive the artists simply as wrappers of buildings, bridges, and other objects. Their work has a powerful philosophy: "revelation through concealment." By hiding the details they allow us to focus entirely on the form. The wrapping sharpens our vision, making us more aware and observant, turning ordinary objects into monumental sculptures and architectural pieces.

In a sense we approached the world in this book following the spirit of Christo and Jeanne-Claude. To look at the networks behind such complex systems as the cell or the society, we concealed all the details. By seeing only nodes and links, we were privileged to observe the architecture of complexity. By distancing ourselves from the particulars, we glimpsed the universal organizing principles behind these complex systems. Concealment revealed the fundamental laws that govern the evolution of the weblike world around us and helped us understand how this tangled architecture affects everything from democracy to curing cancer.

Where do we go from here? The answer is simple. We must remove the wrapping. The goal before us is to understand complexity. To achieve that, we must move beyond structure and topology and start focusing on the dynamics that take place along the links. Networks are only the skeleton of complexity, the highways for the various processes that make our world hum. To describe society we must dress the links of the social network with actual dynamical interactions between people. To understand life we must start looking at the reaction dynamics along the links of the metabolic network. To understand the Internet, we must add traffic to its entangled links. To understand the disappearance of some species in an ecosystem, we have to acknowledge that some prey are easier to catch than others.

In the twentieth century we went as far as we could to uncover and describe the components of complex systems. Our quest to understand nature has hit a glass ceiling because we do not yet know how to fit the pieces together. The complex issues with which we are faced, in fields

from communication systems to cell biology, demand a brand new framework. Embarking on the journey ahead without a map would be hopeless. Fortunately the ongoing network revolution has already provided many of the key maps. Though there are still many "dragons" ahead, the shape of a new world has become discernible, continent by continent. Most important, we have learned the laws of web cartography, allowing us to draw new maps whenever we are faced with new systems. Now we must follow these maps to complete the journey, fitting the pieces to one another, node by node and link by link, and capturing their dynamic interplay. We have ninety-eight years to succeed at this, and make the twenty-first the century of complexity.