5

Centrality and Prestige

One of the primary uses of graph theory in social network analysis is the identification of the "most important" actors in a social network. In this chapter, we present and discuss a variety of measures designed to highlight the differences between important and non-important actors. Definitions of *importance*, or synonymously, *prominence*, have been offered by many writers. All such measures attempt to describe and measure properties of "actor location" in a social network. Actors who are the most important or the most prominent are usually located in strategic locations within the network. As far back as Moreno (1934), researchers have attempted to quantify the notions of sociometric "stars" and "isolates."

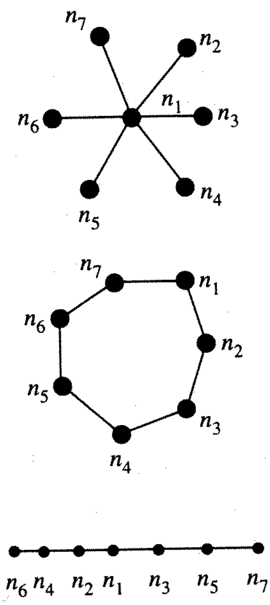
We will discuss the most noteworthy and substantively interesting definitions of importance or prominence along with the mathematical concepts that the various definitions have spawned. Among the definitions that we will discuss in this chapter are those based on *degree*, *closeness*, *betweenness*, *information*, and simply the *differential status* or *rank* of the actors. These definitions yield actor indices which attempt to quantify the prominence of an individual actor embedded in a network. The actor indices can also be aggregated across actors to obtain a single, group-level index which summarizes how variable or differentiated the set of actors is as a whole with respect to a given measure. We will show how to calculate both actor and group indices in this chapter.

Throughout this chapter, we will distinguish between relations that are directional (yielding directed graphs) and those that are not (yielding undirected graphs). The majority of the *centrality* concepts discussed in this chapter are designed for graphs (and thus, symmetric sociomatrices), and most of these, just for dichotomous relations. The notion of *prestige*, however, can only be quantified by using relations for which we can distinguish "choices" sent from choices received by the actors, and therefore, can only be studied with directed graphs. With directional relations, measures such as outdegree and indegree are quite likely to be different, and (as we will see in this chapter) prestigious actors are usually those with large indegrees, or "choices" received. Both centrality and prestige indices are examples of measures of the prominence or importance of the actors in a social network. We will consider definitions of prestige other than the indegree of an actor, and show that prestigious actors not only are chosen or nominated by many actors, but the actors who are doing the choosing must also be prestigious. So, the chapter will be split into two main parts: the first, presenting centrality measures for nondirectional relations, and the second, discussing both centrality and prestige measures for directional relations.

The substantive nature of the relation under study clearly determines which types of measures are appropriate for the network. Directional relations give two types of actor and group measures, based on both centrality and prestige, while nondirectional relations give just one type, based on centrality alone. We describe four well-known varieties of centrality in this chapter, illustrating and defining them first for nondirectional relations. We will then discuss directional relations, and not only show how these four centrality measures can be extended to such relations, but also define three measures of prestige, based on degree, proximity, and status or rank. This latter measure of status or rank has been shown to be quite useful in practice.

All these measures are first defined at the level of the individual actor. The measures can then be aggregated over all actors to obtain a group-level measure of either centralization or group prestige. Such aggregate measures are thus defined at the level of the entire set of actors. They attempt to measure how "centralized" or "prestigious" the set of actors is as a whole. We will present several methods for taking the individual actor indices, and combining them to arrive at a single, group-level index. These methods are as simple as variances, and as complicated as ratios of the average difference of the actor index from its maximum possible value to the maximum of this average difference. The group-level indices are usually between 0 and 1, and thus are not difficult to interpret.

Throughout the chapter, we will apply the actor and group measures to a variety of data, both real and artificial. Three artificial graphs that very nicely highlight the differences among the measures we describe are shown in Figure 5.1. These graphs, all with $g = 7$, will be labeled the star graph (Figure 5.1a), the circle (Figure 5.1b), and the line graph (Figure 5.1c; see



(a) Star graph

0	1	1	1	1	1	1
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0

(b) Circle graph

0	1	0	0	0	0	1
1	0	1	0	0	0	0
0	1	0	1	0	0	0
0	0	1	0	1	0	0
0	0	0	1	0	1	0
0	0	0	0	1	0	1
1	0	0	0	0	1	0

(c) Line graph

0	1	1	0	0	0	0
1	0	0	1	0	0	0
1	0	0	0	1	0	0
0	1	0	0	0	1	0
0	0	1	0	0	0	1
0	0	0	1	0	0	0
0	0	0	0	1	0	0

Fig. 5.1. Three illustrative networks for the study of centrality and prestige

Freeman 1980a). We will refer to these graphs or networks frequently, since the centrality of the actors in these graphs varies greatly, as does the centralization of the graphs. Just a quick glance at these figures shows that the nodes in the graphs are quite different. For example, all nodes in the circle are interchangeable, and hence should be equally central. One node in the star completely outranks the others, while the other six themselves are interchangeable. In the line graph, the nodes' centrality clearly decreases from that for n_1, to n_2 and n_3, and so on, to n_6 and n_7, who are peripheral in this graph.

Many graph theoretic centrality concepts are discussed in Hage and Harary (1983) and in the other general references given in Chapter 4. Based on our understanding of the major concepts of graph theory, as presented in Chapter 4, it should be clear that we can define (maybe even invent) many graph theoretic centrality notions, such as the "center" and "centroid" of a graph, with the goal of quantifying importance or prominence. But the major question still remains unanswered: Are the nodes in the graph center and/or in the graph centroid and/or with maximal degree the most "central" nodes in a substantive sense — that is, does the center, or centroid, of a graph contain the most important actors? In part, this is a question about the validity of the measures of centrality — do they really capture what we substantively mean be "importance" or "prominence"? Can we simply focus on the actors who are "chosen" the most to find the most important actors? Of course, unless we define what we mean by the terms "important" and "prominent," these questions are not answerable.

Thus, we first will define prominence or importance, and discuss how the terms "central" and "prestigious" quantify two important aspects of prominence. We will then answer questions about which actors are the most important, and will find that the best centrality notions are first based primarily on substantive theory, and then use graph theory to be quantified.

5.1 Prominence: Centrality and Prestige

We begin by assuming that one has measurements on a single, dichotomous relation, although some of the measures discussed here are generalizable to other types of network data. We will not be concerned here with a signed or multirelational situation, even though such situations are very interesting (both methodologically and substantively). These types of relations have not been studied using the ideas discussed in this chapter.

We will consider an actor to be *prominent* if the ties of the actor make the actor particularly visible to the other actors in the network. This equating of prominence to visibility was made by Knoke and Burt (1983). Hubbell (1965) and Friedkin (1991) note that prominence should be measured by looking not only at direct or adjacent ties, but also at indirect paths involving intermediaries. This philosophy is maintained throughout. To determine which of the g actors in a group are prominent, one needs to examine not only all "choices" made by an actor and all "choices" received, but indirect ties as well.

If a relation is nondirectional, the ith row of the sociomatrix $\mathbf{X}$, $(X_{i1}, X_{i2}, \ldots, X_{ig})$, is identical to the ith column $(X_{1i}, X_{2i}, \ldots, X_{gi})$. Thus, actor i's prominence within a network is based on the pattern of these $g-1$ possible ties or entries in the sociomatrix, defining the location of actor i. If the relation is directional, the ith row of the sociomatrix differs from the ith column, so that actor i's prominence is based on the $2(g-1)$ entries in the sociomatrix involving i. Some of the specific definitions of prominence will also consider choices made through intermediaries, or third parties, but such choices will almost always be of secondary concern.

This definition of prominence is still rather vague. Are prominent actors the objects of many "choices" from followers, while non-prominent actors (or followers) are not? What properties of these "choices" make an actor more visible than the other actors or the "object of" many ties? And what shall we do about indirect choices? This definition is also relative to the nature of the "choices" made by the other actors. Prominence is difficult to quantify, since many actor indices that are functions of just the ith row and column of the sociomatrix would qualify as measures of prominence.

To allow researchers to define better the important actors as those with more visibility and to understand better the meaning of the concept, Knoke and Burt distinguish two types of visibility, or to us, two classes of prominence — centrality and prestige. Both these types are based on the relational pattern of the row and column entries of the sociomatrix associated with each actor. This dichotomy is very useful and a very important contribution to the extensive literature on prominence. Let us now define both these versions of prominence, after which we will show how they can be quantified first for nondirectional relations, and then for directional ones.

5.1.1 Actor Centrality

Prominent actors are those that are extensively involved in relationships with other actors. This involvement makes them more visible to the others. We are not particularly concerned with whether this prominence is due to the receiving (being the recipient) or the transmission (being the source) of many ties — what is important here is that the actor is simply *involved*. This focus on involvement leads us to consider first nondirectional relations, where there is no distinction between receiving and sending. Thus, for a nondirectional relation, we define a *central* actor as one involved in many ties. However, even though centrality seems most appropriate for nondirectional relations, we will, later in this chapter, show how such indices can also be calculated for directional relations.

This definition of centrality was first developed by Bavelas (1948, 1950). The idea was applied in the late 1940's and early 1950's in laboratory experiments on communication networks (rather than from observed, naturally occurring networks) directed by Bavelas and conducted by

Leavitt (1949, 1951), Smith (1950), and Bavelas and Barrett (1951). As Freeman (1979) reports, these first experiments led to many more experiments in the 1950's and 1960's (see Burgess 1968, Rogers and Agarwala-Rogers 1976, and the citations in Freeman 1979, for reviews). In recent research, Freeman (1977, 1979, 1980a) has advocated the use of centrality measures to understand group structure, by systematically defining the centrality notions we discuss below. At the same time, he introduced a new centrality measure based on betweenness (see below).

As Knoke and Burt (1983) point out, sociological and economic concepts such as access and control over resources, and brokerage of information, are well suited to measurement. These concepts naturally yield a definition of centrality since the difference between the source and the receiver is less important than just participating in many interactions. Assuming that one is studying a relevant relation (such as communication), those actors with the most access or most control or who are the most active brokers will be the most central in the network.

We will employ a simple notation for actor centrality measures, first used by Freeman (1977, 1979). We let C denote a particular centrality measure, which will be a function of a specific n_i. There will be a variety of measures introduced in this chapter, so we will subscript C with an index for the particular measure under study. If we let A be a generic measure, then one of the actor centralities defined below will be denoted by $C_A(n_i)$. We will use a variety of different values for A to distinguish among the different versions of centrality. As usual, the index i will range over the integers from 1 to g.

5.1.2 Actor Prestige

Suppose we can make a distinction between ties sent and ties received, as is true for directional relations. We define a *prestigious* actor as one who is the object of extensive ties, thus focusing solely on the actor as a recipient. Clearly, prestige is a more refined concept than centrality, and cannot always be measured. The prestige of an actor increases as the actor becomes the object of more ties but not necessarily when the actor itself initiates the ties. In other words, one must look at ties directed *to* an actor to study that actor's prestige. Since indegrees are only distinguishable from outdegrees for directional relations, we will not be able to quantify prestige of an actor unless the relation is directional, a point that we discuss in more detail below.

Quantification of prestige, and the separation of the concept from centrality, is somewhat analogous to the distinction frequently made between outdegrees and indegrees (which, as the reader will see, are simple measures of centrality and prestige, respectively). One must look at ties directed to an actor to study that actor's prestige. Since indegrees are only distinguishable from outdegrees for directional relations, we will not be able to quantify prestige unless the relation is directional.

We should note that the term "prestige" is perhaps not the best label for this concept (in some situations). For example, if the relation under study is one of negative affect, such as "despises" or "do not want as a friend," then actors who are prestigious on this relation are not held in very high regard by their peers. Such actors are certainly renowned, but it is for negative feelings, rather than positive. Further, if the relation is "advises," the actors considered prestigious by their peers might be those that are senders, rather than receivers. Nevertheless, the term has become established in the literature, and we will use it, keeping in mind that the substantive nature of the measured relation is quite important when interpreting the property.

Prestige has also been called *status* by authors such as Moreno (1934), Zeleny (1940a, 1940b, 1941, 1960), Proctor and Loomis (1951), Katz (1953), and Harary (1959c). We will introduce several status measures later in this chapter. But we will label these indices *rank* measures, since the term "status" has been used extensively in other network methodology (see Chapters 9 and 10). All these actor prestige measures attempt to quantify the rank that a particular actor has within a set of actors. Other synonyms include *deference*, and simply *popularity*. Recently, Bonacich (1972a, 1972b, 1987) has generalized Katz's (1953), Hubbell's (1965), and Taylor's (1969) ideas, and presented a new family of rank measures. All these rank (or status) indices are examples of prestige measures, and we will discuss them in detail later in the chapter.

We let P denote a particular prestige measure, which will be defined for a specific actor, n_i. There will be three measures introduced in this chapter, so we will subscript P with an index for the particular measure under study.

5.1.3 Group Centralization and Group Prestige

We should note that even though the focus of this chapter is on measures for actors that primarily allow us to quantify importance, one can take many of the measures and combine them across actors to get a group-level measure. These group-level measures allow us to compare different networks easily. When possible in this chapter, we will give formulas for group centralization or prestige measures, although most research on these measures is restricted to centralization.

We should first ask exactly what a group-level index of centralization is measuring. The general index that we introduce below has the property that the larger it is, the more likely it is that a single actor is quite central, with the remaining actors considerably less central. The less central actors might be viewed as residing in the periphery of a centralized system. Thus, this group-level quantity is an index of *centralization*, and measures how variable or heterogeneous the actor centralities are. It records the extent to which a single actor has high centrality, and the others, low centrality. It also can be viewed as a measure of how unequal the individual actor values are. It is (roughly) a measure of variability, dispersion, or spread. Early network researchers interested in centrality, particularly Leavitt (1951), Faucheux and Moscovici (1960), and Mackenzie (1966a), proposed that group-level indices of centralization should reflect such tendencies. Nieminen (1974) and Freeman (1977) also adopt this view, and discuss group centralization measurement.

One can view such a centralized network in Figure 5.1. The star graph is maximally central, since its one central actor has direct contact with all others, who are not in contact with each other. Examining the other two graphs in this figure should indicate that the degree of centralization can vary just by changing a few ties in the network.

Freeman (1979) adopts a convenient, general mathematical definition for a group-level index of centralization. Recall that $C_A(n_i)$ is an actor centrality index. Define $C_A(n^*)$ as the largest value of the particular index that occurs across the g actors in the network; that is, $C_A(n^*) = \max_i C_A(n_i)$.

From these quantities, $\sum_{i=1}^{g}[C_A(n^*) - C_A(n_i)]$ is the sum of the differences between this largest value and the other observed values, while $\max \sum_{i=1}^{g}[C_A(n^*) - C_A(n_i)]$ is the theoretical maximum possible sum of differences in actor centrality, where the differences are taken pairwise between actors. This latter maximum is taken over all possible graphs, with g actors. As we will see, this maximum occurs for the star graph.

The sum of differences becomes the numerator, while the theoretical maximum possible sum becomes the denominator in Freeman's index. The denominator is a theoretical quantity, and is not computed by looking at a specific graph; rather, it is calculated by considering all possible networks, with a fixed g, and then determining analytically how large the sum of differences can actually be. We have the general centralization index:

$$C_A = \frac{\sum_{i=1}^{g} [C_A(n^*) - C_A(n_i)]}{\max \sum_{i=1}^{g} [C_A(n^*) - C_A(n_i)]}. \quad (5.1)$$

The index will always be between 0 and 1. C_A equals 0 when all actors have exactly the same centrality index (that being $C_A(n^*)$), and equals 1 if one actor, "completely dominates or overshadows" the other actors.

Yet another view of graph centralization is offered by Høivik and Gleditsch (1975), who view centralization in a graph more simply than Freeman as the *dispersion* in a set of actor centrality indices. Later in this chapter, we show how such a view is related to Freeman's approach.

We note that one could also construct group-level prestigious measures, but the theoretical maximum values needed in the denominator are usually not calculable (except in special cases). Thus, we usually use something simpler (as we note later in this chapter) like a variance.

In addition to centralization measures, other researchers have proposed graph-level indices based on the *compactness* of a graph. Bavelas (1950), Flament (1963), Beauchamp (1965), and Sabidussi (1966) state that very centralized graphs are also compact, in the sense that the distances between pairs of nodes are small. These authors also proposed an index of actor centrality based on closeness (that is, small distances), as we will discuss later in this chapter.

We will illustrate the quantities defined in this chapter using two examples. First, we will continue to use Padgett's Florentine family network as an example of a network with a nondirectional relation. Second, we introduce the countries trade network as an example of a network of nations, with trade of basic manufactured goods as a directional relation.

5.2 Nondirectional Relations

Suppose that we have a single set of actors, and a single, dichotomous nondirectional relation measured on the pairs of actors. As usual, we let $\mathbf{X}$ refer to the matrix of social network data. For such data, the ith row of the sociomatrix is identical to the ith column. An example of such a matrix can be found in Appendix B, and discussed in Chapter 2. These data measure the alliances among families in 15th century Florence formed by interfamilial marriages. The corresponding sociogram is shown in Chapter 3, where it is discussed at length as an example of a graph with 16 nodes. In order to find the most important actors, we will look for measures reflecting which actors are at the "center" of the set of actors. We will introduce several definitions of this center, including actors with maximum degree, betweenness, closeness, and information.

5.2.1 Degree Centrality

The simplest definition of actor centrality is that central actors must be the most active in the sense that they have the most ties to other actors in the network or graph. Nowhere is this easier to see than by comparing a graph resembling a star to one resembling a circle, shown in Figure 5.1 for networks with g=7 actors. A star graph has the property that exactly one actor has ties to all $g - 1$ other actors, and the remaining $g - 1$ actors have only their single tie to the first actor. The first actor is clearly the most active, and one could view this high level of activity as a large amount of centrality. This very active actor should thus have a maximal centrality index. Here, we measure activity simply as degree. Contrast this star graph with the circle graph also shown in Figure 5.1. A circle has no actor more active than any other actor; indeed, all actors are interchangeable, so all actors should have exactly the same centrality index. Note also that this type of centrality focuses only on direct or adjacent choices. Prominence here is equated to "activity" or simply "degree."

Actor Degree Centrality. The degree of an actor is important; thus, a centrality measure for an individual actor should be the degree of the node, $d(n_i)$. Thus, following suggestions made by Proctor and Loomis (1951) and Shaw (1954), and then many other researchers (Glanzer and Glaser 1959; Faucheux and Moscovici 1960; Garrison 1960; Mackenzie 1964, 1966a; Pitts 1965; Nieminen 1973, 1974; Czepiel 1974; Rogers 1974; and Kajitani and Maruyama 1976; and reviewed by Freeman 1979), we define $C_D(n_i)$ as an actor-level degree centrality index. We let

$$C_D(n_i) = d(n_i) = x_{i+} = \sum_j x_{ij} = \sum_j x_{ji}. \qquad (5.2)$$

We need not comment on the properties of this measure; it is discussed in detail in Chapter 4. We do note that one problem with this measure is that it depends on the group size g; indeed, its maximum value is $g - 1$. Consequently, a proposed standardization of the measure

$$C'_D(n_i) = \frac{d(n_i)}{g-1} \tag{5.3}$$

is the proportion of nodes that are adjacent to n_i. $C'_D(n_i)$ is independent of g, and thus can be compared across networks of different sizes.

Donninger (1986) considers the distribution of equation (5.3), using the probabilistic graph models of Erdös and Renyi (1960). He gives an approximation to the distribution of degrees, which can then be used to place confidence intervals on both the actor- and group-level degree indices.

A related index, one for "ego density," is given by Burt (1982) and Knoke and Kuklinski (1982). An ego density for a nondirectional relation is simply the ratio of the degree of an actor to the maximum number of ties that could occur. Kapferer (1969, 1973) generalizes this, and defines another index, the "span" of an actor, as the percentage of ties in the network that involve the actor or the actors that the primary actor is adjacent to. Thus, the central actor in a star graph has a span of unity.

Refer to the three graphs of Figure 5.1. The degrees for the seven actors in the star graph are 6 (for n_1) and 1 (for $n_2 - n_7$). Thus, the denominator for the standardized actor-level indices $C'_D(n_i)$ is $g - 1 = 6$. The standardized indices have values $\{1.0, 0.167, \ldots, 0.167\}$ — clearly there is one maximally central actor, and six peripheral actors. The degrees for the circle graph are all $d(n_i) = 2$, so that the indices are all equal: $C'_D(n_i) = 0.333$, indicating a low-moderate level of centrality, constant across all actors. Lastly, contrast this network to the line graph, in which $n_1 - n_5$ all have $C'_D(n_i) = 0.333$ also, but the last two actors are less central: $C'_D(n_6) = C'_D(n_7) = 0.167$. The absence of the line between n_6 and n_7 (which is the difference between the circle graph and the line graph) has forced these two actors to be less central than the other five. These centralities and standardized centralities were calculated by hand, although the program *UCINET* calculates these quantities as standard output of its centrality subprogram.

An actor with a high centrality level, as measured by its degree, is "where the action is" in the network. Thus, this measure focuses on the most visible actors in the network (as required by Knoke and Burt's (1983) definition of prominence). An actor with a large degree is in direct contact or is adjacent to many other actors. This actor should then begin to be recognized by others as a major channel of relational information, indeed, a crucial cog in the network, occupying a central location. In contrast, and in accordance with this centrality definition, actors with low degrees are clearly peripheral in the network. Such actors are not active in the relational process. In fact, if the actor is completely isolated (so that $d(n_i) = 0$), then removing this actor from the network has no effect on the ties that are present.

Group Degree Centralization. We now present several degree-based measures of graph centralization. A centralization measure quantifies the range or variability of the individual actor indices. The set of degrees, which represents the collection of actor degree indices, can be summarized in a variety of ways. Freeman (1979) recommends use of the general index (5.1). Applying his general formula for graph centralization here we find

$$C_D = \frac{\sum_{i=1}^{g}[C_D(n^*) - C_D(n_i)]}{\max \sum_{i=1}^{g}[C_D(n^*) - C_D(n_i)]}. \tag{5.4}$$

The $\{C_D(n_i)\}$ in the numerator are the g actor degree indices, while $C_D(n^*)$ is the largest observed value. The denominator of this index can be calculated directly (see Freeman 1979), and equals $(g-1)(g-2)$. Thus,

$$C_D = \frac{\sum_{i=1}^{g}[C_D(n^*) - C_D(n_i)]}{[(g-1)(g-2)]} \tag{5.5}$$

can be used as an index to determine how centralized the degree of the set of actors is. The index is also a measure of the dispersion or range of the actor indices, since it compares each actor index to the maximum attained value.

This index reaches its maximum value of 1 when one actor chooses all other $g-1$ actors, and the other actors interact only with this one, central actor. This is exactly the situation in a star graph. The index attains its minimum value of 0 when all degrees are equal, indicating a *regular* graph (as defined in Chapter 4). This is exactly the situation realized in the circle graph. Graphs that are intermediate to these two (such as the line graph of Figure 5.1) have indices between 0 and 1, indicating varying amounts of centralization of degree. In fact, the line graph has a $C_D = 0.067$.

Another standard statistical summary of the actor degree indices is the variance of the degrees,

$$S_D^2 = \left[\sum_{i=1}^{g}(C_D(n_i) - \overline{C}_D)^2\right] / g, \tag{5.6}$$

where $\overline{C}_D$ is the mean actor degree index. The variance is recommended as a group-level index of centrality by Snijders (1981a, 1981b), reflecting the view of Høivik and Gleditsch (1975) that centralization is synonymous with the *dispersion* or heterogeneity of an actor index. This index attains its minimum value of 0 when all degrees are equal or when the graph is regular.

The maximum value of S_D^2 depends on g and the entire set of degrees. Snijders (1981a, 1981b) recommends that one normalize S_D^2 by the maximum possible variance given the set of degrees actually observed, to obtain a dimensionless index. The formulas for undirected graphs are complicated; we refer those interested to Snijders (1981a, 1981b). The formulas for directed graphs are easier to report, and we do so later in this chapter when we discuss directional relations. One can also test statistically whether a graph is more heterogeneous (with regard to its degree distribution) than expected by chance. Tests such as this one will be described in general in Chapter 13.

Coleman (1964) also recommends the use of S_D^2 as a measure of "hierarchization" (similar to centralization). In fact, Coleman goes on to suggest that one use a more general function of the degrees for this measure; in particular, he chooses the function $x\log(x)$, which yields an information- or entropy-based measure of hierarchization, not unlike those proposed by Mackenzie (1966b) or Stephenson and Zelen (1989) (see below).

There are simpler group-level degree indices. In fact, recognizing that the simplest actor-level index is the degree of the actor, one can take the average of the degrees to get the mean degree, $\overline{C}_D = \sum C_D(n_i)/g = \sum x_{i+}/g$. This quantity varies between 0 and $g-1$, so to standardize it, one should divide by $g-1$. This average degree, divided by $g-1$, is exactly the density of the graph: $\sum C_D(n_i)/g(g-1) = \sum C'_D(n_i)/g = \Delta$. Thus, mathematically, the density is also the average standardized degree. The densities of the three graphs in Figure 5.1 are 0.286 (star), 0.333 (circle), and 0.286 (line).

The density of a graph is perhaps the most widely used group-level index. It is a recommended measure of group cohesion (see Blau 1977), and its use can be traced back at least as far as Kephart (1950) and Proctor and Loomis (1951). Bott (1957) used densities to quantify network "knittedness," while Barnes (1969b) used them to determine how "close-knit" empirical networks were. It is very important in blockmodels and other role-algebraic techniques (see Part IV, particularly Chapter 10). Density takes on values between 0 (empty graph) and 1 (complete graph), and is the average of the standardized actor degree indices, $\{C'_D(n_i)\}$, as well as the fraction of possible ties present in the network for the relation under study. Friedkin (1981) studies the use of density as a summarization tool in network analysis, and concludes that densities can be misleading, especially if the values are small. This result is often due to the fact that as group sizes increase, network density decreases if actor degrees remain unchanged. Friedkin recommends that both density and group size be considered simultaneously, especially if the graph shows tendencies toward subgrouping (see Chapter 7).

The density of a graph is, thus, an overly simplified version of a group-level degree index, constructed by taking the actor degree indices and ignoring Freeman's two principles for group-level indices. It is also an average. As is quite common in data analysis, averages are sometimes difficult to interpret. One also needs information on how dispersed the numbers that make up the average are. So, one frequently computes the variance of these numbers, and reports it along with the average. We therefore recommend the simultaneous use of centralization measures such as S_D^2 and C_D along with average degree and graph density.

It is important to note, however, that indices such as average degree and density are not really centralization measures. As mentioned earlier, centralization should quantify the range or variability of the individual actor indices. Thus, S_D^2, and of course C_D are valid centralization measures, while the average degree or the graph density, which are quantifications of average actor tendencies rather than variability, are not.

Example. Turn now to Padgett's network of Florentine families and examine the marriage relation. The standardized actor degree centralities are shown in the first column of Table 5.1 (along with other actor-centrality and centralization indices which will be discussed later in this chapter). These centralities were calculated using *UCINET*.

One can see that the Medici family (n_9) is the most central family, with respect to degree. For this actor, $C_D'(n_9) = 0.400$, an index considerably larger than the next most central actors (Guadagni and Strozzi families), with $C_D'(n_7) = C_D'(n_{15}) = 0.267$. Six of the families have an index of 0.200; the remaining seven families have small indices. The group-level degree centralization index is $C_D = 0.267$, a rather small value, indicating that the difference between the largest and smallest actor-level indices is not very great. There is little variability. The average degree is $\overline{C}_D = 40/16 = 2.50$, quite small, but not surprising given the nature of the relation (marital ties, something not particularly common). We also note that the variance of the degrees (not the standardized actor

Table 5.1. *Centrality indices for Padgett's Florentine families (*Actor and centralization indices calculated by dropping n_{12} = Pucci from the actor set.)*

	With g = 16 actors		With g = 15 actors			
	$C'_D(n_i)$	$C'_B(n_i)$	$C'_D(n_i)^*$	$C'_C(n_i)^*$	$C'_B(n_i)^*$	$C'_I(n_i)^*$
Acciaiuoli	0.067	0.000	0.071	0.368	0.000	0.049
Albizzi	0.200	0.184	0.214	0.483	0.212	0.074
Barbadori	0.133	0.081	0.143	0.438	0.093	0.068
Bischeri	0.200	0.090	0.214	0.400	0.104	0.074
Castellani	0.200	0.048	0.214	0.389	0.055	0.070
Ginori	0.067	0.000	0.071	0.333	0.000	0.043
Guadagni	0.267	0.221	0.286	0.467	0.255	0.081
Lamberteschi	0.067	0.000	0.071	0.326	0.000	0.043
Medici	0.400	0.452	0.429	0.560	0.522	0.095
Pazzi	0.067	0.000	0.071	0.286	0.000	0.033
Peruzzi	0.200	0.019	0.214	0.368	0.022	0.069
Pucci	0.000	0.000	—	—	—	—
Ridolfi	0.200	0.098	0.214	0.500	0.114	0.080
Salvati	0.133	0.124	0.143	0.389	0.143	0.050
Strozzi	0.267	0.089	0.286	0.438	0.103	0.070
Tornabuoni	0.200	0.079	0.214	0.483	0.092	0.080
Centralization	0.267	0.383	0.257	0.322	0.437	—

degree centrality indices) $S_D^2 = 2.125$, and the density of this relation (which is the average standardized degree) is 0.167, indicating (as noted) a relatively sparse sociomatrix. The density of this relation is quite a bit less than that for the three hypothetical graphs in Figure 5.1, for instance.

5.2.2 Closeness Centrality

The second view of actor centrality is based on closeness or distance. The measure focuses on how *close* an actor is to all the other actors in the set of actors. The idea is that an actor is central if it can quickly interact with all others. In the context of a communication relation, such actors need not rely on other actors for the relaying of information, an idea put forth by Bavelas (1950) and Leavitt (1951). As noted by Beauchamp (1965), actors occupying central locations with respect to *closeness* can be very productive in communicating information to the other actors. If the actors in the set of actors are engaged in problem solving, and the focus is on communication links, efficient solutions occur when one actor has very short communication paths to the others. Thus, this *closeness* view of centrality relies heavily on economic considerations.

Hakimi (1965) and Sabidussi (1966) quantified this notion that central actors are close, by stating that central nodes in a network have "minimum steps" when relating to all other nodes; hence, the geodesics, or shortest paths, linking the central nodes to the other nodes must be as short as possible. With this explanation, researchers began equating closeness with *minimum distance.* The idea is that centrality is inversely related to distance. As a node grows farther apart in distance from other nodes, its centrality will decrease, since there will be more lines in the geodesics linking that node to the other nodes.

Examine the star network in Figure 5.1. The node at the center of this star is adjacent to all the other nodes, has the shortest possible paths to all the other actors, and hence has maximum closeness. There is exactly one actor who can reach all the other actors in a minimum number of steps. This actor need not rely on the other actors for its interactions, since it is tied to all others.

Actor Closeness Centrality. Actor centrality measures reflecting how close an actor is to the other actors in the network have been developed by Bavelas (1950), Harary (1959c), Beauchamp (1965), Sabidussi (1966), Moxley and Moxley (1974), and Rogers (1974). As reviewed by Freeman (1979), the simplest measure is that of Sabidussi (1966), who proposed that actor closeness should be measured as a function of geodesic distances. As mentioned above, as geodesics increase in length, the centrality of the actors involved should decrease; consequently, distances, which measure the length of geodesics, will have to be weighted inversely to arrive at Sabidussi's index. Note how this type of centrality depends not only on direct ties, but also on indirect ties, especially when any two actors are not adjacent.

We let $d(n_i, n_j)$ be the number of lines in the geodesic linking actors i and j; that is, as defined in Chapter 4, $d(\bullet, \bullet)$ is a distance function. The total distance that i is from all other actors is $\sum_{j=1}^{g} d(n_i, n_j)$, where the sum is taken over all $j \neq i$. Thus, Sabidussi's (1966) index of actor closeness is

$$C_C(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j) \right]^{-1}. \tag{5.7}$$

The subscript C is for "closeness." As one can see, the index is simply the inverse of the sum of the distances from actor i to all the other actors.

At a maximum, the index equals $(g-1)^{-1}$, which arises when the actor is adjacent to all other actors. At a minimum, the index attains the value of 0 in its limit, which arises whenever one or more actors are not reachable from the actor in question. A node is said to be *reachable* from another node if there is a path linking the two nodes; otherwise, the nodes are not reachable from each other. Thus, the index is only meaningful for a connected graph.

To verify this assertion, suppose that the graph is disconnected — specifically, let there be one isolated node, with degree 0. The geodesics from all the other nodes to this specific node (n_k) are infinitely long ($d(n_i, n_k) = \infty$ for all $i \neq k$), since the node is not reachable. Hence, the distance sum for every actor is ∞, and the actor closeness indices are all 0. This is a large drawback of this index.

As we have noted, the maximum value attained by this index depends on g; thus, comparisons of values across networks of different sizes are difficult. Beauchamp (1965) made the suggestion of standardizing the indices so that the maximum value equals unity. To do this, we simply multiply $C_C(n_i)$ by $g-1$:

$$\begin{aligned} C'_C(n_i) &= \frac{g-1}{\left[\sum_{j=1}^{g} d(n_i, n_j)\right]} \\ &= (g-1)C_C(n_i). \end{aligned} \tag{5.8}$$

This standardized index ranges between 0 and 1, and can be viewed as the inverse average distance between actor i and all the other actors. It equals unity when the actor is adjacent to all other actors; that is, when the actor is maximally close to all other actors.

Graph theorists have simplified this concept of centrality, and talked about the *center* of a graph, using the graph-theoretic notion of distance (see Chapter 4). Specifically, the *Jordan center* (see Jordan 1869) of a graph is the subset of nodes that have the smallest maximum distance to all other nodes. To find such a center, one can take a $g \times g$ matrix of geodesic distances between pairs of nodes (where the entries are the lengths of the shortest paths or geodesics between all pairs of nodes), and then find the largest entry in each row. These distances (which are sometimes called *eccentricities*) are the maximum distances from every actor to their fellow actors. One then simply finds the smallest of these maximum distances. All nodes that have this smallest maximum distance are part of the center of the graph.

A related notion is the *centroid* of a graph (see Sylvester 1882), which is based on the degrees of the nodes and which is most appropriate for graphs that are trees. The idea is to consider all branches or paths emanating from each node, and define the *weight* of each branch as the number of lines in it. The weight of a node is the maximum weight of any branch at the node. The *centroid* is thus the subset of all nodes that have the smallest weight.

All the graphs in Figure 5.1 are connected, so that all geodesic distances are finite; therefore, the closeness indices can be calculated. For the star graph, $C'_C(n_1) = 1.0$, while the other actors all have indices equal to 0.545. For the circle graph, the actor indices are all equal to 0.5. For the line graph, the indices vary from $C'_C(n_1) = 0.50$ to a low of $C'_C(n_6) = C'_C(n_7) = 0.286$.

We note that there are clever algorithms for finding the geodesics in a graph, and then computing their lengths. We refer the reader to (for example) Flament (1963), and Harary, Norman, and Cartwright (1965). Such algorithms are standard in network computing programs such as *UCINET* and *SNAPS* (see Appendix A).

Group Closeness Centralization. We now consider how to measure group centralization using actor closeness centralities. We first report Freeman's (1979) index, which uses the general graph centralization index, (5.1), given above. We then will consider alternative group closeness indices.

Freeman's general group closeness index is based on the standardized actor closeness centralities, shown in equation (5.8). This index has numerator

$$\sum_{i=1}^{g}[C'_C(n^*) - C'_C(n_i)],$$

where $C'_C(n^*)$ is the largest standardized actor closeness in the set of actors. Freeman shows that the maximum possible value for the numerator is $[(g-2)(g-1)]/(2g-3)$, so that the index of group closeness is

$$C_C = \frac{\sum_{i=1}^{g}[C'_C(n^*) - C'_C(n_i)]}{[(g-2)(g-1)]/(2g-3)}. \tag{5.9}$$

This index, as with the group degree centralization index, reaches its maximum value of unity when one actor "chooses" all other $g-1$ actors (that is, has geodesics of length 1 to all the other actors), and the other actors have geodesics of length 2 to the remaining $(g-2)$ actors. This is exactly the situation realized by a star graph. The proof of this fact is rather complicated, and must be done by induction. We refer the reader to Freeman (1979). The index can attain its minimum value of 0 when the lengths of geodesics are all equal, for example in a complete graph or in a circle graph. For the line graph of Figure 5.1, the index equals 0.277, a relatively small value.

Bolland (1988) proposes a measure (for both actors and groups) that utilizes both degree and closeness of actors. His "continuing flow" centrality index is based on the number of paths (of any length) that originate with each actor. Thus, the measure considers all paths, those of length 1 (that are the focus of C_D) and those indirect (whose distances are reflected in the magnitude of C_C). We discuss this measure in more detail at the end of this chapter.

There are other group-level closeness indices. We may simply summarize the set of g actor-level closeness centralities $\{C'_C(n_i)\}$ by a single statistic, reflecting the tendency toward closeness manifested by all the actors in the set of actors. Such a statistic, to be an effective index, should reach its extremes in the cases of the circle graph (equal distances), and the star graph (one minimally distant actor).

We recommend that one calculate the variance of the standardized actor closeness indices,

$$S_C^2 = \left[\sum_{i=1}^{g} (C'_C(n_i) - \overline{C}_C)^2 \right] / g, \tag{5.10}$$

which summarizes the heterogeneity among the $\{C'_C(n_i)\}$. We note that average normed closeness, $\overline{C}_C = \sum C'_C(n_i)/g$, is simply the mean of the actor-level closeness centralities. The variance attains its minimum value of 0 in a network with equal actor indices (in this case, equal distances between all nodes). Such a network need not be complete (have maximal degree). This index grows as the network becomes less homogeneous (with respect to distances), and thus more centralized. The average normed closeness, $\overline{C}_C$, together with S_C^2, provide simple summary statistics for the entire set of actor closeness indices.

The Example Again. Consider again Padgett's network data, discussed earlier. Actor n_{12} = Pucci (as can be seen from the actor degree centrality value of $C_D(n_{12}) = 0$) is an isolate. Consequently, the distances to this actor from all other actors are infinite, and thus, family Pucci is not reachable and the graph is not connected. Actor closeness centrality indices are then also infinite, and cannot be calculated.

Thus, we dropped family Pucci from the set of actors, giving us a smaller network of $g - 1 = 15$ families, but now we have (for the purpose of demonstrating the calculations of closeness centralities) a connected graph. The actor centralities and centralization indices calculated for this smaller network are shown in Table 5.1 and are indexed with asterisks to distinguish them from indices calculated for the full set of actors. The actor closeness centralities are shown in Column 4, while the actor degree centralities for the smaller set of actors (sans family Pucci) are shown in Column 3. Once again family Medici is the most central actor, but several families are almost as central: Albizzi, Guadagni, Ridolfi, and Tornabuoni.

Note that family Strozzi, which had a rather large actor degree centrality index, has a relatively small actor closeness centrality index. Strozzi has apparently married into a moderately large number of other families, but is not particularly close to the other families; that is, there are many "steps" in the marital linkages from Strozzi to the others.

The closeness indices are much larger than the degree indices, and none of the families have small values. Families Acciaiuoli, Ginori, Lamberteschi, and Pazzi are still the least central. These indices also vary less than the degree indices (from 0.326 to 0.560, as opposed to 0.071 to 0.429 for degree centralities), indicating a much more uniform spread of closenesses. The closeness centralization index is $C_C^* = 0.322$, calculated for the smaller network, and the average closeness centrality and variance are $\overline{C}_C = 0.415$ and $S_C^2 = 0.0056$. This is a small variance, indicating once again the small range of the actor closeness centralities.

5.2.3 Betweenness Centrality

Interactions between two nonadjacent actors might depend on the other actors in the set of actors, especially the actors who lie on the paths between the two. These "other actors" potentially might have some control over the interactions between the two nonadjacent actors. Consider now whether a particular actor might be able to control interactions between pairs of other actors in the network. For example, if the geodesic between actors n_2 and n_3 is $n_2 n_1 n_4 n_3$ — that is, the shortest path between these actors has to go "through" two other actors, n_1 and n_4 — then we could say that the two actors contained in the geodesic might have control over the interaction between n_2 and n_3. Glance again at our star network in Figure 5.1, and note that the most central actor lies on all fifteen geodesics linking the other six actors. This "actor in the middle," the one *between* the others, has some control over paths in the graph. A look at the line network in Figure 5.1 shows that the actors in the middle of this graph might have control over some of the paths, while those at the edge might not. Or, one could state that the "actors in the middle" have more "interpersonal influence" on the others (see Freeman 1979, or Friedkin 1991).

The important idea here is that an actor is central if it lies between other actors on their geodesics, implying that to have a large "betweenness" centrality, the actor must be *between* many of the actors via their geodesics.

Several early centrality researchers recognized the strategic importance of locations on geodesics. Both Bavelas (1948) and Shaw (1954) suggested that actors located on many geodesics are indeed central to the network, while Shimbel (1953) and Cohn and Marriott (1958) noted that such central actors play important roles in the network. None of these researchers, however, were able to quantify this notion of betweenness. It took roughly twenty years, however, until Anthonisse (1971), and later Freeman (1977) and Pitts (1979), suggested that the the locations of actors on geodesics be examined.

Actor Betweenness Centrality. Let us simply quote from Shimbel (1953), reiterated by Pitts (1979), who stated the importance of geodesics and the actors they contain for measuring betweenness and network control:

> Suppose that in order for [actor] i to contact [actor] j, [actor] k must be used as an intermediate station. [Actor] k in such a network has a certain "responsibility" to [actors] i and j. If we count all of the minimum paths which pass through [actor] k, then we have a measure of the "stress" which [actor] k must undergo during the activity of the network. (page 507)

Here, actors who have sufficient stress also possess betweenness, according to this rather political view of network flows.

Specifically, one should first count the number of geodesics linking actors j and k (all these geodesics will be of the same length, $d(n_j, n_k)$), and then determine how many of these geodesics contain actor i, for all distinct indices i, j, k. Shimbel goes on to state that

> A vector giving this [count of minimum paths] for each [actor] of the network would give us a good idea of the stress conditions *throughout the system.* (page 507; emphasis is ours)

Shaw (1954) was the first to recognize that this stress was also betweenness, noting that, in the case of a communication relation where actors could not form new lines, central actors could refuse to pass along messages. Anthonisse (1971) and Freeman (1977) first quantified this idea.

We want to consider the probability that a "communication," or simply a path, from actor j to actor k takes a particular route. We assume that lines have equal weight, and that communications will travel along the shortest route (regardless of the actors along the route). Since we are just considering shortest paths, we assume that such a communication follows one of the geodesics. When there is more than one geodesic between j and k, all geodesics are equally likely to be used. Freeman estimates this probability as follows: Let g_{jk} be the number of geodesics linking the two actors. Then, if all these geodesics are equally likely to be chosen for the path, the probability of the communication using any one of them is simply $1/g_{jk}$. We also consider the probability that a distinct actor, i, is "involved" in the communication between the two actors. We let $g_{jk}(n_i)$ be the number of geodesics linking the two actors that contain actor i. Freeman then estimates this probability by $g_{jk}(n_i)/g_{jk}$, making the critical assumption that geodesics are equally likely to be chosen for this path. (We comment on this assumption later in the chapter.)

The actor betweenness index for n_i is simply the sum of these estimated probabilities over all pairs of actors not including the ith actor:

$$C_B(n_i) = \sum_{j<k} g_{jk}(n_i)/g_{jk} \qquad (5.11)$$

for i distinct from j and k. So, this index, which counts how "between" each of the actors is, is a sum of probabilities. It has a minimum of zero, attained when n_i falls on no geodesics. Its maximum is clearly $(g-1)(g-2)/2$, which is the number of pairs of actors not including n_i. The index reaches the maximum when the ith actor falls on all geodesics. Since the index's values depend on g, we standardize it just like the other actor centrality indices:

$$C'_B(n_i) = C_B(n_i)/[(g-1)(g-2)/2]. \qquad (5.12)$$

Standardized in this way, it now takes on values between 0 and 1, and can easily be compared to the other actor indices, as well as across networks and relations. Unlike the closeness indices, these betweenness indices $\{C'_B(n_i)\}$ can be computed even if the graph is not connected. This is certainly an advantage. As with our other actor indices, algorithms for first finding the geodesics in a graph, and then counting how many of them contain each of the actors, are available, and are implemented in network computer programs such as *UCINET*.

The quantities summed on the right-hand side of equation (5.11) are discussed in more detail in Freeman (1980a). Specifically, if we sum the $g_{jk}(n_i)/g_{jk}$ estimated probabilities over k, we obtain measures of the pair-dependency of actor j on actor i. These values, which can also be viewed as indices of how much "gatekeeping" n_i does for n_j, are crucial components of both the $\{C_B(n_j)\}$ and the $\{C_C(n_j)\}$. Gatekeeping of one actor for another is simply the act of being on geodesics from the latter actor to all other actors, regardless of where the geodesics are going. Actors on whom others are "locally dependent" are central in the network. One can measure the level of gatekeeping for every pair of actors in the network, focusing on how much gatekeeping the second actor does for the first.

Returning again to the graphs of Figure 5.1, we find that for the star graph, $C'_B(n_1) = 1.0$, while $C'_B(n_2) = \cdots = C'_B(n_7) = 0$. This is an idealized situation, since only actor 1 lies on any of the geodesics. The actor betweenness indices in the circle graph are all equal to 0.2, and for the line graph, vary from $C'_B(n_1) = 0.6$ to $C'_B(n_6) = C'_B(n_7) = 0$. In this last graph, actors n_2 and n_3 are almost as central as n_1, since $C'_B(n_2) = C'_B(n_3) = 0.533$.

Group Betweenness Centralization. Group centralization indicies based on betweenness allow a researcher to compare different networks with respect to the heterogeneity of the betweenness of the members of the networks. We first report Freeman's (1979) index for quantifying the overall level of betweenness in the set of actors, which summarizes the actor betweenness indices given in equation (5.11).

Freeman's group betweenness centralization index has numerator $\sum_{i=1}^{g}[C_B(n^*) - C_B(n_i)]$, where $C_B(n^*)$ is the largest realized actor betweenness index for the set of actors. The reason for using the nonstandardized indices rather than the standardized ones (see equation (5.12)) will follow. Freeman shows that the maximum possible value for this sum is $(g-1)^2(g-2)/2$, so that the index of group betweenness is

$$C_B = \frac{2\sum_{i=1}^{g}[C_B(n^*) - C_B(n_i)]}{[(g-1)^2(g-2)]}. \tag{5.13}$$

Freeman (1979) shows that this simplifies to the index given in Freeman (1977):

$$C_B = \frac{\sum_{i=1}^{g}[C'_B(n^*) - C'_B(n_i)]}{(g-1)}; \tag{5.14}$$

that is, the calculation can also be made equivalently with the standardized indices. Freeman (1977) also demonstrates that the index reaches its maximum value (unity) for the star graph. Its minimum value (0) occurs when all actors have exactly the same actor betweenness index — that is, in a network in which all actors are equal in betweenness. The line graph of Figure 5.1 has $C_B = 0.311$.

There are additional group-level betweenness indices, for example the variance of the actor-level betweenness indices. Such centralization indices provide additional summaries of the heterogeneity or variability of betweenness in the entire set of actors.

The Example, Once Again. Actor betweenness centralities are given for Padgett's Florentine families and the marriage relation in Table 5.1, Columns 2 and 5. The second column gives the betweenness centralities calculated for the network consisting of all actors, and the fifth column, for the network without the Pucci family. Note how many actors have 0 indices — families Acciaiuoli, Ginori, Lamberteschi, Pazzi, and of course Pucci — the same actors that had the smallest actor closeness centrality indices. The betweenness indices allow the Medici family, and, to a lesser extent, the Guadagni family, to stand out, just as with the actor degree centralities.

Clearly, families Medici and (perhaps) Guadagni are the most central families in this set of actors on this marital relation. The betweenness centralization index is $C_B = 0.437$, larger than the other centralization indices, reflecting the fact that the Medici family is much more central than any of the other families.

Note how these betweenness indices compare to the other two actor centrality indices. Some actors with moderately large closeness and degree scores have small betweenness indices — families Barbadori and especially Tornabuoni. Family Strozzi, which has a large degree index, has a small betweenness index. Such differences indicate that the betweenness indices can be quite different measures of actor centrality than degree- and closeness-based indices.

5.2.4 ⊗Information Centrality

Of all these indices, Freeman's centrality measure based on betweenness of actors on geodesics has found the most use, because of its generality. But, this index assumes that all geodesics are equally likely when estimating the critical probability that an actor falls on a particular geodesic. That is, if there are g_{ij} geodesics between actors i and j, then the probability that a particular geodesic is "chosen" for the "flow of information" between these two actors is simply $1/g_{ij}$. While this is a justifiable assumption for some purposes, it may not be appropriate here.

Suppose we focus on the actors "contained" in these geodesics. Freeman ignores the fact that if some actors on the geodesics have large degrees, then the geodesics containing these expansive actors are more likely to be used as shortest paths than other geodesics. That is, if an actor has a degree of, say, 10, and this actor is on a geodesic, then this actor is more likely than actors with smaller degrees to be on other geodesics, simply because of its expansiveness. Freeman's assumption is reasonable only if all actors have equal degrees. For such regular graphs, it is not unreasonable to assume that all geodesics between a pair of nodes are equally likely to be "used" for a path. Relaxing this assumption is difficult, and requires a more sophisticated statistical model that allows for unequal probabilities.

A second, more important generalization can also be considered. Freeman, in considering betweenness counts, focuses only on geodesics. That is, paths with distances greater than the minimum path length attained by the geodesics are ignored. Substantively, this might not be realistic. For example, if we consider communication relations, there may be good reasons for actors to choose paths for their communications that are longer than the geodesics. We quote:

> It is quite possible that information will take a more circuitous route either by random communication or [by being] channeled through many intermediaries in order to "hide" or "shield" information. (Stephenson and Zelen 1989, page 3)

So, it may make sense to generalize the notion of betweenness centrality so all paths between actors, with weights depending on their lengths, are considered when calculating betweenness counts.

The index of centrality developed by Stephenson and Zelen (1989) does exactly this. One considers the combined path from one actor to another, by taking all paths, including the geodesics, and assigning them weights. A weighted function of this combined path is then calculated, using as weights the inverses of the lengths of the paths being combined. The weights assigned to the paths making up the combined path are determined so that the "information" in the combined path is maximized.

Geodesics are usually given weights of unity, while paths with lengths longer than the geodesic length receive smaller weights based on the *information* that they contain. The information of a path is defined quite simply as the inverse of its length.

Mackenzie (1966b) was the first to propose the use of *information theory* for the study of centrality, particularly in commnication networks. He defined a "Total Participation Index" for actors in a network, but his rather mathematical presentation has prevented wider adoption of the idea. Bolland's (1988) index of continuing flow also considers all paths originating with each actor, but does not consider betweenness counts.

The concept of information is quite old, and has a rich tradition in statistics (Shannon and Weaver 1949; Khinchin 1957; Kullback 1959; Gokhale and Kullback 1978; see also Coleman 1964; Theil 1967; and Allison 1978, for applications in economics and sociology). It is used extensively in estimation theory and categorical data analysis. Information is usually defined as the inverse of the variance of an estimator. If an estimator has a small variance, it has large information, and thus is considered "good." The opposite is also true: poor estimators with large variances have little information. We can apply this approach to centrality by extending betweenness on geodesics to all possible paths, weighting according to the information that these paths contain. The betweenness counts are then generalized to reflect the information contained in all paths.

Stephenson and Zelen (1989) give a nice discussion of the use of information in statistical estimation as applied to the paths between nodes in a graph. In brief, the length of any path is directly related to the variance of transmitting a signal from one node to another; thus, the information contained in this path is the reciprocal of this variance. Thus, any path (and hence, each and every combined path) has an "information content." Lastly, the information of a node is the harmonic average of the information for the combined paths from the node to all other nodes.

Actor Information Centrality. This version of centrality focuses on the information contained in all paths originating with a specific actor. The information of an actor averages the information in these paths, which, in turn, is inversely related to the variance in the transmission of a signal from one actor to another.

To calculate information centrality indices, Stephenson and Zelen recognized that the information contained in an incidence matrix (see Chapter 4), which codes the nodes and the links between them, is exactly the same as the information contained in the data vector and incidence matrix for an incomplete block design with two treatments per block (see Cochran and Cox 1957). This exact equality allowed Stephenson and Zelen to adopt a statistical model, common in the statistical design of experiments (see St. John and Draper 1975; Box, Hunter, and Hunter 1978; Silvey 1981), designed for such incomplete block designs. The model focuses on all the "signals" flowing between all pairs of nodes (or pairs of rows of the incidence matrix). One estimates the strengths of these signals, and calculates their variances. If V_{jk} is the variance of the estimate of the signal for the path linking nodes n_j and n_k, then the information associated with this path is simply $1/V_{jk}$. The information for an actor is a function of all the information for paths flowing out from the actor. The chosen function is the harmonic average. We refer the reader to the appendix of Stephenson and Zelen (1989) for more mathematical details.

To apply this idea to graphs, the actor information indices are functions of a simple $g \times g$ matrix. We give the most general formulation of the index, which assumes that the relation is nondirectional, but not necessarily dichotomous. A crucial component of the formula is the sum of the strengths or values for the lines incident with a node. This sum is simply a row total (or column total) of the sociomatrix. The sum is the degree of a node when the measured relation is dichotomous, or the sum of the strengths of all ties incident to a node when the relation is valued.

The calculation begins as follows. One first creates a $g \times g$ matrix $\mathbf{A}$, which has diagonal elements

$$a_{ii} = 1 + \text{ sum of values for all lines incident to } n_i \tag{5.15}$$

and off-diagonal elements

$$a_{ij} = \begin{cases} 1 & \text{if nodes } n_i \text{ and } n_j \text{ are not adjacent} \\ 1 - x_{ij} & \text{if nodes } n_i \text{ and } n_j \text{ are adjacent.} \end{cases} \tag{5.16}$$

As usual, x_{ij} is the value of the tie between actors i and j, so that the elements of $\mathbf{A}$ are easily calculated from the sociomatrix. One next calculates the inverse of $\mathbf{A}$: $\mathbf{C} = \mathbf{A}^{-1}$, which has elements $\{c_{ij}\}$.

We should note that not every $\mathbf{A}$ matrix can be inverted. In fact, if the sociomatrix has one (or more) rows (and hence columns) full of zeros, the corresponding $\mathbf{C}$ is not defined. In this instance, actor information centralities cannot be computed. We recommend that the actors who are isolates be dropped from the set of actors, and indices calculated just for the non-isolates.

To get the information indices, one needs two intermediate quantities. These are sums of elements of $\mathbf{C}$: $T = \sum_{i=1}^{g} c_{ii}$ and $R = \sum_{j=1}^{g} c_{ij}$. T is simply the trace or the sum of the diagonal entries of the matrix, while R is any one of the row sums (all the row sums are equal). With these two quantities, and the elements of $\mathbf{C}$, one lastly calculates

$$C_I(n_i) = \frac{1}{c_{ii} + (T - 2R)/g} \tag{5.17}$$

as the information centrality index for actor i.

This index measures how much information is contained in the paths that originate (and end) at a specific actor. The index has a minimum value of 0, but no maximum value; indeed, if $T = 2R$, and $c_{ii} = 0$, the index would equal ∞. Stephenson and Zelen (1989) recommend that one use relative information indices, obtained by dividing each index $C_I(n_i)$ by the total of all indices:

$$C_I'(n_i) = \frac{C_I(n_i)}{\sum_i C_I(n_i)}. \tag{5.18}$$

The relative information indices, $\{C_I'(n_i)\}$, are bounded by 0 and 1, and sum to unity. These indices can be interpreted as the proportion of total "information" flow in a graph controlled by an individual actor. The constraint that the indices sum to unity is unique to this index, and makes comparisons with the other actor-level centrality indices difficult. Necessary calculations are not complicated, and involve manipulations of the sociomatrix, and then a single matrix inversion. One can "program" them with *SAS PROC IML* or *GAUSS*.

Return once again to the graphs of Figure 5.1. We find that for the star graph, $C_I'(n_1) = 0.2340$, while $C_I'(n_2) = \cdots = C_B'(n_7) = 0.1277$. Notice that even though only node n_1 lies on any of the geodesics, the information centralities for the other six nodes are not zero. The actor information indices in the circle graph are all equal to 0.1429, and for the line graph, vary from $C_I'(n_1) = 0.1822$ to $C_I'(n_6) = C_I'(n_7) = 0.1041$. In this last graph, nodes n_2 and n_3 are almost as central as n_1, since $C_I'(n_2) = C_I'(n_3) = 0.1682$. Remember that the actor information centralities are normed differently from the other actor centralities — they must sum to unity, so that if one actor has a large index, the other actors must have smaller indices.

Use of this information index has been limited. Stephenson (1989) and Stephenson and Zelen (1989) apply this methodology to networks of baboons, while Stephenson (1989) and Frey (1989) use this index (and others) to study a network of forty AIDS patients, linked by sexual contact (Auerbach, Darrow, Jaffe, and Curran 1984; Klovdahl 1985; see also Laumann, Gagnon, Michaels, Michael, and Coleman 1989; Morris 1989, 1990). Marsden (1990b) has used information indices in a study of the effect of random sampling on estimation of the parameters of the network effects or social process model (see Erbring and Young 1979; Doreian 1981; Friedkin 1986, 1990; Marsden 1990a; and Burt 1987).

Group Information Centralization. The summary group-level information index proposed by Stephenson and Zelen is simply the average information across actors:

$$\overline{C}_I = \sum_i C_I(n_i). \tag{5.19}$$

This index has limits that depend on g, unfortunately, and so is difficult to compare across networks. As we have mentioned throughout this chapter, averages are not centralization indices. A real group-level information centralization index is the variance of the actor information indices:

$$S_I^2 = \left[\sum_{i=1}^{g}(C_I'(n_i) - \overline{C}_I)^2\right] / g. \tag{5.20}$$

One could also apply Freeman's (1979) general index (5.1) to information indices, although (to our knowledge) no one has calculated the denominator (the maximum possible sum of differences between the observed indices and the largest attained index) for a Freeman information centralization index.

For the graphs of Figure 5.1, the variances are 0.001614, 0.000986, and 0.0, for the star, line, and circle graphs, respectively. Thus, the star graph is most heterogeneous, and the circle, the least.

As mentioned, this information actor-level index of centrality is the only index (that we are aware of) that can be applied to valued relations. Further, as we have discussed, it generalizes Freeman's widely used index of betweenness, since it considers all paths, not just geodesics. We comment further on the differences among all the indices discussed here at the end of the chapter. Further research and application should demonstrate the usefulness of the actor information centrality index (5.18).

Last Look at Padgett's Florentine Families. As we have noted, family Pucci is not married to any other families; it is an isolate, and consequently, the actor information centralities cannot be calculated because the **C** matrix cannot be inverted. Thus, we dropped this actor from the set of actors, and calculated actor information centralities for the other fifteen actors. These indices are shown in Column 6 of Table 5.1.

It is difficult to compare these indices to the others, since only the information centralities are constrained to sum to unity. They must be between 0 and 1, just like all the other types of centrality, but are forced to be smaller in magnitude because of this constraint.

The Medici family is still the most central family, although the Guadagni, Tornabuoni, and Ridolfi families have indices not much smaller than that for Medici. These four families consistently have the largest actor centrality indices. The Pazzi, Ginori, Lamberteschi, and Acciaiuoli families are the least central families; in fact, the ordering of the actors with respect to information centrality is almost identical to that for betweenness centrality. The main difference between the two sets of centralities is the range of values — the range is much smaller for information. The variance of the actor information centralities is $S_I^2 = 0.000297$, quite small, reflecting the small range of the values due to the unity summation constraint.

We now turn to indices that can be applied to social network data consisting of directional relations.

5.3 Directional Relations

In the previous section, we discussed nondirectional relations, and introduced four actor-level indices for centrality (and associated centralization indices). These indices are:

(i) Degree — equation (5.3)
(ii) Closeness — equation (5.8)
(iii) Betweenness — equation (5.12)
(iv) Information — equation (5.18)

We now discuss how these, and other kinds of indices (specifically, those designed to measure prestige), can be calculated for directional relations.

Suppose that we have a single set of actors, and a single, dichotomous directional relation. With such data, we can distinguish between "choices made" and "choices received." An example of such data that we will be analyzing in this section can be found in Chapter 2; specifically, the countries trade network data which show import and export of basic manufactured goods among a collection of g=24 countries. These data are discussed in some detail in Chapter 2, and will be examined at length in Chapters 9–12. Clearly for these data, imports are substantively different from exports, and it is interesting to study which actors are important importers and which are significant exporters. To identify these important actors on this relation, we will examine both aspects of prominence: centrality and prestige.

As mentioned at the beginning of the chapter, centrality indices for directional relations generally focus on choices made, while prestige indices generally examine choices received, both direct and indirect. We will discuss how to calculate centrality indices for directional relations here, but the emphasis in this section will be prestige, and in particular, three types of prestige indices.

We first discuss how the four centrality indices, degree, closeness, betweenness, and information, can be extended to directional relations. For this extension, we examine actors from the perspective of the choices or nominations that are made. Two of the centrality indices are easily applied to directional relations (degree and closeness indices), while the other two (betweenness and information), because of their reliance on nondirected paths and geodesics, are not.

One can also examine the choices received by actors. This allows us to study which actors in the set of actors are prestigious. We will present and discuss three types of prestige indices.

5.3.1 Centrality

To extend to directional relations the centrality indices based on degree, closeness, betweenness, and information, and the group-level indices which aggregate the actor-level indices (equations (5.5), (5.9), (5.13), (5.19)), we must consider how each is computed and how the network properties that are crucial for each are defined for directional relations.

Degree. An actor index for degree centrality can easily be calculated for directional relations. Such indices are meaningful if no restrictions, as in a fixed choice design, are placed on the choices made by the actors. Since centrality indices focus on the choices made, we take the outdegree of each actor, rather than the degree (which we used for nondirectional relations): $C'_D(n_i) = x_{i+}/(g-1)$. A group-level index of degree centralization can be calculated as suggested in equation (5.4). The denominator of this index when the measured relation is directional can be calculated to be $(g-1)^2$. These actor and group-level indices have exactly the same properties as degree indices for nondirectional relations.

Closeness. An actor index for closeness centrality can also be calculated for directional relations. Specifically, we define the distances between any two actors, as discussed in Chapter 4, as the length of the geodesic(s) from n_i to n_j. With a directed graph, the geodesic(s) from n_i to n_j may not be the same as the one(s) from n_j to n_i, so that $d(n_i, n_j)$, the length of the geodesic(s), may not equal $d(n_j, n_i)$. These $\{d(n_i, n_j)\}$ are elements of a $g \times g$ distance matrix.

Actor-level centrality indices for closeness are calculated by taking the sum of row i of the distance matrix to obtain the total distance n_i is from all the other actors, and then dividing by $g - 1$ (the minimum possible total distance). The reciprocal of this ratio gives us an actor-level index for closeness. The formula is exactly the same as for nondirectional relations. Specifically, the actor-level closeness centrality index for directional relations is

$$C'_C(n_i) = (g-1) / \left[\sum_{j=1}^{g} d(n_i, n_j) \right]. \tag{5.21}$$

This index has exactly the same properties as discussed following equation (5.8). A group-level closeness index based on Freeman's general formula (5.1) can be obtained using the standardized indices; however, to our knowledge, no one has calculated the denominator of this index when the measured relation is directional.

One problem with this actor-level centrality index based on closeness is that it is not defined unless the digraph is strongly connected (that is, if there is a directed path from i to j, for all actors i and j); otherwise, some of the $\{d(n_i, n_j)\}$ will be ∞, and equation (5.21) will be undefined. The same problem arises with graphs based on nondirectional relations, as discussed earlier. One remedy to this problem is to consider only those actors that i can reach, ignoring those that are unreachable from i.

This simple index, $C'_C(n_i)$, can be generalized by considering the *influence range* of n_i as the set of actors who are reachable from n_i. This set contains all actors who are reachable from i in a finite number of steps. This notion is common to graph theory, and is related to an idea first used by Lin (1976) to describe the set of actors reachable *to* n_i (see below). We define J_i as the number of actors in the influence range of actor i. This count J_i equals the number of actors who are reachable from n_i. Note that this idea can also be applied to nondirectional relations.

An "improved" actor-level centrality closeness index considers how proximate n_i is to the actors in its influence range. We define closeness now by just focusing on distances from actor i *to* the actors in its influence range. We consider the average distance these actors are from n_i. This average distance, $\sum d(n_i, n_j)/J_i$, where the sum is taken over all actors j in the influence range of actor i, is a refined measure of closeness. Note that this sum ignores actors who are not reachable from n_i, so that unlike the first closeness centrality measures, it is defined even if the graph is not strongly connected. We can define

$$C_C^*(n_i) = \frac{J_i/(g-1)}{\sum d(n_i, n_j)/J_i}, \tag{5.22}$$

where the summation again is just over those actors in the influence range of n_i.

One can see that this index is a ratio of the fraction of the actors in the group who are reachable ($J_i/(g-1)$), to the average distance that these actors are from the actor ($\sum d(n_i, n_j)/J_i$). This index is quite similar to an index for prestige that we discuss in the next section.

Other. The other two centrality indices for nondirectional relations, based on betweenness and information, were derived using theory and algorithms designed specifically for nondirectional relations. Gould (1987) has extended the betweenness index to directional relations, by considering geodesics between any two actors. Gould shows that the algorithm to find actor betweenness indices for nondirectional relations can be applied to directional relations, since the basic algorithm automatically uses ordered (rather than unordered) pairs of actors.

The $\{C_B(n_i)\}$ indices defined in equation (5.11) are thus calculated correctly for both directional and nondirectional relations; however, the $\{C_B'(n_i)\}$ indices defined in equation (5.12) must be multiplied by 2. The maximum value for the index is $(g-1)(g-2)$, so that these standardized scores must be multiplied by a factor of two to be correct (since the maximum for nondirectional relations is $(g-1)(g-2)/2$). We note that Gould's (1987) extension is based on the assumption that a directional relation can be turned into a nondirectional relation by coding all mutual dyads as lines and ignoring asymmetric dyads. Thus, there is a line in the derived undirected graph between two actors if and only if both actors choose each other in the original digraph.

For an information index, we could consider directed geodesics and longer directed paths between actors. All these paths will be directed, given the nature of the data. However, we do not know how to generalize Stephenson and Zelen's (1989) theory for information indices to directional relations.

Thus, we recommend the use of just two centrality indices, $C'_D(n_i)$, and $C'_C(n_i)$ or $C^*_C(n_i)$, for directed graphs. In our later discussions of the countries trade network data, we calculate not only actor prestige indices, but also these two actor centrality indices. Since choices received are usually more interesting than those made, neither of these centrality indices is as useful as the measures of prestige that we discuss below. If the relation allows one to distinguish between choices made and choices received, then the latter, along with prestige indices calculated from them, can give important insights into social structure, as we will demonstrate with our example.

5.3.2 Prestige

With directional relations, choices received are quite interesting to a network analyst. Thus, measures of centrality may not be of as much concern as measures of prestige. We now discuss several prestige measures, which we will illustrate on the countries trade network data. We recommend that both centrality and prestige measures be computed for directional relations, since they do attempt to measure different structural properties.

There has been little research on group-level prestige indices. However, such measures would certainly be welcome and interesting, since they could quantify prestige heterogeneity (and possibly hierarchization or network stratification).

We also note that there has been little work done on applications of prestige measures to actual digraphs. For example, it is not known which digraphs have maximal group-level prestige indices. More research on such important issues is clearly needed.

Degree Prestige. The simplest actor-level measure of prestige is the indegree of each actor, which we denoted by $d_I(n_i)$ in Chapter 4. The idea is that actors who are prestigious tend to receive many nominations or choices (see Alexander 1963). So, we define

$$P_D(n_i) = d_I(n_i) = x_{+i}. \tag{5.23}$$

As with the comparable indices based on outdegrees, equation (5.23) is dependent upon the group size g; thus, the standardization

$$P'_D(n_i) = \frac{x_{+i}}{g-1} \tag{5.24}$$

gives us the proportion of actors who choose actor i, which is sometimes called a "relative indegree." The larger this index is, the more prestigious is the actor. Maximum prestige occurs when $P'_D(n_i) = 1$; that is, when actor i is chosen by all other actors. This index is quite simple to compute, and is usually provided as output from network analysis computer packages, such as *UCINET*.

Proximity Prestige. This simple index, $P'_D(n_i)$, counts only actors who are adjacent to actor i. One can generalize this index by defining the *influence domain* of actor i as the set of actors who are both directly *and* indirectly linked to actor j. Such actors are reachable to i, or alternatively, are those from whom i is reachable. Reachability is discussed in Chapter 4. Thus, the influence domain consists of all actors whose entries in the ith column of the distance matrix or the reachability matrix are finite. This notion was first used by Lin (1976). We define I_i as the number of actors in the influence domain of actor i. This count I_i equals the number of actors who can reach actor i. We use the idea of an influence domain in the next prestige index.

A second actor-level index of prestige considers how proximate n_i is to the actors in its influence domain. We define *proximity* as closeness that focuses on distances *to* rather than *from* each actor. In other words, what matters now is how close all the actors are to n_i. Since the relation is directional, such closeness will no doubt differ from the closeness that n_i is to the other actors. As stressed in Chapter 4, with digraphs, distance *to* a node can be quite different from distance *from*.

We consider the average distance these actors are to n_i. This average distance, $\sum d(n_j, n_i)/I_i$, where the sum is taken over all actors j in the influence domain of actor i, is a crude measure of proximity. Note that it ignores actors who cannot reach n_i, so that unlike our closeness and information centrality measures, it is defined even if the network is not connected (when some actors are not reachable from other actors). This index depends on the size of the group, and is difficult to compare across networks.

But, we can look at the ratio of the proportion of actors who can reach i to the average distance these actors are from i. Thus, a better measure of proximity takes the average distance, standardizes it, and then takes reciprocals. From a suggestion by Lin (1976), we define

$$P_P(n_i) = \frac{I_i/(g-1)}{\sum d(n_j, n_i)/I_i}, \tag{5.25}$$

where the summation again is just over those actors in the influence domain of n_i. One can easily see that this index is a ratio of the fraction of the actors in the set of actors who can reach an actor $(I_i/(g-1))$ to the average distance that these actors are to the actor $(\sum d(n_j, n_i)/I_i)$. As actors who can reach i become closer, on average, then the ratio becomes larger.

This ratio index, based on the average distance actors in an influence domain are to i, has the same properties as the centrality index for actor closeness (see equation (5.7)). The index weights prestige according to closeness or proximity. Note that if all actors are adjacent to n_i, then all the $d(n_j, n_i) = 1$, $I_i = g - 1$, and the average standardized distance is simply $1/(g-1)$. This gives $P_P(n_i) = 1$, the maximum value of the prestige actor proximity index. If an actor is unreachable, then $I_i = 0$, and $P_P(n_i) = 0$. Thus, the limits of this index are 0 and 1, and the magnitude of the index reflects how proximate an actor is from the set of actors as a whole. Similar indices were proposed by Mackenzie (1966a) and Arney (1973).

One could easily take the variance of the $\{P_P(n_i)\}$ to obtain a group-level prestige index based on proximity. In addition, the average of the actor-level indices can be used to summarize the set of actors as a whole. The average is proportional to the average of the reciprocals of the average distances to the actors. These two group-level indices are

$$\overline{P}_P = \sum_{i=1}^{g} \frac{P_P(n_i)}{g} \tag{5.26}$$

and

$$S_P^2 = \sum_{i=1}^{g} \frac{(P_P(n_i) - \overline{P}_P)^2}{g}. \tag{5.27}$$

The average will be between 0 and 1. It equals 1 in a complete directed graph, and 0 in an empty directed graph. The variance will be positive, and measures how much heterogeneity is present in the set of actors, with respect to proximity.

Another index based on proximity was proposed by Harary (1959c), who considered not only the prestige of each actor (which he referred to as *status*, defined as the total distance of actor i to all other actors) but also the *contrastatus* of an actor (defined as the total distance to

n_i of all other actors, not just those in the influence domain). In our terminology, these quantities are $\sum_j d(n_i, n_j)$ (on which the closeness indices for centrality are based) and the sum $\sum_j d(n_j, n_i)$ (which, as just mentioned, is key to the proximity indices for prestige). Using these terms, status (for Harary) is synonymous with actor-level closeness centrality, while contrastatus is similar to actor-level proximity prestige. Harary defines the *net status* of an actor as the difference between these two sums. The idea of constructing an index for prestige that is a difference of two simpler indices was first suggested by Zeleny (1940a, 1940b, 1941, 1960). Zeleny's *sociation index* is the difference of the average of the overall "intensity" of ties in the group (measured by the density of ties in the sociomatrix if the relation is dichotomous) and the number of choices made by actor *i*. Refinements of this idea generate both a social status index and a social adjustment index, measured at the level of the individual actor.

These actor and group-level prestige indices based on proximity or graph distances to each actor can be useful. Actors are judged to be prestigious based on how close or proximate the other actors in the set of actors are to them. However, one should simultaneously consider the prestige of the actors that are proximate to the actor under study. If many prestigious actors "choose" an actor, the actor should be judged more prestigious than an actor who is "chosen" only by peripheral actors. Thus, one should "weight" the distances used in the proximity indices by measures of the prestige of the actors in the influence domain. Seeley (1949) was the first to realize this; using children and friendship as the network actors and relation under study, he states:

> How should we represent each ... child's popularity, as shown by the choices, weighting those choices according to the "popularity" of the source-of-choice child? (page 234)

To answer this question, we turn to yet another class of prestige indices.

⊗**Status or Rank Prestige.** Let us now consider a method to measure the prestige of the actors in a set of actors based on their status or rank within the set of actors. We have described several prestige measures that look at indegrees and distance, but none of these reflects the prominence of the individual actors who are doing the "choosing." We need to combine the numbers of direct "choices" or distances to a specific actor, with the status or rank of the actors involved. If one's influence domain is full of prestigious actors, one's prestige should also be high. If, however, an actor's domain contains only peripheral, or marginally important, actors, then the rank of this actor should be low.

To quantify this idea requires some sophisticated mathematics. An actor's rank depends on the ranks of those who do the choosing; but note that the ranks of those who are choosing depend on the ranks of the actors who choose them, and so on. As Seeley (1949) goes on to state:

> ... both "source" and "target" children are the same children, [so] we seem to be, and indeed we are, involved in an "infinite regress": [*i*'s status] is a function of the [status] of those who choose him; and their [status] is a function of those who choose them, and so *ad infinitum*. (pages 234–235)

Seeley (1949) was the first to propose a solution to this problem. His idea and solution was also discussed by Katz (1953), Hubbell (1965), Taylor (1969), Bonacich (1972a, 1972b, 1987), Coleman (1973), Burt (1982), Mizruchi, Mariolis, Schwartz, and Mintz (1986), and Tam (1989). We discuss this line of research here. We first want to note that researchers usually refer to the property under study as "status" (or even "power"); however, because of the use of this term in the relational analysis of social networks using role algebras (see Part IV), we have chosen to use the term "rank" as a synonym for "status." Thus, actors will be said to be prestigious with respect to their rank within the set of actors if they have large values on the measures described below.

The simplest way to present the solution to this "infinite regress" situation is first to define $P_R(n_i)$ as the actor-level rank prestige measure for actor i within the set of actors. The theory behind prestige as rank states that an actor's rank is a function of the ranks of the actors who choose the actor. Thus, if we take the ith column of the sociomatrix, which contains entries indicating which actors choose n_i, we can multiply these entries by the ranks of the other actors in the set of actors to obtain a linear combination measuring the rank of n_i:

$$P_R(n_i) = x_{1i}P_R(n_1) + x_{2i}P_R(n_2) + \cdots + x_{gi}P_R(n_g). \qquad (5.28)$$

For example, if n_2 is chosen by n_5 and n_7, so that $x_{52} = x_{72} = 1$ and all the other $g - 2$ entries in the second row of the sociomatrix are 0, then the rank index for this actor is defined as $P_R(n_2) = P_R(n_5) + P_R(n_7)$. In this example, if actors n_5 and n_7 are of high rank, so will be n_2. An actor's rank increases if the actor receives choices from high-ranking actors.

Thus, mathematically, we have g equations (5.28), all of which depend on all the indices themselves, the $\{P_R(n_i)\}$. So, we have a system of g linear equations with g unknowns. If we take the entire sociomatrix, $\mathbf{X}$, and put the set of rank indices into a vector $\boldsymbol{p} = (P_R(n_1), P_R(n_2), \ldots, P_R(n_g))'$, we can easily write this system of equations as

$$\boldsymbol{p} = \mathbf{X}'\boldsymbol{p}. \tag{5.29}$$

Or, rearranging terms, we obtain $(\mathbf{I} - \mathbf{X}')\boldsymbol{p} = \mathbf{0}$, where $\mathbf{I}$ is the identity matrix of dimension g, and $\boldsymbol{p}$ and $\mathbf{0}$ are vectors of length g.

This equation is identical to a characteristic equation (used to find the eigensystem of a matrix), in which $\boldsymbol{p}$ is an eigenvector of $\mathbf{X}'$ corresponding to an eigenvalue of 1. One solution to this system is to force $\mathbf{X}'$ to have such an eigenvalue. Thus, to solve this equation, one must put some constraints on either $\mathbf{X}'$, or on the indices themselves; otherwise, as first noted by Katz (1953), equation (5.29) has no finite solution. In fact, many authors, as we will note shortly, have worked on this problem, and all their solutions can be categorized based on the exact constraints that they place on the sociomatrix or on the system (5.29) itself.

Katz (1953) recommends that one first standardize the sociomatrix to have column sums of unity. The effect of this standardization on the system (5.29) is that the system becomes a familiar matrix characteristic equation, with a well-known solution. We also recommend Katz's normalization. Specifically, one finds the eigenvector associated with the largest eigenvalue of the standardized $\mathbf{X}'$. The first eigenvalue of the standardized $\mathbf{X}'$ will be unity (due to the constraint that the sociomatrix have unity column sums), and the eigenvector associated with this eigenvalue will be the vector of rank indices, $\boldsymbol{p}$.

As mentioned, the largest eigenvalue will be unity (if not, one has made a computation error). Call this eigenvector associated with this eigenvalue $\boldsymbol{p}_1$. Then, the elements of this vector are the actor rank prestige indices:

$$\boldsymbol{p}_1 = (P_R(n_1), P_R(n_2), \ldots, P_R(n_g))'.$$

Large rank prestige indices imply that an actor is chosen either by a few other actors who have large rank prestige, or by many others with low to moderate rank prestige. Remember that an actor's rank is a weighted sum of the ranks of those who choose the actor.

There are refinements of this normalization which we now discuss; however, we should note that such refinements are unnecessarily complicated. Katz's simple standardization discussed above, and the extracted eigenvector, are easy to interpret; more intricate refinements give no additional explanatory information. Katz (1953) also proposed that one introduce an "attenuation parameter" a to adjust for the lower "effectiveness" of longer paths in a network. He begins with the matrix $a\mathbf{X} + a^2\mathbf{X}^2 + \ldots + a^k\mathbf{X}^k + \ldots$, which is like an "attenuated number of paths between any two nodes" matrix. The system (5.29) is then modified by considering the column sums of this matrix (as we discuss below); unfortunately, the parameter a is unknown, and must be estimated (actually guessed) for a given sociomatrix.

To solve Katz's modification of the system, we must find a vector $\boldsymbol{p}$ that solves the new system of equations (which arises from the matrix sum mentioned above)

$$\{[(1/a)\mathbf{I} - \mathbf{X}']\boldsymbol{p}\} = \boldsymbol{x}, \qquad (5.30)$$

where $\boldsymbol{x}$ is the vector of indegrees of the unstandardized $\mathbf{X}$. The difference between this modification and the original system (5.29) is the presence of the parameter a, and the fact that the system now is equated to the indegrees, rather than the zero vector. Katz recommends that the reciprocal of the attenuation parameter should be between the largest eigenvalue of the unstandardized $\mathbf{X}$, and twice this largest eigenvalue. That is, if we define λ_1 as this largest eigenvalue, then $\lambda_1 < (1/a) < 2\lambda_1$. It clearly is advantageous from a computing standpoint to choose $(1/a)$ to be equal to an integer. Given such an a and $\mathbf{X}$, a vector of rank indices can easily be computed; one need only solve the equations of the system (5.30). We refer the reader to Katz (1953) for details and an example.

Taylor (1969) reviews Katz (1953) and Harary (1959c), and concludes that one not only needs to standardize the sociomatrix to have column sums of unity, but also to have row sums of unity, thereby adjusting not only for status but also for contrastatus, as does Harary. Taylor's combined measure is derived from an eigenvector of a matrix that has both adjustments (but not the eigenvectors associated with the eigenvalues of unity, which these matrices are forced to have because of the standardizations). Since this index considers both distance to and distance from an actor, as well as the rank of an actor, it can be viewed as a combination of rank, closeness, and proximity. It should be clear that there is a variety of ways to modify systems such as (5.29).

Hubbell (1965) and Bonacich (1972a, 1972b, 1987) proposed methods for identifying cohesive subgroups of actors (see Chapter 7), and by so doing, generalized Seeley's (1949) prestige measure further. Specifically, Hubbell, in searching for an "input-output" model for "clique" detection, derives a "status score" for each actor by taking Seeley's (1949) basic equation (5.28) and adding a constant for each actor. This constant is labeled the "exogenous contribution" of each actor to its own prestige. This assumption yields a matrix equation, which, with suitable constraints on the entries of the sociomatrix (such as unity column sums), can be solved for the vector of indices. Bonacich (1972b) suggests that the prestige vector be normed by multiplying it by a single parameter (with the best choice being the largest eigenvalue). With this normalization, the vector of indices is exactly the eigenvector associated with this largest eigenvalue.

Bonacich (1987), based on his earlier research, proposed a two-parameter family of prestige measures. In addition to the attenuation parameter of Katz (1953), which Bonacich calls a dependence parameter and denotes by β, a scale parameter, α, is introduced into the system of equations. The magnitude of β reflects the degree to which an actor's prestige is a function of the prestige of the actors to whom the actor is connected. The relationship is monotonic, and the parameter can be negative. Bonacich discusses bargaining situations in which prestige (or power, as he refers to it) arises when connections are made to those who are powerless. Bonacich gives an example of an exchange network from Cook, Emerson, Gilmore, and Yamagishi (1983) that has negative dependence. The choice of α depends on the value chosen for the dependence parameter β. Katz's (1953) single parameter prestige indices take $\alpha = 1$. Mathematical details, and examples of the use of this family can be found in Bonacich (1987).

Mizruchi, Mariolis, Schwartz, and Mintz (1986) (see also Mizruchi and Bunting 1981) focus attention on Bonacich's (1972a, 1972b) measure of prestige, and show how his index can be dichotomized as follows: one part due to the amount of prestige that an actor gets from another actor ("derived" prestige), and one due to the prestige that comes back to the original actor after being initially sent to the other actor ("reflected" prestige). This partition of prestige into derived and reflected parts was first suggested by the work of Mintz and Schwartz (1981a, 1981b). The goal of this research is to identify *hubs*, those actors adjacent to many peripheral actors, and *bridges*, those adjacent to few central or prestigious actors. We regret this usage of the term "bridge," which is usually synonymous with a graph theoretic line-cut (see Chapter 4). Hubs have large reflected prestige indices, while bridges have large derived prestige indices. This partition of prestige into derived and reflected parts was first suggested by the work of Mintz and Schwartz (1981a, 1981b). We refer the reader to Mizruchi, Mariolis, Schwartz, and Mintz (1986) for substantive interpretations of hubs and bridges. And, we refer the reader to Tam (1989) for a detailed mathematical study of the relationship between this approach and the more standard actor-level prestige indices.

To our knowledge, the only network computing package that calculates these prestige indices based on rank is *GRADAP* (Sprenger and Stokman 1989). However, the indices themselves are basically the elements of an eigenvector of a matrix based on **X**. Such eigenvectors are not difficult to find, given the available statistical computing packages. We discuss this calculation in more detail in our example. Most of the more complicated indices are elements of eigenvectors of suitably standardized sociomatrices. Thus, all can be calculated using numerical analysis packages such as that provided by *IMSL* and writing short *FORTRAN* computer programs. The *IBM*-compatible personal computer package *GAUSS* (GAUSS 1988), which contains many basic matrix manipulation features, can also do these calculations.

5.3.3 A Different Example

To best understand the use of these centrality and prestige indices, let us look at the Countries Trade Network data, and illustrate the calculation of the $\{P_P(n_i)\}$ and the $\{P_R(n_i)\}$ on these data. As mentioned, we will focus on the directional basic manufactured goods trade relation. Remember that the (i, j)th entry of the sociomatrix for this trade relation is unity if country i exports basic manufacturing goods to country j. Thus, countries are central if they export to others, and countries are prestigious if they import from other countries. In other words, prestigious actors are those with many imports (or those who import from many prestigious actors).

We first calculated actor degree and closeness centralities for the twenty-four countries in this network data set. These indices are shown in Table 5.2. The $\{C'_D(n_i)\}$ for the entire group are given in the first column. Two countries, n_{14} = Liberia, and n_{20} = Syria, export no basic manufactured goods to any of the other countries, so have zero row sums, even though they do import from some of the other countries. Since both these countries have zero outdegree, the directed graph representing this relation is not strongly or unilaterally connected (it is, however, weakly connected), and we cannot calculate closeness indices for the complete group. Thus, we dropped these two countries, and recalculated degree centralities, as well as closeness centralities for this reduced, but unilat-

Table 5.2. *Centrality indices for the countries trade network (*Actor and centralization indices calculated by dropping n_{14} = Liberia and n_{20} = Syria from the actor set.)*

	With g = 24 actors	With g = 22 actors	
	$C'_D(n_i)$	$C'_D(n_i)^*$	$C'_C(n_i)^*$
Algeria	0.174	0.190	0.553
Argentina	0.565	0.619	0.724
Brazil	0.913	0.905	0.913
China	0.913	0.905	0.913
Czechoslovakia	0.913	0.905	0.913
Ecuador	0.087	0.095	0.525
Egypt	0.391	0.429	0.636
Ethiopia	0.087	0.095	0.525
Finland	0.913	0.952	0.955
Honduras	0.043	0.048	0.512
Indonesia	0.609	0.667	0.750
Israel	0.478	0.524	0.667
Japan	1.000	1.000	1.000
Liberia	0.000	—	—
Madagascar	0.043	0.048	0.500
New Zealand	0.478	0.524	0.667
Pakistan	0.565	0.524	0.667
Spain	0.957	0.952	0.955
Switzerland	1.000	1.000	1.000
Syria	0.000	—	—
Thailand	0.609	0.619	0.724
United Kingdom	0.957	0.952	0.955
United States	1.000	1.000	1.000
Yugoslavia	0.783	0.810	0.840

erally connected digraph. These indices are shown in Columns 3 and 4 of Table 5.2.

Focus your attention on the smaller set of countries, those that export (have non-zero outdegrees). There are many "central" exporting countries. In order of decreasing degree centrality (using the smaller group), we have Japan, Switzerland, and United States (all with C'_D = 1.000), Finland, Spain, United Kingdom (these three with an index of 0.952), Brazil, China, Czechoslovakia (all tied at 0.905), Yugoslavia, Indonesia, Thailand, Israel, New Zealand, Pakistan, and so forth. The smallest exporters, and hence least central on this index, are Algeria, Ecuador, Ethiopia, Honduras, and Madagascar. We have almost exactly the same ordering at the top and at the bottom with closeness centrality as with degree centrality. The more developed countries appear to be the most central actors. It is remarkable that these two sets of actor indices agree so well.

The centralization indices for the group of 22 are $C_D^* = 0.333$, and $C_C^* = 0.495$, neither of which is particularly large, reflecting the uniform spread of the indices from the United States, Japan, and Switzerland at the top, to Madagascar at the bottom. The closeness centralities are larger than the degree centralities, and have a smaller range. The variance of the outdegrees is $S_D^2 = 71.64$, rather large (note that the outdegrees have a range of 0 to 23, with a mean of 13.1), so that the variance of the normalized actor degree centralities is 0.135. The variance of the normalized actor closeness centralities is only $S_C^2 = 0.0328$, much smaller than that for the degree indices, indicating more homogeneous actor closeness centralities. This homogeneity is probably due to the fact that the density of this relation is large (0.626) so that one can get from any country to any other country in relatively few steps, giving small distances from country to country on average. We also note that most countries trade with the "biggest" countries, so that even if the smaller countries do not trade with each other, their proximity to the big countries implies that the smaller countries are never very far away from each other (with respect to paths through the digraph).

We now turn to the calculation of the prestige indices. These indices are shown in Table 5.3. Prestige for these countries and this relation is synonymous with high involvement in the importing of basic manufactured goods from other countries. The first column contains the degree prestige indices for all twenty-four countries, and the second, the proximity prestige indices. Notice that even though Liberia and Syria do not export in this group (and hence have outdegrees of zero) we are still able to calculate the proximity prestige indices.

As can be seen from equation (5.24), the standardized degree prestige indices are simply the relative indegrees, standardized by dividing by their maximum possible value, $g-1$. Such quantities are standard output from most network computer packages. The proximity prestige indices can be calculated by first determining the $\{I_i\}$ values, the number of actors who can reach actor i, and then dividing these values by $g - 1$. This ratio is then divided by the average distances of all actors to actor i. Note that these average distances use the columns of the sociomatrix, rather than the rows (as the actor closeness indices do). In fact, if one transposes the sociomatrix, the average distances to an actor become the average distances involving the rows. Thus, the closeness centralities, which use

Table 5.3. *Prestige indices for the countries trade network*

	$P'_D(n_i)$	$P'_P(n_i)$	$P'_R(n_i)$
Algeria	0.565	0.661	0.222
Argentina	0.435	0.599	0.805
Brazil	0.478	0.619	1.000
China	0.652	0.710	0.711
Czechoslovakia	0.565	0.661	0.818
Ecuador	0.391	0.599	0.183
Egypt	0.522	0.599	0.482
Ethiopia	0.435	0.710	0.131
Finland	0.652	0.590	0.758
Honduras	0.391	0.581	0.072
Indonesia	0.609	0.599	0.617
Israel	0.435	0.599	0.682
Japan	0.739	0.767	0.680
Liberia	0.391	0.564	0.000
Madagascar	0.261	0.532	0.106
New Zealand	0.609	0.684	0.461
Pakistan	0.609	0.684	0.525
Spain	0.739	0.767	0.673
Switzerland	0.652	0.710	0.765
Syria	0.522	0.619	0.000
Thailand	0.652	0.710	0.589
United Kingdom	0.695	0.767	0.633
United States	0.826	0.799	0.644
Yugoslavia	0.652	0.710	0.680

the average distances from an actor to all other actors, calculated on the transposed sociomatrix, are exactly the average distances needed for the actor proximity prestige indices.

For the example, we note that all countries are reachable from all countries except Liberia (n_{14}) and Syria (n_{20}). Hence, the influence domain for the countries is the reduced group, giving $I_i = 21$. From equation (5.25), note that this gives us a numerator of 21/23 for all countries.

Examining Table 5.3 we see that the degree prestige indices cover a relatively narrow range of values, from 0.261 (for Madagascar) to 0.826 (for United States). Many countries import from almost all the other countries, and thus have large degree prestige indices: Spain, Japan, United Kingdom, China, Finland, Switzerland, Thailand, and Yugoslavia. The countries with the smallest degree prestige indices (and hence, few imports) are Argentina, Ecuador, Ethiopia, Honduras, Israel, Madagascar, and Liberia. Note that the prestigious countries are similar to the most central, except Thailand and Yugoslavia are prestigious, but not terribly central (import more but export less) and Brazil and Czechoslovakia are central but not prestigious (export more but import less). The least prestigious countries are also the least central.

Column 2 of Table 5.3 gives the actor proximity prestige indices, which have a much smaller range than those based on degree; in fact, the variance of the degree prestige indices is 0.0177, and just 0.0054 for the proximity prestige indices. We have exactly the same countries at the top and at the very bottom. Note, however, that the smallest proximity indices are 0.532 (Madagascar), indicating that even Madagascar is not terribly distant from the other countries. This is probably due to the large density for this relation; most countries do import from the countries in this group. We note that the average actor degree prestige index is 0.562, while the average actor proximity prestige index is 0.660.

Lastly, we turn to the actor status or rank prestige index. We take the sociomatrix, normalize it to have column sums of unity (by dividing by the indegrees), transpose it, and calculate its eigenvalues. Note that this sociomatrix is not symmetric; hence, the standard routines for extracting eigenvalues and eigenvectors, which are designed for symmetric matrices (such as covariance and correlation matrices), cannot be used. We used a small *FORTRAN* program, which calls the *IMSL* routine *EVCRG*. This subroutine extracts eigenvalues and eigenvectors from any real-valued matrix. Such quantities can be complex-valued, so care must be taken in interpreting the output.

As mentioned, the largest eigenvalue of the relevant matrix is unity. The elements of the eigenvector associated with this eigenvalue are the rank-prestige indices. For the countries' basic manufactured goods relation, the indices for the twenty-four countries are shown in Column 3 of Table 5.3. These indices are quite different from the other prestige indices. The ordering of the countries with respect to rank prestige is Brazil, Czechoslovakia, Argentina, Switzerland, Finland, China, Israel, Yugoslavia, and then Spain, United States, and United Kingdom. The addition of Argentina and Israel to this "prestigious subset" is somewhat surprising, since these two countries have small indegrees; but remember, what is important here is not how many countries a country is adjacent to, but the prestige of these countries. Specifically, prestigious countries are those that import goods from nations who in turn import goods. Clearly, Brazil, Czechoslovakia, and Argentina are linked directly to other prestigious countries.

5.4 Comparisons and Extensions

Several authors have compared the performance of the many centrality and prestige indices discussed in this chapter, either on real or simulated data, or both. Earlier researchers, such as Stogdill (1951), concentrated on different measures of actor degrees, thus focusing attention on only one centrality index. Most notable of recent comparative research are studies by Freeman (1979), Freeman, Roeder, and Mulholland (1980), Knoke and Burt (1983), Doreian (1986), Bolland (1988), Stephenson and Zelen (1989), and Friedkin (1991). We now review these comparisons.

The first, extensive study of centrality indices was undertaken by Freeman (1979). Freeman lists all thirty-four possible graphs with $g = 5$ nodes (itemized by Uhlenbeck and Ford 1962), and compares actor- and group-level degree, closeness, and betweenness centrality measures across the graphs. In brief, Freeman demonstrated that the betweenness indices best "captured" the essence of the important actors in the graphs. As we have mentioned throughout this chapter, closeness centrality indices could not be computed for disconnected graphs, and the star graph always attained the largest centralization score, while the circle graph attained the smallest centralization. Other, less obvious findings include:

- The three measures of centrality under review differed noticeably in their rankings of the thirty-four graphs.
- The range of variation in the actor centrality and group centralization scores is greatest for betweenness; that is, betweenness centralities generate the largest actor variances.
- The range of variation in the actor centrality and group centralization scores is least for degree; that is, degree centralities appear to generate the smallest actor variances.

Further, the more theoretical nature of the betweenness indices leads Freeman to recommend their useage over the other two.

Freeman, Roeder, and Mulholland (1980) replicated the MIT experiments, conducted by Bavelas (1950), Smith (1950), and Leavitt (1951), designed to study the effects of the structure of a network on problem solving, perception of leadership, and personal satisfaction (the three variables measured for each actor). Freeman, Roeder, and Mulholland sought to determine which of the three centrality indices (degree, closeness, and betweenness) was most relevant to the same tasks undertaken by the same kinds of networks studied in the earlier experiments. Freeman, Roeder, and Mulholland used four different graphs, all with $g = 5$, and found that betweenness indices best measured which actor in the set of actors was viewed most frequently as a leader. Both the degree and betweenness indices were important indicators of group performance (with respect to efficiency of problem solving). However, the closeness index (based on graph distance) was not even "vaguely related to experimental results" (Freeman, Roeder, and Mulholland 1980).

Knoke and Burt (1983), as part of their classic paper distinguishing between centrality and prestige, studied five centrality indices and five prestige indices. These indices were calculated for the Galesburg, Illinois, physician network studied by Coleman, Katz, and Menzel (1966) to identify diffusion of a medical innovation. Within each set of five indices, two were based on degree (see equation (5.3)), one on closeness (equation (5.8)), and one on either betweenness (for centrality — equation (5.12)) or rank (for prestige — equation (5.28)). The five centrality actor-level indices were calculated on a symmetrized version of the data (so that the graph was nondirected) and the five prestige indices, for the actual data. All these indices are output from the computer program *STRUCTURE* (Burt 1989). For the Galesburg network, the correlations among the centrality and among the prestige indices were high, as expected. In addition, the centrality and prestige indices were also associated. This strong association, which Knoke and Burt (1983) study further by using additional actor attributes (such as the date that the medical innovation was adopted) is described by these researchers as a unique feature of the network under study. It is thus difficult to extend these findings to general network data.

Doreian (1986) reviewed the work of Katz (1953), Harary (1959c), and Hubbell (1965), and focused on measures of "relative standing" of the actors in small networks. He criticized prestige indices based on degree or rank as being arbitrary (which is certainly true of Katz's and Hubbell's prestige indices, since there is not natural choice for scaling or attenuation parameters). Doreian advocated the use of an "iterated Hubbell" index, which converges to a standardized eigenvector of a function of a matrix derived from the sociomatrix. The advantage of this index is that it produces prestige measures that correspond well to the regular equivalences of the actors in the network (see White and Reitz 1983; and Chapter 12).

Bolland (1988) studied four centrality measures: degree, closeness, betweenness, and a new measure, "continuing flow," which combines degree and closeness. Bolland's continuing flow index examines all paths of (at most) a fixed length and counts how many of these paths originate with the ith actor. This count is then standardized, and the fixed length allowed to get as large as possible. Unlike the closeness and betweenness indices, this index considers all paths of any length, not just geodesics.

Bolland examined a network data set giving influence relationships among forty people involved in educational policy-making in Chillicothe, Ohio (see Bolland 1985). In addition to reporting extensive data analyses of this network, he conducted a Monte Carlo analysis by adding random and systematic variation to the network to obtain a number of "noisy" networks. These simulated networks were similar, but not exactly equal to, the original data. Each noisy network was replicated one hundred times to study the validity, robustness, and sensitivity of each of the four centrality indices.

Bolland's findings supported the earlier work of Freeman (1979). Specifically, degree-based measures of centrality are sensitive to small changes in network structure. Betweenness-based measures of centrality are useful and capable of capturing small changes in the network, but are error-prone. Closeness measures are much too sensitive to network change. Lastly, Bolland found the continuing flow index to be relatively insensitive to systematic variation, and useful in most circumstances. He recommends the use of both betweenness and continuing flow indices in practice.

Stephenson and Zelen (1989) compared their information centrality index to the other centrality indices using two data sets — the social network of forty AIDS patients mentioned earlier and a Gelada baboon colony of $g = 12$ animals, before and after the introduction of two additional group members. These latter data, gathered by Dunbar and Dunbar (1975), are analyzed longitudinally by Stephenson (1989). Stephenson and Zelen conducted the only comparison of the degree, closeness, and betweenness centrality measures, with the newer information index. There are several differences between information centrality indices and betweenness centrality indices. Specifically, information indices are much more "continuous" than those based on betweenness, which really are counts, rather than continuous-valued quantities. Thus, information indices can be more sensitive to slight arc changes than betweenness indices. Peripheral actors do not have much effect on the computed values of betweenness indices, since these actors rarely lie on geodesics; however, such actors can have significant effects in a network (especially in networks modeling disease transmission). Information indices are much more likely to measure the impact of these peripheral actors. Degree centrality indices have a limited ability to distinguish among actors with differing centrality. The range of possible values for a degree-based index is quite small, so that such indices are not very sensitive.

Friedkin (1991) offers a different theoretical foundation for the commonly used centrality measures based on a social influence process. He derives degree, closeness, and betweenness centrality measures by assuming that the network effects model (which basically is an application of an autoregressive model for spatially distributed actors or units) is appropriate. This model has been proposed for use in network analysis by Erbring and Young (1979), Doreian (1981), Burt (1987), and Friedkin and Johnsen (1990). The three measures are

(i) Total effects centrality — the total relative effect of an actor on the other actors in the network
(ii) Immediate effects centrality — the rapidity with which an actor's total effects are realized
(iii) Mediative effects centrality — the extent to which particular actors have a role in transmitting the total effects of other actors

Friedkin shows that these measures arise as "side effects" of the network process model of social influence. As can be seen by their definitions, they are congruent with the degree, closeness, and betweenness actor-centrality indices discussed here. Friedkin's work can be extended to directional relations, including real-valued ties, due to the measurement generality of the social process model. Such generalizations would yield new, theoretical rationales for prestige measures.

To gain a better understanding about how important a specific actor is to a network, one can take an actor with a large betweenness index, and drop it from the network (allowing this actor to serve as a "cutpoint"). Counting the number of components generated by this deletion will give an indication of how much "betweenness" this actor exerts over the network. Truly central actors will force many disconnected components to arise. Stephenson (1989) does this for the AIDS network, and finds that four of the actors in this network, which have large betweenness indices, do not "break up" the network when deleted. Betweenness is just one — of many — manifestations of the primary centrality concept. One should not utilize any single centrality measure. Each has its virtues and utility.

We should note that there is a variety of actor- and group-level degree-based indices that can be calculated and examined when more than one relation is measured. For example, one can study how likely it is that an actor chooses another actor on more than one relation. Such an index uses the quantities $x_{ij}(m) = 1$ if at least m of the ties $x_{ij1}, x_{ij2}, \ldots, x_{ijR}$ are equal to 1. An actor-level multiplex index can be calculated by averaging the quantities just over j. A group-level *multiplex index* can be calculated from these quantities, simply by averaging them over all i and j. An index based on network cohesion (for each relation) can be based on the number of dyads that are mutual.

With multirelational data, we suggest that the indices described in this chapter be calculated for *each* relation. We do not recommend (as some authors have, such as Knoke and Burt 1983) that the relations be aggregated into a single sociomatrix, unless there are strong substantive reasons for such aggregations (such as two measures of friendship combined into a single positive affect relation). Further multirelational analyses, designed to measure how similar actors are across relations and how associated the relations are, are discussed in Chapter 16.