CHAPTER 6

Social Networks and Organizations

Brandon Ofem,
Theresa M. Floyd,
and Stephen P. Borgatti

The field of social network analysis (SNA) has exploded in recent years, not only in the social sciences but also in biological and physical sciences. A hallmark of the approach is a focus on the social environment of actors as causative factors in addition to internal characteristics of the social actor. While the roots of this relational or socio-environmental perspective can be traced back to the ancient Greeks, the birth of the modern field of SNA is often credited to the psychiatrist Jacob Moreno, who in the 1930s, developed a field he called sociometry. In a canonical study, Moreno analyzed instances of runaways at a boarding school in upstate New York. He observed that runaways tended to occur in clumps of connected girls: girls were more likely to run away when they resided in the same cabin and were friends with other girls who ran away. He argued that social ties between the girls served as channels for the flow of ideas and influence between them, and therefore, a girl's position within the school social structure was an important determinant of whether and when she ran away.

Sociometry paved the way for even more formalized approaches in the 1940s and 1950s as researchers began translating sociological concepts such as Cooley's primary group into mathematical form using the languages of graph theory and matrix algebra. It was during the 1950s that Alex Bavelas and his student Harold Leavitt at MIT began studying the effects of network structure on the performance of groups.

A Companion to Organizational Anthropology, First Edition.
Edited by D. Douglas Caulkins and Ann T. Jordan.

Bavelas and Leavitt found that, for simple tasks, highly centralized communication structures such as shown in Figure 6.1a were better at solving problems than highly decentralized structures (Figure 6.1b), although people in the decentralized structures had more fun.

Today, the field of SNA is a rich collection of theoretical concepts and empirical methods. The field is not only known for its use of mathematics to define concepts and subsequent measures, but also includes a rich ethnographic and qualitative tradition. Furthermore, while the center of theoretical attention is on the relations among actors, characteristics of the actors are also taken into account. For the study of

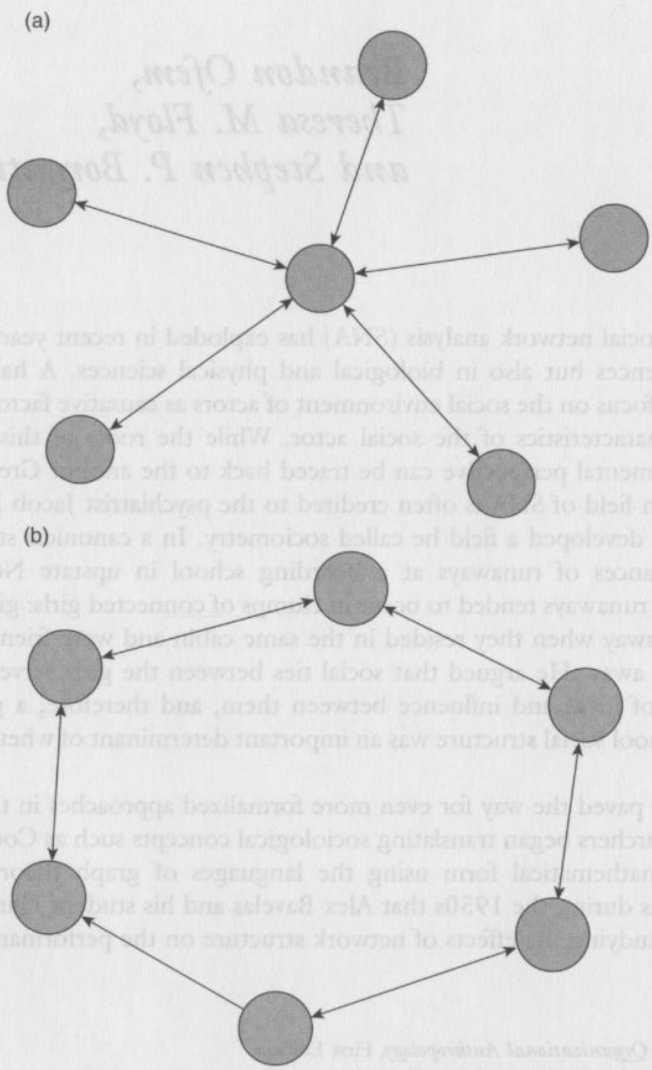

Figure 6.1 Communication structures.

organizations, which consist of a wide range of actors and relations, this flexible lens can be pretty insightful.

The purpose of this chapter is to provide an overview of the field, including both theory and method, as applied to the understanding of organizations. We start with a discussion of fundamental concepts, and end the chapter with a discussion of the methodological tools and challenges.

Network Theory

A traditional mode of explanation in the social sciences has focused on the characteristics of entities to predict their outcomes. For example, since the seminal work of Weber, it has been commonplace to explain the performance of organizations in terms of the characteristics of the organization, such as the degree of bureaucratization, or the quality and quantity of its resources (e.g., the resource-based view of the firm; Wernerfelt 1984). Or consider individual achievement within organizations. A well-known line of research seeks to explain the success of leaders in terms of characteristics of the person, such as personality traits, behaviors, and so on (Hogan *et al.* 1994). These kinds of explanations have an atomistic quality in the sense that the units of analysis are considered in isolation.

Another mode of explanation takes into account the environment of the individual entity. In organizational theory, we can see a general shift over the twentieth century toward an open systems perspective in which the environment of an organization, and particularly the fit between an organization and its environment, is given increasing weight in explaining what happens to that organization. In these approaches, the environment tends to be monolithic in the sense that it is a single thing whose qualities are measured, such as how turbulent it is or how complex it is.

The network approach falls generally in this environmentally oriented perspective, but is distinguished by taking a more complex view of the environment. In the network approach, the environment of an actor consists of other individual actors, each with their own agency, but also connected via various kinds of ties to form a definable structure – the network. The shape of this network, and the position of the actor within the network, is then seen as a key factor in determining actor outcomes (in addition to internal characteristics of the actor).

Thus, what makes SNA distinctive is its focus on the *network* of relationships among a set of actors or nodes. The nodes can be people, teams, departments, organizations, nations, or any other type of active entity.[1] The ties can be any kind of dyadic relation that can exist between the nodes, such as friendships or contracts. It is useful to classify ties into types. Table 6.1, adapted from Borgatti *et al.* (2009) and Borgatti and Halgin (2011), presents a typology that divides possible relations into two broad classes (i.e., state-type and event-type relations) and further into five more specific types (i.e., similarities, social relations, mental relations, interactions, and flows).

State-type relations are relations that have continuity over time. This does not mean they persist indefinitely, but that while they exist, they do so continuously. An example is being someone's spouse. While the marriage lasts, the "is the spouse of" relation holds continuously. This is unlike, say, "yells at," which may occur frequently

Table 6.1 A typology of relational phenomena

State-Type Relations	Similarities	Location	For example, same spatial and temporal space
		Membership	For example, same clubs, same events
		Attribute	For example, same age, gender, ethnicity
	Social relations	Kinship	For example, brother of, daughter of
		Other role	For example, competitor of, friends with, employee of
	Mental relations	Affective	For example, likes, is annoyed by, loves
		Cognitive	For example, knows about, perceives as competitor
Event-Type Relations	Interactions		For example, talked to, advice from, avoided
	Flows		For example, money, information, personnel

over a month's time but is not continuously happening at each moment (or so one hopes). We divide state-type relations into three categories, namely, similarities, social relations, and mental relations. *Similarities* refer to spatial and temporal proximity as well as comembership in groups and events and sharing socially significant attributes, such as ethnicity or gender. We typically think of similarities as relational conditions that enable various other kinds of relations. Hence, being neighbors provides opportunities to interact. *Social relations* are ongoing ties, often role-based, such as kinship and friendship. Social relations typically have institutionalized rights and obligations associated with them, and have a sense of intersubjective reality, meaning that we think there is a right answer to the question "are X and Y friends?" *Mental relations* are perceptions of and attitudes toward others, such as recognizing who someone is, or liking/disliking them. In contrast to social relations, these are usually seen as private and unobservable. They must be disclosed by the actors the analyst is studying.

Event-type relations are relations that have a discrete, transitory nature that occur with a certain frequency over a period of time but do not exist continuously during that time. One category of event-type relations is interactions, such as exchanging e-mails or making a sale. We typically view interactions as being facilitated by and occurring in the context of social/mental relations (and vice versa). For example, friends (social relation) give each other advice (interaction). At the same time, through interactions, social relations may evolve (e.g., friends can become business partners). Another category of event-type relations is flows. *Flows* are those tangible and intangible things that are transmitted through interactions. Ideas are transmitted through communication, viruses and material resources through physical contact, and so on. For practical reasons, flows have typically not been measured directly but rather inferred from interactional and relational data.

Dyadic ties have been studied in other research areas, but the essential difference in the network concept is that ties are not studied in isolation. For example, a psychologist might be interested in the mother–daughter relationship and collect data on several hundred unrelated mother–daughter pairs. But in network analysis, the ties are seen as linking up through common actors to form paths and, ultimately, whole networks. Much of the machinery of network analysis consists in characterizing the properties of the interesting structures these interlinked pairs of node have. Thus, in the analysis, we would normally treat a node's position in the friendship network as a different quantity from their position in the coworker or advice-giving networks. These can then be used together in higher level models to understand node outcomes, such as performance and rewards.

It should be noted that ties can be directed or undirected. Directed ties are those that have a sense of direction, such as "lends money to" or "is the boss of." Mathematically, they correspond to ties being viewed as ordered pairs of nodes. Undirected ties are those that have no sense of direction and correspond to unordered pairs. For example, if nodes A and B are "in communication," then there is no sense of direction, and it does not matter whether one says A is in communication with B or B is in communication with A. The same is true of the "belongs to the same club" relation. Direction is a theoretical property of ties which may or may not be borne out empirically. For example, if A and B are in a romantic relationship, we see this as theoretically undirected. But in actual data, we may find that A admits to the relationship while B does not.

Ties have been mostly frequently measured on a presence/absence scale, as in "reports to" or "is friends with." However, strengths of ties can also be measured, as in "how much do you like this person, on a 1 to 5 scale." For interactions, frequencies may be obtained, as in "how often did you collaborate with this person." In general, valued ties permit the respondent to respond with greater nuance and give dyadic regressions more variance to work with. They can also be dichotomized in the analysis as needed (since some graph-theoretic algorithms are not designed for valued data). A disadvantage of valued surveys is that they can be time-consuming and may be more difficult to administer to some respondents.

A question that often arises regarding the measurement of ties concerns the accuracy and reliability of network questions. In attitude research, for instance, a good practice is to ask a series of questions that all measure the attitude in question, and then – perhaps guided by a factor analysis – construct a composite scale using the items that seem to correlate well with each other. Ideally, we would do the same thing in network research so that to construct the friendship network we would ask a number of questions based on the friendship concept, and then combine these into a best estimate of the friendship between each pair of persons. In practice, however, researchers typically cannot afford to ask too many network questions because they quickly exhaust the respondent. However, some methodological research supports the notion that single-item network questions are adequate with respect to reliability (Marsden 1990).

Relational data describing ties are usually put in matrix form. The most common type of data structure appropriate for SNA is the *adjacency matrix*. This is a square matrix in which the number of row and column vectors is equal and the list of nodes is the same for both. For dichotomous data, a 0 or 1 is entered in each cell to indicate the existence of a tie. For valued data, any value can be entered to reflect the nature/strength of a tie. When data on multiple relations are collected, multiple square matrices are required, with each matrix representing a particular type of tie. Below

(a)

	Mary	Bob	John	Jack	Phil	Mark
Mary	0	0	0	0	1	0
Bob	0	0	0	0	1	0
John	0	0	0	0	1	0
Jack	0	0	0	0	1	0
Phil	1	1	1	1	0	1
Mark	0	0	0	0	1	0

(b)

	Mary	Bob	John	Jack	Phil	Mark
Mary	0	1	0	0	0	0
Bob	1	0	0	0	0	0
John	0	1	1	1	0	0
Jack	0	0	0	0	1	0
Phil	0	0	1	1	0	1
Mark	1	0	0	0	1	0

Figure 6.2 Adjacency matrices.

are two matrices, Figure 6.2a,b, that summarize the relational data used to create the visualizations of Figure 6.1a, b, respectively.

Domains of Network Theorizing

It is important to note that there are two fundamental domains of network research. What we refer to as *network theory* proper is the study of network properties as factors or independent variables to explain node outcomes. These node outcomes can be non-network variables, such as employee performance, but can also be network properties of their own. In contrast, *theory of networks* refers to the explanation of how network properties come to be – why are some nodes more central than others, and why do people have the friends they do. Again, these explanations may also involve non-network factors, such as personality, but can also involve other network processes. Essentially, as shown in Figure 6.3, we posit two overlapping domains of inquiry – networks as antecedents and networks as consequents – whose intersection refers to endogenous models of network change.

Units of Analysis

Up to this point we have been vague about the kinds of things we mean by network properties. For convenience, we divide network properties into three classes, based on unit of analysis. The first level is the dyadic, which is the most iconic for network analysis. Dyadic properties have a value for each pair of nodes in the network. The second level is the node. Node-level variables have a value for each node, and can be simple aggregations of dyadic properties (e.g., the number of ties a node has). The third level is the group or whole network level. Here, there is a single value that describes the whole network, such as the proportion of pairs that trust each other. Table 6.2 gives examples of network concepts at each level of analysis, along with

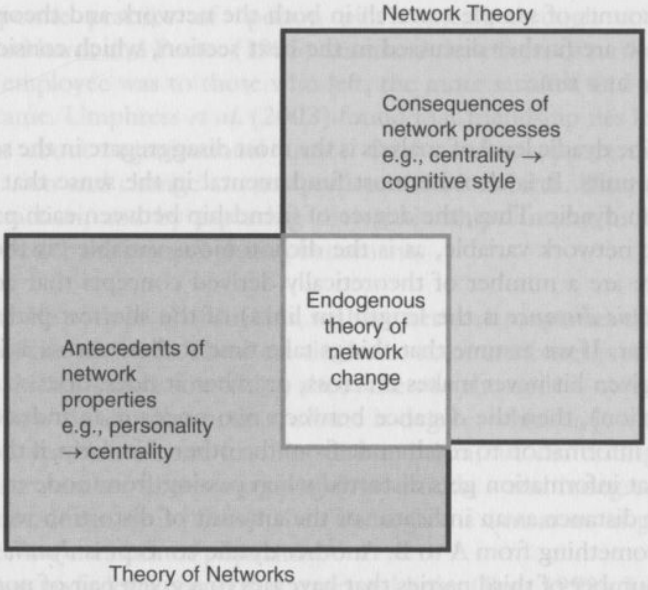

Figure 6.3 Domains of theorizing.

Table 6.2 Network research and concepts

	Theory of Networks (Antecedents)	Social Network Concepts	Network Theory (Consequences)
Dyad level	*Box 1. Selection.* Explaining the pattern of formation and dissolution of ties between nodes.[a]	• Geodesic distance • Dyadic embeddedness • Connectivity • Homophily • Transitivity	*Box 2. Influence.* Explaining similarity in attitudes and behavior as a function of interaction with each other[b]
Node level	*Box 3. Positional Achievement.* Predicting which nodes will develop a given type of centrality[c]	• Centrality (e.g., degree, closeness, betweenness, eigenvector) • Structural holes	*Box 4. Individual Social Capital.* Predicting actor performance and reward outcomes as a function network position[d]
Group or network level	*Box 5. Structuring.* Predicting which groups will develop a given structural characteristic, such as clumpiness[e]	• Density • Centralization • Scale-freeness	*Box 6. Group Social Capital.* Predicting group outcomes as a function of the structure of each group's network[f]

[a]McPherson *et al.* (2001); [b]Krackhardt and Porter (1985); [c]Mehra *et al.* (2001); [d]Burt (1992); [e]Johnson *et al.* (2003); [f]Bavelas (1950).

thumbnail accounts of sample research in both the network and theory of networks domains. These are further discussed in the next section, which considers each level of analysis one at a time.

Dyad Level. The dyadic level of analysis is the most disaggregate in the sense of having the most data units. It is also the most fundamental in the sense that network data is, by definition dyadic. Thus, the degree of friendship between each pair of coworkers is a dyadic network variable, as is the dichotomous variable "is the boss of." In addition, there are a number of theoretically derived concepts that are dyadic. For example, ***geodesic distance*** is the length (in links) of the shortest path from a given node to another. If we assume that things take time to flow across a link (and there is a chance a given bit never makes it across, or, when it does, does so with a certain level of distortion), then the distance between two nodes is an index of the time it could take for information to reach node from the other. Similarly, if there is a certain probability that information gets distorted when passing from node to node, we can view geodesic distance as an indicator of the amount of distortion we can expect in transmitting something from A to B. Another dyadic concept is ***dyadic embeddedness***, which is the number of third parties that have ties to a given pair of nodes. Pairs who have many friends in common may be constrained in their relations with each other, knowing that anything they do will be observed by the mutual friends. Yet another dyadic concept is ***connectivity***, which is the number of third-party nodes (or ties) that would have to be deleted in order to sever all paths from one node to the other. Larger connectivity means greater robustness (i.e., it takes more things going wrong to prevent something from traveling between the two nodes).

Studies in the theory of networks domain that are phrased at the dyadic level (Box 1 in Table 6.2) seek to explain patterns of tie formation and dissolution. One of the best-known findings in this area is the principle of ***homophily*** – the tendency for people to seek (positive) ties with people who are like themselves on socially significant attributes (McPherson *et al.* 2001). For example, Marsden (1987) found that Americans are significantly more likely to discuss confidential matters with people the same age, gender, race, and education level as themselves. Similarly, Allen (1977) found that the probability of communication between workers decreases with the square of the distance between their offices. Finally, studies show that a number of human networks show marked tendencies toward ***transitivity***, meaning that if A has a tie with B, and B has a tie with C, then A and C have an increased probability of having a tie. This is usually explained via balance theory (Heider 1958), which says that if A likes B then they want to be congruent with B, so if B likes C then A would also want to like C.

Studies in Box 2 (in the network theory domain, posed at the dyadic level) attempt to explain node attributes (such as beliefs) as a function of social ties. Such studies are based on the idea that just as similar individuals are more likely to interact, that interaction is likely to make them even more similar (Erickson 1988). Communication between people leads to the flow of ideas, opinions, beliefs, and so on between them. Such flows can affect cognition, and in turn, attitudes and behaviors. For example, Coleman *et al.*'s (1957) classic study found that informal discussions among physicians led to their adoption of tetracycline. Davis (1991) showed that the now-standard corporate practice of "poison pills" spread through corporate board interlocks. Krackhardt and Porter (1985) examined the effects of turnover and found the closer an employee was to those who left, the more satisfied and committed the employee became. Umphress *et al.* (2003) found that friendship ties led to similarity in perceptions about organizational justice. Such studies support the notion that attitude formation and change are not predetermined at birth. They are influenced by social interaction, and by capturing the relationships among individuals, social scientists can better understand these phenomena.

Node Level. At the node level of analysis, we have network properties that are attached to the nodes of the network. These are similar to actor attributes (such as gender or income) but which measure something about the way the node is connected into the network. Perhaps the simplest node attribute is degree, which is just the number of ties of a given type that a node has. For directed data, we can distinguish between outdegree – the number of ties outgoing from a node – and indegree, the number of incoming ties. If the tie is "likes," then outdegree is a measure of gregariousness, while indegree is a measure of popularity.[2]

A well-known node-level property is structural holes (Burt 1992). Loosely, a *structural hole* is the lack of a tie between nodes that are both tied to a given focal node. So in Figure 6.4, node A has three structural holes (because her three contacts do not have ties with each other), while node B has none. It is often suggested that nodes with many structural holes have both control and information benefits. The control benefits derive from the focal node's ability to play their contacts off each other, as in a bidding war. The information benefits derive from the fact that, if one's contacts talk to each other, then to some extent they are redundant: talking to just one of them gets you the information that they all share. In contrast, if one's contacts are isolated from each other, there is a chance that they will give you independent and heterogeneous points of view. Structural holes are often viewed as an aspect of social capital.

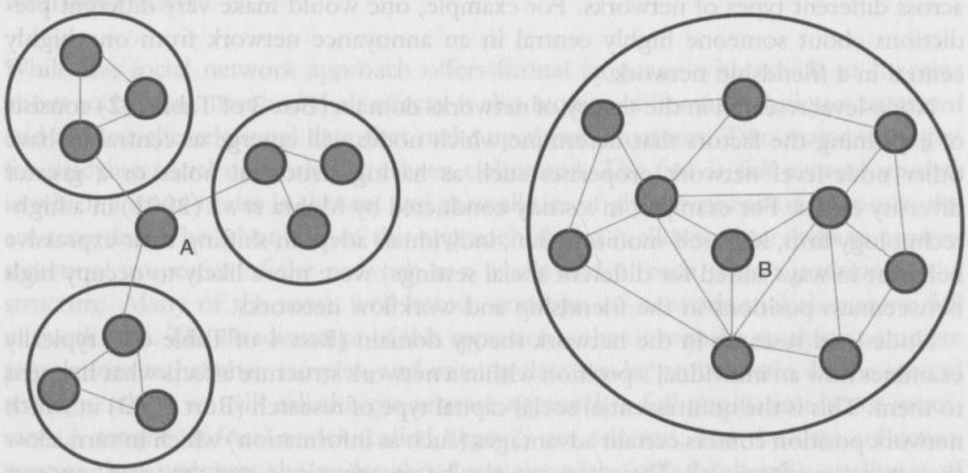

Figure 6.4 Structural holes.

A family of node-level properties is *centrality*. This family of concepts is large enough and heterogeneous enough that it is difficult to characterize. In general, however, centrality either captures some way in which a node's position is advantageous to the node, or the node is somehow important to maintaining the structure of the network. An example of the first kind is closeness centrality, which is defined as the sum of geodesic distances from a given node to all other nodes. To the extent this is a small number, the node is close to many other nodes, and is thus likely to receive whatever is flowing through the network fairly early and without too much distortion. Another example of the first kind is eigenvector centrality, which is a kind of iterated degree. For degree, we count the number of nodes that a node is connected to, weighting them all equally. In eigenvector centrality, we also count the number of nodes a node is connected to, but we weight them by their degree. This means that being connected to three well-connected people counts more than being connected to three poorly connected people. But it does not stop there. We now recount the number of nodes a node is connected to, but weight them according to their centrality on the previous round. We keep doing this until the ratios of the scores stop changing. Thus, in eigenvector centrality, a node gets a high score to the extent they are connected to many nodes who themselves are well connected.

An example of the second kind of centrality is betweenness centrality. Betweenness centrality calculates the proportion of all shortest paths in the network that pass through a given node. It can be seen as the potential for a node to act as a gatekeeper, allowing some information to pass through while filtering or coloring other information. If a node with high betweenness is removed from a network, it tends to disrupt the network significantly, often taking away the only paths between some pairs of points.

It should be noted that every centrality measure makes implicit assumptions about how traffic flows through a network, so in choosing a centrality measure, the researcher should take into consideration the nature of the social system under investigation (Borgatti 2005). Furthermore, centrality measures cannot be interpreted identically across different types of networks. For example, one would make very different predictions about someone highly central in an annoyance network from one highly central in a friendship network.

Node-level research in the theory of networks domain (Box 3 of Table 6.2) consists of explaining the factors that determine which nodes will emerge as central or have other node-level network properties such as having structural holes or a greater diversity of ties. For example, in a study conducted by Mehra *et al.* (2001) in a high-technology firm, high self-monitors (i.e., individuals adept in shifting their expressive behavior in ways suited for different social settings) were more likely to occupy high betweenness positions in the friendship and workflow networks.

Node-level research in the network theory domain (Box 4 of Table 6.2) typically examines how an individual's position within a network structure affects what happens to them. This is the quintessential social capital type of research (Burt 1992) in which network position confers certain advantages (such as information) which in turn allow the actor to perform well. The actors can also be organizations, and the same concepts apply. Indeed, position within interorganizational networks has been shown to relate to organizational performance (Gulati *et al.* 2000).

Network Level. The group or network level is concerned with properties that characterize the network as a whole. An example is *density*, which is defined as the proportion of all dyads that have a tie of a given type. It is the number of ties in a network divided by the number possible. Applied to the appropriate types of ties, density can be seen a measure of the cohesion or well-knittedness of a network. Another network-level construct is *centralization* (Freeman 1979). A network is centralized to the extent that it looks like a star: a single node in the center that is connected to every other node, and all the other nodes have ties only to the center. Still another network-level property is *scale-freeness.* A network is scale-free to the extent that just a few nodes have a very large number of ties, while a large majority of nodes have very few ties. This is similar to wealth distributions, which tend to be highly skewed with a few extraordinarily wealthy people and a very large number of poor.

Network-level research in the theory of networks domain (Box 5 of Table 6.2) focuses on how networks come to have the shapes that they do. For example, why is the World Wide Web a scale-free network? One theory, known as preferential attachment, is that nodes with many ties are highly visible, and when new web sites appear, they tend to form links with these highly visible nodes. As a result, we see a pattern in which, when it comes to web links, the rich get richer.

Network-level studies in the network theory domain (Box 6 of Table 6.2) are concerned with how the network structure of groups affect group outcomes. For example, community-level social capital research fits in this category (Putnam 2000). So does team performance research. For example, Bavelas (1950) and Leavitt (1951) showed how different communication structures affected team problem solving. They found that, for fairly simple tasks, centralized network structures tended to perform better on a series of measures, including speed, accuracy, and efficiency. The premise behind these studies is that some networks with particular properties are better suited to achieve particular objectives than others.

METHODOLOGY

While the social network approach offers formal language with which to theorize about social structures and their effects, it also requires different techniques to record and analyze the relational data that make up those structures. Two major strategies for studying social networks have been elaborated. The first is *full network analysis* in which a set of nodes is chosen, and then all ties of given types among those nodes are recorded. The advantage of this approach is that it allows researchers to capture structural properties of the network as a whole and of each node's position in the structure. Many of the more well-known concepts of network analysis assume full network data. The disadvantage of this approach is that it requires studying complete populations rather than samples, and missing data can be problematic. In the second strategy, which we will call the *ego network approach,* a full population is not necessary. A sample of focal nodes (called "egos") are selected, and the data collection process then uncovers the nodes they have ties with (called "alters"), attributes of these alters and, optionally, ties among the alters (as perceived by ego – the alters are not normally interviewed in an ego network design). The main difference between the ego network approach and the full network approach is that in the ego network approach, the alters are respondents in the study (unless by chance they happen to have been chosen as focal nodes). Thus, no overall picture of the network as a whole is generated, and the analysis proceeds by calculating only local network metrics which do not require knowing the full network. In general, full network studies are costlier to administer and as a result are much more limited in the number of different relations that they collect about (since each kind of tie forms an entire network). Ego network studies permit a great number of relations to be asked about, but are less rich in the sense that many of the conceptual tools of network analysis do not apply.

The choice of strategy depends on the nature of the research question. If the focus is on how attributes of nodes affect who forms ties with whom (Box 1 of Table 6.2), an ego network design is sufficient and the most cost-efficient. If the focus is on how position in the employee communication network affects node outcomes (Box 4 of Table 6.2), a full network design is necessary.

Data Collection

A best practice in network research methodology is called the ***ethnographic sandwich***. The ethnographic sandwich refers to sandwiching a formal, quantitative network study in between two layers of ethnographic work: one at the beginning of the study and one at the end of the study. At the beginning of a study, an ethnographic phase is used to (i) translate the research questions into concepts, terms, and data collection tasks appropriate to the research context, and (ii) to supplement preconceived theoretical frameworks with understandings derived inductively via participant observation. This initial ethnographic work helps ensure that the social relations being asked about are relevant to the population being studied and are expressed in ways that make sense to them. At the end of a study, a second ethnographic piece is often conducted in order to get reactions from the research subjects to the findings. An advantage of social network research is that one of the outputs – a network diagram of who is connected to whom – is easy for informants to understand and fun for them to think about. They can provide a great deal of insight into why the network has the shape it does, and what some of the consequences of that shape might be.

So in the front half of the sandwich the researcher tries to gain familiarity with the people, organizations, and context of interest. This can be done with the aid of semistructured interviews with those people to gain more insight about the relationships that they perceive to matter the most for them and their particular organizational context. Such familiarity enables the researcher to formulate more precise conceptions of the type of relations that have the most consequential effect for whatever organizational phenomenon of interest. Those more precise conceptualizations then allow for more precise measurement of those ties, adding rigor and strength to the methodology. Indeed, in both ego network and full network studies, a key question is which types of ties should be measured. In organizational research, it is common to obtain some kind of expressive relation, such as friendship or liking/disliking, and/or some kind of instrumental tie, such as "seeking advice from" or an "alliance partner with." In addition, researchers often control for positional relations such as "who reports to whom" and "who interacts with whom to get their work done." Ultimately, however, the choice of which ties to measure must depend on the research question and/or theoretical perspective, and on the specific social and cultural setting. Ethnographic work up front should guide those decisions.

Once some familiarity with the context is gained, social network research often proceeds with a survey, and the full and ego network approaches require somewhat different techniques. A typical ego network survey begins with what is called a "name generator," which is a series of open-ended questions aimed at eliciting a set of names (or nicknames) that a respondent (or set of respondents) has relationships with. Depending on the research objectives, the name generator can include questions like "Who do you go to the movies with?" "Which neighbors do you talk to?" The set of unique names generated in this part of the survey is then compiled into a roster. It should be pointed out that the names used by respondents need not be full names or even real names, as long as the respondent can remember which person each nickname referred to.

The second round of an ego network survey is called the "name interpreter." Here, the respondent is systematically asked to characterize their relationship with each alter mentioned in round 1. Some of the questions asked here (such as "Are you friends with this person?"; "Do you work with this person?"; "Does your organization get services from this organization?") may also have been asked in the name generator. The difference is that in the open-ended name generator section, the respondent might have omitted in the early questions some names that came up later. The name interpreter section uses a roster that requires the respondent to react to each name. The name interpreter section also asks about characteristics of the alter, such as how old they are, what race, which department, and so on.

Finally, and optionally, the respondent in an ego network study is asked about ties among the alters, using such questions as "Do Philip and Tommy know each other?" and asking this question for all pairs of alters. This is obviously highly time-consuming and therefore often omitted, but also yields very rich data.

In the full network approach, the initial step of a name generator is normally unnecessary, as the researcher already knows which actors she wants to study, and every one of these is approached to fill out a survey. Note that this does not mean that the respondent has to be given a roster of names to react to. For large networks, it might be impractical to list all possible names. In such cases, the respondent can be asked to provide names in an open-ended fashion. Ideally, however, a hybrid method should be used such that the user is basically confronted with an open-ended task but has a roster available to look up names. For electronic surveys, the name list can be provided as a drop-down menu or through browse/search button. The roster should also contain ancillary information about each name, such as their department, so that the respondent can identify which "Nancy Smith" she is referring to.

Unlike ego network studies, in full network studies there is usually no sampling – the researcher attempts to get data from every actor in the population frame. The result is a set of ties among all respondents. In practice, of course, it is difficult to obtain 100% response rates. Nonresponse creates a problem as each missing respondent corresponds to an entire row missing in the n-by-n data matrix that is built from the questionnaire data. A variety of techniques have been developed for coping with missing data in network analyses. For example, one can simply ignore nonrespondents as if they were not part of the population at all. This can reduce the value of the analysis, particularly when the missing nodes were of special interest. Another approach is to try to impute the missing ties. A number of sophisticated Bayesian methods have been developed to do this, but a simple approach is as follows. With networks that are known to be logically undirected (such as "interacted with"), the judgment of whether a tie exists between A and B is actually made by two separate observers, namely A and B. If A does not fill out the survey, we simply rely on B's judgment alone. This takes care of much of the missing data, but only works for ties between respondents and nonrespondents and not between nonrespondents and nonrespondents. It also does not work when the relation is not necessarily logically symmetric. For example, advice-giving in organizational settings is not necessarily symmetric. As a result, it does not make sense to fill in the (A,B) cell with the (B,A) information because they are not estimates of the same relation. For nonsymmetric relations, the best solution is to have asked the advice question two ways: who do you get advice from, and who do you give advice to. This provides two estimates, one from each node, of both the (A,B) advice tie and the (B,A) advice tie. If A does not fill out the survey, we can estimate whether A gives advice to B by looking to see if B has said she gets advice from A. Similarly, we can estimate whether A gets advice from B by seeing if B claims to give advice to A. A third approach to handling missing data is to ask actors to estimate the ties among all pairs of actors, so that A comments not only on her own ties, but B's ties with C. This yields NR estimates for each tie in the network, where NR is the number of actors who actually responded.

Despite the existence of these various options for coping with missing data, it is still advisable to do whatever is necessary to maximize the response rate. Survey research literature (e.g., Dillman *et al.* 2009) can offer insight into how to do so. Some general rules of thumb to follow include building as much rapport with the respondents as possible, showing individualized attention in letters or emails promoting the study, communicating the value and contribution of the study, and tirelessly hounding stragglers to get their surveys in.

Network data can also be obtained through other means, such as through archival and/or electronic sources. For example, telephone companies might disclose anonymized records of who has called whom among their customers. Organizations might provide records of who has e-mailed whom via the organization's servers. A great deal of data are also freely available on the Internet, particularly data in the form of affiliations – memberships of actors in groups or participation of actors in events. These data can be converted into actor by actor co-affiliations indicating the number of times that each pair of actors has been linked by comembership or coparticipation in groups or events. These are then used to infer social ties. Again, the choice of which networks to create from such sources should be guided by the research question.

Visualization

Once network data have been collected and formatted appropriately, analysis usually begins by creating a graphical representation of the network. This graphical representation allows the network analyst to qualitatively look for patterns in the structure of the network. Visualization tools, such as NETDRAW in the UCINET network analysis package (Borgatti *et al.* 2002), allow the researcher to change the appearance and create different visualizations of the network based on properties of the nodes and ties. Those different visualizations can allow for richer analysis and theorizing, as patterns and/or interesting observations can be uncovered and explored, and possibly explained and confirmed with further analysis. This ability to hide and reveal data with different visualizations allows for a fun way to understand what could be very complex social structures (Hanneman and Riddle 2005). Various characteristics of the points and lines that make up the illustrations, such as color, size, and shape, can be changed to communicate different information about the nodes and the relationships between them.

For example, Figure 6.5 depicts the friendship network of managers from a high-tech company (Krackhardt 1987). The nodes in the diagram are colored based on department. As one can see, managers within the same department tend to have friendships with each other more so than managers in different departments. It can also be observed that some individuals, like the node near the center of the diagram, tend to have friendship ties with managers in all of the different departments. Such observations might offer insight about the workings of the organization, and can be followed with the back end of the ethnographic sandwich. Individuals from the study can be interviewed, and richer insight can be gleaned about their structural positions.

Quantitative Analysis

Much of the quantitative part of SNA consists of measuring the constructs – at all three levels of analysis – mentioned in our discussion of network theorizing. Once

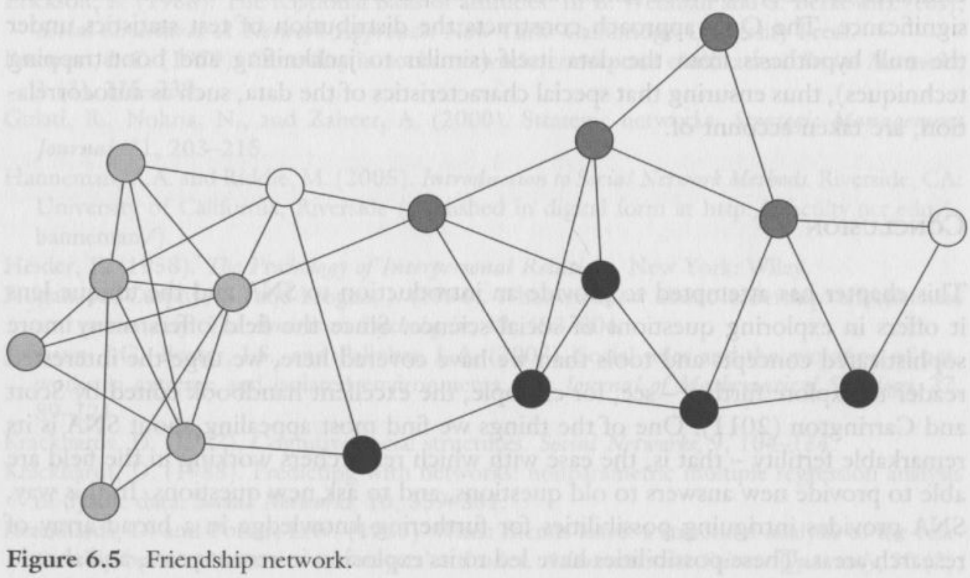

Figure 6.5 Friendship network.

such metrics are calculated, they then become variables in the researcher's database and can be used to test hypotheses regarding the antecedents and/or consequences of the social network concepts associated with each variable. Such testing of hypotheses proceeds, in a very general sense, as it does in any social science study. For example, we might theorize that teams with a more centralized structure make faster decisions. To test this, the researcher would collect network and performance data on a sample of teams, and then use ordinary least square (OLS) regression to relate team structure centralization to decision-making speed. For another example, if we hypothesize that more friendships within teams are associated with higher levels of overall team satisfaction, we can test this by studying a sample of, say, 100 teams, and for each one measuring the density of the team friendship network, and the average satisfaction level. We then correlate these two variables across teams. Or, at the node level, we might predict that a person's neuroticism score is negatively related to their centrality in a friendship network. This would be tested by (i) measuring each person's neuroticism using standard personality inventories, (ii) collecting the friendship network, (iii) constructing centrality measures for each person, and finally, (iv) statistically relating neuroticism to centrality (net of appropriate controls, of course).

However, there are a few caveats that should be kept in mind in analyzing relational data. Depending on how the data are collected and the level of analysis, network data can violate the assumptions of many inferential statistical techniques. For example, suppose we are testing a dyadic hypothesis: we believe that a friendship between two nodes facilitates knowledge sharing between those two nodes. To test this, we would want to correlate friendship ties with information sharing data. But the observations (which are dyads) are not independent because a given node is part of many dyads. Thus, if a node happens to be highly neurotic, it could affect every dyad in which it is involved, both in terms of friendship and knowledge sharing. One way to deal with such autocorrelation is with the use of the quadratic assignment procedure (QAP) family of methods (Krackhardt 1988). A QAP regression yields the same parameter estimates as OLS, but uses randomization or permutation tests to assess statistical significance. The QAP approach constructs the distribution of test statistics under the null hypothesis from the data itself (similar to jackknifing and bootstrapping techniques), thus ensuring that special characteristics of the data, such as autocorrelation, are taken account of.

Conclusion

This chapter has attempted to provide an introduction to SNA and the unique lens it offers in exploring questions of social science. Since the field offers many more sophisticated concepts and tools than we have covered here, we urge the interested reader to explore further – see, for example, the excellent handbook edited by Scott and Carrington (2011). One of the things we find most appealing about SNA is its remarkable fertility – that is, the ease with which researchers working in the field are able to provide new answers to old questions, and to ask new questions. In this way, SNA provides intriguing possibilities for furthering knowledge in a broad array of research areas. These possibilities have led to its explosion in recent years, and should continue to fuel its growth in the future.

NOTES

1 We refer here to *social* networks. We can of course define any relational system as a network, even if the nodes are not agentic entities, such as a network of words or concepts.

2 Despite the positive connotations of "gregariousness" and "popularity," we should keep in mind that whether these are "good" things or not depends on the kind of tie. If the relation measured is "dislikes," then a node with high gregariousness is one who dislikes many people and a node with high popularity is one who is disliked by many.

REFERENCES

Allen, T.J. (1977). *Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R&D Organization*. Cambridge, MA: MIT Press.

Bavelas, A. (1950). Communication patterns in task oriented groups. *The Journal of the Acoustical Society of America*, 22, 271–282.

Borgatti, S.P. (2005). Centrality and network flow. *Social Networks*, 27 (1), 55–71.

Borgatti, S.P. and Halgin, D. (2011,). Network theorizing. *Organization Science*, 22, 1168–1181.

Borgatti, S.P., Everett, M.G., and Freeman, L.C. (2002). *Ucinet 6 for Windows: Software for Social Network Analysis*. Havard, MA: Analytic Technologies.

Borgatti, S.P., Mehra, A., Brass, D., and Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323, 892–895.

Burt, R.S. (1992). *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Coleman, J.S., Katz, E., and Menzel, H. (1957). The diffusion of innovation among physicians. *Sociometry*, 20, 253–270.

Davis, G.F. (1991). Agents without principles? The spread of the poison pill through the intercorporate network. *Administrative Science Quarterly*, 36 (4), 583–613.

Dillman, D.A., Smyth, J.D., and Christian, L.M. (2009). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, NJ: Wiley and Sons.

Erickson, B. (1988). The relational basis of attitudes. In B. Wellman and S. Berkowitz (eds), *Social Structures: A Network Approach*. New York: Cambridge University Press.

Freeman, L.C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, 1 (3), 215–239.

Gulati, R., Nohria, N., and Zaheer, A. (2000). Strategic networks. *Strategic Management Journal*, 21, 203–215.

Hanneman, R.A. and Riddle, M. (2005). *Introduction to Social Network Methods*. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/).

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.

Hogan, R., Curphy, G., and Hogan, J. (1994). What we know about leadership: effectiveness and personality. *The American Psychologist*, 49, 493–504.

Johnson, J.C., Boster, J.S., and Palinkas, L.A. (2003). Social roles and the evolution of networks in extreme and isolated environments. *The Journal of Mathematical Sociology*, 27, 89–121.

Krackhardt, D. (1987). Cognitive social structures. *Social Networks*, 9, 104–134.

Krackhardt, D. (1988). Predicting with networks: nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10, 359–381.

Krackhardt, D. and Porter, L.W. (1985). When friends leave: a structural analysis of the relationship between turnover and stayers' attitudes. *Administrative Science Quarterly*, 30 (2), 242–261.

Leavitt, H.J. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal Psychology*, 46, 38–50.

Marsden, P.V. (1987). Core discussion networks of Americans. *American Sociological Review*, 52, 122–131.

Marsden, P.V. (1990). Network data and measurement. *Annual Review of Sociology*, 16, 435–463.

McPherson, M., Smith-Lovin, L., and Cook, J.M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 27, 415–444.

Mehra, A., Kilduff, M., and Brass, D.J. (2001). The social networks of high and low self-monitors: implications for workplace performance. *Administrative Science Quarterly*, 46, 121–146.

Putnam, R.D. (2000). *Bowling Alone: The Collapse and Revival of American Community.* New York: Simon and Schuster.

Scott, J. and Carrington, P. (2011). *The Sage Handbook of Social Network Analysis.* Thousand Oaks, CA: Sage Publications.

Umphress, E., Labianca, G., Brass, D.J., *et al.* (2003). The role of instrumental and expressive social ties in employees' perceptions of organizational justice. *Organization Science*, 14, 738–753.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, 5, 171–180.