CHAPTER 10

Centrality

In an academic department the regular faculty may be away frequently doing their own research or working at home. Because they may be in the department only infrequently to teach and because their research interests may be quite different, they may not see much of each other. On the other hand, the staff of a department, especially the staff in the front office, is there all day five days a week, and they may be the only ones to talk to all members of the department frequently. As a result, they may be the best informed people in the department, better aware than most members of the department about successes, significant personal events, illness, and tragedies.

This was brought home to the senior author (Bonacich) a few years ago when his father died. He had to spend some time in San Francisco settling his father's affairs. Some of his responsibilities had to be transferred to others, and this required coordination with the staff. As a result, the staff all knew about the father's death before most of the faculty. Soon afterward he received a condolence card signed by all the staff, but sympathy from other faculty members trickled in as they found out.

Information flows through a social network, and some individuals will be better positioned within this network to learn more information and to learn it earlier. It should also be clear that in such a network it's not just how many people you talk to but who you talk to; regularly having lunch with someone who is himself well informed will provide you with better and more current office gossip. Being well placed in such a network need not correspond to formal status or rank in the organization. There are firms that specialize in diagnosing communication problems in organizations. An important executive who should be well informed in order to make good decisions may have been shut out of the informal communication network of an organization. There may be no communication between members of two divisions of a firm that should be coordinating their activities.

Information is not the only thing that flows through networks. Electricity flows through power grids, and those who plan such grids should have a good idea of where the system is in danger of becoming overloaded and where it is especially vulnerable to catastrophic failure if a piece of equipment fails. Contagious epidemics also flow through social networks. If a vaccination exists, it may be especially important to vaccinate those whose removal from the network would especially disrupt the spread of the disease, not just those who are most vulnerable. Health workers come in contact with many sick individuals and thus are particularly likely to catch the disease. If children are especially vulnerable, it may be important to inoculate teachers and parents of small children even if they themselves are not especially at risk.

Social scientists and non–social scientists have recently become interested in *social capital*, which refers to the resources available to someone through others in their social networks. Social capital means different things in different circumstances. To a poor person social capital may refer to the availability among her friends of a car she can borrow, free babysitting, money she can borrow in an emergency, and emotional support in times of stress. To an entrepreneur his social capital may refer to the skills available among people he plans to work with.

Rankings of all sorts that we are exposed to daily can be better understood if they are described and understood as networks. For example, consider network search engines. The early search engines were not deficient in the number of webpages they covered. They all had webcrawlers that explored the web by following links from the websites they had discovered. The primary deficiency was in presenting the results. The output from early search engines was page after page of mostly irrelevant websites sprinkled occasionally with a useful site. Who today remembers AltaVista, Magellan, HotBot, or Excite? (You do? Obviously you aren't a young college student.)

Google became the dominant search engine not because it had a better web crawler but because of an algorithm, PageRank, that ordered results in a useful way. The primary innovation behind PageRank is to consider webpages as nodes in a network with links between pages as arcs in a digraph. PageRank locates highly central nodes in this network. These highly central webpages will appear at the top of the list in a Google search. The central idea is to weight links by the importance of the sites that link to them.

Computerized sports ranking systems also make use a network approach. The vertices are sports teams and the arcs represent who beat whom. Victories and losses are weighted according to the rankings of a team's opponent, and the results of one game can have implications that spread throughout the network indirectly affecting the rankings of many other teams.

The centrality measures we examine in this chapter will fall into two broad categories: those that measure how important a node is in the flow through the network of information of some other resource and those that measure the status or prominence of a position by its integration with other prominent positions. The first set of measures are usually based

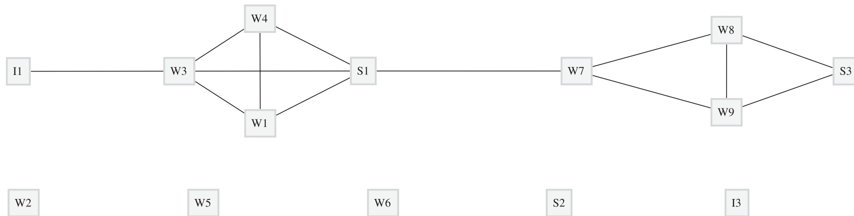

Figure 10.1. Friends in the bank wiring room.

on geodesics, and the latter on connections with other well-connected nodes. A lowly assistant professor with a joint appointment can be the only connection between two departments without being in the ruling clique in either department.

In this chapter we will be using two real networks as examples. The first one involves friendship patterns among a set of workers in a room in the factory, the "bank wiring room" (Figure 10.1). These workers produced telephone equipment. There were wiremen who wired connections, soldermen who soldered the connections, and inspectors (see Homans 1958; Roethlisberger et al. 1939). Lots of things flow through friendship networks: information, reciprocal gifts and favors, and emotional support.

The second network (Figure 10.2), also well studied by social scientists (Padgett and Ansell, 1993), is of marriage ties between important families in the Italian Renaissance city of Florence in the 15th century. Through these ties might flow political, financial, and social support. Social scientists have attempted to account for the power and influence of the Medici family in this period through their positions in this and other networks.

Centrality refers to characteristics of a node's position in a network in which some type of resource flows from node to node. There are many ways in which this rather vague idea could be implemented. We will describe a few of them in this chapter. In what follows we will assume that the network is represented by a symmetric binary (0 and 1) adjacency matrix A. This means that the relation is symmetric and that it either exists or does not exist between any two nodes. There are more complex ways of handling more realistically complex situations (for more methods, see Nooy et al. 2005; Wasserman and Faust 1994), but these assumptions make the presentation manageably simple.

All the measures we are looking at here (with the possible exception of degree centrality) make sense only for sets of vertices that are in the same component, so that their positions within the component can be meaningfully compared. The measures should be computed separately for each component.

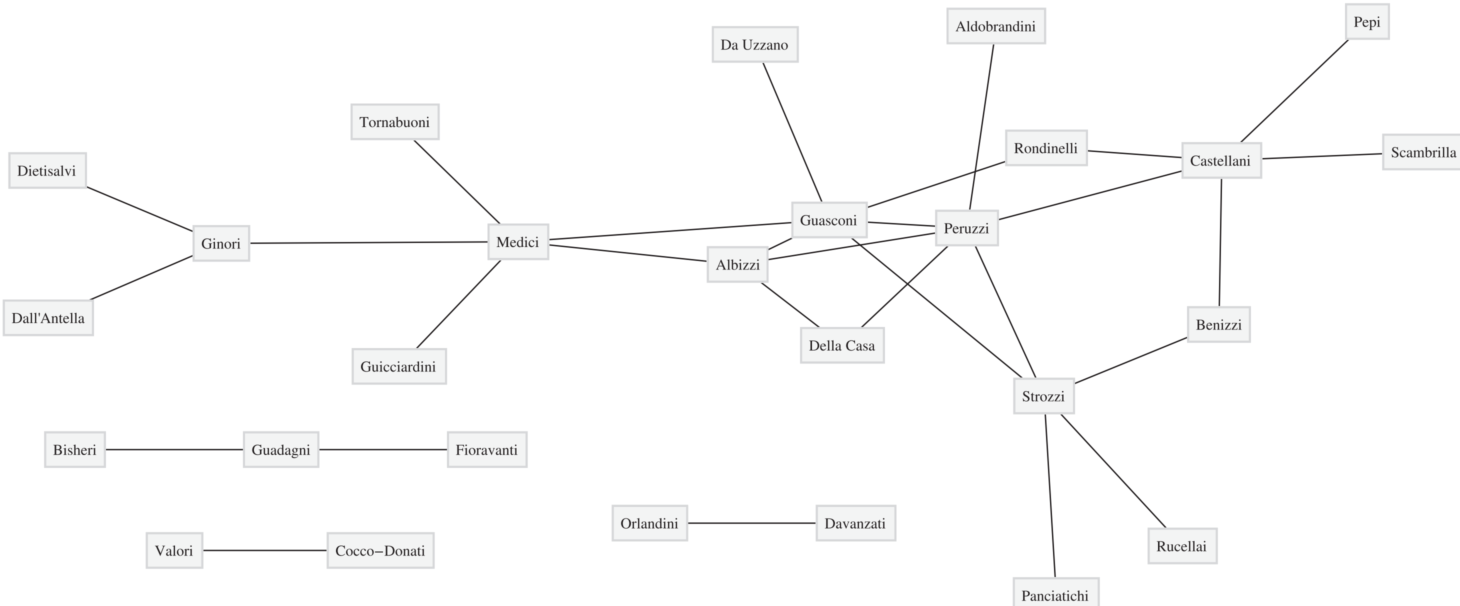

Figure 10.2. Marriage network in Renaissance Florence.

DEGREE CENTRALITY

One of the most basic forms of centrality is degree centrality. Degree centrality, D, is simply a node's total number of connections.

$$D_i = \sum_k a_{ik} \tag{10.1}$$

Everything else being equal (which it probably isn't), the amount of information an individual is exposed to or the resources he has access to is proportional to the number of his contacts. Among the Florentine families, for example, the Guasconi and Peruzzi families were tied to the most other families (six) by marriage, and this gave them an advantage in terms of information and alliances. Degree centrality is often used as a rough measure of popularity or importance. In a network of friends, people with a high degree have the most friends. In a food-chain network where the edges represent predator and prey relationship, species with high degree centrality are important to the overall network.

Degree centrality is popular because it's easy to measure. Large representative random surveys that may not be designed to facilitate network data can include a question or two, such as "How many friends do you have," to approximate degree centrality. However, this aspect of the measure is also its biggest weakness. Degree centrality really isn't a "true" network centrality measure because it does not take into account the structure of the entire network.

GRAPH CENTER

Suppose you ran a large retail store that sold many different items. Your store just received a new barcode scanner that would make life easier for customers who wanted to check the prices on items. Here's the catch—your store received only one of these blasted scanners. Now you are tasked with finding out the best location for the scanner. Well, if you imagine the aisles in your store as a network, you would want to place the scanner at a location that is central to all products. If your store's building is perfectly square, and your aisles follow a neat grid pattern, the most logical place to meet this criterion would be right in the center of the store. In networks, closeness centrality measures this aspect of power. Individuals with high closeness centrality can be thought of having high influence. If they had to pass along information, orders, or resources through their network, they would be able to reach a large number of people fairly quickly.

A crude centrality measure that tries to operationalize the concept above is the graph center.

If everyone in a network has some information to contribute and transmits her information to her neighbors, then the person in the network who possesses the entire group's information first will be the person whose maximum distance from others in the network is least. For example, let's look at the largest component in Figure 10.1, containing nine workers. Just to keep it simple, suppose that each person transmits all he knows to his neighbors in one time period. It will take five time periods for Inspector 1 and Solderman 3 to hear the information known by those on the opposite side of the network. At the other extreme, Solderman 1 and Wireman 7 are no more than three links away from everyone else in the component. They are the centers of this graph. If d_{ij} is the graph distance between nodes i and j, then the center (or centers) is the node k whose maximum distance from other nodes is the minimum. You should be able to show that the Guasconi and Albizzi families are the centers of the network of Florentine families.

$$\text{Vertex k is a graph center if } \max_j\{d_{kj}\} = \min_i \max_j\{\mathrm{d}_{ij}\} \tag{10.2}$$

Another advantage of graph centers is that any information they have will become completely distributed to all members of the group the quickest. Solderman 1 and Wireman 7 in Figure 10.1 are in the nest positions to coordinate the activities of a group. If they suggest some group activity, all members of the largest component hear it first. If Solderman 3 were to say that he planned to go to Bar A after work for a drink and Solderman 1 that he planned to go to Bar B and if Figure 10.1 describes communication links, everyone will hear the second suggestion before the first.

CLOSENESS CENTRALITY

Closeness centrality C can be thought of as a refinement the graph center. For each node, it is the reciprocal of the total distance of all other nodes from that node. In other words, given a vertex, its farness centrality is the distance between that vertex and all other vertices—all $(n-1)$ of them. If a vertex is in a remote location, the sum of all these numbers is going to be very big. If the vertex is in a location close to all else, this number will be small. In order to turn farness to closeness, network researchers take the multiplicative inverse (aka divide by 1) of farness in order produce closeness.

$$C_j = \frac{1}{\sum_k d_{jk}} \tag{10.3}$$

Table 10.1 shows the distances between all pairs of vertices in the largest component of the bank wiring room.

The total distances are given by the vector (23, 17, 16, 17, 14, 18, 18, 13, 24). The closeness centralities are, therefore, (1/23, 1/17, 1/16, 1/17,

Table 10.1.
Distances between vertices in the bank wiring room

	I1	**W1**	**W3**	**W4**	**W7**	**W8**	**W9**	**S1**	**S3**
I1	0	2	1	2	3	4	4	2	5
W1	2	0	1	1	2	3	3	1	4
W3	1	1	0	1	2	3	3	1	4
W4	2	1	1	0	2	3	3	1	4
W7	3	2	2	2	0	1	1	1	2
W8	4	3	3	3	1	0	1	2	1
W9	4	3	3	3	1	1	0	2	1
S1	2	1	1	1	1	2	2	0	3
S3	5	4	4	4	2	1	1	3	0

1/14, 1/18, 1/18, 1/13, 1/24) = (.04, .06, .06, .06, .07, .06, .06, .08, .04). Solderman 1 is the most central.

EIGENVECTOR CENTRALITY

Eigenvector centrality is the refinement of degree centrality. As you remember, degree centrality is "blind" in the respect that it doesn't take into account anything past immediate friends. In many networks the importance of a node is dependent on the importance of the nodes to which it is connected, not the sheer number. In a high school an individual who is the friend of one very high status person becomes popular because of that one connection. The current men's basketball head coach at Oregon State University, Craig Robinson, enjoys a cache because he is President Obama's brother-in-law. In a communications network a node is exposed to lots of information if her immediate neighbors possess lots of information because they themselves are well connected.

An eigenvector **e** is a mathematical solution to this problem, the details of which we need not go into here. It is enough to say:

$$e_i \propto \sum_k a_{ik} e_k \text{ where " } \propto \text{ " means "proportional to."} \tag{10.4}$$

The actual computation of eigenvector centrality is beyond the scope of this book, but the results are easy to illustrate. Consider, for example, the nine workers in the largest component of the bank wiring room, Figure 10.2. The centrality of each worker is proportional to the sum of the centralities of the actors to which he is connected. The measure is basically degree centrality, but it weighs each person's friends by the amount of friends they have, and then the amount of friends the friends of their friends have, and so on.

$$3.22\begin{pmatrix}1.49\\4.43\\4.78\\4.43\\2.57\\1.61\\1.61\\5.04\\1\end{pmatrix}=\begin{pmatrix}4.78\\4.78+4.43+5.04\\1.78+4.43+4.43+5.04\\4.78+4.53+5.04\\5.04+1.61+1.61\\2.57+1+1.61\\2.57+1+1.61\\4.43+4.78+4.43+2.57\\1.61+1.61\end{pmatrix}=\begin{pmatrix}4.78\\14.25\\15.68\\14.25\\8.26\\5.18\\5.18\\16.21\\3.22\end{pmatrix}\tag{10.5}$$

In Figure 10.1, Solderman 1, who is connected only to other nodes of relatively high degree, is the most central by this measure.

BETWEENNESS CENTRALITY

Remember that centrality refers to two different phenomena: the degree to which a node in a network is advantageously placed and the degree to which a node is important for the functioning of the network as a means for distributing resources, where edges are transmission routes. Betweenness centrality is the latter type of centrality score.

Let's go back to your large retail store. Instead of placing your price scanner in a nice spot, suppose you have a new product that you want to promote. If you used the criterion represented through closeness centrality, you would place this product in the center of the store. However, the center of your store is not necessarily the place with the highest traffic, it's only the place that's the closest if someone from any random location in your store suddenly decides to confirm that you are really trying to hawk them an old Rick Astley CD for $15. What a rip-off. If you wanted to promote a product, you would probably want to place it in a location where most people would eventually pass by it. Any location by the checkout scanners or the entrance of your store would fit this new criterion.

In networks, betweenness centrality tries to find vertices that are in locations with high traffic. It represents a type of power that is distinct from the popularity of degree centrality or the influence of closeness centrality. For example, consider a network with three vertices: a factory, a wholesaler, and a retailer. The wholesaler's entire profitability is based on his betweenness centrality. If there were a direct connection from the factory or retailer, and no need to go through the wholesaler, the wholesaler would go out of business. That's the premise behind the

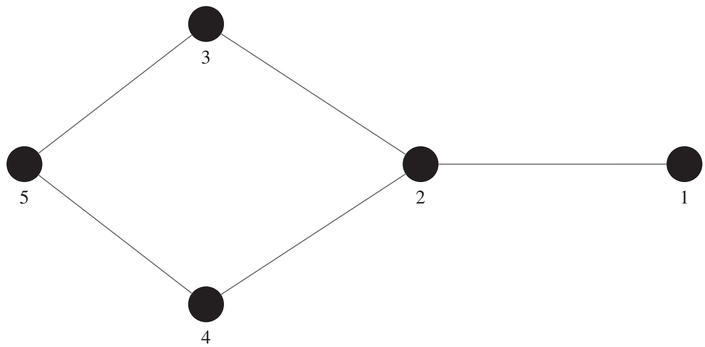

Figure 10.3. Example network for betweenness centrality.

marketing strategy of Costco, a wholesale retail store. Consumers believe they're getting better prices because they are bypassing the retail end of the chain, and buying wholesale. A newspaper editor has a lot of power because she is the gatekeeper between reporters and readers. She can selectively choose which stories to run and which stories to leave out. In social networks, betweenness centrality represents those who are in positions that offer unique advantages in the transmission or flow of goods or ideas.

Like closeness centrality, betweenness centrality is based on geodesics, or shortest paths. It measures the degree to which communication (or other quantities) flows through different vertices when members use the network to communicate with one another indirectly. To calculate betweenness centrality for a node i in a component of size n go through the following steps.

1. For each of the $(n-1)(n-2)/2$ pairs $\{j, k\}$ of vertices that do not include i, calculate all the geodesics connecting j and k that include and exclude vertex i. Take the proportion of geodesics connecting j and k that flow through node i.
2. Add up all these $(n-1)(n-2)/2$ proportions. This is the closeness centrality of vertex i. It's range is from 0 to $(n-1)(n-2)/2$.

To take a very simple example, look at the network in Figure 10.3.

Let us first consider Vertex 4. The following labeled matrix shows the proportion of the geodesics between every pair of vertices that pass through Vertex 4. For example, there is one geodesic between Vertices 3 and 4, and it does not pass through Vertex 4. Therefore the element in the row labeled 1 and the column labeled 3 is $0/1 = 0$. There are two geodesics between Vertices 1 and 5: 1-2-3-5 and 1-2-4-5. One of the two includes 4, so the element in the row labeled 1 and the column labeled 5 is 1/2.

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 5 \end{array} \\ \begin{pmatrix} X & \frac{0}{1} & \frac{0}{1} & \frac{1}{2} \\ X & X & \frac{0}{1} & \frac{1}{2} \\ X & X & X & \frac{0}{1} \\ X & X & X & X \end{pmatrix} \end{array} \quad \text{for Vertex 4} \tag{10.6}$$

The sum of all the numbers in Equation 10.6 is $0+0+1/2+0+1/2+0 = 1$, the betweenness centrality of Vertex 4. The entries in the matrix follow the ordering of vertex numbers, omitting the vertex for which the score is given. Equations 10.7 and 10.8 show the calculations for the betweenness centralities of Vertices 2 and 5.

$$\begin{array}{c} \\ 1 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} X & \frac{1}{1} & \frac{1}{1} & \frac{2}{2} \\ X & X & \frac{1}{2} & \frac{0}{1} \\ X & X & X & \frac{0}{1} \\ X & X & X & X \end{pmatrix} \end{array} \quad \text{for Vertex 2} \tag{10.7}$$

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} X & \frac{0}{1} & \frac{0}{1} & \frac{0}{1} \\ X & X & \frac{0}{1} & \frac{0}{1} \\ X & X & X & \frac{1}{2} \\ X & X & X & X \end{pmatrix} \end{array} \quad \text{for Vertex 5} \tag{10.8}$$

Adding up the values in the matrices for the different vertices, we find that the centrality scores for the five vertices is $(0, 3\frac{1}{2}, 1, 1, \frac{1}{2})$. Vertex 2 is by far the most central.

Looking at the bank wiring room, Solderman 1 in Figure 10.1 is "between" all the four nodes to his right and all the four vertices to his left. His betweenness centrality score is therefore 16. At the other extreme, Inspector 1 and Solderman 3 are not on any geodesics. The removal of a node that is high on betweenness centrality would eliminate many geodesics and therefore would tend to increase the distances between other pairs of nodes in the network. The removal of a node with a score of zero would have no effect on distances between other pairs.

Cut Points

Cut points are easy to define: the removal of a cut point increases the number of unconnected components in a graph, just like the elimination of an edge that is a bridge. In a communication network vertices that could communicate with each other indirectly cannot if the node is eliminated. Thus, cut points are a crude index of the importance of a node for the functioning of the network. In Figure 10.1 the removal Solderman 1 or Wireman 7 would divide the largest component in the bank wiring room into two almost equal parts. Wireman 3 is also a cut point because his removal would isolate on Inspector 1.

The following tables show all the measures of centrality for all the nodes in Figures 10.1 and 10.2.

At this point, please play around with the demonstration *Centrality* to become familiar with the properties of the measures. Pay particular attention to networks for which the different measures do not agree. The sizes of the vertices are proportional to their centralities using six centrality measures: degree, closeness, betweenness, cut points, eigenvector, and centers.

CENTRALIZATION

Centralization refers to the degree of inequality of centrality in a network. Centrality is a property of nodes, but centralization is a property of networks. Centralization is the degree to which centrality is monopolized by a small set of nodes in the network. If a network is centralized, there is a small core of highly central nodes and a large periphery of low centrality nodes. As we shall see later in the chapter on scale-free networks, the existence of a core within a network can have profound consequences for the operation of the network. Highly centralized networks can be more efficient in the distribution of information (once information reaches a core member it can easily and quickly be made available to all), but highly centralized networks can be very vulnerable if a core member is disabled. Inequality can also diminish the effectiveness of a network if participation by all members of a group is desirable; if all decisions are made by a small set of members, a group may not be able to take advantage of the knowledge and skills possessed by all its members. Highly centralized networks can also be very unequal in status and power, which may be considered undesirable on its own; a community in which only a small subset of interconnected individuals participate in all the governmental and nongovernmental organizations may be unhealthily and undemocratically centralized.

Once one has chosen an appropriate measure of centrality for the nodes, there is a standard way of computing the centralization of a network for that measure of centrality. The measure takes the star network (one central vertex connected to all others, none of which are connected to each other) as most centralized and measures the degree of deviation from that criterion. In the numerator of the measure of centralization is the difference between the centrality of the most central vertex and all other vertices. In

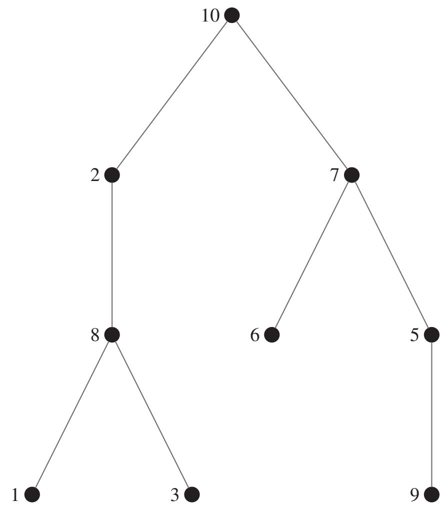

Figure 10.4. Network for Exercise 2.

the denominator is the sum of these differences in the star network with the same number of vertices, the most centralized structure. Suppose there are g actors in network G and actor $i*$ is the most central according to centrality measure A. Let $C_A(G)$ be the centralization of network G using measure A, and let $C_A(i)$ be the centrality of actor i using measure A. Then,

$$C_A(G) = \frac{\sum_j (C_A(i*) - C_A(j))}{\max \sum_j (C_A(i*) - C_A(j))} \tag{10.9}$$

The "max" refers to the distribution of centrality scores in a star network of the same size. For example, consider two networks: G_1 is a line network A-B-C-D-E; G_2 is the complete network with five vertices. Let's calculate degree centralization C_D for these two networks. In the star network with five vertices the degrees are 4, 1, 1, 1, and 1. The value of the denominator of Equation 10.9 is $(4-4)+(4-1)+(4-1)+(4-1)+(4-1)=12$.

Chapter Demonstrations

- *Centrality* illustrates various measures of network centrality applied to random networks
- *Centrality Game* tests your understanding of centrality by asking you to create networks with desired properties

Table 10.2.
Centrality scores for the bank wiring room network in Figure 10.1

	I1	W1	W3	W4	W7	W8	W9	S1	S3
Degree	1	3	4	3	3	3	3	4	2
Betweenness	0	0	7	0	15	3	3	16	0
Closeness	.043	.059	.062	.0591	.071	.056	.056	.077	.042
Graph center	0	0	0	0	1	0	0	1	0
Cut point	0	0	1	0	1	0	0	1	0
Eigenvector	1.49	4.43	4.78	4.43	2.57	1.61	1.61	5.04	1

Table 10.3.
Centrality scores for the marriage network in Figure 10.2

	Degree	Betweenness	Closeness	Center	Cut point	Eigenvector
Dall'Antella	1	0	.014	0	0	0.187
Ginori	3	35	.018	0	1	0.711
Dietisalvi	1	0	.014	0	0	0.187
Medici	5	44	.024	0	1	2.326
Guiciardini	1	0	.017	0	0	0.612
Tornabuoni	1	0	.017	0	0	0.612
Guasconi	6	73.5	.028	1	1	3.806
Albizzi	4	19.3	.024	1	0	3.092
Da Uzzano	1	0	.019	0	0	1.002
Rondinelli	2	9	.021	0	0	1.531
Della Casa	2	0	.020	0	0	1.531
Castellani	5	39.2	.020	0	1	2.008
Pepi	1	0	.015	0	0	0.529
Scambrilla	1	0	.015	0	0	0.529
Benizzi	0	4.5	.019	0	0	1.241
Peruzzi	6	51.3	.026	0	1	3.798
Strozzi	5	45.2	.026	0	1	2.703
Panciatichi	1	0	.017	0	0	0.712
Rucellai	1	0	.017	0	0	0.712
Aldobrandini	1	0	.018	0	0	1

EXERCISES

1. Using the demonstration *Centrality Game*, answer the question posed on the demonstration.
2. Using Figure 10.4, compute betweenness, closeness, and degree centralities for the following network. Hint about betweenness: In a tree there is only one geodesic between any pair of vertices.
3. Calculate closeness centralization for the centrality scores in Table 10.2.
4. Calculate betweenness centralization for the centrality scores in Table 10.3.