1
Introduction

Learning Outcomes

1 Understand what constitutes a social network
2 Identify and describe different levels of analysis
3 Formulate problems in terms of network variables

1.1 Why networks?

An obvious question to ask is why anyone would want to analyze social network data. The incontestable answer, of course, is because they want to. But what are some sensible-sounding reasons that a researcher could use in polite company? One is that much of culture and nature seems to be structured as networks – from brains (e.g., neural networks) and organisms (e.g., circulatory systems) to organizations (e.g., who reports to whom), economies (e.g., who sells to whom) and ecologies (e.g., who eats whom). Furthermore, a generic hypothesis of network theory is that an actor's position in a network determines in part the constraints and opportunities that he or she will encounter, and therefore identifying that position is important for predicting actor outcomes such as performance, behavior or beliefs. Similarly, there is an analogous generic hypothesis at the group level stating that what happens to a group of actors is in part a function of the structure of connections among them. For example, a sports team may consist of a number of talented individuals, but they need to collaborate well to make full use of that talent.

1.2 What are networks?

Networks are a way of thinking about social systems that focus our attention on the relationships among the entities that make up the system, which we call actors or

1

nodes. The nodes have characteristics – typically called 'attributes' – that distinguish among them, and these can be categorical traits, such as being male, or quantitative attributes, such as being 56 years of age. The relationships between nodes also have characteristics, and in network analysis we think of these as ties or links. Thus, the relationships between Bill (male, 47 years old) and Jane (female, 43 years old) may be characterized by being married, living together, co-owners of a business, and a multitude of other social ties. These relational characteristics can also be continuously or ordinally valued, as in having known each other for 12.5 years and having fights 3–5 times a year.

Of special interest in network analysis is the fact that ties interlink through common nodes (e.g., the A→B link shares a node in common with the B→C link), which creates chains or paths of nodes and links whose endpoints are now connected indirectly by this means. This in turn creates the connected web that we think of as a network.[1] Part of the power of the network concept is that it provides a mechanism – indirect connection – by which disparate parts of a system may affect each other.

The nodes in a network can be almost anything, although when we talk about *social* networks we normally expect the nodes to be active agents rather than, say, inanimate objects.[2] Most often, nodes are individuals, such as individual persons or chimpanzees. But they can also be collectivities, such as teams, firms, cities, countries or whole species.

Whether actors are collectivities or individuals should not be confused with levels of analysis. In network analysis, it is useful to distinguish between three levels of analysis: the dyad, the node, and the network (see Table 1.1). At the dyad level of analysis, we study pairwise relations between actors and ask research questions like 'do pairs of actors with business ties tend to develop affective ties?'. The dyad level is the fundamental unit of network data collection, and is the unit with the greatest frequency. In Table 1.1, the notation $O(n^2)$ indicates that the number of dyads in a network is of order n^2, where n is the number of nodes in the network.[3] At the node level of analysis, we ask questions like 'do actors with more friends tend to have stronger immune systems?'. Most network-level node properties are aggregations of dyad-level measurements, as when we count the number of ties that a node has. The number of nodes in the network is, of course, of order n.

---

[1] However, it should be understood that we do not require a network to be connected, nor to have any ties at all. This is important when analyzing networks over time, as initially a set of actors (say, a new task force charged with investigating unethical behavior in an organization) may have no ties at all to each other, but will develop ties over time. If the data is collected over time, we will be able to see the network connect up.

[2] But this gets more complicated in the case of two-mode networks. See Chapter 13 for more on this.

[3] The use of this notation to represent levels of analysis is due to David Krackhardt (personal communication).

Table 1.1   Examples of research questions by level of analysis and type of node.

Level of analysis	Type of node	
	Individuals	**Collectivities**
Dyad level $O(n^2)$	Are employees whose offices are near each other more likely to develop friendships than employees whose offices are further apart?	Are firms with similar organizational cultures more likely to form joint ventures with each other?
Node level $O(n^1)$	Are employees who are more central in their organization's friendship network less likely to leave for another company?	Are firms with more diverse technology partners more likely to introduce innovative products into the market?
Network level $O(n^0)$	When a network of employees is characterized by many redundant paths between all pairs of persons, is the network less disrupted by individuals leaving the firm?	When a network of firms is densely connected, does this place the network at greater risk of catastrophic failure (because of cascade effects)?

At the network level, we ask questions like 'do well-connected networks tend to diffuse ideas faster?'. The number of objects at this level of analysis is of order n^0, which is to say, 1. This means, for example, that if we have a friendship network, a variable at this level of analysis will consist of a single number that characterizes the network as a whole (e.g., how densely connected it is). Note that at each level of analysis, the actors could be individuals or collectivities, as shown in Table 1.1.

It is worth noting that the 'micro' versus 'macro' terminology used in many of the social sciences can refer to either the rows or the columns of Table 1.1. For instance, in the management literature, micro refers to studies in which the cases are persons and macro refers to studies in which the cases are firms. But in economics, it is more common to use micro to refer to the study of actor-level behavior (whether the actors are individuals or firms) and macro to refer to studies of the economy as a whole (i.e., the network level of analysis).

1.3 Types of relations

Relations among actors can be of many different kinds, and each type gives rise to a corresponding network. So if we measure friendship ties, we have a friendship network, and if we also measure kinship ties among the same people, we have both a friendship network and a kinship network. In the analysis we may choose to combine the networks in various ways, but in reality we have two networks. Perhaps the most commonly studied ties for persons are friendship ties, advice giving, communication and, the most basic of all, simple acquaintanceship (who knows whom). Acquaintanceship is especially important in large networks, such as a firm of 160,000 employees or society as a whole. The latter is the basis for

the famous Milgram (1967) 'six degrees of separation' study. The process of how individuals become acquainted has been the subject of considerable research, including Newcomb's (1961) seminal book, *The Acquaintance Process*.

Table 1.2 provides a useful taxonomy of types of ties among persons. Inspired by Atkin's (1977) distinction between backcloth and traffic, the principal division in the table is between the relational states (on the left) and the relational events (on the right). Relational states refer to continuously persistent relationships between nodes, such as being someone's brother or friend. 'Continuously persistent' does not mean that the relationship will never end, but rather that, while it does exist, it exists continuously over that time. This contrasts with relational events, such as selling a house. Although the process may take months to execute, the concept of a sale is a discrete event. (Of course, we can always define a relational state based on a relational event simply by casting it in a timeless way. For example, if Bill sells a house to Jim, it is an event, but the relation 'has ever received a house from' is a state.) Events that recur can also be counted up, as in the number of emails that X sent to Y last month. We often use recurring relational events as evidence of an underlying relational state, as in assuming that a frequent lunch partner is a friend. We may also regard recurring events as antecedents of relational states, so that if we frequently have lunch together (perhaps for work-related reasons), we may develop a friendship. It is difficult to develop friendships without any interactions at all.

Within relational events, the table distinguishes between interactions and flows. Interactions are behaviors with respect to others and often observable by third parties. Flows are the outcomes of interactions, and interactions form the medium that enables things to flow. Flows may be intangibles, such as beliefs, attitudes, norms, and so on, that are passed from person to person. They can also consist of physical resources such as money or goods.

Within relational states, the table distinguishes between similarities, relational roles and relational cognition. Taking these in reverse order, relational cognition refers to thoughts and feelings that people have about each other. This includes acquaintance – who knows whom. Relational cognitions are essentially unobservable by other network members except as inferred from interactions.

Table 1.2   Taxonomy of types of relations.

Relational states							Relational events	
Similarities			Relational roles		Relational cognition			
Location	Participation	Attribute	Kinship	Other role	Affective	Perceptual	Interactions	Flows
Same spatial and temporal space	Same clubs, same events	Same gender, same attitude	Mother of, sibling of	Friend of, boss of, student of, competitor	Likes, hates	Knows, knows of, sees as happy	Sold to, talked to, helped, fought with	Information, beliefs, money

A highly consequential example of relational cognition is the trust relation, which can determine whether transactions will take place, and at what cost.

The relational roles category includes some of the most permanent of human relations, such as 'parent of' and 'sibling of'. Typically, the persons we have these relationships with are named or categorized by the relationship. Hence the person we have a friendship tie with is called a friend and is seen as enacting the friend role. When these relationships are asymmetric (such as 'mother of'), our culture typically provides us with named reciprocal roles. Hence we have parents and children, students and teachers, bosses and subordinates, and so on.

The similarities category refers to relational phenomena that are not quite social ties but can be treated as such methodologically, and which are often seen as both antecedents and consequences of social ties. For example, physical proximity (i.e., similarity in physical location) provides opportunities for face-to-face interactions. At the same time, certain social relations (e.g., romantic) often lead to radical increases in proximity (as in moving in together). Co-membership in groups (such as universities, gyms, teams, workplaces) provides many opportunities for interaction. Co-participation in events (such as attending the same conference or the same political rally) also provides opportunities for interaction. We can also define similarities in terms of attributes of nodes, such as gender and race. An enduring finding in social psychology is homophily – the tendency for people to like people who are similar to themselves on socially significant attributes.

It is worth pointing out that when nodes are collectivities, such as firms, there are two different kinds of ties possible. First, there are ties among the firms qua firms – that is, ties that are explicitly between the firms as single entities, such as a joint venture between two firms, an alliance, a purchase agreement, and so on. Second, there are ties between the individual members of the firms. Even though these are not 'official' ties between the organizations, they may serve all the same functions. For example, if the chief executive officers of two companies are friends, they may well share considerable information about each other's organization, constituting a flow of information between the firms. Table 1.3

Table 1.3 Relations among firms.

Type	Firms as entities	Via individuals
Similarities	Joint membership in trade association; co-located in Silicon Valley	CEO of organization A sits on same board as CEO of organization B
Relations	Joint ventures; alliances; distribution agreements; own shares in; regards as competitor	Chief scientist of A is friends with chief scientist of B
Interactions	Sells product to; makes competitive move in response to	Representatives of A attend same conference as representatives of B
Flows	Technology transfers; cash infusions	Employee of A leaks information to employee of B

provides examples of both kinds of ties among firms, cross-classified using the typology in Table 1.1.

1.4 Goals of analysis

Network analyses can be applied or basic.[4] By 'applied' we mean that the study consists of calculating a number of metrics to describe the structure of the network or capture aspects of individuals' positions in the network. The results are then interpreted and acted upon directly. For example, in an applied setting such as public health, we might use a centrality analysis of a network of drug addicts to detect good candidates for costly training in healthful practices, with the hope that these individuals would then diffuse the practices through the network. Or in management consulting, we might detect groups of employees from one organization in a merger situation who are not integrating well with the other company, and create some kind of intervention with them. Applied studies are basically univariate in the sense that the variables measured are not correlated with each other. Rather, the correlations are assumed – because they have been observed or deduced in other, basic, research. For example, in the drug addict case, we choose to identify central players because previous research has suggested that getting central players to adopt a behavior will have add-on effects through diffusion to others. The causal relationships have been established, so we need only measure the predictor variables.

In contrast, basic research studies are multivariate and correlative – they try to describe the variance in certain variables as a function of others. The objective is to understand the dependent variables (i.e., outcomes) as the result of a causal process acting on a set of starting conditions. The independent variables serve to capture the initial conditions as well as traces of the theorized process. These are the kinds of studies we usually see in academic research. The function of network analysis in these studies is often to generate the variables that will be correlated, either as independent/explanatory variables or as dependent/outcome variables. As an example of the former, we might construct a measure of the centrality of each actor in a network, and use that to predict each actor's ability to get things done (i.e., their power). Studies of this type seek to create a network theory of ____, where we fill in the blank with the dependent variable, such as aggression or status attainment, yielding a 'network theory of aggression' or a 'network theory of status attainment'. As an example of using network variables as dependent variables, we might use the similarity of actors on attitudinal and behavioral variables (e.g., political views and smoking behavior) to predict who

---

[4] Some might use 'descriptive' or 'explanatory', but explanation is theory and a theory is a description of how a system works.

Table 1.4 Types of network studies classified by direction of causality and level of analysis.

	Network variables as independent/explanatory	Network variables as dependent/outcomes
Dyad level	Friendship between pairs of farmers to predict which pairs of farmers make the same decision about going organic	Similarity of interests (e.g., sky diving) to predict who becomes friends with each other
Node level	Centrality in organizational trust network to predict who is chosen for promotion	Extraversion to predict who becomes central in friendship network
Network level	Shortness of paths in a group's communication network to predict group's ability to solve problems	Type of organizational culture (emphasizing either cooperation or competition) to predict structure of the trust network

becomes friends with whom. Studies of this type seek to generate a _____ theory of networks, where we fill in the blank with a mechanism relating to the independent variables, such as a 'utility-maximization theory of network tie formation' or a 'balance theory perspective on network change'.

Whether we use network variables as the independent variables in our analyses or as the dependent variables, they can be at any of the three levels of analysis discussed earlier. Table 1.4 gives examples of studies representing six possible combinations.[5]

1.5 Network variables as explanatory variables

When network variables are used as independent variables, the researcher is implicitly or explicitly using network theory to explain outcomes. These outcomes can be highly varied given that networks are studied in so many different fields – anything from individual weight gain to firm profitability. But because network processes are being used to explain these outcomes, there is a certain amount of unity in the logic that is used to predict the outcomes.

Most network theorizing is based on a view of ties as conduits through which things flow – material goods, ideas, instructions, diseases, and so on. Atkin (1977) referred to this as the backcloth and traffic model, where the backcloth is a medium, like a road system, that enables some kind of traffic to flow between locations. Within this basic conception, however, there are many different mechanisms that have been proposed to relate flows to outcomes. To discuss these, it

---

[5] For simplicity, the table excludes cases where network variables are both the independent and dependent variables, as when friendship ties are used to predict business ties, or one kind of node centrality is used to predict another.

is helpful to classify the outcomes being studied into a few broad categories. One basic category of outcomes consists of some sort of achievement, performance or benefit, either for individual nodes or for whole networks. Studies of this sort are known as social capital studies. An example is social resource theory (Lin, 2001), which argues that a node's achievement is in part a function of the resources that their social ties enable them to access. Thus, an entrepreneur who is well connected to people who control a variety of important resources (e.g., money, power, knowledge) should be better positioned to succeed than one who has only his own resources to draw on. Thus the key here is the inflow of resources that the entrepreneur's ties afford him. Another perspective, which we refer to as *arbitrage* theory,[6] argues that a node B can benefit if it has ties to A and C, who are otherwise unconnected and who have achieved differing levels of progress toward a common goal. For example, if C has already solved something that A is still struggling with, B can make herself useful by bringing C's solution to A (for a price!). Here, the benefit is derived from the combination of an inflow and an outflow. Yet another network mechanism linking networks to achievement is *auctioning*. Here, if B has something that both A and C need, B can play them off against each other to bid up the price or extract favors from them both. In this case, the benefit comes from the potential outflow from B to her customers. In all of these cases, achievement is some sort of function of social ties. That is, the structure of the network and the position of individual nodes within it are crucial factors in predicting outcomes. This is very clear in the last two examples, in which a node B occupies a position between two others. But it is also true of the first case (social resource theory), because the resources of a node's connections may themselves be a function of their connections.

Another basic category of outcomes is what we might call 'style'. Unlike achievement, where one outcome is 'better' than another, style is about choices. Studies in this category look at things like political views, decisions to adopt an innovation, acquisition of practices and behaviors, and so on. These outcomes are often phrased in dyadic terms, so that what we are trying to explain is why, say, two firms have adopted similar internal structures, or why two people have made the same decision on the kind of smartphone to buy. The classic network explanation for these observed similarities is diffusion or influence. Through interactions, actors affect each other and come to hold similar views or become aware of similar bits of information. This is a perspective that clearly stems from a view of ties as conduits. But it is not necessarily the case that node A resembles node B because they influenced each other. It could be that a third party is tied to both of them and is influencing them both. It could also be something more subtle. For example, consider predicting employees' reaction to their phone ringing.

---

[6] Arbitrage is our term for one specific mechanism in Burt's (2004) discussion of brokerage.

Suppose some people cringe when it rings and others enjoy it. It could be that the people who cringe are those who are highly central in the advice network, meaning that lots of people are constantly calling to get their help, and this gets annoying. Notice it is not that the central people are infecting each other with a bad attitude toward the phone, or even that third parties are infecting both of them with that attitude. It is a reaction that both have to the same situation, namely receiving so much flow. Essentially, the argument is that nodes are shaped by their social environments, hence nodes that have similar environments (such as both being central) will have similar outcomes.

1.6 Network variables as outcome variables

It is often asserted that there is more research examining the consequences of network variables than the antecedents. This could be true, but it could also be a misperception due to the fact that the various factors that impinge on network variables will come from a wide variety of different fields and will not have any particular theoretical unity. This is especially clear when you consider that the network properties being explained can be at different levels of analysis (i.e., the dyad, the node and the whole network), and that they may not be talked about using network terms. For example, there is a large and venerable literature on the acquaintance process (Newcomb 1961) that never uses the term 'network'.

One of the oldest and most frequently replicated findings in social psychology is homophily – the tendency for people to have positive ties to those who are similar to themselves on socially significant attributes such as gender, race, religion, ethnicity and class. One way of thinking about these findings is in terms of a logistic regression in which the cases are dyads, the dependent variable is whether or not the nodes in the dyad have a positive tie, and the independent variables are things like *samegender* (a variable that is 1 if the nodes in the dyad are the same gender and 0 otherwise) and *agediff* (the absolute value of the difference between their ages).[7] In predicting most kinds of positive ties (but not marriage or other romantic relationships) we find a positive coefficient for *samegender* and a negative coefficient for *agediff*.

It is worth noting that having positive ties with people similar to oneself need not be solely the result of a preference. It could also reflect the availability of suitable partners. For example, if most people in an organization were women, we would expect most of these women's work friends to be women as well, simply because of availability. We would not want to conclude from such data that men are homophilous whereas women are not. One of the historical roots of

---

[7] See Chapter 8 for a discussion of how to deal with issues of non-independence of observations that arise in an analysis of this type.

social network analysis is in structuralist sociology, which, in the name of parsimony, urges us to seek answers in opportunities and constraints before turning to preferences.

This suggests two basic types of factors in tie formation – opportunity and preference – and these are often intertwined. As an example of an opportunity-based mechanism, another well-known finding in the literature is that one tie leads to another. For example, business ties can lead to friendship ties, and vice versa. The presence of one tie sets up the opportunity for another kind of tie to form. More generally, as discussed in the third section of this chapter, we often expect relational states like friendship to lead to interactions (e.g., talking) through which things like information can flow, and which in turn can change the relationship (e.g., sharing intimacies deepens the relationship).

An example of a preference-based mechanism is balance theory (Festinger 1957; Heider 1958). In this theory a person tries to be congruent with those she likes. So if Jane likes Sally, and Sally likes Mary, it would cause Jane cognitive dissonance to dislike Mary. So based on balance theory, we would expect either that Jane's estimation of Mary would rise, or her estimation of Sally would decline. Note that an opportunity-based perspective would also predict the development of a positive tie between Jane and Mary because both of them are friends with Sally, and Sally might well invite both to the same event, or at least make each of them aware of the other.

1.7 Conclusion

Network analysis is about structure and position. To this end, the field has developed an impressive array of concepts to characterize position and structure. In large part, the field has been able to express these concepts formally (i.e., in mathematical terms). This is a huge advantage because it means we can program computers to detect and measure these concepts in data, which in turn allows us to test hypotheses empirically. One downside, however, has been that some social scientists, unfamiliar with formal theorizing, have misconceived of the field as a methodology. It does indeed have a distinctive methodology that is born of its fundamentally relational view of social phenomena. But the theoretical concepts that are so emblematic of the field, such as centrality and structural equivalence, are just that: theoretical concepts that are part of a distinctive approach to explaining the social world (Borgatti and Halgin 2011).