The SAGE Handbook of Social Network Analysis

Concepts and Measures for Basic Network Analysis

Contributors: Robert A. Hanneman & Mark Riddle
Edited by: John Scott & Peter J. Carrington
Book Title: The SAGE Handbook of Social Network Analysis
Chapter Title: "Concepts and Measures for Basic Network Analysis"
Pub. Date: 2014
Access Date: December 17, 2018
Publishing Company: SAGE Publications Ltd
City: London
Print ISBN: 9781847873958
Online ISBN: 9781446294413
DOI: http://dx.doi.org/10.4135/9781446294413.n24
Print pages: 340-369

Concepts and Measures for Basic Network Analysis

Robert A.Hanneman and MarkRiddle

Networks and Actors

The social network perspective emphasizes multiple levels of analysis: differences among actors are traced to the constraints and opportunities that arise from how they are embedded in networks; the structure and behavior of networks are grounded in and enacted by local interactions among actors.

In this chapter we will examine basic concepts and measures used in formal social network analysis. Despite the simplicity of the ideas, there are good theoretical reasons (and considerable empirical evidence) to believe that these basic properties of social networks have very important consequences for both individuals and the larger social structures of which they are parts.

There are many ways that one could organize this survey, and we will only be able to provide an introduction. For more extended treatments, the reader may wish to consult John Scott's *Social Network Analysis* (2000), Stanley Wasserman and Katherine Faust's *Social Network Analysis: Methods and Applications* (1994), David Knoke and Song Yang's *Network Analysis* (2008), and our own text (Hanneman and Riddle, 2005).

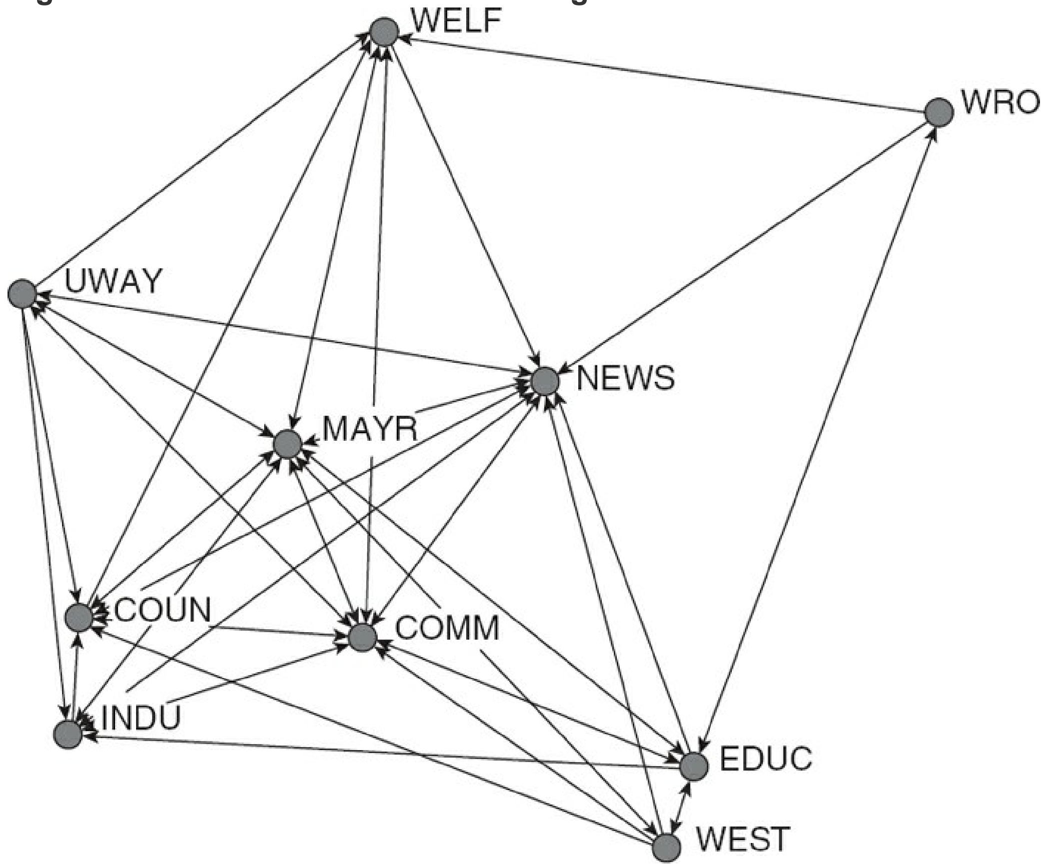
In the first half of this chapter we will focus on the network as a whole; this sort of “top-down” perspective enables us to see and measure aspects of whole social structures (e.g., families, groups, communities, markets, polities) that may be predictive of their dynamics. For example, networks where a high percentage of all possible ties among actors are actually present (high “density”) are often more prone to rapid “information cascades” than graphs with low density (Watts, 2003). Network analysts have developed a considerable number of measures to describe the “texture” of the “social fabric.” At the broadest level, the concern is with the solidarity and robustness of the whole structure and with the presence of – and relations among – substructures.

In the second half of this chapter we will shift our perspective to the “bottom up.” We will focus on some concepts and measures that provide insight into the ways that individuals are embedded in networks. For example, it is often the case that actors who have many others in their “neighborhood” (i.e., “alters” to whom “ego” has a direct connection) are more influential but sometimes also more constrained than those who have fewer connections. We will examine a number of related approaches that identify different aspects of the “ego network” of actors as potential sources of opportunity and constraint (Wellman et al., 1988).

Throughout this chapter we'll illustrate the concepts and measures discussed with output from the analysis of a small, binary, directed graph. These analyses are performed with UCINET (Borgatti et al., 2002), but there are many excellent software packages that provide similar tools (see, particularly, Huisman and van Duijin, this volume).

Our illustrations will examine the flow of information among 10 formal organizations concerned with social welfare issues in one midwestern U.S. city (Knoke and Wood, 1981). Without reading further, take just a moment to examine the graph of this network in Figure 24.1.

Figure 24.1 Knoke information exchange



The formal methods that we will be discussing in this chapter are really just ways of indexing features readily seen in the graph. From the "top down," we can see, for example, that there are a limited number of actors here (size), and all of them are connected (a single component). Not every possible connection is present (density), and there are thin spots (structural holes) in the fabric. There are some sets of organizations where everyone exchanges information with everyone (cliques), and these cliques are connected by overlapping membership (bridges).

Looking at the same graph from the "bottom up," there appear to be some differences among the actors in how connected they are; compare the newspaper to the welfare rights advocacy organization. If you look closely, you can see that some actors' connections are likely to be reciprocated (that is, if A shares information with B, B also shares information with A), while other actors are more likely to be senders than receivers of information. As a result of the variation in how connected individuals are, and whether the ties are reciprocated, some actors may be quite "central" and others less so. Actors may have "equivalent" positions in the network, which constitute "types" or "roles."

Social network analysts and graph theorists have developed a large number of formal algorithms to index these kinds of features about whole networks and about the positions of individuals within them. The remainder of this chapter is a quick tour of some of the most commonly used approaches.

The Whole Network

Since networks are defined by their actors and the connections among them, it is useful to begin our description by examining these very simple properties. Small groups differ from large groups in many important ways; indeed, population size figures largely in much sociological analysis (Finsveen and van Oorschot, 2008; Totterdell et al., 2008). The extent to which actors are connected may be a key indicator of the “cohesion” (Burris, 2005; Fominaya, 2007; Moody and White, 2003; Sanders and Nauta, 2004), “solidarity,” “moral density,” and “complexity” of the social organization (Crossley, 2008; Schnegg, 2007; Urry, 2006).

Size and Density

The size of a network is often very important. Imagine a group of 12 students in a seminar. It would not be difficult for each of the students to know each of the others fairly well and to build up exchange relationships (e.g., sharing reading notes). Now imagine a large lecture class of 300 students. It would be extremely difficult for any student to know all of the others, and it would be virtually impossible for there to be a single face-to-face network for exchanging reading notes. Size is critical for the structure of social relations because of the limited resources and capacities that each actor has for building and maintaining ties. Our example network has 10 actors. The size of a network is indexed simply by counting the number of nodes.

In any network there are (k * k–1) unique ordered pairs of actors (i.e., AB is different from BA, and self-ties are ignored), where k is the number of actors. You may wish to verify this for yourself with some small networks. So, in our network of 10 actors, with directed data, there are 90 logically possible relationships. If we had undirected (i.e., symmetric) ties, the number would be 45, since the relationship AB would be the same as BA. The number of logically possible relationships then grows exponentially as the number of actors increases linearly. It follows from this that the range of logically possible social structures increases (or, by one definition, “complexity” increases) exponentially with size.

The density of a binary network is simply the proportion of all possible ties that are actually present. For a valued network, density is defined as the sum of the ties divided by the number of possible ties (i.e., the ratio of all tie strength that is actually present to the number of possible ties). The density of a network may give us insights into such phenomena as the speed at which information diffuses among the nodes and the extent to which actors have high levels of social capital and/or social constraint.

In our example data, we see that there are 10 nodes; therefore, there are 90 possible connections. Of these, 49 are actually present. The network density then is .5444.

Connections

Size and density give us the overall sense of the range of possible social structures that could be present in a population, but what really matters is the pattern or “texture” of these connections. There are many widely used indexes that summarize various aspects of the structure of connections in a graph.

Reachability

An actor is “reachable” by another if there is a set of connections by which we can trace from the source to the target actor, regardless of how many others fall between them. If the data are asymmetric or directed, it is possible that actor A can reach actor B, but that actor B cannot reach actor A. With symmetric or undirected data, of course, each pair of actors either is or is not reachable to one another. If some actors in a network cannot reach others, there is potential for a division of the network. Or it may indicate that the population we are studying is really composed of more than one subpopulation. In the Knoke information exchange data set, it turns out that all actors are reachable by all others. A message or signal originating anywhere in the network could potentially be received by all other nodes.

Connectivity

Even though one actor may be able to reach another, the connection may not be a strong one. If there are many different pathways that connect two actors, they have high "connectivity" in the sense that there are multiple ways for a signal to reach from one to the other (Burris, 2005; Crossley, 2008; Finsveen and van Oorschot, 2008; Fominaya, 2007; Haythornthwaite, 2005; Hermann, 2008; Kien, 2008; Kratke and Brandt, 2009). The measure "connectivity" counts the number of nodes that would have to be removed in order to make one actor unreachable by another. Figure 24.2 shows the point connectivity for the flow of information among the 10 Knoke organizations.

Figure 24.2 Point connectivity of Knoke information exchange

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	5	5	3	4	5	1	6	4	4	3
2	5	8	3	5	8	1	6	5	3	4
3	3	3	4	4	3	1	4	3	3	3
4	5	5	3	5	5	1	5	4	3	4
5	5	8	3	5	8	1	6	5	3	5
6	1	1	1	1	1	1	2	1	2	1
7	5	6	3	5	6	1	6	4	2	3
8	5	5	3	5	5	1	5	5	4	4
9	3	3	3	3	3	1	3	3	3	3
10	4	5	3	4	5	1	4	4	3	5

The result demonstrates the tenuousness of organization 6's (the welfare rights organization) connection as both a source (row) or receiver (column) of information. To get its message to most other actors, organization 6 has only one alternative; should a single organization refuse to pass along information, organization 6 would receive none at all! Point connectivity can be a useful measure to get at notions of dependency and vulnerability.

Distance

The properties of the network that we have examined so far primarily deal with adjacencies, or the direct connections from one actor to the next. But the way that people are embedded in networks is more complex than this. Two persons, call them A and B, might each have five friends. But suppose that none of person A's friends have any friends except A. Person B's five friends, in contrast, each have five friends. The information available to B, and B's potential for influence is far greater than A's. That is, sometimes being a “friend of a friend” may be quite consequential.

To capture this aspect of how individuals are embedded in networks, one common approach is to examine the distance between actors. If two actors are adjacent, the distance between them is one (that is, it takes one step for a signal to go from the source to the receiver). If A tells B, and B tells C (and A does not tell C), then actors A and C are at a distance of two. The distances among actors in a network may be an important macro-characteristic of the network as a whole. Where distances are great, it may take a long time for information to diffuse across a population. It may also be that some actors are quite unaware of and not influenced by others; even if they are technically reachable, the costs may be too high to conduct exchanges.

The most commonly used definition of the distance between two actors in a network is *geodesic distance*. For binary data, the geodesic distance is the number of relations in the shortest possible pathway from one actor to another. The geodesic distance from one actor to another that is not reachable is usually treated as infinite, or equal to the largest observed distance in the graph. Valued data are often dichotomized in order to calculate the geodesic distance.

For valued networks, there are several alternative approaches to defining distances. Where we have measures of the strengths of ties (e.g., the dollar volume of trade between two nations), the “nearness (the opposite of distance)” between two actors is often defined as the strength of the weakest path between them. If A sends 6 units to B, and B sends 4 units to C, the “strength” of the path from A to C (assuming A to B to C is the shortest path) is 4. Where we have a measure of the cost of making a connection (as in an “opportunity cost” or “transaction cost” analysis), the “distance” between two actors is defined as the sum of the costs along the shortest pathway. Where we have a measure of the probability that a link will be used, the “distance” between two actors is defined as the product along the pathway (as in path analysis in statistics).

Indices of nearness or distance may also be weighted in various ways. For example, we might imagine that the value or potency of a signal decays exponentially, rather than linearly, as it passes through more and more intervening nodes between two actors.

In our example, we are using simple directed adjacencies, and the results (Figure 24.3) are quite straightforward.

Figure 24.3 Geodesic distances for Knoke information exchange

Geodesic Distances

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	1	2	2	1	3	1	2	1	2
2	1	0	1	1	1	2	1	1	1	2
3	2	1	0	1	1	1	1	2	2	1
4	1	1	2	0	1	3	1	2	2	2
5	1	1	1	1	0	2	1	1	1	1
6	3	2	1	2	2	0	1	3	1	2
7	2	1	2	1	1	3	0	2	2	2
8	1	1	2	1	1	3	1	0	1	2
9	2	1	2	2	1	3	1	2	0	2
10	1	1	1	2	1	2	1	2	2	0

Because the network is moderately dense, the geodesic distances are generally small. This suggests that information may travel pretty quickly in this network.

For each actor, that actor's largest geodesic distance is called its *eccentricity*, a measure of how far an actor is from the furthest other. For the network as a whole, the *diameter* is defined as the largest eccentricity. The mean (or median) geodesic distance and the standard deviation in geodesic distances may be used to summarize overall distance and the heterogeneity of distances in a network.

The use of geodesic paths to examine properties of the distances between individuals and for the whole network often makes a great deal of sense. There may be other cases, however, for which the distance between two actors and the connectedness of the graph as a whole is best thought of as involving all connections, not just the most efficient ones. For example, if I start a rumor, it will pass through a network by all pathways, not just the most efficient ones. How much credence another person gives my rumor may depend on how many times they hear it from different sources, and not merely how soon they hear it (Frank, 1996; Gallie, 2009; Gurrieri, 2008; Lai and Wong, 2002; Rycroft, 2007). For uses of distance like this, we need to take into account all of the connections among actors.

One approach to measuring how connected two actors are is to ask how many different actors in the neighborhood of a source lie on pathways to a target. If I need to get a message to you, and there is only one other person to whom I can send this for retransmission, my connection is weak, even if the person I send it to may have many ways of reaching you. If, on the other hand, there are four people to whom I can send my message, each of whom has one or more ways of retransmitting my message to you, then my connection is stronger. The "flow" approach suggests that the strength of my tie to you is no stronger than the weakest link in the chain of connections, where weakness means a lack of alternatives. For our directed information flow data, the results of UCINET's count of maximum flow are shown in Figure 24.4.

Figure 24.4 UCINET "maximum flow" for Knoke information network

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
	—	—	—	—	—	—	—	—	—	—
1	0	4	3	4	4	1	4	2	4	2
2	5	0	3	5	7	1	7	2	5	2
3	5	6	0	5	6	1	6	2	5	2
4	4	4	3	0	4	1	4	2	4	2
5	5	8	3	5	0	1	8	2	5	2
6	3	3	3	3	3	0	3	2	3	2
7	3	3	3	3	3	1	0	2	3	2
8	5	6	3	5	6	1	6	0	5	2
9	3	3	3	3	3	1	3	2	0	2
10	5	5	3	5	5	1	5	2	5	0

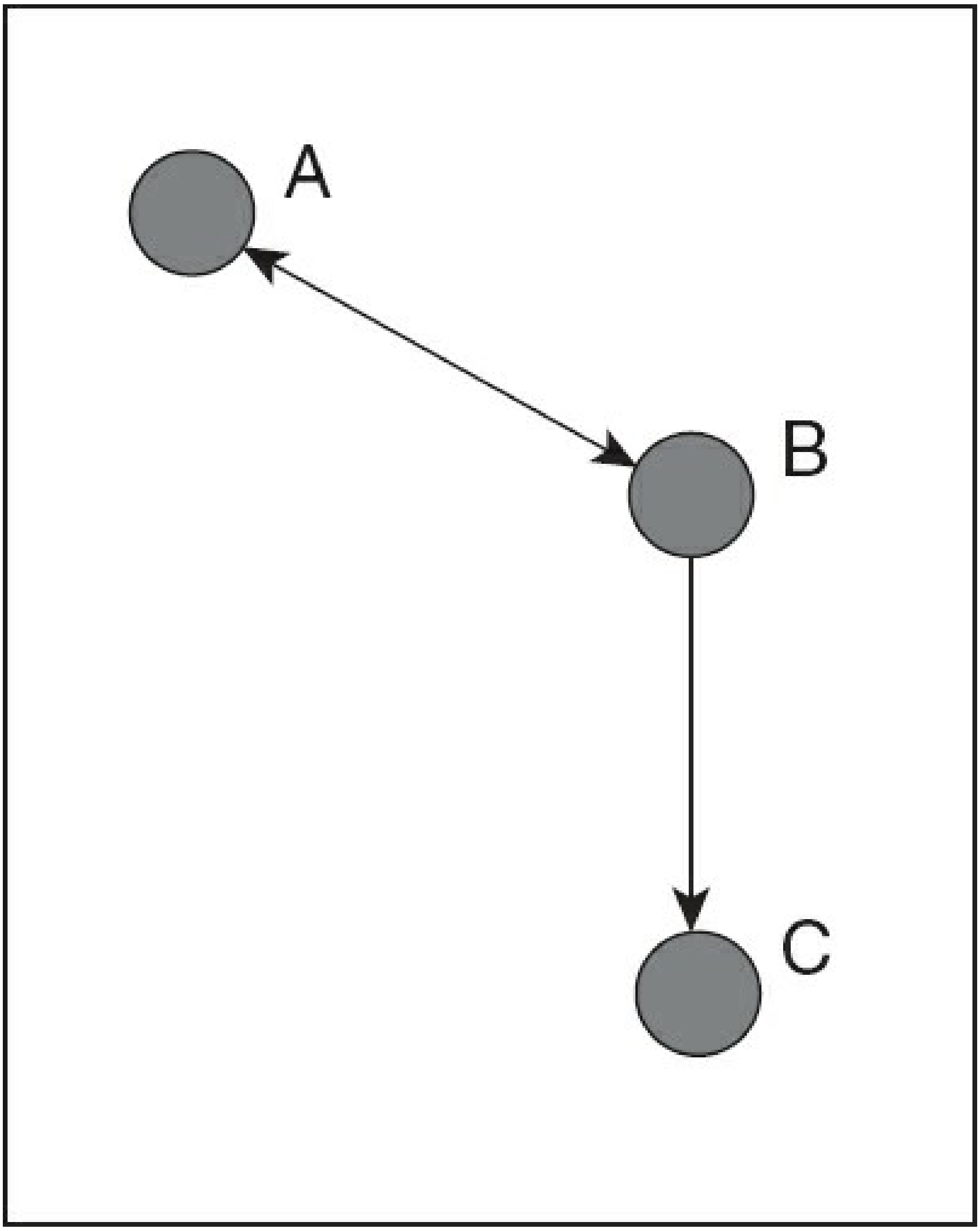
You should verify for yourself that, for example, there are four alternative routes in flows from actor 1 to actor 2, but five such points in the flow from actor 2 to actor 1. The higher the number of flows from one actor to another, the greater the likelihood that communication will occur, and the less "vulnerable" the connection. Note that actors 6, 7, and 9 are relatively disadvantaged. In particular, actor 6 has only one way of obtaining information from all other actors (the column vector of flows to actor 6).

Reciprocity

The smallest "social structure" represented in a graph is a "dyad," the relation between two actors. With symmetric dyadic data, two actors are either connected or they are not. Density tells us pretty much all there is to know. If we are considering a directed relation, there are three kinds of dyads (no tie, a single tie, or ties in both directions). The extent to which a population is characterized by "reciprocated" ties may tell us about the degree of cohesion in populations. Some theorists feel that there is an equilibrium tendency toward dyadic relationships to be either null or reciprocated and that asymmetric ties may be unstable. A network that has a predominance of null or reciprocated ties over asymmetric connections may be a more "equal" or "stable" network than one with a predominance of asymmetric connections (which might be more of a hierarchy).

There are several different approaches to indexing the degree of reciprocity in a population. Consider the very simple network shown in Figure 24.5. Actors A and B have reciprocated ties, actors B and C have a nonreciprocated tie, and actors A and C have no tie.

Figure 24.5 Definitions of reciprocity



What is the prevalence of reciprocity in this network? One approach is to focus on dyads, and ask which pairs have a reciprocated tie between them. This would yield one such tie for three possible pairs (AB, AC, BC), or a reciprocity rate of .333. More commonly, analysts are concerned with the ratio of the number of pairs with a reciprocated tie relative to the number of pairs with any tie. In large populations, most actors are linked directly to relatively few other actors, and it may be more sensible to focus on the degree of reciprocity among pairs that have any ties between them. In our simple example, this would yield one reciprocated pair divided by two tied pairs, or a reciprocity rate of .500. In the Knoke information tie network, the proportion of all dyads having a tie that has a reciprocated tie is .5313. This is neither “high” nor “low” in itself but does seem to suggest a considerable degree of institutionalized horizontal connection within this organizational population. One may also focus on relations, rather than on dyads, by asking what the proportion is of all relations in the graph as part of reciprocated relations. For our example data, this approach yields a result of .6939. That is, of all the relations in the graph, 69% are parts of reciprocated ties.

Transitivity

The smallest social structure that has the true character of a "society" is the triad, any "triple" {A, B, C} of actors. Such a structure "embeds" dyadic relations in a structure where "other" is present along with "ego" and "alter." In (directed) triads, we can see the emergence of tendencies toward equilibrium and consistency of social structures ("institutionalization"), as well as features such as balance and transitivity. Triads are also the simplest structures in which we can see the emergence of hierarchy.

With undirected data, there are four possible types of triadic relations (no ties, one tie, two ties, or all three ties). Counts of the relative prevalence of these four types of relations across all possible triples (that is, a "triad census") can give a good sense of the extent to which a population is characterized by "isolation," "couples only," "structural holes" (i.e., where one actor is connected to two others who are not connected to each other), or "clusters."

With directed data, there are actually 16 possible types of relations among three actors, including relationships that exhibit hierarchy, equality, and the formation of exclusive groups (e.g., where two actors connect and exclude the third). Thus, small-group researchers suggest that all of the really fundamental forms of social relationships can be observed in triads. Because of this interest, we may wish to conduct a "triad census" for each actor and for the network as a whole.

Of the 16 possible types of directed triads, six involve zero, one, or two relations, and can't display transitivity because there are not enough ties to do so. One type with three relations (AB, BC, CB) does not have any ordered triples (AB, BC) and hence can't display transitivity. In three more types of triads, there are ordered triples (AB, BC) but the relation between A and C is not transitive. The remaining types of triads display varying degrees of transitivity.

Figure 24.6 displays the results of one type of transitivity analysis of the Knoke information data.

Figure 24.6 Transitivity results for Knoke information network

```
TRANSITIVITY
--------------------------------------------------------------------------------

Type of transitivity:          ADJACENCY
Input dataset:                 C:\Program Files\Ucinet 6\DataFiles\KNOKBUR

Relation: KNOKI
---------------

Number of non-vacuous transitive ordered triples: 146
Number of triples of all kinds: 720
Number of triples in which i-->j and j-->k: 217

Percentage of all ordered triples: 20.28%
Transitivity: % of ordered triples in which i-->j and j-->k that are transitive: 67.28%
```

With 10 nodes, there are 720 triads. However, only 146 have enough ties to display transitivity. That is, there are 146 cases where, if AB and BC are present, then AC is also present. There are a number of different ways in which we could try to norm this count so that it becomes more meaningful. One approach is to divide the number of transitive triads by the total number of triads of all kinds (720). This shows that 20.28 percent of all triads are transitive. Perhaps more meaningful is to norm the number of transitive triads by the number of cases where a single link could complete the triad. That is, norm the number of {AB, BC, AC} triads by the number of {AB, BC, anything} triads. Seen in this way, about two-thirds or all relations that could easily be transitive actually are.

Clustering

Most of the time, most people interact with a fairly small set of others, many of whom know one another. The extent of local "clustering" in populations can be quite informative about the texture of everyday life. Watts (1999) and many others have noted that in large, real-world networks (of all kinds of things) there is often a structural pattern that seems somewhat paradoxical.

On one hand, in many large networks (like, for example, the Internet), the average geodesic distance between any two nodes is relatively short (Field et al., 2006; Hampton and Wellman, 1999). The "six degrees" of distance phenomenon is an example of this. So, most of the nodes in even very large networks may be fairly close to one another. The average distance between pairs of actors in large empirical networks is often much shorter than in random graphs of the same size.

On the other hand, most actors live in local neighborhoods where most others are also connected to one another. That is, in most large networks, a very large proportion of the total number of ties are highly "clustered" into local neighborhoods. That is, the density in local neighborhoods of large graphs tends to be much higher than we would expect for a random graph of the same size.

Most of the people we know may also know one another, which gives the impression that we live in a very narrow social world. Yet, at the same time, we can be at quite short distances to vast numbers of people whom we don't know at all. The "small-world" phenomena – a combination of short average path lengths over the entire graph, coupled with a strong degree of "clique-like" local neighborhoods – seems to have evolved independently in many large networks.

We've already discussed one part of this phenomenon. The average geodesic distance between all actors in a graph gets at the idea of how close actors are together. The other part of the phenomenon is the tendency toward dense local neighborhoods, or what is now thought of as "clustering."

One common way of measuring the extent to which a graph displays clustering is to examine the local neighborhood of an actor (that is, all the actors who are directly connected to ego), and to calculate the density in this neighborhood (leaving out ego). After doing this for all actors in the whole network, we can characterize the degree of clustering as an average of all the neighborhoods in the whole graph.

Figure 24.7 shows the clustering of the Knoke information network.

Figure 24.7 Clustering coefficient of Knoke information network

```
Input dataset:                 C:\Program Files\Ucinet 6

Relation: KNOKI
---------------
Overall graph clustering coefficient: 0.607
Weighted Overall graph clustering coefficient: 0.599
```

Two alternative measures are presented. The "overall" graph clustering coefficient is simply the average of the densities of the neighborhoods of all of the actors. The "weighted" version gives weight to the neighborhood densities proportional to their size; that is, actors with larger neighborhoods get more weight in computing the average density. Since larger graphs are generally (but not necessarily) less dense than smaller ones, the weighted average neighborhood density (or clustering coefficient) is usually less than the unweighted version. In our example, we see that all of the actors are surrounded by local neighborhoods that are fairly dense; our organizations can be seen as embedded in dense local neighborhoods to a fairly high degree. Lest we overinterpret, we must remember that the overall density of the entire graph in this population is rather high (.54). So, the density of local neighborhoods is not really much higher than the density of the whole graph. In assessing the degree of clustering, it is usually wise to compare the clustering coefficient to the overall density.

Connections Among Groups

In addition to dyads, triads, and local clustering, the texture of connections in a network can be affected by “categorical social units” or “subpopulations” defined either by shared attributes or contexts. Persons of the same gender may be more likely to form friendship ties; persons who attend the same school are more likely to be acquainted. The extent to which these subpopulations are open or closed (i.e., the extent to which most individuals have most of their ties within the boundaries of these groups) may be a telling dimension of social structure.

Block Density

In an organizational community, we might suppose that there may be competition (expressed as information hoarding) between organizations of the same type, and cooperation between organizations of different, complementary types. We have used an attribute or partition to divide the cases in Knoke information exchange data into three subpopulations (governmental agencies, nongovernmental generalists, and welfare specialists) so that we can see the amount of connection within and between groups. We can then examine the patterns of ties within and between “blocks” of nodes of the same type. Consider the results in Figure 24.8.

Figure 24.8 Block density of three subpopulations in Knoke information network

```
Column
Block Old Code        Members:
----- --------     ---------------
    1        1     COUN EDUC MAYR
    2        2     COMM INDU NEWS
    3        3     WRO UWAY WELF WEST


Relation: KNOKI

                                1
      1 3 5   2 4 7   6 8 9 0
      C E M   C I N   W U W W
     -------------------------
  1 |     1 | 1   1 |     1   |
  3 |     1 | 1 1 1 | 1     1 |
  5 | 1 1   | 1 1 1 |   1 1 1 |
```

```
 5 | 1 1   | 1 1 1 |     1 1 1 |
   ---------------------------
 2 | 1 1 1 |   1 1 |     1 1   |
 4 | 1   1 | 1   1 |           |
 7 |     1 | 1 1   |           |
   ---------------------------
 6 |   1   |     1 |       1   |
 8 | 1   1 | 1 1 1 |       1   |
 9 |     1 | 1   1 |           |
10 | 1 1 1 | 1   1 |           |
    --------------------------


Density / average value within blocks

                  1      2      3
                  1      2      3
             ------ ------ ------
   1 1   0.6667 0.8889 0.5000
   2 2   0.6667 1.0000 0.1667
   3 3   0.5833 0.6667 0.1667

Standard Deviations within blocks

                  1      2      3
                  1      2      3
             ------ ------ ------
   1 1   0.4714 0.3143 0.5000
   2 2   0.4714 0.0000 0.3727
   3 3   0.4930 0.4714 0.3727
```

The density in the 1,1 block is .6667. That is, of the six possible directed ties among actors 1, 3, and 5, four are actually present (ignoring the diagonal, which is the most common approach). We can see that the three subpopulations appear to have some differences. Governmental generalists (block 1) have quite dense in- and out-ties to one another, and to the other populations; nongovernment generalists (block 2) have out-ties among themselves and with block 1 and have high densities of in-ties with all three subpopulations. The welfare specialists have high density of information sending to the other two blocks (but not within their block), and receive more input from governmental than from nongovernmental organizations.

The extent to which these simple characterizations of blocks characterize all the individuals within those blocks – essentially the validity of the blocking – can be assessed by looking at the standard deviations within the partitions. The standard deviations measure the lack of homogeneity within the partition, or the extent to which the actors vary.

Group-External and Group-Internal Ties

Krackhardt and Stern (1988) developed a very simple and useful measure of group embedding based on comparing the numbers of ties within groups to those between groups. The E-I (external–internal) index takes the number of ties of group members to outsiders, subtracts the number of ties to other group members, and divides by the total number of ties. The resulting index ranges from −1 (all ties are internal to the group) to +1 (all ties are external to the group). Since this measure is concerned with any connection between members, the directions of ties are ignored (i.e., either an out-tie or an in-tie constitutes a tie between two actors).

The E-I index can be applied at three levels: the entire population, each group, and each individual. That is, the network as a whole (all the groups) can be characterized in terms of the boundedness and closure of its subpopulations. We can also examine variation across the groups in their degree of closure; each individual can be seen as more or less embedded in its group.

To assess whether a given E-I index value is significantly different than what would be expected by random mixing (i.e., no preference for within- or without-group ties by group members), a permutation test can be performed. A large number of trials are run in which the blocking of groups is maintained, and the overall density of ties is maintained, but the actual ties are randomly distributed. A sampling distribution of the numbers of internal and external ties under the assumption that ties are randomly distributed is calculated and used to assess the frequency with which the observed result would occur by sampling from a population in which ties were randomly distributed. Results for the blocked Knoke data are shown as Figure 24.9.

Figure 24.9 E-I index output for the Knoke information network

```
Density matrix

            1     2     3
            1     2     3
        ----- ----- -----
  1 1   0.667 1.000 0.667
  2 2   1.000 1.000 0.667
  3 3   0.667 0.667 0.333

64 ties.

Whole Network Results

                    1      2      3      4
                 Freq    Pct Possib Densit
               ------ ------ ------ ------
  1 Internal   14.000  0.219 24.000  0.583
  2 External   50.000  0.781 66.000  0.758
  3      E-I   36.000  0.563 42.000  0.467


Max possible external ties: 66.000
Max possible internal ties: 24.000

E-I Index: 0.563
Expected value for E-I index is: 0.467

Max possible E-I given density & group sizes: 1.000
Min possible E-I given density & group sizes: 0.250

Re-scaled E-I index: -0.167


Permutation Test
Number of iterations = 5000

                     1       2       3       4       5       6       7
                   Obs     Min     Avg     Max      SD P >= Ob P <= Ob
               ------- ------- ------- ------- ------- ------- -------
  1 Internal     0.219   0.625   0.733   0.844   0.039   1.000   0.000
  2 External     0.781   0.156   0.267   0.375   0.039   0.000   1.000
  3      E-I     0.563   0.250   0.467   0.688   0.078   0.203   0.953
```

The observed block densities are presented first. Since any tie (in or out) is regarded as a tie, the densities in this example are quite high. The densities off the main diagonal (out-group ties) appear to be slightly more prevalent than the densities on the main diagonal (in-group ties).

Next, we see the numbers of internal ties (14, or 22 percent) and external ties (50, or 78 percent) that yield a raw (not rescaled) E-I index of +.563. That is, this graph displays a preponderance of external over internal ties. Also shown are the maximum possible numbers of internal and external ties given the group sizes and density. Note that, due to these constraints, the result of a preponderance of external ties is not unexpected: under a random distribution, the E-I index would be expected to have a value of .467, which is not very much different from the observed value.

We see that, given the group sizes and density of the graph, the maximum possible value of the index (1.0) and its minimum value (+.25) are both positive. If we re-scale the observed value of the E-I index (.563) to fall into this range, we obtain a re-scaled index value of –.167. This suggests that, given the demographic constraints and overall density, there is a very modest tendency toward group closure.

The last portion of the results gives the values of the permutation-based sampling distribution. Most important here is the standard deviation of the sampling distribution of the index, or its standard error (.078). This suggests that the value of the raw index is expected to vary by this much from trial to trial (on the average) just by chance. Given this result, we can compare the observed value in our sample (.563) to the expected value (.467) relative to the standard error. The observed difference of about .10 could occur fairly frequently just by sampling variability (p = .203). Most analysts would not reject the null hypothesis that the deviation from randomness was not “significant.” That is, we cannot be confident that the observed mild bias toward group closure is not random variation.

Substructures

One of the most common interests of structural analysts is in the “substructures” that may be present in a network. The dyads, triads, and ego-centered neighborhoods that we examined earlier can all be thought of as substructures. In this section, we'll consider some approaches to identifying larger groupings.

Many of the approaches to understanding the structure of a network emphasize how dense connections are built up from simpler dyads and triads to more extended dense clusters, such as “cliques.” This view of social structure focuses attention on how solidarity and connection of large social structures can be built up out of small and tight components, a sort of “bottom up” approach. Network analysts have developed a number of useful definitions and algorithms that identify how larger structures are compounded from smaller ones: cliques, N-cliques, N-clans, K-plexes, and K-cores all look at networks this way.

We can also look for substructure from the “top down.” Looking at the whole network, we can think of substructures as areas of the graph that seem to be locally dense but are separated, to some degree, from the rest of the graph. This idea has been applied in a number of ways: components, blocks/cutpoints, K-cores, Lambda sets and bridges, factions, and f-groups will be discussed here. It is important to note (Moody and White, 2003), that bottom-up and top-down approaches to substructures in graphs often do not identify the same groupings. The choice of method should be informed by the researcher's definition of a meaningful substructure for the purposes of analysis.

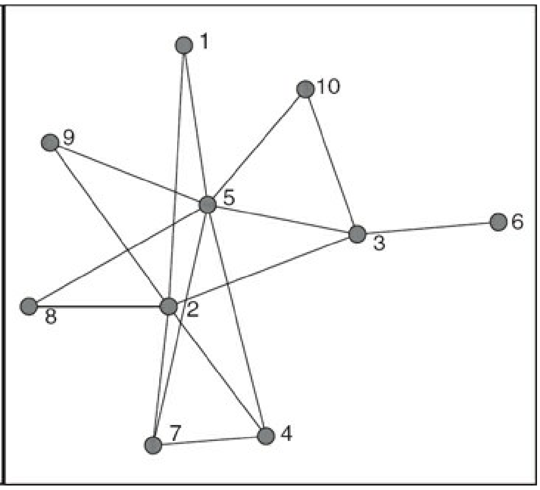
The idea that some regions of a graph may be less connected to the whole than others may lead to insights into lines of cleavage and division. Weaker parts in the “social fabric” also create opportunities for brokerage and less constrained action. So the numbers and sizes of regions, and their “connection topology” may be consequential for predicting both the opportunities and constraints facing groups and actors, as well as for predicting the evolution of the graph itself.

Most computer algorithms for locating substructures operate on binary symmetric data. We will use the Knoke information exchange data for most of the illustrations that follow. Where algorithms allow it, the directed form of the data will be used. Where symmetric data are called for, we will analyze “strong ties.” That is, we will symmetrize the data by insisting that ties must be reciprocated in order to count (i.e., a tie only exists if xy and yx are both present). The reciprocity-symmetric data matrix and graph are shown in Figure 24.10.

Figure 24.10 Reciprocated relations in the Knoke information network

Matrix #1: KNOKI

		1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	1	1	1	0	0	1	0	0	0	0	1
2	2	1	1	1	1	1	0	1	1	1	0
3	3	0	1	1	0	1	1	0	0	0	1
4	4	0	1	0	1	1	0	1	0	0	0
5	5	1	1	1	1	1	0	1	1	1	1
6	6	0	0	1	0	0	1	0	0	0	0
7	7	0	1	0	1	1	0	1	0	0	0
8	8	0	1	0	0	1	0	0	1	0	0
9	9	0	1	0	0	1	0	0	0	1	0
10	10	1	0	1	0	1	0	0	0	0	1



Bottom-Up Approaches: Clique, Clan, Plex, Core, and F-Group

In a sense, all networks are composed of groups (or subgraphs). When two actors have a tie, they form a "group." One approach to thinking about the group structure of a network begins with this most basic group and seeks to see how far this kind of close relationship can be extended. This is a useful way of thinking, because sometimes more complex social structures evolve, or emerge, from very simple ones.

Cliques are the (maximal) subgraphs of nodes that have all possible ties present among themselves. That is, a clique is the largest possible collection of nodes (more than two) in which all actors are directly connected to all others. Figure 24.11 shows part of the UCINET analysis of cliques in our symmetrized data.

Figure 24.11 Cliques in the reciprocity-symmetric Knoke information network

7 cliques found.

1: COMM INDU MAYR NEWS
2: COMM EDUC MAYR
3: COUN COMM MAYR
4: COMM MAYR UWAY
5: COMM MAYR WELF
6: COUN MAYR WEST
7: EDUC MAYR WEST

Clique Proximities: Prop. of clique members adjace

		1	2	3	4	5	6	7
1	1	0.500	0.667	1.000	0.667	0.667	1.000	0.667
2	2	1.000	1.000	1.000	1.000	1.000	0.667	0.667
3	3	0.500	1.000	0.667	0.667	0.667	0.667	1.000
4	4	1.000	0.667	0.667	0.667	0.667	0.333	0.333
5	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	6	0.000	0.333	0.000	0.000	0.000	0.000	0.333
7	7	1.000	0.667	0.667	0.667	0.667	0.333	0.333
8	8	0.500	0.667	0.667	1.000	0.667	0.333	0.333
9	9	0.500	0.667	0.667	0.667	1.000	0.333	0.333
10	10	0.250	0.667	0.667	0.333	0.333	1.000	1.000

There are seven maximally complete subgraphs present in these data. The largest one is composed of 4 of the 10 actors (2, 4, 5, and 7), and all of the other smaller cliques share some overlap with some part of the largest clique. The second panel shows how "adjacent" each actor (row) is to each clique (column). Actor 1, for example, is adjacent to two-thirds of the members of clique 5. There is a very high degree of common membership in these data.

We can look at the extent to which the cliques overlap with one another, as measured by the numbers of members in common, as in Figure 24.12.

Figure 24.12 Clique overlap in the reciprocity-symmetric Knoke information network

```
Clique-by-Clique Actor Co-membership matrix

      1 2 3 4 5 6 7
      - - - - - - -
   1  4 2 2 2 2 1 1
   2  2 3 2 2 2 1 2
   3  2 2 3 2 2 2 1
   4  2 2 2 3 2 1 1
   5  2 2 2 2 3 1 1
   6  1 1 2 1 1 3 2
   7  1 2 1 1 1 2 3

HIERARCHICAL CLUSTERING OF OVERLAP MATRIX

Level    1 2 3 4 5 6 7
-----    - - - - - - -
2.000    XXXXXXXXX XXX
1.072    XXXXXXXXXXXXX
```

A cluster analysis of the closeness of the cliques shows that cliques 6 and 7 are (a little) separate from the other cliques. That is, there is a tendency toward one larger "clique of cliques" and one smaller one.

N-Cliques

The strict clique definition (maximally connected subgraph) may be too strong for many purposes. It insists that every member of a subgroup should have a direct tie with each and every other member. One alternative is to define an actor as a member of a clique if they are connected to every other member of the group at some distance greater than one. Usually, a path distance of two is used. This corresponds to being "a friend of a friend." This approach to defining substructures is called an N-clique, where *n* stands for the maximum length of paths to all other members (Figure 24.13).

Figure 24.13 N-cliques of reciprocity-symmetric Knoke information network (N) = 2)

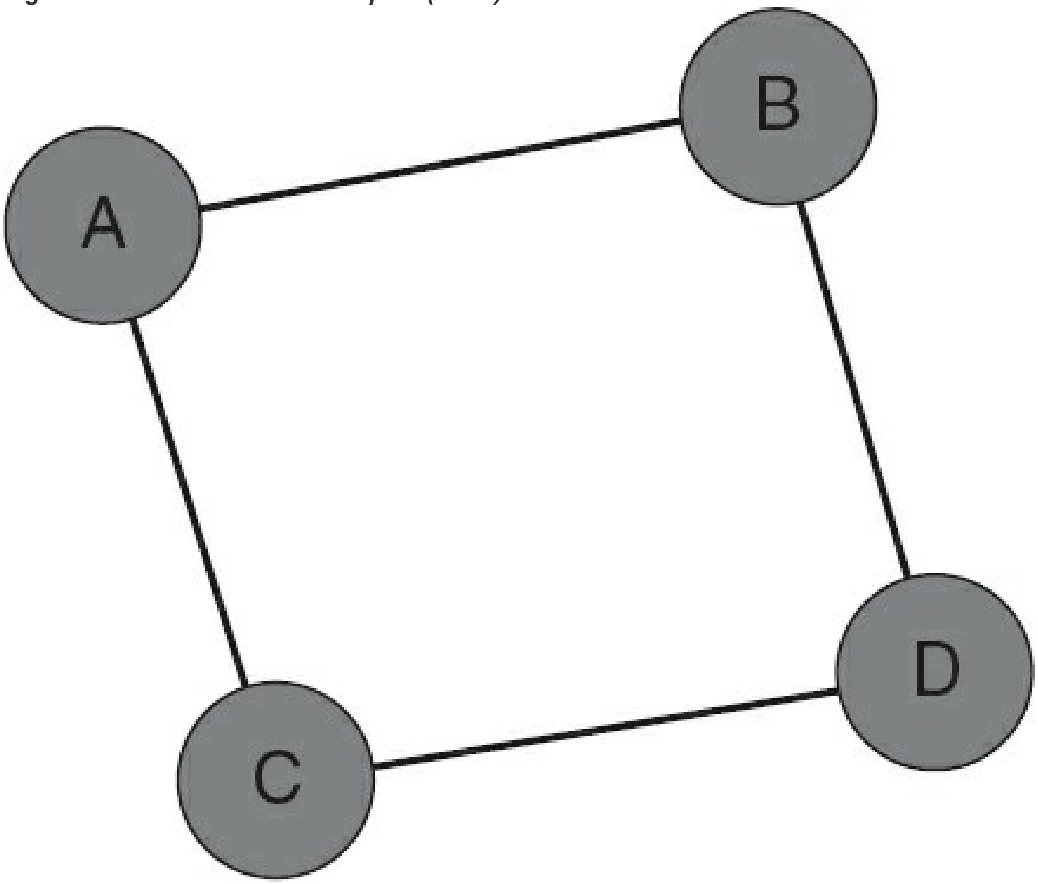
```
2 2-cliques found.

   1:   COUN COMM EDUC INDU MAYR NEWS UWAY WELF WEST
   2:   COMM EDUC MAYR WRO WEST


                                    1
        1  2  3  4  5  6  7  8  9  0
        C  C  E  I  M  W  N  U  W  W
        -  -  -  -  -  -  -  -  -  -
  1  1  1  1  1  1  1  0  1  1  1  1
  2  2  1  2  2  1  2  1  1  1  1  2
  3  3  1  2  2  1  2  1  1  1  1  2
  4  4  1  1  1  1  1  0  1  1  1  1
  5  5  1  2  2  1  2  1  1  1  1  2
  6  6  0  1  1  0  1  1  0  0  0  1
  7  7  1  1  1  1  1  0  1  1  1  1
  8  8  1  1  1  1  1  0  1  1  1  1
  9  9  1  1  1  1  1  0  1  1  1  1
 10 10  1  2  2  1  2  1  1  1  1  2

HIERARCHICAL CLUSTERING OF OVERLAP MATRIX

        C  I  N  U  W     M  E  C  W
        O  N  E  W  E  W  A  D  O  E
        U  D  W  A  L  R  Y  U  M  S
        N  U  S  Y  F  O  R  C  M  T

                                    1
Level   1  4  7  8  9  6  5  3  2  0
-----   -  -  -  -  -  -  -  -  -  -
2.000   .  .  .  .  .  .  XXXXXXXXXXX
1.000   XXXXXXXXXXXXX  XXXXXXXXXXXXXX
0.833   XXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

The cliques that we saw before have been made more inclusive by the relaxed definition of group membership. The first N-clique includes everyone but actor 6. The second is more restricted, and includes 6 (WRO), along with two elements of the core. With larger and fewer subgroups, the mayor (5) no longer appears to be quite so critical. With the more relaxed definition, there is now an “inner circle” of actors that are members of both larger groupings. This can be seen in the co-membership matrix and by clustering.

In some cases, N-cliques can be found that have a property that is undesirable for many purposes: it is possible for members of N-cliques to be connected by actors who are not, themselves, members of the clique. For most sociological applications, this is quite troublesome. To overcome this problem, some analysts have suggested a related grouping, the N-clan. Members of the “clan” are all connected at a distance *n* (or less), and all intermediate actors must also be members of the clan.

The K-plex is an alternative way of relaxing the requirement for clique membership (where members form a maximally complete subgraph). It allows actors to be members of a clique if they have ties to all but *k* other members. For example, if A has ties with B and C, but not D; while both B and C have ties with D, all four actors could still be in a clique under the K-plex approach, as in Figure 24.14.

Figure 24.14 Illustration of a K-plex (K = 2)



This approach says that a node is a member of a clique of size *n* if it has direct ties to *n–k* members of that clique. In Figure 24.15, we have allowed *k* to be equal to 2 but insisted that each K-plex should include at least four members.

Figure 24.15 K-plex groups in Knoke reciprocity-symmetric information network

```
15 k-plexes found.

   1:   COUN COMM EDUC MAYR WEST
   2:   COUN COMM INDU MAYR
   3:   COUN COMM MAYR NEWS
   4:   COUN COMM MAYR UWAY
   5:   COUN COMM MAYR WELF
   6:   COMM EDUC INDU MAYR
   7:   COMM EDUC MAYR NEWS
   8:   COMM EDUC MAYR UWAY
   9:   COMM EDUC MAYR WELF
  10:   COMM INDU MAYR NEWS
  11:   COMM INDU MAYR UWAY
  12:   COMM INDU MAYR WELF
  13:   COMM MAYR NEWS UWAY
  14:   COMM MAYR NEWS WELF
  15:   COMM MAYR UWAY WELF
```

```
             1  2  3  4  5  6  7  8  9 10
            CO CO ED IN MA WR NE UW WE WE
            -- -- -- -- -- -- -- -- -- --
   1  1      5  5  1  1  5  0  1  1  1  1
   2  2      5 15  5  5 15  0  5  5  5  1
   3  3      1  5  5  1  5  0  1  1  1  1
   4  4      1  5  1  5  5  0  1  1  1  0
   5  5      5 15  5  5 15  0  5  5  5  1
   6  6      0  0  0  0  0  0  0  0  0  0
   7  7      1  5  1  1  5  0  5  1  1  0
   8  8      1  5  1  1  5  0  1  5  1  0
   9  9      1  5  1  1  5  0  1  1  5  0
  10 10      1  1  1  0  1  0  0  0  0  1

HIERARCHICAL CLUSTERING OF OVERLAP MATRIX

              I E C M C N U W W
            W N D O A O E W E E
            R D U U Y M W A L S
            O U C N R M S Y F T

                              1
 Level      6 4 3 1 5 2 7 8 9 0
------      - - - - - - - - - -
15.000      . . . . XXX . . . .
 5.000      . . . XXXXX . . . .
 4.000      . . XXXXXXX . . . .
 3.400      . XXXXXXXXX . . . .
 3.000      . XXXXXXXXXXX . . .
 1.909      . XXXXXXXXXXXXX . .
 1.420      . XXXXXXXXXXXXXXX .
 0.082      . XXXXXXXXXXXXXXXXX
 0.000      XXXXXXXXXXXXXXXXXXX
```

The K-plex method of defining cliques tends to find “overlapping social circles” when compared to the maximal or N-clique method. The K-plex approach to defining substructures makes a good deal of sense for many problems. It requires that members of a group have ties to (most) other group members and that a tie by way of nonclique intermediaries (which are permissible in the N-clique approach) does not qualify a node for membership. The picture of group structure that emerges from K-plex approaches can be rather different from that of N-clique analysis.

K-cores are a variation on K-plexes that may be particularly helpful with larger numbers of actors. The K-core approach allows actors to join the group if they are connected to *k* members, regardless of how many other members they may not be connected to. The K-core definition is intuitively appealing for some applications. If an actor has ties to a sufficient number of members of a group, they may feel tied to that group even if they don't know many (or even most) members. It may be that identity depends on connection, rather than on immersion in a subgroup.

Top-Down Approaches: Component, Cutpoint, Block, and Faction

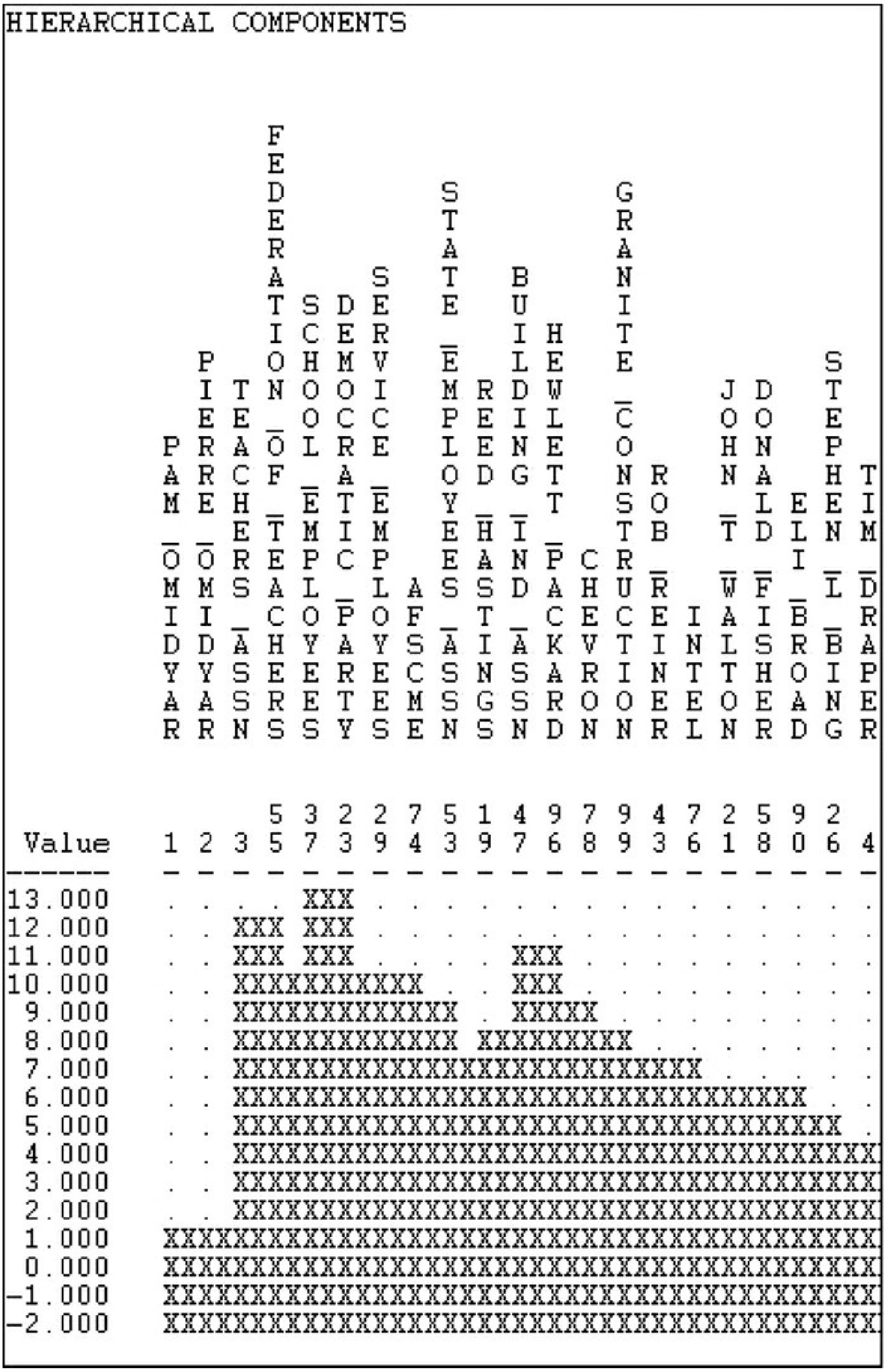
The approaches we've examined to this point start with the dyad, and extend this kind of tight structure outward. Overall structure of the network is seen as “emerging” from overlaps and couplings of smaller components. Alternatively, one might start with the entire network and identify “substructures” as parts that are locally denser or thinner than the field as a whole. Places where the social fabric is more thinly woven may define lines of division or cleavage in the network and can point to how it might be decomposed into smaller units. This top-down perspective leads us to think of dynamics that operate at the level of group selection and to focus on the constraints under which actors construct networks.

There are numerous ways that one might define the divisions and “weak spots” in a network. Below are some of the most common approaches.

Components of a graph are subgraphs that are connected within – but disconnected between – subgraphs. If a graph contains one or more “isolates,” these actors are components. More interesting components are those that divide the network into separate parts with each having several actors. For directed graphs we can define two different kinds of components. A weak component is a set of nodes that is connected, regardless of the direction of ties. A strong component requires that there be a directed path from A to B in order for the two to be in the same component.

Because the Knoke information network has a single component, it isn't very interesting as an example. Let's look instead at the network of large donors to California political campaigns, where the strength of the relation between two actors is defined by the number of times that they contributed on the same side of an issue (Figure 24.16).

Figure 24.16 Weak component hierarchy for California political donors



If we set a very high cut-off value of 13 issues in common to define membership in the same component, then our graph has only nonisolate components (made up of the Democratic Party and the School Employees union). Progressively lower cut-offs produce multiple, separate components until we reach a value of seven issues in common. At this point, the nonisolated nodes all become connected into a single component.

Blocks and cutpoints (bi-components) are an alternative approach to finding the key "weak" spots in the graph. If a node were removed, would the structure become divided into unconnected parts? If there are such nodes, they are called "cutpoints." One can imagine that such cutpoints may be particularly important actors. The divisions into which cutpoints divide a graph are called blocks (not the same usage of the term as in "blockmodels" or "block images"). Another name for a block is a "bi-component." We apply the bi-component idea to the Knoke data in Figure 24.17.

Figure 24.17 Cutpoints and blocks in the Knoke information network

```
2 blocks found.


BLOCKS:

Block     1:  EDUC WRO
Block     2:  COUN COMM EDUC INDU MAYR NEWS UWAY WELF WEST

Articulation points

                 1
          CutPoint
          --------
  1  1           0
  2  2           0
  3  3           1
  4  4           0
  5  5           0
  6  6           0
  7  7           0
  8  8           0
  9  9           0
 10 10           0
```

Two blocks are identified, with EDUC a member of both. This means that if EDUC (node 3) were removed, the WRO would become isolated.

Moody and White (2003) provide new algorithms for identifying nested cut-sets, and make a strong case for the close correspondence of their approach to graph substructure to the concept of "structural cohesion." Their approach identifies hierarchies of nested cohesive groups and is particularly sensitive to identifying the robustness of groups in the face of the removal of individual nodes, and the identification of K-components (maximal K-connected subgraphs).

Lambda sets and bridges are alternative approaches to the issue of connectivity. Here we ask if there are certain *connections* (rather than nodes) in the graph that, if removed, would result in a disconnected structure. In our example, the only relationship that qualifies is that between EDUC and WRO. The Lambda set approach ranks each of the relationships in the network in terms of importance by evaluating how much of the flow among actors in the net goes through each link. It then identifies sets of relationships, which, if disconnected, would most greatly disrupt the flow among all of the actors. The math and computation is rather extreme, though the idea is fairly simple. We apply this to our Knoke data in Figure 24.18.

Figure 24.18 Lambda sets in the Knoke information network

```
LAMBDA SETS

HIERARCHICAL LAMBDA SET PARTITIONS

                U W C E I M C N W
              W W E O D N A O E E
              R A L U U D Y M W S
              O Y F N C U R M S T

                                1
Lambda        6 8 9 1 3 4 5 2 7 0
------        - - - - - - - - - -
     7        . . . . . . XXX . .
     3        . . . XXXXXXXXXXXXX
     2        . XXXXXXXXXXXXXXXXX
     1        XXXXXXXXXXXXXXXXXXX
```

This approach identifies the 2 to 5 (MAYR to COMM) linkage as the most important one in the graph: it carries a great deal of traffic, and the graph would be most disrupted if it were removed.

M. E. J. Newman (2006) has advanced the closely related idea of “modularity” as an approach to identifying substructures in graphs. In Newman's approach, substructures are defined by having more ties within, and fewer ties between, groups than would be expected on the basis of the degrees of the nodes. This is an important advance on earlier approaches, which seek to minimize the number of bridging ties between groups but do not take account of group size or node degree.

Factions: Imagine a society in which each person was closely tied to all others in their own subpopulation (i.e., all subpopulations are cliques), and there are no connections at all between subpopulations (i.e., each subpopulation is a component). Most real populations do not look like this, but the “ideal type” of complete connection within and complete disconnection between subgroups is a useful reference point for assessing the degree of “factionalization” in a population.

If we took all the members of each “faction” in this ideal-typical society and put their rows and columns together in an adjacency matrix (i.e., permuted the matrix so all members of the same group occupied adjacent rows and columns), we would see a distinctive pattern of “1-blocks” and “0-blocks.” All connections among actors within a faction would be present; all connections between actors in different factions would be absent. With valued data, the average tie strength within a block would be high; the average tie strength between blocks would be low.

We applied this idea to the Knoke data. After running several alternative numbers of blocks, we settled on four as meaningful for our purposes. This result is shown in Figure 24.19.

Figure 24.19 Four-faction solution for the directed Knoke information network

```
Initial number of errors:  39

  Number of errors:  27
  Number of errors:  27
  Number of errors:  27
  Number of errors:  29
  Number of errors:  27
  Number of errors:  27
  Number of errors:  27
  Number of errors:  27
  Number of errors:  27
  Number of errors:  27

Final number of errors:  27

Group Assignments:

    1:   1 2 4 5 7 8
    2:   10
    3:   3 6
    4:   9

Grouped Adjacency Matrix

                    1
      1 2 8 4 5 7   0   6 3   9
      C C U I M N   W   W E   W
     ---------------------------
  1 |   1     1 1 |   |     | 1 |
  2 | 1   1 1 1 1 |   |   1 | 1 |
  8 | 1 1   1 1 1 |   |     | 1 |
  4 | 1 1     1 1 |   |     |   |
  5 | 1 1 1 1   1 | 1 |   1 | 1 |
  7 |   1   1 1   |   |     |   |
```

```
     -----------------------------
 10 | 1 1     1 1 |   |   1 |   |
     -----------------------------
  6 |           1 |   |   1 | 1 |
  3 |   1   1 1 1 | 1 | 1   |   |
     -----------------------------
  9 |   1     1 1 |   |     |   |
     -----------------------------
```

Density Table

	1	2	3	4
1	0.83	0.17	0.17	0.67
2	0.67		0.50	0.00
3	0.42	0.50	1.00	0.50
4	0.50	0.00	0.00	

The "final number of errors" can be used as a measure of the "goodness of fit" of the "blocking" of the matrix. This count (27) is the sum of the number of zeros within factions (where all the ties are supposed to be present in the ideal type) plus the number of ones in the nondiagonal blocks (ties between members of different factions, which are supposed to be absent in the ideal type). Since there are 49 total ties in our data, being wrong 27 times is not a terribly good fit. It is, however, the best we can do with four "factions." The four factions are identified, and we note that two of them are individuals (10, 9), and one is a dyad (3, 6).

The "blocked" or "grouped" adjacency matrix shows a picture of the solution. We can see that there is quite a lot of density "off the main diagonal" where there shouldn't be any. The final panel of the results reports the "block densities" as the number of ties that are present in blocks as proportions of all possible ties.

Core-periphery and other blockmodels extend the idea of factions to identify groups or "types" (or, technically, equivalence classes) of cases based on their patterns of ties (Boyd et al., 2006; Clark, 2008; Raval and Kral, 2004). Patterns of core and periphery are often found in sociological data; in this blockmodel, there are many ties among members of the core, few ties among members of the periphery, and some ties (definitions vary) between core and periphery members. Grouping cases into types based on similarity of their positions or roles in the graph has proven to be one of the most important approaches to identifying substructures in social structures. Ferligoj et al. (this volume) cover this topic in depth.

The "Embedded" Individual: Ego Networks

The approaches to exploring networks that we've examined so far tend to be views from the "top down." That is, they focus attention on the whole network's structure, texture, and substructures. For many problems, it can be useful to view social networks from the "bottom up," focusing attention on individuals and their connections. Describing and indexing the variation across individuals in the way they are embedded in "local" social structures is the goal of the analysis of *ego networks*. We need some definitions.

"Ego" is an individual "focal" node. A complete network has as many egos as it has nodes. However, our data may also consist of one or many ego networks that are not connected to one another. Egos can be persons, groups, organizations, or whole societies.

A one-step neighborhood consists of ego and all nodes to whom ego has a direct connection. Importantly, the neighborhood also includes all of the ties among all of the actors to whom ego has a connection. Neighborhoods of greater path length than one are rarely used in social network analysis. When we use the term "neighborhood" here, we mean the one-step neighborhood. The N-step neighborhood expands the definition of the size of ego's neighborhood by including all nodes to whom ego has a connection at a path length of *N* or less, and all the connections among all of these actors.

In and Out, and Other Kinds of Neighborhoods

Most of the analyses of ego networks use simple graphs (i.e., graphs that are symmetric and show only the presence or absence of connections, but not their direction). If we are working with a directed graph, it is possible to define different kinds of ego-neighborhoods. An *out neighborhood* would include all the actors to whom ties are directed *from* ego. An *in neighborhood* would include all the actors who send ties directly to ego. It is also possible to define a neighborhood of only those actors with whom ego has reciprocated ties. These are just a few of the ways of defining ego neighborhoods; there isn't a single "right" way for every research question.

Strong and Weak Tie Neighborhoods

Most analyses of ego networks use binary data (actors are connected or they aren't), which makes defining the ego neighborhood fairly straightforward. If we have measured the strength of the relation between two actors or its valence (positive or negative), however, we need to make choices about the definition of "neighbor." With ties that are measured as strengths or probabilities, a reasonable approach is to define a cut-off value (or, better, explore several reasonable alternatives). When ties are characterized as positive or negative, the most common approach is to analyze the positive tie neighborhood and the negative tie neighborhood separately.

Ego network data commonly arise in two ways: surveys may be used to collect information on ego networks. We can ask each research subject to identify all of the actors to whom they have a connection, and to report to us (as an informant) the links among these other actors. Alternatively, we could use a snowball method: first ask ego to identify others to whom ego has a tie, then ask each of those identified about their ties to additional others. With each stage, the size of the network increases, until all members of the component originally sampled have been included.

Data collected in this way cannot directly inform us about the overall embeddedness of the networks in a population, but it can tell us about the prevalence of various kinds of ego networks in even very large populations. This kind of investigation results in a data structure that is composed of a collection of networks. As the actors in each network are likely to be different, the networks need to be treated as separate actor-by-actor matrices stored as different data sets.

The second major way in which ego network data arise is by “extracting” them from regular complete network data. This is the approach that we will take in our example. Rather than treating the Knoke information exchange network as a single network, we will treat it at 10 ego networks (which happen to be connected and overlapping). One might, for example, extract all the ego networks from a full network where ego was male and compare their structures to all the ego networks where ego was female. When you create a sample of ego networks by extracting them from a single full network, you need to remember that these are not independent samples from a population, and normal statistical sampling assumptions don't apply.

Connections

There are quite a few characteristics of the ego-neighborhoods that may be of interest. Figure 24.20 displays a collection of many of the most commonly used measures of the texture of ego's neighborhood. In this example, we are looking at the one-step out neighborhood of each of the 10 egos in the Knoke information exchange data. That is, each ego's neighborhood is defined by those actors to whom ego sends information. A parallel analysis of in neighborhoods might also be of interest.

Figure 24.20 Ego network connections for Knoke information out neighborhoods

Density Measures

	1 Size	2 Ties	3 Pairs	4 Densit	5 AvgDis	6 Diamet	7 nWeakC	8 pWeakC	9 2StepR	10 ReachE	11 Broker	12 nBroke	13 EgoBet	14 nEgoBe
1	4.00	11.00	12.00	91.67	1.08	2.00	1.00	25.00	100.00	29.03	0.50	0.04	0.00	0.00
2	7.00	24.00	42.00	57.14	1.43	2.00	1.00	14.29	100.00	18.75	9.00	0.21	8.17	19.44
3	6.00	17.00	30.00	56.67			1.00	16.67	100.00	23.08	6.50	0.22	8.25	27.50
4	4.00	11.00	12.00	91.67	1.08	2.00	1.00	25.00	100.00	28.13	0.50	0.04	0.33	2.78
5	8.00	29.00	56.00	51.79	1.57	3.00	1.00	12.50	100.00	16.98	13.50	0.24	14.67	26.19
6	3.00	2.00	6.00	33.33			1.00	33.33	100.00	42.86	2.00	0.33	1.00	16.67
7	3.00	6.00	6.00	100.00	1.00	1.00	1.00	33.33	88.89	36.36	0.00	0.00	0.00	0.00
8	6.00	24.00	30.00	80.00	1.20	2.00	1.00	16.67	100.00	20.45	3.00	0.10	0.00	0.00
9	3.00	6.00	6.00	100.00	1.00	1.00	1.00	33.33	100.00	36.00	0.00	0.00	0.00	0.00
10	5.00	16.00	20.00	80.00	1.20	2.00	1.00	20.00	100.00	23.68	2.00	0.10	0.33	1.67

1. Size. Size of ego network.
2. Ties. Number of directed ties.
3. Pairs. Number of ordered pairs.
4. Density. Ties divided by Pairs.
5. AvgDist. Average geodesic distance.
6. Diameter. Longest distance in egonet.
7. nWeakComp. Number of weak components.
8. pWeakComp. NWeakComp divided by Size.
9. 2StepReach. # of nodes within 2 links of ego.
10. ReachEffic. 2StepReach divided Size.
11. Broker. # of pairs not directly connected.
12. Normalized Broker. Broker divided by number of pairs.
13. Ego Betweenness. Betweenness of ego in own network.
14. Normalized Ego Betweenness. Betweenness of ego in own network.

Some measures of the structure of ego networks are parallel to those for complete networks. Many others, though, reflect the particular interests of “bottom-up” analysis, and describe ego's opportunities and constraints due to how they are embedded in their local structure of connections.

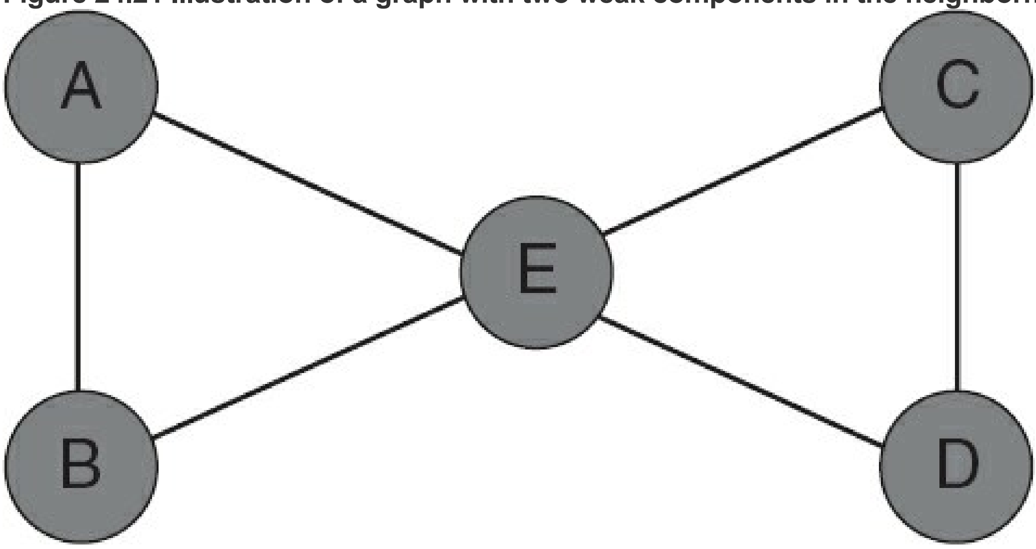
The size of an ego network is the number of nodes that are one-step neighbors of ego, plus ego itself. Actor 5 has the largest ego network; actors 6, 7, and 9 have the smallest networks. The number of directed ties is the number of connections among all the nodes in the ego network. Among the four actors in ego 1's network, there are 11 ties. The number of ordered pairs is the number of possible directed ties in each ego network. In node 1's network there are four actors, so there are 4*3 possible directed ties. The density is the number of actual ties divided by the number of pairs (i.e., possible ties). Note that actors 7 and 9 live in neighborhoods where all actors send information to all other actors; they are embedded in very dense local structures. The welfare rights organization (node 6) lives in a small world where the members are not tightly connected. This kind of difference in the constraints and opportunities facing actors in their local neighborhoods may be very consequential, as we shall see in examining “structural holes” below.

The average geodesic distance is the mean of the shortest path lengths among all connected pairs in the ego network. Where everyone is directly connected to everyone (e.g., nodes 7 and 9), this distance is one. In our example, the largest average path length for connected neighbors is for actor 5 (average distances among members of the neighborhood is 1.57). The diameter of an ego network is the length of the longest path between connected actors (just as it is for any network). The idea of a network diameter is to index the span or extensiveness of the network: how far apart are the two furthest actors? In the current example, they are not very far apart in the ego networks of most actors.

The size, density, and distances in an ego neighborhood are very much parallel to the same ideas for whole networks. In addition to these fairly basic and reasonably straightforward measures, ego network analysts have developed a number of approaches to understanding the role that ego plays in connecting the neighborhood and to understanding ego's positional advantage and disadvantage.

One interesting feature is the extent to which ego's neighborhood consists of separate components of factions. To what extent does ego play a critical role in connecting others? A weak component is the largest number of actors who are connected, disregarding the direction of the ties (a strong component pays attention to the direction of the ties for directed data). In Figure 24.21, if ego (E) was connected to A and B (who are connected to one another), and ego is connected to C and D (who are connected to one another), but A and B are not connected in any way to C and D (except by way of everyone being connected to ego) then there would be two “weak components” in ego's neighborhood.

Figure 24.21 Illustration of a graph with two weak components in the neighborhood of actor E



In our example, there are no such cases – each ego is embedded in a single component neighborhood. That is, there are no cases where ego is the only connection between otherwise disjointed sets of actors. The likelihood that there would be more than one weak component in ego's neighborhood would be a function of neighborhood size if connections were random. So, to get a sense of whether ego's role in connecting components is “unexpected” given the size of the network, it is useful to normalize the count of components by size. In our example, since there are no cases of multiple components, this is a pretty meaningless exercise.

The two-step reach goes beyond ego's one-step neighborhood to report the percentage of all actors in the whole network that are within two directed steps of ego. In our example, only node 7 cannot get a message to all other actors within “friend-of-a-friend” distance. The reach efficiency (two-step reach divided by size) norms the two-step reach by dividing it by size. The idea here is, how much (nonredundant) secondary contact do I get for each unit of primary contact? If reach efficiency is high, then I am getting a lot of “bang for my buck” in reaching a wider network for each unit of effort invested in maintaining a primary contact. On the other hand, if I share many contacts with my neighbors, I have low efficiency.

Ego may be the “go-between” for pairs of other actors. In an ego network, ego is connected to every other actor (by definition). If these others are not connected directly to one another, ego may be a “broker” if ego falls on the paths between the others. One item of interest is simply how much potential for brokerage there is for each actor (how many times pairs of neighbors in ego's network are not directly connected). In our example, actor 5, who is connected to almost everyone, is in a position to broker many connections. Normalized brokerage (brokerage divided by number of pairs) assesses the extent to which ego's role is that of a broker. One can be in a brokering position a number of times, but this is a small percentage of the total possible connections in a network (e.g., the network is large). Given the large size of actor 5's network, the relative frequency with which actor 5 plays the broker role is not so exceptional.

Betweenness is an aspect of the larger concept of “centrality.” In an ego network, ego is “between” two other actors if ego lies on the shortest directed path from one to the other. The ego betweenness measure indexes the percentage of all geodesic paths from neighbor to neighbor that pass through ego. Normalized betweenness compares the actual betweenness of ego to the maximum possible betweenness in neighborhood of the size and connectivity of ego's. The “maximum” value for betweenness would be achieved where ego is the center of a “star” network; that is, no neighbors communicate directly with one another, and all directed communications between pairs of neighbors go through ego.

The ideas of brokerage and betweenness are slightly differing ways of indexing just how central or powerful ego is within its own neighborhood. This aspect of how an actor's embedding may provide it with strategic advantage has received a great deal of attention. The next two sections, on structural holes and brokerage elaborate on ways of looking at positional opportunity and constraint of individual actors.

Structural Holes

Ronald Burt (1992) coined and popularized the term “structural holes” to refer to some very important aspects of positional advantage and disadvantage of individuals that result from how they are embedded in neighborhoods. Burt's formalization of these ideas, and his development of a number of measures (including the computer program *Structure*, which provides these measures and other tools), has facilitated a great deal of further thinking about how and why the ways that an actor is connected affect their constraints and opportunities, and hence their behavior.

The basic idea is simple, as good ideas often are. Imagine a network of three actors (A, B, and C), in which each is connected to each of the others. Suppose that actor A wanted to influence or exchange with another actor. Assume that both B and C may have some interest in interacting or exchanging, as well. Actor A will not be in a strong bargaining position in this network, because both of A's potential exchange partners (B and C) have alternatives to treating with A; they could isolate A and then exchange only with one another.

Now imagine that we open a "structural hole" between actors B and C. That is, a relation or tie is "absent" such that B and C cannot exchange (perhaps they are not aware of one another, or there are very high transaction costs involved in forming a tie). In this situation, actor A has an advantaged position because he or she has two alternative exchange partners; actors B and C have only one choice, if they choose to (or must) enter into an exchange. Ego A now has power with respect to two dependent alters and is not constrained by the threat of being excluded from an exchange opportunity.

Burt developed a number of measures related to structural holes that can be computed on both valued and binary data. The normal practice in sociological research has been to use binary (a relation is present or not). Interpretation of the measures becomes quite difficult with valued data. The structural holes measures may be computed for either directed or undirected data – and the interpretation, of course, depends on which is used. Here, we've used the directed binary data. Figure 24.22 shows UCINET output for a "structural holes" analysis of the neighborhoods of each of our 10 egos.

Figure 24.22 Structural holes analysis for Knoke information exchange ego networks

Dyadic redundancy

	1 COUN	2 COMM	3 EDUC	4 INDU	5 MAYR	6 WRO	7 NEWS	8 UWAY	9 WELF	10 WEST
1	0.00	0.72	0.00	0.61	0.78	0.00	0.72	0.61	0.56	0.39
2	0.43	0.00	0.33	0.47	0.87	0.00	0.57	0.40	0.33	0.40
3	0.00	0.50	0.00	0.50	0.60	0.05	0.70	0.00	0.00	0.35
4	0.61	0.78	0.56	0.00	0.78	0.00	0.61	0.61	0.00	0.00
5	0.44	0.81	0.38	0.44	0.00	0.00	0.56	0.38	0.31	0.31
6	0.00	0.00	0.13	0.00	0.00	0.00	0.38	0.00	0.13	0.00
7	0.54	0.71	0.58	0.46	0.75	0.13	0.00	0.50	0.46	0.38
8	0.69	0.75	0.00	0.69	0.75	0.00	0.75	0.00	0.63	0.00
9	0.63	0.63	0.00	0.00	0.63	0.06	0.69	0.63	0.00	0.00
10	0.50	0.86	0.50	0.00	0.71	0.00	0.64	0.00	0.00	0.00

Dyadic Constraint

	1 COUN	2 COMM	3 EDUC	4 INDU	5 MAYR	6 WRO	7 NEWS	8 UWAY	9 WELF	10 WEST
1	0.00	0.13	0.00	0.04	0.15	0.00	0.06	0.04	0.04	0.03
2	0.05	0.00	0.04	0.05	0.11	0.00	0.06	0.04	0.04	0.02
3	0.00	0.09	0.00	0.03	0.10	0.04	0.06	0.00	0.00	0.06
4	0.04	0.13	0.03	0.00	0.13	0.00	0.10	0.04	0.00	0.00
5	0.05	0.09	0.04	0.04	0.00	0.00	0.06	0.04	0.03	0.03
6	0.00	0.00	0.27	0.00	0.00	0.00	0.11	0.00	0.07	0.00
7	0.03	0.10	0.04	0.06	0.11	0.01	0.00	0.03	0.03	0.02
8	0.05	0.15	0.00	0.05	0.15	0.00	0.06	0.00	0.05	0.00
9	0.05	0.13	0.00	0.00	0.13	0.02	0.06	0.05	0.00	0.00
10	0.04	0.08	0.12	0.00	0.17	0.00	0.06	0.00	0.00	0.00

Structural Hole Measures

	1 EffSize	2 Efficie	3 Constra	4 Hierarc
1	2.611	0.373	0.481	0.103
2	4.200	0.525	0.401	0.052
3	3.300	0.550	0.386	0.044
4	2.056	0.343	0.479	0.082
5	4.375	0.547	0.387	0.032
6	2.375	0.792	0.454	0.139
7	4.500	0.500	0.424	0.097
8	1.750	0.292	0.514	0.079
9	2.750	0.458	0.436	0.101
10	1.786	0.357	0.486	0.072

Dyadic redundancy means that ego's tie to an alter is “redundant.” If A is tied to both B and C, and B is tied to C, A's tie to B is redundant, because A can influence B by way of C. The dyadic redundancy measure calculates, for each actor in ego's neighborhood, how many of the other actors in the neighborhood are also tied to a given alter. The larger the proportion of others in the neighborhood who are tied to a given “alter,” the more “redundant” is ego's direct tie. In the example, we see that actor 1's (COUN) tie to actor 2 (COMM) is largely redundant, as 72% of ego's other neighbors also have ties with COMM. Actors that display high dyadic redundancy are actors who are embedded in local neighborhoods where there are few structural holes.

Dyadic constraint is a measure that indexes the extent to which the relationship between ego and each alter in ego's neighborhood “constrains” ego. A full description is given in Burt's 1992 monograph, and the construction of the measure is somewhat complex. At the core though, A is constrained by its relationship with B to the extent that A does not have many alternatives (has few other ties except that to B), and A's other alternatives are also tied to B. If A has few alternatives to exchanging with B, and if those alternative exchange partners are also tied to B, then B is likely to constrain A's behavior. In our example, constraint measures are not very large, as most actors have several ties. COMM and MAYR are, however, exerting constraint over a number of others and are not very constrained by them. This situation arises because COMM and MAYR have considerable numbers of ties, and many of the actors to whom they are tied do not have many independent sources of information.

The effective size of the network (EffSize) is the number of alters that ego has, minus the average number of ties that each alter has to other alters. Suppose that A has ties to three other actors. Suppose that none of these three has ties to any of the others. The effective size of ego's network is three. Alternatively, suppose that A has ties to three others and that all of the others are tied to one another. A's network size is three, but the ties are “redundant” because A can reach all three neighbors by reaching any one of them. The average degree of the others in this case is two (each alter is tied to two other alters). So, the effective size of the network is its actual size (three), reduced by its redundancy (two), to yield an efficient size of one.

The efficiency (Efficie) norms the effective size of ego's network by its actual size; that is, it measures the proportion of ego's ties to its neighborhood that are “nonredundant.” The effective size of ego's network may tell us something about ego's total impact; efficiency tells us how much impact ego is getting for each unit invested in using ties. An actor can be effective without being efficient, and an actor can be efficient without being effective.

The constraint (Constra) is a summary measure of the extent to which ego's connections are to others who are connected to one another. If ego's potential trading partners all have one another as potential trading partners, ego is highly constrained. If ego's partners do not have other alternatives in the neighborhood, they cannot constrain ego's behavior. The logic is pretty simple, but the measure itself is not (see Burt's 1992 book). The idea of constraint is an important one because it points out that actors who have many ties to others may actually lose freedom of action rather than gain it, depending on the relationships among the other actors. This is the same basic insight as Bonacich's analysis of the difference between influence and power.

The hierarchy is another quite complex measure that describes the nature of the constraint on ego. If the total constraint on ego is concentrated in a single other actor, the hierarchy measure will have a higher value; if the constraint arises from multiple actors in ego's neighborhood, hierarchy will have a lower value. The hierarchy measure does not assess the degree of constraint directly but, given some level of constraint on ego, it measures an important property of dependency: inequality in the distribution of constraints on ego across the alters in its neighborhood.

Brokerage Among Groups

The extent to which ego is “between” alters is the focus of brokerage, betweenness, and structural holes analyses; it is a major theme in the analysis of ego networks. Gould and Fernandez (1989) extended these ideas in an interesting way by taking into account the possibility that egos and alters might also be affiliated with social groups.

Suppose that ego's network is composed of both men and women (or any qualitative difference). We might be interested in the extent to which ego is “between” men and women in the network, rather than simply whether ego has high betweenness, overall. Gould and Fernandez's “brokerage” notions examine ego's relations with its neighborhood from the perspective of ego acting as an agent in relations among groups (or categories).

To examine the brokerage roles played by a given actor, we find every instance where that actor lies on the directed path between two others. So each actor may have many opportunities to act as a “broker.” For each one of the instances where ego is a broker, we examine which *kinds* of actors are involved. That is, what are the group memberships of each of the three actors? There are five possible combinations.

If ego falls on a directed path between two members of the same category as themselves (e.g., a woman falling between two other women in a path), ego is called a coordinator. If ego falls on the path between two members of a group of which they are not a part (e.g., a man falling on a path from one woman to another), the members are called consultants. If ego falls on the path from a member of another group to a member of its own group (e.g., ego, a man, falls on a path from a woman to another man), the ego is called a gatekeeper. If ego falls on the path from another member of its own group to a member of another group, ego is a representative. Lastly, if ego falls on a path from a member of one group to another but is not a member of either of those groups, ego is a liaison.

As an example, we've taken the Knoke information exchange network and have classified each of the organizations as either a general government organization (group 1), a private nonwel-fare organization (group 2), or an organizational specialist (group 3).

Figure 24.23 shows the results of a basic analysis of brokerage roles for each of the 10 egos in the Knoke directed information network.

Figure 24.23 Brokerage role scores for the Knoke information network

	1 Coordinat	2 Gatekeepe	3 Represent	4 Consultan	5 Liaison	6 Total
1	0	0	0	1	1	2
3	0	1	1	2	5	9
5	2	6	5	5	9	27
2	0	3	7	5	6	21
4	0	0	1	1	0	2
7	0	5	0	0	1	6
6	0	1	0	0	0	1
8	0	0	0	0	0	0
9	0	0	2	0	0	2
10	0	0	0	1	0	1

The actors have been grouped together into “partitions” for presentation; actors 1, 3, and 5, for example, are the general government organizations group. Each row counts the raw number of times that each actor plays each of the five roles in the whole graph. While we have analyzed the entire graph here, the analysis could be restricted to the one-step neighborhood of each ego. Two actors (5 and 2) are the main sources of interconnection among the three organizational populations. Organizations in the third population (6, 8, 9, 10), the welfare specialists, have overall low rates of brokerage. Organizations in the first population (1, 3, 5), the government organizations, seem to be more heavily involved as liaisons than as other roles. Organizations in the second population (2, 4, 7), nongovernmental generalists, play more diverse roles. Overall, there is very little coordination within each of the populations.

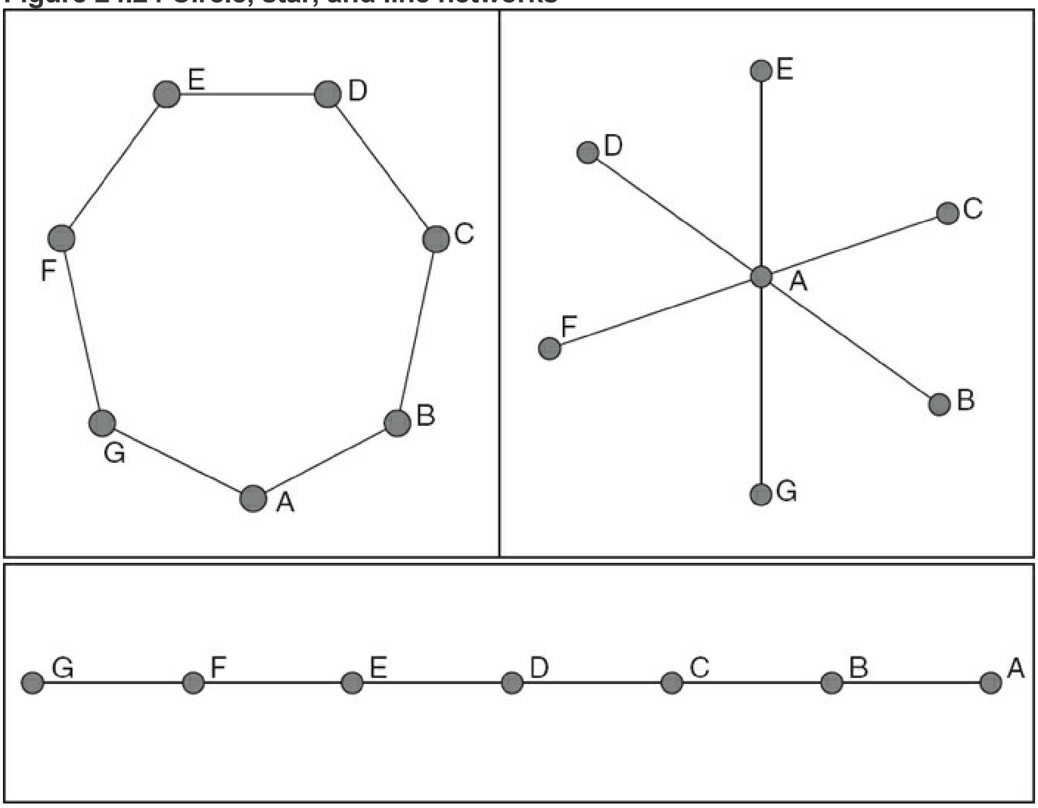
Centrality

Network analysts often describe the way that an actor is embedded in a relational network as imposing constraints on the actor and as offering the actor opportunities (Granovetter, 1982). Actors that face fewer constraints, and have more opportunities than others, are in favorable structural positions. Having a favored position means that an actor may extract better bargains in exchanges, have greater influence, and may be a focus for deference and attention from those in less favored positions.

The question of what we mean by having a "favored position," "more opportunities," or "fewer constraints" has no single correct and final answer. As we have seen above, having structural holes in one's neighborhood may confer advantage; being in a position to act as a broker between substructures may also provide a structurally favorable position.

The most widely used approach to understanding the structural sources of individuals' advantage and disadvantage relative to their neighbors is that of "centrality." The core idea is very simple: actors who are more "central" to social structures are more likely to be influential or powerful (but possibly also more constrained). But the simple idea of being central turns out to not be so simple. Consider the situation of actor A in the three simple networks in Figure 24.24.

Figure 24.24 Circle, star, and line networks



A moment's inspection suggests that actor A has a highly favored structural position in the star network (upper right). The star network shows a neighborhood of maximum inequality: A is central, and everyone else is equally peripheral. In the circle network (upper left), all actors in the neighborhood are in equivalent positions, and A is no more or less central than anyone else. In the line network at bottom, A would seem to be marginalized, and the overall distribution of advantage in the neighborhood is between the two extremes of the star and circle.

But what are the sources of the advantage or disadvantage of the egos in the figures above? The centrality of an ego relative to its alters has been approached in three major ways by network analysts. One approach focuses on the actor's degree. Actors who have more ties, that is, a higher degree (or ties to the “right” others), may be advantaged. Degree-based approaches to centrality are closely connected to the notion of “social capital.” A second approach, based on closeness, argues that egos who can “reach” more alters with less effort have an advantaged position. A third major approach suggests that egos who bridge gaps *between* alters have an advantage.

Degree Centrality: Connectedness

In undirected data, egos differ from one another only in how many connections they have. With directed data, however, it can be important to distinguish centrality based on in-degree from centrality based on out-degree. If an actor receives many ties, they are often said to be prominent, or to have high prestige. That is, many other actors seek direct ties to them, and this may indicate their importance. Actors who have unusually high out-degrees are actors who are able to exchange with many others or to make many others aware of their views. Actors who display high out-degree centrality are often said to be influential actors.

Linton Freeman (1979) developed basic measures of the centrality of actors based on their degree. Figure 24.25 shows the out-degree and in-degree centrality of each of the egos in the Knoke information network.

Figure 24.25 Degree centrality of egos in the Knoke information network

	1 OutDegree	2 InDegree	3 NrmOutDeg	4 NrmInDeg
1	4.000	5.000	44.444	55.556
2	7.000	8.000	77.778	88.889
3	6.000	4.000	66.667	44.444
4	4.000	5.000	44.444	55.556
5	8.000	8.000	88.889	88.889
6	3.000	1.000	33.333	11.111
7	3.000	9.000	33.333	100.000
8	6.000	2.000	66.667	22.222
9	3.000	5.000	33.333	55.556
10	5.000	2.000	55.556	22.222

DESCRIPTIVE STATISTICS

		1 OutDegree	2 InDegree	3 NrmOutDeg	4 NrmInDeg
1	Mean	4.900	4.900	54.444	54.444
2	Std Dev	1.700	2.625	18.889	29.165
3	Sum	49.000	49.000	544.444	544.444
4	Variance	2.890	6.890	356.790	850.617
5	SSQ	269.000	309.000	33209.875	38148.148
6	MCSSQ	28.900	68.900	3567.901	8506.173
7	Euc Norm	16.401	17.578	182.236	195.316
8	Minimum	3.000	1.000	33.333	11.111
9	Maximum	8.000	9.000	88.889	100.000

Network Centralization (Outdegree) = 38.272%
Network Centralization (Indegree) = 50.617%

Actors 5 and 2 have the greatest out-degrees and might be regarded as the most influential (though it might matter to whom they are sending information; this measure does not take that into account). Actors 5 and 2 are joined by 7 (the newspaper) when we examine in-degree.

Other organizations might share information with these three in an effort to exert influence. This could be seen as an act of deference, or a recognition that the positions of actors 5, 2, and 7 might be worth trying to influence. If we were interested in comparing influence in networks of different sizes or densities, it might be useful to “standardize” the measures of in- and out-degrees. In the last two columns of the first panel, all the degree counts have been expressed as percentages of the number of actors in the network, less one (ego).

If we are analyzing the egos in a complete network (as we are here) instead of separate ego neighborhoods (as we would if we had collected information about the ego network of a sample of individuals from some population) we can also examine the distribution of ego centralities to learn more about the population as a whole. In Figure 24.22, we see that, on average, actors have a degree of 4.9, which is quite high given that there are only nine other actors. The range of in-degree is slightly larger (minimum and maximum) than that of out-degree, and there is more variability across the actors in in-degree than out-degree (based on their standard deviations and variances). The range and variability of degree (and other network properties) can be quite important, because they describe whether the population is homogeneous or heterogeneous in structural positions. Finally, Freeman's graph centralization measures describe the population as a whole, the macro level. The graph centralization measure expresses the degree of inequality or variance in our network as a percentage of that of a perfect star network of the same size. In the current case, the out-degree graph centralization is 51% and the in-degree graph centralization is 38% of these theoretical maximums. We would arrive at the conclusion that there is a substantial amount of concentration or centralization in this whole network. That is, the power of individual actors varies rather substantially, and this means that, overall, positional advantages are rather unequally distributed in this network.

Degree Centrality: Influence and Power

Phillip Bonacich (1987) proposed a modification of the degree centrality approach. Suppose that Bill and Fred each have five close friends. Bill's friends, however, happen to be pretty isolated folks and don't have many other friends, save Bill. In contrast, Fred's friends each have lots of friends, who have lots of friends, and so on. Who is more central? One argument would be that Fred is likely to be more influential because he can quickly reach a lot of other actors, but if the actors that one is connected to are themselves well connected, they are not highly dependent on you. Bonacich argued that being connected to others who are connected makes an actor influential but not powerful. Somewhat ironically, being connected to others that are not well connected makes one powerful, because these other actors are dependent on you, whereas well-connected actors are not.

Let's examine the Bonacich influence and power positions of the egos in our information exchange data. The Bonacich influence or power indexes are calculated by using an attenuation factor (beta) or weight to show whether the index increases with the degree of those to whom ego is connected (influence) or decreases with the degree of those to whom ego is connected (power). The results for our information exchange network are shown in Figures 24.26 and 24.27.

Figure 24.26 Bonacich influence in the Knoke network (beta = + .50)

Actor Power

	1 Power
1	-2.732
2	-3.938
3	-3.235
4	-2.855
5	-4.428
6	-1.167
7	-2.610
8	-3.526
9	-2.488
10	-3.472

STATISTICS

		1 Power
1	Mean	-3.045
2	Std Dev	0.856
3	Sum	-30.452
4	Variance	0.732
5	SSQ	100.056
6	MCSSQ	7.321
7	Euc Norm	10.003
8	Minimum	-4.428
9	Maximum	-1.167

Figure 24.27 Bonacich power in the Knoke information network (beta = – .50)

Actor Power

	1 Power
1	4.667
2	-9.333
3	12.667
4	6.000
5	-8.000
6	-11.333
7	8.667
8	1.333
9	7.333
10	0.667

STATISTICS

		1 Power
1	Mean	1.267
2	Std Dev	7.828
3	Sum	12.667
4	Variance	61.284
5	SSQ	628.888
6	MCSSQ	612.843
7	Euc Norm	25.078
8	Minimum	-11.333
9	Maximum	12.667

If we look at the absolute value of the index scores, we see a familiar story. Actors 5 and 2 are clearly the most central. This is because they have high degree, and because they are connected to each other and to other actors with high degree. Actors 8 and 10 also appear to have high centrality by this measure; this is a new result. In this case, it is because the actors are connected to all of the other actors with high degree. Actors 8 and 10 don't have extraordinary numbers of connections, but they have “the right connections.”

Let's take a look at the power side of the index, which is calculated by the same algorithm, but gives negative weights to connections with well-connected others and positive weights for connections to weakly connected others.

Not surprisingly, these results are very different from many of the others we've examined. Egos 2 and 6 are distinguished because their ties are mostly ties to alters with high degree, making actors 2 and 6 “weak” by having powerful neighbors. Egos 3, 7, and 9 have more ties to alters who have few ties, making them “strong” by virtue of having weak neighbors.

Closeness Centrality

Degree centrality measures might be criticized because they take into account only an actor's immediate ties, or the ties of the actor's neighbors, rather than indirect ties to all others. One actor might be tied to a large number of others, but those others might be relatively disconnected from the network as a whole. In a case like this, the actor could be quite central but only in a local neighborhood. Of course, if the only data we have available are for the one- or two-step neighborhoods of egos, this is the best we can do. If we have full network data, however, there are a number of influence centrality measures available that are based on overall closeness of each ego to all others in the graph (see Hanneman and Riddle, 2005, Chapter 10). Closeness measures such as the geodesic path distance, eigenvector centrality, reach centrality, Hubbell, Katz, Taylor, and Stephenson and Zelen all extend the influence centrality idea to larger networks (or the full graph).

Betweenness Centrality

Suppose that I want to influence you by sending you information, or I want to make a deal to exchange some resources, but in order to talk to you, I must go through one or more intermediaries. This gives the people who lie “between” me and others power with respect to me. To the extent that I can use multiple pathways to reach others, however, my dependency on any one intermediary is reduced (Cook et al., 1983).

The extent to which connections between alters in a neighborhood depend on ego can vary. In a star network, ego mediates all connections; in a clique, each alter can reach each other without ego's help. The extent of ego's “betweenness” can be an important dimension of relative power.

For networks with binary relations, Freeman (1979) created some measures of the centrality of individual actors based on their geodesic path betweenness, as well as overall graph centralization. Freeman et al. (1991) extended the basic approach to deal with all paths between actors.

With binary data, betweenness centrality views an actor as being in a favored position to the extent that the actor falls on the geodesic paths between other pairs of actors in the network. That is, the more people depend on me to make connections with other people, the more power I have. If, however, two actors are connected by more than one geodesic path, and I am not on all of them, I lose some power. If we add up, for each actor, the proportion of times that they are “between” alters we get a measure of ego “betweenness” centrality. We can norm this measure by expressing it as a percentage of the maximum possible betweenness that an actor could have. The results for the Knoke information network are shown in Figure 24.28.

Figure 24.28 Freeman betweenness for Knoke information network

	1 Betweenness	2 nBetweenness
5	17.833	24.769
2	12.333	17.130
3	11.694	16.242
7	2.750	3.819
9	1.222	1.698
4	0.806	1.119
1	0.667	0.926
10	0.361	0.502
6	0.333	0.463
8	0.000	0.000

DESCRIPTIVE STATISTICS FOR EACH MEASURE

		1 Betweenness	2 nBetweenness
1	Mean	4.800	6.667
2	Std Dev	6.220	8.639
3	Sum	48.000	66.667
4	Variance	38.689	74.632
5	SSQ	617.290	1190.760
6	MCSSQ	386.890	746.316
7	Euc Norm	24.845	34.507
8	Minimum	0.000	0.000
9	Maximum	17.833	24.769

Network Centralization Index = 20.11%

These results in Figure 24.28 are based on treating the whole network as the neighborhood of each actor. Betweenness, though, could be calculated on each ego's *n*-step neighborhood. We see here that ego 5 is most central and that there is a clear “inner circle” (egos 5, 2, and 3). There is considerable variation in the betweenness of the egos (mean of 4.8, with a standard deviation of 6.2, which yields a coefficient of variation of 130).

Another way to think about betweenness is to ask which *relations* are most central, rather than which *actors*. Freeman's definition can be easily applied: a relation is *between* to the extent that it is part of the geodesic path between pairs of actors. Using this idea, we can calculate a measure of the extent to which each relation in a binary graph is *between*.

Suppose that two actors want to have a relationship, but the geodesic path between them is blocked by a reluctant broker. If there exists another pathway, the two actors are likely to use it, even if it is longer and “less efficient.” In general, actors may use all of the pathways connecting them, rather than just geodesic paths. The flow betweenness (Freeman et al.,

1991) approach to centrality expands the notion of betweenness centrality. It assumes that actors will use all pathways that connect them, in proportion to the length of those pathways. Betweenness is measured by the proportion of the entire flow between two actors (that is, through all of the pathways connecting them) that occurs on paths of which a given actor is a part. For each actor, then, the measure adds up how involved an actor is in all of the flows between all other pairs of actors (the amount of computation with more than a couple actors can be pretty intimidating!). Because the magnitude of this index number would be expected to increase with the sheer size of the network and with network density, it is useful to standardize it by calculating the flow betweenness of each actor in ratio to the total flow betweenness that does not involve the actor. Figure 24.29 shows the flow betweenness of each ego in the information network.

Figure 24.29 Flow betweenness centrality for Knoke information network

	1 FlowBet	2 nFlowBet
1	3.854	5.352
2	20.783	28.866
3	16.954	23.547
4	4.220	5.861
5	25.876	35.939
6	1.500	2.083
7	8.401	11.668
8	2.954	4.102
9	4.054	5.630
10	4.092	5.683

Network Centralization Index = 25.629%

DESCRIPTIVE STATISTICS FOR EACH MEASURE

		1 FlowBet	2 nFlowBet
1	Mean	9.269	12.873
2	Std Dev	8.230	11.430
3	Sum	92.687	128.732
4	Variance	67.725	130.642
5	SSQ	1536.335	2963.609
6	MCSSQ	677.249	1306.421
7	Euc Norm	39.196	54.439
8	Minimum	1.500	2.083
9	Maximum	25.876	35.939

By this more complete measure of betweenness centrality, actors 2 and 5 are clearly the most important mediators. Actor 3, who was fairly important when we considered only geodesic flows, appears to be rather less important. While the overall picture does not change a great deal, the elaborated definition of betweenness does give us a somewhat different impression of who is most central in this network.

Conclusion

The social network analysis approach to social science emphasizes the relations between actors as being equally important (or perhaps more important) than the attributes of actors. The social network analysis approach also strongly emphasizes the interactions between individuals and their social context. Individuals make and enact social structure by their agency, but the choices they make are strongly conditioned by their locations in the texture of the larger social fabric in which they are embedded.

In this chapter, we've examined some of the most widely used concepts and measures found in quantitative approaches to basic social network analysis. At the core of the often bewildering complexity of specific approaches are a few basic ideas about social structures. Size, density, connection, distance, substructures (including groups and equivalence classes), and their interconnections are the common themes that emerge as major structural properties of social structures, whether global or local. As social network analysts have demonstrated again and again, these ideas, though simple, are powerful, and they help to provide insight into the behavior of both whole social structures and the individuals that form and enact them.

References

Bonacich, P. (1987) ‘Power and centrality: A family of measures’, American Journal of Sociology92: 1170–82. http://dx.doi.org/10.1086/228631
Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002) UCINET for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.
Boyd, J.P., Fitzgerald, W.J. and Beck, R.J. (2006) ‘Computing core/periphery structures and permutation tests for social relations data’, Social Networks28: 165–78. http://dx.doi.org/10.1016/j.socnet.2005.06.003
Burris, V. (2005) ‘Interlocking directorates and political cohesion among corporate elites’, American Journal of Sociology111: 249–83. http://dx.doi.org/10.1086/428817
Burt, R.S. (1992) Structural Holes: The Social Structure of Competition. Cambridge: Harvard University Press.
Clark, R. (2008) ‘Dependency, network integration, and development’, Sociological Perspectives51: 629–48. http://dx.doi.org/10.1111/j.1533-8525.1996.tb00744.x
Cook, K.S., Emerson, R.M., Gillmore, M.R. and Yamagishi, T. (1983) ‘The distribution of power in exchange networks: Theory and experimental results’, American Journal of Sociology89: 275–305. http://dx.doi.org/10.1086/227866
Crossley, N. (2008) ‘Small-world networks, complex systems and sociology’, Sociology42: 261–77. http://dx.doi.org/10.1177/0038038507087353
Field, S., Frank, K.A., Schiller, K., Riegle-Crumb, C. and Muller, C. (2006) ‘Identifying positions from affiliation networks: Preserving the duality of people and events’, Social Networks28: 97–123. http://dx.doi.org/10.1016/j.socnet.2005.04.005
Finsveen, E. and Wim vanOorschot. (2008) ‘Access to resources in networks: A theoretical and empirical critique of networks as a proxy for social capital’, Acta Sociologica51: 293–307. http://dx.doi.org/10.1177/0001699308097375

Fominaya, C.F. (2007) 'The role of humour in the process of collective identity formation in autonomous social movement groups in contemporary Madrid', International Review of Social History52: 243–58.
Frank, K.A. (1996) 'Mapping interactions within and between cohesive subgroups', Social Networks18: 93–119. http://dx.doi.org/10.1016/0378-8733%2895%2900257-X
FreemanL.C. (1979) 'Centrality in social networks: Conceptual clarification', Social Networks1: 215–39.
FreemanL.C., BorgattiS.P. and White, D.R. (1991) 'Centrality in valued graphs: A measure of betweenness based on network flow', Social Networks13: 141–54. http://dx.doi.org/10.1016/0378-8733%2891%2990017-N
Gallie, E.-P. (2009) 'Is geographical proximity necessary for knowledge spillovers within a cooperative technological network? The case of the French biotechnology sector', Regional Studies43: 33–42. http://dx.doi.org/10.1080/00343400701652818
Gould, J., and Fernandez, J. (1989) 'Structures of mediation: A formal approach to brokerage in transaction networks', Sociological Methodology: 89–126. http://dx.doi.org/10.2307/270949
Granovetter, M.S. (1982) 'Alienation reconsidered: The strength of weak ties', Connections5: 4–16.
Gurrieri, A. Rosa. (2008) 'Knowledge network dissemination in a family-firm sector', The Journal of Socio-Economics37: 2380–89. http://dx.doi.org/10.1016/j.socec.2008.04.005
Hampton, K.N. and Wellman, B. (1999) 'Netville online and offline: Observing and surveying a wired suburb', American Behavioral Scientist43: 475–92. http://dx.doi.org/10.1177/00027649921955290
Hanneman, R. and Riddle, M. (2005) Introduction to Social Network Methods. Riverside, CA: University of California.
Haythornthwaite, C. (2005) 'Social networks and Internet connectivity effects', Information, Communication & Society8: 125–47. http://dx.doi.org/10.1080/13691180500146185
Hermann, R.J. (2008) 'National security challenges and competition: US defense and space R&D in a strategic context', Technology in Society30: 371–81. http://dx.doi.org/10.1016/j.techsoc.2008.04.003
Knoke, D. and Yang, S. (2008) Network Analysis.
2nd ed.
Thousand Oaks, CA: Sage.
Knoke, D. and Wood, J. (1981) Organized for Action: Commitment in Voluntary Associations. New Brunswick, NJ: Rutgers University Press.
Kien, G. (2008) 'Beijing, 2006: International connectivity the way it is supposed to be', Qualitative Inquiry14: 1264–71. http://dx.doi.org/10.1177/1077800408321720
Krackhardt, D. and Stern, R.N. (1988) 'Informal networks and organizational crises: An experimental simulation', Social Psychology Quarterly51(2): 123–40. http://dx.doi.org/10.2307/2786835
Kratke, S. and Brandt, A. (2009) 'Knowledge networks as a regional development resource: A network analysis of the interlinks between Scientific Institutions and regional firms in the metropolitan region of Hanover, Germany', European Planning Studies17: 43–63. http://dx.doi.org/10.1080/09654310802513930
Lai, G. and Wong, O. (2002) 'The tie effect on information dissemination: The spread of a commercial rumor in Hong Kong', Social Networks24: 49–75. http://dx.doi.org/10.1016/S0378-8733%2801%2900050-8
Moody, J. and White, D.R. (2003) 'Structural cohesion and embeddedness: A hierarchical concept of social groups', American Sociological Review68: 103–27. http://dx.doi.org/10.2307/3088904
Newman, M.E.J. (2006) 'Modularity and community structure in networks', Proceedings of the National Academy of Sciences [PNAS]103(23): 8577–82.

http://www.pnas.org/cgi/doi/10.1073/pnas.0601602103. http://dx.doi.org/10.1073/pnas.0601602103
Raval, V.V. and Kral, M.J. (2004) 'Core versus periphery: Dynamics of personhood over the life-course for a Gujarati Hindu woman', Culture & Psychology10: 162–94. http://dx.doi.org/10.1177/1354067X04040927
Rycroft, R.W. (2007) 'Does cooperation absorb complexity? Innovation networks and the speed and spread of complex technological innovation', Technological Forecasting and Social Change74: 565–78. http://dx.doi.org/10.1016/j.techfore.2006.10.005
Sanders, K. and Nauta, A. (2004) 'Social cohesiveness and absenteeism: The relationship between characteristics of employees and short-term absenteeism within an organization', Small Group Research35: 724–41. http://dx.doi.org/10.1177/1046496404267186
Schnegg, M. (2007) 'Blurred edges, open boundaries: The long-term development of a peasant community in rural Mexico', Journal of Anthropological Research63: 5–32. http://dx.doi.org/10.3998/jar.0521004.0063.103
Scott, J. (2000) Social Network Analysis: A Handbook.
2nd ed.
Thousand Oaks, CA: Sage.
Totterdell, P., Holman, D. and Hukin, A. (2008) 'Social networkers: Measuring and examining individual differences in propensity to connect with others', Social Networks30: 283–96. http://dx.doi.org/10.1016/j.socnet.2008.04.003
Urry, J. (2006) 'Complexity', Theory, Culture & Society23: 111–15. http://dx.doi.org/10.1177/0263276406062818
Wasserman, S. and Faust, K. (1994) Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511815478
Watts, D.J. (1999) Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton, N.J.: Princeton University Press.
Watts, D.J. (2003) Six Degrees: The Science of a Connected Age. New York: W.W. Norton.
Wellman, B., Carrington, P.J. and Hall, A. (1988) 'Networks as personal communities', in BarryWellman and S.D.Berkowitz (eds), Social Structures: A Network Approach. Cambridge: Cambridge University Press.

- ego
- cliques
- actors
- structural holes
- nodes
- information networks
- neighborhoods

http://dx.doi.org/10.4135/9781446294413.n24