3

Notation for Social Network Data

Social network data consist of measurements on a variety of relations for one or more sets of actors. In a network data set we may also have recorded information on attributes of the actors. We will need notation for the set of actors, the relations themselves, and the actor attributes so that we can refer to important network concepts in a unified manner.

In this chapter, we introduce notation and illustrate with examples. We start by defining notation for a single, dichotomous relation. We then move to more complicated network data sets involving more than one set of actors and/or more than one relation. We will need a notational system flexible enough to handle the wide range of network data sets that are encountered in practice. We note that the only type of structural variable discussed in this chapter is relational. Chapter 8 presents notation and methodology for affiliational networks.

For the reader who already is familiar with social networks and the ways in which social network data can be denoted, or the reader who is only interested in specific techniques, we recommend a quick reading of the material in this chapter. Specifically, such readers can glance at Section 2 and the examples used in this chapter (perhaps skipping the material on multiple relations), and return to this chapter as needed.

There are many ways to describe social network data mathematically. We will introduce three different notational schemes. These schemes can each be adapted to represent a wide range of network data. However, for some forms of data and some types of network methods, one notation scheme may be preferred to the others, because of its appropriateness, clarity, or efficiency. The notations are:

- Graph theoretic
- Sociometric

- Algebraic

Each scheme will be described and illustrated in detail. We will show how these schemes overlap, and discuss when a specific scheme is more useful than the others. Graph theoretic notation is most useful for centrality and prestige methods, cohesive subgroup ideas, as well as dyadic and triadic methods. Sociometric notation is often used for the study of structural equivalence and blockmodels. Algebraic notation is most appropriate for role and positional analyses and relational algebras. We should note that there are other ways to denote social network data, some of which are used to study specific statistical models. Such schemes will be mentioned, when needed, in later chapters.

The graph theoretic notation scheme can be viewed as an elementary way to represent actors and relations. It is the basis of the many concepts of graph theory used since the late 1940's to study social networks. The notation provides a straightforward way to refer to actors and relations, and is completely consistent with the notation from the other three schemes. Mathematicians and statisticians such as Bock, Harary, Katz, and Luce were among the first to view networks as directed and undirected graphs (see Forsyth and Katz 1946; Katz 1947; Luce and Perry 1949; Bock and Husain 1950, 1952; Harary and Norman 1953). Graph theory texts such as Flament (1963) and Harary (1969) describe social network applications. We should also direct the reader to other texts on graph theory and social networks, such as Harary, Norman, and Cartwright (1965), and Hage and Harary (1983), that present graph theoretic notation for social network data. Mathematical sociology texts, such as Coleman (1964), Fararo (1973), and Leik and Meeker (1975), contain elementary discussions of the use of graph theory in social network analysis.

The second notation scheme, sociometric notation, is by far the most common in the social network literature. One presents the data for each relation in a two-way matrix, termed a *sociomatrix*, where the rows and columns refer to the actors making up the pairs. Sociomatrices began to be used more than fifty years ago after their introduction by Moreno (1934) in his pioneering research in sociometry (see also Moreno and Jennings 1938).

Most major computer software packages for social network data analyze network information presented in sociomatrices. Further, many methods are defined for sociomatrices. This notational scheme is probably the most useful for readers interested in the methods discussed in Parts III and IV of the book. Sociomatrices are *adjacency matrices* for *graphs*, and consequently, this second notational scheme is directly related to the first.

The third notational scheme, algebraic notation, is used to study multiple relations. This notation is useful for studying network role structures and relational algebras. Such analyses use algebraic techniques to compare and contrast measured relations, and derived compound relations. A compound relation is the composition or combination of two or more relations. For example, if we have measured two relations, "is a friend of" and "is an enemy of," for a set of people, then we might be interested in the composition of these two relations: "friends' enemies." The focus of such algebraic techniques is on the associations among the relations measured on pairs of actors, across the entire set of actors. This notation is designed for one-mode networks, and was first used by White (1963) and Boyd (1969).

We now turn to each of these notations, show how they are related, discuss when each is useful, and illustrate each with examples.

3.1 Graph Theoretic Notation

A network can be viewed in several ways. One of the most useful views is as a *graph*, consisting of *nodes* joined by *lines*. Chapter 4 discusses graph theory at length. Here, we introduce some simple graph theoretic notation, and show how this notation can be used to label the actors and relations in a network data set.

Suppose we have a set of actors. We will refer to this set as $\mathscr{N}$. The set $\mathscr{N}$ contains g actors in number, which we will denote by $\mathscr{N} = \{n_1, n_2, \ldots, n_g\}$. The symbol $\mathscr{N}$ is commonly used to stand for the set, since the graph theory literature frequently refers to this set as a collection of *nodes* of a graph. For example, consider a collection of $g = 6$ second-grade children: Allison, Drew, Eliot, Keith, Ross, and Sarah. We have $\mathscr{N} = \{$Allison, Drew, Eliot, Keith, Ross, Sarah$\}$, a collection of six actors, so that we can refer to the children by their symbols: n_1 = Allison, n_2 = Drew, n_3 = Eliot, n_4 = Keith, n_5 = Ross, and n_6 = Sarah.

3.1.1 A Single Relation

We now assume that we have a single relation for the set of actors $\mathscr{N}$. That is, we record whether each actor in $\mathscr{N}$ relates to every other actor on this relation. To start, we will let the relation be dichotomous and directional. Thus, n_i either relates to n_j or does not. For now, we do not consider the strength of this interaction or how frequently n_i interacts with n_j.

Consider an ordered pair of actors, n_i and n_j. Either the first actor in the ordered pair relates to the second or it does not. Since the relation is directional, the pair of actors n_i and n_j is distinct from the pair n_j and n_i (that is, order matters). If a tie is present, then we say that the ordered pair is an element of a special collection of pairs, which we will refer to as $\mathscr{L}$. If an ordered pair is in $\mathscr{L}$, then the first actor in the pair relates to the second on the relation under consideration.

Note that there can be as many as $g(g-1)$ elements (the total number of ordered pairs in $\mathscr{L}$), and as few as 0.

If the ordered pair under consideration is $< n_i, n_j >$, and if there is a tie present, we will write $n_i \rightarrow n_j$. The elements, or ordered pairs, of relating actors in $\mathscr{L}$ will be denoted by the symbol l. Let us assume that there are L entries in $\mathscr{L}$, so that $\mathscr{L} = \{l_1, l_2, \ldots, l_L\}$. The elements in $\mathscr{L}$ can be represented graphically by drawing a line from the first actor in the element to the second. It is customary to refer to such a graph as a *directed graph*, since the lines have a direction. Directed lines are referred to as *arcs*. We use the symbol $\mathscr{L}$ to refer to the set of directed Lines and the symbol l to refer to the individual directed *lines* in the set. We will frequently refer to such ordered pairs of relating actors as *directed lines* or *arcs*.

Since a graph consists of a set of nodes $\mathscr{N}$, and a set of lines $\mathscr{L}$, it can be described mathematically by the two sets, $(\mathscr{N}, \mathscr{L})$. We will use the symbol $\mathscr{G}$ to denote a graph. It is important to note that for the graph theoretic notation scheme, these two sets (a set of actors, and a set of ordered pairs of actors, or arcs) suffice for mathematical descriptions of the crucial components in a network on which a single dichotomous relation is measured.

On some relations, an individual actor does not usually relate to itself. When studying such relations, one does not consider *self-choices*.

There are relations that are *nondirectional*, that is, we cannot distinguish between the line from n_i to n_j and the line from n_j to n_i. For example, we may consider a set of actors, and record whether they "live near each other." Clearly, this is a nondirectional relation — if n_i lives near n_j, then n_j lives near n_i. There is only one measurement to be made for each pair, rather than two as with a directional relation. The two ordered pairs have identical relational interactions. The set $\mathscr{L}$ now contains at most $g(g-1)/2$ pairs. The order of the pair of actors in these relating pairs no longer matters, since both actors relate to each other in the same way.

Return to our example, and suppose that the single, dichotomous directional relation is "friendship," so that we consider whether each child views every other child as a friend. Suppose further that eight of the possible thirty ordered pairs are friendships (that is, eight of the thirty possible arcs are present) and that the other twenty-two are not friendships (or that there are twenty-two arcs absent). Let these $L = 8$ pairs be <Allison, Drew>, <Allison, Ross>, <Drew, Sarah>, <Drew, Eliot>, <Eliot, Drew>, <Keith, Ross>, <Ross, Sarah>, and <Sarah, Drew>. Thus, for the elements of $\mathscr{L}$, l_1 = <Allison, Drew>, l_2 = <Allison, Ross>, ... , and l_8 = <Sarah, Drew>. The data tell us that Allison views Drew as a friend, Allison also views Ross as a friend, Drew states that Sarah is his friend, and so forth. It is also interesting to note that this friendship is not reciprocal; that is, if n_i states that n_j is his friend (or $n_i \rightarrow n_j$), it is possible that this sentiment is not returned — n_j may not "choose" n_i as a friend (or $n_j \not\rightarrow n_i$).

A graph can be presented as a diagram in which nodes are represented as points in a two-dimensional space and arcs are represented by directed arrows between points. Thus, these six children can be represented as points in a two-dimensional space. It is important to note that the actual location of points in this two-dimensional space is irrelevant. We can take these points, and draw in the eight arcs representing these eight ordered pairs of children who are friends. This directed graph or *sociogram* is shown in Figure 3.1.

3.1.2 Multiple Relations

We may have more than one relation in a social network data set. Graph theoretic notation can be generalized to multirelational networks, which could include both directional and nondirectional relations. For example, we may study whether the corporations in a metropolitan area do business with each other — does n_i sell to n_j, for example — and whether they interlock through their boards of directors — does an officer of corporation n_i sit on the board of directors of corporation n_j? Given the notation presented for the case of a single dichotomous relation, it is easy to generalize it to multiple relations.

Suppose that we are interested in more than one relation defined on pairs of actors taken from $\mathscr{N}$. Let R be the number of relations.

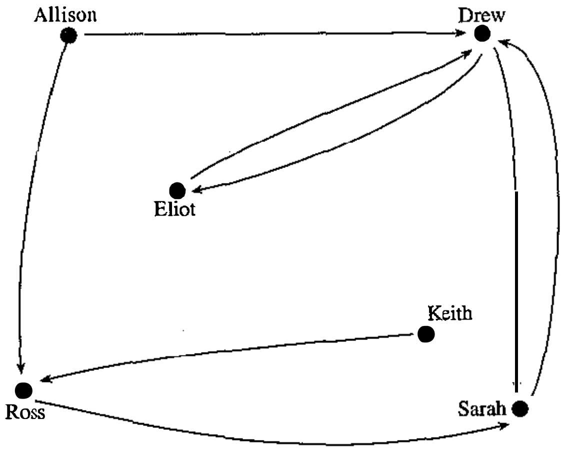

Fig. 3.1. The six actors and the directed lines between them — a sociogram

Each of these relations can be represented as a graph or directed graph; hence, each has associated with it a set of lines or arcs, specifying which (directed) lines are present in the (directed) graph for the relation (or, which (ordered) pairs are "relating"). Thus, each relation has a corresponding set of arcs, $\mathscr{L}_r$, which contains L_r ordered pairs of actors as elements. Here, the subscript r ranges from 1 to R, the total number of relations.

Each of these R sets defines a directed graph on the nodes in $\mathscr{N}$. These directed graphs can be viewed in one or more figures. So, each relation is defined on the same set of nodes, but each has a different set of arcs. Thus, we can quantify the rth relation by $(\mathscr{N}, \mathscr{L}_r)$, for $r = 1, 2, \ldots, R$.

For example, return to our second-graders, and now consider $R = 3$ relations: 1) who chooses whom as a friend, measured at the beginning of the school year; 2) who chooses whom as a friend, measured at the end of the school year; and 3) who lives near whom. The first two relations are directional, while the last is nondirectional. Suppose that $L_1 = 8$ ordered pairs of actors, $L_2 = 11$, and $L_3 = 12$. Below, we list these three sets.

Relation 1 Friendship at Beginning	Relation 2 Friendship at End	Relation 3 Lives Near
<Allison, Drew>	<Allison, Drew>	(Allison, Ross)
<Allison, Ross>	<Allison, Ross>	(Allison, Sarah)
<Drew, Sarah>	<Drew, Sarah>	(Drew, Eliot)
<Drew, Eliot>	<Drew, Eliot>	(Keith, Ross)
<Eliot, Drew>	<Drew, Ross>	(Keith, Sarah)
<Keith, Ross>	<Eliot, Ross>	(Ross, Sarah)
<Ross, Sarah>	<Keith, Drew>	
<Sarah, Drew>	<Keith, Ross>	
	<Ross, Keith>	
	<Ross, Sarah>	
	<Sarah, Drew>	

For a nondirectional relation, such as "lives near," measurements are made on unordered rather than ordered pairs. Clearly, when one actor relates to a second, the second relates to the first; therefore, since Allison lives near Ross, Ross lives near Allison. When listing the pairs of relating actors (or arcs) for a nondirectional relation, each pair can be listed no more than once. We use $(\bullet, \bullet)$ to denote pairs of actors for whom a tie is present on a nondirectional relation, and use $< \bullet, \bullet >$ to denote ties on a directional relation.

Examining such lists can be difficult. An alternative way to present the three sets $\mathscr{L}_1$, $\mathscr{L}_2$, and $\mathscr{L}_3$ is graphically. We can place the arcs for directed graphs or lines for undirected graphs on three figures (one for each relation), or on a single figure containing points representing the six actors and arcs or lines for all relations, simultaneously. We use different types of lines in Figure 3.2 for the different relations: solid, for relation 1 (friend at beginning); dashed, for relation 2 (friend at end); and dotted, for relation 3 (lives near). Since friendship is a directional relation, there are arrowheads indicating the directionality of an arc. Since "lives near" is nondirected, there are no arrowheads on these lines. This figure is an example of a *multivariate directed graph*; such graphs are described in more detail in Chapter 4.

3.1.3 Summary

To review, we have assumed that there is just one set of actors. This assumption will be relaxed in a later section of this chapter. In this simple

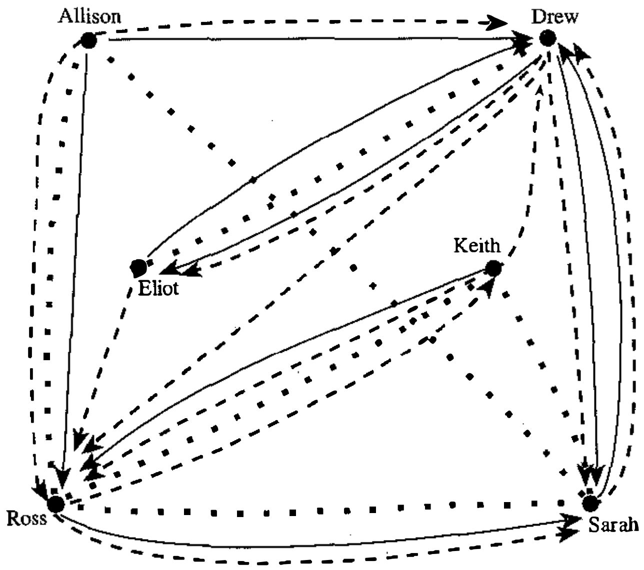

Fig. 3.2. The six actors and the three sets of directed lines — a multivariate directed graph

situation, there is just a single kind of pair of actors, those with both actors in the single set $\mathcal{N}$. The number of actors in $\mathcal{N}$ is g. Assuming that we have R relations, we have a set of arcs associated with each relation, $\mathscr{L}_1, \mathscr{L}_2, \ldots, \mathscr{L}_R$. Each set of arcs can have as many as $g(g-1)$ entries in these sets. The entries in each set are exactly those ordered pairs for which the first actor relates to the second actor on the relation in question. Thus, one needs to specify the set $\mathcal{N}$ and the R sets of arcs to describe completely the network data set.

We should mention that this notation scheme does not extend well to valued relations. Graph theory is not well designed for data sets that record the strength or frequency of the interaction for a pair of actors. One can use special graphs, such as signed graphs and valued graphs (see Chapter 4), to represent valued relations, but many of the more elegant results from graph theory do not apply to this extension. However, sociometric notation is general enough to handle valued relations.

3.2 Sociometric Notation

Sociometry is the study of positive and negative affective relations, such as liking/disliking and friends/enemies, among a set of people. A social network data set consisting of people and measured affective relations between people is often referred to as *sociometric*.

Relational data are often presented in two-way matrices termed *sociomatrices*. The two dimensions of a sociomatrix are indexed by the sending actors (the rows) and the receiving actors (the columns). Thus, if we have a one-mode network, the sociomatrix will be square.

A sociomatrix for a dichotomous relation is exactly the *adjacency matrix* for the graph (or *sociogram*) quantifying the ties between the actors for the relation in question. So, this notation can be viewed as complementary to graph theoretic notation described in the previous section. In these next paragraphs, we describe the history of sociomatrices and sociograms. We then show how social network data can be denoted by a set of sociomatrices.

Sociometry has grown and expanded over the past half century, so that such studies are now usually called simply sociological or occasionally social psychological. The first sociometricians published much of their research in the journal *Sociometry*, which was renamed first *Social Psychology* and then *Social Psychology Quarterly* in the late 1970's. Moreno was *Sociometry*'s founding editor (in 1937). Moreno and other researchers developed a very useful notation for social networks, which we will refer to as sociometric notation. We describe this classic notation in this section.

Sociograms and sociomatrices were first used by Moreno (1934), who demonstrated how they could represent the relational interactions pictured in a sociogram. The focus of Moreno's research, and much of the sociometric literature of the 1930's and 1940's, was how advantageous it was to picture interpersonal interactions using sociograms, even for sets with many actors. In fact, Moreno (see "Emotions Mapped," 1933) aspired to draw a sociometric "map" of New York City, but the best he could do was a sociogram for a community of size 435 (included as a foldout in Moreno 1934). Both Moreno (1934) and Northway (1940) proposed rules for drawing sociograms. These pioneering sociometricians looked for techniques to show the acceptability of each actor relative to the set of actors as a whole and to determine which "choices" were the most important to the group structure. Lindzey and Byrne (1968), building on Moreno's original guidelines, provide a very good discussion of the measurement of relations. Recently, because of innovations in computing, there has been renewed interest in the graphical representations of social network data (Klovdahl 1986).

Moreno actually preferred the use of sociograms to sociomatrices, and had several arguments in print with proponents of sociomatrices, such as Katz. Moreno used his position as editor of *Sociometry* frequently to interject editor's notes into articles in his journal.

Even with the growing interest in figures such as sociograms, researchers were unhappy that different investigators using the same data could produce as many different sociograms (in appearance) as there were investigators. As we have mentioned, the placement of actors and lines in the two-dimensional space is completely arbitrary. Consequently, the use of the sociomatrix to denote social network data increased in the 1940's. The literature in the 1940's presented a variety of methods for analyzing and manipulating sociomatrices (see Dodd 1940; Katz 1947; Festinger 1949; Luce and Perry 1949; and Harary, Norman, and Cartwright 1965). For example, Dodd (1940) described simple algebraic operations for square sociomatrices indexed by the set of actors. He also showed how rows and columns of such matrices could be aggregated to highlight the relationships among sets of actors, rather than the individual actors themselves. Forsyth and Katz (1946) advocated the use of sociomatrices over sociograms to standardize the quantification of social interactions and to represent network data "more objectively" (page 341). This research appears to be the first to focus on derived subgroupings of actors. Katz (1947) proposed a "canonical" decomposition of a sociomatrix to facilitate the comparison of an observed sociomatrix to a target sociomatrix, an idea first proposed by Northway (1940, 1951, 1952). He also showed how sociomatrices could be rearranged using permutation matrices to identify subgroups of actors, and how choices made by a particular actor could be viewed as a multidimensional vector. Festinger (1949) applied matrix multiplication to sociomatrices and described how products of a sociomatrix (particularly squares and cubes) can be used to find cliques or subgroups of similar actors (see also Chabot 1950). Since such powers have simple graph theoretic interpretation (see Chapter 4's discussion of 2- and 3-step walks), this research helped begin the era of graph theoretic approaches to social network analysis. Luce and Perry (1949) and Luce (1950) proposed one of the first techniques to find cliques or subgroups of actors using (for that time) rather sophisticated sociomatrix calculations backed up with an elaborate set of theorems describing the properties and uniqueness of their approach (which was termed *n-clique analysis*; see Chapter 7). Bavelas (1948, 1950) and Leavitt (1951) introduced the notion of centrality (see Chapter 5) into social network analysis. By the end of the decade, researchers had begun to think about electronic calculations for sociometric data (Beum and Criswell 1947; Katz 1950; Beum and Brundage 1950) consisting of a collection of sociomatrices. Research of Katz (1953), MacRae (1960), Wright and Evitts (1961), Coleman (1964), Hubbell (1965), and methods discussed by Mitchell (1969) rely extensively on computers to find various graph theoretic measures. The 1950's and early 1960's became the era of graph theory in sociometry.

The line between sociometric and graph theoretic approaches to social network analysis began to become blurred during the early history of the discipline, as computers began to play a bigger role in data analysis. Sociograms waned in importance as sociomatrices became more popular and as more mathematical and statistical indices were invented that used sociomatrices, much to the dismay of Moreno (1946).

History is certainly on the side of this notational scheme. In fact, most research papers and books on social network methodology begin with the definition of a sociomatrix. Readers who are interested in the topics in Parts II and III will find this notation most useful. For most social network methods, sociometric notation is probably the only notation necessary. It is also the scheme preferred by most network analysis computer programs. It is important to note, however, that sociometric notation can not easily quantify or denote actor attributes, and thus is limited. It is useful when actor attributes are not measured. The relationship between sociometric notation and the more general graph theoretic notation contributes to the popularity of this approach.

As is done throughout this chapter, we split our discussion of sociometric notation and sociomatrices into several parts. We first describe how to construct these two-dimensional sociometric arrays when only one set of actors and one relation is present, and then, when one set of actors and two (or more) relations are measured. Our discussion of two (or more) sets of actors can be found at the end of the chapter.

3.2.1 Single Relation

Let us suppose that we have a single relation measured on one set of g actors in $\mathcal{N} = \{n_1, n_2, \ldots, n_g\}$. We let $\mathcal{X}$ refer to this single valued, directional relation. This relation is measured on the ordered pairs of actors that can be formed from the actors in $\mathcal{N}$.

Consider now the measurements taken on each ordered pair of actors. Define x_{ij} as the value of the tie from the ith actor to the jth actor on the single relation. We now place these measurements into a *sociomatrix*. Rows and columns of this sociomatrix index the individual actors, arranged in identical order. Since there are g actors, the matrix is of size $g \times g$. Sociometric notation uses such matrices to denote measurements on ties.

For the relation $\mathscr{X}$, we define $\mathbf{X}$ as the associated sociomatrix. This sociomatrix has g rows and g columns. The value of the tie from n_i to n_j is placed into the (i, j)th element of $\mathbf{X}$. The entries are defined as:

$$\begin{aligned} x_{ij} \quad = \quad & \text{the value of the tie from } n_i \text{ to } n_j \\ & \text{on relation } \mathscr{X}, \end{aligned} \tag{3.1}$$

where i and j $(i \neq j)$ range over all integers from 1 to g. An example will be given shortly. One can think of the elements of $\mathbf{X}$ as the coded values of the relation $\mathscr{X}$. If the relation is dichotomous, then the values for the tie are simply 0 and 1.

Pairs listing the same actor twice, (n_i, n_i), $i = 1, 2, \ldots, g$, are called "self-choices" for a specific relation and are usually undefined. These self-choices lie along the main diagonal of the sociomatrix; consequently, the main diagonal of a sociomatrix is usually full of undefined entries. However, there are situations in which self-choices do make sense. In such cases, the entries $\{x_{ii}\}$ of the sociomatrix are defined. Usually, we will assume undefined sociomatrix diagonals since most methods ignore these elements.

Assume now that this relation is valued and discrete. We will then assume that the possible values for the relation come from the set $\{0, 1, 2, \ldots, C-1\}$, for $C = 2, 3, \ldots$. If the relation is dichotomous, then $C = 2$ possible values. Thus, C is defined as the number of different values the tie can take on. If the relation is valued and discrete, but takes on other than integer values from 0 to $C - 1$, then we can easily transform the actual values into the values for this set. For example, if the relation can take on the values $-1, 0, 1$, then we can map -1 to 0, 0 to 1, and $+1$ to 2 (so that $C = 3$). One nice feature of sociometric notation is its ability to handle valued relations.

Since the case of a single relation is just a special case of the multirelational situation, we now turn to this more general case.

3.2.2 Multiple Relations

Suppose that we have R relations $\mathscr{X}_1$, $\mathscr{X}_2$, ..., $\mathscr{X}_R$ measured on a single set of actors. We assume that we have R relations indexed by $r = 1, 2, \ldots, R$. As with a single relation, these relations are valued, and the values for relation $\mathscr{X}_r$ come from the set $\{0, 1, 2, \ldots, C_r - 1\}$.

Consider now the measurements on each possible ordered pair of actors. We define x_{ijr} as the strength of the tie from the ith actor to the jth actor on the rth relation. We now place these measurements into a collection of sociomatrices, one for each relation. Rows and columns of each sociomatrix index the individual actors, arranged in identical order. Thus, the rows and columns of all the sociomatrices are labeled identically. Each matrix is of size $g \times g$.

Consider one of the relations, say $\mathscr{X}_r$, and define $\mathbf{X}_r$ as the sociomatrix associated with this relation. The value of the tie from n_i to n_j is placed into the (i,j)th element of $\mathbf{X}_r$. The entries are defined as:

$$x_{ijr} = \text{the value of the tie from } n_i \text{ to } n_j \text{ on relation } \mathscr{X}_r, \qquad (3.2)$$

where i and j $(i \neq j)$ range over all integers from 1 to g, and $r = 1, 2, \ldots, R$. As mentioned, x_{ijr} takes on integer values from 0 to $C_r - 1$. One can think of the elements of $\mathbf{X}_r$ as the coded values of the relation $\mathscr{X}_r$. There are R, $g \times g$ sociomatrices, one for each relation defined for the actors in $\mathscr{N}$. In fact, one can view these R sociomatrices as the layers in a three-dimensional matrix of size $g \times g \times R$. The rows of these sociomatrices index the sending actors, the columns index the receiving actors, and the layers index the relations. Sometimes, this matrix is referred to as a *super-sociomatrix*, representing the information in a multirelational network.

Consider again our example, consisting of a collection of $g = 6$ children and $R = 3$ relations: 1) Friendship at beginning of the school year; 2) Friendship at end of the school year; and 3) Lives near. All three relations are dichotomous, so that $C_1 = C_2 = C_3 = 2$. These three relations are pictured in a single multivariate or multirelational sociogram in Figure 3.2. In Table 3.1 below, we give the three 6×6 dichotomous sociomatrices for the three relations. Note how in Figure 3.2, a "1" in entry (i, j) for the rth sociomatrix indicates that $n_i \rightarrow n_j$ on relation $\mathscr{X}_r$ (or, $n_i \xrightarrow{\mathscr{X}_r} n_j$, for short).

To illustrate, look at the first relation and the first arc in $\mathscr{L}_1$. In Section 3.1, we said that this arc is l_1 = <Allison, Drew>. Allison $\rightarrow$ Drew is

Table 3.1. *Sociomatrices for the six actors and three relations of Figure 3.2*

Friendship at Beginning of Year

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	0	1
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	0
Ross	0	0	0	0	-	1
Sarah	0	1	0	0	0	-

Friendship at End of Year

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	1	0	0	1	0
Drew	0	-	1	0	1	1
Eliot	0	0	-	0	1	0
Keith	0	1	0	-	1	0
Ross	0	0	0	1	-	1
Sarah	0	1	0	0	0	-

Lives Near

	Allison	Drew	Eliot	Keith	Ross	Sarah
Allison	-	0	0	0	1	1
Drew	0	-	1	0	0	0
Eliot	0	1	-	0	0	0
Keith	0	0	0	-	1	1
Ross	1	0	0	1	-	1
Sarah	1	0	0	1	1	-

represented by the arc l_1. Thus, there is an arc from Allison to Drew in the sociogram for the first relation, indicating that Allison chooses Drew as a friend at the beginning of the school year. The first entry in $\mathcal{L}_1$ is exactly this arc. This arc is how this tie is denoted by graph theoretic notation. Consider now how this single tie is coded with sociometric notation. Consider the first sociomatrix in Table 3.1. Consider the entry which quantifies Allison (n_1) as a sender (the first row) and Drew (n_2) as a receiver (the second column) on relation $\mathcal{X}_1$. This entry is in the $(1, 2)$ cell of this sociomatrix, and contains a 1 indicating that

$$\begin{aligned} x_{121} &= \text{the value of the tie from } n_1 \text{ to } n_2 \text{ on relation } \mathcal{X}_1 \\ &= 1. \end{aligned}$$

Note also that $x_{211} = 0$, indicating that Drew does not choose Allison as a friend at the beginning of the school year; that is, Drew $\not\rightarrow$ Allison. This friendship is clearly one-sided, and is not reciprocated.

As one can see, sociometric notation is simple, once one gets used to reading information from two-dimensional sociomatrices. Also note how the diagonals of all three sociomatrices in Table 3.1 are undefined — by design, children are not allowed to choose themselves as friends, and we do not record whether a child lives near himself or herself.

These sociomatrices are the adjacency matrices for the two directed graphs and one undirected graph for the three dichotomous relations. The graphs and the sociomatrices represent exactly the same information. In graph theoretic notation, there are two sets of arcs and one set of lines, $\mathscr{L}_1$, $\mathscr{L}_2$, $\mathscr{L}_3$, which list the ordered pairs of children that are tied for the first two relations and the pairs of children that are tied for the third. If an ordered pair is included in the first or second $\mathscr{L}$ set, then there is an arc drawn from the first child in the pair (the sender) to the second (the receiver). And if an unordered pair of actors is included in the third line set, then there is a line between the two children in the pair. In sociometric notation, the entry in the corresponding cell of the sociomatrix is unity.

We also want to note that the third relation in this network data set is nondirectional; that is, there is a line from n_i to n_j whenever there is a line from n_j to n_i, and vice versa. Note how we were able to code this relation in the sociomatrix given in Table 3.1. Also note that the sociomatrix for a nondirectional relation is symmetric; that is, $x_{ij} = x_{ji}$. One very nice feature of sociometric notation is that it can easily handle both directional and nondirectional relations.

3.2.3 Summary

As we have stated in this section, sociometric notation is the oldest, and perhaps the easiest, way to denote the ties among a set of actors. A single two-dimensional sociomatrix is defined for each relation, and the entries of this matrix code the ties between pairs of actors. Generalizing to valued relations is also easy — the entries in a sociomatrix are the values of the ties, not simply 0's and 1's.

Sociometric notation is very common, the notation of choice for network computing, and will be our first choice of a notational scheme throughout this book. However, as we have mentioned, there are network data sets for which sociometric notation is more difficult to use — specifically, those which contain information on the attributes of the actors. For example consider our second-graders. If we knew their ethnicity (coded on some nominal scale), it would be difficult to include this information in the three sociomatrices (but see Frank and Harary 1982, for an alternative representational scheme).

To conclude, we will frequently use sociomatrices to present network data. These arrays are very convenient (and space-saving!) devices to denote network data sets.

3.3 ○Algebraic Notation

Let us now focus on relations in multirelational networks. In order to present algebraic methods and models for multiple relations (such as relational algebras) in Chapters 11 and 12, it is useful to employ a notation that is different from, though consistent with, the sociometric and graph theoretic notations that we have just discussed. We will refer to this scheme as *algebraic notation.* Algebraic notation is most useful for multirelational networks since it easily denotes the "combinations" of relations in these networks. However, it can also be used to describe data for single relational networks.

There are two major differences between algebraic notation and sociometric notation. First, one refers to relations with distinct capital letters, rather than with subscripted $\mathscr{X}$'s. For example, we could use F to denote the relation "is a friend of" and E for the relation "is an enemy of." Second, we will record the presence of a tie from actor i to actor j on relation F as iFj. This is a shorthand for the sociometric and graph theoretic notation. Rather than indicating ties as $i \rightarrow j$, we will replace the $\rightarrow$ with the letter label for the relation.

In general, $x_{ijF} = 1$ if $n_i \rightarrow n_j$ on the relation labeled $\mathscr{X}_F$ (or F for short). This tie will be denoted by $i \stackrel{F}{\rightarrow} j$, or shortened even further to iFj. This latter notation, iFj, is algebraic.

Referring to our example, we label the relation "is a friend of at the beginning of the school year" as F. We would record the tie implied by "child i chooses child j as a friend at the beginning of the school year" as iFj. In sociometric notation, iFj means that $x_{ijF} = 1$, and implies that there is a "1" in the cell at row i and column j of the sociomatrix for this relation.

Algebraic notation is especially useful for dichotomous relations, since it codes the presence of ties on a given relation. Extensions to valued relations can be difficult. However, the limitation to dichotomous relations presents no problem for us, since the models that use algebraic notation are specific to dichotomous relations. The advantages of this notation are that it allows us to distinguish several distinct relations using letter designations, and to record *combinations* of relations, such as "friends' enemy," or "mother's brother," or a "friend's neighbor." Unfortunately, this notational scheme can not handle valued relations or actor attributes.

3.4 ○Two Sets of Actors

A network may include two sets of actors. Such a network is a two-mode network, with each set of actors constituting one of the modes. A researcher studying such a network might focus on how the actors in one set relate to each other, how the actors in the other set relate to each other, and/or how actors in one set relate to the actors in the other set. In this situation, we need to distinguish between the two sets of actors and the different types of ties. We note that relations defined on two sets of actors often yield complicated network data sets. It is thus quite complicated to give "hard-and-fast" notation rules to apply to every and all situations. We recommend that for multirelational data sets one make an inventory of measured relations and modify the rules given below to apply to the situation at hand.

There are many social networks that involve two sets of actors. For example, we might have a collection of teachers and students who are interacting with each other. Consider the relations "is a student of" and "attends faculty meetings together." The relation "is a student of" can only exist between a student and a teacher. The relation "attends faculty meetings together" is defined only for pairs of teachers.

We will call the first actor in the pair the *sender* and the second actor the *receiver*. Other authors have called these actors *originators* and *recipient*, or simply, *actors* and *partners*. With this understanding, we can distinguish between the two actors in the pair. If the relation is defined on a single set of actors, both actors in the pair can be senders and both can be receivers. The interesting "wrinkle" that arises if there are two sets of actors is that the senders might come only from the first set and the receivers only from the second.

We will let $\mathscr{N}$ refer to the first set of actors and $\mathscr{M}$ refer to the second set. The set $\mathscr{N}$ contains g actors and the second set $\mathscr{M}$ contains h actors. The set $\mathscr{M}$ contains elements $\{m_1, m_2, \ldots, m_h\}$, so that m_i is a typical actor in the second set. Further, there are $\binom{h}{2}$ dyads that can be formed from actors in $\mathscr{M}$.

In this section, we will first discuss the two types of pairs that can arise when relations are measured on two (or even more) sets of actors. We present only sociometric notation, since it is sufficient.

To illustrate the notation, we return to our collection of six second-grade children, and now consider a second set of actors, $\mathscr{M}$, consisting of $h = 4$ adults. We define m_1 = Mr. Jones, m_2 = Ms. Smith, m_3 = Mr. White, and m_4 = Ms. Davis. In total, we have ten actors, which are grouped into these two sets. Considering just the actors in $\mathscr{M}$, there are $4(4-1)/2 = 6$ additional unordered pairs.

3.4.1 ⊗Different Types of Pairs

With two sets of actors, there can be two types of pairs — those that consist of actors from the same set and those that consist of actors from different sets. We will call the former *homogeneous* and the latter *heterogeneous*. Thus, in homogeneous pairs the senders and receivers are from the same set, while in heterogeneous pairs actors are from different sets. We discuss each of these types, beginning with homogeneous pairs.

We can further distinguish between two kinds of homogeneous pairs by noting that there are two sets from which the actors can come. The two kinds of homogeneous pairs are:

- Sender and Receiver both belong to $\mathscr{N}$
- Sender and Receiver both belong to $\mathscr{M}$

In a data set with just one set of actors, the pairs are all homogeneous. However, when there are two sets of actors, there are two kinds of homogeneous pairs.

Of more interest when there are two sets of actors are the pairs that contain one actor from each set. These heterogeneous pairs are also of two kinds, depending on the sets to which the sender and receiver belong. Assuming the relation for the heterogeneous pairs is directional, the originating actor must belong to a different set than the receiving actor. Since there are two sets of actors, we get two kinds of heterogeneous pairs:

- Sender belongs to $\mathscr{N}$ and Receiver belongs to $\mathscr{M}$
- Sender belongs to $\mathscr{M}$ and Receiver belongs to $\mathscr{N}$

It is important to distinguish between these two collections of heterogeneous pairs. Relations defined on the first collection of pairs can be quite different from those defined on the second. For example, if

$\mathcal{N}$ is a set of major corporations in a large city and $\mathcal{M}$ is a set of non-profit organizations (such as churches, arts organizations, charitable institutions, etc.), then we could study how the corporations in $\mathcal{N}$ make charitable contributions to the non-profits in $\mathcal{M}$. Such a relation would not be defined for the other collection of heterogeneous pairs, since it is virtually impossible for non-profits to contribute money to the welfare of the corporations.

3.4.2 ○Sociometric Notation

We now turn our attention to sociometric notation and sociomatrices for the relations defined for both homogeneous *and* heterogeneous pairs. The notation will have to allow for the fact that the sending and receiving actors could come from different sets. We assume that we have a number of relations. The measurements for a specific relation can be placed into a sociomatrix, and there is one sociomatrix for each relation.

A sociomatrix is indexed by the set of originating actors (for its rows) and the set of receiving actors (for its columns) and gives the values of the ties from the row actors to the column actors. If the relation is defined for actors from different sets, then in general, its sociomatrix will not be square. Rather, it will be *rectangular.*

Let us pick one of the relations, say $\mathcal{X}_r$, and suppose that it is defined on a collection of heterogeneous pairs in which the originating actor is from $\mathcal{N}$ and the receiving actor is from $\mathcal{M}$. The sociomatrix $\mathbf{X}_r$, giving the measurements on $\mathcal{X}_r$, has dimensions $g \times h$. The (i,j)th cell of this matrix gives the measurement on this rth relation for the pair of actors (n_i, m_j). The (i,j)th entry of the sociomatrix $\mathbf{X}_r$ is defined as:

$$x_{ijr} = \text{the value of the tie from } n_i \text{ to } m_j \text{ on the relation } \mathcal{X}_r. \tag{3.3}$$

The actor index i ranges over all integers from 1 to g, while j ranges over all integers from 1 to h, and $r = 1, 2, \ldots, R$. As with relations defined on a single set of actors, x_{ijr} takes on integer values from 0 to $C_r - 1$.

Here, i can certainly equal j, since these two indices refer to different sets. The value of x_{iir} is meaningful.

When there are two sets of actors, there are four possible types of sociomatrices, each of which might be of a different size. The rows and columns of the sociomatrices will be labeled by the actors in the sets involved: the rows for the sending actor set and the columns for

Table 3.2. *The sociomatrix for the relation "is a student of" defined for heterogeneous pairs from $\mathcal{N}$ and $\mathcal{M}$*

		$\mathcal{M}$			
		Mr. Jones	Ms. Smith	Ms. Davis	Mr. White
	Allison	1	0	0	0
	Drew	0	1	0	0
$\mathcal{N}$	Eliot	0	0	1	0
	Keith	0	0	0	1
	Ross	0	0	1	0
	Sarah	0	1	0	0

the receiving actor set. We will denote the sociomatrices by using their sending and receiving actor sets, so, for example, the sociomatrix $\mathbf{X}^{\mathcal{N}\mathcal{M}}$ contains measurements on a relation defined from actors in $\mathcal{N}$ to actors in $\mathcal{M}$. These sociomatrices and their sizes are:

- $\mathbf{X}_r^{\mathcal{N}}$, dimensions $= g \times g$
- $\mathbf{X}_r^{\mathcal{M}}$, dimensions $= h \times h$
- $\mathbf{X}_r^{\mathcal{N}\mathcal{M}}$, dimensions $= g \times h$
- $\mathbf{X}_r^{\mathcal{M}\mathcal{N}}$, dimensions $= h \times g$

The second two types are, in general, rectangular. As always, in each sociomatrix, x_{ijr} is the value of the tie from actor i to actor j on the rth relation of that particular type.

Clearly, this notational scheme can accommodate multiple relations. However, since there may be a different number of relations defined for the four different types of pairs of actors, there may be different numbers of sociomatrices of each type.

To illustrate, consider an example with two sets of actors: students and teachers. Suppose there are four adults, second-grade teachers at the elementary school that is attended by six children. Define a relation, "is a student of." This relation is defined for heterogeneous pairs of actors for which the sender belongs to $\mathcal{N}$ and the receiver belongs to $\mathcal{M}$; that is, a child "is a student of" an adult teacher, but not vice versa. Table 3.2 gives the sociomatrix for the two-mode relation "is a student of" from our network of second-grade children. This relation is defined for the heterogeneous pairs consisting of a child as the sender and an adult as a receiver. This is a dichotomous relation ($C = 2$), and is measured on the $6 \times 4 = 24$ heterogeneous pairs of children and teachers.

Note that there is only one 1 in every row of this matrix, since a child can have only one teacher. The entries in a specific column give the children that are taught by each teacher. Note how easily this array codes the information in the directional relation between two sets of actors. It is important to note that with sociometric notation all we need is one sociomatrix (with the proper dimensions) for each relation.

3.5 Putting It All Together

We conclude this chapter by pulling together all three notations into a single, more general framework. To begin, we note that the collection of actors, the relational information on pairs of actors, and possible attributes of the actors constitute a collection of data that can be referred to as a *social relational system.* Such a system is a conceptualization of the actors, pairs, relations, and attributes found in a social network.

As we have shown in this chapter, the data for a social relational system can be denoted in a variety of ways. It is important to stress that when dichotomous relations are considered, the three notational systems discussed in this chapter are capable of representing the entire data set.

We will use the symbols "$n_i \rightarrow n_j$" as shorthand notation for n_i "chooses" n_j on the single relation in question; that is, the arc from n_i to n_j is contained in the set $\mathscr{L}$, so that there is a tie present for the ordered pair $< n_i, n_j >$. If this arc is an element of $\mathscr{L}$, then there is a directed line from node i to node j in the directed graph or sociogram representing the relationships between pairs of actors on the relation. Sometimes we will replace "$n_i \rightarrow n_j$" with "$i \rightarrow j$" if no confusion could arise. With algebraic notation, if we label this relation by, say, F, we can also state that iFj. And with sociometric notation, we record this tie as $x_{ij} = 1$ in the proper sociomatrix.

As we have mentioned in our discussion of graph theoretic notation, if one has a single set of g actors, $\mathscr{N}$, then there are g(g − 1) ordered pairs of actors. In addition to $\mathscr{N}$, the set $\mathscr{L}$ contains the collection of ordered pairs of actors for which ties are present.

Some social network methodologists refer to the set of actors and the set of arcs as the *algebraic structure* $S = <\mathscr{N}, \mathscr{L}>$ (Freeman 1989). S is the standard representation of the simplest possible social network. For us, this is the graph theoretic representation.

One can define a graph from S by stating that the directed graph $\mathscr{G}_d$ is the ordered pair $<\mathscr{N}, \mathscr{L}>$, where the elements of $\mathscr{N}$ are nodes in the graph, and the elements of $\mathscr{L}$ are the ordered pairs of nodes for which there is a tie from n_i to n_j ($n_i \rightarrow n_j$).

Nodes and arcs are the basic building blocks for graph theoretic notation. To relate these concepts to the elements of sociometric notation, we consider again the collection of all ordered pairs of actors in $\mathcal{N}$. Sometimes this collection is denoted $\mathcal{N} \times \mathcal{N}$, a Cartesian product of sets. We define a binary quantity x_{ij} to be equal to 1 if the ordered pair $< n_i, n_j >$ is an element of $\mathcal{L}$ (that is, if there is a tie from n_i to n_j) and equal to 0 if the ordered pair is not an element of $\mathcal{L}$. This quantity is a mapping from the elements of the collection of ordered pairs to the set containing just 0 and 1. These quantities are exactly the elements of the binary $g \times g$ sociomatrix $\mathbf{X}$.

A relation is the collection of all ordered pairs for which $n_i \rightarrow n_j$. It is thus a subset of $\mathcal{N} \times \mathcal{N}$. In algebraic notation, capital letters (such as F) are used to refer to specific relations and to denote which ties are present. A relation is thus the set of all pairs of actors for which $n_i \rightarrow n_j$, or $x_{ij} = 1$, or iFj.

Thus, one can see the equivalence between the graph theoretic notation, and the sociometric notation (built on sociomatrices), and the algebraic notation (dependent on relations such as F). Freeman (1989) views the triple consisting of the algebraic structure S, the directed graph or sociogram $\mathcal{G}_d$, and the adjacency matrix or sociomatrix $\mathbf{X}$ as a social network:

$$\mathcal{S} = < S,\ \mathcal{G}_d,\ \mathbf{X} > .$$

This triple provides a nice abstract definition of the central concept of this book. And, it shows how these notational schemes are usually viewed together as providing the three essential components of the simplest form of a social network:

- A set of nodes and a set of arcs (from graph theoretic notation)
- A sociogram or graph (produced from the sets of nodes and arcs)
- A sociomatrix (from sociometric notation)

It is important to note that most of the generalizations of this simple social network $\mathcal{S}$, such as to valued relations, multiple relations, more than one set of actors, and relations measured over time, can be viewed in just the same way as the situation described here (single dichotomous relation measured on a single set of actors). The only wrinkle is that actor attributes are not easily quantified by using these concepts. The best one can do is to define a new matrix, $\mathbf{A}$, of dimensions (number of actors) $\times$ (number of attributes) to hold the measurements on the