Basic Concepts

In: Social Network Analysis and Education: Theory, Methods & Applications

By: Brian V. Carolan
Pub. Date: 2013
Access Date: July 6, 2021
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks
Print ISBN: 9781412999472
Online ISBN: 9781452270104
DOI: https://dx.doi.org/10.4135/9781452270104
Print pages: 43-64

Basic Concepts

Objectives

Building on the introductory concepts briefly noted in Chapter 1 as well as the theoretical and methodological breakthroughs summarized in Chapter 2, this chapter reviews the basic means through which network data and their characteristics at either the ego or complete network level are represented. As such, this chapter provides you with an overview of the foundational concepts that are required building blocks for subsequent chapters. Using a number of graphs and matrices, this chapter emphasizes how network data differ from traditional social science actor-by-attribute data while also sharing some similarities. Because others have covered this material extensively in more comprehensive texts, the emphasis is on directing you to these other sources. To illustrate these basic concepts and provide you with opportunities to work with these concepts, this chapter focuses on one network data set: Pittinsky's Middle School Science Classroom Friendship Nominations (referred to as the “Peer Groups” data set).

Revisiting Key Assumptions

Recall that Freeman's (2004) four hallmarks of contemporary social network analysis include: (1) an emphasis on structuralism based on ties among actors; (2) firmly grounded in empirical data; (3) the use of graphical imagery; and (4) is mathematically based. This chapter incorporates all four of these features by introducing the ways in which graph theory and sociometry are used to mathematically and visually represent the ties among actors that are embedded in an empirical context. The distinctions between these mathematical tools have blurred as computers have assumed a bulk of the analytical responsibilities. The reason for preferencing sociometric over graph theoretic notation will become more evident as the chapter progresses, but suffice it to say that it is the easiest, and perhaps oldest, way to denote properties of social networks.

Regardless of how networks are represented, the collection and analysis of social network data is predicated on three assumptions offered by Knoke and Yang (2008) and noted in the preceding chapter. First, relations are critically important when attempting to explain one's behaviors or attitudes. Second, social networks affect one's behaviors and attitudes through a variety of direct and indirect contacts. Third, relations within and between networks are dynamic. Keeping these three assumptions in mind, this chapter provides an introduction to how the core concepts of social network analysis can be represented and how these representations allow us to see things more clearly.

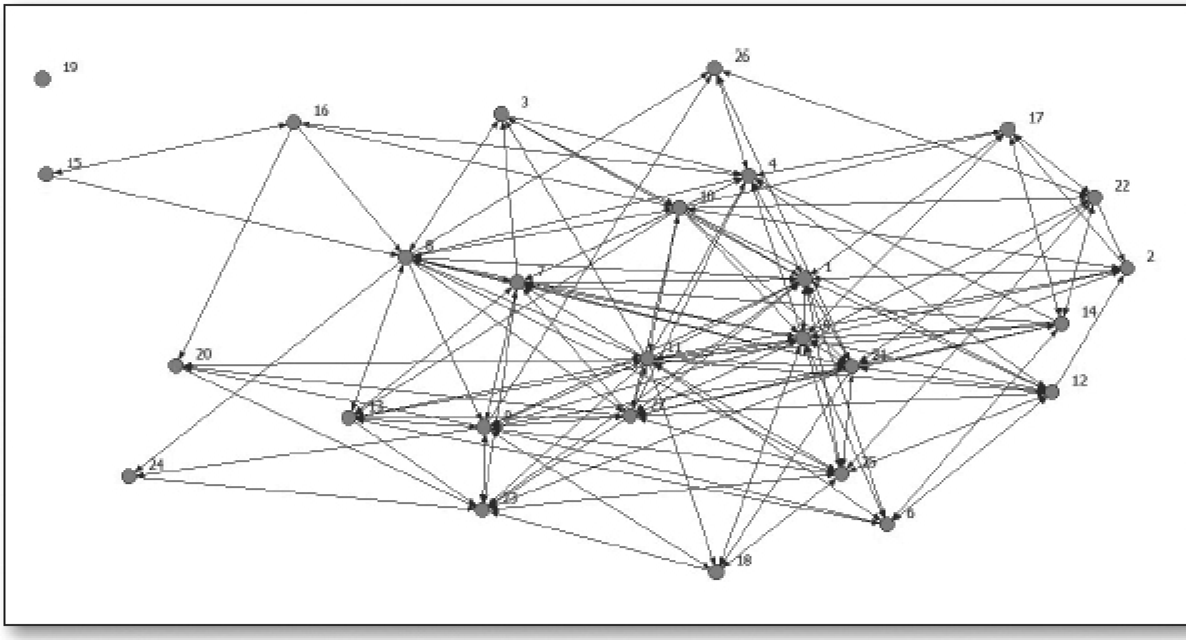
These concepts will be introduced using the Peer Groups data, which collected relational and attribute data from middle school students in four science classes taught by the same teacher. Keeping with the introductory theme of this chapter, only one network from one point in time will be discussed: student-reported friendship relations among students in one class in the fall semester, the beginning of the school year. Using these data, this chapter will introduce (1) the distinguishing features of network data; (2) the types of variables collected in network studies; (3) how networks are represented as graphs and matrices; and (4) how these representations allow us to analyze network phenomena at different analytic levels. Some of these concepts were introduced in Chapter 1. This chapter goes further by discussing these concepts using sociometric notation and the corresponding logic of graph theory.

Graphs and Network Data

Social network data at its most basic level consists of at least one relational measure among a set of actors. Substantive concerns and theories that drive a specific network study usually inform which variables to measure and the frequency with which they are measured. For example, the Peer Groups data set was motivated by several questions that shaped the collection of the network data. First, there was an interest in measuring how relations among peers change throughout the duration of the school year. This interest required that data be collected at a minimum two points in time, in this case, the fall and spring of the academic year. In addition, because the analytic focus was on peer relations, Pittinsky asked each student (ego) to rate their friendship on six-point scale (1 = best friend; 2 = friend, 3 = know-like; 4 = know, 5 = know-dislike; 6 = strongly dislike) with every other student (alter) in the class. In addition, the teacher was asked to do the same. This last feature allows the comparison between the students' and teacher's reports, a nice analytical move that allows the data to explore whether the teacher's and students' reports (dis)agree and how this level of (dis)agreement changes over time. Rather than rely on observations or archival or historical records, Pittinsky used a questionnaire to elicit responses. Because the interest was on the whole network (i.e., the class), the entire set of actors, in this case students, was surveyed, not just a sample of them.

Therefore, the Peer Groups network data can be represented as a graph, a common way to represent social networks, consisting of two dimensions: actors (students) sending relations, represented as rows, and receiving actors (also students) represented as columns. Therefore, the network, X, consists of n actors, or in the language of graph theory, nodes. For example, consider the collection of n = 27 students in the Peer Groups network data set. We have N = 27 (Student 1, Student 2, … Student 27), a collection of 27 actors, so that each student can be referred by symbol: n_1 = Student 1, n_2 = Student 2 … n_{27} = Student 27. Figure 3.1 shows how these data are represented in a graph. Each actor can send or receive ties from every other actor.

Figure 3.1 Peer Groups Data Network Graph. This graph consists of 27 students, and each directed line indicates that a "best friend" or "friend" nomination has been sent from one student to another. Because these ties are directed, they are referred to as arcs.



Graphs such as the one in Figure 3.1 can immediately highlight several important features of overall network structure and provide a good entry point for introducing key network concepts. Graphs consist of two features: (1) nodes, *n*, which represent actors, and (2) lines, which represent relations. Each circle represents a student (node), and each line represents whether a student "sent" a friendship nomination to another student (relation); that is, they sent a 1 or 2 to another student ("best friend" or "friend"). Students can be connected by either arcs or edges. Arcs represent those relations that are directed from one student to another, meaning that the friendship nomination has not necessarily been reciprocated. Edges, on the other hand, are those lines that do not have arrowheads (since friendships are directed, there are no edges in Figure 3.1), which are appropriate when the relation is by definition reciprocated (e.g., "studies with"). Therefore, Figure 3.1 is what is referred to as a directed graph. Conversely, an undirected graph would consist exclusively of edges, nondirected relations between nodes.

Looking at the entire graph, there are several properties that become evident, which highlights how a relatively simple graph can immediately suggest some important structural features. Specifically, it is evident that there is an unevenness in the network's pattern of friendship relations; predictably, friendship nominations among these adolescents are distributed unequally. A small number of students receive a large number of nominations, and the overall connectivity across the entire graph is relatively low. In addition, clearly there are some students on the periphery who receive very few nominations.

Individual students occupy different positions in the graph, indicating students are embedded in the network in quite varied ways. For example, Student 1 (middle, Figure 3.1) sends and receives many nominations

(13 and 19 nominations, respectively), suggesting that that student gets along with others in ways that are quite different than Student 15 (upper left, Figure 3.1), who sends and receives very few nominations. Also, consider Student 11, who is located between numerous pairs of students. Such a position may enable this student to serve as a broker between two different students or groups of students.

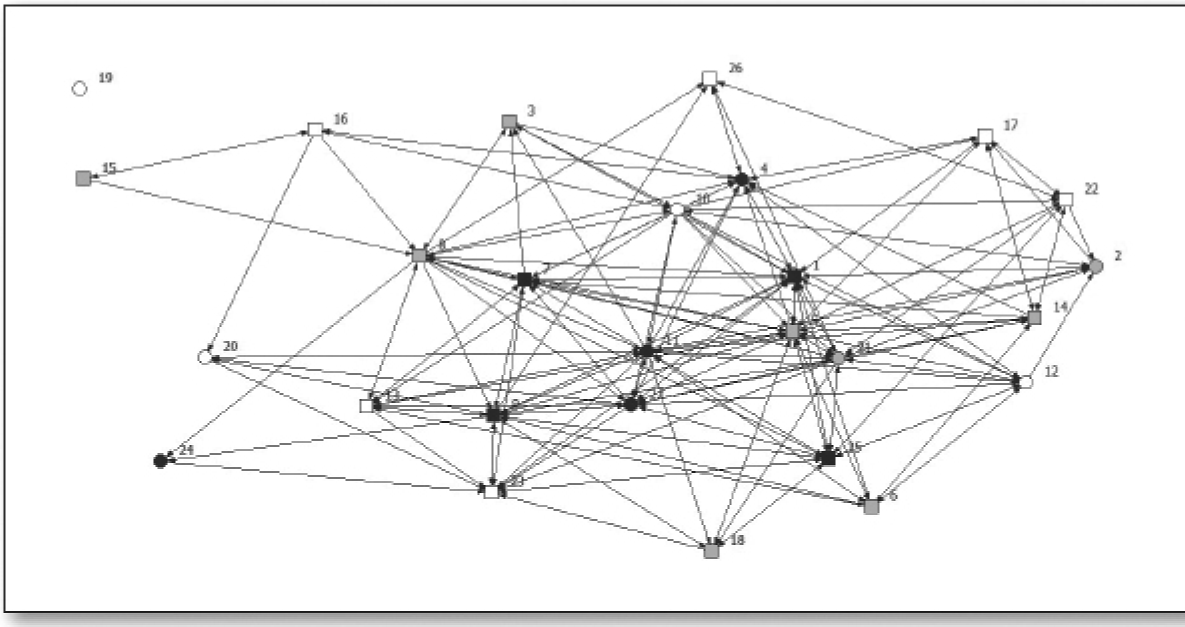
Graphs such as Figure 3.1 can provide useful snapshots by simply representing actors as nodes and relations as lines that are either edges (undirected) or arcs (directed). The book's final chapter proves an extended treatment of network visualization, including an overview of best practices and widely used applications.

Graphing Actors (Nodes) and Relations (Lines)

Graphs can easily incorporate colors and shapes to convey attributes of each node. Homophily, for example, might suggest that relations among students of the same gender would be more common than relations between students of the opposite gender. Peer influence theory suggests that a division between high- and low-achieving students might influence these friendship patterns.

Figure 3.2 shows how these two attributes can be incorporated into a graph, making it much more informative about the hypotheses of differential friendship patterns among students. Girls are represented as square nodes and boys as circles. Light-colored nodes are average achievers and dark nodes are high achievers. Those nodes without color are low achievers. From this graph, there appears to be evidence that friendship nominations are possibly influenced by homophily and peer influence, with high achievers occupying positions in the center of the graph and low achievers on the periphery.

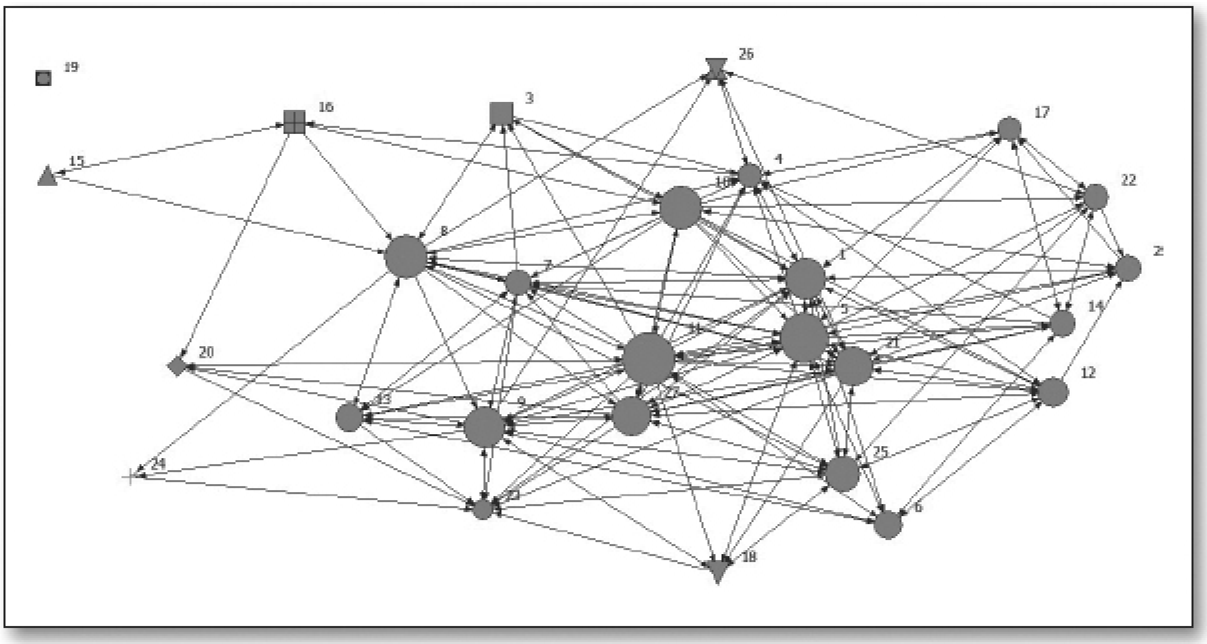
Figure 3.2 Peer Groups Data Network Graph with Node Shape and Color Representing Gender and Achievement Level. Girls are represented as square nodes and boys as circles. Light-colored nodes are average achievers and dark nodes are high achievers. Those nodes without color are low achievers. From this graph, there appears to be evidence that friendship nominations are possibly influenced by homophily and peer influence, with high achievers occupying positions in the center of the graph and low achievers on the periphery.



Not only may nodes differ by attribute, they may also vary quantitatively based on their relations with other nodes. In the Peer Groups data, for example, it might be insightful to make each node's size proportional to the number of friendship nominations received (in-degrees) or sent (out-degrees). Or the size or color of the node can be adjusted to reflect the group with which one is affiliated. Figure 3.3 incorporates these two dimensions by first adjusting the size of each node to reflect the number of friendship nominations that node has received. In addition, the graph also shows the different subgroups within the network, with each subgroup representing a *K*-core (this is one approach to identifying subgroups through relational data discussed in Chapter 5). Each shape represents one of nine different subgroups.

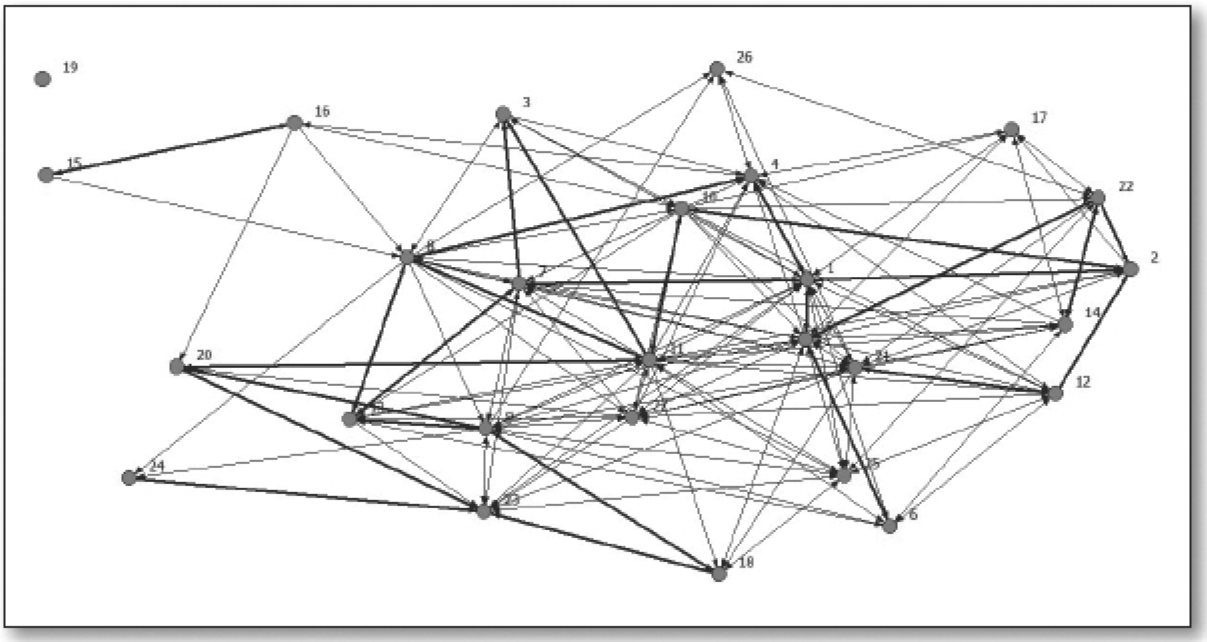
This graph demonstrates a few interesting properties. First, students who receive many nominations, most of whom are near the center of the graph, are also members of the same subgroup. This subgroup is large (18 members), while the network's remaining nine students are split among eight subgroups that are small (1–2 members). This large subgroup also hints at the network having what is referred to as a core-periphery structure: a group of well-connected actors at the center of the network with a set of actors residing on the periphery whose only connections are typically to those actors affiliated with this core group.

Figure 3.3 Peer Groups Data Network Graph With Node Size and Shape Reflecting In-degree and Group Membership. Node size reflects number of friendship nominations received (bigger nodes equal more nominations), and node shape reflects same subgroup membership (*K*-core). This graph demonstrates a few interesting properties. For example, students who receive many nominations, most of whom are near the center of the graph, are also members of the same subgroup. This subgroup is large (18 members), while the network's remaining nine students are split among eight subgroups that are small (1–2 members).



The relations, shown in graphs as either edges (undirected) or arcs (directed), can also be represented in ways that reflect several interesting properties. It can also be very helpful to use color and thickness to indicate a difference in kind and amount among the relations. When ties are measured as a value, as they have been in the original Peer Groups data (values range from 1–6), the strength of the tie can be indicated by using thicker lines to represent stronger ties and vice versa. For example, in Figure 3.4, thicker lines represent "best friend" nominations, originally coded as 1s, and regular lines "friend," originally coded as 2s. These types of relations, therefore, are referred to as "valued," as opposed to the binary (yes/no) relations shown in Figure 3.1. Colored and dashed lines can also be used to indicate different kinds of relations (e.g., who studies with whom or seeks advice from whom), allowing what is referred to as multiplex data to be displayed in one graph. Multiplex data, discussed later in this chapter, are those network data that measure more than one kind of relation, which most contemporary network studies incorporate.

Figure 3.4 Peer Groups Data Network Graph With Line Width Representing Tie Strength. Thick lines represent “best friend” nominations, thinner lines “friends.” In addition to altering the lines’ thickness, it is possible to use colors and dashes to indicate different kinds of relations (e.g., who studies with whom or seeks advice from whom), allowing what is referred to as multiplex data to be displayed in one graph.



Understanding Graphs through Ego Networks

A helpful way to parse complicated network graphs—and it should be evident from Figures 3.1 through 3.4 that things can become complicated quickly—is to see how they emerge from the immediate ties of individual focal actors (Hanneman & Riddle, 2011a). The network formed by targeting one node, including all other nodes to which that node is connected (friends) and all the other connections among those other nodes (friends of friends), is referred to as the ego network, also called the one-step neighborhood of a node. These neighborhoods can also be extended for two (the friends of friends’ friends) or more degrees from ego.

To graph the way in which individual nodes are embedded in the whole network, visual representations of ego networks can be useful. One of the fundamental insights of the social network perspective established in Chapter 1 is that an actor's behaviors and attitudes are influenced by those with whom that actor has relations, and these relations, in turn, are shaped by one's own behaviors and attitudes. Graphing ego-level neighborhoods and comparing them can provide hints into the similarities and differences among the network's actors. For example, using the Peer Groups data, you could ask whether the ego neighborhoods of high-achieving students are bigger than those of low-achieving students or which students have neighborhoods that consist mainly of reciprocated ties.

Questions such as these are descriptive or even exploratory in nature, which, as noted in Chapter 2, reflects a majority of empirical social network studies. Only recently has confirmatory hypothesis testing been employed more regularly in network studies, and this is primarily due to recent advances in inferential statistics that incorporate the nonindependence of observations (see Chapter 8). For small networks such as those represented by the Peer Groups data, graphs can provide an intuitive means through which you can inspect the network's structure and get an intuitive sense of how individual actors are incorporated into its larger structure. Chapters 5 and 6 provide a more formal way of representing these intuitive properties by showing how certain complete and ego network features are precisely calculated.

Social network analysis relies extensively on graphs to represent social structure. As noted by Hanneman and Riddle (2011a), a well-constructed graph can be very useful, perhaps even more useful than words, for communicating a network's properties. Large networks (> 25 actors), however, are not so easy to study visually. In addition, because social network analysis is often concerned with more then one type of relation among actors, graphs are limited in what they can “do.” Therefore, the formal description of network properties and testing hypotheses about them requires that graphs be converted to numbers. This is where we shift from representing networks as graphs and turn toward the uses of sociometry and matrices.

Matrices and Network Data

When there are many actors or many kinds of relations, graphs are not terribly useful. The more common and flexible way of representing networks is through matrices. The advantages of representing network data in this fashion will become clear, but for now, keep in mind that a matrix is simply an array of data. This section briefly reviews the most commonly used matrix representations of social network data. It also introduces some language associated with matrices and matrix operations that differs from the language used when dealing with networks as graphs. This language is important when working with network data. While it is unlikely that you will have to do the math associated with matrix operations, as computers do the bulk of the heavy lifting, the mathematical concepts to be introduced will provide an efficient means through which you can understand the logic of network data and how it is ultimately analyzed (Hanneman & Riddle, 2011a).

When network data are organized in an array, it means that they are in a list. For example, you might call the list of 27 students in the Peer Groups data ***X***, as an array is typically denoted with a bold letter. Each element (student) in the array can then be indexed by its place in the list: Student 1, Student 2 … Student 27. This obviously is the simplest type of array, but there are several different types that are used—typically in combination with each other—in social network analysis. Following Hanneman and Riddle's (2011a) lead, this section considers the five most often-used matrices in social network analysis, with a special emphasis on how these matrices are both different from and similar to the actor-by-attribute data matrices that constitute most social science research.

Vectors

When a matrix has a single dimension, it is referred to as a vector. For example, the simple list of students' names noted earlier is vector. Row vectors are horizontal lists of elements, and column vectors are vertical. In social network analysis, row and column vectors are most typically used to present information about the attributes of actors, referred to later in this chapter as attribute variables. These are the types of variables with which most are familiar. For example, Table 3.1 shows a column vector that vertically lists the 27 different elements, in this case, students who are part of the Peer Groups data set, and two other columns that include an ordinal measure of achievement (1 = high, 2 = average, 3 = low) and gender (1 = girl, 0 = boy). The first column is an ID column and the second and third columns contain attribute variables. Therefore, this data array can be considered 27-by-2 (the number of rows by number of columns), a rectangular matrix that is sometimes referred to as a "list of lists."

Most social network analyses include arrays of variables that describe attributes of variables, ones that are either categorical (e.g., sex, race, etc.) or continuous in nature (e.g., test scores, number of times absent, etc.). Rather than have a separate vector for each attribute variable, it is more efficient to include all attribute variables in a rectangular array that mimics the actor-by-attribute that is the dominant convention in social science: Rows represent cases, columns represent variables, and cells consist of values on those variables.

Rectangular arrays such as Table 3.1 are used to provide information on each actor. This information can come from a number of sources, including standard survey

Table 3.1 Rectangular Matrix With ID and Attribute (Achievement and Gender) Vectors. This table shows a column vector that vertically lists the 27 different elements, in this case, students, who are part of the Peer Groups data set and two other columns that include a measure of achievement (1 = high, 2 = average, 3 = low), and gender (1 = girl, 0 = boy). The first column is an ID column and the second and third columns contain attribute variables. Therefore, this data array can be considered 27-by-2 (the number of rows by number of columns), a rectangular matrix that is sometimes referred to as a "list of lists."

	1	2
1	1	1
2	2	0
3	2	1
4	1	0
5	2	1
6	2	1
7	1	1
8	2	1
9	1	1
10	3	0
11	1	0
12	3	0
13	3	1
14	2	1
15	2	1
16	3	1
17	3	1
18	2	1
19	3	0
20	3	0
21	2	0
22	3	1

	1	2
23	3	1
24	1	0
25	1	1
26	3	1
27	1	0

instruments or observations (these and other data-collection techniques are discussed at length in the next chapter). Each actor can also be described by a variable that has been derived from that actor's relations with others. Referred to later in this chapter as relational variables, information about these variables can also be included as a vector in a rectangular data matrix. For example, you could imagine a separate vector for the Peer Groups data that includes information about the number of friendship nominations sent or received (out- or in-degree). Another way in which attribute vectors are used is to indicate the “group” to which an actor belongs. Here, groups are not defined by some exogenous attribute such as grade level or sports team affiliation but rather, one's group membership is determined by the relations that have been measured. Recall how Figure 3.3 visually shows each student's membership in a *K*-core (one of a number of techniques social network analysts employ to identify cohesive groups discussed in Chapter 5). An attribute vector could be created from these groups to indicate the group in which each of the 27 students belongs. This is what is known as a partitioning vector, which can be used to select subsets of actors, reorganize the data, and calculate more advanced summary measures.

Single-Mode (Square) Matrices

While vectors represent an array of variables for a set of elements (most typically attribute and relational variables), square matrices represent how the elements relate to each other on some measured tie—that is, for example, whether n_1 and n_2 are connected. These matrices are square, meaning that the columns and rows consist of the same nodes. The cells, therefore, indicate whether any two nodes share an arc (undirected) or edge (directed). Table 3.2 shows a simple representation of the Peer Groups data in a square matrix.

Table 3.2 is the simplest and most common network data matrix. First, the relationship is binary; if a tie exists, a 1 is entered into the cell, a 0 if there is no tie. As an aside, the Peer Groups data had to be transformed in order for them to be represented in this fashion. Both nodes had to rate the other as either a 1 or 2 (friend or best friend) in order for the cell to be coded as a 1. Square matrices with

Table 3.2 Binary and Undirected Peer Croups Data Network Data in Square Matrix. A 1 indicates an edge between two nodes, which are therefore adjacent to each other. A 0 indicates that there is no friendship tie between two students. For example, Students 1 and 2 are adjacent because they have an edge between them, indicating that they are friends.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	0	1	0	1	0	1	0	1
2	1	0	0	0	1	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0
3	1	0	0	1	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	0	1	0	1	0	1	1	0	1	1	1	0	1	0	1	1	0	0	0	1	0	0	0	0	1	1
5	1	1	0	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	0	1
6	1	0	0	0	1	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	0	1	1	1	0	0	1	1	1	1	0	1	1	0	0	0	0	0	0	1	0	1	0	0	0	1
8	1	0	1	1	1	0	1	0	1	1	1	0	1	0	1	1	0	0	0	0	1	0	0	1	1	1	1
9	1	0	0	0	1	1	1	1	0	0	1	0	1	0	0	0	0	1	0	1	1	0	1	1	1	1	1
10	1	1	1	1	1	0	1	1	0	0	1	1	1	0	0	1	1	0	0	0	1	1	0	0	0	0	1
11	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	0	1	1	0	1	0	1	0	1
12	1	1	0	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1
13	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
14	0	0	0	1	1	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1
15	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
17	1	1	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
18	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1
21	1	1	0	1	1	0	1	1	1	1	1	1	0	1	0	0	0	1	0	0	0	1	1	0	1	1	1
22	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	1	0
23	1	0	0	0	0	0	1	0	1	0	1	0	1	0	0	0	0	1	0	1	1	0	0	1	1	0	1
24	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
25	1	0	0	0	1	0	0	1	1	0	1	1	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1
26	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
27	1	0	0	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	1	1	0	1	0	1	0	0

binary relations such as these—all the ties are edges and represented as 1s—are typically the starting point for most social network analyses. This type of matrix is also referred to as an “adjacency matrix” because it represents who is near (adjacent) whom in the social structure. Second, this is also referred to as a “single-mode” matrix, as the rows and columns represent the same set of nodes.

However, matrices can also consist of arcs. Whereas Table 3.2 consists entirely of edges and is therefore considered symmetric, the relational data in Table 3.3 are asymmetric: n_1 could send a tie to n_2, but this tie need not be reciprocated. Convention dictates that in a directed (i.e., asymmetric) matrix consisting of arcs, the sender is the row and the target is the column. The data in Table 3.3 highlight this point. For example, consider Student 3, who “sends” a friendship tie to Students 1, 4, 8, and 10. In return, Student 3 “receives” friendship nominations from Students 7, 8, 10, and 11. Asymmetric ties are not necessarily reciprocated, even though they may very well be (Student 3 has a reciprocated tie with Student 8). Hanneman and Riddle (2011a) note that symmetric matrices typically represent “bonded ties” or “co-membership” of any kind of tie in which, if n_1 is tied to n_2, n_2 must be tied to n_1. In educational research, you might be interested in symmetric relations such as membership in the same class or participation in the same extracurricular activities. However, if the focus is on relations such as “advice,” or “support” that are not necessarily reciprocal, it is likely that the data will first be recorded as asymmetric arcs such as the friendship data in Table 3.3. In terms of the language introduced earlier, these lines would be arcs between pairs of actors.

Each single-mode, square matrix reflects a specific relation between all pairs of actors in the network. The quantity in each cell of the matrix indicates the direction (arcs are directed, edges are undirected) and strength (binary or valued) of the tie between any two actors. The strength of relations can be measured a few different ways. In its simplest form, it can be binary (either the relation exists or not) and shown as an adjacency matrix or measured at higher levels, which is discussed in the following chapter, which includes a detailed treatment of measurement issues.

One question arises when representing network data as a single-mode, square matrix: Is it possible for n_1 to send or receive a tie from itself ? The question is what to do with the information on the main diagonal, or trace, of the matrix. Oftentimes, the values on this diagonal are meaningless and ignored. In these instances, the values are left blank or filled with 0s, as in Table 3.3. Sometimes, however, the information on this diagonal is important and is included in the matrix. Consider participation patterns in online classrooms. Isn't it possible to send a question, only to eventually answer it yourself? In this case, the diagonal would certainly reveal something interesting about participation patterns and would likely have to be included in the data matrix.

Table 3.3 Binary and Directed Peer Croups Data Network Data in Square Matrix. A 1 indicates an arc between two nodes, which are therefore adjacent to each other. A 0 indicates that there is no arc between two students. For example, Student 3 sends a friendship nomination to Student 1 but does not receive a nomination in return. This relationship is asymmetric.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	0	1	0	1	1	1	1	1	1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	1	0	1
2	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
3	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0
5	1	1	0	1	0	1	1	1	1	0	1	1	1	1	0	0	1	1	0	0	1	1	0	0	1	0	1
6	1	0	0	0	1	0	0	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
8	1	0	1	1	1	0	1	0	1	1	1	0	1	0	0	0	0	0	0	0	1	0	0	1	1	1	1
9	1	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	1	0	1	1	0	1	1	1	1	1
10	1	1	1	1	1	0	1	1	0	0	1	1	1	0	0	0	1	0	0	0	1	1	0	0	0	0	1
11	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	0	1	1	0	1	0	1	0	1
12	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1
13	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
14	0	0	0	1	0	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
15	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
17	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
18	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
21	1	1	0	1	1	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	1	1	1
22	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	1	0
23	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
25	1	0	0	0	1	0	0	0	1	0	1	1	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1
26	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
27	1	0	0	1	1	0	1	0	1	0	1	1	1	1	0	0	0	0	0	1	0	0	1	0	1	0	0

Multiplex Matrices and Relations

It is true that most network studies across the social sciences and within the field of educational research focus on a single relationship among actors: friendship, support, advice, and the like. However, because social relations are more complex than this and contemporary analytical tools so flexible and efficient, social network analysis is increasingly concerned with a number of different relations among the same set of actors. Students, for example, are connected in many ways other than friendship simultaneously. Students may have emotional ties and exchange relations, co-memberships, and other kinds of connections concurrently. If schools are the actors, they too have multiple relations with other schools, including information exchange, alliances, partnerships, and other connections.

When different kinds of relational data are collected on the same set of actors, this is referred to as multiplex data. Multiplex data consist of a set of matrices, each being a single-mode, square matrix that describes one type of tie. The data in each matrix can either be undirected or directed and either binary or valued. The two matrices in Table 3.4 are two different looks at the Peer Groups data (for the purposes of presentation, only the relations between Students 1–10 are shown). The top panel shows the friendship network from the perspectives of the students themselves in the form of a directed, binary matrix (same as those data in Table 3.3). The matrix in the bottom panel is also directed and binary but includes information on the students' friendship nominations from the teacher's perspective. While the same relation has technically been measured, the fact that the data matrices with the same set of actors (same rows and columns) are generated from two different sources makes these data multiplex in nature. This is a special kind of multiplex data referred to as cognitive social structure (Krackhardt, 1987a). Data such as these illustrate the potentially different ways in which, different actors perceive ties between the same pairs of actors. For example, in Table 3.4, Student 2 sends a friendship tie to Student 1 (as reported by the student, top panel), but the teacher does not perceive that same tie (bottom panel).

Working with multiplex data induces a few analytic choices. You can use all the tools of social network analysis to analyze each data matrix separately. For example, in the Peer Groups data, you may ask whether the subgroups are more distinct in the teacher's report of friendship nominations or if there is more overlap among subgroups in the students' reports? Or does the teacher report greater or fewer friendship ties than the students themselves? But you may also want to combine the information from more than one relation among the same set of actors. This leads to one of two approaches (Hanneman & Riddle, 2011a). The first of these is the reductionist approach, which collapses information about multiple relations among the same set of actors into one single relation that captures the quantity of ties. This could be as simple as "adding" the matrices together and, therefore, the cells of the new matrix represent sums. The combination approach is similar in that the goal is to create a single new matrix derived from multiple relations, but the focus is on quality, not quantity. The relational algebra behind this transformation is too advanced for this introduction to basic network concepts, but is revisited in the book's final chapter when newer advances such as these are discussed.

Table 3.4 Binary and Directed Peer Groups Data Network Data in Square Matrix Generated From Students' (top panel) and Teacher's Reports (bottom panel). While the same friendship relation has been measured, the fact that the data matrices with the same set of actors (same rows and columns) are generated from two different sources makes these data multiplex in nature. Only data from Students 1 through 10 are presented.

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

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

Two-Mode (Affiliation) Matrices

Thus far, the discussion on matrices has focused on those relational data sets that consist of the same set of actors. However, an important focus for educational researchers and social scientists in general is one's location in larger social structures such as neighborhoods, professional associations, communities, and other “identity” categories. Social network analysis can be used to examine the relations within and between multiple analytical levels. For example, students affiliate with groups, and groups are linked by their overlapping memberships.

When relational data represent the connections between two different sets of actors, the data is considered “two–mode.” In two-mode data, the rows and columns represent different sets of actors. Typically, rows represent individual actors and columns represent events, organizations, or some other identity category. Because the number of rows and columns is likely to differ, these matrices are no longer considered square. In addition, the actors are not considered adjacent but rather affiliated. Hence, these two-mode matrices are known as affiliation networks (Borgatti & Everett, 1997). Affiliation networks reflect the connections of two different sets of actors. Similar to single-mode adjacency matrices, the relational data in two-mode matrices can be binary or valued, with rows sending relations to the columns. Because affiliation data consist of two sets of actors, the techniques used to analyze these data are different and somewhat more complex. The two basic approaches (conversion and direct) to the analysis of affiliation data are also discussed in Chapter 7.

For now, it is important to note that the conversion approach is most frequently employed. In this approach, two-mode data are collapsed into a single-mode adjacency matrix. For example, you might transform an undirected (binary) matrix that displays which students were members of certain school-sponsored extracurricular organizations into a matrix that shows how many times each pair of students happened to be co-members. Or you can transform the same data into a single-mode matrix that shows how many times each pair of organizations was connected by overlapping student memberships. The idea is simple: Convert a two-mode data matrix into one prior to analysis. However, in doing so, much information is “lost,” and because of this, the direct approach is gaining traction among social network analysts.

Regardless of whether you rely on graphs or matrices to represent relations, ultimately these representations include variables, most likely many of them. The relations that are measured among actors also determine which analytic methods are ultimately appropriate. Therefore, it is important to consider the different types of variables that can be employed in social network analysis. These variables were briefly noted throughout the previous sections, but it is important to emphasize the distinctions among them as well as the importance of including them in analyses of social networks.

Network Variables

Generally, there are three different types of variables used in the analysis of social networks (Valente, 2010). The first two are derived from the network data themselves, and the third type is the one with which most are familiar. The first type of variable is relational, which is constructed from the respondents’ set of direct ties with others. An example of this variable at the ego level would be density, the degree to which an ego actor's alters know one another. So, in the Peer Groups data set, each student could have a measure of his or her ego network density, that is, a variable that captures the degree to which the friends of n_1 are also friends with each other. Examples of other relational variables include connectedness, reach, in-out-degree, constraint, and several others that will be discussed in subsequent chapters.

The second type of variable employed in social network analysis is similar to relational variables in that these, too, are also derived from network data. The difference is that these variables, referred to as structural variables, are constructed from the entire network of connections, that is, they are calculated on *N*. This distinction, however, is not rigid, since relational and structural measures are often associated (Valente, 2010). Examples of structural variables include centrality and positions. Centrality, measured a number of different ways, generally refers to the degree to which relations are concentrated or dependent on one or more actors. Positional analyses, on the other hand, parse the network's structure by grouping actors based on their (dis)similarity of network ties. Similar to centrality, there are several different ways in which to identify positions within a network and the relations between and within these positions. Both types of structural variables are discussed in Chapter 5, which focuses on these measures as derived from complete networks.

The third type, attribute variables, is increasingly employed in social network analysis and has been the cornerstone of social science research for some time. Also referred to as compositional variables (Wasserman & Faust, 1994), these variables capture properties of individual actors. As such, they are defined at the level of individual actors, *n*. For example, the Peer Groups data set also includes information about students' race, sex, and achievement test scores. Table 3.1 shows how two attribute variables are recorded as vectors, in this case an ordinal measure for achievement and an indictor (dummy) variable for gender. Similar to conventional social science research, these, too, can be incorporated into the statistical analysis of social networks (introduced in Chapter 8).

Summary

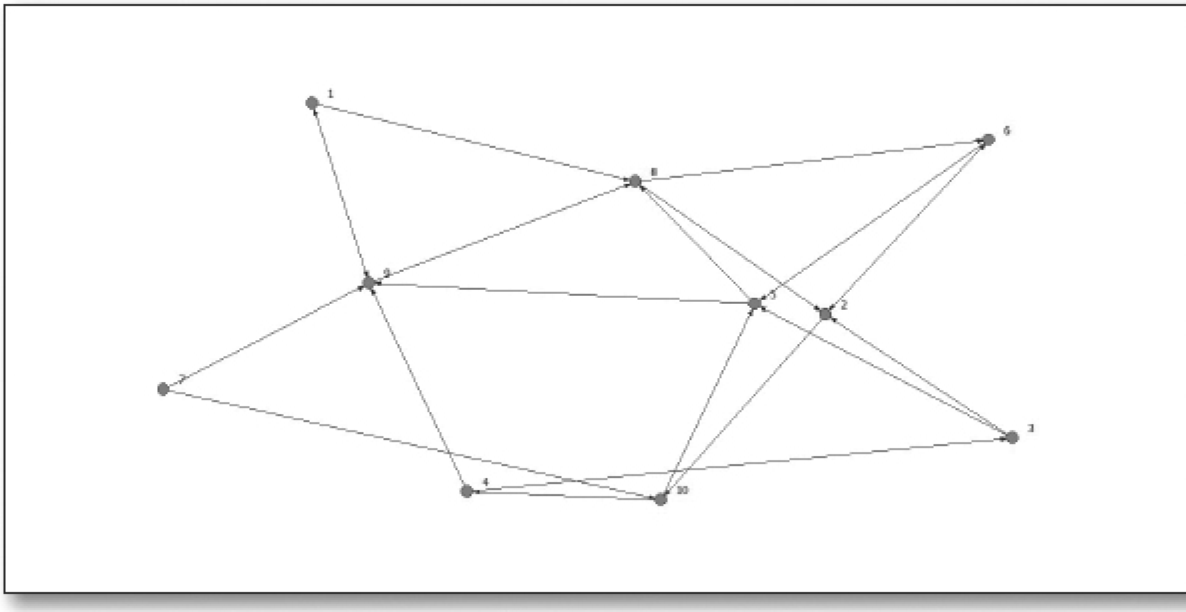
This chapter used the Peer Groups data to show how a pattern of relations among a set of actors (in this case, friendship among students) can be represented in two different formal ways: graphs or matrices. These representations can then be used to define numerous ideas about a network's social structure in precise mathematical terms. Prior to moving on to that task, however, the next chapter addresses the measurement, collection, storage, and manipulation of network data.

Chapter Follow-Up

To answer these questions, please refer to the graph in Figure 3.5.

- Is this graph directed or undirected? Are the lines arcs or edges? Given your responses to these questions, what type of relation might this graph represent? Based on what you see in this graph, is this relation valued or binary?
- Create a matrix that represents the relational data evident in the graph. How many actors are in this matrix? How many rows and columns are in this matrix? What type of data are in the cells of this matrix? Is this matrix symmetric? Explain the values on the diagonal of the matrix.
- What attribute data might you want to incorporate into analysis? Create a rectangular data matrix that represents this variable as a vector. Be sure to describe what the values on this variable represent.

Figure 3.5 Example Graph for Follow-Up Questions.



Essential Reading

Hanneman, R. A., & **Riddle, M.**(2005). Introduction to social network methods. Retrieved from http://faculty.ucr.edu/~hanneman/

Hanneman, R. A., & **Riddle, M.**(2011). A brief introduction to analyzing social network data. In Edited by: **J.Scott** & **P. J.Carrington** (Eds.), The Sage handbook of social network analysis (pp. 331–339). Thousand Oaks, CA: Sage Publications.

Krackhardt, D. Cognitive social structures. *Social Networks,*(1987).9(2),109–134.

http://dx.doi.org/10.4135/9781452270104.n3