Chapter 1
Introduction

Welcome to the world of social network analysis. This book is intended as a general introduction the analysis of social networks. As such, the focus is on the concepts and methodology of social network analysis. It is about how to actually analyze network data. Of course, since methodology and theory are deeply intertwined, this book is also about network theory. What it is not is a literature review of empirical work on social networks.

The book is also not meant to be technically rigorous. We try not to simplify to the point of misleading, but when forced to make a choice, we have opted for intelligibility and transmitting the spirit of an idea rather than be right in every nuance. Each case, however, we provide pointers to the appropriate technical literature so that the reader can get the full picture if desired.

Doing network analysis implies using network analysis software. A number of packages exist, such as UCINET and Pajek. Since the authors of this book include the authors of UCINET, we use UCINET for all examples in the book. However, we wanted the book to be more than a tutorial on UCINET. Therefore, we focus on generic data analysis issues and give UCINET-specific information only in separate inserts giving examples using real data.

Social networks are studied in many traditional academic departments, including virtually all of the social sciences, physics, biology and the professions, such as medicine and management. As a result, this book will draw examples from a number of different fields.

Why Networks?

An obvious question to ask is why anyone would want to analyze social network data. The real answer, of course, is because they want to. But what are some sensible-sounding reasons that a Ph.D. student can use in polite company? One reason is that much of culture and nature seems to be structured as networks. From brains (e.g., neural networks) to organisms (e.g., circulatory systems) to organizations (e.g., who reports to whom) to economies (e.g., who sells to whom) to ecologies (e.g., who eats whom). Furthermore, there is a generic hypothesis in network theory that an actor's position in a network determines in part the constraints and opportunities that they will encounter, and therefore identifying that position is important for predicting actor attributes such as performance or behavior or beliefs. So for example, There is also a group-level corollary that states that what happens to a group of actors is a function of the structure of connections among them.

What are Networks?

Networks consist of a set of actors (also called nodes or, in the technical literature, vertices or points) together with a set of ties (also called links or, in the technical literature, edges, arcs or lines) that link pairs of actors. The fact ties can share actors (e.g., the A→B link shares an actor with the B→C link) creates the connected web that we think of as a network. [1]

In addition, we normally assume that actors have attributes, which

The ties that link actors can be of many different kinds. Perhaps the most commonly studied are friendship ties, advice giving, communication and, the most basic of all, simple acquaintanceship (who knows whom). Figure xx provides a taxonomy of types of ties. The principal division is between the relations, on the left, and the events, on the right. The basic difference is that relations are potentially enduring, whereas events are necessarily transitory. In addition, the distinction corresponds loosely to the difference between roads and traffic (or backcloth and traffic, as it was famously put by Atkin (19xx)). Roads are like conduits that enable and (in their absence) constrain the flow of traffic.

Among the types of ties we have called relations, we have further subdivided them into (i) role-based relations (sometimes called "proper social relations") such as "is a friend of" and "reports to", (ii) affective or attitudinal relations such as likes, dislikes, or trusts, and (iii) perceptual relations such as "is acquainted with" or "knows the area of expertise of"

Among the ties we have labeled events, we have two broad types. First we have the interactions and transactions, which include such things as "had sex with", "talked to", "gave advice to", "sold to" and so on. These are transitory events that can be counted up to obtain a frequency over a certain period, as in the number of times that a pair of actors had lunch with each other. These frequencies are often regarded as a proxy for some underlying relation, such as friendship, because the events are enabled by such underlying relations. In this vein, we can also explicitly ask respondents to report on "who do you give advice to" or "who do you talk in a typical day", both of which are ways to convert something transitory into a more relation-like tie. (At the same time, when measuring relations like friendship, we often assume that interactions such as talking, take place.)

---

[1] Although it should be understood that, from a methodological point of view, a network does not have to be connected. Even a network that has no ties (yet) is still a network, just a very very sparse one. 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 you collect data over time, you will see the network connect up.

From Borgatti, Everett, Johnson and Stevenson. 2008. Analyzing Social Networks. Publisher TBA.
Please do not distribute.

The other kinds of events are the flows. These are the traffic that is moved during an interaction . These are things like gossip or resources or viruses. We might think of them as streams of particles or bits of information that move from one actor to the other.

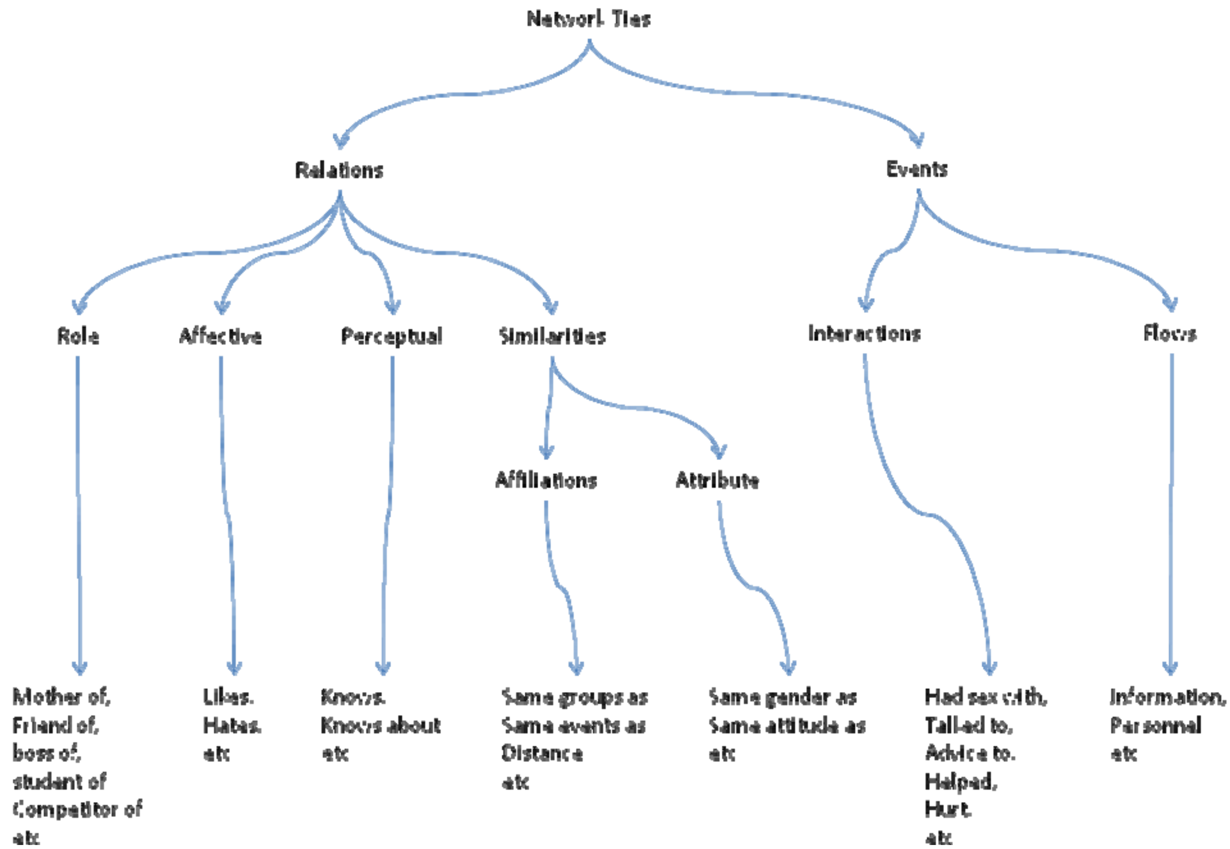

Figure 1. Taxonomy of types of tie forming networks.

[need to remove similarities for the table later to make sense]

One class of relational data that we have left out of Figure 1 are similarities. Similarities are measures we can calculate for pairs of actors, much like social ties, but which are not, strictly speaking, social ties. Because they are dyadic we can conveniently analyze them using the methods of social network analysis. Also, similarities may be highly correlated with social ties, and could therefore be used as proxies for ties. One particularly important type of this kind of tie is co-memberships, such as co-membership in groups (e.g., belonging to the same corporate boards, or working on the same client projects), or in events (e.g., attending the same parties). In many cases, we assume that comembership implies social ties such as knowing each other or communicating with each other. Of course, as groups or events get large, this assumption becomes increasingly dubious.

Another type of similarity is based on a single attribute of an actor, such as a belief (e.g., the world was created 6 thousand years ago) or behavior (e.g., smoking). For any categorical attribute, we can construct a dyadic variable that indicates, for each pair of actors, whether they have the same attribute of not. For continuous attributes, such as age, we can calculate the difference in attribute values for each pair of actors. [show figure] More generally, if we can describe each actor in terms of a profile of scores across variables (e.g., a battery of attitude variables), we can correlate the profiles and create, for each pair of actors, a correlation that indicates the overall similarity in their attitudes.

It should also be mentioned that some kinds of ties are logically symmetric, while others are not. For example, "had discussions with" would normally be considered a symmetric type of tie because if A had discussions with B, then clearly B had discussions with A. On the other hand, the relation "gives advice to" is not necessarily symmetric. Just because I give you advice doesn't mean you necessarily give me advice, though it might happen. Some kinds of ties are not just non-symmetric, they are anti-symmetric, which means they can't be reciprocated. An example is the "is the grandfather of" relation: Except in a old folk song, if A is the grandfather of B, then B cannot be the grandfather of A.

Of course, kinds of tie that are logically symmetric or anti-symmetric may not be so in the actual data. This can be because of data entry errors, or missing data, or, in survey data, respondent recall errors (e.g., A mentions his friend B, but B forgets to mention A), and differing interpretations (e.g., what is a friend?). In practice, what A regards as a conversation may not be what B regards as a conversation.

Network Variables and Cases

Ultimately, all quantitative research comes down to correlating variables measured on cases. What distinguishes different fields are the kinds of variables measured, and what the units are that these variables are measured on – the cases. In network analysis, the fundamental unit of data (but not necessarily the analysis) is the dyad – a pair of actors. When we measure a social tie, such as friendship, we assign a value to each pair of actors. Often, the value is 1/0, with a 1 indicating the presence of a tie, and a 0 indicating the absence of a tie. However, it can also be any other kind of value, such as an ordinal ranking like "not at all friendly", "somewhat friendly", and "very friendly", or even a ratio-level quantity such as the number of times the pair talked on the phone. Note that for each pair of actors, we can simultaneously measure many different kinds of ties, giving us multiple dyadic variables measured on the same set of cases, just as in non-network research we would normally measure several variables on the same sample of persons. Or we can measure the same kind of tie at different points in time. Figure xx shows one way to record such data.

Actor 1	Actor 2	Friends	Phone Calls	Distance
001	002	1	27	3
001	003	0	0	NA
001	004	0	1	5
002	003	0	1	5
002	004	0	0	6
003	004	1	10	2

Figure xx. Network with 3 kinds of ties among 4 actors.

One way network analysis can proceed is very much like in conventional monadic data analysis, in which we use certain variables to predict others. For example, we may be interested, in whether certain kinds of ties create the conditions necessary for other kinds of ties, as in social and family ties paving the way for business ties. For the data in Figure xx, we might use distance to predict the number of phone calls between pairs.

However, just because the raw data are dyadic doesn't mean we have to analyze the data at the dyadic level. Once we have built the network, we can construct actor-level indices such as the number of ties that each actor has, and correlate these actor-level variables with other variables we may have collected on the actors, such as their age, performance, beliefs, and so on. Nor do we have to stop at the actor level. Consider, for example, a network analysis of a corporation that has multiple departments. For each department, we could count up the proportion of pairs of actors that have a certain kind of tie, such as trust. This yields a department-level variable in which we have estimated the level of trust among members. We can then correlate this with other department-level variables, such as the compensation and incentive system used, or the importance of the department's function. Similarly, we could collect trust data on the top management teams of hundreds of corporations, calculate the proportion of team members who trust each other, and relate this proportion to organizational profit levels.

Goals of Analysis

Network analyses can be descriptive or explanatory. By descriptive, we mean that purpose of the analysis is to describe the structure of the network and to measure different 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 network analysis of a network of drug addicts to detect good candidates for diffusion healthful practices such as bleaching needles. Or in management consulting, we might detect groups of employees in a merger situation who are not integrating well with the other company, and create some kind of intervention with them.

In contrast, explanatory studies are basically correlative, trying to understand the distribution of certain variables as a function of others. These are the kinds of studies we usually see in academic research. The function of network analysis in these studies is 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 a network theory of turnover, 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 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 maximizing theory of network tie formation.

Whether we use network variables as the independent variables in our analyses or as the dependent variables, the variables can be at any of the three levels of analysis discussed earlier. Table xx gives examples of studies representing six possible combinations.[2]

	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
Actor level	Centrality in organizational trust network to predict who is chosen for promotion	Extraversion to predict who becomes central in friendship network
Group level	Shortness of paths in a group's communication network to predict group's ability to solve problems	Heterogeneity with respect to respect to national culture to predict proportion of actors in multinational group who have conflicts

Figure x.

Most work to date has been in the first column, using network variables as explanatory. The dependent variables in such cases are widely varied since they come from many fields. However, it is useful to note that a number of studies of this type basically seek to provide network explanations of some sort of performance or success. We refer to these as "social capital studies". The essential idea of social capital studies is that ties have benefits that actors or groups of actors can exploit for gain. For example, a social

[2] 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.

capital study might propose that having many weak ties gives a person access to novel information which in turns enables them to outperform those with less information.

Ego Network versus Full Network Analysis

When we talk about networks, we almost always mean full networks – the set of ties among a set of actors. However, for both practical and theoretical reasons, we don't necessarily have to collect data on the entire network in order to accomplish analytical goals. Particularly for large networks (such as the set of US consumers), we often use what are called ego networks instead. An ego network is defined as consisting of a focal actor called ego, along with a set of actors with whom ego has some kind of tie, called alters, and optionally a set of ties linking the alters to each other.