

Introduction to Social Network Methods

Robert A. Hanneman
Department of Sociology
University of California, Riverside

2001

Introduction to Social Network Methods

Table of Contents

This page is the starting point for an on-line textbook supporting Sociology 157, an undergraduate introductory course on social network analysis. Robert A. Hanneman of the Department of Sociology teaches the course at the University of California, Riverside. Feel free to use and reproduce this textbook (with citation). For more information, or to offer comments, you can send me e-mail.

About this Textbook

This on-line textbook introduces many of the basics of formal approaches to the analysis of social networks. It provides very brief overviews of a number of major areas with some examples. The text relies heavily on the work of Freeman, Borgatti, and Everett (the authors of the UCINET software package). The materials here, and their organization, were also very strongly influenced by the text of Wasserman and Faust, and by a graduate seminar conducted by Professor Phillip Bonacich at UCLA in 1998. Errors and omissions, of course, are the responsibility of the author.

Table of Contents

1. Social network data
2. Why formal methods?
3. Using graphs to represent social relations
4. Using matrices to represent social relations
5. Basic properties of networks and actors
6. Centrality and power
7. Cliques and sub-groups
8. Network positions and social roles: The analysis of equivalence
9. Structural equivalence
10. Automorphic equivalence
11. Regular equivalence

A bibliography of works about, or examples of, social network methods

1. Social Network Data

Introduction: What's different about social network data?

On one hand, there really isn't anything about social network data that is all that unusual. Networkers do use a specialized language for describing the structure and contents of the sets of observations that they use. But, network data can also be described and understood using the ideas and concepts of more familiar methods, like cross-sectional survey research.

On the other hand, the data sets that networkers develop usually end up looking quite different from the conventional rectangular data array so familiar to survey researchers and statistical analysts. The differences are quite important because they lead us to look at our data in a different way -- and even lead us to think differently about how to apply statistics.

"Conventional" sociological data consists of a rectangular array of measurements. The rows of the array are the cases, or subjects, or observations. The columns consist of scores (quantitative or qualitative) on attributes, or variables, or measures. Each cell of the array then describes the score of some actor on some attribute. In some cases, there may be a third dimension to these arrays, representing panels of observations or multiple groups.

<i>Name</i>	<i>Sex</i>	<i>Age</i>	<i>In-Degree</i>
Bob	Male	32	2
Carol	Female	27	1
Ted	Male	29	1
Alice	Female	28	3

The fundamental data structure is one that leads us to compare how actors are similar or dissimilar to each other across attributes (by comparing rows). Or, perhaps more commonly, we examine how variables are similar or dissimilar to each other in their distributions across actors (by comparing or correlating columns).

"Network" data (in their purest form) consist of a square array of measurements. The rows of the array are the cases, or subjects, or observations. The columns of the array are -- and note the key difference from conventional data -- the same set of cases, subjects, or observations. In each cell of the array describes a relationship between the actors.

Who reports liking whom?				
	Choice:			
Chooser:	Bob	Carol	Ted	Alice
Bob	---	0	1	1
Carol	1	---	0	1
Ted	0	1	---	1
Alice	1	0	0	---

We could look at this data structure the same way as with attribute data. By comparing rows of the array, we can see which actors are similar to which other actors in whom they choose. By looking at the columns, we can see who is similar to whom in terms of being chosen by others. These are useful ways to look at the data, because they help us to see which actors have similar positions in the network. This is the first major emphasis of network analysis: seeing how actors are located or "embedded" in the overall network.

But a network analyst is also likely to look at the data structure in a second way -- holistically. The analyst might note that there are about equal numbers of ones and zeros in the matrix. This suggests that there is a moderate "density" of liking overall. The analyst might also compare the cells above and below the diagonal to see if there is reciprocity in choices (e.g. Bob chose Ted, did Ted choose Bob?). This is the second major emphasis of network analysis: seeing how the whole pattern of individual choices gives rise to more holistic patterns.

It is quite possible to think of the network data set in the same terms as "conventional data." One can think of the rows as simply a listing of cases, and the columns as attributes of each actor (i.e. the relations with other actors can be thought of as "attributes" of each actor). Indeed, many of the techniques used by network analysts (like calculating correlations and distances) are applied exactly the same way to network data as they would be to conventional data.

While it is possible to describe network data as just a special form of conventional data (and it is), network analysts look at the data in some rather fundamentally different ways. Rather than thinking about how an actor's ties with other actors describes the attributes of "ego," network analysts instead see a structure of connections, within which the actor is embedded. Actors are described by their relations, not by their attributes. And, the relations themselves are just as fundamental as the actors that they connect.

The major difference between conventional and network data is that conventional data focuses on actors and attributes; network data focus on actors and relations. The difference in emphasis is consequential for the choices that a researcher must make in deciding on research design, in

conducting sampling, developing measurement, and handling the resulting data. It is not that the research tools used by network analysts are different from those of other social scientists (they mostly are not). But the special purposes and emphases of network research do call for some different considerations.

In this chapter, we will take a look at some of the issues that arise in design, sampling, and measurement for social network analysis. Our discussion will focus on the two parts of network data: nodes (or actors) and edges (or relations). We will try to show some of the ways in which network data are similar to, and different from more familiar actor by attribute data. We will introduce some new terminology that makes it easier to describe the special features of network data. Lastly, we will briefly discuss how the differences between network and actor-attribute data are consequential for the application of statistical tools.

Nodes

Network data are defined by actors and by relations (or nodes and ties, etc.). The nodes or actors part of network data would seem to be pretty straight-forward. Other empirical approaches in the social sciences also think in terms of cases or subjects or sample elements and the like. There is one difference with most network data, however, that makes a big difference in how such data are usually collected -- and the kinds of samples and populations that are studied.

Network analysis focuses on the relations among actors, and not individual actors and their attributes. This means that the actors are usually not sampled independently, as in many other kinds of studies (most typically, surveys). Suppose we are studying friendship ties, for example. John has been selected to be in our sample. When we ask him, John identifies seven friends. We need to track down each of those seven friends and ask them about their friendship ties, as well. The seven friends are in our sample because John is (and vice-versa), so the "sample elements" are no longer "independent."

The nodes or actors included in non-network studies tend to be the result of independent probability sampling. Network studies are much more likely to include all of the actors who occur within some (usually naturally occurring) boundary. Often network studies don't use "samples" at all, at least in the conventional sense. Rather, they tend to include all of the actors in some population or populations. Of course, the populations included in a network study may be a sample of some larger set of populations. For example, when we study patterns of interaction among students in classrooms, we include all of the children in a classroom (that is, we study the whole population of the classroom). The classroom itself, though, might have been selected by probability methods from a population of classrooms (say all of those in a school).

The use of whole populations as a way of selecting observations in (many) network studies makes it important for the analyst to be clear about the boundaries of each population to be studied, and how individual units of observation are to be selected within that population. Network data sets also frequently involve several levels of analysis, with actors embedded at the lowest level (i.e. network designs can be described using the language of "nested" designs).

Populations, samples, and boundaries

Social network analysts rarely draw samples in their work. Most commonly, network analysts will identify some population and conduct a census (i.e. include all elements of the population as units of observation). A network analyst might examine all of the nouns and objects occurring in a text, all of the persons at a birthday party, all members of a kinship group, of an organization, neighborhood, or social class (e.g. landowners in a region, or royalty).

Survey research methods usually use a quite different approach to deciding which nodes to study. A list is made of all nodes (sometimes stratified or clustered), and individual elements are selected by probability methods. The logic of the method treats each individual as a separate "replication" that is, in a sense, interchangeable with any other.

Because network methods focus on relations among actors, actors cannot be sampled independently to be included as observations. If one actor happens to be selected, then we must also include all other actors to whom our ego has (or could have) ties. As a result, network approaches tend to study whole populations by means of census, rather than by sample (we will discuss a number of exceptions to this shortly, under the topic of sampling ties).

The populations that network analysts study are remarkably diverse. At one extreme, they might consist of symbols in texts or sounds in verbalizations; at the other extreme, nations in the world system of states might constitute the population of nodes. Perhaps most common, of course, are populations of individual persons. In each case, however, the elements of the population to be studied are defined by falling within some boundary.

The boundaries of the populations studied by network analysts are of two main types. Probably most commonly, the boundaries are those imposed or created by the actors themselves. All the members of a classroom, organization, club, neighborhood, or community can constitute a population. These are naturally occurring clusters, or networks. So, in a sense, social network studies often draw the boundaries around a population that is known, *a priori*, to be a network. Alternatively, a network analyst might take a more "demographic" or "ecological" approach to defining population boundaries. We might draw observations by contacting all of the people who are found in a bounded spatial area, or who meet some criterion (having gross family incomes over \$1,000,000 per year). Here, we might have reason to suspect that networks exist, but the entity being studied is an abstract aggregation imposed by the investigator -- rather than a pattern of institutionalized social action that has been identified and labeled by its participants.

Network analysts can expand the boundaries of their studies by replicating populations. Rather than studying one neighborhood, we can study several. This type of design (which could use sampling methods to select populations) allows for replication and for testing of hypotheses by comparing populations. A second, and equally important way that network studies expand their scope is by the inclusion of multiple levels of analysis, or modalities.

Modality and levels of analysis

The network analyst tends to see individual people nested within networks of face-to-face relations with other persons. Often these networks of interpersonal relations become "social facts" and take on a life of their own. A family, for example, is a network of close relations among a set of people. But this particular network has been institutionalized and given a name and reality beyond that of its component nodes. Individuals in their work relations may be seen as nested within organizations; in their leisure relations they may be nested in voluntary associations. Neighborhoods, communities, and even societies are, to varying degrees, social entities in and of themselves. And, as social entities, they may form ties with the individuals nested within them, and with other social entities.

Often network data sets describe the nodes and relations among nodes for a single bounded population. If I study the friendship patterns among students in a classroom, I am doing a study of this type. But a classroom exists within a school - which might be thought of as a network relating classes and other actors (principals, administrators, librarians, etc.). And most schools exist within school districts, which can be thought of as networks of schools and other actors (school boards, research wings, purchasing and personnel departments, etc.). There may even be patterns of ties among school districts (say by the exchange of students, teachers, curricular materials, etc.).

Most networkers think of individual persons as being embedded in networks that are embedded in networks that are embedded in networks. Networkers describe such structures as "multi-modal." In our school example, individual students and teachers form one mode, classrooms a second, schools a third, and so on. A data set that contains information about two types of social entities (say persons and organizations) is a two mode network.

Of course, this kind of view of the nature of social structures is not unique to social networkers. Statistical analysts deal with the same issues as "hierarchical" or "nested" designs. Theorists speak of the macro-meso-micro levels of analysis, or develop schema for identifying levels of analysis (individual, group, organization, community, institution, society, global order being perhaps the most commonly used system in sociology). One advantage of network thinking and method is that it naturally predisposes the analyst to focus on multiple levels of analysis simultaneously. That is, the network analyst is always interested in how the individual is embedded within a structure and how the structure emerges from the micro-relations between individual parts. The ability of network methods to map such multi-modal relations is, at least potentially, a step forward in rigor.

Having claimed that social network methods are particularly well suited for dealing with multiple levels of analysis and multi-modal data structures, it must immediately be admitted that networkers rarely actually take much advantage. Most network analyses does move us beyond simple micro or macro reductionism -- and this is good. Few, if any, data sets and analyses, however, have attempted to work at more than two modes simultaneously. And, even when working with two modes, the most common strategy is to examine them more or less separately

(one exception to this is the conjoint analysis of two mode networks).

Relations

The other half of the design of network data has to do with what ties or relations are to be measured for the selected nodes. There are two main issues to be discussed here. In many network studies, all of the ties of a given type among all of the selected nodes are studied -- that is, a census is conducted. But, sometimes different approaches are used (because they are less expensive, or because of a need to generalize) that sample ties. There is also a second kind of sampling of ties that always occurs in network data. Any set of actors might be connected by many different kinds of ties and relations (e.g. students in a classroom might like or dislike each other, they might play together or not, they might share food or not, etc.). When we collect network data, we are usually selecting, or sampling, from among a set of kinds of relations that we might have measured.

Sampling ties

Given a set of actors or nodes, there are several strategies for deciding how to go about collecting measurements on the relations among them. At one end of the spectrum of approaches are "full network" methods. This approach yields the maximum of information, but can also be costly and difficult to execute, and may be difficult to generalize. At the other end of the spectrum are methods that look quite like those used in conventional survey research. These approaches yield considerably less information about network structure, but are often less costly, and often allow easier generalization from the observations in the sample to some larger population. There is no one "right" method for all research questions and problems.

Full network methods require that we collect information about each actor's ties with all other actors. In essence, this approach is taking a census of ties in a population of actors -- rather than a sample. For example we could collect data on shipments of copper between all pairs of nation states in the world system from IMF records; we could examine the boards of directors of all public corporations for overlapping directors; we could count the number of vehicles moving between all pairs of cities; we could look at the flows of e-mail between all pairs of employees in a company; we could ask each child in a play group to identify their friends.

Because we collect information about ties between all pairs or dyads, full network data give a complete picture of relations in the population. Most of the special approaches and methods of network analysis that we will discuss in the remainder of this text were developed to be used with full network data. Full network data is necessary to properly define and measure many of the structural concepts of network analysis (e.g. between-ness).

Full network data allows for very powerful descriptions and analyses of social structures. Unfortunately, full network data can also be very expensive and difficult to collect. Obtaining data from every member of a population, and having every member rank or rate every other member can be very challenging tasks in any but the smallest groups. The task is made more manageable by asking respondents to identify a limited number of specific individuals with whom they have ties. These lists can then be compiled and cross-connected. But, for large groups

(say all the people in a city), the task is practically impossible.

In many cases, the problems are not quite as severe as one might imagine. Most persons, groups, and organizations tend to have limited numbers of ties -- or at least limited numbers of strong ties. This is probably because social actors have limited resources, energy, time, and cognitive capacity -- and cannot maintain large numbers of strong ties. It is also true that social structures can develop a considerable degree of order and solidarity with relatively few connections.

Snowball methods begin with a focal actor or set of actors. Each of these actors is asked to name some or all of their ties to other actors. Then, all the actors named (who were not part of the original list) are tracked down and asked for some or all of their ties. The process continues until no new actors are identified, or until we decide to stop (usually for reasons of time and resources, or because the new actors being named are very marginal to the group we are trying to study).

The snowball method can be particularly helpful for tracking down "special" populations (often numerically small sub-sets of people mixed in with large numbers of others). Business contact networks, community elites, deviant sub-cultures, avid stamp collectors, kinship networks, and many other structures can be pretty effectively located and described by snowball methods. It is sometimes not as difficult to achieve closure in snowball "samples" as one might think. The limitations on the numbers of strong ties that most actors have, and the tendency for ties to be reciprocated often make it fairly easy to find the boundaries.

There are two major potential limitations and weaknesses of snowball methods. First, actors who are not connected (i.e. "isolates") are not located by this method. The presence and numbers of isolates can be a very important feature of populations for some analytic purposes. The snowball method may tend to overstate the "connectedness" and "solidarity" of populations of actors. Second, there is no guaranteed way of finding all of the connected individuals in the population. Where does one start the snowball rolling? If we start in the wrong place or places, we may miss whole sub-sets of actors who are connected -- but not attached to our starting points.

Snowball approaches can be strengthened by giving some thought to how to select the initial nodes. In many studies, there may be a natural starting point. In community power studies, for example, it is common to begin snowball searches with the chief executives of large economic, cultural, and political organizations. While such an approach will miss most of the community (those who are "isolated" from the elite network), the approach is very likely to capture the elite network quite effectively.

Ego-centric networks (with alter connections)

In many cases it will not be possible (or necessary) to track down the full networks beginning with focal nodes (as in the snowball method). An alternative approach is to begin with a selection of focal nodes (egos), and identify the nodes to which they are connected. Then, we determine which of the nodes identified in the first stage are connected to one another. This can be done by contacting each of the nodes; sometimes we can ask ego to report which of the nodes that it is tied to are tied to one another.

This kind of approach can be quite effective for collecting a form of relational data from very

large populations, and can be combined with attribute-based approaches. For example, we might take a simple random sample of male college students and ask them to report who are their close friends, and which of these friends know one another. This kind of approach can give us a good and reliable picture of the kinds of networks (or at least the local neighborhoods) in which individuals are embedded. We can find out such things as how many connections nodes have, and the extent to which these nodes are close-knit groups. Such data can be very useful in helping to understand the opportunities and constraints that ego has as a result of the way they are embedded in their networks.

The ego-centered approach with alter connections can also give us some information about the network as a whole, though not as much as snowball or census approaches. Such data are, in fact, micro-network data sets -- samplings of local areas of larger networks. Many network properties -- distance, centrality, and various kinds of positional equivalence cannot be assessed with ego-centric data. Some properties, such as overall network density can be reasonably estimated with ego-centric data. Some properties -- such as the prevalence of reciprocal ties, cliques, and the like can be estimated rather directly.

Ego-centric networks (ego only)

Ego-centric methods really focus on the individual, rather than on the network as a whole. By collecting information on the connections among the actors connected to each focal ego, we can still get a pretty good picture of the "local" networks or "neighborhoods" of individuals. Such information is useful for understanding how networks affect individuals, and they also give a (incomplete) picture of the general texture of the network as a whole.

Suppose, however, that we only obtained information on ego's connections to alters -- but not information on the connections among those alters. Data like these are not really "network" data at all. That is, they cannot be represented as a square actor-by-actor array of ties. But doesn't mean that ego-centric data without connections among the alters are of no value for analysts seeking to take a structural or network approach to understanding actors. We can know, for example, that some actors have many close friends and kin, and others have few. Knowing this, we are able to understand something about the differences in the actors places in social structure, and make some predictions about how these locations constrain their behavior. What we cannot know from ego-centric data with any certainty is the nature of the macro-structure or the whole network.

In ego-centric networks, the alters identified as connected to each ego are probably a set that is unconnected with those for each other ego. While we cannot assess the overall density or connectedness of the population, we can sometimes be a bit more general. If we have some good theoretical reason to think about alters in terms of their social roles, rather than as individual occupants of social roles, ego-centered networks can tell us a good bit about local social structures. For example, if we identify each of the alters connected to an ego by a friendship relation as "kin," "co-worker," "member of the same church," etc., we can build up a picture of the networks of social positions (rather than the networks of individuals) in which egos are embedded. Such an approach, of course, assumes that such categories as "kin" are real and meaningful determinants of patterns of interaction.

Multiple relations

In a conventional actor-by-trait data set, each actor is described by many variables (and each variable is realized in many actors). In the most common social network data set of actor-by-actor ties, only one kind of relation is described. Just as we often are interested in multiple attributes of actors, we are often interested in multiple kinds of ties that connect actors in a network.

In thinking about the network ties among faculty in an academic department, for example, we might be interested in which faculty have students in common, serve on the same committees, interact as friends outside of the workplace, have one or more areas of expertise in common, and co-author papers. The positions that actors hold in the web of group affiliations are multi-faceted. Positions in one set of relations may re-enforce or contradict positions in another (I might share friendship ties with one set of people with whom I do not work on committees, for example). Actors may be tied together closely in one relational network, but be quite distant from one another in a different relational network. The locations of actors in multi-relational networks and the structure of networks composed of multiple relations are some of the most interesting (and still relatively unexplored) areas of social network analysis.

When we collect social network data about certain kinds of relations among actors we are, in a sense, sampling from a population of possible relations. Usually our research question and theory indicate which of the kinds of relations among actors are the most relevant to our study, and we do not sample -- but rather select -- relations. In a study concerned with economic dependency and growth, for example, I could collect data on the exchange of performances by musicians between nations -- but it is not really likely to be all that relevant.

If we do not know what relations to examine, how might we decide? There are a number of conceptual approaches that might be of assistance. Systems theory, for example, suggests two domains: material and informational. Material things are "conserved" in the sense that they can only be located at one node of the network at a time. Movements of people between organizations, money between people, automobiles between cities, and the like are all examples of material things which move between nodes -- and hence establish a network of material relations. Informational things, to the systems theorist, are "non-conserved" in the sense that they can be in more than one place at the same time. If I know something and share it with you, we both now know it. In a sense, the commonality that is shared by the exchange of information may also be said to establish a tie between two nodes. One needs to be cautious here, however, not to confuse the simple possession of a common attribute (e.g. gender) with the presence of a tie (e.g. the exchange of views between two persons on issues of gender).

Methodologies for working with multi-relational data are not as well developed as those for working with single relations. Many interesting areas of work such as network correlation, multi-dimensional scaling and clustering, and role algebras have been developed to work with multi-relational data. For the most part, these topics are beyond the scope of the current text, and are best approached after the basics of working with single relational networks are mastered.

Scales of measurement

Like other kinds of data, the information we collect about ties between actors can be measured (i.e. we can assign scores to our observations) at different "levels of measurement." The different levels of measurement are important because they limit the kinds of questions that can be examined by the researcher. Scales of measurement are also important because different kinds of scales have different mathematical properties, and call for different algorithms in describing patterns and testing inferences about them.

It is conventional to distinguish nominal, ordinal, and interval levels of measurement (the ratio level can, for all practical purposes, be grouped with interval). It is useful, however, to further divide nominal measurement into binary and multi-category variations; it is also useful to distinguish between full-rank ordinal measures and grouped ordinal measures. We will briefly describe all of these variations, and provide examples of how they are commonly applied in social network studies.

Binary measures of relations: By far the most common approach to scaling (assigning numbers to) relations is to simply distinguish between relations being absent (coded zero), and ties being present (coded one). If we ask respondents in a survey to tell us "which other people on this list do you like?" we are doing binary measurement. Each person from the list that is selected is coded one. Those who are not selected are coded zero.

Much of the development of graph theory in mathematics, and many of the algorithms for measuring properties of actors and networks have been developed for binary data. Binary data is so widely used in network analysis that it is not unusual to see data that are measured at a "higher" level transformed into binary scores before analysis proceeds. To do this, one simply selects some "cut point" and rescores cases as below the cutpoint (zero) or above it (one). Dichotomizing data in this way is throwing away information. The analyst needs to consider what is relevant (i.e. what is the theory about? is it about the presence and pattern of ties, or about the strengths of ties?), and what algorithms are to be applied in deciding whether it is reasonable to recode the data. Very often, the additional power and simplicity of analysis of binary data is "worth" the cost in information lost.

Multiple-category nominal measures of relations: In collecting data we might ask our respondents to look at a list of other people and tell us: "for each person on this list, select the category that describes your relationship with them the best: friend, lover, business relationship, kin, or no relationship." We might score each person on the list as having a relationship of type "1" type "2" etc. This kind of a scale is nominal or qualitative -- each person's relationship to the subject is coded by its type, rather than it's strength. Unlike the binary nominal (true-false) data, the multiple category nominal measure is multiple choice.

The most common approach to analyzing multiple-category nominal measures is to use it to create a series of binary measures. That is, we might take the data arising from the question described above and create separate sets of scores for friendship ties, for lover ties, for kin ties,

etc. This is very similar to "dummy coding" as a way of handling multiple choice types of measures in statistical analysis. In examining the resulting data, however, one must remember that each node was allowed to have a tie in at most one of the resulting networks. That is, a person can be a friendship tie or a lover tie -- but not both -- as a result of the way we asked the question. In examining the resulting networks, densities may be artificially low, and there will be an inherent negative correlation among the matrices.

This sort of multiple choice data can also be "binarized." That is, we can ignore what kind of tie is reported, and simply code whether a tie exists for a dyad, or not. This may be fine for some analyses -- but it does waste information. One might also wish to regard the types of ties as reflecting some underlying continuous dimension (for example, emotional intensity). The types of ties can then be scaled into a single grouped ordinal measure of tie strength. The scaling, of course, reflects the predisposition of the analyst -- not the reports of the respondents.

Grouped ordinal measures of relations: One of the earliest traditions in the study of social networks asked respondents to rate each of a set of others as "liked" "disliked" or "neutral." The result is a grouped ordinal scale (i.e., there can be more than one "liked" person, and the categories reflect an underlying rank order of intensity). Usually, this kind of three-point scale was coded -1, 0, and +1 to reflect negative liking, indifference, and positive liking. When scored this way, the pluses and minuses make it fairly easy to write algorithms that will count and describe various network properties (e.g. the structural balance of the graph).

Grouped ordinal measures can be used to reflect a number of different quantitative aspects of relations. Network analysts are often concerned with describing the "strength" of ties. But, "strength" may mean (some or all of) a variety of things. One dimension is the frequency of interaction -- do actors have contact daily, weekly, monthly, etc. Another dimension is "intensity," which usually reflects the degree of emotional arousal associated with the relationship (e.g. kin ties may be infrequent, but carry a high "emotional charge" because of the highly ritualized and institutionalized expectations). Ties may be said to be stronger if they involve many different contexts or types of ties. Summing nominal data about the presence or absence of multiple types of ties gives rise to an ordinal (actually, interval) scale of one dimension of tie strength. Ties are also said to be stronger to the extent that they are reciprocated. Normally we would assess reciprocity by asking each actor in a dyad to report their feelings about the other. However, one might also ask each actor for their perceptions of the degree of reciprocity in a relation: Would you say that neither of you like each other very much, that you like X more than X likes you, that X likes you more than you like X, or that you both like each other about equally?

Ordinal scales of measurement contain more information than nominal. That is, the scores reflect finer gradations of tie strength than the simple binary "presence or absence." This would seem to be a good thing, yet it is frequently difficult to take advantage of ordinal data. The most commonly used algorithms for the analysis of social networks have been designed for binary data. Many have been adapted to continuous data -- but for interval, rather than ordinal scales of measurement. Ordinal data, consequently, are often binarized by choosing some cut-point and rescaling. Alternatively, ordinal data are sometimes treated as though they really were interval. The former strategy has some risks, in that choices of cutpoints can be consequential; the latter

strategy has some risks, in that the intervals separating points on an ordinal scale may be very heterogeneous.

Full-rank ordinal measures of relations: Sometimes it is possible to score the strength of all of the relations of an actor in a rank order from strongest to weakest. For example, I could ask each respondent to write a "1" next to the name of the person in the class that you like the most, a "2" next to the name of the person you like next most, etc. The kind of scale that would result from this would be a "full rank order scale." Such scales reflect differences in degree of intensity, but not necessarily equal differences -- that is, the difference between my first and second choices is not necessarily the same as the difference between my second and third choices. Each relation, however, has a unique score (1st, 2nd, 3rd, etc.).

Full rank ordinal measures are somewhat uncommon in the social networks research literature, as they are in most other traditions. Consequently, there are relatively few methods, definitions, and algorithms that take specific and full advantage of the information in such scales. Most commonly, full rank ordinal measures are treated as if they were interval. There is probably somewhat less risk in treating fully rank ordered measures (compared to grouped ordinal measures) as though they were interval, though the assumption is still a risky one. Of course, it is also possible to group the rank order scores into groups (i.e. produce a grouped ordinal scale) or dichotomize the data (e.g. the top three choices might be treated as ties, the remainder as non-ties). In combining information on multiple types of ties, it is frequently necessary to simplify full rank order scales. But, if we have a number of full rank order scales that we may wish to combine to form a scale (i.e. rankings of people's likings of other in the group, frequency of interaction, etc.), the sum of such scales into an index is plausibly treated as a truly interval measure.

Interval measures of relations: The most "advanced" level of measurement allows us to discriminate among the relations reported in ways that allow us to validly state that, for example, "this tie is twice as strong as that tie." Ties are rated on scales in which the difference between a "1" and a "2" reflects the same amount of real difference as that between "23" and "24."

True interval level measures of the strength of many kinds of relationships are fairly easy to construct, with a little imagination and persistence. Asking respondents to report the details of the frequency or intensity of ties by survey or interview methods, however, can be rather unreliable -- particularly if the relationships being tracked are not highly salient and infrequent. Rather than asking whether two people communicate, one could count the number of email, phone, and inter-office mail deliveries between them. Rather than asking whether two nations trade with one another, look at statistics on balances of payments. In many cases, it is possible to construct interval level measures of relationship strength by using artifacts (e.g. statistics collected for other purposes) or observation.

Continuous measures of the strengths of relationships allow the application of a wider range of mathematical and statistical tools to the exploration and analysis of the data. Many of the algorithms that have been developed by social network analysts, originally for binary data, have been extended to take advantage of the information available in full interval measures. Whenever possible, connections should be measured at the interval level -- as we can always move to a less

refined approach later; if data are collected at the nominal level, it is much more difficult to move to a more refined level.

Even though it is a good idea to measure relationship intensity at the most refined level possible, most network analysis does not operate at this level. The most powerful insights of network analysis, and many of the mathematical and graphical tools used by network analysts were developed for simple graphs (i.e. binary, undirected). Many characterizations of the embeddedness of actors in their networks, and of the networks themselves are most commonly thought of in discrete terms in the research literature. As a result, it is often desirable to reduce even interval data to the binary level by choosing a cutting -point, and coding tie strength above that point as "1" and below that point as "0." Unfortunately, there is no single "correct" way to choose a cut-point. Theory and the purposes of the analysis provide the best guidance. Sometimes examining the data can help (maybe the distribution of tie strengths really is discretely bi-modal, and displays a clear cut point; maybe the distribution is highly skewed and the main feature is a distinction between no tie and any tie). When a cut-point is chosen, it is wise to also consider alternative values that are somewhat higher and lower, and repeat the analyses with different cut-points to see if the substance of the results is affected. This can be very tedious, but it is very necessary. Otherwise, one may be fooled into thinking that a real pattern has been found, when we have only observed the consequences of where we decided to put our cut-point.

A note on statistics and social network data

Social network analysis is more a branch of "mathematical" sociology than of "statistical or quantitative analysis," though networkers most certainly practice both approaches. The distinction between the two approaches is not clear cut. Mathematical approaches to network analysis tend to treat the data as "deterministic." That is, they tend to regard the measured relationships and relationship strengths as accurately reflecting the "real" or "final" or "equilibrium" status of the network. Mathematical types also tend to assume that the observations are not a "sample" of some larger population of possible observations; rather, the observations are usually regarded as the population of interest. Statistical analysts tend to regard the particular scores on relationship strengths as stochastic or probabilistic realizations of an underlying true tendency or probability distribution of relationship strengths. Statistical analysts also tend to think of a particular set of network data as a "sample" of a larger class or population of such networks or network elements -- and have a concern for the results of the current study would be reproduced in the "next" study of similar samples.

In the chapters that follow in this text, we will mostly be concerned with the "mathematical" rather than the "statistical" side of network analysis (again, it is important to remember that I am over-drawing the differences in this discussion). Before passing on to this, we should note a couple main points about the relationship between the material that you will be studying here, and the main statistical approaches in sociology.

In one way, there is little apparent difference between conventional statistical approaches and network approaches. Univariate, bi-variate, and even many multivariate descriptive statistical tools are commonly used in the describing, exploring, and modeling social network data. Social

network data are, as we have pointed out, easily represented as arrays of numbers -- just like other types of sociological data. As a result, the same kinds of operations can be performed on network data as on other types of data. Algorithms from statistics are commonly used to describe characteristics of individual observations (e.g. the median tie strength of actor X with all other actors in the network) and the network as a whole (e.g. the mean of all tie strengths among all actors in the network). Statistical algorithms are very heavily used in assessing the degree of similarity among actors, and in finding patterns in network data (e.g. factor analysis, cluster analysis, multi-dimensional scaling). Even the tools of predictive modeling are commonly applied to network data (e.g. correlation and regression).

Descriptive statistical tools are really just algorithms for summarizing characteristics of the distributions of scores. That is, they are mathematical operations. Where statistics really become "statistical" is on the inferential side. That is, when our attention turns to assessing the reproducibility or likelihood of the pattern that we have described. Inferential statistics can be, and are, applied to the analysis of network data. But, there are some quite important differences between the flavors of inferential statistics used with network data, and those that are most commonly taught in basic courses in statistical analysis in sociology.

Probably the most common emphasis in the application of inferential statistics to social science data is to answer questions about the stability, reproducibility, or generalizability of results observed in a single sample. The main question is: if I repeated the study on a different sample (drawn by the same method), how likely is it that I would get the same answer about what is going on in the whole population from which I drew both samples? This is a really important question -- because it helps us to assess the confidence (or lack of it) that we ought to have in assessing our theories and giving advice.

To the extent the observations used in a network analysis are drawn by probability sampling methods from some identifiable population of actors and/or ties, the same kind of question about the generalizability of sample results applies. Often this type of inferential question is of little interest to social network researchers. In many cases, they are studying a particular network or set of networks, and have no interest in generalizing to a larger population of such networks (either because there isn't any such population, or we don't care about generalizing to it in any probabilistic way). In some other cases we may have an interest in generalizing, but our sample was not drawn by probability methods. Network analysis often relies on artifacts, direct observation, laboratory experiments, and documents as data sources -- and usually there are no plausible ways of identifying populations and drawing samples by probability methods.

The other major use of inferential statistics in the social sciences is for testing hypotheses. In many cases, the same or closely related tools are used for questions of assessing generalizability and for hypothesis testing. The basic logic of hypothesis testing is to compare an observed result in a sample to some null hypothesis value, relative to the sampling variability of the result under the assumption that the null hypothesis is true. If the sample result differs greatly from what was likely to have been observed under the assumption that the null hypothesis is true -- then the null hypothesis is probably not true.

The key link in the inferential chain of hypothesis testing is the estimation of the standard errors

of statistics. That is, estimating the expected amount that the value a statistic would "jump around" from one sample to the next simply as a result of accidents of sampling. We rarely, of course, can directly observe or calculate such standard errors -- because we don't have replications. Instead, information from our sample is used to estimate the sampling variability.

With many common statistical procedures, it is possible to estimate standard errors by well validated approximations (e.g. the standard error of a mean is usually estimated by the sample standard deviation divided by the square root of the sample size). These approximations, however, hold when the observations are drawn by independent random sampling. Network observations are almost always non-independent, by definition. Consequently, conventional inferential formulas do not apply to network data (though formulas developed for other types of dependent sampling may apply). It is particularly dangerous to assume that such formulas do apply, because the non-independence of network observations will usually result in under-estimates of true sampling variability -- and hence, too much confidence in our results.

The approach of most network analysts interested in statistical inference for testing hypotheses about network properties is to work out the probability distributions for statistics directly. This approach is used because: 1) no one has developed approximations for the sampling distributions of most of the descriptive statistics used by network analysts and 2) interest often focuses on the probability of a parameter relative to some theoretical baseline (usually randomness) rather than on the probability that a given network is typical of the population of all networks.

Suppose, for example, that I was interested in the proportion of the actors in a network who were members of cliques (or any other network statistic or parameter). The notion of a clique implies structure -- non-random connections among actors. I have data on a network of ten nodes, in which there are 20 symmetric ties among actors, and I observe that there is one clique containing four actors. The inferential question might be posed as: how likely is it, if ties among actors were purely random events, that a network composed of ten nodes and 20 symmetric ties would display one or more cliques of size four or more? If it turns out that cliques of size four or more in random networks of this size and degree are quite common, I should be very cautious in concluding that I have discovered "structure" or non-randomness. If it turns out that such cliques (or more numerous or more inclusive ones) are very unlikely under the assumption that ties are purely random, then it is very plausible to reach the conclusion that there is a social structure present.

But how can I determine this probability? The method used is one of simulation -- and, like most simulation, a lot of computer resources and some programming skills are often necessary. In the current case, I might use a table of random numbers to distribute 20 ties among 10 actors, and then search the resulting network for cliques of size four or more. If no clique is found, I record a zero for the trial; if a clique is found, I record a one. The rest is simple. Just repeat the experiment several thousand times and add up what proportion of the "trials" result in "successes." The probability of a success across these simulation experiments is a good estimator of the likelihood that I might find a network of this size and density to have a clique of this size "just by accident" when the non-random causal mechanisms that I think cause cliques are not, in fact, operating.

This may sound odd, and it is certainly a lot of work (most of which, thankfully, can be done by computers). But, in fact, it is not really different from the logic of testing hypotheses with non-network data. Social network data tend to differ from more "conventional" survey data in some key ways: network data are often not probability samples, and the observations of individual nodes are not independent. These differences are quite consequential for both the questions of generalization of findings, and for the mechanics of hypothesis testing. There is, however, nothing fundamentally different about the logic of the use of descriptive and inferential statistics with social network data.

The application of statistics to social network data is an interesting area, and one that is, at the time of this writing, at a "cutting edge" of research in the area. Since this text focuses on more basic and commonplace uses of network analysis, we won't have very much more to say about statistics beyond this point. You can think of much of what follows here as dealing with the "descriptive" side of statistics (developing index numbers to describe certain aspects of the distribution of relational ties among actors in networks). For those with an interest in the inferential side, a good place to start is with the second half of the excellent Wasserman and Faust textbook.

2. Why Formal Methods?

Introduction to chapter 2

The basic idea of a social network is very simple. A social network is a set of actors (or points, or nodes, or agents) that may have relationships (or edges, or ties) with one another. Networks can have few or many actors, and one or more kinds of relations between pairs of actors. To build a useful understanding of a social network, a complete and rigorous description of a pattern of social relationships is a necessary starting point for analysis. That is, ideally we will know about all of the relationships between each pair of actors in the population.

One reason for using mathematical and graphical techniques in social network analysis is to represent the descriptions of networks compactly and systematically. This also enables us to use computers to store and manipulate the information quickly and more accurately than we can by hand. For small populations of actors (e.g. the people in a neighborhood, or the business firms in an industry), we can describe the pattern of social relationships that connect the actors rather completely and effectively using words. To make sure that our description is complete, however, we might want to list all logically possible pairs of actors, and describe each kind of possible relationship for each pair. This can get pretty tedious if the number of actors and/or number of kinds of relations is large. Formal representations ensure that all the necessary information is systematically represented, and provides rules for doing so in ways that are much more efficient than lists.

A related reason for using (particularly mathematical) formal methods for representing social networks is that mathematical representations allow us to apply computers to the analysis of network data. Why this is important will become clearer as we learn more about how structural analysis of social networks occurs. Suppose, for a simple example, that we had information about trade-flows of 50 different commodities (e.g. coffee, sugar, tea, copper, bauxite) among the 170 or so nations of the world system in a given year. Here, the 170 nations can be thought of as actors or nodes, and the amount of each commodity exported from each nation to each of the other 169 can be thought of as the strength of a directed tie from the focal nation to the other. A social scientist might be interested in whether the "structures" of trade in mineral products are more similar to one another than, the structure of trade in mineral products are to vegetable products. To answer this fairly simple (but also pretty important) question, a huge amount of manipulation of the data is necessary. It could take, literally, years to do by hand. It can be done by a computer in a few minutes.

The third, and final reason for using "formal" methods (mathematics and graphs) for representing social network data is that the techniques of graphing and the rules of mathematics themselves suggest things that we might look for in our data — things that might not have occurred to us if we presented our data using descriptions in words. Again, allow me a simple example.

Suppose we were describing the structure of close friendship in a group of four people: Bob, Carol, Ted, and Alice. This is easy enough to do with words. Suppose that Bob likes Carol and

Ted, but not Alice; Carol likes Ted, but neither Bob nor Alice; Ted likes all three of the other members of the group; and Alice likes only Ted (this description should probably strike you as being a description of a very unusual social structure).

We could also describe this pattern of liking ties with an actor-by-actor matrix where the rows represent choices by each actor. We will put in a "1" if an actor likes another, and a "0" if they don't. Such a matrix would look like:

	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

There are lots of things that might immediately occur to us when we see our data arrayed in this way, that we might not have thought of from reading the description of the pattern of ties in words. For example, our eye is led to scan across each row; we notice that Ted likes more people than Bob, than Alice and Carol. Is it possible that there is a pattern here? Are men are more likely to report ties of liking than women are (actually, research literature suggests that this is not generally true). Using a "matrix representation" also immediately raises a question: the locations on the main diagonal (e.g. Bob likes Bob, Carol likes Carol) are empty. Is this a reasonable thing? Or, should our description of the pattern of liking in the group include some statements about "self-liking"? There isn't any right answer to this question. My point is just that using a matrix to represent the pattern of ties among actors may let us see some patterns more easily, and may cause us to ask some questions (and maybe even some useful ones) that a verbal description doesn't stimulate.

Summary of chapter 2

There are three main reasons for using "formal" methods in representing social network data:

Matrices and graphs are compact and systematic.

They summarize and present a lot of information quickly and easily; and they force us to be systematic and complete in describing patterns of social relations.

Matrices and graphs allow us to apply computers to analyzing data.

This is helpful because doing systematic analysis of social network data can be extremely tedious if the number of actors or number of types of relationships among the actors is large. Most of the work is dull, repetitive, and uninteresting, but requires accuracy. This is exactly the sort of thing

that computers do well, and we don't.

Matrices and graphs have rules and conventions.

Sometimes these are just rules and conventions that help us communicate clearly. But sometimes the rules and conventions of the language of graphs and mathematics themselves lead us to see things in our data that might not have occurred to us to look for if we had described our data only with words.

So, we need to learn the basics of representing social network data using matrices and graphs. That's what the next chapter is about.

3. Using Graphs to Represent Social Relations

Introduction: Representing Networks with Graphs

Social network analysts use two kinds of tools from mathematics to represent information about patterns of ties among social actors: graphs and matrices. On this page, we will learn enough about graphs to understand how to represent social network data. On the next page, we will look at matrix representations of social relations. With these tools in hand, we can understand most of the things that network analysts do with such data (for example, calculate precise measures of "relative density of ties").

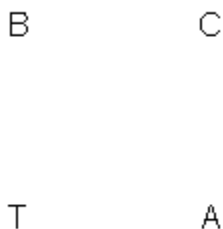
There is a lot more to these topics than we will cover here; mathematics has whole sub-fields devoted to "graph theory" and to "matrix algebra." Social scientists have borrowed just a few things that they find helpful for describing and analyzing patterns of social relations.

A word of warning: there is a lot of specialized terminology here that you do need to learn. It's worth the effort, because we can represent some important ideas about social structure in quite simple ways, once the basics have been mastered.

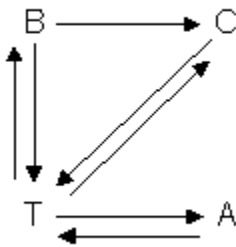
Graphs and Sociograms

There are lots of different kinds of "graphs." Bar charts, pie charts, line and trend charts, and many other things are called graphs and/or graphics. Network analysis uses (primarily) one kind of graphic display that consists of points (or nodes) to represent actors and lines (or edges) to represent ties or relations. When sociologists borrowed this way of graphing things from the mathematicians, they re-named their graphics "sociograms." Mathematicians know the kind of graphic displays by the names of "directed graphs" "signed graphs" or simply "graphs."

There are a number of variations on the theme of sociograms, but they all share the common feature of using a labeled circle for each actor in the population we are describing, and line segments between pairs of actors to represent the observation that a tie exists between the two. Let's suppose that we are interested in summarizing who nominates whom as being a "friend" in a group of four people (Bob, Carol, Ted, and Alice). We would begin by representing each actor as a "node" with a label (sometimes nodes are represented by labels in circles or boxes).



We collected our data about friendship ties by asking each member of the group (privately and confidentially) who they regarded as "close friends" from a list containing each of the other members of the group. Each of the four people could choose none to all three of the others as "close friends." As it turned out, in our (fictitious) case, Bob chose Carol and Ted, but not Alice; Carol chose only Ted; Ted chose Bob and Carol and Alice; and Alice chose only Ted. We would represent this information by drawing an arrow from the chooser to each of the chosen, as in the next graph:



Kinds of Graphs

Now we need to introduce some terminology to describe different kinds of graphs. This particular example above is a *binary* (as opposed to a signed or ordinal or valued) and *directed* (as opposed to a co-occurrence or co-presence or bonded-tie) graph. The social relations being described here are also *simplex* (as opposed to multiplex).

Levels of Measurement: Binary, Signed, and Valued Graphs

In describing the pattern of who describes whom as a close friend, we could have asked our question in several different ways. If we asked each respondent "is this person a close friend or not," we are asking for a binary choice: each person is or is not chosen by each interviewee. Many social relationships can be described this way: the only thing that matters is whether a tie exists or not. When our data are collected this way, we can graph them simply: an arrow represents a choice that was made, no arrow represents the absence of a choice. But, we could have asked the question a second way: "for each person on this list, indicate whether you like, dislike, or don't care." We might assign a + to indicate "liking," zero to indicate "don't care" and - to indicate dislike. This kind of data is called "signed" data. The graph with signed data uses a + on the arrow to indicate a positive choice, a - to indicate a negative choice, and no arrow to indicate neutral or indifferent. Yet another approach would have been to ask: "rank the three people on this list in order of who you like most, next most, and least." This would give us "rank order" or "ordinal" data describing the strength of each friendship choice. Lastly, we could have asked: "on a scale from minus one hundred to plus one hundred - where minus 100 means you hate this person, zero means you feel neutral, and plus 100 means you love this person - how do you feel about...". This would give us information about the value of the strength of each choice on a (supposedly, at least) ratio level of measurement. With either an ordinal or valued graph, we would put the measure of the strength of the relationship on the arrow in the diagram.

Directed or "Bonded" Ties in the Graph

In our example, we asked each member of the group to choose which others in the group they regarded as close friends. Each person (ego) then is being asked about ties or relations that they themselves direct toward others (alters). Each alter does not necessarily feel the same way about each tie as ego does: Bob may regard himself as a good friend to Alice, but Alice does not necessarily regard Bob as a good friend. It is very useful to describe many social structures as being composed of "directed" ties (which can be binary, signed, ordered, or valued). Indeed, most social processes involve sequences of directed actions. For example, suppose that person A directs a comment to B, then B directs a comment back to A, and so on. We may not know the order in which actions occurred (i.e. who started the conversation), or we may not care. In this example, we might just want to know that "A and B are having a conversation." In this case, the tie or relation "in conversation with" necessarily involves both actors A and B. Both A and B are "co-present" or "co-occurring" in the relation of "having a conversation." Or, we might also describe the situation as being one of an the social institution of a "conversation" that by definition involves two (or more) actors "bonded" in an interaction (Berkowitz).

"Directed" graphs use the convention of connecting nodes or actors with arrows that have arrowheads, indicating who is directing the tie toward whom. This is what we used in the graphs above, where individuals (egos) were directing choices toward others (alters). "Co-occurrence" or "co-presence" or "bonded-tie" graphs use the convention of connecting the pair of actors involved in the relation with a simple line segment (no arrowhead). Be careful here, though. In a directed graph, Bob could choose Ted, and Ted choose Bob. This would be represented by headed arrows going from Bob to Ted, and from Ted to Bob, or by a double-headed arrow. But, this represents a different meaning from a graph that shows Bob and Ted connected by a single line segment without arrowheads. Such a graph would say "there is a relationship called close friend which ties Bob and Ted together." The distinction can be subtle, but it is important in some analyses.

Simplex or Multiplex Relations in the Graph

The information that we have represented about the social structure of our group of four people is pretty simple. That is, it describes only one type of tie or relation - choice of a close friend. A graph that represents a single kind of relation is called a simplex graph. Social structures, however, are often multiplex. That is, there are multiple different kinds of ties among social actors. Let's add a second kind of relation to our example. In addition to friendship choices, lets also suppose that we asked each person whether they are kinfolk of each of the other three. Bob identifies Ted as kin; Ted identifies Bob; and Ted and Alice identify one another (the full story here might be that Bob and Ted are brothers, and Ted and Alice are spouses). We could add this information to our graph, using a different color or different line style to represent the second type of relation ("is kin of...").

We can see that the second kind of tie, "kinship" re-enforces the strength of the relationships between Bob and Ted and between Ted and Alice (or, perhaps, the presence of a kinship tie explains the mutual choices as good friends). The reciprocated friendship tie between Carol and Ted, however, is different, because it is not re-enforced by a kinship bond.

Of course, if we were examining many different kinds of relationships among the same set of actors, putting all of this information into a single graph might make it too difficult to read, so we might, instead, use multiple graphs with the actors in the same locations in each. We might also want to represent the multiplexity of the data in some simpler way. We could use lines of different thickness to represent how many ties existed between each pair of actors; or we could count the number of relations that were present for each pair and use a valued graph.

Summary of chapter 3

A graph (sometimes called a sociogram) is composed of nodes (or actors or points) connected by edges (or relations or ties). A graph may represent a single type of relations among the actors (simplex), or more than one kind of relation (multiplex). Each tie or relation may be directed (i.e. originates with a source actor and reaches a target actor), or it may be a tie that represents co-occurrence, co-presence, or a bonded-tie between the pair of actors. Directed ties are represented with arrows, bonded-tie relations are represented with line segments. Directed ties may be reciprocated (A chooses B and B chooses A); such ties can be represented with a double-headed arrow. The strength of ties among actors in a graph may be nominal or binary (represents presence or absence of a tie); signed (represents a negative tie, a positive tie, or no tie); ordinal (represents whether the tie is the strongest, next strongest, etc.); or valued (measured on an interval or ratio level). In speaking the position of one actor or node in a graph to other actors or nodes in a graph, we may refer to the focal actor as "ego" and the other actors as "alters."

Review questions for chapter 3

1. What are "nodes" and "edges"? In a sociogram, what is used for nodes? for edges?
2. How do valued, binary, and signed graphs correspond to the "nominal" "ordinal" and "interval" levels of measurement?
3. Distinguish between directed relations or ties and "bonded" relations or ties.
4. How does a reciprocated directed relation differ from a "bonded" relation?
5. Give an example of a multi-plex relation. How can multi-plex relations be represented in graphs?

Application questions for chapter 3

1. Think of the readings from the first part of the course. Did any studies present graphs? If they did, what kinds of graphs were they (that is, what is the technical description of the kind of graph or matrix). Pick one article and show what a graph of its data would look like.
2. Suppose that I was interested in drawing a graph of which large corporations were networked with one another by having the same persons on their boards of directors. Would it make more sense to use "directed" ties, or "bonded" ties for my graph? Can you think of a kind of relation among large corporations that would be better represented with directed ties?
3. Think of some small group of which you are a member (maybe a club, or a set of friends, or

people living in the same apartment complex, etc.). What kinds of relations among them might tell us something about the social structures in this population? Try drawing a graph to represent one of the kinds of relations you chose. Can you extend this graph to also describe a second kind of relation? (e.g. one might start with "who likes whom?" and add "who spends a lot of time with whom?").

4. Make graphs of a "star" network, a "line," and a "circle." Think of real world examples of these kinds of structures where the ties are directed and where they are bonded, or undirected. What does a strict hierarchy look like? What does a population that is segregated into two groups look like?

4. Using Matrices to Represent Social Relations

Introduction to chapter 4

Graphs are very useful ways of presenting information about social networks. However, when there are many actors and/or many kinds of relations, they can become so visually complicated that it is very difficult to see patterns. It is also possible to represent information about social networks in the form of matrices. Representing the information in this way also allows the application of mathematical and computer tools to summarize and find patterns. Social network analysts use matrices in a number of different ways. So, understanding a few basic things about matrices from mathematics is necessary. We'll go over just a few basics here that cover most of what you need to know to understand what social network analysts are doing. For those who want to know more, there are a number of good introductory books on matrix algebra for social scientists.

What is a Matrix?

To start with, a matrix is nothing more than a rectangular arrangement of a set of elements (actually, it's a bit more complicated than that, but we will return to matrices of more than two dimensions in a little bit). Rectangles have sizes that are described by the number of rows of elements and columns of elements that they contain. A "3 by 6" matrix has three rows and six columns; an "I by j" matrix has I rows and j columns. Here are empty 2 by 4 and 4 by 2 matrices:

2 by 4

1,1	1,2	1,3	1,4
2,1	2,2	2,3	2,4

4 by 2

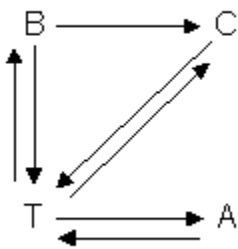
1,1	1,2
2,1	2,2
3,1	3,2
4,1	4,2

The elements of a matrix are identified by their "addresses." Element 1,1 is the entry in the first row and first column; element 13,2 is in the 13th row and is the second element of that row. The cell addresses have been entered as matrix elements in the two examples above. Matrices are often represented as arrays of elements surrounded by vertical lines at their left and right, or square brackets at the left and right. In html (the language used to prepare web pages) it is easier to use "tables" to represent matrices. Matrices can be given names; these names are usually presented as capital bold-faced letters. Social scientists using matrices to represent social networks often dispense with the mathematical conventions, and simply show their data as an array of labeled rows and columns. The labels are not really part of the matrix, but are simply for clarity of presentation. The matrix below, for example, is a 4 by 4 matrix, with additional labels:

	Bob	Carol	Ted	Alice
Bob	---	1	0	0
Carol	1	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

The "Adjacency" Matrix

The most common form of matrix in social network analysis is a very simple one composed of as many rows and columns as there are actors in our data set, and where the elements represent the ties between the actors. The simplest and most common matrix is binary. That is, if a tie is present, a one is entered in a cell; if there is no tie, a zero is entered. This kind of a matrix is the starting point for almost all network analysis, and is called an "adjacency matrix" because it represents who is next to, or adjacent to whom in the "social space" mapped by the relations that we have measured. By convention, in a directed graph, the sender of a tie is the row and the target of the tie is the column. Let's look at a simple example. The directed graph of friendship choices among Bob, Carol, Ted, and Alice looks like this:



We can since the ties are measured at the nominal level (that is, the data are binary choice data), we can represent the same information in a matrix that looks like:

	B	C	T	A
B	---	1	1	0
C	0	---	1	0
T	1	1	---	1
A	0	0	1	---

Remember that the rows represent the source of directed ties, and the columns the targets; Bob chooses Carol here, but Carol does not choose Bob. This is an example of an "asymmetric" matrix that represents directed ties (ties that go from a source to a receiver). That is, the element i,j does not necessarily equal the element j,i . If the ties that we were representing in our matrix were "bonded-ties" (for example, ties representing the relation "is a business partner of" or "co-occurrence or co-presence," (e.g. where ties represent a relation like: "serves on the same board of directors as") the matrix would necessarily be symmetric; that is element i,j would be equal to element j,i .

Binary choice data are usually represented with zeros and ones, indicating the presence or absence of each logically possible relationship between pairs of actors. Signed graphs are represented in matrix form (usually) with -1, 0, and +1 to indicate negative relations, no or neutral relations, and positive relations. When ties are measured at the ordinal or interval level, the numeric magnitude of the measured tie is entered as the element of the matrix. As we discussed in chapter one, other forms of data are possible (multi-category nominal, ordinal with more than three ranks, full-rank order nominal). These other forms, however, are rarely used in sociological studies, and we won't give them very much attention.

In representing social network data as matrices, the question always arises: what do I do with the elements of the matrix where $i = j$? That is, for example, does Bob regard himself as a close friend of Bob? This part of the matrix is called the *main diagonal*. Sometimes the value of the main diagonal is meaningless, and it is ignored (and left blank). Sometimes, however, the main diagonal can be very important, and can take on meaningful values. This is particularly true when the rows and columns of our matrix are "super-nodes" or "blocks." More on that in a minute.

It is often convenient to refer to certain parts of a matrix using shorthand terminology. If I take all of the elements of a row (e.g. who Bob chose as friends: 1,1,1,0) I am examining the "*row vector*" for Bob. If I look only at who chose Bob as a friend (the first column, or 1,0,1,0), I am examining the "*column vector*" for Bob. It is sometimes useful to perform certain operations on

row or column vectors. For example, if I summed the elements of the column vectors in this example, I would be measuring how "popular" each node was (in terms of how often they were the target of a directed friendship tie).

Matrix Permutation, Blocks, and Images

It is also helpful, sometimes, to rearrange the rows and columns of a matrix so that we can see patterns more clearly. Shifting rows and columns (if you want to rearrange the rows, you must rearrange the columns in the same way, or the matrix won't make sense for most operations) is called "permutation" of the matrix.

Our original data look like:

	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

Let's rearrange (permute) this so that the two males and the two females are adjacent in the matrix. *Matrix permutation* simply means to change the order of the rows and columns. Since the matrix is symmetric, if I change the position of a row, I must also change the position of the corresponding column.

	Bob	Ted	Carol	Alice
Bob	---	1	1	0
Ted	1	---	1	1
Carol	0	1	---	0
Alice	0	1	0	---

None of the elements have had their values changed by this operation or rearranging the rows and columns, we have just shifted things around. We've also highlighted some sections of the matrix. Each colored section is referred to as a *block*. Blocks are formed by passing dividing lines through the matrix (e.g. between Ted and Carol) rows and columns. Passing these dividing lines through the matrix is called *partitioning the matrix*. Here we have partitioned by the sex of the actors. Partitioning is also sometimes called "blocking the matrix," because partitioning produces blocks.

This kind of grouping of cells is often done in network analysis to understand how some sets of actors are "embedded" in social roles or in larger entities. Here, for example, we can see that all occupants of the social role "male" choose each other as friends; no females choose each other as

friends, and that males are more likely to choose females (3 out of 4 possibilities are selected) than females are to choose males (only 2 out of 4 possible choices). We have grouped the males together to create a "partition" or "super-node" or "social role" or "block." We often partition social network matrices in this way to identify and test ideas about how actors are "embedded" in social roles or other "contexts."

We might wish to dispense with the individual nodes altogether, and examine only the positions or roles. If we calculate the proportion of all ties within a block that are present, we can create a *block density matrix*. In doing this, we have ignored self-ties in the current example.

Block Density Matrix		
	Male	Female
Male	1.00	0.75
Female	0.50	0.00

We may wish to summarize the information still further by using *block image* or *image matrix*. If the density in a block is greater than some amount (we often use the average density for the whole matrix as a cut-off score, in the current example the density is .58), we enter a "1" in a cell of the blocked matrix, and a "0" otherwise. This kind of simplification is called the "image" of the blocked matrix.

Image Matrix		
	Male	Female
Male	1	1
Female	0	0

Images of blocked matrices are powerful tools for simplifying the presentation of complex patterns of data. Like any simplifying procedure, good judgement must be used in deciding how to block and what cut-off to use to create images -- or we may lose important information.

Doing Mathematical Operations on Matrices

Representing the ties among actors as matrices can help us to see patterns by performing simple manipulations like summing row vectors or partitioning the matrix into blocks. Social network

analysts use a number of other mathematical operations that can be performed on matrices for a variety of purposes (matrix addition and subtraction, transposes, inverses, matrix multiplication, and some other more exotic stuff like determinants and eigenvalues and vectors). Without trying to teach you matrix algebra, it is useful to know at least a little bit about some of these mathematical operations, and what they are used for in social network analysis.

Transposing a matrix

This simply means to exchange the rows and columns so that i becomes j , and *vice versa*. If we take the transpose of a directed adjacency matrix and examine its row vectors (you should know all this jargon by now!), we are looking at the sources of ties directed at an actor. The degree of similarity between an adjacency matrix and the transpose of that matrix is one way of summarizing the degree of symmetry in the pattern of relations among actors. That is, the correlation between an adjacency matrix and the transpose of that matrix is a measure of the degree of reciprocity of ties (think about that assertion a bit). Reciprocity of ties can be a very important property of a social structure because it relates to both the balance and to the degree and form of hierarchy in a network.

Taking the inverse of a matrix

This is a mathematical operation that finds a matrix which, when multiplied by the original matrix, yields a new matrix with ones in the main diagonal and zeros elsewhere (which is called an identity matrix). Without going any further into this, you can think of the inverse of a matrix as being sort of the "opposite of" the original matrix. Matrix inverses are used mostly in calculating other things in social network analysis. They are sometimes interesting to study in themselves, however. It is sort of like looking at black lettering on white paper versus white lettering on black paper: sometimes you see different things.

Matrix addition and matrix subtraction

These are the easiest of matrix mathematical operations. One simply adds together or subtracts each corresponding i,j element of the two (or more) matrices. Of course, the matrices that this is being done to have to have the same numbers of i and j elements (this is called "conformable" to addition and subtraction) - and, the values of i and j have to be in the same order in each matrix. Matrix addition and subtraction are most often used in network analysis when we are trying to simplify or reduce the complexity of multiplex data to simpler forms. If I had a symmetric matrix that represented the tie "exchanges money" and another that represented the relation "exchanges goods" I could add the two matrices to indicate the intensity of the exchange relationship. Pairs with a score of zero would have no relationship, those with a "1" would be involved in either barter or commodity exchange, and those with a "2" would have both barter and commodity exchange relations. If I subtracted the "goods" exchange matrix from the "money exchange" matrix, a score of -1 would indicate pairs with a barter relationship; a score of zero would indicate either no relationship or a barter and commodity tie; a score of +1 would indicate pairs with only a commodified exchange relationship. For different research questions, either or both approaches might be useful.

Matrix correlation and regression

Correlation and regression of matrices are ways to describe association or similarity between the matrices. Correlation looks at two matrices and asks, "how similar are these?" Regression uses the scores in one matrix to predict the scores in the other. If we want to know how similar matrix A is to matrix B, we take each element i,j of matrix A and pair it with the same element i,j of matrix B, and calculate a measure of association (which measure one uses, depends upon the level of measurement of the ties in the two matrices). Matrix regression does the same thing with the elements of one matrix being defined as the observations of the dependent variable and the corresponding i,j elements of other matrices as the observations of independent variables. These tools are used by network analysts for the same purposes that correlation and regression are used by non-network analysts: to assess the similarity or correspondence between two distributions of scores. We might, for example, ask how similar is the pattern of friendship ties among actors to the pattern of kinship ties. We might wish to see the extent to which one can predict which nations have full diplomatic relations with one another on the basis of the strength of trade flows between them.

Matrix multiplication and Boolean matrix multiplication

Matrix multiplication is a somewhat unusual operation, but can be very useful for the network analyst. You will have to be a bit patient here. First we need to show you how to do matrix multiplication and a few important results (like what happens when you multiply an adjacency matrix times itself, or raise it to a power). Then, we will try to explain why this is useful.

To multiply two matrices, they must be "conformable" to multiplication. This means that the number of rows in the first matrix must equal the number of columns in the second. Usually network analysis uses adjacency matrices, which are square, and hence, conformable for multiplication. To multiply two matrices, begin in the upper left hand corner of the first matrix, and multiply every cell in the first row of the first matrix by the values in each cell of the first column of the second matrix, and sum the results. Proceed through each cell in each row in the first matrix, multiplying by the column in the second. To perform a Boolean matrix multiplication, proceed in the same fashion, but enter a zero in the cell if the multiplication product is zero, and one if it is not zero.

Suppose we wanted to multiply these two matrices:

0	1
2	3
4	5

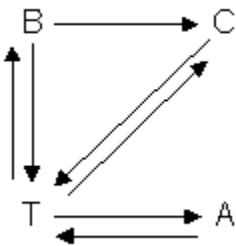
times

6	7	8
9	10	11

The result is:

$(0*6)+(1*9)$	$(0*7)+(1*10)$	$(0*8)+(1*11)$
$(2*6)+(3*9)$	$(2*7)+(3*10)$	$(2*8)+(3*11)$
$(4*6)+(5*9)$	$(4*7)+(5*10)$	$(4*8)+(5*11)$

The mathematical operation in itself doesn't interest us here (any number of programs can perform matrix multiplication). But, the operation is useful when applied to an adjacency matrix. Consider our four friends again:



The adjacency matrix for the four actors B, C, T, and A (in that order) is:

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

Another way of thinking about this matrix is to notice that it tells us whether there is a path from each actor to each actor. A one represents the presence of a path, a zero represents the lack of a path. The adjacency matrix is exactly what its name suggests -- it tells us which actors are adjacent, or have a direct path from one to the other.

Now suppose that we multiply this adjacency matrix times itself (i.e. raise the matrix to the 2nd power, or square it).

$(0*0)+(1*0)+(1*1)+(0*0)$	$(0*1)+(1*0)+(1*1)+(0*0)$	$(0*1)+(1*1)+(1*0)+(0*1)$	$(0*0)+(1*0)+(1*1)+(0*0)$
$(0*0)+(0*0)+(1*1)+(0*0)$	$(0*1)+(0*0)+(1*1)+(0*0)$	$(0*1)+(0*1)+(1*0)+(0*1)$	$(0*0)+(0*0)+(1*1)+(0*0)$
$(1*0)+(1*0)+(0*1)+(1*0)$	$(1*1)+(1*0)+(0*1)+(1*0)$	$(1*1)+(1*1)+(0*0)+(1*1)$	$(1*0)+(1*0)+(0*1)+(1*0)$
$(0*0)+(0*0)+(1*1)+(0*0)$	$(0*1)+(0*0)+(1*1)+(0*0)$	$(0*1)+(0*1)+(1*0)+(0*1)$	$(0*0)+(0*0)+(1*1)+(0*0)$

or:

1	1	1	1
1	1	0	1
0	1	3	0
1	1	0	1

This matrix (i.e. the adjacency matrix squared) counts the number of pathways between two nodes that are of length two. Stop for a minute and verify this assertion. For example, note that actor "B" is connected to each of the other actors by a pathway of length two; and that there is no more than one such pathway to any other actor. Actor T is connected to himself by pathways of length two, three times. This is because actor T has reciprocal ties with each of the other three actors. There is no pathway of length two from T to B (although there is a pathway of length one).

So, the adjacency matrix tells us how many paths of length one are there from each actor to each other actor. The adjacency matrix squared tells us how many pathways of length two are there from each actor to each other actor. It is true (but we won't show it to you) that the adjacency matrix cubed counts the number of pathways of length three from each actor to each other actor. And so on...

If we calculated the Boolean product, rather than the simple matrix product, the adjacency matrix squared would tell us whether there was a path of length two between two actors (not how many such paths there were). If we took the Boolean squared matrix and multiplied it by the adjacency

matrix using Boolean multiplication, the result would tell us which actors were connected by one or more pathways of length three. And so on...

Now, finally: why should you care?

Some of the most fundamental properties of a social network have to do with how connected the actors are to one another. Networks that have few or weak connections, or where some actors are connected only by pathways of great length may display low solidarity, a tendency to fall apart, slow response to stimuli, and the like. Networks that have more and stronger connections with shorter paths among actors may be more robust and more able to respond quickly and effectively. Measuring the number and lengths of pathways among the actors in a network allow us to index these important tendencies of whole networks.

Individual actors' positions in networks are also usefully described by the numbers and lengths of pathways that they have to other actors. Actors who have many pathways to other actors may be more influential with regard to them. Actors who have short pathways to more other actors may be more influential or central figures. So, the number and lengths of pathways in a network are very important to understanding both individual's constraints and opportunities, and for understanding the behavior and potentials of the network as a whole.

There are many measures of individual position and overall network structure that are based on whether there are pathways of given lengths between actors, the length of the shortest pathway between two actors, and the numbers of pathways between actors. Indeed, most of the basic measures of networks (chapter 5), measures of centrality and power (chapter 6), and measures of network groupings and substructures (chapter 7) are based on looking at the numbers and lengths of pathways among actors.

Summary of chapter 4

Matrices are collections of elements into rows and columns. They are often used in network analysis to represent the adjacency of each actor to each other actor in a network. An adjacency matrix is a square actor-by-actor ($i=j$) matrix where the presence of pairwise ties are recorded as elements. The main diagonal or "self-tie of an adjacency matrix is often ignored in network analysis.

Sociograms, or graphs of networks can be represented in matrix form, and mathematical operations can then be performed to summarize the information in the graph. Vector operations, blocking and partitioning, and matrix mathematics (inverses, transposes, addition, subtraction, multiplication and Boolean multiplication), are mathematical operations that are sometimes helpful to let us see certain things about the patterns of ties in social networks.

Social network data are often multiplex (i.e. there are multiple kinds of ties among the actors). Such data are represented as a series of matrices of the same dimension with the actors in the same position in each matrix. Many of the same tools that we can use for working with a single matrix (matrix addition and correlation, blocking, etc.) Are helpful for trying to summarize and see the patterns in multiplex data.

Once a pattern of social relations or ties among a set of actors has been represented in a formal way (graphs or matrices), we can define some important ideas about social structure in quite precise ways using mathematics for the definitions. In the remainder of the readings on the pages in this site, we will look at how social network analysts have formally translated some of the core concepts that social scientists use to describe social structures.

Review questions for chapter 4

1. A matrix is "3 by 2." How many columns does it have? How many rows?
2. Adjacency matrices are "square" matrices. Why?
3. There is a "1" in cell 3,2 of an adjacency matrix representing a sociogram. What does this tell us?
4. What does it mean to "permute" a matrix, and to "block" it?

Application questions for chapter 4

1. Think of the readings from the first part of the course. Did any studies present matrices? If they did, what kinds of matrices were they (that is, what is the technical description of the kind of graph or matrix). Pick one article, and show what the data would look like, if represented in matrix form.
2. Think of some small group of which you are a member (maybe a club, or a set of friends, or people living in the same apartment complex, etc.). What kinds of relations among them might tell us something about the social structures in this population? Try preparing a matrix to represent one of the kinds of relations you chose. Can you extend this matrix to also describe a second kind of relation? (E.g. one might start with "who likes whom?" and add "who spends a lot of time with whom?").
3. Using the matrices you created in the previous question, does it make sense to leave the diagonal "blank," or not, in your case? Try permuting your matrix, and blocking it.
4. Can you make an adjacency matrix to represent the "star" network? what about the "line" and "circle." Look at the ones and zeros in these matrices -- sometimes we can recognize the presence of certain kinds of social relations by these "digital" representations. What does a strict hierarchy look like? What does a population that is segregated into two groups look like?

5. Basic Properties of Networks and Actors

Introduction: Basic Properties of 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 grounded in, and enacted by local interactions among actors. As we examine some of the basic concepts and definitions of network analysis in this and the next several chapters, this duality of individual and structure will be highlighted again and again.

In this chapter we will examine some of the most obvious and least complex ideas of formal network analysis methods. Despite the simplicity of the ideas and definitions, there are good theoretical reasons (and some empirical evidence) to believe that these basic properties of social networks have very important consequences. For both individuals and for structures, one main question is connections. Typically, some actors have lots of connections, others have fewer. Particularly as populations become larger, not all the possible connections are present -- there are "structural holes." The extent to which individuals are connected to others, and the extent to which the network as a whole is integrated are two sides of the same coin.

Differences among individuals in how connected they are can be extremely consequential for understanding their attributes and behavior. More connections often mean that individuals are exposed to more, and more diverse information. Highly connected individuals may be more influential, and may be more influenced by others. Differences among whole populations in how connected they are can be quite consequential as well. Disease and rumors spread more quickly where there are high rates of connection. But, so to does useful information. More connected populations may be better able to mobilize their resources, and may be better able to bring multiple and diverse perspectives to bear to solve problems. In between the individual and the whole population, there is another level of analysis -- that of "composition." Some populations may be composed of individuals who are all pretty much alike in the extent to which they are connected. Other populations may display sharp differences, with a small elite of central and highly connected persons, and larger masses of persons with fewer connections. Differences in connections can tell us a good bit about the stratification order of social groups.

Because most individuals are not usually connected directly to most other individuals in a population, it can be quite important to go beyond simply examining the immediate connections of actors, and the overall density of direct connections in populations. The second major (but closely related) set of approaches that we will examine in this chapter have to do with the idea of the distance between actors (or, conversely how close they are to one another). Some actors may be able to reach most other members of the population with little effort: they tell their friends, who tell their friends, and "everyone" knows. Other actors may have difficulty being heard. They may tell people, but the people they tell are not well connected, and the message doesn't go far. Thinking about it the other way around, if all of my friends have one another as friends, my

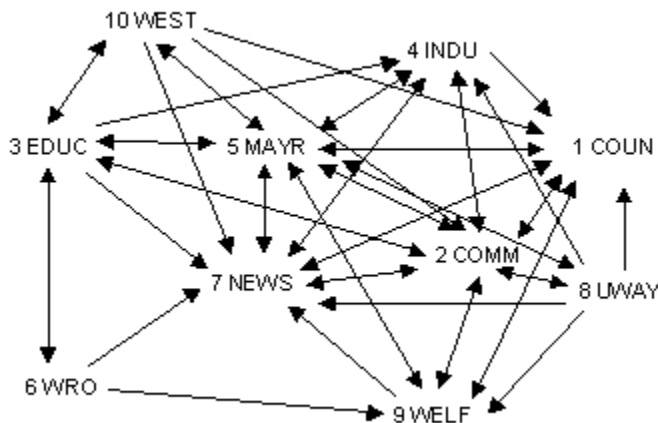
network is fairly limited -- even though I may have quite a few friends. But, if my friends have many non-overlapping connections, the range of my connection is expanded. If individuals differ in their closeness to other actors, then the possibility of stratification along this dimension arises. Indeed, one major difference among "social classes" is not so much in the number of connections that actors have, but in whether these connections overlap and "constrain" or extent outward and provides "opportunity." Populations as a whole, then, can also differ in how close actors are to other actors, on the average. Such differences may help us to understand diffusion, homogeneity, solidarity, and other differences in macro properties of social groups.

Social network methods have a vocabulary for describing connectedness and distance that might, at first, seem rather formal and abstract. This is not surprising, as many of the ideas are taken directly from the mathematical theory of graphs. But it is worth the effort to deal with the jargon. The precision and rigor of the definitions allow us to communicate more clearly about important properties of social structures -- and often lead to insights that we would not have had if we used less formal approaches.

An Example

The basic properties of networks are easier to learn and understand by example. Studying an example also shows sociologically meaningful applications of the formalisms. In this chapter, we will look at a single directed binary network that describes the flow of information among 10 formal organizations concerned with social welfare issues in one mid-western U.S. city (Knoke and Burke). Of course, network data come in many forms (undirected, multiple ties, valued ties, etc.) and one example can't capture all of the possibilities. Still, it can be rather surprising how much information can be "squeezed out" of a single binary matrix by using basic graph concepts.

For small networks, it is often useful to examine graphs. Here is the di-graph for the Knoke information exchange data:



Your trained eye should immediately perceive a number of things in looking at the graph. There are a limited number of actors here (ten, actually), and all of them are "connected." But, clearly

not every possible connection is present, and there are "structural holes" (or at least "thin spots" in the fabric). There appear to be some differences among the actors in how connected they are (compare actor number 7, a newspaper, to actor number 6, a welfare rights advocacy organization). If you look closely, you can see that some actor's connections are likely to be reciprocated (that is, if A shares information with B, B also shares information with A); some other actors (e.g. 6 and 10, 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 at quite some "distance" from other actors. There appear to be groups of actors who differ in this regard (1, 2, 4, 5, and 6 seem to be in the center of the action, 6, 9, and 10 seem to be more peripheral).

A careful look at the graph can be very useful in getting an intuitive grasp of the important features of a social network. With larger populations or more connections, however, graphs may not be much help. Looking at a graph can give a good intuitive sense of what is going on, but our descriptions of what we see are rather imprecise (the previous paragraph is an example of this). To get more precise, and to use computers to apply algorithms to calculate mathematical measures of graph properties, it is necessary to work with the adjacency matrix instead of the graph.

	1COUN	2COMM	3EDUC	4INDU	5MAYR	6WRO	7NEWS	8UWAY	9WELF	10WEST
1COUN	---	1	0	0	1	0	1	0	1	0
2COMM	1	---	1	1	1	0	1	1	1	0
3EDUC	0	1	---	1	1	1	1	0	0	1
4INDU	1	1	0	---	1	0	1	0	0	0
5MAYR	1	1	1	1	---	0	1	1	1	1
6WRO	0	0	1	1	1	---	1	0	1	0
7NEWS	0	1	0	1	1	0	---	0	0	0
8UWAY	1	1	0	1	1	0	1	---	1	0
9WELF	0	1	0	0	1	0	1	0	---	0
10WEST	1	1	1	0	1	0	1	0	0	---

There are ten rows and columns, the data are binary, and the matrix is asymmetric. As we mentioned in the chapter on using matrices to represent networks, the row is treated as the source of information and the column as the receiver. By doing some very simple operations on this matrix it is possible to develop systematic and useful index numbers, or measures, of some of the network properties that our eye discerns in the graph.

Connections

Since networks are defined by their actors and the connections among them, it is useful to begin our description of networks by examining these very simple properties. Focusing first on the network as a whole, one might be interested in the number of actors, the number of connections that are possible, and the number of connections that are actually present. Differences in the size of networks, and how connected the actors are tell us two things about human populations that are critical. Small groups differ from large groups in many important ways -- indeed, population size is one of the most critical variables in all sociological analyses. Differences in how connected the actors in a population are may be a key indicator of the solidarity, "moral density," and "complexity" of the social organization of a population.

Individuals, as well as whole networks, differ in these basic demographic features. Individual actors may have many or few ties. Individuals may be "sources" of ties, "sinks" (actors that receive ties, but don't send them), or both. These kinds of very basic differences among actor's immediate connections may be critical in explaining how they view the world, and how the world views them. The number and kinds of ties that actors have are a basis for similarity or dissimilarity to other actors -- and hence to possible differentiation and stratification. The number and kinds of ties that actors have are keys to determining how much their embeddedness in the network constrains their behavior, and the range of opportunities, influence, and power that they have.

It is possible that a network is not completely connected. This is the question of reachability. There may be two or more disconnected groups in the population. If it is not possible for all actors to "reach" all other actors, then our population consists of more than one group. The groups may occupy the same space, or have the same name, but not all members are connected. Obviously, such divisions in populations may be sociologically significant. To the extent that a network is not connected, there may be a structural basis for stratification and conflict. At the individual level, the degree to which an actor can reach others indicates the extent to which that individual is separated from the whole, or the extent to which that actor is isolated. Such isolation may have social-psychological significance. If an actor cannot reach, or cannot be reached by another, then there can be no learning, support, or influence between the two.

Another useful way to look at networks as a whole, and the way in which individuals are embedded in them, is to examine the local structures. The most common approaches here has been to look at dyads (i.e. sets of two actors) and triads (i.e. sets of three actors).

With directed data, there are four possible dyadic relationships: A and B are not connected, A sends to B, B sends to A, or A and B send to each other (with undirected data, there are only two possible relationships - no tie or tie). It may be useful to look at each actor in terms of the kinds of dyadic relationships in which they are involved. An actor that sends, but does not receive ties may be quite different from one who both sends and receives. A common interest in looking at dyadic relationships is the extent to which ties are reciprocated. 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. Of course, one can examine the entire network, as well as individual differences. In one sense, 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).

Small group theorists argue that many of the most interesting and basic questions of social structure arise with regard to triads. Triads allow for a much wider range of possible sets of relations (with directed data, there are actually 64 possible types of relations among 3 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, 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. In particular, we may be interested in the proportion of triads that are "transitive" (that is, display a type of balance where, if A directs a tie to B, and B directs a tie to C, then A also directs a tie to C). Such transitive or balanced triads are argued by some theorists to be the "equilibrium" or natural state toward which triadic relationships tend (not all theorists would agree!).

So, there is really quite a lot that can be learned both about individual actors embeddedness, and about the whole network structure just by examining the adjacencies. Let's turn to our data on organizations in the social welfare field, and apply these ideas.

Size, density and degree

The size of a network is often a 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 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 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. As a group gets bigger, the proportion of all of the ties that could (logically) be present -- density -- will fall, and the more likely it is that differentiated and partitioned groups will emerge.

Our example network has ten actors. Usually 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 (that is AB is different from BA, and leaving aside self-ties), 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, or 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.

Fully saturated networks (i.e. one where all logically possible ties are actually present) are empirically rare, particularly where there are more than a few actors in the population. It is useful to look at how close a network is to realizing this potential. That is, to examine the *density of*

ties, which is defined as the proportion of all ties that could be present that actually are. Here is the relevant output from UCINET's univariate statistics routine.

(For the whole Matrix)

Mean	0.54
Std Dev	0.50
Sum	49
Variance	0.25
Minimum	0
Maximum	1
N of Obs	90

Notes: The mean is .54. That is, the mean strength of ties across all possible ties (ignoring self-ties) is .54. Since the data are binary, this means that 54% of all possible ties are present (i.e. the density of the matrix). The standard deviation is a measure of how much variation there is among the elements. If all elements were one, or all were zero, the standard deviation would be zero -- no variation. Here, the average variability from one element to the next is .50, almost as large as the mean. So, we would say that there is, relatively, a great deal of variation in ties. With binary data, the maximum variability in ties -- or the maximum uncertainty about whether any given tie is likely to be present or absent -- is realized at a density of .50. As density approaches either zero or unity, the standard deviation and variance in ties will decline.

Since the data in our example are asymmetric (that is directed ties), we can distinguish between ties being sent and ties being received. Looking at the density for each row and for each column can tell us a good bit about the way in which actors are embedded in the overall density. Here are the "row-wise" univariate statistics for our example from UCINET:

(for the rows, or sending of ties)

	Mean	Std D	Sum	Var
	----	----	----	----
1	0.44	0.50	4.00	0.25
2	0.78	0.42	7.00	0.17
3	0.67	0.47	6.00	0.22
4	0.44	0.50	4.00	0.25
5	0.89	0.31	8.00	0.10
6	0.33	0.47	3.00	0.22
7	0.33	0.47	3.00	0.22
8	0.67	0.47	6.00	0.22
9	0.33	0.47	3.00	0.22
10	0.56	0.50	5.00	0.25

Notes: Statistics on the rows tell us about the role that each actor plays as a "source" of ties (in a directed graph). The sum of the connections from the actor to others (e.g. actor #1 sends information to four others) is called the *out-degree* of the point (for symmetric data, of course, each node simply has *degree*, as we cannot distinguish *in-degree* from *out-degree*). The degree

of points is important because it tells us how many connections an actor has. With out-degree, it is usually a measure of how influential the actor may be. We can see that actor #5 sends ties to all but one of the remaining actors; actors #6, #7 and #9 send information to only three other actors. Actors #2, #3, #5, and #8 are similar in being sources of information for large portions of the network; actors #1, #6, #7, and #9 as being similar as not being sources of information. We might predict that the first set of organizations will have specialized divisions for public relations, the latter set might not. Actors in the first set have a higher potential to be influential; actors in the latter set have lower potential to be influential; actors in "the middle" will be influential if they are connected to the "right" other actors, otherwise, they might have very little influence. So, there is variation in the roles that these organizations play as sources of information. We can norm this information (so we can compare it to other networks of different sizes, by expressing the out-degree of each point as a proportion of the number of elements in the row. That is, calculating the mean. Actor #10, for example, sends ties to 56% of the remaining actors. This is a figure we can compare across networks of different sizes.

Another way of thinking about each actor as a source of information is to look at the row-wise variance or standard deviation. We note that actors with very few out-ties, or very many out-ties have less variability than those with medium levels of ties. This tells us something: those actors with ties to almost everyone else, or with ties to almost no-one else are more "predictable" in their behavior toward any given other actor than those with intermediate numbers of ties. In a sense, actors with many ties (at the center of a network) and actors at the periphery of a network (few ties) have patterns of behavior that are more constrained and predictable. Actors with only some ties can vary more in their behavior, depending on to whom they are connected.

We can also look at the data column-wise. Now, we are looking at the actors as "sinks" or receivers of information. The sum of each column in the adjacency matrix is the *in-degree of the point*. That is, how many other actors send information or ties to the one we are focusing on. Actors that receive information from many sources may be prestigious (other actors want to be known by the actor, so they send information). Actors that receive information from many sources may also be more powerful -- to the extent that "knowledge is power." But, actors that receive a lot of information could also suffer from "information overload" or "noise and interference" due to contradictory messages from different sources.

(for the columns, or ties received)

```
*****
      1   2   3   4   5   6   7   8   9  10
      COUN COMM EDUC INDU MAYR  WRO NEWS UWAY WELF WEST
      ---- ---- ---- ---- ---- ---- ---- ---- ----
Mean  0.56 0.89 0.44 0.56 0.89 0.11 1.00 0.22 0.56 0.22
Std Dev 0.50 0.31 0.50 0.50 0.31 0.31 0.00 0.42 0.50 0.42
Sum    5.00 8.00 4.00 5.00 8.00 1.00 9.00 2.00 5.00 2.00
Variance 0.25 0.10 0.25 0.25 0.10 0.10 0.00 0.17 0.25 0.17
```

Notes: Looking at the means, we see that there is a lot of variation -- more than for information sending. We see that actors #2, #5, and #7 are very high. #2 and #5 are also high in sending

information -- so perhaps they act as "communicators" and "facilitators" in the system. Actor #7 receives a lot of information, but does not send a lot. Actor #7, as it turns out is an "information sink" -- it collects facts, but it does not create them (at least we hope so, since actor #7 is a newspaper). Actors #6, #8, and #10 appear to be "out of the loop" -- that is, they do not receive information from many sources directly. Actor #6 also does not send much information -- so #6 appears to be something of an "isolate." Numbers #8 and #10 send relatively more information than they receive. One might suggest that they are "outsiders" who are attempting to be influential, but may be "clueless."

We can learn a great deal about a network overall, and about the structural constraints on individual actors, and even start forming some hypotheses about social roles and behavioral tendencies, just by looking at the simple adjacencies and calculating a few very basic statistics. Before discussing the slightly more complex idea of distance, there are a couple other aspects of "connectedness" that are sometimes of interest.

Reachability

An actor is "reachable" by another if there exists any 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 are or are not reachable to one another. If some actors in a network cannot reach others, there is the potential of a division of the network. Or, it may indicate that the population we are studying is really composed of more than one separate sub-population.

In the Knoke information exchange data set, it turns out that all actors are reachable by all others. This is something that you can verify by eye. See if you can find any pair of actors in the diagram such that you cannot trace from the first to the second along arrows all headed in the same direction (don't waste a lot of time on this, there is no such pair!).

Reciprocity and Transitivity

It might be useful, in some cases, to classify the dyadic relationships of each actor and for the network as a whole. We've done this in the table below.

Node	No ties	In ties	Out ties	Reciprocated ties	Neighborhood size
1COUN	2	3	0	4 (58%)	7
2COMM	1	1	0	7 (88%)	8
3EDUC	3	0	2	4 (67%)	6
4INDU	3	2	1	3 (50%)	6
5MAYR	1	0	0	8 (100%)	8
6WRO	6	0	2	1 (33%)	3
7NEWS	0	5	0	4 (44%)	9
8UWAY	3	0	4	2 (33%)	6
9WELF	3	2	1	3 (50%)	6
10WEST	4	0	3	2 (40%)	5
Whole Network	39	13	13	38 (75%)	51

The neighborhood size for each actor is the number of other actors to whom they are adjacent. When we array the data about actor's adjacencies as we have here, there is a tendency to characterize actors in terms of the similarity of their tie profiles (much more on this, later). Some actors are predominantly "sources" (that is, they have a tendency to send more than to receive -- actors #3, #6, and #8). Some actors may be predominantly "sinks" (that is, they have a tendency to receive more than send -- actors #1, #5). Other actors may be "transmitters" that both send and receive, but to different others. Perhaps most interestingly, the actors differ quite markedly in their involvement in mutual or reciprocated relationships. Actors #2 and #5 stand out particularly as having predominantly "symmetric" or balanced relationships with other actors.

For the network as a whole, we note that a very considerable proportion of the relationships are reciprocated. Such a structure is one in which a good deal of unmediated pair-wise

communication can occur. This might suggest a rather non-hierarchical structuring of the organizational field, and a field in which there are many local and particular pair-wise relationships, rather than a single monolithic structure or "establishment."

It is also useful to examine the connections in terms of triadic relations. The number of triads that exist with 10 actors is very large; and, with directed data, each of the triads could be classified into one of 64 types. Most commonly, interest has focused on transitive triads. The transitivity principle holds that, if A is tied to B, and B is tied to C, then A should be tied to C. The idea, like "balance" and "reciprocity" is that triadic relationships (where there are ties among the actors) should tend toward transitivity as an equilibrium condition. A special form of this notion is what is called "balance theory." Balance theory deals specifically with relationships of positive and negative affect. It argues that if A likes B, and B likes C, then A should come to like C. Or, if A likes B and B dislikes C, then A should come to dislike C.

For the Knoke information exchange network, UCINET reports that 20.3% of all of the triads that could be transitive (i.e. those triads with three or more ties that connect all actors) are transitive. This is not a particularly high level, and suggests either that the theory of a tendency toward transitivity in triads has not been well realized in these data (perhaps because the system has not been operating very long), or that the theory does not apply in this case (more likely).

Distance

The properties of the network that we have examined so far primarily deal with adjacencies -- 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 main approach is to examine the distance that an actor is from others. 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. How many actors are at various distances from each actor can be important for understanding the differences among actors in the constraints and opportunities they have as a result of their position. Sometimes we are also interested in how many ways there are to connect between two actors, at a given distance. That is, can actor A reach actor B in more than one way? Sometimes multiple connections may indicate a stronger connection between two actors than a single connection.

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 influenced by others -- even if they are technically reachable, the costs may be too high to conduct exchanges. The variability across the actors in the distances that they have from other actors may be a basis for differentiation and even stratification. Those actors who are closer to more others may be able to exert more power than those who are more distant. We will have a good deal more to say

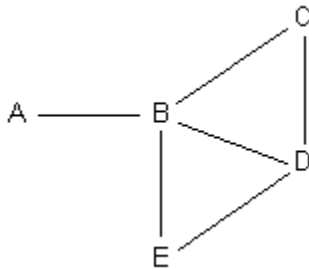
about this aspect of variability in actor distances in the next chapter.

For the moment, we need to learn a bit of jargon that is used to describe the distances between actors: walks, paths, and semi-paths. Using these basic definitions, we can then develop some more powerful ways of describing various aspects of the distances among actors in a network.

Walks etc.

To describe the distances between actors in a network with precision, we need some terminology. And, as it turns out, whether we are talking about a simple graph or a directed graph makes a good bit of difference. If A and B are adjacent in a simple graph, they have a distance of one. In a directed graph, however, A can be adjacent to B while B is not adjacent to A -- the distance from A to B is one, but there is no distance from B to A. Because of this difference, we need slightly different terms to describe distances between actors in graphs and digraphs.

Simple graphs: The most general form of connection between two actors in a graph is called a *walk*. A walk is a sequence of actors and relations that begins and ends with actors. A *closed walk* is one where the beginning and end point of the walk are the same actor. Walks are unrestricted. A walk can involve the same actor or the same relation multiple times. A cycle is a specially restricted walk that is often used in algorithms examining the neighborhoods (the points adjacent) of actors. A cycle is a closed walk of 3 or more actors, all of who are distinct, except for the origin/destination actor. The length of a walk is simply the number of relations contained in it. For example, consider this graph:



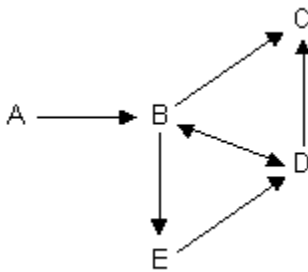
There are many walks in a graph (actually, an infinite number if we are willing to include walks of any length -- though, usually, we restrict our attention to fairly small lengths). To illustrate just a few, begin at actor A and go to actor C. There is one walk of length 2 (A,B,C). There is one walk of length three (A,B,D,C). There are several walks of length 4 (A,B,E,D,C; A,B,D,B,C; A,B,E,B,C). Because these are unrestricted, the same actors and relations can be used more than once in a given walk. There are no cycles beginning and ending with A. There are some beginning and ending with actor B (B,D,C,B; B,E,D,B; B,C,D,E,B).

It is usually more useful to restrict our notion of what constitutes a connection somewhat. One possibility is to restrict the count only walks that do not re-use relations. A *trail* between two actors is any walk that includes a given relation no more than once (the same other actors, however, can be part of a trail multiple times). The length of a trail is the number of relations in it.

All trails are walks, but not all walks are trails. If the trail begins and ends with the same actor, it is called a *closed trail*. In our example above, there are a number of trails from A to C. Excluded are tracings like A,B,D,B,C (which is a walk, but is not a trail because the relation BD is used more than once).

Perhaps the most useful definition of a connection between two actors (or between an actor and themselves) is a *path*. A path is a walk in which each other actor and each other relation in the graph may be used at most one time. The single exception to this is a *closed path*, which begins and ends with the same actor. All paths are trails and walks, but all walks and all trails are not paths. In our example, there are a limited number of paths connecting A and C: A,B,C; A,B,D,C; A,B,E,D,C.

Directed graphs: Walks, trails, and paths can also be defined for directed graphs. But there are two flavors of each, depending on whether we want to take direction into account or not. *Semi-walks, semi-trails, and semi-paths* are the same as for undirected data. In defining these distances, the directionality of connections is simply ignored (that is, arcs - or directed ties are treated as though they were edges - undirected ties). As always, the length of these distances is the number of relations in the walk, trail, or path. If we do want to pay attention to the directionality of the connections we can define walks, trails, and paths in the same way as before, but with the restriction that we may not "change direction" as we move across relations from actor to actor. Consider a directed graph:



In this di-graph, there are a number of walks from A to C. However, there are no walks from C (or anywhere else) to A. Some of these walks from A to C are also trails (e.g. A,B,E,D,B,C). There are, however, only three paths from A to C. One path is length 2 (A,B,C); one is length three (A,B,D,C); one is length four (A,B,E,D,C).

The various kinds of connections (walks, trails, paths) provide us with a number of different ways of thinking about the distances between actors. The main reason that networkers are concerned with these distances is that they provide a way of thinking about the strength of ties or relations. Actors that are connected at short lengths or distances may have stronger connections; actors that are connected many times (for example, having many, rather than a single path) may have stronger ties. Their connection may also be less subject to disruption, and hence more stable and reliable.

Let's look briefly at the distances between pairs of actors in the Knoke data on directed information flows:

of walks of length 1

	1	2	3	4	5	6	7	8	9	0
1	0	1	0	0	1	0	1	0	1	0
2	1	0	1	1	1	0	1	1	1	0
3	0	1	0	1	1	1	1	0	0	1
4	1	1	0	0	1	0	1	0	0	0
5	1	1	1	1	0	0	1	1	1	1
6	0	0	1	0	0	0	1	0	1	0
7	0	1	0	1	1	0	0	0	0	0
8	1	1	0	1	1	0	1	0	1	0
9	0	1	0	0	1	0	1	0	0	0
10	1	1	1	0	1	0	1	0	0	0

of walks of length 2

	1	2	3	4	5	6	7	8	9	0
1	2	3	2	3	3	0	3	2	2	1
2	3	7	1	4	6	1	6	1	3	2
3	4	4	4	3	4	0	5	2	3	1
4	2	3	2	3	3	0	3	2	3	1
5	4	7	2	4	8	1	7	1	3	1
6	0	3	0	2	3	1	2	0	0	1
7	3	2	2	2	2	0	3	2	2	1
8	3	5	2	3	5	0	5	2	3	1
9	2	2	2	3	2	0	2	2	2	1
10	2	4	2	4	4	1	4	2	3	2

of walks of length 3

	1	2	3	4	5	6	7	8	9	10
1	12	18	7	13	18	2	18	6	10	5
2	20	26	16	21	27	1	28	13	18	7
3	14	26	9	19	26	4	25	8	14	8
4	12	19	7	13	19	2	19	6	10	5
5	21	30	17	25	29	2	31	15	21	10
6	9	8	8	8	8	0	10	6	7	3
7	9	17	5	11	17	2	16	4	9	4
8	16	24	11	19	24	2	24	10	15	7
9	10	16	5	10	16	2	16	4	8	4
10	16	23	11	16	23	2	24	8	13	6

Total number of walks (lengths 1, 2, 3)										
	1	2	3	4	5	6	7	8	9	10
1	14	21	9	16	21	2	21	8	12	6
2	23	33	17	25	33	2	34	14	21	9
3	18	30	13	22	30	4	30	10	17	9
4	14	22	9	16	22	2	22	8	13	6
5	25	37	19	29	37	3	38	16	24	11
6	9	11	8	10	11	1	12	6	7	4
7	12	19	7	13	19	2	19	6	11	5
8	19	29	13	22	29	2	29	12	18	8
9	12	18	7	13	18	2	18	6	10	5
10	18	27	13	20	27	3	28	10	16	8

The inventory of the total connections among actors is primarily useful for getting a sense of how "close" each pair is, and for getting a sense of how closely coupled the entire system is. Here, we can see that using only connections of two steps (e.g. "A friend of a friend"), there is a great deal of connection in the graph overall; we also see that there are sharp differences among actors in their degree of connectedness, and who they are connected to. These differences can be used to understand how information moves in the network, which actors are likely to be influential on one another, and a number of other important properties.

Diameter and Geodesic distance

One particular definition of the distance between actors in a network is used by most algorithms to define more complex properties of individual's positions and the structure of the network as a whole. This quantity is the geodesic distance. For both directed and undirected data, the geodesic distance is the number of relations in the shortest possible walk from one actor to another (or, from an actor to themselves, if we care, which we usually do not).

The geodesic distance is widely used in network analysis. There may be many connections between two actors in a network. If we consider how the relation between two actors may provide each with opportunity and constraint, it may well be the case that not all of these ties matter. For example, suppose that I am trying to send a message to Sue. Since I know her e-mail address, I can send it directly (a path of length 1). I also know Donna, and I know that Donna has Sue's email address. I could send my message for Sue to Donna, and ask her to forward it. This would be a path of length two. Confronted with this choice, I am likely to choose the geodesic path (i.e. directly to Sue) because it is less trouble and faster, and because it does not depend on Donna. That is, the geodesic path (or paths, as there can be more than one) is often the "optimal" or most "efficient" connection between two actors. Many algorithms in network analysis assume that actors will use the geodesic path when alternatives are available.

Using UCINET, we can easily locate the lengths of the geodesic paths in our directed data on information exchanges.

Geodesic distances for information exchanges

	1	2	3	4	5	6	7	8	9	0
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

Notes: Because the network is dense, the geodesic path distances are generally small. This suggests that information may travel pretty quickly in this network. Also note that there is a geodesic distance for each xy and yx pair -- that is, the graph is fully connected, and all actors are "reachable" from all others (that is, there exists a path of some length from each actor to each other actor). When a network is not fully connected, we cannot exactly define the geodesic distances among all pairs. The standard approach in such cases is to treat the geodesic distance between unconnected actors as a length greater than that of any real distance in the data. For each actor, we could calculate the mean and standard deviation of their geodesic distances to describe their closeness to all other actors. For each actor, that actor's largest geodesic distance is called the *eccentricity* -- a measure of how far a actor is from the furthest other.

Because the current network is fully connected, a message that starts anywhere will eventually reach everyone. Although the computer has not calculated it, we might want to calculate the mean (or median) geodesic distance and the standard deviation in geodesic distances for the matrix, and for each actor row-wise and column-wise. This would tell us how far each actor is from each other as a source of information for the other; and how far each actor is from each other actor who may be trying to influence them. It also tells us which actors behavior (in this case, whether they've heard something or not) is most predictable and least predictable.

In looking at the whole network, we see that it is connected, and that the average geodesic distance among actors is quite small. This suggests a system in which information is likely to reach everyone, and to do so fairly quickly. To get another notion of the size of a network, we might think about its diameter. The *diameter of a network* is the largest geodesic distance in the (connected) network. In the current case, no actor is more than three steps from any other -- a very "compact" network. The diameter of a network tells us how "big" it is, in one sense (that is, how many steps are necessary to get from one side of it to the other). The diameter is also a

useful quantity in that it can be used to set an upper bound on the lengths of connections that we study. Many researchers limit their explorations of the connections among actors to involve connections that are no longer than the diameter of the network.

Sometimes the redundancy of connection is an important feature of a network structure. If there are many efficient paths connecting two actors, the odds are improved that a signal will get from one to the other. One index of this is a count of the number of geodesic paths between each pair of actors. Of course, if two actors are adjacent, there can only be one such path. I've edited the display below to include only those pairs where there is more than one geodesic path:

Of Geodesic Paths in Information Exchanges

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

Notes: We see that most of the geodesic connections among these actors are not only short distance, but that there are very often multiple shortest paths from x to y. This suggests a couple things: information flow is not likely to break down, because there are multiple paths; and, it will be difficult for any individual to be a powerful "broker" in this structure because most actors have alternative efficient ways of connection to other actors that can by-pass any given actor.

Flow, cohesion, and influence,

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. But, there may be other cases where 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. If I start a rumor, for example, 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 how soon they hear it. For uses of distance like this, we need to take into account all of the connections among actors.

Several approaches have been developed for counting the amount of connection between pairs of actors that take into account all connections between them. These measures have been used for a number of different purposes, and these differences are reflected in the algorithms used to calculate them. We will examine three such ideas.

Maximum flow: One notion of how totally connected two actors are (called maximum flow by UCINET) asks how many different actors in the neighborhood of a source lead to 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. This "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. This approach to connection between actors is closely connected to the notion of between-ness that we will examine in the next chapter. It is also logically close to the idea that the number of pathways, not their length may be important in connecting people. For our directed information flow data, the results of UCINET's count of maximum flow are shown below. You should verify for yourself that, for example, there are four "choke points" in flows from actor 1 to actor 2, but five such points in the flow from actor 2 to actor 1.

MAXIMUM FLOW

		1	2	3	4	5	6	7	8	9	0
	COCOEDINMAWRNEUWWEWE	-	-	-	-	-	-	-	-	-	-
1	COUN	0	4	3	4	4	1	4	2	4	2
2	COMM	5	0	3	5	7	1	7	2	5	2
3	EDUC	5	6	0	5	6	1	6	2	5	2
4	INDU	4	4	3	0	4	1	4	2	4	2
5	MAYR	5	8	3	5	0	1	8	2	5	2
6	WRO	3	3	3	3	3	0	3	2	3	2
7	NEWS	3	3	3	3	3	1	0	2	3	2
8	UWAY	5	6	3	5	6	1	6	0	5	2
9	WELF	3	3	3	3	3	1	3	2	0	2
10	WEST	5	5	3	5	5	1	5	2	5	0

SINGLE-LINK HIERARCHICAL CLUSTERING

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

Level	6	8	1	3	4	5	7	2	9	0
8	XXXXX
5	.	.	XXXXXXXXXXXXXXXXX
2	.	XXXXXXXXXXXXXXXXXXXXX
1	XXXXXXXXXXXXXXXXXXXXX

notes: As with almost any other measure of the relations between pairs of actors, it may sometimes be useful to see if we can classify actors as being more or less similar to one another

in terms of the pattern of their relation with all the other actors. Cluster analysis is one useful technique for finding groups of actors who have similar relationships with other actors. We will have much more to say on this topic in later chapters.

In the current example, we see that actors #2, 5, and 7 are relatively similar to one another in terms of their maximum flow distances from other actors. A quick check of the graph reveals why this is. These three actors have many different ways of reaching one another, and all other actors in the network. In this regard, they are very different from actor 6, who can depend on (at most) only three other actors to make a connection.

Hubbell and Katz cohesion: The maximum flow approach focuses on the vulnerability or redundancy of connection between pairs of actors - kind of a "strength of the weakest link" argument. As an alternative approach, we might want to consider the strength of all links as defining the connection. If we are interested in how much two actors may influence one another, or share a sense of common position, the full range of their connections should probably be considered.

Even if we want to include all connections between two actors, it may not make a great deal of sense (in most cases) to consider a path of length 10 as important as a path of length 1. The Hubbell and Katz approaches count the total connections between actors (ties for undirected data, both sending and receiving ties for directed data). Each connection, however, is given a weight, according to its length. The greater the length, the weaker the connection. How much weaker the connection becomes with increasing length depends on an "attenuation" factor. In our example, below, we have used an attenuation factor of .5. That is, an adjacency receives a weight of one, a walk of length two receives a weight of .5, a connection of length three receives a weight of .5 squared (.25) etc. The Hubbell and Katz approaches are very similar. Katz includes an identity matrix (a connection of each actor with themselves) as the strongest connection; the Hubbell approach does not. As calculated by UCINET, both approaches "norm" the results to range from large negative distances (that is, the actors are very close relative to the other pairs, or have high cohesion) to large positive numbers (that is, the actors have large distance relative to others)

Method: HUBBELL

	1	2	3	4	5	6	7	8	9	10
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
1 COUN	-0.67	-0.67	2.00	-0.33	-0.67	2.00	0.33	-1.33	-1.33	1.33
2 COMM	-0.92	-0.17	1.50	-0.08	-0.67	1.50	0.08	-0.83	-1.08	0.83
3 EDUC	5.83	3.33	-11.00	0.17	3.33	-11.00	-2.17	6.67	8.17	-7.67
4 INDU	-1.50	-1.00	3.00	0.50	-1.00	3.00	0.50	-2.00	-2.50	2.00
5 MAYR	1.25	0.50	-2.50	-0.25	1.00	-2.50	-0.75	1.50	1.75	-1.50
6 WRO	3.83	2.33	-8.00	0.17	2.33	-7.00	-1.17	4.67	6.17	-5.67
7 NEWS	-1.17	-0.67	2.00	0.17	-0.67	2.00	0.83	-1.33	-1.83	1.33
8 UWAY	-3.83	-2.33	7.00	-0.17	-2.33	7.00	1.17	-3.67	-5.17	4.67
9 WELF	-0.83	-0.33	1.00	-0.17	-0.33	1.00	0.17	-0.67	-0.17	0.67
10 WEST	4.33	2.33	-8.00	-0.33	2.33	-8.00	-1.67	4.67	5.67	-4.67

Method:		KATZ									
		1	2	3	4	5	6	7	8	9	10
		COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
		-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1	COUN	-1.67	-0.67	2.00	-0.33	-0.67	2.00	0.33	-1.33	-1.33	1.33
2	COMM	-0.92	-1.17	1.50	-0.08	-0.67	1.50	0.08	-0.83	-1.08	0.83
3	EDUC	5.83	3.33	-12.00	0.17	3.33	-11.00	-2.17	6.67	8.17	-7.67
4	INDU	-1.50	-1.00	3.00	-0.50	-1.00	3.00	0.50	-2.00	-2.50	2.00
5	MAYR	1.25	0.50	-2.50	-0.25	0.00	-2.50	-0.75	1.50	1.75	-1.50
6	WRO	3.83	2.33	-8.00	0.17	2.33	-8.00	-1.17	4.67	6.17	-5.67
7	NEWS	-1.17	-0.67	2.00	0.17	-0.67	2.00	-0.17	-1.33	-1.83	1.33
8	UWAY	-3.83	-2.33	7.00	-0.17	-2.33	7.00	1.17	-4.67	-5.17	4.67
9	WELF	-0.83	-0.33	1.00	-0.17	-0.33	1.00	0.17	-0.67	-1.17	0.67
10	WEST	4.33	2.33	-8.00	-0.33	2.33	-8.00	-1.67	4.67	5.67	-5.67

As with all measures of pairwise properties, one could analyze the data much further. We could see which individuals are similar to which others (that is, are there groups or strata defined by the similarity of their total connections to all others in the network?). Our interest might also focus on the whole network, where we might examine the degree of variance, and the shape of the distribution of the dyad connections. For example, a network in with the total connections among all pairs of actors might be expected to behave very differently than one where there are radical differences among actors.

Taylor's Influence: The Hubbell and Katz approach may make most sense when applied to symmetric data, because they pay no attention to the directions of connections (i.e. A's ties directed to B are just as important as B's ties to A in defining the distance or solidarity -- closeness-- between them). If we are more specifically interested in the influence of A on B in a directed graph, the Taylor influence approach provides an interesting alternative.

The Taylor measure, like the others, uses all connections, and applies an attenuation factor. Rather than standardizing on the whole resulting matrix, however, a different approach is adopted. The column marginals for each actor are subtracted from the row marginals, and the result is then normed (what did he say?!). Translated into English, we look at the balance between each actor's sending connections (row marginals) and their receiving connections (column marginals). Positive values then reflect a preponderance of sending over receiving to the other actor of the pair -- or a balance of influence between the two. Note that the newspaper (#7) shows as being a net influencer with respect to most other actors in the result below, while the welfare rights organization (#6) has a negative balance of influence with most other actors.

Method: TAYLOR

	1	2	3	4	5	6	7	8	9	10
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
1 COUN	0.00	-0.02	0.23	-0.07	0.12	0.11	-0.09	-0.15	0.03	0.18
2 COMM	0.02	0.00	0.11	-0.06	0.07	0.05	-0.05	-0.09	0.05	0.09
3 EDUC	-0.23	-0.11	0.00	0.17	-0.36	0.18	0.26	0.02	-0.44	-0.02
4 INDU	0.07	0.06	-0.17	0.00	0.05	-0.17	-0.02	0.11	0.14	-0.14
5 MAYR	-0.12	-0.07	0.36	-0.05	0.00	0.30	0.01	-0.23	-0.13	0.23
6 WRO	-0.11	-0.05	-0.18	0.17	-0.30	0.00	0.19	0.14	-0.32	-0.14
7 NEWS	0.09	0.05	-0.26	0.02	-0.01	-0.19	0.00	0.15	0.12	-0.18
8 UWAY	0.15	0.09	-0.02	-0.11	0.23	-0.14	-0.15	0.00	0.28	0.00
9 WELF	-0.03	-0.05	0.44	-0.14	0.13	0.32	-0.12	-0.28	0.00	0.31
10 WEST	-0.18	-0.09	0.02	0.14	-0.23	0.14	0.18	-0.00	-0.31	0.00

As with most other measures, the various approaches to the distance between actors and in the network as a whole provide a menu of choices. No one definition to measuring distance will be the "right" choice for a given purpose. Sometimes we don't really know, before hand, what approach might be best, and we may have to try and test several.

Summary of chapter 5

There is a great deal of information about both individuals and the population in a single adjacency matrix. In this chapter you have learned a lot of terminology for describing the connections and distances between actors, and for whole populations.

One focus in basic network analysis is on the immediate neighborhood of each actor: the dyads and triads in which they are involved. The degree of an actor, and the in-degree and out-degree (if the data are directed) tell us about the extent to which an actor may be constrained by, or constrain others. The extent to which an actor can reach others in the network may be useful in describing an actor's opportunity structure. We have also seen that it is possible to describe "types" of actors who may form groups or strata on the basis of their places in opportunity structures -- e.g. "isolates" "sources" etc.

Most of the time and effort of most social actors is spent in very local contexts -- interacting in dyads and triads. In looking at the connections of actors, we have suggested that the degree of "reciprocity" and "balance" and "transitivity" in relations can be regarded as important indicators of the stability and institutionalization (that is, the extent to which relations are taken for granted and are norm governed) of actor's positions in social networks.

The local connections of actors are important for understanding the social behavior of the whole population, as well as for understanding each individual. The size of the network, it's density, whether all actors are reachable by all others (i.e. is the whole population connected, or are there multiple components?), whether ties tend to be reciprocal or transitive, and all the other properties that we examined for individual connections are meaningful in describing the whole

population. Both the typical levels of characteristics (e.g. the mean degree of points), and the amount of diversity in characteristics (e.g. the variance in the degree of points) may be important in explaining macro behavior. Populations with high density respond differently to challenges from the environment than those with low density; populations with greater diversity in individual densities may be more likely to develop stable social differentiation and stratification.

In this chapter we also examined some properties of individual's embeddedness and of whole networks that look at the broader, rather than the local neighborhoods of actors. A set of specialized terminology was introduced to describe the distances between pairs of actors: walks, trails, and paths. We noted that there are some important differences between un-directed and directed data in applying these ideas of distance.

One of the most common and important approaches to indexing the distances between actors is the geodesic. The geodesic is useful for describing the minimum distance between actors. The geodesic distances between pairs of actors are the most commonly used measure of closeness. The average geodesic distance for an actor to all others, the variation in these distances, and the number of geodesic distances to other actors may all describe important similarities and differences between actors in how and how closely they are connected to their entire population.

The geodesic distance, however, examines only a single connection between a pair of actors (or, in some cases several, if there are multiple geodesics connecting them). Sometimes the sum of all connections between actors, rather than the shortest connection may be relevant. We have examined approaches to measuring the vulnerability of the connection between actors (flow), the pairwise solidarity of actors (Hubbell and Katz), and the potential influence of actors on others (Taylor). Again, it important to remember that differences and similarities among individual persons in these aspects of their distances from others may form a structural basis for differentiation and stratification. And, as always, these more complex patterns of distances of individuals from all others are also a characteristic of the whole population -- as well as describing the constraints and opportunities facing individuals. For example, a population in which most actors have multiple geodesic distances to others will, most likely, behave quite differently from one in which there is little redundancy of efficient communication linkages (think about some different kinds of bureaucratic organizations -- say a medical clinic versus an automobile factory).

We have seen that there is a great deal of information available in fairly simple examinations of an adjacency matrix. Life, of course, can get more complicated. We could have multiple layers, or multiplex data; we could have data that gave information on the strength of ties, rather than simple presence or absence. Nonetheless, the methods that we've used here will usually give you a pretty good grasp of what is going on in more complicated data.

Now that you have a pretty good grasp of the basics of connection and distance, you are ready to use these ideas to build some concepts and methods for describing somewhat more complicated aspects of the network structures of populations. We will turn to two of the most important of these in the next chapters. We will first (chapter 6) examine the concept of centrality and centralization, which are important to understanding power, stratification, ranking, and inequality in social structures. Then, in chapter 7, we will turn our attention to examining sub structures

(larger than dyads and triads). Populations of social structures are almost always divided and differentiated into cliques, factions, and groups (which may, or may not be ranked). Building on these ideas, we will conclude our introductory survey of network concepts (chapters 8-11) with attention to the more abstract, but theoretically important task of describing network positions and social roles that are central concepts in sociological analysis.

Review questions for chapter 5

1. Explain the differences among the "three levels of analysis" of graphs (individual, aggregate, whole).
2. How is the size of a network measured? Why is population size so important in sociological analysis?
3. You have a network of 5 actors, assuming no self-ties, what is the potential number of directed ties? What is the potential number of un-directed ties?
4. How is density measured? Why is density important in sociological analysis?
5. What is the "degree of a point?" Why might it be important, sociologically, if some actors have high degree and other actors have lower degree? What is the difference between "in-degree" and "out-degree?"
6. If actor "A" is reachable from actor "B" does that necessarily mean that actor "B" is reachable from actor "A?" Why or why not?
7. For pairs of actors with directed relations, there are four possible configurations of ties. Can you show these? Which configurations are "balanced?" For a triad with undirected relations, how many possible configurations of ties are there? Which ones are balanced or transitive?
8. What are the differences among walks, trails, and paths? Why are "paths" the most commonly used approach to inter-actor distances in sociological analysis?
9. What is the "geodesic" distance between two actors? Many social network measures assume that the geodesic path is the most important path between actors -- why is this a plausible assumption?
10. I have two populations of ten actors each, one has a network diameter of 3, and the other has a network diameter of 6. Can you explain this statement to someone who doesn't know social network analysis? Can you explain why this difference in diameter might be important in understanding differences between the two populations?
11. How do "weighted flow" approaches to social distance differ from "geodesic" approaches to social distance?
12. Why might it matter if two actors have more than one geodesic or other path between them?

Application questions for chapter 5

1. Think of the readings from the first part of the course. Which studies used the ideas of connectedness and density? Which studies used the ideas of distance? What specific approaches did they use to measure these concepts?

2. Draw the graphs of a "star" a "circle" a "line" and a "hierarchy." Describe the size, potential, and density of each graph. Examine the degrees of points in each graph -- are there differences among actors? Do these differences tell us something about the "social roles" of the actors? Create a matrix for each graph that shows the geodesic distances between each pair of actors. Are there differences between the graphs in whether actors are connected by multiple geodesic distances?

3. Think about a small group of people that you know well (maybe your family, neighbors, a study group, etc.). Who helps whom in this group? What is the density of the ties? Are ties reciprocated? Are triads transitive?

4. Chrysler Corporation has called on you to be a consultant. Their research division is taking too long to generate new models of cars, and often the work of the "stylists" doesn't fit well with the work of the "manufacturing engineers" (the people who figure out how to actually build the car). Chrysler's research division is organized as a classical hierarchical bureaucracy with two branches (stylists, manufacturing) coordinated through group managers and a division manager. Analyze the reasons why performance is poor. Suggest some alternative ways of organizing that might improve performance, and explain why they will help.

6. Centrality and Power

Introduction: Centrality and Power

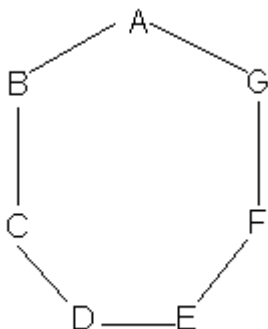
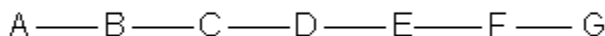
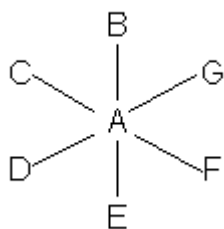
All sociologists would agree that power is a fundamental property of social structures. There is much less agreement about what power is, and how we can describe and analyze its causes and consequences. In this chapter we will look at some of the main approaches that social network analysis has developed to study power, and the closely related concept of centrality.

Network thinking has contributed a number of important insights about social power. Perhaps most importantly, the network approach emphasizes that power is inherently relational. An individual does not have power in the abstract, they have power because they can dominate others -- ego's power is alter's dependence, and vice versa. Because power is a consequence of patterns of relations, the amount of power in social structures can vary. If a system is very loosely coupled (low density) not much power can be exerted; in high-density systems there is the potential for greater power. Power is both a systemic (macro) and relational (micro) property. The amount of power in a system and its distribution across actors are related, but are not the same thing. Two systems can have the same amount of power, but it can be equally distributed in one and unequally distributed in another. Power in social networks may be viewed either as a micro property (i.e. it describes relations between actors) or as a macro property (i.e. one that describes the entire population); as with other key sociological concepts, the macro and micro are closely connected in social network thinking.

Network analysts often describe the way that an actor is embedded in a relational network as imposing constraints on the actor and offering the actor opportunities. 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, and that the actor will be a focus for deference and attention from those in less favored positions. But what do we mean by "having a favored position" and having "more opportunities" and "fewer constraints?" There are no single correct and final answers to these difficult questions. But, network analysis has made important contributions in providing precise definitions and concrete measures of several different approaches to the notion of the power that attaches to positions in structures of social relations.

The several faces of power

To understand the approaches that network analysis uses to study power, it is useful to first think about some very simple systems. Consider these three networks, which are called the "star," "line," and "circle."



A moment's thought ought to suggest that actor A has a highly favored structural position in the star network, if the network is describing a relationship such as resource exchange or resource sharing. But, exactly why is it that actor A has a "better" position than all of the others in the star network? What about the position of A in the line network? Is being at the end of the line an advantage or a disadvantage? Are all of the actors in the circle network really in exactly the same structural position? We need to think about why structural location can be advantageous or disadvantageous to actors. Let's focus our attention on why actor A is so obviously at an advantage in the star network.

Degree: First, actor A has more opportunities and alternatives than other actors. If actor D elects to not provide A with a resource, A has a number of other places to go to get it; however, if D elects to not exchange with A, then D will not be able to exchange at all. The more ties an actor has then, the more power they (may) have. In the star network, Actor A has degree six, all other actors have degree one. This logic underlies measures of centrality and power based on *actor degree*, which we will discuss below. Actors who have more ties have greater opportunities because they have choices. This autonomy makes them less dependent on any specific other actor, and hence more powerful.

Now, consider the circle network in terms of degree. Each actor has exactly the same number of alternative trading partners (or degree), so all positions are equally advantaged or disadvantaged. In the line network, matters are a bit more complex. The actors at the end of the line (A and G) are actually at a structural disadvantage, but all others are apparently equal (actually, it's not really quite that simple). Generally, though, actors that are more central to the structure, in the sense of having higher degree or more connections, tend to have favored positions, and hence more power.

Closeness: The second reason why actor A is more powerful than the other actors in the star network is that actor A is *closer* to more actors than any other actor. Power can be exerted by direct bargaining and exchange. But power also comes from acting as a "reference point" by which other actors judge themselves, and by being a center of attention whose views are heard by larger numbers of actors. Actors who are able to reach other actors at shorter path lengths, or who are more reachable by other actors at shorter path lengths have favored positions. This structural advantage can be translated into power. In the star network, actor A is at a geodesic distance of one from all other actors; each other actor is at a geodesic distance of two from all other actors (but A). This logic of structural advantage underlies approaches that emphasize the distribution of closeness and distance as a source of power.

Now consider the circle network in terms of actor closeness. Each actor lies at different path lengths from the other actors, but all actors have identical distributions of closeness, and again would appear to be equal in terms of their structural positions. In the line network, the middle actor (D) is closer to all other actors than are the set C,E, the set B,F, and the set A,G. Again, the actors at the ends of the line, or at the periphery, are at a disadvantage.

Betweenness: The third reason that actor A is advantaged in the star network is because actor A lies *between* each other pair of actors, and no other actors lie between A and other actors. If A wants to contact F, A may simply do so. If F wants to contact B, they must do so by way of A. This gives actor A the capacity to broker contacts among other actors -- to extract "service charges" and to isolate actors or prevent contacts. The third aspect of a structurally advantaged position then is in being between other actors.

In the circle network, each actor lies between each other pair of actors. Actually, there are two pathways connecting each pair of actors, and each third actor lies on one, but not on the other of them. Again, all actors are equally advantaged or disadvantaged. In the line network, our end points (A,G) do not lie between any pairs, and have no brokering power. Actors closer to the middle of the chain lie on more pathways among pairs, and are again in an advantaged position.

Each of these three ideas -- degree, closeness, and betweenness -- has been elaborated in a number of ways. We will examine three such elaborations briefly here. The eigenvector of geodesics approach builds on the notion of closeness/distance. The flow approach (which is very similar to the flow approach discussed in the last chapter) modifies the idea of betweenness. The Bonacich power measure is an important and widely used generalization of degree-based approaches to power. In the previous chapter we discussed several approaches to distance that spoke to the influence of points on one another (Hubbell, Katz, Taylor). These approaches, which we will not discuss again here, can also be seen as generalizations of the closeness logic that measure the power of actors.

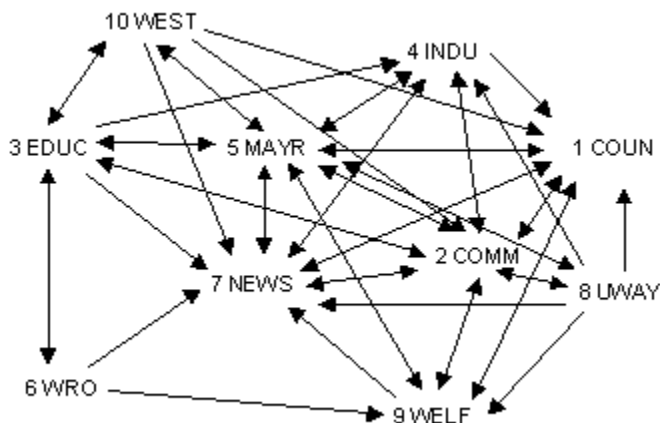
Network analysts are more likely to describe their approaches as descriptions of centrality than of power. Each of the three approaches (degree, closeness, betweenness) describe the locations of individuals in terms of how close they are to the "center" of the action in a network -- though the definitions of what it means to be at the center differ. It is more correct to describe network approaches this way -- measures of centrality -- than as measures of power. But, as we have suggested here, there are several reasons why central positions tend to be powerful positions.

Degree centrality

Actors who have more ties to other actors may be advantaged positions. Because they have many ties, they may have alternative ways to satisfy needs, and hence are less dependent on other individuals. Because they have many ties, they may have access to, and be able to call on more of the resources of the network as a whole. Because they have many ties, they are often third-parties and deal makers in exchanges among others, and are able to benefit from this brokerage. So, a very simple, but often very effective measure of an actor's centrality and power potential is their degree.

In undirected data, actors 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 to direct ties to them, and this may indicate their importance. Actors who have unusually high out-degree are actors who are able to exchange with many others, or make many others aware of their views. Actors who display high out-degree centrality are often said to be *influential* actors.

Recall the data on information exchanges among organizations operating in the social welfare field (Knoke) which we have been examining.



Let's examine the in-degrees and out-degrees of the points as a measure of who is "central" or "influential" in this network. UCINET has been used to do the counting, and some additional calculations and standardizations that were suggested by Linton Freeman.

FREEMAN'S DEGREE CENTRALITY MEASURES

	1	2	3	4
	OutDegree	InDegree	NrmOutDeg	NrmInDeg
1 COUN	4.00	5.00	44.44	55.56
2 COMM	7.00	8.00	77.78	88.89
3 EDUC	6.00	4.00	66.67	44.44
4 INDU	4.00	5.00	44.44	55.56
5 MAYR	8.00	8.00	88.89	88.89
6 WRO	3.00	1.00	33.33	11.11
7 NEWS	3.00	9.00	33.33	100.00
8 UWAY	6.00	2.00	66.67	22.22
9 WELF	3.00	5.00	33.33	55.56
10 WEST	5.00	2.00	55.56	22.22

DESCRIPTIVE STATISTICS

	1	2	3	4
	OutDegree	InDegree	NrmOutDeg	NrmInDeg
Mean	4.90	4.90	54.44	54.44
Std Dev	1.70	2.62	18.89	29.17
Variance	2.89	6.89	356.79	850.62
Minimum	3.00	1.00	33.33	11.11
Maximum	8.00	9.00	88.89	100.00

Network Centralization (Outdegree) = 43.056%

Network Centralization (Indegree) = 56.944%

notes: 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. That other organizations share information with these three would seem to indicate a desire on the part of others to exert influence. This is an act of deference, or recognition that the positions of actors 5, 2, and 7 might be worth trying to influence. If we were interested in comparing across networks of different sizes or densities, it might be useful to "standardize" the measures of in and out-degree. In the last two columns of the first panel of results above, all the degree counts have been expressed as percentages of the largest degree count in the data set (i.e. #7's out-degree of 9).

The next panel of results speaks to the "meso" level of analysis. That is, what does the distribution of the actor's degree centrality scores look like? On the average, actors have a degree of 4.9, which is quite high, given that there are only nine other actors. We see that the range of in-degree is slightly larger (minimum and maximum) than that of out-degree, and that there is more variability across the actors in in-degree than out-degree (standard deviations and variances). The range and variability of degree (and other network properties) can be quite

important, because it describes whether the population is homogeneous or heterogeneous in structural positions. One could examine whether the variability is high or low relative to the typical scores by calculating the coefficient of variation (standard deviation divided by mean, times 100) for in-degree and out-degree. By the rules of thumb that are often used to evaluate coefficients of variation, the current values (35 for out-degree and 53 for in-degree) are moderate. Clearly, however, the population is more homogeneous with regard to out-degree (influence) than with regard to in-degree (prominence).

The last bit of information provided by the output above are Freeman's *graph centralization measures*, which describe the population as a whole -- the macro level. These are very useful statistics, but require a bit of explanation.

Remember our "star" network from the discussion above (if not, go review it)? The star network is the most centralized or most unequal possible network for any number of actors. In the star network, all the actors but one have degree of one, and the "star" has degree of the number of actors, less one. Freeman felt that it would be useful to express the degree of variability in the degrees of actors in our observed network as a percentage of that in a star network of the same size. This is how the Freeman graph centralization measures can be understood: they express 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 43% and the in-degree graph centralization 57% 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.

Now that we understand the basic ideas, we can be a bit briefer in looking at the other approaches to centralization and power.

Closeness centrality

Degree centrality measures might be criticized because they only take into account the immediate ties that an actor has, rather than indirect ties to all others. One actor might be tied to a large number of others, but those others might be rather disconnected from the network as a whole. In a case like this, the actor could be quite central, but only in a local neighborhood.

Closeness centrality approaches emphasize the distance of an actor to all others in the network by focusing on the geodesic distance from each actor to all others. One could consider either directed or undirected geodesic distances among actors. In our current example, we have decided to look at undirected ties. The sum of these geodesic distances for each actor is the "farness" of the actor from all others. We can convert this into a measure of nearness or closeness centrality by taking the reciprocal (that is one divided by the farness) and norming it relative to the most central actor. Here are the UCINET results for our information exchange data.

	Farness	Closeness
1	11.00	81.82
2	10.00	90.00
3	12.00	75.00
4	12.00	75.00
5	10.00	90.00
6	15.00	60.00
7	9.00	100.00
8	12.00	75.00
9	12.00	75.00
10	13.00	69.23

	Farness	Closeness
Mean	11.60	79.10
Std Dev	1.62	11.01
Sum	116.00	791.05
Variance	2.64	121.13
Minimum	9.00	60.00
Maximum	15.00	100.00

Network Centralization = 49.34%

We can see that actor #7 is the closest or most central actor using this method, because the sum of #7's geodesic distances to other actors (a total of 9, across 9 other actors) is the least (in fact, it is the minimum possible sum of geodesic distances). Two other actors (#2 and #5) are nearly as close, the Welfare Rights Organization (#6) has the greatest farness. In a small network with high density, it is not surprising that centrality based on distance is very similar to centrality based on adjacency (since many geodesic distances in this network are adjacencies!). In bigger and/or less dense networks, the two approaches can yield rather different pictures of who are the most central actors.

Distance based centrality can also be used to characterize the centralization of the entire graph. That is, how unequal is the distribution of centrality across the population? Again, a useful way of indexing this property of the whole graph is to compare the variance in the actual data to the variance in a star network of the same size. Again, the star network is one in which the distribution of farness of actors displays the maximum possible concentration (with one actor being maximally close to all others and all others being maximally distant from one another). The graph centralization index based on closeness shows a somewhat more modest, but still substantial degree of concentration in the whole network.

Betweenness centrality

Suppose that I want to influence you by sending you information, or make a deal to exchange some resources. But, in order to talk to you, I must go through an intermediary. For example, let's suppose that I wanted to try to convince the Chancellor of my university to buy me a new

computer. According to the rules of our bureaucratic hierarchy, I must forward my request through my department chair, a dean, and an executive vice chancellor. Each one of these people could delay the request, or even prevent my request from getting through. This gives the people who lie "between" me and the Chancellor power with respect to me. To stretch the example just a bit more, suppose that I also have an appointment in the school of business, as well as one in the department of sociology. I might forward my request to the Chancellor by both channels. Having more than one channel makes me less dependent, and, in a sense, more powerful.

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. Using the computer, it is quite easy to locate the geodesic paths between all pairs of actors, and to count up how frequently each actor falls in each of these pathways. If we add up, for each actor, the proportion of times that they are "between" other actors for the sending of information in the Knoke data, we get the a measure of actor centrality. We can norm this measure by expressing it as a percentage of the maximum possible betweenness that an actor could have had. The results from UCINET are:

	Between	Normed Between
	-----	-----
1	0.67	0.93
2	12.33	17.13
3	11.69	16.24
4	0.81	1.12
5	17.83	24.77
6	0.33	0.46
7	2.75	3.82
8	0.00	0.00
9	1.22	1.70
10	0.36	0.50

	Between	nBetween
	-----	-----
Mean	4.80	6.67
Std Dev	6.22	8.64
Sum	48.00	66.67
Variance	38.69	74.63
Minimum	0.00	0.00
Maximum	17.83	24.77

Network Centralization Index = 20.11%

Notes: We can see that there is a lot of variation in actor between-ness (from zero to 17.83), and that there is quite a bit of variation (std. dev. = 6.2 relative to a mean between-ness of 4.8). Despite this, the overall network centralization is relatively low. This makes sense, because we

know that fully one half of all connections can be made in this network without the aid of any intermediary -- hence there cannot be a lot of "between-ness." In the sense of structural constraint, there is not a lot of "power" in this network. Actors #2, #3, and #5 appear to be relatively a good bit more powerful than others are by this measure. Clearly, there is a structural basis for these actors to perceive that they are "different" from others in the population. Indeed, it would not be surprising if these three actors saw themselves as the movers-and-shakers, and the dealmakers that made things happen. In this sense, even though there is not very much betweenness power in the system, it could be important for group formation and stratification.

Each of the three basic approaches to centrality and power (degree, closeness, and betweenness) can, and has been elaborated to capture additional information from graphs. Reviewing a few of these elaborations may suggest some other ways in which power and position in networks are related.

Eigenvector of the geodesic distances

The closeness centrality measure described above is based on the sum of the geodesic distances from each actor to all others (farness). In larger and more complex networks than the example we've been considering, it is possible to be somewhat misled by this measure. Consider two actors, A and B. Actor A is quite close to a small and fairly closed group within a larger network, and rather distant from many of the members of the population. Actor B is at a moderate distance from all of the members of the population. The farness measures for actor A and actor B could be quite similar in magnitude. In a sense, however, actor B is really more "central" than actor A in this example, because B is able to reach more of the network with same amount of effort.

The eigenvector approach is an effort to find the most central actors (i.e. those with the smallest farness from others) in terms of the "global" or "overall" structure of the network, and to pay less attention to patterns that are more "local." The method used to do this (factor analysis) is beyond the scope of the current text. In a general way, what factor analysis does is to identify "dimensions" of the distances among actors. The location of each actor with respect to each dimension is called an "eigenvalue," and the collection of such values is called the "eigenvector." Usually, the first dimension captures the "global" aspects of distances among actors; second and further dimensions capture more specific and local sub-structures.

For illustration, we have used UCINET to calculate the eigenvectors of the distance matrix for the Knoke information exchange data. In this example, we have measured the distance between two actors as the longer of the directed geodesic distances between them. If we were interested in information "exchange" relationships, rather than simply sources or receivers of information, we might want to take this approach -- because it says that, in order to be close, a pair of actors must have short distances between themselves in both directions. Here's the output:

BONACICH CENTRALITY

WARNING: This version of the program cannot handle asymmetric data.
 Matrix symmetrized by taking larger of Xij and Xji.

EIGENVALUES

FACTOR	VALUE	PERCENT	CUM %	RATIO
1:	6.766	74.3	74.3	5.595
2:	1.209	13.3	87.6	1.282
3:	0.944	10.4	97.9	5.037
4:	0.187	2.1	100.0	
=====	9.106	100.0		=====

Bonacich Eigenvector Centralities

	1	2
	Eigenvec	nEigenvec
-----	-----	-----
1 COUN	0.343	48.516
2 COMM	0.379	53.538
3 EDUC	0.276	38.999
4 INDU	0.308	43.522
5 MAYR	0.379	53.538
6 WRO	0.142	20.079
7 NEWS	0.397	56.124
8 UWAY	0.309	43.744
9 WELF	0.288	40.726
10 WEST	0.262	37.057

Descriptive Statistics

	1	2
	Eigenven	Eigenv
-----	-----	-----
1 Mean	0.31	43.58
2 Std Dev	0.07	10.02
3 Sum	3.08	435.84
4 Variance	0.01	100.41
5 Euc Norm	1.00	141.42
6 Minimum	0.14	20.08
7 Maximum	0.40	56.12
8 N of Obs	10.00	10.00

Network centralization index = 20.90%

The first set of statistics, the eigenvalues, tell us how much of the overall pattern of distances among actors can be seen as reflecting the global pattern (the first eigenvalue), and more local, or additional patterns (the additional patterns). We are interested in the percentage of the overall

variation in distances that is accounted for by the first factor. Here, this percentage is 74.3%. This means that about 3/4 of all of the distances among actors are reflective of the main dimension or pattern. If this amount is not large (say over 70%), great caution should be exercised in interpreting the further results, because the dominant pattern is not doing a very complete job of describing the data. The first eigenvalue should also be considerably larger than the second (here, the ratio of the first eigenvalue to the second is about 5.6 to 1). This means that the dominant pattern is, in a sense, 5.6 times as "important" as the secondary pattern.

Next, we turn our attention to the scores of each of the cases on the 1st eigenvector. Higher scores indicate that actors are "more central" to the main pattern of distances among all of the actors, lower values indicate that actors are more peripheral. The results are very similar to those for our earlier analysis of closeness centrality, with actors #7, #5, and #2 being most central, and actor #6 being most peripheral. Usually the eigenvalue approach will do what it is supposed to do: give us a "cleaned-up" version of the closeness centrality measures, as it does here. It is a good idea to examine both, and to compare them.

Last, we examine the overall centralization of the graph, and the distribution of centralities. There is relatively little variability in centralities (standard deviation .07) around the mean (.31). This suggests that, overall, there are not great inequalities in actor centrality or power, when measured in this way. Compared to the pure "star" network, the degree of inequality or concentration of the Knoke data is only 20.9% of the maximum possible. This is much less than the network centralization measure for the "raw" closeness measure (49.3), and suggests that some of the apparent differences in power using the raw closeness approach may be due more to local than to global inequalities.

The factor analysis approach could be used to examine degree or betweenness as well, but we will not burden you with these exercises. Our main point is this: geodesic distances among actors are a reasonable measure of one aspect of centrality -- or positional advantage. Sometimes these advantages may be more local, and sometimes more global. The factor-analytic approach is one approach that may sometimes help us to focus on the more global pattern. Again, it is not that one approach is "right" and the other "wrong." Depending on the goals of our analysis, we may wish to emphasize one or the other aspects of the positional advantages that arise from centrality.

Flow centrality

The betweenness centrality measure we examined above characterizes actors as having positional advantage, or power, to the extent that they fall on the shortest (geodesic) pathway between other pairs of actors. The idea is that actors who are "between" other actors, and on whom other actors must depend to conduct exchanges, will be able to translate this broker role into power.

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 approach to centrality expands the notion of betweenness centrality. It assumes that actors will use all pathways that connect them,

proportionally to the length of the 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 that 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!). Since the magnitude of this index number would be expected to increase with 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.

FLOW BETWEENNESS CENTRALITY

	FlowBet	nFlowBet
	-----	-----
1	10.00	7.69
2	49.00	39.20
3	17.00	12.50
4	14.00	10.53
5	48.00	37.80
6	2.00	1.34
7	23.00	16.67
8	8.00	6.06
9	9.00	6.34
10	12.00	9.09

DESCRIPTIVE STATISTICS FOR EACH MEASURE

	FlowBet	nFlowBet
	-----	-----
Mean	19.20	14.72
Std Dev	15.57	12.50
Sum	192.00	147.21
Variance	242.56	156.33
Minimum	2.00	1.34
Maximum	49.00	39.20

Network Centralization Index = 27.199

By this more complete measure of betweenness centrality, actors #2 and #5 are clearly the most important mediators (with the newspaper, #7, being a distant third). 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.

Some actors are clearly more central than others, and the relative variability in flow betweenness of the actors is fairly great (the standard deviation of normed flow betweenness is 14.7 relative to a mean of 12.5, giving a coefficient of relative variation of 85%). Despite this relatively high amount of variation, the degree of inequality, or concentration in the distribution of flow betweenness centralities among the actors is fairly low -- relative to that of a pure star network

(the network centralization index is 27.199). This is slightly higher than the index for the betweenness measure that was based only on geodesic distances.

The Bonacich power index

Phillip Bonacich proposed a modification of the degree centrality approach that has been widely accepted as superior to the original measure. Bonacich's idea, like most good ones, is pretty simple. The original degree centrality approach argues that actors who have more connections are more likely to be powerful because they can directly affect more other actors. This makes sense, but having the same degree does not necessarily make actors equally important.

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. Fred's friends each also have lots of friends, who have lots of friends, and so on. Who is more central? We would probably agree that Fred is, because the people he is connected to are better connected than Bill's people. Bonacich argued that one's centrality is a function of how many connections one has, and how many the connections the actors in the neighborhood had.

While we have argued that more central actors are more likely to be more powerful actors, Bonacich questioned this idea. Compare Bill and Fred again. Fred is clearly more central, but is he more powerful? One argument would be that one is likely to be more influential if one is connected to central others -- because one can quickly reach a lot of other actors with one's message. But if the actors that you are connected to are, themselves, well connected, they are not highly dependent on you -- they have many contacts, just as you do. If, on the other hand, the people to whom you are connected are not, themselves, well connected, then they are dependent on you. Bonacich argued that being connected to connected others makes an actor central, 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.

Bonacich proposed that both centrality and power were a function of the connections of the actors in one's neighborhood. The more connections the actors in your neighborhood have, the more central you are. The fewer the connections the actors in your neighborhood, the more powerful you are. There would seem to be a problem with building an algorithm to capture these ideas. Suppose A and B are connected. Actor A's power and centrality are functions of her own connections, and also the connections of actor B. Similarly, actor B's power and centrality depend on actor A's. So, each actor's power and centrality depends on each other actor's power simultaneously.

There is a way out of this chicken-and-egg type of problem. Bonacich showed that, for symmetric systems, an iterative estimation approach to solving this simultaneous equations problem would eventually converge to a single answer. One begins by giving each actor an estimated centrality equal to their own degree, plus a weighted function of the degrees of the actors to whom they were connected. Then, we do this again, using the first estimates (i.e. we again give each actor an estimated centrality equal to their own first score plus the first scores of those to whom they are connected). As we do this numerous times, the relative sizes (not the

absolute sizes) of all actors scores will come to be the same. The scores can then be re-expressed by scaling by constants.

Let's examine the centrality and power scores for our information exchange data. First, we examine the case where the score for each actor is a positive function of their own degree, and the degrees of the others to whom they are connected. We do this by selecting a positive weight (Beta parameter) for the degrees of others in each actor's neighborhood. This approach leads to a centrality score.

BONACICH CENTRALITY

Beta parameter: 0.500000
 WARNING: Data matrix symmetrized by taking larger of Xij and Xji.

Actor Centrality

	1
	Centrality

1 COUN	-2.68
2 COMM	-3.74
3 EDUC	-2.76
4 INDU	-2.95
5 MAYR	-4.06
6 WRO	-1.16
7 NEWS	-2.67
8 UWAY	-3.49
9 WELF	-2.89
10 WEST	-2.95

STATISTICS

	1
	Centrality

1 Mean	-2.93
2 Std Dev	0.74
3 Sum	-29.32
4 Variance	0.55
5 Euc Norm	9.57
6 Minimum	-4.06
7 Maximum	-1.16

Network Centralization Index = 17.759

Notes: If we look at the absolute value of the index scores, we see the 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. The United Way (#8) also appears

to have high centrality by this measure -- this is a new result. In this case, it is because the United Way is connected to all of the other high degree points by being a source to them -- and we have counted being either a source or a sink as a link.

When we take into account the connections of those to whom each actor is connected, there is less variation in the data than when we examine only adjacencies. The standard deviation of the actor's centralities is fairly small relative to the mean score. This suggests relatively little dispersion or inequality in centralities, which is confirmed by comparing the distribution of the data to that of a perfect star network (i.e. Network centralization index = 17.759).

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.

```
BONACICH POWER
Beta parameter:          -0.500000
WARNING: Data matrix symmetrized by taking larger of Xij and Xji.
```

Actor Power

	1
	Power

1 COUN	-1.00
2 COMM	2.00
3 EDUC	1.00
4 INDU	1.00
5 MAYR	3.00
6 WRO	-0.00
7 NEWS	3.00
8 UWAY	1.00
9 WELF	2.00
10 WEST	1.00

STATISTICS

	1
	Power

1 Mean	1.30
2 Std Dev	1.19
3 Sum	13.00
4 Variance	1.41
5 Euc Norm	5.57
6 Minimum	-1.00
7 Maximum	3.00

Network Centralization Index = 17.000

Notes: Not surprisingly, these results are rather different from many of the others we've examined. I say this is not surprising because most of the measures we've looked at equate centrality with power, whereas the Bonacich power index views power as quite distinct from centrality.

The mayor (#5) and Comm (#2) are both central (from previous results) and powerful. Both of these actors are connected to almost everyone else -- including both the well connected and the weakly connected. Actor #9 (WELF) is identified by this index as being among the more powerful actors. This is probably because of their tie to the WRO (#6) a weakly connected actor. Actor #1, which has high degree, is seen here as not being powerful. If you scan the diagram again, you can see that actor #1's connections are almost all to other highly connected actors.

The Bonacich approach to degree-based centrality and degree-based power are fairly natural extensions of the idea of degree centrality based on adjacencies. One is simply taking into account the connections of one's connections, in addition to one's own connections. The notion that power arises from connection to weak others, as opposed to strong others is an interesting one, and points to yet another way in which the positions of actors in network structures endow them with different potentials.

Summary of chapter 6

Social network analysis methods provide some useful tools for addressing one of the most important (but also one of the most complex and difficult), aspects of social structure: the sources and distribution of power. The network perspective suggests that the power of individual actors is not an individual attribute, but arises from their relations with others. Whole social structures may also be seen as displaying high levels or low levels of power as a result of variations in the patterns of ties among actors. And, the degree of inequality or concentration of power in a population may be indexed.

Power arises from occupying advantageous positions in networks of relations. Three basic sources of advantage are high degree, high closeness, and high betweenness. In simple structures (such as the star, circle, or line), these advantages tend to covary. In more complex and larger networks, there can be considerable disjuncture between these characteristics of a position-- so that an actor may be located in a position that is advantageous in some ways, and disadvantageous in others.

We have reviewed three basic measures of "centrality" of individual positions, and some elaboration on each of the three main ideas of degree, closeness, and betweenness. This review is not exhaustive. The question of how structural position confers power remains a topic of active research and considerable debate. As you can see, different definitions and measures can capture different ideas about where power comes from, and can result in some rather different insights about social structures.

In the last chapter and this one, we have emphasized that social network analysis methods give us, at the same time, views of individuals and of whole populations. One of the most enduring and important themes in the study of human social organization, however, is the importance of

social units that lie between the two poles of individuals and whole populations. In the next chapter, we will turn our attention to how network analysis methods describe and measure the differentiation of sub-populations.

Review questions for chapter 6

1. What is the difference between "centrality" and "centralization?"
2. Why is an actor who has higher degree a more "central" actor?
3. How does Bonacich's influence measure extend the idea of degree centrality?
4. Can you explain why an actor who has the smallest sum of geodesic distances to all other actors is said to be the most "central" actor, using the "closeness" approach?
5. How does the "flow" approach extend the idea of "closeness" as an approach to centrality?
6. What does it mean to say that an actor lies "between" two other actors? Why does betweenness give an actor power or influence?
7. How does the "flow" approach extend the idea of "betweenness" as an approach to centrality?
8. Most approaches suggest that centrality confers power and influence. Bonacich suggests that power and influence are not the same thing. What is Bonacich' argument? How does Bonacich measure the power of an actor?

Application questions for chapter 6

1. Think of the readings from the first part of the course. Which studies used the ideas of structural advantage, centrality, power and influence? What kinds of approach did each use: degree, closeness, or betweenness?
2. Can you think of any circumstances where being "central" might make one less influential? Less powerful?
3. Consider a directed network that describes a hierarchical bureaucracy, where the relationship is "gives orders to." Which actors have highest degree? are they the most powerful and influential? Which actors have high closeness? Which actors have high betweenness?
4. Can you think of a real-world example of an actor who might be powerful but not central? Who might be central, but not powerful?

7. Cliques and Sub-Groups

Introduction: Groups and Sub-structures

One of the most common interests of structural analysts is in the "sub-structures" that may be present in a network. Dyads, Triads, and ego-centered circles can all be thought of as sub-structures. Networks are also built up, or developed out of the combining of dyads and triads into larger, but still closely connected structures. Many of the approaches to understanding the structure of a network emphasize how dense connections are compounded and extended to develop larger "cliques" or sub-groupings. 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, and k-plexes all look at networks this way.

Divisions of actors into cliques or "sub-groups" can be a very important aspect of social structure. It can be important in understanding how the network as a whole is likely to behave. For example, suppose the actors in one network form two non-overlapping cliques; and, suppose that the actors in another network also form two cliques, but that the memberships overlap (some people are members of both cliques). Where the groups overlap, we might expect that conflict between them is less likely than when the groups don't overlap. Where the groups overlap, mobilization and diffusion may spread rapidly across the entire network; where the groups don't overlap, traits may occur in one group and not diffuse to the other.

Knowing how an individual is embedded in the structure of groups within a net may also be critical to understanding his/her behavior. For example, some people may act as "bridges" between groups (cosmopolitans, boundary spanners). Others may have all of their relationships within a single clique (locals or insiders). Some actors may be part of a tightly connected and closed elite, while others are completely isolated from this group. Such differences in the ways that individuals are embedded in the structure of groups within a network can have profound consequences for the ways that these actors see the work, and the behaviors that they are likely to practice.

The idea of sub-structures or groups or cliques within a network is a powerful tool for understanding social structure and the embeddedness of individuals. The general definition of a clique is pretty straight-forward: a clique is simply a sub-set of actors who are more closely tied to each other than they are to actors who are not part of the group. But, when one wants to get more precise about cliques, and to apply these ideas rigorously to looking at data, things get a little more complicated.

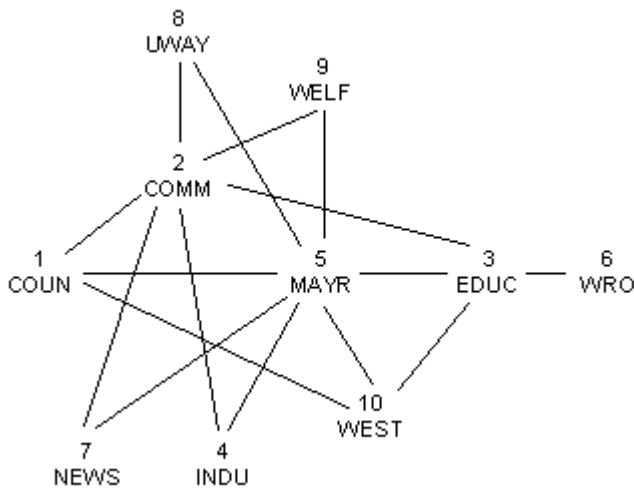
To see some of the main ideas, and to examine some of the major ways in which structural analysts have defined cliques, let's work with an example. Again, all analyses are performed using UCINET software on a PC.

Most computer algorithms for defining cliques operate on binary symmetric data. We will use the Knoke information exchange data again. 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; that is, a tie only exists if xy and yx are both present.

The resulting symmetric data matrix looks like this:

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

Insisting that information move in both directions between the parties in order for the two parties to be regarded as "close" makes theoretical sense, and substantially lessens the density of the matrix. Matrices that have very high density, almost by definition, are likely to have few distinctive sub-groups or cliques. It might help to graph these data.



Notes: The diagram suggests a number of things. Actors #5 and #2 appear to be in the middle of the action -- in the sense that they are members of many of the groupings, and serve to connect them, by co-membership. The connection of sub-graphs by actors can be an important feature. We can also see that there is one case (#6) that is not a member of any sub-group (other than a dyad). If you look closely, you will see that dyads and triads are the most common sub-graphs

here -- and despite the substantial connectivity of the graph, tight groupings larger than this seem to be few. It is also apparent from visual inspection that most of the sub-groupings are connected -- that groups overlap.

The main features of a graph, in terms of its cliques or sub-graphs, may be apparent from inspection:

- How separate are the sub-graphs (do they overlap and share members, or do they divide or factionalize the network)?
- How large are the connected sub-graphs? Are there a few big groups, or a larger number of small groups?
- Are there particular actors that appear to play network roles? For example, act as nodes that connect the graph, or who are isolated from groups?

The formal tools and concepts of sub-graph structure help to more rigorously define ideas like this. Various algorithms can then be applied to locate, list, and study sub-graph features. Obviously, there are a number of possible groupings and positions in sub-structures, depending on our definitions. Below, we will look at the most common of these ideas.

Bottom-up Approaches

In a sense, all networks are composed of groups (or sub-graphs). 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. A clique extends the dyad by adding to it members who are tied to all of the members in the group. This strict definition can be relaxed to include additional nodes that are not quite so tightly tied (n-cliques, n-clans, k-plexes and k-cores). The notion, however, is to build outward from single ties to "construct" the network. A map of the whole network can be built up by examining the sizes of the various cliques and clique-like groupings, and noting their size and overlaps.

These kinds of approaches to thinking about the sub-structures of networks tend to emphasize how the macro might emerge out of the micro. They tend to focus our attention on individuals first, and try to understand how they are embedded in the web of overlapping groups in the larger structure. I make a point of these seemingly obvious ideas because it is also possible to approach the question of the sub-structure of networks from the top-down. Usually, both approaches are worthwhile and complementary. We will turn our attention first to "bottom-up" thinking.

Cliques

The idea of a clique is relatively simple. At the most general level, a clique is a sub-set of a network in which the actors are more closely and intensely tied to one another than they are to other members of the network. In terms of friendship ties, for example, it is not unusual for people in human groups to form "cliques" on the basis of age, gender, race, ethnicity, religion/ideology, and many other things. The smallest "cliques" are composed of two actors: the

dyad. But dyads can be "extended" to become more and more inclusive -- forming strong or closely connected components in graphs. A number of approaches to finding groups in graphs can be developed by extending the close coupling of dyads to larger structures.

The strongest possible definition of a clique is some number of actors (more than two, usually three is used) who have all possible ties present among themselves. A "Maximal Complete Sub-Graph" is such a grouping, expanded to include as many actors as possible. We asked UCINET to find all of the cliques in the Knoke data that satisfy the "maximal complete sub-graph" criterion for the definition of a clique.

- 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

Notes: There are seven maximal complete sub-graphs present in these data (see if you can find them in figure one). The largest one is composed of four of the ten actors and all of the other smaller cliques share some overlap with some part of the core. We might be interested in the extent to which these sub-structures overlap, and which actors are most "central" and most "isolated" from the cliques. We can examine these questions by looking at "co-membership."

Group Co-Membership Matrix

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

Notes: It is immediately apparent that actor #6 is a complete isolate, and that actors #2 and #5 overlap with almost all other actors in at least one clique. We see that actors #2 and #5 are "closest" in the sense of sharing membership in five of the seven cliques. We can take this kind of analysis one step further by using single linkage agglomerative cluster analysis to create a "joining sequence" based on how many clique memberships actors have in common.

HIERARCHICAL CLUSTERING

Level	6	4	7	8	9	1	3	5	2	0
5	XXX	.	.
2	XXXXXXXXXX	.	.	.
1	.	XXXXXXXXXXXXXXXXXXXX
0	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX

Notes: We see that actors 2 and 5 are "joined" first as being close because they share 5 clique memberships in common. At the level of sharing only two clique memberships in common, actors #1, #3, and #10 "join the core." If we require only one clique membership in common to define group membership, then all actors are joined except #6.

Insisting that every member of a clique be connected to every other member is a very strong definition of what we mean by a group. There are a number of ways in which this restriction could be relaxed. Two major approaches are the N-clique/N-clan approach and the k-plex approach.

N-cliques

The strict clique definition (maximal fully connected sub-graph) may be too strong for many purposes. It insists that every member or a sub-group have a direct tie with each and every other member. You can probably think of cases of "cliques" where at least some members are not so tightly or closely connected. There are two major ways that the "clique" definition has been "relaxed" to try to make it more helpful and general.

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 a distance greater than one. Usually, the path distance two is used. This corresponds to being "a friend of a friend." This approach to defining sub-structures is called N-clique, where N stands for the length of the path allowed to make a connection to all other members. When we apply the N-Clique definition to our data, we get the following result.

```

Max Distance (n-):      2
Minimum Set Size:      3

2 2-cliques found.

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

Notes: 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. Because our definition

of how closely linked actors must be to be members of a clique has been relaxed, there are fewer maximal cliques. With larger and fewer sub-groups, 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.

Group Co-Membership Matrix

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

HIERARCHICAL CLUSTERING

											1
Level	1	4	6	7	8	9	5	3	2	0	
	-	-	-	-	-	-	-	-	-	-	-
2	XXXXXXX				
1	XXXXXXXXXXXXXXXXXXXXX										

Notes: An examination of the clique co-memberships and a clustering of closeness under this definition of a clique gives a slightly different picture than the strong-tie approach. Actors #2, #3, #5, and #10 form a "core" or "inner circle" in this method -- as opposed to the 2,5,4,7 core by the maximal method (with #5 as a clear "star").. One cannot say that one result or another is correct.

N-Clans

The N-clique approach tends to find long and stringy groupings rather than the tight and discrete ones of the maximal approach. In some cases, N-cliques can be found that have a property that is probably un-desirable 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.

For these reasons, some analysts have suggested restricting N-cliques by insisting that the total span or path distance between any two members of an N-clique also satisfy a condition. The kind of a restriction has the effect of forcing all ties among members of an n-clique to occur by way of other members of the n-clique. This approach is the n-clan approach. In the current case, the N-

clans approach (i.e. N-cliques with an additional condition limiting the maximum path length within a clique) does not differ from the N-clique approach.

N-CLANS

Max Distance (n-): 2
 Minimum Set Size: 3

2 2-clans found.

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

Group Co-Membership Matrix

		1	2	3	4	5	6	7	8	9	0
	COCOEDINMAWRNEUWWEWE										
1	COUN	1	1	1	1	1	0	1	1	1	1
2	COMM	1	2	2	1	2	1	1	1	1	2
3	EDUC	1	2	2	1	2	1	1	1	1	2
4	INDU	1	1	1	1	1	0	1	1	1	1
5	MAYR	1	2	2	1	2	1	1	1	1	2
6	WRO	0	1	1	0	1	1	0	0	0	1
7	NEWS	1	1	1	1	1	0	1	1	1	1
8	UWAY	1	1	1	1	1	0	1	1	1	1
9	WELF	1	1	1	1	1	0	1	1	1	1
10	WEST	1	2	2	1	2	1	1	1	1	2

SINGLE-LINK HIERARCHICAL CLUSTERING

	C	I	N	U	W	M	E	C	W
	O	N	W	E	W	E	A	D	O
	U	D	R	W	A	L	Y	U	M
	N	U	O	S	Y	F	R	C	M
Level	1	4	6	7	8	9	5	3	2
-----	-	-	-	-	-	-	-	-	-
2	XXXXXXX		
1	XXXXXXXXXXXXXXXXXXXX								

notes: The n-clique and n-clan approaches provide an alternative to the stricter "clique" definition, and this more relaxed approach often makes good sense with sociological data. In essence, the n-clique approach allows an actor to be a member of a clique even if they do not have ties to all other clique members; just so long as they do have ties to some member, and are no further away than n steps (usually 2) from all members of the clique. The n-clan approach is a relatively minor modification on the n-clique approach that requires that the all the ties among actors occur through other members of the group.

If one is uncomfortable with regarding the friend of a clique member as also being a member of the clique (the n-clique approach), one might consider an alternative way of relaxing the strict assumptions of the clique definition -- the K-plex approach.

K-plexes

An alternative way of relaxing the strong assumptions of the "Maximal Complete Sub-Graph" is to allow that actors may be members of a clique even 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 fall in clique under the K-Plex approach. 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. The k-plex approach would seem to have quite a bit in common with the n-clique approach, but k-plex analysis often gives quite a different picture of the sub-structures of a graph. Rather than the large and "stringy" groupings sometimes produced by n-clique analysis, k-plex analysis tends to find relatively large numbers of smaller groupings. This tends to focus attention on overlaps and co-presence (centralization) more than solidarity and reach.

In our example, below, we have allowed k to be equal to two. That is, an actor is considered to be a member of a clique if that actor has ties to all but two others in that clique. With this definition, there are many cliques of size three (indeed, in our graph there are 35 K-plex cliques, but only six of them have more than three members). With K=2, a node needs only to be connected to one other member of a clique of three in order to qualify. This doesn't sound like much of a "clique" to me, so I've dropped these groupings of three cases. The remaining K-cliques are:

K-PLEXES :

```
COUN COMM EDUC MAYR WEST
COUN COMM EDUC WEST
COUN COMM INDU MAYR
COUN COMM MAYR NEWS
COUN COMM MAYR UWAY
COUN COMM MAYR WELF
COMM EDUC INDU MAYR
COMM EDUC MAYR NEWS
COMM EDUC MAYR UWAY
COMM EDUC MAYR WELF
COMM INDU MAYR NEWS
COMM INDU MAYR UWAY
COMM INDU MAYR WELF
COMM MAYR NEWS UWAY
COMM MAYR NEWS WELF
COMM MAYR UWAY WELF
```

Notes: The COMM is present in every k-component; the MAYR is present in all but one. Clearly these two actors are "central" in the sense of playing a bridging role among multiple slightly different social circles. Again we note that organization #6 (WRO) is not a member of any K-

plex clique. 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 sub-structures makes a good deal of sense for many problems. It requires that members of a group have ties to (most) other group members -- ties by way of intermediaries (like the n-clique approach) do 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. Again, it is not that one is "right" and the other "wrong." Depending on the goals of the analysis, both can yield valuable insights into the sub-structure of groups.

K-cores

A k-core is a maximal group of actors, all of whom are connected to some number (k) of other members of the group. To be included in a k-plex, an actor must be tied to all but k other actors in the group. The k-core approach is more relaxed, allowing actors to join the group if they are connected to k members, regardless of how many other members they may not be connected to. By varying the value of k (that is, how many members of the group do you have to be connected to), different pictures can emerge. K-cores can be (and usually are) more inclusive than k-plexes. And, as k becomes smaller, group sizes will increase.

In our example data, if we require that each member of a group have ties to 3 other members (a 3-core), a rather large central group of actors is identified {1,2,3,4,5,7,10}. Each of the seven members of this core has ties to at least three others. If we relax the criterion to require only two ties, actors 8 and 9 are added to the group (and 6 remains an isolate). If we require only one tie (really, the same thing as a component), all actors are connected.

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 sub-group.

Top-down Approaches

The approaches we've examined to this point start with the dyad, and see if this kind of tight structure can be extended outward. Overall structure of the network is seen as "emerging" from overlaps and couplings of smaller components. Certainly, this is a valid way of thinking about large structures and their component parts.

Some might prefer, however, to start with the entire network as their frame of reference, rather than the dyad. Approaches of this type tend to look at the "whole" structure, and identify "sub-structures" as parts that are locally denser than the field as a whole. In a sense, this more macro lens is looking for "holes" or "vulnerabilities" or "weak spots" in the overall structure or solidarity of the network. These holes or weak spots define lines of division or cleavage in the larger group, and point to how it might be de-composed into smaller units.

These verbal descriptions of the lenses of "bottom up" and "top down" approaches greatly overstate the differences between the approaches, when we look closely at specific definitions

and concrete algorithms that can be used to locate and study sub-structure. Still, the logic, and some of the results of studying networks from the top-down and from the bottom-up can lead to different (and usually complementary) insights.

Components

Components of a graph are parts that are connected within, but disconnected between sub-graphs. If a graph contains one or more "isolates," these actors are components. More interesting components are those which divide the network into separate parts, and where each part has several actors who are connected to one another (we pay no attention to how closely connected).

In the example that we have been using here, the symmetrized Knoke information exchange data, there is only a single component. That is, all of the actors are connected. This may often be the case, even with fairly large networks with seemingly obvious sub-structures. Rather as the strict definition of a "clique" may be too strong to capture the meaning of the concept, the strong notion of a component is usually too strong to find meaningful weak-points, holes, and locally dense sub-parts of a larger graph. So, we will examine some more flexible approaches.

Blocks and Cutpoints

One approach to finding these key spots in the diagram is to ask if a node were removed, would the structure become divided into un-connected systems? If there are such nodes, they are called "cutpoints." And, one can imagine that such cutpoints may be particularly important actors -- who may act as brokers among otherwise disconnected groups. The divisions into which cutpoints divide a graph are called blocks. We can find the maximal non-separable sub-graphs (blocks) of a graph by locating the cutpoints. That is, we try to find the nodes that connect the graph (if there are any).

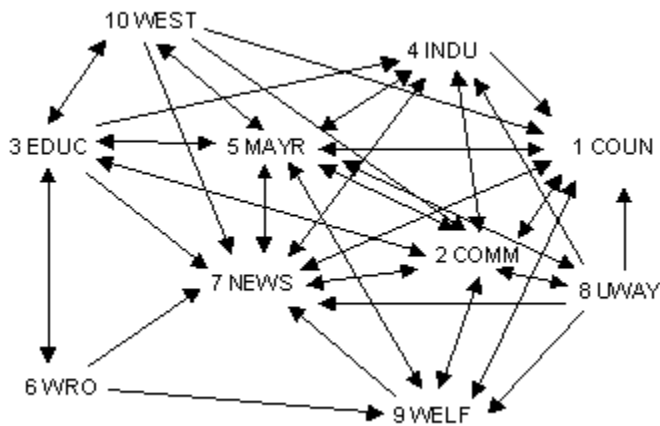
In the data set we've been examining, there is one (and only one) cutpoint -- though it is not really very interesting.

```
2 blocks found.
```

```
BLOCKS:
```

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

Notes: We see that the graph is de-composable (or dis-connectable) into two blocks. EDUC is the member that spans the two separable sub-graphs, and hence is the "cut-point." You can confirm this with a quick glance at the figure (for these analyses, we have used the directed, rather than symmetrized data).



Here you can see that EDUC is the only point which, if removed, would result in a dis-connected structure (that is, WRO would no longer be reachable by other actors). In this case, it turns out that EDUC plays the role of connecting an otherwise isolated single organization (WRO) to the remaining members of the network.

Lambda Sets and Bridges

An alternative approach is to ask if there are certain connections in the graph which, if removed, would result in a disconnected structure. That is, are there certain key relationships (as opposed to certain key actors). In our example, the only relationship that qualifies is that between EDUC and WRO. But, since this would only lop-off one actor, rather than really altering the network, it is not very interesting. However, it is possible to approach the question in a more sophisticated way. 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 go through each link. It then identifies sets of actors who, 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. For our data, the results of a Lambda set analysis are:

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	.	.	.	XXXXXXXXXXXXXXXXXX							
2	.	XXXXXXXXXXXXXXXXXXXX									
1	XXXXXXXXXXXXXXXXXXXX										

Notes: This approach identifies the #2 to #5 (MAYR to COMM) linkage as the most important one in the graph - in the sense that it carries a great deal of traffic, and the graph would be most disrupted if it were removed. This result can be confirmed by looking at figure one, where we see that most actors are connected to most other actors by way of the linkage between #2 and #5. Considerably less critical are linkages between 2 and 5 and actors 1, 3, 4, 7, and 10. Again, a glance at the figure shows these organizations to be a sort of "outer circle" around the core.

The lambda set idea has moved us quite far away from the strict components idea. Rather than emphasizing the "decomposition" or separation of the structure into un-connected components, the lambda set idea is a more "continuous" one. It highlights points at which the fabric of connection is most vulnerable to disruption. Again, there is no "right" or "wrong" way to approach the notion that large structures may be non-homogeneous and more subject to division and disruption in some locations than others. Our last approach, in a sense, combines the logic of both "top down" approaches.

Factions

In network terms, actors are said to be equivalent to the extent that they have the same profiles of ties to other actors. It follows that we might define partitions of the network on the basis of grouping together actors on the basis of similarity in who they are tied to. At one extreme, we could have a graph that is divisible into components, and where each component is a clique (that is, within a group, everyone is tied to everyone; between groups, there are no ties at all). Somewhat less strict divisions would allow that there might be some ties between groups, and less than full density within groups.

Using the power of the computer, it is possible to search for partitions of a graph into sub-graphs that maximize the similarity of the patterns of connections of actors within each group. It is also possible to then assess "how good" the partitioning is by comparing the result to a pure "ideal typical" partitioning where actors within each group have maximum similarity, and actors across groups have maximum dissimilarities of ties.

The first step in such an approach is to try to decide what number of groups or divisions provide a useful and interesting picture of the data. We tried 2 through 5 groups, and settled on 4. The decision was made on how clean and interpretable the resulting partition of the adjacency matrix looked.

FACTIONS

```

Diagonal valid?      YES
Use geodesics?      NO
Method:              CORRELATION
Treat data as:      SIMILARITIES
Number of factions:  4
  
```

Fit: -0.503

Group Assignments:

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

Grouped Adjacency Matrix

		6	3	2	4	5	1	7	8	9	0
		W	E	C	I	M	C	N	U	W	W
6	WRO	1	1					1		1	
3	EDUC	1	1	1	1	1		1			1
2	COMM		1	1	1	1	1	1	1	1	
4	INDU			1	1	1	1	1			
5	MAYR		1	1	1	1	1	1	1	1	1
1	COUN			1		1	1	1		1	
7	NEWS			1	1	1		1			
8	UWAY			1	1	1	1	1	1	1	
9	WELF			1		1		1		1	
10	WEST		1	1	1	1	1				1

This approach corresponds nicely to the intuitive notion that the groups of a graph can be defined by a combination of local high density, and the presence of "structural holes" between some sets of actors and others. In this case, the picture is one of stronger and weaker parts of the fabric, rather than actual "holes." Since we have used directed data for this analysis, note that the division is based on both who sends information to whom and who receives it from whom. The grouping {6, 3}, for example sends information to other groups, but is divided from them because other groups do not send as much information back. The picture then not only identifies

actual or potential factions, but also tells us about the relations among the factions -- potential allies and enemies, in some cases.

Summary of chapter 7

One of the most interesting thing about social structures is their sub-structure in terms of groupings or cliques. The number, size, and connections among the sub-groupings in a network can tell us a lot about the likely behavior of the network as a whole. How fast will things move across the actors in the network? Will conflicts most likely involve multiple groups, or two factions. To what extent do the sub-groups and social structures over-lap one another? All of these aspects of sub-group structure can be very relevant to predicting the behavior of the network as a whole.

The location of individuals in nets can also be thought of in terms of cliques or sub-groups. Certain individuals may act as "bridges" among groups, others may be isolates; some actors may be cosmopolitans, and others locals in terms of their group affiliations. Such variation in the ways that individuals are connected to groups or cliques can be quite consequential for their behavior as individuals.

In this section we have briefly reviewed some of the most important definitions of "sub-groups" or "cliques." and examined the results of applying these definitions to a set of data. We have seen that different definitions of what a clique is can give rather different pictures of the same reality.

Review questions for chapter 7

1. Can you explain the term "maximal complete sub-graph?"
2. How do N-cliques and N-clans "relax" the definition of a clique?
3. Give an example of when it might be more useful to use a N-clique or N-clan approach instead of a strict clique.
4. How do K-plexes and K-cores "relax" the definition of a clique?
5. Give an example of when it might be more useful to us a K-plex or K-core approach instead of a strict clique.
6. What is a component of a graph?
7. How does the idea of a "block" relax the strict definition of a component?
8. Are there any cut points in the "star" network? in the "line" network? in the "circle" network?
9. How does the idea of a lambda set relax the strict definition of a component?
10. Are there any "bridges" in a strict hierarchy network?

Application questions for chapter 7

1. Think of the readings from the first part of the course. Which studies used the ideas of group sub-structures? What kinds of approaches were used: cliques, clans, plexes, etc.?
2. Try to apply the notion of group sub-structures at different levels of analysis. Are their sub-structures within the kinship group of which you are a part? How is the population of Riverside divided into sub-structures? Are there sub-structures in the population of Universities in the United States? Are the nations in the world system divided into sub-structures in some way?
3. How might the lives of persons who are "cut points" be affected by having this kind of a structural position? Can you think of an example?
4. Can you think of a real-world (or literary) example of a population with sub-structures? How might the sub-structures in your real world case be described using the formal concepts (are the sub structures "clans" or "factions" etc.).

8. Network Positions and Social Roles: The Idea of Equivalence

Introduction to positions and roles

We have been examining some of the ways that structural analysts look at network data. We began by looking for patterns in the overall structure (e.g. connectedness, density, etc.) And the embeddedness of each actor (e.g. geodesic distances, centrality). A second major way of going about examining network data is to look for "sub-structures," or groupings of actors that are closer to one another than they are to other groupings. For example, we looked at the meaning of "cliques" "blocks" and "bridges" as ways of thinking about and describing how the actors in a network may be divided into sub-groups on the basis of their patterns of relations with one another. All of this, while sometimes a bit technical, is pretty easy to grasp conceptually. The central node of a "star" network is "closer" to all other members than any other member -- a simple (if very important) idea that we can grasp. A clique as a "maximal complete subgraph" sounds tough, but, again, is easy to grasp. It is simply the biggest collection of folks who all have connections with everyone else in the group. Again, the idea is not difficult to grasp, because it is really quite concrete: we can see and feel cliques.

Now we are going to turn our attention to somewhat more abstract ways of making sense of the patterns of relations among social actors: the analysis of "positions." Being able to define, theorize about, and analyze data in terms of positions is important because we want to be able to make generalizations about social behavior and social structure. That is, we want to be able to state principles that hold for all groups, all organizations, all societies, etc. To do this, we must think about actors not as individual unique persons (which they are), but as examples of categories. As an empirical task, we need to be able to group together actors who are the most similar, and to describe what makes them similar; and, to describe what makes them different, as a category, from members of other categories.

Sociological thinking uses abstract categories routinely. "Working class, middle class, upper class" are one such set of categories that describe social positions. "Men and Women" are really labels for categories of persons who are more similar within category than between category -- at least for the purposes of understanding and predicting some aspects of their social behavior. When categories like these are used as parts of sociological theories, they are being used to describe the "social roles" or "social positions" typical of members of the category.

Many of the category systems used by sociologists are based on "attributes" of individual actors that are in common across actors. If I state that "European-American males, ages 45-64 are likely to have relatively high incomes" I am talking about a group of people who are demographically similar -- they share certain attributes (maleness, European ancestry, biological age, and income). Structural analysis is not particularly concerned with systems of categories (i.e. variables), that are based on descriptions of similarity of individual attributes (some radical structural analysts would even argue that such categories are not really "sociological" at all). Structural analysts seek to define categories and variables in terms of similarities of the patterns of relations among

actors, rather than attributes of actors. That is, the definition of a category, or a "social role" or "social position" depends upon its relationship to another category. Social roles and positions, structural analysts argue, are inherently "relational." That's pretty abstract in itself. Some examples can make the point.

What is the social role "husband?" One useful way to think about it is as a set of patterned interactions with a member or members of some other social categories: "wife" and "child" (and probably others). Each one of these categories (i.e. husband, wife, child) can only be defined by regularities in the patterns of relationships with members of other categories (there are a number of types of relations here -- monetary, emotional, ritual, sexual, etc.). That is, family and kinship roles are inherently relational.

What is a "worker?" We could mean a person who does labor (an attribute, actually one shared by all humans). A more sociologically interesting definition was given by Marx as a person who sells control of their labor power to a capitalist. Note that the meaning of "worker" depends upon a capitalist -- and vice versa. It is the relation (in this case, as Marx would say, a relation of exploitation) between occupants of the two roles that defines the meaning of the roles.

The point is: to the structural analyst, the building blocks of social structure are "social roles" or "social positions." These social roles or positions are defined by regularities in the patterns of relations among actors, not attributes of the actors themselves. We identify and study social roles and positions by studying relations among actors, not by studying attributes of individual actors. Even things that appear to be "attributes of individuals" such as race, religion, and age can be thought of as short-hand labels for patterns of relations. For example, "white" as a social category is really a short-hand way of referring to persons who typically have a common form of relationships with members of another category -- "non-whites." Things that might at first appear to be attributes of individuals are really just ways of saying that an individual falls in a category that has certain patterns of characteristic relationships with members of other categories.

Approaches to Network Positions and Social Roles

Because "positions" or "roles" or "social categories" are defined by "relations" among actors, we can identify and empirically define social positions using network data. In an intuitive way, we would say that two actors have the same "position" or "role" to the extent that their pattern of relationships with other actors is the same. But, there are a couple things about this intuitive definition that are troublesome.

First, what relations do we take into account, among whom, in seeking to identify which actors are similar and which are not. The relations that I have with the University are similar in some ways to the relations that you have with the University: we are both governed by many of the same rules, practices, and procedures. The relations I have with the University are very different from yours in some ways (e.g. they pay me, you pay them). Which relations should count and which ones not, in trying to describe the roles of "professor" and "student?" Indeed, why am I examining relations among you, me, and the University, instead of including, say, members of the state legislature? There is no simple answer about what the "right relations" are to examine; and, there is no simple answer about who the relevant set of "actors" are. It all depends upon the

purposes of our investigation, the theoretical perspective we are using, and the populations to which we would like to be able to generalize our findings.

The second problem with our intuitive definition of a "role" or "position" is this: assuming that I have a set of actors and a set of relations that make sense for studying a particular question, what do I mean that actors who share the same position are similar in their pattern of relationships or ties? The idea of "similarity" has to be rather precisely defined. Again, there is no single and clear "right" answer for all purposes of investigation. But, there are rigorous ways of thinking about what it means to be "similar" and there are rigorous ways of actually examining data to define social roles and social positions empirically. These are the issues to which we will devote the remainder of this (somewhat lengthy) section.

Defining Equivalence or Similarity

What do we mean when we say that two actors have "similar" patterns of relations, and hence are both members of a particular social role or social position? There are at least two quite different things that we might mean by "similar." Some analysts have labeled these types of similarity as "structural equivalence," and "regular equivalence." The two types of similarity differ in their degrees of abstraction, with structural equivalence being most concrete and regular equivalence being most abstract.

Structural Equivalence

Two nodes are said to be exactly structurally equivalent if they have the same relationships to all other nodes. Structural equivalence is easy to grasp (though it can be operationalized in a number of ways) because it is very concrete. Two actors are equivalent to the extent that they have the same relationships with all other actors. If A likes B and C likes B, A and C are structurally equivalent (note that whether A and C like each other doesn't matter, because A and C have identical patterns of ties either way). If two nodes are exactly structurally equivalent, they will also be automorphically and regularly equivalent; this is because "structurally equivalent" really means the same thing as "identical" or "substitutable." Borgatti, et al. refer to structural equivalence as "equivalence of network location." Because exact structural equivalence is likely to be rare (particularly in large networks), we often are interested in examining the degree of structural equivalence, rather than the simple presence or absence of exact equivalence.

Automorphic Equivalence

Regular Equivalence

Two nodes are said to be regularly equivalent if they have the same profile of ties with members of other sets of actors that are also regularly equivalent. This is a complicated way of saying something that we recognize intuitively. Two mothers, for example, are "equivalent" because each has a certain pattern of ties with a husband, children, and in-laws (for one example -- but one that is very culturally relative). The two mothers do not have ties to the same husband (usually) or the same children or in-laws. That is, they are not "structurally equivalent." But they are similar because they have the same relationships with some member or members of another

set of actors (who are themselves regarded as equivalent because of the similarity of their ties to a member of the set "mother"). This is an obvious notion, but a critical one. Regular equivalence sets describe the social roles or social types that are the basic building blocks of all social structures. Actors that are regularly equivalent do not necessarily fall in the same network positions or locations with respect to other individual actors; rather, they have the same kinds of relationships with some members of other sets of actors.

Actors who are "structurally equivalent" are necessarily "regularly equivalent." Actors who are "regularly equivalent" are not necessarily "structurally equivalent." Structural equivalence is easier to examine empirically, because it involves specific individual actors; regular equivalence is more difficult to examine empirically, because we must develop abstract categories of actors in relation to other abstract categories.

Review questions for chapter 8

1. How are network roles and social roles different from network "sub-structures" as ways of describing social networks?
2. Explain the differences among structural, automorphic, and regular equivalence.
3. Actors who are structurally equivalent have the same patterns of ties to the same other actors. How do correlation, distance, and match measures index this kind of equivalence or similarity?
4. If the adjacency matrix for a network can be blocked into perfect sets of structurally equivalent actors, all blocks will be filled with zeros or with ones. Why is this?
5. If two actors have identical geodesic distances to all other actors, they are (probably) automorphically equivalent. Why does having identical distances to all other actors make actors "substitutable" but not necessarily structurally equivalent?
6. Regularly equivalent actors have the same pattern of ties to the same kinds of other actors -- but not necessarily the same distances to all other actors, or ties to the same other actors. Why is this kind of equivalence particularly important in sociological analysis?

Application questions for chapter 8

1. Think of the readings from the first part of the course. Did any studies use the idea of structural equivalence or network role? Did any studies use the idea of regular equivalence or social role?
2. Think about the star network. How many sets of structurally equivalent actors are there? What are the sets of automorphically equivalent actors? Regularly equivalent actors? What about the circle network?
3. Examine the line network carefully -- this one's a little trickier. Describe the structural equivalence and regular equivalence sets in a line network.
4. Consider our classical hierarchical bureaucracy, defined by a network of directed ties of "order

giving" from the top to the bottom. Make an adjacency matrix for a simple bureaucracy like this. Block the matrix according to the regular equivalence sets; block the matrix according to structural equivalence sets. How (and why) do these blockings differ? How do the permuted matrices differ?

5. Think about some social role (e.g. "mother") what would you say are the kinds of ties with what other social roles that could be used to identify which persons in a population were "mothers" and which were not? Note the relational character of social roles -- one social role can only be defined with respect to others. Provide some examples of social roles from an area of interest to you.

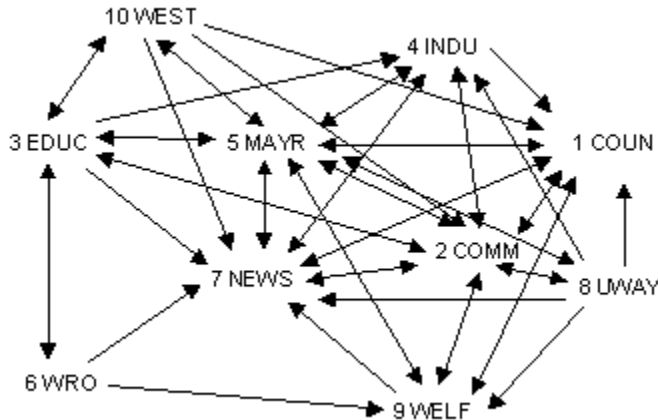
9. Measures of similarity and structural equivalence

Introduction to chapter 9

In this section we will examine some of the ways in which we can empirically define and measure the degree of structural equivalence, or similarity of network position among actors. We will then examine a few of the possible approaches to analyzing the patterns of structural equivalence based on the measures of similarity..

Recall that by structural equivalence, we mean that actors are "substitutable." That is, actors have the same pattern of relationships with all other actors. Exact structural equivalence of actors is rare in most social structures (as a mental exercise, try to imagine what a social structure might be like in which most actors were "substitutable" for one another). Consequently, we often compute measures of the degree to which actors are similar, and use this as the basis for seeking to identify sets of actors that are very similar to one another, and distinct from actors in other sets.

For illustrative purposes, we will analyze the data on information sending and receiving from the Knoke bureaucracies data set. These data show directed ties, and are measured at the nominal (binary) level.



This is a very restricted analysis, and should not be taken very seriously. But, the data are useful to provide an illustration of the main approaches to measuring structural equivalence, and identifying actors who have similar network positions.

Measuring structural similarity

We might try to assess which nodes are most similar to which other nodes intuitively by looking at the diagram. We would notice some important things. It would seem that actors 2,5, and 7

might be structurally similar in that they seem to have reciprocal ties with each other and almost everyone else. Actors 6, 8, and 10 are "regularly" similar in that they are rather isolated; but they are not structurally similar because they are connected to quite different sets of actors. But, beyond this, it is really rather difficult to assess structural similarity rigorously by just looking at a diagram.

We can be a lot more precise in assessing structural similarity if we use the matrix representation of the network instead of the diagram. This also lets us use the computer to do some of the quite tedious jobs involved in calculating index numbers to assess similarity. The original data matrix has been reproduced below. Many of the features that were apparent in the diagram are also easy to grasp in the matrix. If we look across the rows and count out-degrees, and if we look down the columns (to count in-degree) we can see who the central actors are and who are the isolates. But, even more generally, we can see that two actors are structurally equivalent to extent that the profile of scores in their rows and columns are similar.

	1 Coun	2 Comm	3 Educ	4 Indu	5 Mayr	6 WRO	7 News	8 UWay	9 Welf	10 West
1 Coun	---	1	0	0	1	0	1	0	1	0
2 Comm	1	---	1	1	1	0	1	1	1	0
3 Educ	0	1	---	1	1	1	1	0	0	1
4 Indu	1	1	0	---	1	0	1	0	0	0
5 Mayr	1	1	1	1	---	0	1	1	1	1
6 WRO	0	0	1	0	0	---	1	0	1	0
7 News	0	1	0	1	1	0	---	0	0	0
8 UWay	1	1	0	1	1	0	1	---	1	0
9 Welf	0	1	0	0	1	0	1	0	---	0
10 West	1	1	1	0	1	0	1	0	0	---

Two actors may be said to be structurally equivalent to if they have the same patterns of ties with other actors. This means that the entries in the rows and columns for one actor are identical to those of another. If the matrix were symmetric, we would need only to scan pairs of rows (or columns). But, since these data are on directed ties, we should examine the similarity of sending and receiving of ties (of course, we might be interested in structural equivalence with regard to only sending, or only receiving ties). We can see the similarity of the actors if we expand the matrix a bit by listing the row vector followed by the column vector for each actor as a column,

as we have below:

1 Coun	2 Comm	3 Educ	4 Indu	5 Mayr	6 WRO	7 News	8 UWay	9 Welf	10 West
---	1	0	1	1	0	0	1	0	1
1	---	1	1	1	0	1	1	1	1
0	1	---	0	1	1	0	0	0	1
0	1	1	---	1	0	1	1	0	0
1	1	1	1	---	0	1	1	1	1
0	0	1	0	0	---	0	0	0	0
1	1	1	1	1	1	---	1	1	1
0	1	0	0	1	0	0	---	0	0
1	1	0	0	1	1	0	1	---	0
0	0	1	0	1	0	0	0	0	---
---	1	0	0	1	0	1	0	1	0
1	---	1	1	1	0	1	1	1	0
0	1	---	1	1	1	1	0	0	1
1	1	0	---	1	0	1	0	0	0
1	1	1	1	---	0	1	1	1	1
0	0	1	0	0	---	1	0	1	0
0	1	0	1	1	0	---	0	0	0
1	1	0	1	1	0	1	---	1	0
0	1	0	0	1	0	1	0	---	0
1	1	1	0	1	0	1	0	0	---

To be structurally equivalent, two actors must have the same ties to the same other actors -- that

is, the entries in their columns above must be identical (with the exception of self-ties, which we would probably ignore in most cases). Having arrayed the data in this way, there are some pretty obvious things that we could do to provide an index number to summarize how close to perfect structural equivalence each pair of actors are. With some imagination, you can come up with other ideas, but probably the four most common approaches to indexing pairwise structural equivalence are correlation, squared Euclidean distance, matches, and positive matches. Below, each of these results is shown as a "similarity" or "distance" matrix.

Pearson correlation coefficients

The correlation measure of similarity is particularly useful when the data on ties are "valued," that is, tell us about the strength, rather than simple presence or absence. Pearson correlations range from -1.00 (meaning that the two actors have exactly the opposite ties to each other actor), through zero (meaning that knowing one actor's tie to a third party doesn't help us at all in guessing what the other actor's tie to the third party might be), to +1.00 (meaning that the two actors always have exactly the same tie to other actors - perfect structural equivalence). Pearson correlations are often used to summarize pairwise structural equivalence because the statistic (called "little r") is widely used in social statistics. If the data on ties are truly nominal, or if density is very high or very low, correlations can sometimes be a little troublesome, and matches (see below) should also be examined. Different statistics, however, usually give very much the same answers. Here are the Pearson correlations among the concatenated row and column vectors, as calculated by UCINET.

	1	2	3	4	5	6	7	8	9	10
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
1	1.00									
2	0.42	1.00								
3	0.10	-0.48	1.00							
4	0.50	0.42	0.13	1.00						
5	0.33	0.79	-0.38	0.33	1.00					
6	-0.07	0.15	0.00	-0.07	0.00	1.00				
7	0.40	0.29	0.22	0.42	0.10	-0.37	1.00			
8	0.63	0.37	0.26	0.63	0.29	-0.00	0.05	1.00		
9	0.63	0.04	0.26	0.38	-0.10	-0.10	0.59	0.49	1.00	
10	0.26	0.29	0.40	0.52	0.25	0.36	-0.03	0.38	0.13	1.00

Note: Only one-half of the matrix is needed, as the similarity of X to Y is the same as the similarity of Y to X. There is quite a range of similarities in the matrix. Actors 2 and 5 are most similar ($r = .79$); actors 3 and 5 are most dissimilar ($r = -.38$). Most of the numbers are positive, and many are substantial. This result is consistent with (and a result of) the relatively high density and reciprocity in this particular data set. Where densities are very low, and ties are not reciprocated, correlations can be very small.

One way of getting an even clearer summary of the results is to perform a cluster analysis on the "similarity matrix." What this does is to group together the nodes that are most similar first (in

this case, actors #2 and #5). Taking the score in each cluster that is closest to some other cluster (single-link, or nearest neighbor method), similarities are re-calculated, and the next most similar pair are then joined (it could be two other actors, or the pair 2 and 5 with some other actor). This process continues until all the actors are joined together. Here is a summary of the similarity (or proximity) matrix above, generated by a single-link cluster analysis.

HIERARCHICAL CLUSTERING OF CORRELATION MATRIX

	5	2	4	1	8	7	9	6	3	0
Level										
0.787	XXX
0.630	XXX	.	XXX
0.595	XXX	.	XXX	XXX
0.587	XXX	XXXXXX	XXX
0.409	XXX	XXXXXXXXXX
0.405	XXX	XXXXXXXXXX	.	XXX
0.242	XXX	XXXXXXXXXX	XXXXX
0.140	XXXXXXXXXXXXXXXX	XXXXX
0.101	XXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXX

Note: This result says that actors #5 and #2 are most similar (at .787 using the particular method of clustering chosen here). This pair remains separate from all the others until quite late in the aggregation process. Next most similar are the pair #1 and #8 (at .630); then the pair #7 and #9 (at .595). At a similarity of .587, a larger group is created, composed of #4, #1, and #8; the process continues until all actors are joined. Note that actors #6, #3, and #10 are joined together into a (loose) group, and are rather distant from the remaining actors.

How many groups of structurally equivalent nodes are there in this network? There is no one "correct" answer. Theory may lead us to predict that information flows always produce three groups; or that groups will always divide into two factions, etc. If we have a theory of how many sets of structurally equivalent actors there ought to be, we can evaluate how well it fits the data. More often, we don't have a strong prior theory of how many sets of structurally equivalent actors there are likely to be. In this case, the results above can provide some guidance.

Note that, at a very high level of similarity (.787), our ten actors are partitioned into 9 groups (5 and 2, and each of the others). If we allow the members of groups to still be "equivalent" even if they are similar at only the .60 level (.587, actually), we get a pretty efficient picture: there are 6 groups among the 10 actors. If we drop back to similarities of about .40, there are three groups and one outlier (not surprisingly, our "near isolate" the WRO, #6). We could draw a picture of the number of groups (on the Y axis) against the level of similarity in joining (on the X axis). This diagram is actually a "marginal returns" kind of diagram (called a Scree plot in factor analysis). Its inflection point is the "most efficient" number of groups. More usually, one seeks a picture of the number of structurally equivalent groups that is simple enough to understand, while being statistically strong enough (having a relatively high similarity) to be defensible.

The Pearson correlation coefficient, by the way it is calculated, gives considerable weight to

large differences between particular scores in the profiles of actors (because it squares the difference in scores between the vectors). This can make the correlation coefficient somewhat sensitive to extreme values (in valued data) and to data errors. The Pearson correlation also measures only linear association, and, in some cases this may be too restrictive a notion.

Euclidean distances

A related, but somewhat less sensitive measure is the Euclidean distance. This measure is a measure of dis-similarity, in that it is the root of the sum the squared differences between the actor's vectors (that is, the columns of adjacencies shown above). In many cases, analyses of Euclidean distances and Pearson correlations give the same substantive answer. It is always a good idea to examine both, however.

Here are the results based on Euclidean distances.

EUCLIDEAN DISTANCE MATRIX

	1	2	3	4	5	6	7	8	9	10
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
1	0.00									
2	2.45	0.00								
3	2.65	3.32	0.00							
4	2.00	2.45	2.65	0.00						
5	2.65	1.00	3.16	2.65	0.00					
6	3.00	3.32	2.83	3.00	3.46	0.00				
7	2.24	2.24	2.45	2.24	2.45	3.46	0.00			
8	1.73	2.65	2.45	1.73	2.83	2.83	2.83	0.00		
9	1.73	3.00	2.45	2.24	3.16	2.83	2.00	2.00	0.00	
10	2.45	2.83	2.24	2.00	3.00	2.24	3.00	2.24	2.65	0.00

Note: The patterns are much the same. Actors #2 and #5, who have the strongest correlation (a measure of similarity) have the smallest distance (a measure of dissimilarity). Unlike the correlation coefficient, the sheer size of a Euclidean distance cannot be interpreted. All distances will be positive (or at least non-negative), and the ideas of "positive association" "no-association" and "negative association" that we have with the Pearson coefficient cannot be used with distances. Note that the range of values of the Euclidean distances is, relatively less than that of the correlations we saw before. That is, for example, the biggest distance is not as many times the smallest as the biggest correlation is to the smallest. This results from the way that distance statistic is calculated - giving less weight to extreme cases.

HIERARCHICAL CLUSTERING OF EUCLIDEAN DISTANCE MATRIX

```

Level   5 2 7 4 1 8 9 6 3 0
----- - - - - - - - - -
1.000   XXX . . . . .
1.732   XXX . . XXX . . .
1.821   XXX . XXXXXX . . .
1.992   XXX . XXXXXXXX . . .
2.228   XXX XXXXXXXXXXXX . . .
2.236   XXX XXXXXXXXXXXX . XXX
2.434   XXX XXXXXXXXXXXX XXXXX
2.641   XXX XXXXXXXXXXXXXXXXXXXX
3.014   XXXXXXXXXXXXXXXXXXXXXXXX
    
```

Note: in this case, the analysis of the Euclidean distances gives us exactly the same impression of which sets of actors are most structurally equivalent to one another. This is frequently the case with binary data. With valued data, the results from correlation and Euclidean distance analysis can sometimes be quite different. The graphic representation of the cluster analysis result clearly suggests three groupings of cases ({2,5}, {7,4,1,8,9}, {6,3,10}). It also suggests, again, that our "core" actors [5,2] have a high degree of structural equivalence - and are rather substitutable for one another.

Percent of Exact Matches

In some cases, the ties we are examining may be measured at the nominal level. That is, each pair of actors may have a tie of one of several types (perhaps coded as "a" "b" "c" or "1" "2" "3"). To apply Euclidean distance or correlation to such data would be misleading. Rather, we are interested in the degree to which the a tie for actor X is an exact match of that the corresponding tie of actor Y. With binary data the "percentage of exact matches" asks, "across all actors for which we can make comparisons, what percentage of the time do X and Y have the same tie with alter?" This is a very nice measure of structural similarity for binary data because it matches our notion of the meaning of structural equivalence very well.

Percent of Exact Matches

	1	2	3	4	5	6	7	8	9	10
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
1	1.00									
2	0.63	1.00								
3	0.56	0.31	1.00							
4	0.75	0.63	0.56	1.00						
5	0.56	0.94	0.38	0.56	1.00					
6	0.44	0.31	0.50	0.44	0.25	1.00				
7	0.69	0.69	0.63	0.69	0.63	0.25	1.00			
8	0.81	0.56	0.63	0.81	0.50	0.50	0.50	1.00		
9	0.81	0.44	0.63	0.69	0.38	0.50	0.75	0.75	1.00	
10	0.63	0.50	0.69	0.75	0.44	0.69	0.44	0.69	0.56	1.00

Note: These results show similarity in another way, but one that is quite easy to interpret. The number .63 in the cell 2,1 means that, in comparing actor #1 and #2, they have the same tie (present or absent) to other actors 63% of the time. The measure is particularly useful with multi-category nominal measures of ties; it also provides a nice scaling for binary data.

HIERARCHICAL CLUSTERING OF EXACT MATCH MATRIX

											1
Level	5	2	4	1	8	7	9	6	3	0	
0.938	XXX
0.813	XXX	.	XXX
0.792	XXX	XXXXXX
0.750	XXX	XXXXXX	XXX
0.698	XXX	XXXXXXXXXX
0.688	XXX	XXXXXXXXXX	.	XXX
0.625	XXX	XXXXXXXXXX	XXXXX
0.554	XXX	XXXXXXXXXXXXXXXXXX
0.432	XXXXXXXXXXXXXXXXXXXX

Note: In this case, the results using matches are very similar to correlations or distances (as they often are, in practice). The clustering does suggest, however, that the set [2,5] are quite different from all of the other actors. Correlations and Euclidean distances tended to emphasize the difference of [6,3,10] and to downplay somewhat the uniqueness of [2,5].

Jaccard coefficients

In some networks connections are very sparse. Indeed, if one were looking at ties of personal acquaintance in very large organizations, the data might have very low density. Where density is very low, the "matches" "correlation" and "distance" measures can all show relatively little variation among the actors, and may cause difficulty in discerning structural equivalence sets (of course, in very large, low density networks, there may really be very low levels of structural equivalence).

One approach to solving this problem is to calculate the number of times that both actors report a tie (or the same type of tie) to the same third actors as a percentage of the total number of ties reported. That is, we ignore cases where neither X or Y are tied to Z, and ask, of the total ties that are present, what percentage are in common. This measure is called the "percent of positive matches" (by UCINET), or the "Jaccard coefficient" (by, for example, SPSS). Here are the result of using it on our data:

Percent of Positive Matches (Jaccard coefficients)

	1	2	3	4	5	6	7	8	9	10
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
1	1.00									
2	0.54	1.00								
3	0.46	0.31	1.00							
4	0.60	0.54	0.42	1.00						
5	0.50	0.93	0.38	0.50	1.00					
6	0.18	0.27	0.11	0.18	0.25	1.00				
7	0.58	0.64	0.54	0.55	0.60	0.08	1.00			
8	0.67	0.46	0.50	0.67	0.43	0.20	0.38	1.00		
9	0.67	0.36	0.50	0.55	0.33	0.11	0.64	0.56	1.00	
10	0.40	0.43	0.44	0.60	0.36	0.38	0.31	0.50	0.36	1.00

HIERARCHICAL CLUSTERING OF JACCARD COEFFICIENT MATRIX

Level	6	2	5	3	4	1	8	7	9	0
0.929	.	XXX
0.667	.	XXX	.	.	XXX
0.644	.	XXX	.	XXXXXX
0.636	.	XXX	.	XXXXXX	XXX
0.545	.	XXX	.	XXXXXXXXXX
0.497	.	XXX	.	XXXXXXXXXXXXXX
0.435	.	XXXXXXXXXXXXXXXXXXXXXX
0.403	.	XXXXXXXXXXXXXXXXXXXXXX
0.272	XXXXXXXXXXXXXXXXXXXXXX

Note: Again the same basic picture emerges, suggesting that exactly which statistic one uses to index pairwise structural similarity of actors may not matter too much -- at least for painting a broad picture of things. The clustering of these distances this time emphasizes the uniqueness of actor #6. Actor six is more unique by this measure because of the relatively small number of total ties that it has -- this results in a lower level of similarity when "joint absence" of ties are ignored. Where data are sparse, and where there very substantial differences in the degrees of points, the positive match coefficient is a good choice for binary or nominal data.

With some inventiveness, you can probably think of some other reasonable ways of indexing the degree of structural similarity between actors. You might look at the program "Proximities" by SPSSx, which offers a large collection (and a good discussion about) measures of similarity. The choice of a measure should be driven by a conceptual notion of "what about" the similarity of two tie profiles is most important for the purposes of a particular analysis. Often, frankly, it makes little difference, but that is hardly sufficient grounds to ignore the question.

Describing Structural Equivalence Sets: Block Models and Images

The approaches that we have examined thus far seek to provide summaries of how similar or different each pair of actors are, using the yardstick of structural equivalence. These "similarity" or "proximity" (or sometimes the opposite: "distance") matrices provide a complete pair-wise description of this aspect of network positions. But, it is very difficult to see the forest for the trees -- distance or similarity matrices are almost as big and dense as the original adjacency or tie matrices from which they were taken.

One very useful approach to getting the big picture is to apply cluster analysis to attempt to discern how many structural equivalence sets there are, and which actors fall within each set. We have seen several examples above of how cluster analysis can be used to present a simpler or summary picture of patterns of similarity or dissimilarity in the structural positions of actors. Cluster analysis (and there are many flavors of this technique) is strongly grounded in pairwise and hierarchical comparisons. Cluster analysis also assumes a unidimensionality underlying the similarities data. It is often helpful to consider some other approaches that have "different biases." One such approach is multidimensional scaling, but we will not discuss it here -- as it is rarely used for making sense of structural equivalence data. We will examine three more common approaches -- CONCOR, principle components analysis, and tabu search.

What the similarity matrix and cluster analysis do not tell us is what similarities make the actors in each set "the same" and which differences make the actors in one set "different" from the actors in another. A very useful approach to understanding the bases of similarity and difference among sets of structurally equivalent actors is the block model, and a summary based on it called the image matrix. Both of these ideas have been explained elsewhere. We will take a look at how they can help us to understand the results of CONCOR and tabu search.

CONCOR

CONCOR is an approach that has been used for quite some time. Although the algorithm (that is what calculations it does in what order) of concor is now regarded as a bit peculiar, the technique usually produces meaningful results. CONCOR begins by correlating each pair of actors (as we did above). Each row of this actor-by-actor correlation matrix is then extracted, and correlated with each other row. In a sense, the approach is asking "how similar is the vector of similarities of actor X to the vector of similarities of actor Y?" This process is repeated over and over. Eventually the elements in this "iterated correlation matrix" converge on a value of either +1 or -1 (if you want to convince yourself, give it a try!).

CONCOR then divides the data into two sets on the basis of these correlations. Then, within each set (if it has more than two actors) the process is repeated. The process continues until all actors are separated (or until we lose interest). The result is a binary branching tree that gives rise to a final partition.

For illustration, we have asked CONCOR to show us the groups that best satisfy this property

when we believe that there are four groups. All blocking algorithms require that we have a prior idea about how many groups there are.

Blocked Matrix:

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

Note: The blocked matrix has rearranged the rows and columns (and inserted dividing lines or "partitions") to try to put actors with similar rows and columns together (given the constraint of forcing four groups). The blocked matrix allows us to see the similarities that define the four groups (and also where the groupings are less than perfect). The first group [1,4,8,9] have no consistent pattern of sending or receiving information among themselves -- they are not a clique. The relationship of this group [1,4,8,9] to the next group [7,3] is interesting. All actors in the first group send information to 7 but not to 3. Only one actor in the group (1) receives information from [7,3], and gets it from both members. Each member of the set [1,4,8,9] both sends and receives information from both members of [5,2]. Finally, the set [1,4,8,9] are similar in that they do not send or get information from [6,10]. Actors 5 and 2 have nearly identical profiles of reciprocal communication with every member of every other group, with the exception of [6,10]. Finally, [6,10] are largely isolated, but send to both of [7,3]; they also share the fact that only [3] reciprocates their attentions.

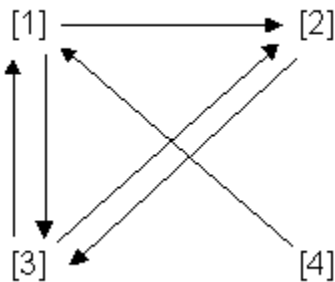
The blocked diagram lets us see pretty clearly which actors are structurally equivalent, and, more importantly, what about the patterns of their ties define the similarity. In some cases we may wish to summarize the information still further by creating an "image" of the blocked diagram. To do this, we combine the elements on the rows and columns into groups, and characterize the

relations in the matrix as "present" if they display more than average density (remember that our overall density is about .50 in this matrix). If fewer ties are present, a zero is entered. The image of the blocking above is:

	[1]	[2]	[3]	[4]
[1]	0	1	1	0
[2]	0	0	1	0
[3]	1	1	0	0
[4]	0	1	0	0

Note: This "image" of the blocked diagram has the virtue of extreme simplicity, though a good bit of information has been lost about the relations among particular actors. It says that the actors in block 1 tend to not send and receive to each other or to block 4; they tend to send information to blocks 2 and 3, but receive information only from 3. The second block of actors appears to be "receivers." They send only to block 3, but receive information from each of the other groups. The third block of actors send and receive to blocks one and two, but not block 4. Lastly, as before, block 4 actors send to block 2 actors, but do not receive information from any other block.

One can carry the process of simplification of the presentation one step further still. Using partitioned sets of actors as nodes, and the image as a set of adjacencies, we can return to a graph to present the result.



Note: The much simplified graph suggests the marginality of [4], the reciprocity and centrality of the relations of [3], and the asymmetry between [1] and [2].

We've gone on rather long in presenting the results of the CONCOR. This is really to illustrate the utility of permuting and blocking, calculating the image, and graphing the result. CONCOR

can generate peculiar results that are not reproduced by other methods. If you use CONCOR, it is wise to cross-check your results with other algorithms, as well. One very good cross-check is the tabu search (which can also generate peculiar results, and also needs cross-checking).

The goodness of fit of a block model can be assessed by correlating the permuted matrix (the block model) against a "perfect" model with the same blocks (i.e. one in which all elements of one blocks are ones, and all elements of zero blocks are zeros). For the CONCOR two-split (four group) model, this r-squared is .50. That is, about 1/2 of the variance in the ties in the CONCOR model can be accounted for by a "perfect" structural block model. This might be regarded as OK, but is hardly a wonderful fit (there is no real criterion for what is a good fit).

Tabu Search

This method of blocking has been developed more recently, and relies on extensive use of the computer. Tabu search uses a more modern (and computer intensive) algorithm than CONCOR, but is trying to implement the same idea of grouping together actors who are most similar into a block. Tabu search does this by searching for sets of actors who, if placed into a blocks, produce the smallest sum of within-block variances in the tie profiles. That is, if two actors have similar ties, their variance around the block mean profile will be small. So, the partitioning that minimizes the sum of within block variances is minimizing the overall variance in tie profiles. In principle, this method ought to produce results similar (but not necessarily identical) to CONCOR. In practice, this is not always so. Below are results from this method where we have specified 3 groups.

Blocked Adjacency Matrix

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

Note: The blocking produced is really very similar to the CONCOR result. One set of actors [1,4,8,9] is identical; the CONCOR block [7,3] is divided in the new results, with actor 7 being grouped with [5,2] and actor 3 being grouped with [6,10]. Looking at the permuted and blocked adjacency matrix, we can see that big regions of the matrix do indeed now have minimum variance -- all entries are the same. This is what this blocking is really trying to do: produce regions of homogeneity of scores that are as large as possible. We could interpret these results directly, or we could produce the block image:

	[1]	[2]	[3]
[1]	1	0	1
[2]	1	1	0
[3]	1	0	0

Note: This image suggests that two of the three blocks are somewhat "solidaristic" in that the members send and receive information among themselves (blocks [1] and [2]). The actors in block 3 are similar because they don't send information with one another directly. What we are seeing here is something of a hierarchy between blocks one and two, with block 3 being somewhat more isolated. These patterns can also be represented in a graph.



The goodness of fit of a block model can be assessed by correlating the permuted matrix (the block model) against a "perfect" model with the same blocks (i.e. one in which all elements of one blocks are ones, and all elements of zero blocks are zeros). For the tabu search model for three groups, this r-squared is .47. This is not great. We do note that it is very nearly as good a fit as the CONCOR model with one additional group. Of course, one could fit models for various numbers of groups, and scree plot the r-squared statistics to try to decide which block model was optimal.

Factor (Principal components, actually) Analysis

All approaches to structural equivalence focus on the similarity in the pattern of ties of each actor with each other. If patterns of ties are very similar, then actors are very equivalent in the structural sense.

The cluster analysis, tabu search, and CONCOR approaches to the matrix of similarities among actors share the commonality that they assume that a singular or global similarity along a single dimension is sought. However, there is no guarantee that the matrix of similarities among the tie

profiles of actors is unidimensional. Factor analysis and multi-dimensional scaling of the similarity (or distance) matrix relaxes this assumption. In a sense, these scaling methods suggest that there may be more than one "aspect" or "form" or "fundamental pattern" of structural equivalence underlying the observed similarities -- and that actors who are "structurally similar" in one regard (or dimension) may be dissimilar in another.

UCINET's implementation of the principal components approach searches for similarity in the profiles of distances from each actor to others. The other approaches we looked at above examined both ties from and to actors in directed data. Of course, factor analysis may be applied to any symmetric rectangular matrix, so the UCINET's choice is only one of many possible approaches.

Correlations between pairs of rows of the distance matrix are calculated. Then a principal components analysis is performed, and loadings are rotated (UCINET's documentation does not describe the method used to decide how many factors to retain; the rotation method is also not described). Here is (part of) the output from UCINET's CATIJ routine.

CATIJ

Type of data: ADJACENCY
 Loadings cutoff: 0.60

CATIJ Matrix (length of optimal paths between nodes)

		1	2	3	4	5	6	7	8	9	0
	COCOEDINMAWRNEUWWEWE	-	-	-	-	-	-	-	-	-	-
1	COUN	0	1	2	2	1	3	1	2	1	2
2	COMM	1	0	1	1	1	2	1	1	1	2
3	EDUC	2	1	0	1	1	1	1	2	2	1
4	INDU	1	1	2	0	1	3	1	2	2	2
5	MAYR	1	1	1	1	0	2	1	1	1	1
6	WRO	3	2	1	2	2	0	1	3	1	2
7	NEWS	2	1	2	1	1	3	0	2	2	2
8	UWAY	1	1	2	1	1	3	1	0	1	2
9	WELF	2	1	2	2	1	3	1	2	0	2
10	WEST	1	1	1	2	1	2	1	2	2	0

Correlations Among Rows of CATIJ Matrix

	1	2	3	4	5	6	7	8	9	10
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
1 COUN	1.00	0.65	-0.10	0.88	0.71	0.38	0.71	0.89	1.00	0.54
2 COMM	0.65	1.00	-0.45	0.65	0.65	0.06	0.60	0.82	0.60	0.38
3 EDUC	-0.10	-0.45	1.00	0.05	-0.29	0.37	0.29	-0.31	0.22	0.26
4 INDU	0.88	0.65	0.05	1.00	0.71	-0.16	0.87	0.89	0.88	0.79
5 MAYR	0.71	0.65	-0.29	0.71	1.00	0.00	0.71	0.80	0.71	0.38
6 WRO	0.38	0.06	0.37	-0.16	0.00	1.00	-0.00	-0.22	0.37	0.12
7 NEWS	0.71	0.60	0.29	0.87	0.71	-0.00	1.00	0.80	0.87	0.38
8 UWAY	0.89	0.82	-0.31	0.89	0.80	-0.22	0.80	1.00	0.80	0.32
9 WELF	1.00	0.60	0.22	0.88	0.71	0.37	0.87	0.80	1.00	0.68
10 WEST	0.54	0.38	0.26	0.79	0.38	0.12	0.38	0.32	0.68	1.00

Rotated Factor Loadings of Actors on Groups

	1	2	3
1 COUN	-0.90	-0.17	-0.37
2 COMM	-0.71	-0.53	-0.10
3 EDUC	-0.04	0.96	-0.19
4 INDU	-1.00	0.07	0.17
5 MAYR	-0.79	-0.36	-0.02
6 WRO	-0.02	0.16	-1.00
7 NEWS	-0.90	0.14	0.04
8 UWAY	-0.90	-0.38	0.14
9 WELF	-0.95	0.12	-0.33
10 WEST	-0.68	0.37	-0.06

Group 1:
 COUN COMM INDU MAYR NEWS UWAY WELF WEST
 Group 2:
 EDUC
 Group 3:
 WRO

Notes: The factor analysis identifies one clear strong dimension, and two additional dimensions that are essentially "item" factors. The second and third factors are due to uniqueness in the patterns of EDUC's distances and uniqueness in the patterns of WRO's distances. These distances cannot be accounted for by the same underlying pattern as the others -- which suggests that these two organizations cannot be ranked along the same dimensions of "similarity" as the others.

This result is quite unlike that of the blocking methods. In part this is due to the use of distances, rather than adjacencies. The more important reason for the difference, however, is that the factoring (and multi-dimensional scaling) methods are searching for dimensionality in the data, rather than assuming unidimensionality. The result above suggests that organizations 3 and 6 are not structurally equivalent to the other organizations -- and that their ties may differ in "meaning" or function from the ties among the other organizations. In a sense, the result suggests that

"information exchange" for these two organizations may be a different kind of tie than it is for the other organizations.

If we look at the locations of the nodes along the first dimension, it can be seen that the organizations are, in fact ranked in a way similar to those of our other results. One might imagine a grouping {4,9,1,7,8}, {5}, {2,10}, {3,6} by selecting cut points along the first dimension. This grouping is rather similar (but not identical) to that found in other analyses here.

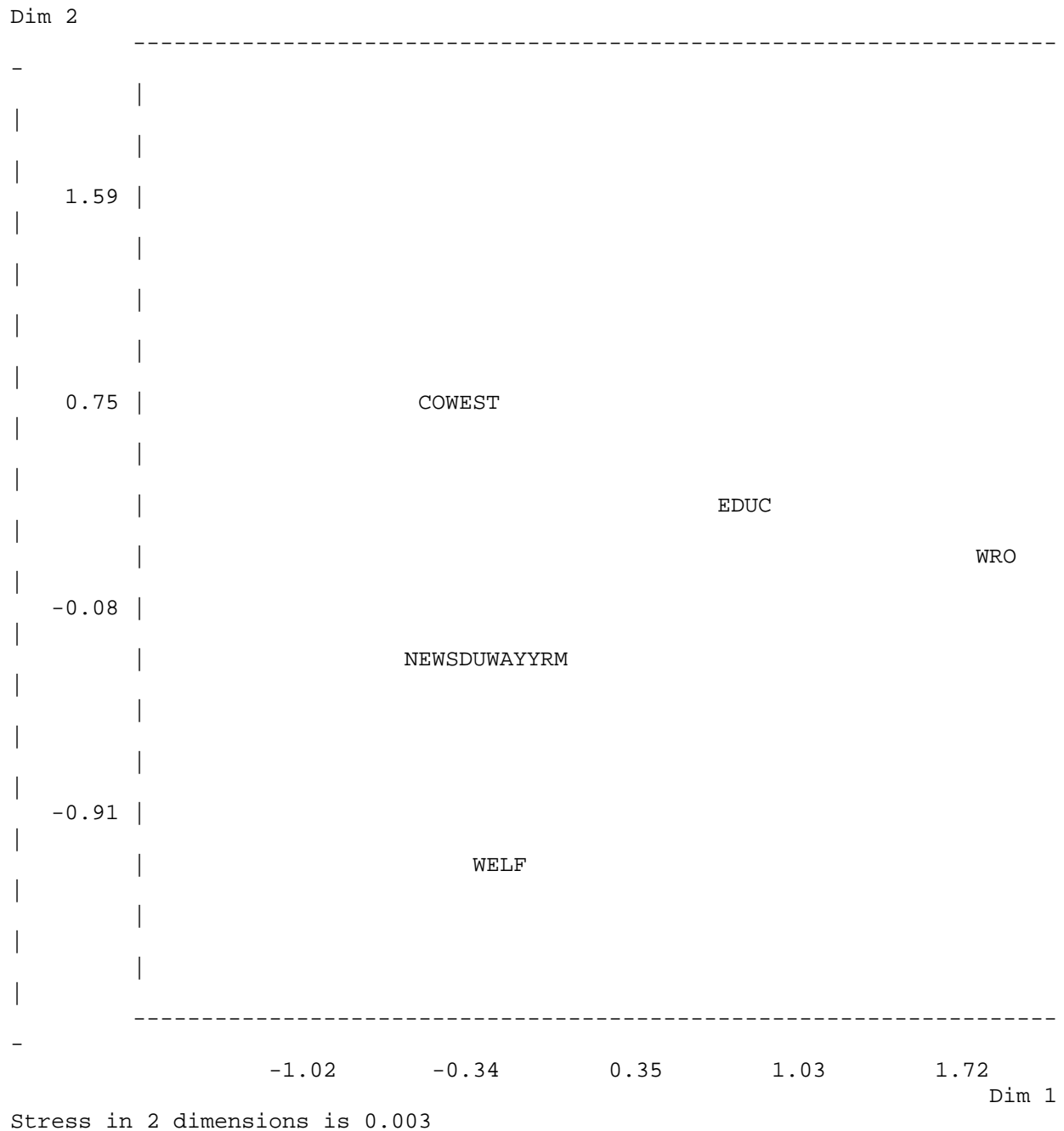
This use of factor analysis to "purify" the grouping by extracting and removing secondary dimensions is suggestive. However, like most other uses of factor analysis, it should be used thoughtfully. Results may be very similar, or very different to those of unidimensional scaling approaches. As usual, neither is inherently more "right."

Multi-Dimensional Scaling

A variation on the same theme as the CATIJ algorithm is the application of multi-dimensional scaling. MDS, like factor and cluster analysis, is actually a large family of somewhat diverse techniques. Perhaps most usefully, MDS represents the patterns of similarity or dissimilarity in the tie profiles among the actors (when applied to adjacency or distances) as a "map" in multi-dimensional space. This map lets us see how "close" actors are, whether they "cluster" in multi-dimensional space, and how much variation there is along each dimension. Here are the results from UCINET's non-metric routine applied to generate a two-dimensional map of the adjacency matrix.

NON-METRIC MULTIDIMENSIONAL SCALING

	1	2
1 COUN	-0.49	0.94
2 COMM	-0.12	-0.24
3 EDUC	0.77	0.59
4 INDU	-0.42	-0.26
5 MAYR	-0.14	-0.24
6 WRO	1.91	0.23
7 NEWS	-0.58	-0.27
8 UWAY	-0.25	-0.25
9 WELF	-0.27	-1.47
10 WEST	-0.40	0.96



Notes: The stress in two dimensions is very low (that is, the goodness-of-fit is high). This suggests that the third dimension identified by the factor analysis previously may not have really been necessary. EDUC and WRO are again clearly identified as identifying a different dimension of the similarities (in this case, dim. 1).

The clustering of cases along the main dimension (here, dim. 2) does not correspond very closely

to either the blocking or factoring methods {1,10}, {2,4,5,7,8}, {9}. There are many reasons why this might occur (use of Euclidean distances instead of correlations, use of two rather than three dimensions, etc.). Again, it is not that one solution is correct and the others not. The MDS approach is showing us something different and additional about the approximate structural equivalence of these points.

Summary of chapter 9

In this section we have discussed the idea of "structural equivalence" of actors, and seen some of the methodologies that are most commonly used to measure structural equivalence, find patterns in empirical data, and describe the sets of "substitutable" actors.

Structural equivalence of two actors is the degree to which the two actors have the same profile of relations across alters (all other actors in the network). Exact structural equivalence is rare in most social structures (one interpretation of exact structural equivalence is that it represents systematic redundancy of actors; which may be functional in some way to the network).

While it is sometimes possible to see patterns of structural equivalence "by eye" from an adjacency matrix or diagram, we almost always use numerical methods. Numerical methods allow us to deal with multiplex data, large numbers of actors, and valued data (as well as the binary type that we have examined here).

The first step in examining structural equivalence is to produce a "similarity" or a "distance" matrix for all pairs of actors. This matrix summarizes the overall similarity (or dissimilarity) of each pair of actors in terms of their ties to alters. While there are many ways of calculating such index numbers, the most common are the Pearson Correlation, the Euclidean Distance, the proportion of matches (for binary data), and the proportion of positive matches (Jaccard coefficient, also for binary data).

A number of methods may be used to identify patterns in the similarity or distance matrix, and to describe those patterns. Cluster analysis (generally, there are many methods) groups together the two most similar actors, recalculates similarities, and iterates until all actors are combined. What is produced is a "joining sequence" or map of which actors fall into a hierarchy of increasingly inclusive (and hence less exactly equivalent) groups. Multi-dimensional scaling and factor analysis can be used to identify what aspects of the tie profiles are most critical to making actors similar or different, and can also be used to identify groups. Groupings of structurally equivalent actors can also be identified by the divisive method of iterating the correlation matrix of actors (CONCOR), and by the direct method of permutation and search for perfect zero and one blocks in the adjacency matrix (Tabu search).

Once the number of groupings that are useful has been determined, the data can be permuted and blocked, and images calculated. These techniques enable us to get a rather clear picture of how the actors in one set are "approximately equivalent" and why different sets of actors are different. That is, they enable us to describe the meaning of the groups and the place of group members in the overall network in a general way.

Structural equivalence analysis often produces interesting and revealing findings about the

patterns of ties and connections among the individual actors in a network. The structural equivalence concept aims to operationalize the notion that actors may have identical or nearly identical positions in a network -- and hence be directly "substitutable" for one another. An alternative interpretation is that actors who are structurally equivalent face nearly the same matrix of constraints and opportunities in their social relationships.

Sociological analysis is not really about individual people. And, structural analysis, is primarily concerned with the more general and abstract idea of the roles or positions that define the structure of the group -- rather than the locations of specific actors with regard to specific others. For such analysis, we turn to a related set of tools for studying replicate sub structures ("automorphic equivalence") and social roles ("regular equivalence").

10. Automorphic Equivalence

Defining automorphic equivalence

Automorphic equivalence is not as demanding a definition of similarity as structural equivalence, but is more demanding than regular equivalence. There is a hierarchy of the three equivalence concepts: any set of structural equivalencies are also automorphic and regular equivalencies. Any set of automorphic equivalencies are also regular equivalencies. But, not all regular equivalencies are necessarily automorphic or structural; and not all automorphic equivalencies are necessarily structural.

Formally "Two vertices u and v of a labeled graph G are automorphically equivalent if all the vertices can be re-labeled to form an isomorphic graph with the labels of u and v interchanged. Two automorphically equivalent vertices share exactly the same label-independent properties." (Borgatti, Everett, and Freeman, 1996: 119).

More intuitively, actors are automorphically equivalent if we can permute the graph in such a way that exchanging the two actors has no effect on the distances among all actors in the graph. If we want to assess whether two actors are automorphically equivalent, we first imagine exchanging their positions in the network. Then, we look and see if, by changing some other actors as well, we can create a graph in which all of the actors are the same distance that they were from one another in the original graph.

In the case of structural equivalence, two actors are equivalent if we can exchange them one-for-one, and not affect any properties of the graph. Automorphically equivalent actors are actors that can be exchanged with no effect on the graph -- given that other actors are also moved. If the concept is still a bit difficult to grasp at this point, don't worry. Read on, and then come back after you've looked at a few examples.

Uses of the concept

Structural equivalence focuses our attention on pairwise comparisons of actors. By trying to find actors who can be swapped for each other, we are really paying attention to the positions of the actors in a particular network. We are trying to find actors who are clones or substitutes.

Automorphic equivalence begins to change the focus of our attention, moving us away from concern with individuals network positions, and toward a more abstracted view of the network. Automorphic equivalence asks if the whole network can be re-arranged, putting different actors at different nodes, but leaving the relational structure or skeleton of the network intact.

Suppose that we had 10 workers in the University Avenue McDonald's restaurant, who report to one manager. The manager, in turn, reports to a franchise owner. The franchise owner also controls the Park Street McDonald's restaurant. It too has a manager and 8 workers. Now, if the owner decided to transfer the manager from University Avenue to the Park Street restaurant (and

vice versa), the network has been disrupted. But if the owner transfers both the managers and the workers to the other restaurant, all of the network relations remain intact. Transferring both the workers and the managers is a permutation of the graph that leaves all of the distances among the pairs of actors exactly as it was before the transfer. In a sense, the "staff" of one restaurant is equivalent to the staff of the other, though the individual persons are not substitutable.

The hypothetical example of the restaurants actually does suggest the main utility of the automorphic equivalence concept. Rather than asking what individuals might be exchanged without modifying the social relations described by a graph (structural equivalence), the somewhat more relaxed concept of automorphic equivalence focuses our attention on sets of actors who are substitutable as sub-graphs, in relation to other sub-graphs. In many social structures, there may well be sub-structures that are equivalent to one another. The number, type, and relations among such sub-structures might be quite interesting. Many structures that look very large and complex may actually be composed (at least partially) of multiple identical sub-structures; these sub-structures may be "substitutable" for one another. Indeed, a McDonalds is a McDonalds is a McDonalds...

Finding Equivalence Sets

In principle, the automorphisms in a graph can be identified by the *brute force method* of examining every possible permutation of the graph. With a small graph, and a fast computer, this is a useful thing to do. Basically, every possible permutation of the graph is examined to see if it has the same tie structure as the original graph. For graphs of more than a few actors, the number of permutations that need to be compared becomes extremely large. For the graphs of many "real" networks (not the examples dreamed up by theorists), there may well be no exact automorphisms. For complicated graphs, and particularly directed or valued graphs, the amount of computational power necessary can also be rather daunting, and it can almost be assured that there will be few or any exact equivalencies.

As with structural equivalence, it may be very useful to identify "approximate" automorphisms. Approximate automorphisms can arise when the relations are measured with error, when sampling variability has given us a view of the network that is incomplete, or when the relations generating the network are not in equilibrium (i.e. processes like transitivity and balance building are not "complete" at the time we collect the data). Or, of course, it may be the case that a given network truly contains sets of actors that are similar to one another, but not exactly equivalent. UCINET provides several approaches to identifying sets of actors that are approximately automorphically equivalent.

Geodesic equivalence focuses on similarity in the profiles of actor's geodesic distances to other actors. The idea can be applied to directed or undirected data, and to binary or valued data. First, the geodesic distance matrix is calculated. Second, each actor's geodesic distance vector is extracted and the elements are ranked. Third, the Euclidean distance measure of dissimilarity between these sorted vectors are calculated. The resulting distance matrix summarizes the pairwise dissimilarity in the profile of actor's geodesic distances. The key element here is that it is the profile of geodesic distances, and not the geodesic distances to the same targets that are compared between two actors. So long as two actors have the same mix of geodesic distances,

they are equivalent.

Once a matrix of dissimilarities has been generated, cluster analysis or dimensional scaling can be used to identify approximately equivalent partitions of actors.

The *maxsim* algorithm in UCINET is an extension of the geodesic equivalence approach that is particularly useful for valued and directed data. The algorithm begins with a distance matrix (or a concatenation of both "distance from" and "distance to" for directed data). The distances of each actor are sorted from low to high, and the Euclidean distance is used to calculate the dissimilarity between the distance profiles of each pair of actors. The algorithm scores actors who have similar distance profiles as more automorphically equivalent. Again, the focus is on whether actor *u* has a similar set of distances, regardless of which distances, to actor *v*. Again, dimensional scaling or clustering of the distances can be used to identify sets of approximately automorphically equivalent actors.

Tabu search is a numerical method for finding the best division of actors into a given number of partitions on the basis of approximate automorphic equivalence. In using this method, it is important to explore a range of possible numbers of partitions (unless one has a prior theory about this), to determine how many partitions are useful. Having selected a number of partitions, it is useful to re-run the algorithm a number of times to insure that a global, rather than local minimum has been found.

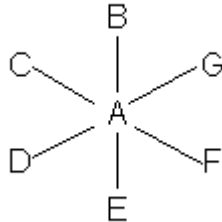
The method begins by randomly allocating nodes to partitions. A measure of badness of fit is constructed by calculating the sums of squares for each row and each column within each block, and calculating the variance of these sums of squares. These variances are then summed across the blocks to construct a measure of badness of fit. Search continues to find an allocation of actors to partitions that minimizes this badness of fit statistic.

What is being minimized is a function of the dissimilarity of the variance of scores within partitions. That is, the algorithm seeks to group together actors who have similar amounts of variability in their row and column scores within blocks. Actors who have similar variability probably have similar profiles of ties sent and received within, and across blocks -- though they do not necessarily have the same ties to the same other actors.

Unlike the other methods mentioned here, the Tabu search produces a partitioned matrix, rather than a matrix of dissimilarities. It also provides an overall badness of fit statistic. Both of these would seem to recommend the approach, perhaps combined with other methods.

Some Examples

The Star Network analyzed by geodesic equivalence



We know that the partition $\{A\}\{B,C,D,E,F,G\}$ defines structurally equivalent sets. Therefore, this partition is also an automorphic partition. Even though this result is obvious, it will serve as an example to help understand algorithms for identifying automorphic equivalences. Here is the output from UCINET's Geodesic equivalence algorithm for identifying automorphic equivalence sets.

```

GEODESIC EQUIVALENCE
Node by pathlength frequency matrix
      1      2
---- ----
1      6      0
2      1      5
3      1      5
4      1      5
5      1      5
6      1      5
7      1      5
  
```

note: The actors are listed on the rows, the number of their geodesics of various lengths on the columns

```

Geodesic Equivalence Matrix (dissimilarities)
      1      2      3      4      5      6      7
---- ---- ---- ---- ---- ---- ----
1  0.00  7.07  7.07  7.07  7.07  7.07  7.07
2  7.07  0.00  0.00  0.00  0.00  0.00  0.00
3  7.07  0.00  0.00  0.00  0.00  0.00  0.00
4  7.07  0.00  0.00  0.00  0.00  0.00  0.00
5  7.07  0.00  0.00  0.00  0.00  0.00  0.00
6  7.07  0.00  0.00  0.00  0.00  0.00  0.00
7  7.07  0.00  0.00  0.00  0.00  0.00  0.00
  
```

note: the dissimilarities are measured as Euclidean distances

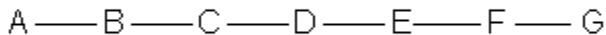
HIERARCHICAL CLUSTERING OF (NON-)EQUIVALENCE MATRIX

```
Level   1 2 3 4 5 6 7
----- - - - - - - -
0.000   . XXXXXXXXXXXXX
7.071   XXXXXXXXXXXXXXX
```

note: The clustering of the distances the clear division into two sets (The center and the periphery).

The Line network analyzed by Maxsim

The line network is interesting because of the differing centralities and betweenness of the actors.



Here is the output from UCINET's maxsim algorithm (node 1 = "A", node 2 = "B", etc.)

AUTOMORPHIC EQUIVALENCE VIA MAXSIM

NOTE: Binary adjacency matrix converted to reciprocals of geodesic distances.

note: maxsim is most useful for valued data, so reciprocals of distances rather than adjacency was analyzed.

Distances Among Actors

	1	2	3	4	5	6	7
1	0.00	3.22	3.62	3.73	3.62	3.22	0.00
2	3.22	0.00	1.16	1.37	1.16	0.00	3.22
3	3.62	1.16	0.00	0.50	0.00	1.16	3.62
4	3.73	1.37	0.50	0.00	0.50	1.37	3.73
5	3.62	1.16	0.00	0.50	0.00	1.16	3.62
6	3.22	0.00	1.16	1.37	1.16	0.00	3.22
7	0.00	3.22	3.62	3.73	3.62	3.22	0.00

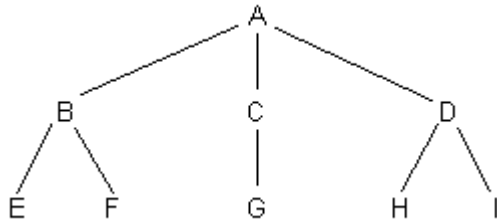
HIERARCHICAL CLUSTERING OF (NON-)EQUIVALENCE MATRIX

```
Level   4 5 3 2 6 1 7
----- - - - - - - -
0.000   . XXX XXX XXX
0.500   XXXXX XXX XXX
1.212   XXXXXXXXXXX XXX
3.383   XXXXXXXXXXXXXXX
```

notes: The clustering first separates the set of the two "end" actors as equivalent, then the next inner pair, the next inner pair, and the center. You should be convinced yourself that this is a valid (indeed, an exact) automorphism for the graph.

The Wasserman-Faust network analyzed by all permutations search

The graph presented by Wasserman and Faust is an ideal one for illustrating the differences among structural, automorphic, and regular equivalence. Before looking at the results, see if you can write down what the automorphic equivalence positions in this network.



Here is the output from UCINET's examination of all permutations of this graph.

AUTOMORPHIC EQUIVALENCE VIA DIRECT SEARCH

Number of permutations examined: 362880

Number of automorphisms found: 8

Hit rate: 0.00%

ORBITS:

Orbit #1: 1 (i.e. actor A)

Orbit #2: 2 4 (i.e. actors B and D)

Orbit #3: 3 (i.e. actor C)

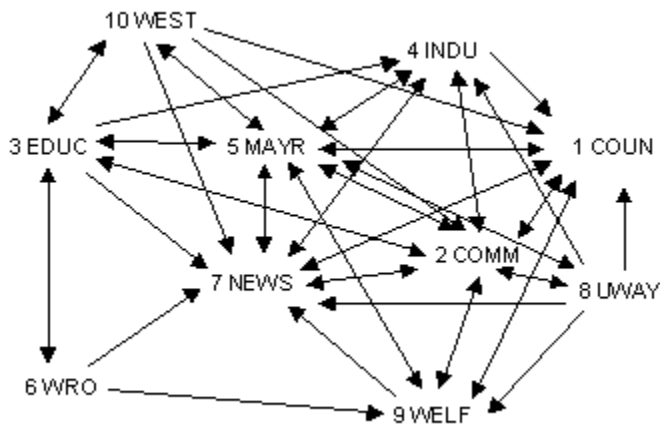
Orbit #4: 5 6 8 9 (i.e. actors E, F, H, and I)

Orbit #5: 7 (i.e. actor G)

notes: The algorithm examined over three hundred thousand possible permutations of the graph. We see that the two branches (B, E, F and D, H, I) are "switchable" as whole sub-structures.

The Knoke bureaucracies information exchange network analyzed by TABU Search

Now we turn our attention to some more complex data. In the Knoke information data there are no exact automorphisms. This is not really surprising, given the complexity of the pattern (and particularly if we distinguish ties in from ties out) of connections.



Like structural equivalence, however, it may be useful to examine approximate automorphic equivalence. One useful tool is the TABU search algorithm. We select a number of partitions to evaluate, and the program seeks to find the minimum error grouping of cases.

Here are the results of analyzing the Knoke data with UCINET's TABU search algorithm. First we examine the badness of fit of different numbers of partitions for the data.

Partitions	Fit
1	11.970
2	21.132
3	16.780
4	15.965
5	14.465
6	13.563
7	10.367
8	9.521
9	1.500
10	0.000

notes: There is no "right" answer about how many automorphisms there are here. There are two trivial answers: those that group all the cases together into one partition and those that separate each case into it's own partition. In between, one might want to follow the logic of the "scree" plot from factor analysis to select a meaningful number of partitions. Look first at the results for three partitions:

```

AUTOMORPHIC BLOCKMODELS VIA TABU SEARCH
Diagonal valid?          NO
Use geodesics?          YES
Fit: 16.780
Block Assignments:
  1:  COUN COMM EDUC INDU NEWS UWAY WELF WEST
  2:  WRO
  3:  MAYR
Blocked Distance Matrix
          1
          1 2 3 4 0 8 7 9  6  5
          C C E I W U N W  W  M
-----
 1 COUN |   1 2 2 2 2 1 1 | 3 | 1 |
 2 COMM |  1  1 1 2 1 1 1 | 2 | 1 |
 3 EDUC |  2 1  1 1 2 1 2 | 1 | 1 |
 4 INDU |  1 1 2  2 2 1 2 | 3 | 1 |
10 WEST |  1 1 1 2  2 1 2 | 2 | 1 |
 8 UWAY |  1 1 2 1 2  1 1 | 3 | 1 |
 7 NEWS |  2 1 2 1 2 2  2 | 3 | 1 |
 9 WELF |  2 1 2 2 2 2 1  | 3 | 1 |
-----
 6 WRO  |  3 2 1 2 2 3 1 1 |  | 2 |
-----
 5 MAYR |  1 1 1 1 1 1 1 1 | 2 |  |
-----

```

note: the rows and columns for WRO have similar variances (high ones), and that the rows and columns for the mayor have similar variances (low ones). It appears that the algorithm has done a good job of blocking the matrix to produce rows and columns with similar variance within blocks. One very simple (but not too crude) view of the data is that the mayor and the WRO are quite unique, and have similar relationships with more-or-less interchangeable actors in the large remaining group.

Of course, with more partitions, one can achieve better goodness-of-fit. The analysts will have to judge for themselves, given the purposes of their study, whether the better fit is a better model -- or just a more complicated one. Here's the result for **seven** partitions:

AUTOMORPHIC BLOCKMODELS VIA TABU SEARCH

Diagonal valid? NO
 Use geodesics? YES
 Fit: 10.367

Block Assignments:

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

Blocked Distance Matrix

		1	7	9	4	6	8	3	2	0	5
		C	N	W	I	W	U	E	C	W	M
1	COUN		1	1	2	3	2	2	1	2	1
7	NEWS	2		2	1	3	2	2	1	2	1
9	WELF	2	1		2	3	2	2	1	2	1
4	INDU	1	1	2		3	2	2	1	2	1
6	WRO	3	1	1	2		3	1	2	2	2
8	UWAY	1	1	1	1	3		2	1	2	1
3	EDUC	2	1	2	1	1	2		1	1	1
2	COMM	1	1	1	1	2	1	1		2	1
10	WEST	1	1	2	2	2	2	1	1		1
5	MAYR	1	1	1	1	2	1	1	1	1	

notes: In this case, further partitions of the data result in separating additional individual actors from the larger "center." This need not be the case. Factions and splits of partitions, as well as separation of individual actors could have resulted. We are left with a view of this network as one in which there is a core of about 1/2 of the actors who are more or less substitutable in their relations with a number (six, in the result shown) of other positions.

Summary of chapter 10

The kind of equivalence expressed by the notion of automorphism falls between structural and regular equivalence, in a sense. Structural equivalence means that individual actors can be substituted one for another. Automorphic equivalence means that sub-structures of graphs can be substituted for one another. As we will see next, regular equivalence goes further still, and seeks to deal with classes or types of actors--where each member of any class has similar relations with some member of each other.

The notion of structural equivalence corresponds well to analyses focusing on how individuals are embedded in networks -- or network positional analysis. The notion of regular equivalence focuses our attention on classes of actors, or "roles" rather than individuals or groups.

Automorphic equivalence analysis falls between these two more conventional foci, and has not received as much attention in empirical research. Still, the search for multiple substitutable substructures in graphs (particularly in large and complicated ones) may reveal that the complexity of very large structures is more apparent than real; sometimes very large structures are decomposable (or partially so) into multiple similar smaller ones.

11. Regular Equivalence

Defining regular equivalence

Regular equivalence is the least restrictive of the three most commonly used definitions of equivalence. It is, however, probably the most important for the sociologist. This is because the concept of regular equivalence, and the methods used to identify and describe regular equivalence sets correspond quite closely to the sociological concept of a "role." The notion of social roles is a centerpiece of most sociological theorizing.

Formally, "Two actors are regularly equivalent if they are equally related to equivalent others." (Borgatti, Everett, and Freeman, 1996: 128). That is, regular equivalence sets are composed of actors who have similar relations to members of other regular equivalence sets. The concept does not refer to ties to specific other actors, or to presence in similar sub-graphs; actors are regularly equivalent if they have similar ties to any members of other sets.

The concept is actually easier to grasp intuitively than formally. Susan is the daughter of Inga. Deborah is the daughter of Sally. Susan and Deborah form a regular equivalence set because each has a tie to a member of the other set. Inga and Sally form a set because each has a tie to a member of the other set. In regular equivalence, we don't care which daughter goes with which mother; what is identified by regular equivalence is the presence of two sets (which we might label "mothers" and "daughters"), each defined by its relation to the other set. Mothers are mothers because they have daughters; daughters are daughters because they have mothers.

Uses of the concept

Most approaches to social positions define them relationally. For Marx, capitalists can only exist if there are workers, and vice versa. The two "roles" are defined by the relation between them (i.e. capitalists expropriate surplus value from the labor power of workers). Husbands and wives; men and women; minorities and majorities; lower caste and higher caste; and most other roles are defined relationally.

The regular equivalence approach is important because it provides a method for identifying "roles" from the patterns of ties present in a network. Rather than relying on attributes of actors to define social roles and to understand how social roles give rise to patterns of interaction, regular equivalence analysis seeks to identify social roles by identifying regularities in the patterns of network ties -- whether or not the occupants of the roles have names for their positions.

Regular equivalence analysis of a network then can be used to locate and define the nature of roles by their patterns of ties. The relationship between the roles that are apparent from regular equivalence analysis and the actor's perceptions or naming of their roles can be problematic. What actors label others with role names and the expectations that they have toward them as a result (i.e. the expectations or norms that go with roles) may pattern -- but not wholly determine

actual patterns of interaction. Actual patterns of interaction, in turn, are the regularities out of which roles and norms emerge.

These ideas: interaction giving rise to culture and norms, and norms and roles constraining interaction, are at the core of the sociological perspective. The identification and definition of "roles" by the regular equivalence analysis of network data is possibly the most important intellectual development of social network analysis.

Finding Equivalence Sets

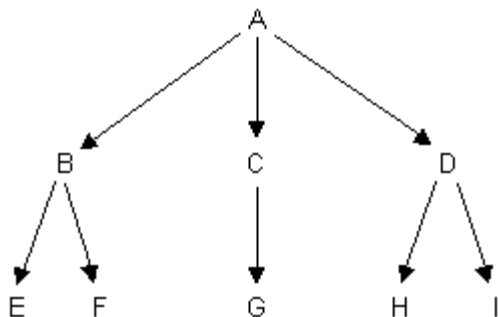
The formal definition says that two actors are regularly equivalent if they have similar patterns of ties to equivalent others. Consider two men. Each has children (though they have different numbers of children, and, obviously have different children). Each has a wife (though again, usually different persons fill this role with respect to each man). Each wife, in turn also has children and a husband (that is, they have ties with one or more members of each of those sets). Each child has ties to one or more members of the set of "husbands" and "wives."

In identifying which actors are "husbands" we do not care about ties between members of this set (actually, we would expect this block to be a zero block, but we really don't care). What is important is that each "husband" have at least one tie to a person in the "wife" category and at least one person in the "child" category. That is, husbands are equivalent to each other because each has similar ties to some member of the sets of wives and children.

But there would seem to be a problem with this fairly simple definition. If the definition of each position depends on its relations with other positions, where do we start?

There are a number of algorithms that are helpful in identifying regular equivalence sets. UCINET provides some methods that are particularly helpful for locating approximately regularly equivalent actors in valued, multi-relational and directed graphs. Some simpler methods for binary data can be illustrated directly.

Consider, again, the Wasserman-Faust example network. Imagine, however, that this is a picture of order-giving in a simple hierarchy. That is, all ties are directed from the top of the diagram downward. We will find a regular equivalence characterization of this diagram.



For a first step, characterize each position as either a "source" (an actor that sends ties, but does

not receive them), a "repeater" (an actor that both repeats and sends), or a "sink" (an actor that receives ties, but does not send). The source is A; repeaters are B, C, and D; and sinks are E, F, G, H, and I. There is a fourth logical possibility. An "isolate" is a node that neither sends nor receives ties. Isolates form a regular equivalence set in any network, and should be excluded from the regular equivalence analysis of the connected sub-graph.

Since there is only one actor in the set of senders, we cannot identify any further complexity in this "role."

Consider the three "repeaters" B, C, and D. In the neighborhood (that is, adjacent to) actor B are both "sources" and "sinks." The same is true for "repeaters" C and D, even though the three actors may have different numbers of sources and sinks, and these may be different (or the same) specific sources and sinks. We cannot define the "role" of the set {B, C, D} any further, because we have exhausted their neighborhoods. That is, the sources to whom our repeaters are connected cannot be further differentiated into multiple types (because there is only one source); the sinks to whom our repeaters send cannot be further differentiated, because they have no further connections themselves.

Now consider our "sinks" (i.e. actors E, F, G, H, and I). Each is connected to a source (although the sources may be different). We have already determined, in the current case, that all of these sources (actors B, C, and D) are regularly equivalent. So, E through I are equivalently connected to equivalent others. We are done with our partitioning.

The result of {A}{B, C, D} {E, F, G, H, I} satisfies the condition that each actor in each partition have the same pattern of connections to actors in other partitions. The permuted adjacency matrix looks like:

	A	B	C	D	E	F	G	H	I
A	---	1	1	1	0	0	0	0	0
B	0	---	0	0	1	1	0	0	0
C	0	0	---	0	0	0	1	0	0
D	0	0	0	---	0	0	0	1	1
E	0	0	0	0	---	0	0	0	0
F	0	0	0	0	0	---	0	0	0
G	0	0	0	0	0	0	---	0	0
H	0	0	0	0	0	0	0	---	0
I	0	0	0	0	0	0	0	0	---

It is useful to block this matrix and show its image. Here, however, we will use some special rules for determining zero and 1 blocks. If a block is all zeros, it will be a zero block. If each actor in a partition has a tie to any actor in another, then we will define the joint block as a 1-block. Bear with me a moment. The image, using this rule is:

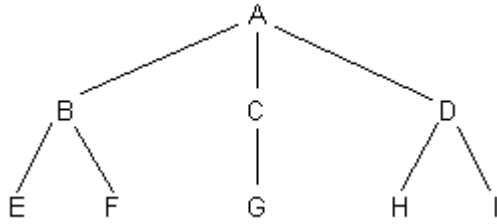
	A	B,C,D	E,F,G,H,I
A	---	1	0
B,C,D	0	---	1
E,F,G,H,I	0	0	---

A sends to one or more of BCD but to none of EFGHI. BCD do not send to A, but each of BCD sends to at least one of EFGHI. None of EFGHI send to any of A, or of BCD. The image, in fact, displays the characteristic pattern of a strict hierarchy: ones on the first off-diagonal vector and zeros elsewhere. The rule of defining a 1 block when each actor in one partition has a relationship with any actor in the other partition is a way of operationalizing the notion that the actors in the first set are equivalent if they are connected to equivalent actors (i.e. actors in the other partition), without requiring (or prohibiting) that they be tied to the same other actors. This image rule is the basis of the *tabu search algorithm*. The neighborhood search algorithm is the basis for *REGE's approach to categorical data* (see discussions and examples below).

To review: we begin a neighborhood search by characterizing each actor as a source, repeater, sink, or isolate. Looking within each category, we then examine the kinds of actors that are present in the neighborhood of each of the actors in our initial partitions. If the kinds of actors present in each of their neighborhoods are the same, we are finished and can move on to the next initial group; if the actors have different neighborhood composition, we can sub-divide them, and repeat the process. In principle, searching neighborhoods can continue outward from each actor until paths of all lengths have been examined. In practice, analysts go no further than 3 steps. For most problems, differences in the composition of actor's neighborhoods more distant than that are not likely to be substantively important.

The neighborhood search rule works well for binary, directed data. It can be extended to integer valued and multi-relational data (*REGE categorical* algorithm). If the strength of directed ties has been measured (or if one wants to use the reciprocal of the geodesic distance between actors as a proxy for tie strength), the *REGE continuous* algorithm can be applied to iteratively re-weight in the neighborhood search to identify approximately regularly equivalent actors.

Many sets of network data show non-directed ties. Applying regular equivalence algorithms can be troublesome in these cases. Consider the Wasserman-Faust network in its undirected form.

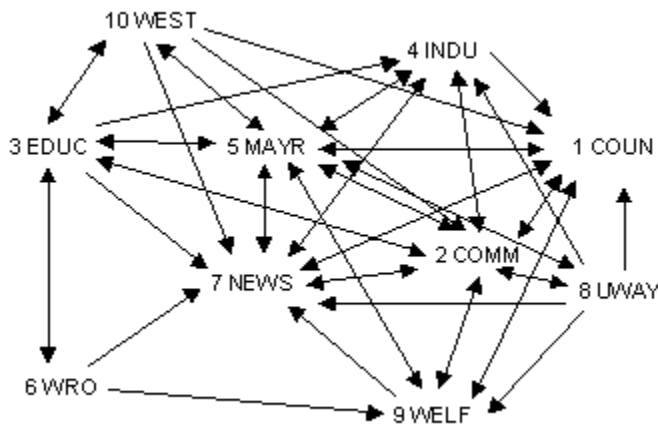


Without noting direction, we cannot divide the data into "sources" "repeaters" and "sinks," so the neighborhood search idea breaks down. In this case, it might make sense to examine the geodesic distance matrix, rather than the adjacency matrix. One can then use categorical REGE (which treats each distinct geodesic distance value as a qualitatively different neighbor). Or, one can use continuous REGE (probably a more reasonable choice), regarding the inverse of distance as a measure of tie strength.

Let's look at some examples of these approaches with real data. First, we will illustrate the basic neighborhood search approach with REGE categorical. Then, we will compare the categorical and continuous treatments of geodesics as values for an undirected graph. Lastly, we will apply the tabu search method to some directed binary data.

Categorical REGE for directed binary data (Knoke information exchange)

The Knoke information exchange network is directed and binary, though many ties are reciprocated.



The REGE algorithm, when applied to binary data, adopts the approach of first categorizing actors as sources, repeaters, and sinks. It then attempts to sub-divide the actors in each of these categories according to the types of actors in their neighborhood. The process continues (usually not for long) until all actors are divided, on no further differentiation can be drawn by extending the neighborhood. Here are the results:

CATEGORICAL REGULAR EQUIVALENCE

HIERARCHICAL CLUSTERING

	C	I	W	C	N	E	U	W	M	
	O	N	E	O	E	D	W	W	E	A
	U	D	L	M	W	U	R	A	S	Y
	N	U	F	M	S	C	O	Y	T	R
										1
Level	1	4	9	2	7	3	6	8	0	5
-----	-	-	-	-	-	-	-	-	-	-
	3
	2	XXXXX	XXX	XXXXXXXX	.					
	1	XXXXXXXXXXXXXXXXXXXXX								

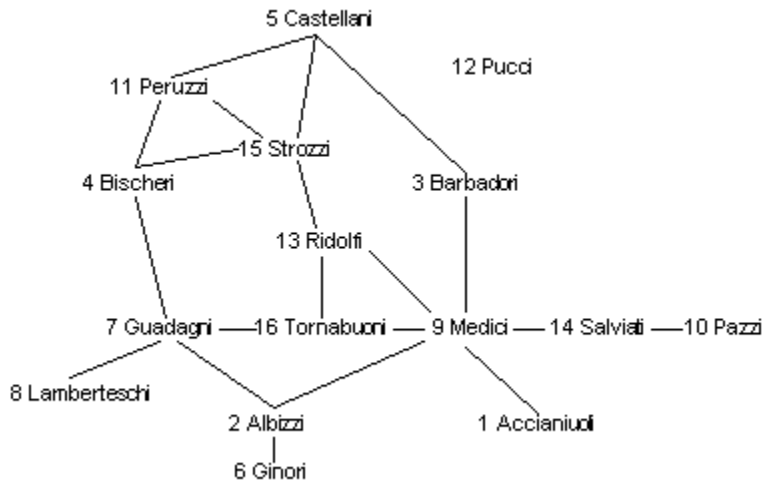
Actor-by-actor similarities

										1											
		1	2	3	4	5	6	7	8	9	0										
		C	O	C	O	E	D	I	N	M	A	W	R	N	E	U	W	W	E	W	E
		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	COUN	3	1	1	2	1	1	1	1	1	2	1									
2	COMM	1	3	1	1	1	1	2	1	1	1										
3	EDUC	1	1	3	1	1	2	1	2	1	2										
4	INDU	2	1	1	3	1	1	1	1	2	1										
5	MAYR	1	1	1	1	3	1	1	1	1	1										
6	WRO	1	1	2	1	1	3	1	2	1	2										
7	NEWS	1	2	1	1	1	1	3	1	1	1										
8	UWAY	1	1	2	1	1	2	1	3	1	2										
9	WELF	2	1	1	2	1	1	1	1	3	1										
10	WEST	1	1	2	1	1	2	1	2	1	3										

note: This approach suggests a solution of {1, 4, 9}, {2, 7}, {3, 6, 8, 10}, {5}. We will see, below, that this partition is similar (but not identical) to the solution produced by tabu search. In this case, actors and their neighborhoods are quite similar in the crude sense of "sources" "repeaters" and "sinks." Where the actor's roles are very similar in this crude sense, simple neighborhood search may produce uninteresting results.

Categorical REGE for geodesic distances (Padgett's marriage data)

The Padgett data on marriage alliances among leading Florentine families are of low to moderate density, and are undirected. There are considerable differences among the positions of the families.



The categorical REGE algorithm can be used to identify regularly equivalent actors by treating the elements of the geodesic distance matrix as describing "types" of ties -- that is different geodesic distances are treated as "qualitatively" rather than "quantitatively" different. Two nodes are more equivalent if each has an actor in their neighborhood of the same "type" in this case, that means they are similar if they each have an actor that is at the same geodesic distance from themselves. With many data sets, the levels of similarity of neighborhoods can turn out to be quite high -- and it may be difficult to differentiate the positions of the actors on "regular" equivalence grounds.

CATEGORICAL REGULAR EQUIVALENCE

WARNING: Data converted to geodesic distances before processing.

HIERARCHICAL CLUSTERING

	A	B	C	L	T
C	A	B	A	G	A
S	A	R	R	I	S
I	T	L	I	B	S
A	R	B	D	A	C
I	O	I	O	D	H
U	Z	Z	L	O	E
O	Z	Z	F	R	R
L	I	I	I	I	I

Level	1	1												
-----	1	5	2	3	3	4	5	6	7	8	9	0	1	2
3
2	XXX	XXX
1	XX													

Actor-by-actor similarities

		1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6
	ACALB	BABIC	GAGIG	ULAME	PEPE	PURIS	SAST	TO									
1	ACCIAIUOL	3	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1
2	ALBIZZI	1	3	1	1	1	1	1	1	1	1	1	1	2	1	1	1
3	BARBADORI	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1
4	BISCHERI	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1
5	CASTELLAN	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1
6	GINORI	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1
7	GUADAGNI	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1
8	LAMBERTES	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1
9	MEDICI	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1
10	PAZZI	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1
11	PERUZZI	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1
12	PUCCI	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1
13	RIDOLFI	1	2	1	1	1	1	1	1	1	1	1	1	3	1	1	1
14	SALVIATI	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	1
15	STROZZI	2	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1
16	TORNABUON	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3

notes: The use of REGE with undirected data, even substituting geodesic distances for binary values, can produce rather unexpected results. It may be more useful to combine a number of different ties to produce continuous values. The main problem, however, is that with undirected data, most cases will appear to be very similar to one another (in the "regular" sense), and no algorithm can really "fix" this. If geodesic distances can be used to represent differences in the types of ties (and this is a conceptual question), and if the actors do have some variability in their distances, this method can produce meaningful results. But, in my opinion, it should be used cautiously, if at all, with undirected data.

Continuous REGE for geodesic distances (Padgett's marriage data)

An alternative approach to the undirected Padgett data is to treat the different levels of geodesic distances as measures of (the inverse of) strength of ties. Two nodes are said to be more equivalent if they have an actor of similar distance in their neighborhood (similar in the quantitative sense of "5" is more similar to "4" than 6 is). By default, the algorithm extends the search to neighborhoods of distance 3 (though less or more can be selected).

REGULAR EQUIVALENCE VIA WHITE/REITZ REGE ALGORITHM

REGE similarities (3 iterations)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	ACCIA	BIBAR	BISCA	CASTG	INO	GUADL	LAMB	MEDI	PAZZ	PERUP	PUCC	RIDOS	SALV	STRO	TORN	
1 ACCIAIUOL	100	93	43	52	51	29	64	33	46	60	45	0	91	67	92	73
2 ALBIZZI	93	100	53	62	62	42	73	47	56	72	56	0	94	81	94	79
3 BARBADORI	43	53	100	96	95	70	95	98	91	71	96	0	52	94	57	94
4 BISCHERI	52	62	96	100	99	76	99	98	97	71	100	0	67	94	70	98
5 CASTELLAN	51	62	95	99	100	76	98	97	97	70	99	0	66	93	70	97
6 GINORI	29	42	70	76	76	100	75	83	73	92	78	0	54	74	53	79
7 GUADAGNI	64	73	95	99	98	75	100	97	97	71	99	0	76	93	79	97
8 LAMBERTES	33	47	98	98	97	83	97	100	91	84	98	0	47	95	53	96
9 MEDICI	46	56	91	97	97	73	97	91	100	66	97	0	62	86	65	92
10 PAZZI	60	72	71	71	70	92	71	84	66	100	73	0	72	73	77	78
11 PERUZZI	45	56	96	100	99	78	99	98	97	73	100	0	60	94	64	98
12 PUCCI	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0
13 RIDOLFI	91	94	52	67	66	54	76	47	62	72	60	0	100	76	100	86
14 SALVIATI	67	81	94	94	93	74	93	95	86	73	94	0	76	100	78	93
15 STROZZI	92	94	57	70	70	53	79	53	65	77	64	0	100	78	100	87
16 TORNABUON	73	79	94	98	97	79	97	96	92	78	98	0	86	93	87	100

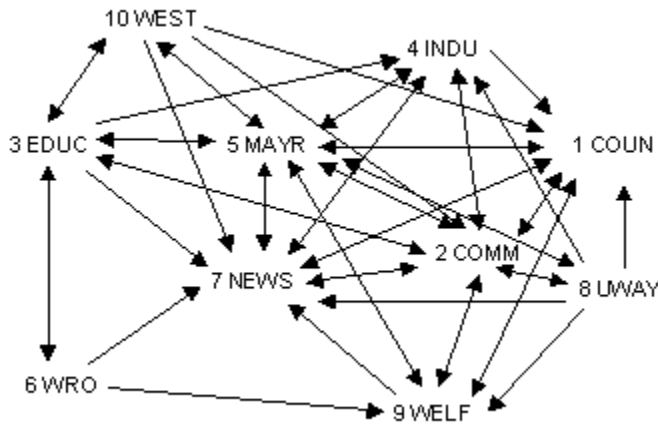
HIERARCHICAL CLUSTERING OF EQUIVALENCE MATRIX

	A			B	L	C		T								
	C		S	A	A	G	A	B	O							
	C	A	R	S	A	R	M	U	S	I	P	R				
	I	L	I	T	G	L	M	B	B	A	T	S	E	N		
	P	A	B	D	R	I	P	V	E	A	E	D	E	C	R	A
	U	I	I	O	O	N	A	I	D	D	R	A	L	H	U	B
	C	U	Z	L	Z	O	Z	A	I	O	T	G	L	E	Z	U
	C	O	Z	F	Z	R	Z	T	C	R	E	N	A	R	Z	O
	I	L	I	I	I	I	I	I	I	S	I	N	I	I	I	N
	1		1	1						1	1					
Level	2	1	2	3	5	6	0	4	9	3	8	7	5	4	1	6
-----	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
99.860	XXX	.
99.787	.	.	.	XXX	XXX	.
99.455	.	.	.	XXX	XXXXXX	.
98.609	.	.	.	XXX	XXXXXXXX	.
97.895	.	.	.	XXX	XXX	XXXXXXXX
97.676	.	.	.	XXX	XXX	XXXXXXXXXX
96.138	.	.	.	XXX	XXXXXXXXXXXXXXXX
93.756	.	.	.	XXXXXX	XXXXXXXXXXXXXXXX
93.673	.	.	.	XXXXXX	XXXXXXXXXXXXXXXX
92.796	.	.	.	XXXXXX	XXXXXXXXXXXXXXXX
92.059	.	.	.	XXXXXX	XXX	XXXXXXXXXXXXXXXXXXXXXXXX
91.911	.	.	.	XXXXXXXX	XXX	XXXXXXXXXXXXXXXXXXXXXXXX
75.816	.	.	.	XXXXXXXX	XXXXXXXXXXXXXXXXXXXXXXXX
67.386	.	.	.	XXXXXXXXXXXXXXXXXXXXXXXX
0.000	XXXXXXXXXXXXXXXXXXXXXXXX

notes: The continuous REGE algorithm applied to geodesic distances for the undirected data is probably a better choice than the categorical approach. The result still shows very high regular equivalence among the actors, and the solution is only modestly similar to that of the categorical approach.

The Knoke bureaucracies information exchange network analyzed by Tabu search

Above, we examined the Knoke information network using neighborhood search. The tabu method is an iterative search for a permutation and partition of the graph (given a prior decision about how many partitions) that produces the fewest possible "exceptions" to zero and one block coding for regular equivalence. That is, blocks are zero if all entries are zero and one if there is at least one element in each row and column. Here we apply this method to the same data set that we used earlier to look at the simple neighborhood search approach. The result is:



REGULAR BLOCKMODELS VIA TABU SEARCH

Number of blocks: 4

Fit: 2.000

Block Assignments:

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

Blocked Adjacency Matrix

		5 2		1 6 0 3			4 7 1 9				8
		M C		W W E		I N C W				U	
5	MAYR		1		1 1		1 1 1 1		1		
2	COMM		1		1		1 1 1 1		1		
6	WRO				1		1 1				
10	WEST		1 1		1		1 1				
3	EDUC		1 1		1 1		1 1				
4	INDU		1 1				1 1				
7	NEWS		1 1				1				
1	COUN		1 1				1 1				
9	WELF		1 1				1				
8	UWAY		1 1				1 1 1 1				

notes: The method produces a fit statistic, and solutions for different numbers of partitions should be compared. In the current case, a 5 group model fits somewhat better than the 4 group solution shown here. The four group solution is reported, so that it can be compared to the neighborhood search result, above.

The blocked adjacency matrix for the 4 group solution is, however, quite convincing. Of the 12 blocks of interest (the blocks on the diagonal are not usually treated as relevant to "role" analysis) 10 satisfy the rules for zero or one blocks perfectly.

The solution is also an interesting one substantively. The first set (2,5) for example, are pure "repeaters" sending and receiving from all other roles. The set { 6, 10, 3 } send to only two other types (not all three other types) and receive from only one other type. And so on.

The tabu search method can be very useful, and usually produces quite nice results. It is an iterative search algorithm, however, and can find local solutions. Many networks have more than one valid partitioning by regular equivalence, and there is no guarantee that the algorithm will always find the same solution. It should be run a number of times with different starting configurations.

Summary of chapter 11

The regular equivalence concept is a very important one for sociologists using social network methods, because it accords well with the notion of a "social role." Two actors are regularly equivalent if they are equally related to equivalent (but not necessarily the same, or same number of) equivalent others. Regular equivalencies can be exact or approximate. Unlike the structural and automorphic equivalence definitions, there may be many valid ways of classifying actors into regular equivalence sets for a given graph -- and more than one may be meaningful.

There are a number of algorithmic approaches for performing regular equivalence analysis. All are based on searching the neighborhoods of actors and profiling these neighborhoods by the presence of actors of other "types." To the extent that actors have similar "types" of actors at similar distances in their neighborhoods, they are regularly equivalent. This seemingly loose definition can be translated quite precisely into zero and one-block rules for making image matrices of proposed regular equivalence blockings. The "goodness" of these images is perhaps the best test of a proposed regular equivalence partitioning. And, the images themselves are the best description of the nature of each "role" in terms of its' expected pattern of ties with other roles.

We have only touched the surface of regular equivalence analysis, and the analysis of roles in networks. One major extensions that make role analysis far richer is the inclusion of multiple kinds of ties (that is, stacked or pooled matrices of ties). Another extension is "role algebra" which seeks to identify "underlying" or "generator" or "master" relations from the patterns of ties in multiple tie networks (rather than simply stacking them up or adding them together).

A Working Bibliography on Social Network Analysis Methods

- Alba, Richard D. 1973. "A graph-theoretic definition of a sociometric clique" *Journal of Mathematical Sociology*, 3: 113-126
- Alba, Richard D. and Gwen Moore. 1983. "Elite social circles" Chapter 12 in Burt and Minor (eds). *Applied network analysis: A methodological introduction*. Beverly Hills: Sage.
- Anheier, Helmut K. 1987. "Structural analysis and strategic research design: Studying politicized interorganizational networks" *Sociological Forum* 2: 563-582
- Arabie, P., S.A. Boorman, and P.R. Levitt. 1978. "Constructing blockmodels: How and why" *Journal of Mathematical Psychology*, Vol. 17: 21-63
- Baker, Wayne E. 1986. "Three-dimensional blockmodels" *Journal of Mathematical Sociology*, 12: 191-223.
- Baker, Wayne E. 1990. "Market Networks and Corporate Behavior" *American Journal of Sociology*. 96: 589-625.
- Barnes, J.A. 1983. "Graph theory in network analysis" *Social Networks* 5: 235-244.
- Berkowitz, S.D. 1988. "Markets and market-areas: Some preliminary formulations" pp. 261-303 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press
- Berkowitz, S.D. 1988. Afterword: Toward a formal structural sociology, pp. 477-497 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Berkowitz, S.D. 1982. *An introduction to structural analysis: The network approach to social research* Toronto: Butterworths
- Berkowitz, S.D., Peter J. Carrington, Yehuda Kotowitz, and Leonard Waverman. 1978-79. The determination of enterprise groupings through combined ownership and directship ties, *Social Networks* 1: 391-413.
- Blau, Peter and Robert K. Merton. 1981. *Continuities in structural inquiry* London and Beverly Hills: Sage.
- Blau, Peter. 1981. Introduction: Diverse views of social structure and their common denominator. in Blau and merton (eds.) *Continuities in structural inquiry*. London and Beverly Hills: Sage.
- Bodemann, Y. Michal. 1988. Relations of production and class rule: The hidden basis of patron-clientage. pp. 198-220 in Wellman and Berkowitz (eds.) *Social structures: A*

network approach. Cambridge: Cambridge University Press

- Bonacich, Phillip. 1972. Technique for analyzing overlapping memberships. in Herbert Costner (ed.) *Sociological Methodology 1972*. San Francisco: Jossey-Bass
- Bonacich, Phillip. 1972. Factoring and weighting approaches to status scores and clique detection, *Journal of Mathematical Sociology*, 2: 113-120
- Boorman, S.A. 1975. A combinatorial optimization model for transmission of job information through contact networks, *Bell Journal of Economics*, 6: 216-249
- Borgatti, Steven, Martin Everett, and Linton Freeman. 1992. *UCINET IV Version 1.0 User's Guide*. Columbia, SC: Analytic Technologies.
- Bott, Elizabeth. 1957. *Family and social network* London: Tavistock Publications
- Bougon, Michel, Karl Weick, and Din Binkhorst. 1977. Cognition in organizations: An analysis of the Utrecht jazz orchestra, *Administrative Science Quarterly*, Vol. 22: 606-639
- Boyd, John Paul, John H. Haehl and Lee D. Sailer. 1972. Kinship systems and inverse semigroups, *Journal of Mathematical Sociology*, 2: 37-61
- Breiger, Ronald. 1979. Toward an operational theory of community elite structure, *Quality and Quantity*, 13: 21-47.
- Breiger, Ronald L. and Philippa E. Pattison. 1978. The joint role structure of two communities' elites, *Sociological Methods and Research*, 7: 213-26.
- Breiger, Ronald L. and Philippa E. Pattison. 1986. Cumulated social roles: The duality of persons and their algebras, *Social Networks*, 8: 215-256.
- Breiger, R.L. 1976. Career attributes and network structure: A blockmodel study of a biomedical research specialty, *American Sociological Review*, 41: 117-135
- Breiger, Ronald L. 1988. The duality of persons and groups. pp. 83-98 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Brym, Robert J. 1988. Structural location and ideological divergence: Jewish Marxist intellectuals in turn-of-the-century Russia. pp. 332-358 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Burkhardt, Marlene E. and Daniel J. Brass. 1989. Changing patterns or patterns of change: A longitudinal investigation of the interplay of technology, structure, and power, *Pennsylvania State University*, mimeo.
- Burt, Ronald S. 1982. *Toward a Structural Theory of Action*. New York: Academic Press.

- Burt, Ronald S. 1983. Distinguishing relational contents. pp. 35-74 in Burt and Minor (eds.) *Applied network analysis: A methodological introduction*. Beverly Hills: Sage
- Burt, Ronald S. 1983. Network data from informant interviews. Chapter 7 in Burt and Minor (eds.) *Applied network analysis: A methodological introduction*
- Burt, Ronald S. 1983. Cohesion versus structural equivalence as a basis for network subgroups. Chapter 13 in Burt and Minor (eds.) *Applied network analysis: A methodological introduction*. Beverly Hills: Sage.
- Burt, Ronald and M.J. Minor (eds.) 1983. *Applied network analysis: A methodological introduction* Beverly Hills: Sage
- Burt, Ronald S. 1982. *Toward a structural theory of action: Network models of social structure, perception, and action*. New York: Academic Press.
- Burt, Ronald S. 1983. *Corporate profits and cooptation: networks of market constraints and directorate ties in the American economy*. New York: Academic Press.
- Burt, Ronald S. 1980. Models of network structure, *Annual Review of Sociology* 6: 79-141
- Burt, Ronald S. 1983. Network data from archival records. Chapter 8 in Burt and Minor (eds.) *Applied network analysis: A methodological introduction*. Beverly Hills: Sage
- Burt, Ronald S. 1983. Range, Chapter 9 in Burt and Minor (eds.) *Applied network analysis: A methodological introduction*
- Burt, Ronald S. 1976. Position in networks, *Social Forces*, 55: 93-122.
- Burt, Ronald S. 1987. Social Contagion and Innovation. *American Journal of Sociology*. 92: 1287-1335.
- Burt, Ronald S. 1992. *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.
- Carrington, Peter J. and Greg H. Heil. 1981. COBLOC: A hierarchical method for blocking network data, *Journal of Mathematical Sociology*, 8: 103-131
- Carrington, Peter J., Greg H. Heil, and Stephen D. Berkowitz. 1979-80. A goodness-of-fit index for blockmodels, *Social Networks*, 2: 219-234
- Cartwright, Dorwin. 1979. Balance and clusterability: An overview, pp. 25-50 in P. Holland and S. Leinhardt (eds.) *Perspectives on social network research*. New York: Academic Press.
- Cartwright, Dorwin and Frank Harary. 1977. A graph theoretic approach to the investigation of system-environment relationships, *Journal of Mathematical Sociology*,

Vol. 5: 87-111.

- Cartwright, D. and F. Harary. 1956. Structural balance: A generalization of Heider's theory, *Psychological Review*, 63: 277-92.
- Clawson, Dan and Alan Neustadtl. 1989. Interlocks, PACs, and Corporate Conservatism. *American Journal of Sociology*. 94: 749-73.
- Clawson, Dan, Alan Neustadtl, and James Bearden. 1986. The Logic of Business Unity: Corporate Contributions to the 1980 Congressional Elections. *American Sociological Review*. 51: 797-811.
- Coleman, James, Elihu Katz, and Herbert Menzel. 1957. The diffusion of an innovation among physicians, *Sociometry*, 20: 253-270
- Cook, Karen. 1982. Network structures from an exchange perspective, in Peter Marsden and Nan Lin (eds.). *Social structure and network analysis*. Beverly Hills: Sage
- Crane, Diana. 1969. Social structure in a group of scientists: A test of the 'Invisible college' hypothesis, *American Sociological Review*, 34: 335-352
- Davis, J. 1963. Structural balance, mechanical solidarity, and interpersonal relations, *American Journal of Sociology*, 68: 444-62
- Davis, James A. 1967. Clustering and structural balance in graphs, *Human Relations*, 30: 181-187
- Delany, John. 1988. Social networks and efficient resource allocation: Computer models of job vacancy allocation through contacts, pp. 430-451 in Wellman and Berkowitz (eds.). *Social structures: A network approach*. Cambridge: Cambridge University
- Doreian, Patrick. 1974. On the connectivity of social networks, *Journal of Mathematical Sociology*, 3: 245-258.
- Doreian, Patrick. 1988. Equivalence in a social network, *Journal of Mathematical Sociology*, 13: 243-282.
- Emirbayer, Musafa and Jeff Goodwin. 1994. Network analysis, culture, and the problems of agency. *American Journal of Sociology*. 99: 1411-54.
- Erikson, Bonnie. 1988. The relational basis of attitudes, pp. 99-122 in Wellman and Berkowitz (eds.), *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Everett, M.G. 1982. Graph theoretic blockings K-Plexes and K-cutpoints, *Journal of Mathematical Sociology*, 9: 75-84
- Everett, Martin G. 1982. A graph theoretic blocking procedure for social networks, *Social*

Networks, 4: 147-167

- Everett, Martin and Juhani Nieminen. 1980. Partitions and homomorphisms in directed and undirected graphs, *Journal of Mathematical Sociology*, 7: 91-111
- Feinberg, S.E. and Wasserman, S. 1981. Categorical data analysis of single sociometric relations, in S. Leinhardt (ed.) *Sociological Methodology 1981*. San Francisco: Jossey-Bass.
- Fernandez, Roberto M. and Roger V. Gould. 1994. A Dilemma of State Power: Brokerage and Influence in the National Health Policy Domain. *American Journal of Sociology*. 99: 1455-91.
- Fischer, Claude. 1982. *To Dwell among Friends: Personal Networks in Town and City*. Chicago: University of Chicago Press.
- Fiskel, Joseph. 1980. Dynamic evolution in societal networks, *Journal of Mathematical Sociology*, 7: 27-46.
- Flament, C. 1963. *Applications of graph theory to group structure* Englewood Cliffs: Prentice-Hall
- Frank, Ove, Maureen Hallinan, and Krzysztof Nowicki. 1985. Clustering of dyad distributions as a tool in network modeling, *Journal of Mathematical Sociology*, 11: 47-64
- Frank, Ove. 1980. Transitivity in stochastic graphs and digraphs, *Journal of Mathematical Sociology*, 7: 199-213
- Freeman, Linton C. 1984. Turning a profit from mathematics: The case of social networks, *Journal of Mathematical Sociology*, 10: 343-360
- Freeman, Linton. 1979. Centrality in social networks: Conceptual clarification, *Social Networks*, 1: 215-39.
- French Jr., John R.P. 1956. A formal theory of social power, *Psychological Review*, 63: 181-194
- Friedkin, Noah E. 1986. A formal theory of social power, *Journal of Mathematical Sociology*, 12: 103-126
- Friedmann, Harriet. 1988. Form and substance in the analysis of the world economy, pp. 304-326 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Galaskiewicz, Joseph and Stanley Wasserman. 1981. Change in a Regional Corporate Network. *American Sociological Review*. 46: 475-84.

- Gould, Roger V. 1991. Multiple Networks and Mobilization in the Paris Commune, 1871. *American Sociological Review*. 56: 716-29.
- Gould, Roger V. 1993. Collective Action and Network Structure. *American Journal of Sociology*. 58: 182-96.
- Gould, Roger V. 1993. Trade Cohesion, Class Unity, and Urban Insurrection: Artisanal Activism in the Paris Commune. *American Journal of Sociology*. 98: 721-54.
- Granovetter, Mark. 1973. The strength of weak ties. *American Journal of Sociology*, 78: 1360-80.
- Granovetter, Mark. 1985. Economic action and social structure: the problem of embeddedness. *American Journal of Sociology*. 91: 481-510.
- Granovetter, Mark. 1994. *Getting a Job*. Cambridge, MA: Harvard University Press.
- Hage, Per and Frank Harary. 1983. *Structural models in anthropology* Cambridge: Cambridge University Press
- Harary, Frank and Helene J. Kimmel. 1979. Matrix measures for transitivity and balance, *Journal of Mathematical Sociology*, Vol 6.: 199-210
- Harary, Frank. 1971. Demiarc: An atomistic approach to relational systems and group dynamics, *Journal of Mathematical Sociology*, Vol. 1: 195-205
- Harary, F., R. Norman, and D. Cartwright. 1965. *Structural models* New York: Wiley
- Heider, Fritz. 1979. On balance and attribution, pp: 11-24 in Holland, P. and S. Leinhardt *Perspectives on Social Network research*. New York: Academic Press.
- Hoivik, Tord and Nils Petter Gleditsch. 1975. Structural parameters of graphs: A theoretical investigation, in Blalock et al. (eds.). *Quantitative Sociology*. New York: Academic Press.
- Holland, Paul W. and Samuel Leinhardt. 1977. A dynamic model for social networks, *Journal of Mathematical Sociology*, 5: 5-20
- Holland, Paul W. and Samuel Leinhardt (eds.) 1979. *Perspectives on social network research* New York: Academic Press
- Holland, Paul W. and Samuel Leinhardt. 1976. Local structure in social networks, pp. 1-45 in David Heise (ed.) *Sociological Methodology, 1976*. San Francisco: Josey-Bass
- Howard, Leslie. 1988. Work and community in industrializing India, pp. 185-197 in Wellman and Berkowitz (Eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.

- Howell, Nancy. 1988. Understanding simple social structure: Kinship units and ties, pp. 62-82 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Hubbell, Charles H. 1965. An input-output approach to clique identification, *Sociometry*, 28: 377-399
- Knoke, David and Ronald S. Burt. 1983. Prominence, Chapter 10 in Burt and Minor (eds.) *Applied network analysis: A methodological introduction*. Beverly Hills: Sage
- Knoke, D. and J. H. Kuklinski. 1981. *Network analysis* Beverly Hills: Sage
- Krackhardt, David. 1989 Graph theoretical dimensions of informal organizations, Washington, D.C. presented at the *Academy of Management* meetings.
- Laumann, Edward O. 1973. *Bonds of Pluralism: The Form and Substance of Urban Social Networks*. New York: Wiley.
- Laumann, Edward O., Peter V. Marsden, and Joseph Galaskiewicz. 1977. *Community Influence Structures: Extension and Replication of a Network Approach*. *American Journal of Sociology*. 83: 594-631.
- Laumann, Edward O., Peter V. Marsden and David Prensky. 1983. The boundary specification problem in network analysis, pp. 18-34 in Burt and Minor (eds.) *Applied network analysis: A methodological introduction*. Beverly Hills: Sage.
- Laumann, Edward O. and Peter V. Marsden. 1982. Microstructural analysis in interorganizational systems, *Social Networks*, 4: 329-48.
- Laumann, Edward O., et al. 1994. *The Social Organization of Sexuality: Sexual Practices in the United States*. Chicago: University of Chicago Press.
- Leik, Robert K. and B.F. Meeker. 197?. Graphs, matrices, and structural balance, pp. 53-73 in Leik and Meeker. *Mathematical Sociology*. Englewood Cliffs: Prentice-Hall
- Leinhardt, Samuel (ed.) 1977. *Social networks: A developing paradigm* New York: Academic Press
- Levine, Joel H. 1972. The sphere of influence, *American Sociological Review*, 37: 14-27
- Levine, Joel H. and John Spadaro. 1988. Occupational mobility: A structural model, pp. 452-476 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Lorrain, Françoise and Harrison C. White. 1971. The structural equivalence of individuals in social networks, *Journal of Mathematical Sociology*, 1: 49-80.
- Mandel, Michael J. 1983. Local roles and social networks, *American Sociological*

Review, 48: 376-386

- Marsden, Peter and Nan Lin (eds.) 1982. *Social structure and network analysis*. Beverly Hills: Sage.
- Marwell, Gerald, Pamela Oliver, and Ralph Prahl. 1988. Social Networks and Collective Action: A Theory of the Critical Mass. III. *American Journal of Sociology*. 94: 502-34.
- Mayer, Thomas F. 1984. Parties and networks: Stochastic models for relationship networks, *Journal of Mathematical Sociology*, 10: 51-103.
- McCallister, Lynne and Claude S. Fischer. 1983. A procedure for surveying personal networks, pp. 75-88 in Burt and Minor (eds.) *Applied network analysis: A methodological introduction*. Beverly Hills: Sage
- Minor, Michael J. 1983. Panel data on ego networks: A longitudinal study of former heroin addicts, chapter 4 in Burt and Minor (eds.) *Applied network analysis: A methodological introduction*. Beverly Hills: Sage.
- Minor, Michael J. 1983. New directions in multiplexity analysis, Chapter 11 in Burt and Minor (Eds.) *Applied network analysis: A methodological introduction*. Beverly Hills: Sage
- Mintz, Beth and Michael Schwartz. 1981. Interlocking directorates and interest group formation, *American Sociological Review*, 46: 851-69.
- Mintz, Beth and Michael Schwartz. 1985. *The power structure of American business* Chicago: University of Chicago Press.
- Mitchell, Clyde. 1969. *Social networks in urban situations: Analyses of personal relationships in central African towns* Manchester: Manchester University Press.
- Mizruchi, Mark S. 1989. Similarity of Political Behavior among American Corporations. *American Journal of Sociology*. 95: 401-24.
- Mizruchi, Mark S., Peter Mariolis, Michael Schwartz, and Beth Mintz. 1986. Techniques for disaggregating centrality scores in social networks, pp. 26-48 in Nancy Tuma (ed.) *Sociological Methodology 1986*. San Francisco: Jossey-Bass.
- Neustadtl, Alan and Dan Clawson. 1988. Corporate Political Groupings: Does Ideology Unify Business Political Behavior? *American Sociological Review*. 53: 172-90.
- Oliver, Melvin. 1988. The Urban Black Community as Network: Toward a Social Network Perspective. *Sociological Quarterly*. 29: 623-645.
- Padgett, John F. and Christopher K. Ansell. 1993. Robust Action and the Rise of the Medici, 1400-1434. *American Journal of Sociology*. 98: 1259-1319.

- Peay, Edmund R. 1976. A note concerning the connectivity of social networks, *Journal of Mathematical Sociology*, 4: 319-321
- Peay, Edmund R. 1982. Structural models with qualitative values, *Journal of Mathematical Sociology*, 8: 161-192
- Rapoport, A. 1963. Mathematical models of social interaction, pp. 493-579 in R. Luce, R. Bush and E. Galanter (eds.), *Handbook of mathematical psychology*, Vol. II. New York: Wiley
- Rogers, Everett M. 1979. Network analysis of the diffusion of innovations, pp. 137-164 in P. Holland and S. Leinhardt (eds.) *Perspectives on social network research*.
- Roy, William. 1983. The Interlocking Directorate Structure of the United States. *American Sociological Review*. 42: 248-57.
- Sailer, Lee Douglas. 1978. Structural equivalence: Meaning and definition, computation and application, *Social Networks*, I: 73-90
- Scott, John. 1991. *Social Network Analysis: A Handbook*. Newbury Park, CA: Sage Publications.
- Seidman, Stephen B. 1985. Structural consequences of individual position in nondyadic social networks, *Journal of Mathematical Psychology*, 29: 367-386
- Smith, David A. and Douglas R. White n.d. Structure and dynamics of the global economy: Network analysis of international trade 1965-1980, *University of California, Irvine*, mimeo.
- Snyder, David and Edward L. Kick. 1979. Structural position in the world system and economic growth 1955-1970: A multiple-network analysis of transnational interactions, *American Journal of Sociology*, 84: 1096-1126
- Stark, Rodney and William Sims Bainbridge. 1980. Networks of faith: Interpersonal bonds and recruitment to cults and sects, *American Journal of Sociology*, 85: 1376-95
- Tepperman, Lorne. 1988. Collective mobility and the persistence of dynasties, pp. 405-429 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Tilly, Charles. 1988. Misreading, then reading, nineteenth-century social change, pp. 332-358 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Travers, Jeffrey and Stanley Milgram. 1969. An experimental study of the small world problem, *Sociometry* 32: 425-443
- Turk, Herman. 1970. Interorganizational networks in urban society: Initial perspectives

and comparative research, *American Sociological Review*, 35: 1-20.

- Uehara, Edwina. 1990. Dual Exchange Theory, Social Networks, and Informal Social Support. *American Journal of Sociology*. 96: 521-57.
- Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Wellman, Barry, Peter J. Carrington, and Alan Hall. 1988. Networks as personal communities, pp. 130-184 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Wellman, Barry and S.D. Berkowitz (eds.) 1988. *Social structures: A network approach* Cambridge: Cambridge University Press.
- Wellman, Barry. 1988. Networks as Personal Communities. Pp. 130-184 in Wellman and Berkowitz (Eds.) *Social Structures: A Network Approach*. New York: Cambridge University Press.
- Wellman, Barry. 1988. Structural analysis: From method and metaphor to theory and substance, pp. 19-61 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Wellman, Barry. 1979. The community question: The intimate networks of East Yorkers, *American Journal of Sociology*, 84: 1201-31.
- Wellman, Barry and S.D. Berkowitz. 1988. Introduction: Studying social structures, pp. 1-14 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- Wellman, Barry, and Scot Wortley. 1990. Different Strokes from Different Folks: Community Ties and Social Support. *American Journal of Sociology*. 96: 558-88.
- White, Douglas R. and H. Gilman McCann. 1988. Cites and fights: Material entailment analysis of the eighteenth-century chemical revolution, pp. 359-379 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- White, Harrison. 1988. Varieties of markets, pp. 226-260 in Wellman and Berkowitz (eds.) *Social structures: A network approach*. Cambridge: Cambridge University Press.
- White, Douglas R. and Karl P. Reitz. 1983. Graph and semigroup homomorphisms on networks and relations, *Social Networks*, 5: 193-234
- White, Harrison, S. Boorman, and R. Breiger. 1976. Social structure from multiple networks. I.: Blockmodels of roles and positions, *American Journal of Sociology*, 81: 730-780.

- Winship, Christopher and Michael Mandel. 1983. Roles and positions: A critique and extension of the blockmodeling approach, pp. 314-344 in Samuel Leinhardt (ed.) *Sociological Methodology 1983-1984*. San Francisco: Jossey-Bass
- Witz, Klaus and John Earls. 1974. A representation of systems of concepts by relational structures, pp. 104-120 in Paul A. Ballonoff (ed.) *Mathematical models of social and cognitive structures: Contributions to the mathematical development of anthropology*.
- Wu, Lawrence. 1983. Local blockmodel algebras for analyzing social networks, pp. 272-313 in Samuel Leinhardt (ed.) *Sociological Methodology 1983-84*. San Francisco: Jossey-Bass.
- Yamagishi, Toshio and Karen S. Cook. 1993. Generalized Exchange and Social Dilemmas. *Social Psychology Quarterly*. 56: 235-48.

Index

- agglomerative cluster analysis 80
- balance theory 46
- block 29, 30, 86
 - image 30
- Bonacich power index 72
- Boolean multiplication of matrix 32
- clique 77
 - bottom-up approach 79
 - agglomerative cluster analysis 80
 - K-core 85
 - K-plexes 84
 - Maximal optimal sub-graph 80
 - N-clan 82
 - N-clique 81
 - top-down approach 85
 - block 86
 - component of graph 86
 - cutpoints 86
 - factions 88
 - Lambda sets 87
- cohesion
 - Hubbel, Katz, Taylor 54, 55
- CONCOR 106
- connections 40
- cutpoints 86
- data collection 7
 - full network 7
 - snowball 8
 - ego-centric (with alter) 8
 - ego-centric (ego only) 9
- degree of node (actor)
 - in-degree 43
 - out-degree 43
 - power 61
- density of network 41
- diameter 50
- distance 46
- ego-centric network data (ego only) 9
- ego-centric network data (alter) 8
- eigenvector 68
- equivalence
 - automorphic 117
 - regular 94, 127
 - structural 94, 98
- Euclidean distance 102
- factions 88
- Factor analysis 110
- flow 52
- centrality 70
- geodesic distance 50
- graphs 22-23
 - component of 86
 - Maximal Complete sub-graph 80
- Hubbel 54, 55
- image 107
- Jaccard coefficient 104
- Katz 54, 55
- K-core 85
- K-plexes 84
- Lambda sets 87
- matrix, 26
 - addition 31
 - adjacency 27
 - asymmetrical 28
 - correlation 32
 - image 30
 - multiplication 32
 - Boolean 32
 - partitioning 29
 - regression 32
 - subtraction 31
 - transposing 31
- Maximal Complete sub-graph 80
- maxism 119
- measurement scales 11
 - binary 11
 - multiple-category nominal 11
 - grouped ordinal 12
 - full-rank ordinal 13
 - interval 13
- Multi-dimensional scaling 113
- N-clan 82
- N-clique 81
- nodes 4
- paths 47
- Percent of exact matches 103
- Person correlation coefficient 100
- power 60
 - betweenness 62
 - Bonacich index 72
 - centrality 63-65
 - closeness 62
 - degree 61
- Principle components analysis 110
- reachability 44
- reciprocity 45
- REGE algorithm 130 ppf
- similarity 97
- sinks 40
- size of network 41
- snowball 8
- sociograms 22
- sources 40
- structural holes 37
- super-node 30
- Tabu search 109, 119, 130, 136
- Taylor 54, 55
- trail 47
- transitivity 41, 45
- triad census 41
- vector 28
 - column, row
- walks 47