10. *Network Models*

Network theory is a whole branch of science, but it's relatively new in terms of the last 20 or 30 years. We haven't had a chance to take all that theory out of the universities and apply it to ask: "What kinds of networks should we build, and for what purposes?"

—Anne-Marie Slaughter

In this chapter, we cover models of networks. A comprehensive study of networks would require multiple books. We have more modest goals. We want to understand the basics of networks, to be able to name their parts, and to ask why they matter for modeling. The answer we arrive at will be that networks almost always matter. Any model we construct, be it of a market, the spread of a disease, or the transmission of information, can be enriched by embedding the actors in a network.[1]

Networks are ubiquitous. People talk of trade networks, terrorist networks, and networks of volunteers. Species organize into food webs, a form of network. Firms build supply chain networks. As already noted, the financial system is usefully thought of as a network of promises to pay. Networks have always been important to understanding social relations. During much of human history, social networks were constrained by geography and difficult to map. Due to technological advances, many social interactions and economic transactions now take place over virtual networks which can be analyzed using models.

The organization of this chapter follows the same structure-logic-function format we applied to distributions. We first characterize the structure of networks using statistical measures of degree, path length, clustering coefficient, and community structure. Then we discuss common

classes of networks: random networks, hub-and-spoke networks, geographic networks, small-world networks, and power-law networks. After that we turn to the logic of how networks form. We construct micro-level processes that produce the network structures we see. Last, we take up function, the question of why network structure matters. Here we focus on five implications. We begin with the friendship paradox, and then describe the six degrees of separation phenomenon and the strength of weak ties property. Last, we take up the robustness of networks to node or edge failure and the aggregation of information over networks. The chapter concludes with a discussion of how networks influence model outcomes.

Network Structure

A network consists of *nodes* and *edges* that connect them. We refer to nodes connected by an edge as *neighbors* and to a network as *connected* if it is possible to get from any one node to any other along edges. Networks can be represented as graphs, as lists of edges, or as matrices of zeros and ones, where a one in row *A* and column *B* denotes an edge between node *A* and node *B*. Though people prefer graphical representations of networks, lists and matrices are better for representations for calculating network statistics.

The edges in a network can be *directed*—that is, pointing from one node to another. In an information network, a directed edge denotes that one person gets information from another. In an ecosystem network, a directed edge from a red-tailed hawk to a gray squirrel represents that the hawk eats the squirrel. Edges can also be *undirected*. Edges that connect friends are drawn in this way. In an undirected network, the *degree of a node* equals the number of edges that connect to it. Networks are characterized by a set of network statistics. For each statistic, we can compute the network average and the distribution across all nodes. The *average degree* of a friendship network tells us, on average, how many friends each person has. The *degree distribution* tells us if some nodes are more connected than others. Social networks have more equal distributions than the World Wide Web, the internet, and citation networks, all of which have long tails.

Network Statistics

Degree: The number of neighbors (also the number of edges) of a node.

Path length: The minimum number of edges that must be traversed to get from one node to another.

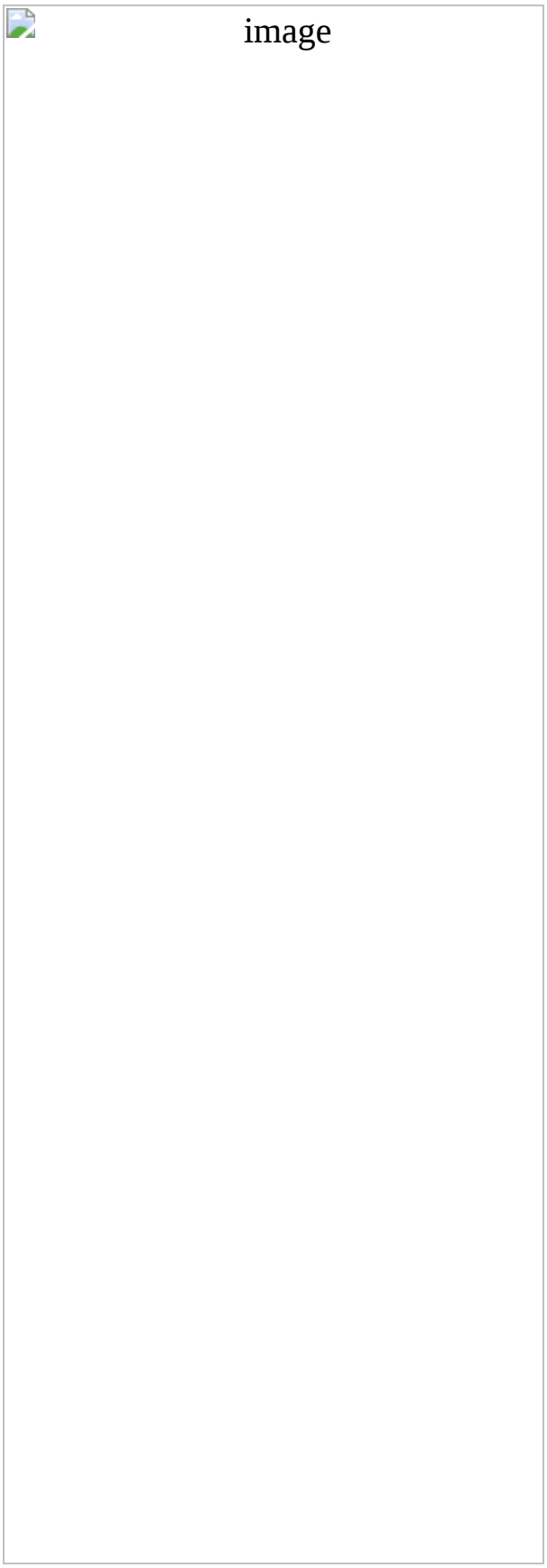
Betweenness: The number of paths of minimal length connecting two other nodes that pass through a node.

Clustering coefficient: The percentage of a node's pairs of neighbors that are also connected by a edge.

Path length, the minimal distance between two nodes, varies inversely with degree. As we add edges, we shorten the average length between nodes. In an airline flight network, path length corresponds to the number of flights, on average, a person needs to take to get from one city in the network to another. Given a choice between two airlines, all else equal (namely, prices), a traveler would prefer the one with lower average path length. Average path length also correlates with information loss. Information that passes through several people is more likely to suffer distortion than information passed between only two people. The nodes on minimal paths play critical roles in networks. If information takes the shortest route, then it goes through the nodes on a minimal path. A node's *betweenness* score equals the percentage of minimal paths that go through a node. In a social network, people with high betweenness scores know more information and wield more power.

The final statistic, the *clustering coefficient*, equals the proportion of a node's pairs of neighbors who are also neighbors of one another. For example, a person with 10 friends has 45 pairs of friends. If 15 of those 45 pairs are themselves friends, then the person's clustering coefficient equals  . If all 45 friendships existed, then the person's clustering coefficient

would equal 1, the maximal possible value. The clustering coefficient for the entire network equals the average of the clustering coefficients of the individual nodes.



image

Figure 10.1: A Hub-and-Spoke Network and a Geographic Network

Figure 10.1 shows two networks with thirteen nodes: a hub-and-spoke network and a geographic network. In the *hub-and-spoke network*, the hub has degree 12 and all other nodes have degree 1, for an average degree of less than 2. The degree distribution is unequal. The hub has a distance of 1 to every node. All other nodes have a distance of 1 to the hub and a distance of 2 to the other nodes. It follows that average path length is also less than 2. The hub, which lies on every minimal path between any two other nodes, has a betweenness score of 1. The spoke nodes do not lie on any minimal paths connecting other nodes, so they have a betweenness of 0. Finally, in the hub-and-spoke network, no nodes connected to a node are connected to one another. Therefore, the network has a clustering coefficient of zero.

In the *geographic network*, each node is connected to the two nodes to its right and its left, so the average degree equals 4. Each node is distance 1 from four nodes, distance 2 from four nodes, and distance 3 from four nodes. So the average distance equals exactly 2. The degree and distance distributions for this graph are *degenerate*—every node has the same degree and the same average distance. It can be shown that the betweenness of each node equals  . Each node has four neighbors, creating six pairs. Of those six pairs, exactly three are connected: the two nodes to the immediate left and right are each connected to an outer node and to each other. Therefore, the clustering coefficient equals  .

An alternative method for capturing clustering is to partition the nodes into *communities*. In a junior high friendship network, the communities might correspond to teenagers interested in the arts, athletics, or science. Or they could be defined by race and gender. A network of political alliances might partition into regional or ideological allies. Multiple methods exist for determining communities. One approach sequentially removes edges with the highest betweenness, as edges with high betweenness are more likely to connect distinct clusters. Other approaches take the number of communities as given and seek an optimal partitioning given an objective function such as minimizing the number of edges between the communities or maximizing the proportion of edges within communities.[3]

We can use community detection algorithms to ask questions of network

data. Studies show that people may reside in *online bubbles*. That is, we may belong to communities of people who get their news from similar sources. If so, that has implications for social cohesion. Prior to the creation of the internet, that may have been true as well, but demonstrating it with data would have been hard. Now data scientists can scrape the web to identify the news sources that people frequent and tell us that, yes, in fact we do live in bubbles to an extent. Models provide the formal definitions of communities. Data tells us the strength of those communities. Using judgment we can make wise inferences based on what the data say.

Common Network Structures

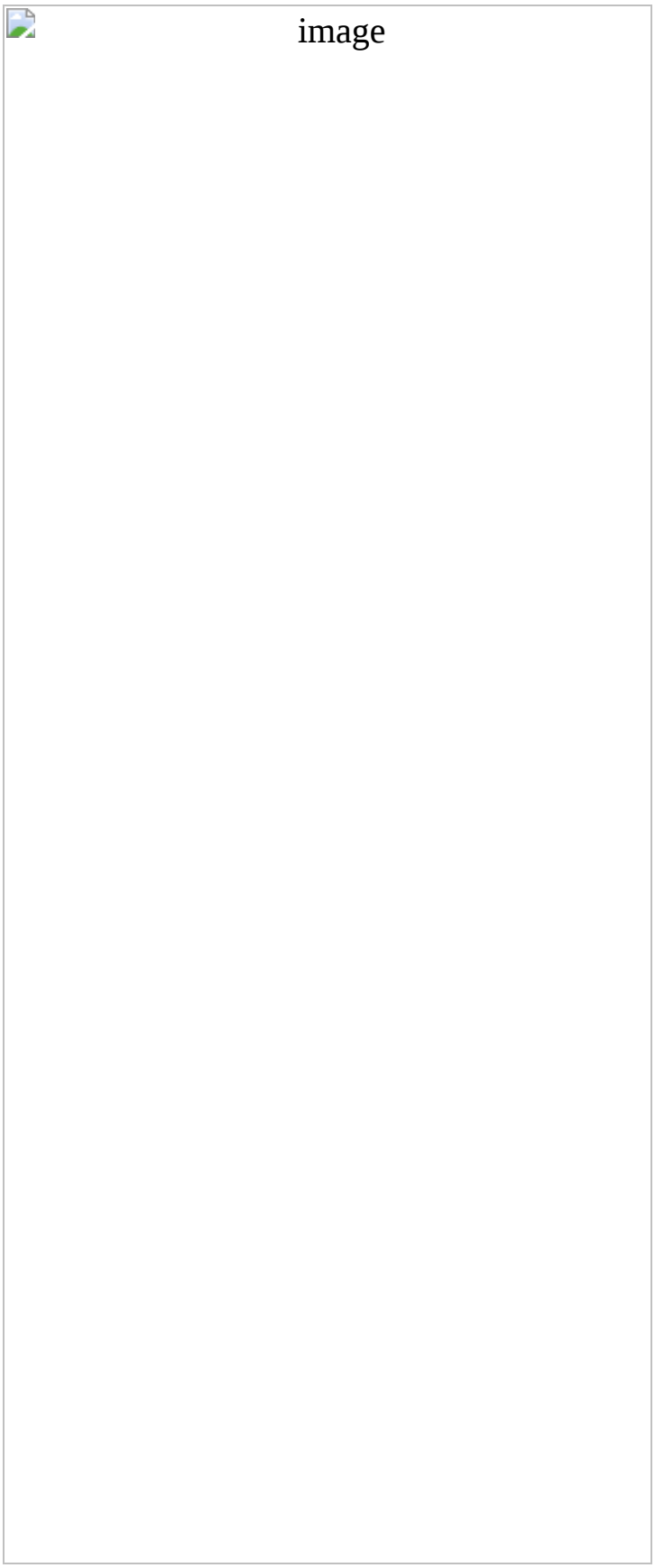
In analyzing networks, we encounter a problem of variety. A handful of network statistics are incapable of pinning down the specific network structure: one can construct billions of distinct networks with ten nodes and an average degree of 2. An alternative approach to characterizing a network is to test whether its statistical measures differ significantly from those of a common network structure. For example, a scholar might gather data on judicial citations and put it in network form by drawing an edge when one judge cites another judge's opinions. The graph of that network may appear to possess interesting structures and clusters. We can test whether a network is random by comparing the network's statistics to those of a random network that has the same number of nodes and edges. A *random network* clustering coefficient equals the probability of a random edge because two neighbors of a node are no more likely to contain an edge than any other randomly chosen node.

Monte Carlo Method for Random Networks

To test whether a network with *N* nodes and *E* edges is random, we create a large number of random networks with *N* nodes and *E* edges and calculate distributions for degree, path length, clustering coefficient, and betweenness. We then perform standard statistical tests to accept or reject the hypothesis that the network's statistics could have been drawn from the simulated distributions.[4]

Theoretical models often assume a particular network structure. Many assume random networks, while others assume regular geographic networks such as when the nodes are arranged in a circle and each is connected to the nearest nodes in each direction. Other geographic networks arrange the nodes on a checkerboard and connect each node to its neighbors to the north, south, east, and west. Most of the common geographic networks have low degree—they connect to only the local neighbors—and relatively high average path length. On geographic networks, betweenness and clustering coefficient have no variation.

A third common type of network, a *power-law network*, has a power-law degree distribution. A handful of nodes has many connections, but most nodes have very few networks. A fourth type of network, a *small-world network*, combines features of geographic and random networks.[5] To construct a small-world network, we begin with a geographic network and then "rewire" it by randomly selecting an edge and replacing one of the nodes it connects with a random node. If the rewiring probability equals zero, we have a geographic network. If it equals 1, we have a random network. In between, we have a small-world network, distinguished by small clusters from the geographic network connected by random links to other clusters. Social networks look similar to small worlds. Each person has a cluster of friends as well as random friends.



image

Figure 10.2: Random, Geographic, Power-Law, and Small-World Networks

Network Formation: Logic

We now briefly describe models of network formation. These models provide logic to explain network structures. Most of the network structures that we encounter *emerge* from choices of individual actors to make connections. That is true of friendship networks, the World Wide Web, and power grids. These networks are not planned. Other networks, such as supply chain networks, do result from planning. We would expect planned networks to be robust to the failure of nodes. The fact that emergent network structures are robust is more of a puzzle.

We have already discussed how to create random networks and small-world networks. We create the former by randomly creating a set of nodes and then drawing edges connecting random pairs of nodes. We create a small-world network by first constructing a regular geographic network (often by arranging nodes in a circle and connecting k neighbors in each direction) and then randomly "rewiring" a proportion of the edges.

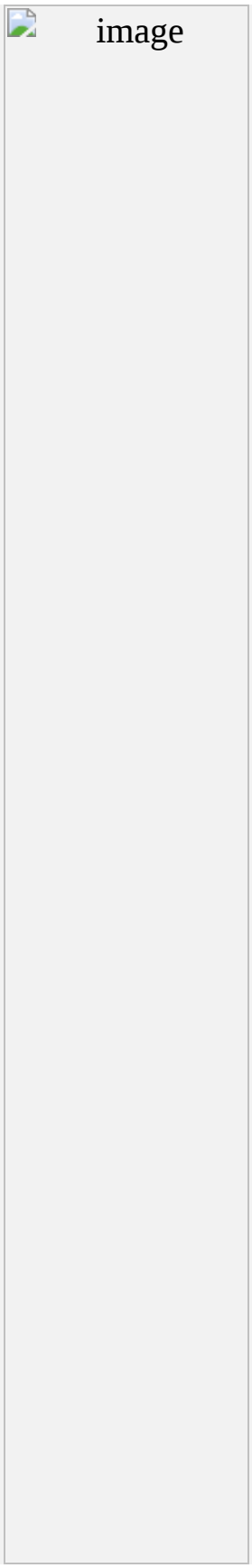
Models of the formation of the power grid rely on economic and engineering principles. The network must deliver power to homes, businesses, and the government. Whether the producers are for-profit companies or public utilities, they have little incentive to create high clustering, as it would be inefficient. This lack of clusters reduces the robustness of the network. Economic and engineering considerations also rule out long leaps: connections that reach far across the network. Power companies do not build direct connections from Chicago to Dallas. People and businesses, however, do. A Chicagoan might strike up a friendship with someone from Dallas. A firm in Singapore might trade with a firm in Detroit. As we see in the next section, these long leaps contribute to network robustness.

To create a network with a long-tailed distribution, we can apply a version of the preferential attachment model. We create nodes randomly and then draw edges from new nodes to existing nodes. If we let the probability of connecting to a node be proportional to its degree, we produce a power-law degree distribution. In that model, early-arriving nodes will be far more likely to be of high degree. A shortcoming of the

model is that it does not allow for any difference in node quality. Higher quality nodes should have higher degree. The *quality and degree network formation model* corrects that omission while also producing a long-tailed distribution.

Quality and Degree Network Formation Model

Create d disconnected nodes. In each period t create a new node with quality Q_t drawn from a distribution F. Connect that node to d other nodes based on the degree of those nodes. If D_{it} denotes the degree of node i at time t, the probability of choosing node i given N nodes equals:



If the quality of new nodes has a low mean and low variance, the model resembles the standard preferential attachment model. If the quality distribution has a long tail, then new nodes of very high quality can grow to have large degree.[6]

Why Networks Matter: Function

In Chapter 1, we mentioned the friendship paradox, the fact that on any network, on average, people cannot have more friends than their friends do. The logic for why this holds can be shown using the hub-and-spoke network. In that network, twelve people have one friend and one person has twelve friends. The twelve people with one friend are all connected to the hub, and the hub has twelve friends. That feature—the fact that high-degree people are connected to more people—drives the result. On the hub network, people, on average, have fewer than two friends. Yet, on average, each person's friends have more than eleven friends.

The friendship paradox holds for any network: academic citation networks, email networks, sexual contact networks, banking networks, and international trade networks. On average, the references cited by an academic article receive more citations than the article itself; a country's trading partners, on average, trade with more countries than the country itself; and the multiple species connected to a single species in a food network have, on average more connections than the single species itself. The disparity between the number of friends and the number of friends of friends becomes more pronounced on networks with degree distributions that are more dispersed. One analysis of friendships on Facebook found that the average person has around two hundred friends and their friends, on average, have more than six hundred friends.[7]

The Friendship Paradox

If any two nodes in a network differ in their degree, on average a node has lower degree than its neighbors. In other words, on average, people's friends are more popular than they are.[8]

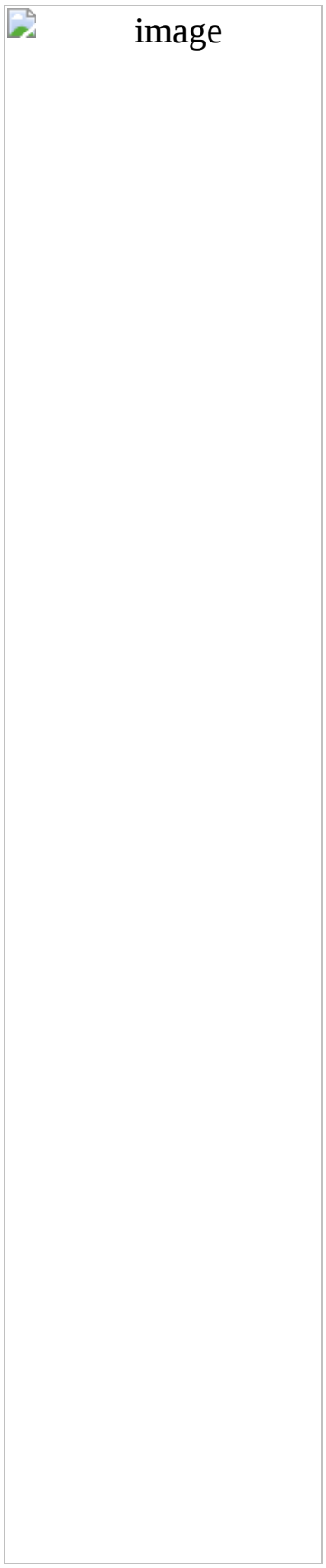
The logic of the friendship paradox extends to any attribute that correlates with the number of friends. If active, happy, intelligent, wealthy, and kind people have, on average, more friends, then a person's friends will be, on average, more active, happier, more intelligent, wealthier, and more beautiful.[9] Imagine a network in which 90% of unhappy people have four friends and 10% have ten friends. Reverse the proportions for happy people: 10% have four friends and 90% have ten friends. People's friends will disproportionately consist of people with ten friends. A large majority of those people will be happy, so most people's friends will be happier than they are.

We now show the *Six Degrees of Separation* phenomenon, the claim that any two people on the earth can be connected through six friends or fewer. While the friendship paradox holds for any network, six degrees of separation only holds for some types of networks. The phenomenon's name derives from an experiment carried out by Stanley Milgram in the 1960s. Milgram sent packets to 296 individuals in Omaha, Nebraska, and Wichita, Kansas, that were to be forwarded to an individual in Boston, Massachusetts. The recipients had to follow the same rules. These participants were only allowed to send the packets via mail to people that they knew personally and whom they believed might have a greater chance of knowing the target person in Boston, with instructions to do the same. Individuals signed a roster to record the path and mailed postcards to the researchers so the researchers could track breaks in the chain. Sixty-four of the letters arrived in Boston. Of those that did, the average path length was slightly less than six, hence the phrase "six degrees of separation."

A second experiment run fifty years later on a much grander scale using

email created eighteen global targets and sent them to more than 20,000 people. The median path length of email chains was between five and seven, depending on the geographic distance between the source and the target. The length of paths found does not equal the minimal path length between participants. The evidence therefore suggests that most people are linked by fewer than six degrees.[10]

We construct a simplified version of the small-world network to give intuition about the six degrees of separation phenomenon. Our version assumes that individuals have a small cluster of *clique friends*, who all know one another, and that they also have friends outside of those cliques, whom we call *random friends*.[11] Figure 10.3 shows an individual (denoted by the black circle) with five clique friends and two random friends. It also shows a selection of friends of the node's friends (light gray circles).



image

Figure 10.3: A Node's Clique Friends (*C*) and Random Friends (*R*)

These random friends might also be thought of as *weak ties*—people who connect you to other communities of people. Our weak ties, the random friends in our network, play an important informational role by connecting communities with diverse interests and information. Hence, sociologists speak of the *strength of weak ties*.[12]

This construction allows us to calculate the number of neighbors of degree two (the friends of friends), by adding up all of the friends of random friends but only adding the random friends of the clique friends. We do not count the clique friends of the clique friends, as they are members of the node's clique. We calculate the number of *friends of friends of friends* similarly. We add in all of the clique friends' random friends' friends, but we do not add in a random friend's clique friends' clique friends, as they have already been counted as neighbors of degree two. To produce the six degrees of separation phenomenon, we apply the same logic to a network with 100 clique friends and 20 random friends.

Six Degrees of Separation

Assume each node has 100 **clique friends** (*C*), all of whom are friends with one another, and 20 **random friends** (*R*), who have no friends in common with the node.

Degree one: $C + R = 120$

Degree two: $CR + RC + RR = 2000 + 2000 + 400 = 4400$

Degree three: $CRC + CRR + RCR + RRC + RRR = 328{,}000$

Degree four: 17, 360,000 [13]

Degree five: > 1 billion

Degree six: > 20 billion

By assuming no overlap in the friends of the random friends, the model implicitly assumes an infinite population. An actual social network will have overlap in friends as the degree increases. In a network that includes overlap and other realistic features such as heterogeneity in the number of friends, the values will differ from those calculated above. The relative magnitudes of the number of neighbors of each degree will remain similar. A person will have many more neighbors of degree three (friends of friends of friends) than of degree two (friends of friends).

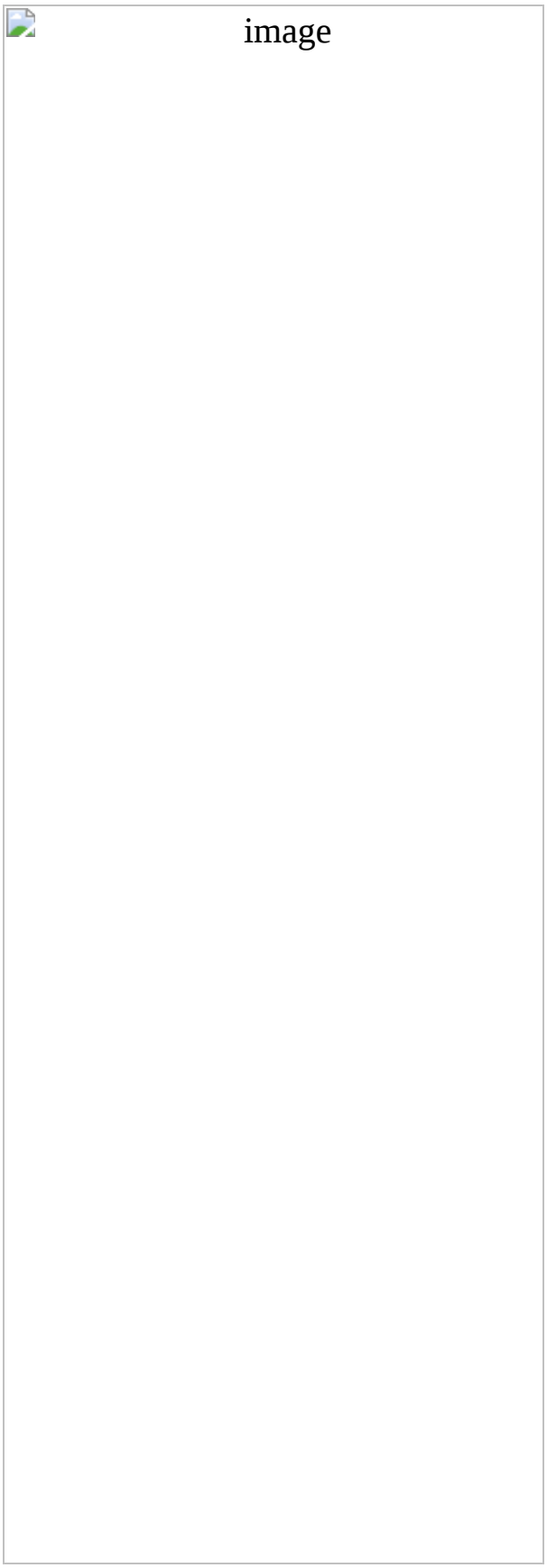
The large number of friends of degree three, over a quarter million in our example, can be consequential. Unlike a person's clique friends, a person's friends of degree three tend to live in different cities, attend different schools, and have different information. They are more diverse. They are also near enough for trust to be established: a friend of a friend of a friend could be your roommate's mother's coworker, or your sister's boyfriend's aunt. The number of friends of degree three, their diversity, and

their relative proximity make them an important asset. They can provide new information and job opportunities. These are the people most likely to help a person find a job, facilitate a move to a new city, or become a life or business partner.

Network Robustness

Our last implication of network structure evaluates the *robustness* of network properties, or how close the network is to node (or edge) failure. The most essential property of a network is whether it remains connected. We can use models to calculate the probability that a network remains connected as a function of the number of nodes removed. We could also ask what happens to average path length as nodes are removed. Applied to an airline network, an analysis of path length robustness would tell us how many extra flights would be needed if an airport were to shut down due to weather or a power failure.

Here, we consider the question of how the size of the largest connected component of the network, the *giant component*, changes as nodes randomly fail. Figure 10.4 shows the size of the giant component for a large random network and a large small-world network. In the random network, the size of the giant component falls linearly at first. At a critical value where the probability of an edge equals 1 divided by the number of nodes, the size of the largest component falls to an arbitrarily small proportion of the original network size. The small-world network shows no such abrupt change. A majority of connections exist within the geographic clusters. Each cluster can withstand the failure of multiple nodes. This feature combined with the random links prevents the entire network from collapsing.



image

Figure 10.4: Size of the Giant Component (G) as a Function of Node Failure

From figure 10.4 we can infer that sparse networks that lack local clustering are susceptible to failure. We can apply that insight to the power grid. It lacks the long leaps and the tight clusters that make the small-world network robust. In a power grid, the failure of a node or a link cannot be overcome by other links in a cluster or by a long connection to a working node far away. Local failures can cascade through the network.[14] In contrast, the internet, which has a long-tailed degree distribution, is robust to random node failure. The degree distribution implies that the vast majority of nodes have few connections. Even if they fail, the network remains connected.

Up to now, we have assumed random node failure. We can also consider strategic node removal. Networks with long tails, like the internet, now become non-robust. Strategic removal of the nodes of highest degree destroys the network. The logic can be seen by considering the hub-and-spoke network. When nodes are removed randomly, the network remains connected unless the hub node is removed, a low-probability event. Strategic removal, wiping out the hub, disconnects the network in one step.

With some networks, such as terrorist networks and drug supply networks, we might want to disconnect a network. If those networks are sparse, like the power grid, or have a long-tailed degree distribution, they can be disconnected through strategic node removal. For the terrorist network, this would entail arresting the most connected members. If those networks resemble small worlds, they will be robust, even to the strategic removal of nodes. Attempts to cut off any geographic segment of the network will fail because of the random reconnections that connect the segment to the rest of the network.

Summary

When we build network models of people, we often do so to capture social influences, where the success, behavior, information, or beliefs of a person in the network influence the success, behavior, information, or beliefs of their friends. Behavior can be contextual or intrinsic; so too can a person's value or contribution to a collective enterprise. A person's value or contribution could be due to properties of that person, such as her brilliance, her effort level, or her good fortune. A person's success could also be due to his network of friends and colleagues. This is an age-old question: Does success depend on what you know or whom you know?

Imagine a group of scientists working together in a research lab. They share advice, ideas, and knowledge. The number of academic papers, patents, or scientific breakthroughs produced by a scientist depends on what she knows, but it can also be influenced by whom she knows, on her interactions with other scientists. By thinking in terms of contextual features (friendship networks) as well as intrinsic attributes (individual abilities), we can determine how much of a scientist's success to attribute to each.

Investment firms that hire away superstar fund managers based on the belief that investment success depends mostly on talent have not had very promising results. Empirical evidence shows that top investors also depend on networks of colleagues who provide them with specific types of information.[15] That specific finding can be viewed within the lens of a much larger literature (some of which is model-based) showing how a person's position in an organization influences success.

Success still correlates with ability. A business idea that makes its investors millions was probably a good one. A scientist who publishes hundreds of papers and receives numerous awards has high ability. At the same time, those best positioned in the network make the largest contributions. We can measure a person's position using betweenness and other measures of centrality. The people who occupy high-betweenness positions in a network fill what Ron Burt calls *structural holes* between

communities, which we can identify using algorithms.[16] Access to information and ideas from multiple communities gives people who fill structural holes power and influence. Filling a structural hole requires certain talents and abilities. A person cannot just jump in and fill any hole. She must build trust and understanding within each community. And she must be conversant in the knowledge base of each community.

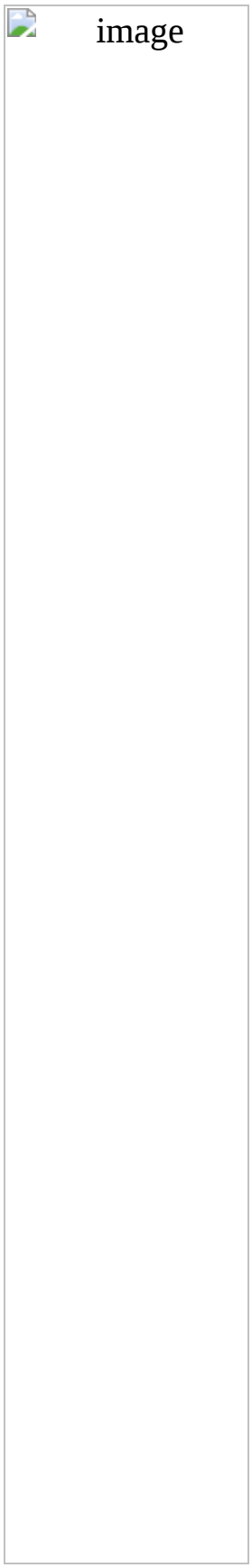
We can apply nearly identical logic to assess the value of firms and assign power to countries. We can see a firm's value as intrinsic and take a balance sheet perspective by looking at assets and liabilities. We can also look at the context in which the firm operates, such as its position in the supply chain. Similarly, the power of a country depends on its resources and its alliances. For both firms and countries, intrinsic attributes and connectedness correlate. Those who occupy powerful positions in the network also possess important attributes.

Our analysis as well as most of the literature considers the node as the unit of analysis. The edges can be critical as well. Taking an even broader perspective, the network itself may be an appropriate unit of analysis. For example, teacher networks that allow ideas and information to flow between classrooms can improve educational outcomes, and so a well-connected administrator can effectively coordinate curriculum reform. Similarly, a second-grade teacher knows a lot of information about the students from his class who are going on to third grade; that information could help the third-grade teacher. A mathematics teacher knows what concepts students have yet to grasp; that information could help the science teacher structure her lessons. Good schools, therefore, have strong faculty networks. This is just one example of how network models can improve our thinking.[17]

Myerson Value and Burt's Structural Holes

People who fill structural holes connect communities in a network and have more influence. A variety of statistical measures of a network, such as betweenness, should correlate with occupying a structural hole. An alternative measure of influence in a network, *Myerson value*, relies on the logic of Shapley values. To compute Myerson values, we construct a cooperative game on a network but only allow coalitions that include connected components.

Consider three individuals arranged in a line. Assume that their locations represent political ideologies with person *B* in the center, as shown below. If we restrict coalitions to immediate neighbors, then *A*, the left-most person, cannot connect to *C*, the right-most person, unless person *B* also belongs to the coalition. To compute each player's Myerson value, we first assign added values to all feasible coalitions. We then compute Shapley values for each possible coalition, treating each as a distinct game. Last, we add up the Shapley values for each coalition game to obtain Myerson values.



image

As an example, suppose that any coalition of two players produces an output of value 10 and all three players together produce output of value 14, we obtain the following: Players 1 and 3 have Myerson values of 3, and Player 2 has a Myerson value of 8.[18]

Centrality measures such as betweenness are based only on the network. Myerson values depend on a value function. By having both measures we can disentangle the dependence of power on a person's position in the network and on the functions she performs. In our example, the Myerson values for the three players, (3, 8, 3), correlate perfectly with their betweenness scores (0, 1, 0). That will not always be the case, particularly for more complicated networks and value functions.