Structural analysis in the social sciences

Mark Granovetter, editor

Other books in the series:

Ronald L. Breiger, ed., *Social Mobility and Social Structure*
John L. Campbell, J. Rogers Hollingsworth, and Leon N. Lindberg, eds., *Governance of the American Economy*
David Knoke, *Political Networks: The Structural Perspective*
Kyriakos Kontopoulos, *The Logics of Social Structure*
Mark S. Mizruchi and Michael Schwartz, eds., *Intercorporate Relations: The Structural Analysis of Business*
Philippa Pattison, *Algebraic Models for Social Networks*
Barry Wellman and S. D. Berkowitz, eds., *Social Structures: A Network Approach*

The series *Structural Analysis in the Social Sciences* presents approaches that explain social behavior and institutions by reference to *relations* among such concrete entities as persons and organizations. This contrasts with at least four other popular strategies: (a) reductionist attempts to explain by a focus on individuals alone; (b) explanations stressing the causal primacy of such abstract concepts as ideas, values, mental harmonies, and cognitive maps (thus, "structuralism" on the Continent should be distinguished from structural analysis in the present sense); (c) technological and material determinism; (d) explanations using "variables" as the main analytic concepts (as in the "structural equation" models that dominated much of the sociology of the 1970s), where structure is that connecting variables rather than actual social entities.

The social network approach is an important example of the strategy of structural analysis; the series also draws on social science theory and research that is not framed explicitly in network terms, but stresses the importance of relations rather than the atomization of reductionism or the determinism of ideas, technology, or material conditions. Though the structural perspective has become extremely popular and influential in all the social sciences, it does not have a coherent identity, and no series yet pulls together such work under a single rubric. By bringing the achievements of structurally oriented scholars to a wider public, the *Structural Analysis* series hopes to encourage the use of this very fruitful approach.

Mark Granovetter

SOCIAL NETWORK ANALYSIS: METHODS AND APPLICATIONS

STANLEY WASSERMAN
University of Illinois

KATHERINE FAUST
University of South Carolina

CAMBRIDGE
UNIVERSITY PRESS

1

Social Network Analysis in the Social and Behavioral Sciences

The notion of a *social network* and the methods of social network analysis have attracted considerable interest and curiosity from the social and behavioral science community in recent decades. Much of this interest can be attributed to the appealing focus of social network analysis on *relationships* among social entities, and on the patterns and implications of these relationships. Many researchers have realized that the network perspective allows new leverage for answering standard social and behavioral science research questions by giving precise formal definition to aspects of the political, economic, or social structural environment. From the view of social network analysis, the social environment can be expressed as patterns or regularities in relationships among interacting units. We will refer to the presence of regular patterns in relationship as *structure*. Throughout this book, we will refer to quantities that measure structure as *structural variables*. As the reader will see from the diversity of examples that we discuss, the relationships may be of many sorts: economic, political, interactional, or affective, to name but a few. The focus on relations, and the patterns of relations, requires a set of methods and analytic concepts that are distinct from the methods of traditional statistics and data analysis. The concepts, methods, and applications of social network analysis are the topic of this book.

The focus of this book is on methods and models for analyzing social network data. To an extent perhaps unequaled in most other social science disciplines, social network methods have developed over the past fifty years as an integral part of advances in social theory, empirical research, and formal mathematics and statistics. Many of the key structural measures and notions of social network analysis grew out of keen insights of researchers seeking to describe empirical phenomena and are motivated by central concepts in social theory. In addition, methods have developed to test specific hypotheses about network structural properties arising in the course of substantive research and model testing. The result of this symbiotic relationship between theory and method is a strong grounding of network analytic techniques in both application and theory. In the following sections we review the history and theory of social network analysis from the perspective of the development of methodology.

Since our goal in this book is to provide a compendium of methods and applications for both veteran social network analysts, and for naive but curious people from diverse research traditions, it is worth taking some time at the outset to lay the foundations for the social network perspective.

1.1 The Social Networks Perspective

In this section we introduce social network analysis as a distinct research perspective within the social and behavioral sciences; distinct because social network analysis is based on an assumption of the importance of relationships among interacting units. The social network perspective encompasses theories, models, and applications that are expressed in terms of relational concepts or processes. That is, relations defined by linkages among units are a fundamental component of network theories. Along with growing interest and increased use of network analysis has come a consensus about the central principles underlying the network perspective. These principles distinguish social network analysis from other research approaches (see Wellman 1988a, for example). In addition to the use of relational concepts, we note the following as being important:

- Actors and their actions are viewed as interdependent rather than independent, autonomous units
- Relational ties (linkages) between actors are channels for transfer or "flow" of resources (either material or nonmaterial)
- Network models focusing on individuals view the network structural environment as providing opportunities for or constraints on individual action
- Network models conceptualize structure (social, economic, political, and so forth) as lasting patterns of relations among actors

In this section we discuss these principles further and illustrate how the social network perspective differs from alternative perspectives in practice. Of critical importance for the development of methods for

social network analysis is the fact that the unit of analysis in network analysis is not the individual, but an entity consisting of a collection of individuals and the linkages among them. Network methods focus on dyads (two actors and their ties), triads (three actors and their ties), or larger systems (subgroups of individuals, or entire networks). Therefore, special methods are necessary.

Formal Descriptions. Network analysis enters into the process of model development, specification, and testing in a number of ways: to express relationally defined theoretical concepts by providing formal definitions, measures and descriptions, to evaluate models and theories in which key concepts and propositions are expressed as relational processes or structural outcomes, or to provide statistical analyses of multirelational systems. In this first, descriptive context, network analysis provides a vocabulary and set of formal definitions for expressing theoretical concepts and properties. Examples of theoretical concepts (properties) for which network analysis provides explicit definitions will be discussed shortly.

Model and Theory Evaluation and Testing. Alternatively, network models may be used to test theories about relational processes or structures. Such theories posit specific structural outcomes which may then be evaluated against observed network data. For example, suppose one posits that tendencies toward reciprocation of support or exchange of materials between families in a community should arise frequently. Such a supposition can be tested by adopting a statistical model, and studying how frequently such tendencies arise empirically.

The key feature of social network theories or propositions is that they require concepts, definitions and processes in which social units are linked to one another by various relations. Both statistical and descriptive uses of network analysis are distinct from more standard social science analysis and require concepts and analytic procedures that are different from traditional statistics and data analysis.

Some Background and Examples. The network perspective has proved fruitful in a wide range of social and behavioral science disciplines. Many topics that have traditionally interested social scientists can be thought of in relational or social network analytic terms. Some of the topics that have been studied by network analysts are:

- Occupational mobility (Breiger 1981c, 1990a)

- The impact of urbanization on individual well-being (Fischer 1982)
- The world political and economic system (Snyder and Kick 1979; Nemeth and Smith 1985)
- Community elite decision making (Laumann, Marsden, and Galaskiewicz 1977; Laumann and Pappi 1973)
- Social support (Gottlieb 1981; Lin, Woelfel, and Light 1986; Kadushin 1966; Wellman, Carrington, and Hall 1988; Wellman and Wortley 1990)
- Community (Wellman 1979)
- Group problem solving (Bavelas 1950; Bavelas and Barrett 1951; Leavitt 1951)
- Diffusion and adoption of innovations (Coleman, Katz, and Menzel 1957, 1966; Rogers 1979)
- Corporate interlocking (Levine 1972; Mintz and Schwartz 1981a, 1981b; Mizruchi and Schwartz 1987, and references)
- Belief systems (Erickson 1988)
- Cognition or social perception (Krackhardt 1987a; Freeman, Romney, and Freeman 1987)
- Markets (Berkowitz 1988; Burt 1988b; White 1981, 1988; Leifer and White 1987)
- Sociology of science (Mullins 1973; Mullins, Hargens, Hecht, and Kick 1977; Crane 1972; Burt 1978/79a; Michaelson 1990, 1991; Doreian and Fararo 1985)
- Exchange and power (Cook and Emerson 1978; Cook, Emerson, Gillmore, and Yamagishi 1983; Cook 1987; Markovsky, Willer, and Patton 1988)
- Consensus and social influence (Friedkin 1986; Friedkin and Cook 1990; Doreian 1981; Marsden 1990a)
- Coalition formation (Kapferer 1969; Thurman 1980; Zachary 1977)

The fundamental difference between a social network explanation and a non-network explanation of a process is the inclusion of concepts and information on *relationships* among units in a study. Theoretical concepts are relational, pertinent data are relational, and critical tests use distributions of relational properties. Whether the model employed seeks to understand individual action in the context of structured relationships, or studies structures directly, network analysis operationalizes structures in terms of networks of linkages among units. Regularities or patterns in

interactions give rise to *structures*. "Standard" social science perspectives usually ignore the relational information.

Let us explore a couple of examples. Suppose we are interested in corporate behavior in a large, metropolitan area, for example, the level and types of monetary support given to local non-profit and charitable organizations (see, for example, Galaskiewicz 1985). Standard social and economic science approaches would first define a population of relevant units (corporations), take a random sample of them (if the population is quite large), and then measure a variety of characteristics (such as size, industry, profitability, level of support for local charities or other non-profit organizations, and so forth).

The key assumption here is that the behavior of a specific unit does not influence any other units. However, network theorists take exception to this assumption. It does not take much insight to realize that there are many ways that corporations decide to do the things they do (such as support non-profits with donations). Corporations (and other such actors) tend to look at the behaviors of other actors, and even attempt to mimic each other. In order to get a complete description of this behavior, we must look at corporate to corporate relationships, such as membership on each others' boards of directors, acquaintanceships of corporate officers, joint business dealings, and other relational variables. In brief, one needs a network perspective to fully understand and model this phenomenon.

As another example, consider a social psychologist studying how groups make decisions and reach consensus (Hastie, Penrod, and Pennington 1983; Friedkin and Cook 1990; Davis 1973). The group might be a jury trying to reach a verdict, or a committee trying to allocate funds. Focusing just on the outcome of this decision, as many researchers do, is quite limiting. One really should look how members influence each other in order to make a decision or fail to reach consensus. A network approach to this study would look at interactions among group members in order to better understand the decision-making process. The influences a group member has on his/her fellow members are quite important to the process. Ignoring these influences gives an incomplete picture.

The network perspective differs in fundamental ways from standard social and behavioral science research and methods. Rather than focusing on attributes of autonomous individual units, the associations among these attributes, or the usefulness of one or more attributes for predicting the level of another attribute, the social network perspective views characteristics of the social units as arising out of structural or relational processes or focuses on properties of the relational systems themselves. The task is to understand properties of the social (economic or political) structural environment, and how these structural properties influence observed characteristics and associations among characteristics. As Collins (1988) has so aptly pointed out in his review of network theory,

> Social life is relational; it's only because, say, blacks and whites occupy particular kinds of patterns in networks in relation to each other that "race" becomes an important variable. (page 413)

In social network analysis the observed attributes of social actors (such as race or ethnicity of people, or size or productivity of collective bodies such as corporations or nation-states) are understood in terms of patterns or structures of ties among the units. Relational ties among actors are primary and attributes of actors are secondary.

Employing a network perspective, one can also study patterns of relational structures directly without reference to attributes of the individuals involved. For example, one could study patterns of trade among nations to see whether or not the world economic system exhibits a core-periphery structure. Or, one could study friendships among high school students to see whether or not patterns of friendships can be described as systems of relatively exclusive cliques. Such analyses focus on characteristics of the network as a whole and must be studied using social network concepts.

In the network analytic framework, the ties may be any relationship existing between units; for example, kinship, material transactions, flow of resources or support, behavioral interaction, group co-memberships, or the affective evaluation of one person by another. Clearly, some types of ties will be relevant or measurable for some sorts of social units but not for others. The relationship between a pair of units is a property of the pair and not inherently a characteristic of the individual unit. For example, the number (or dollar value) of Japanese manufactured automobiles exported from Japan to the United States is part of the trade relationship between Japan and the United States, and not an intrinsic characteristic of either one country or the other. In sum, the basic unit that these relational variables are measured on is the pair of actors, not one or the other individual actors. It is important for methods described in this book, that we assume that one has measurements on interactions between all possible pairs of units (for example, trade among all pairs of nations).

It is important to contrast approaches in which networks and structural properties are central with approaches that employ network ideas and measurements in standard individual-level analyses. A common usage of network ideas is to employ network measurements, or statistics calculated from these network measurements, as variables measured at the individual actor level. These derived variables are then incorporated into a more standard "cases by variables" analysis. For example, the range of a person's social support network may be used as an actor-level variable in an analysis predicting individual mental well-being (see Kadushin 1982), or occupational status attainment (Lin and Dumin 1986; Lin, Ensel, and Vaughn 1981; Lin, Vaughn, and Ensel 1981). We view analyses such as these as auxiliary network studies. Network theories and measurements become explanatory factors or variables in understanding individual behavior. We note that such an analysis still uses individual actors as the basic modeling unit. Such analyses do not focus on the network structure or network processes directly.

Our approach in this book is that network measurements are *central*. We do not discuss how to use network measurements, statistics, model parameter estimates, and so forth, in further modeling endeavors. These usual data analytic concerns are treated in existing standard statistics and methods texts.

The Perspective. Given a collection of actors, social network analysis can be used to study the structural variables measured on actors in the set. The relational structure of a group or larger social system consists of the pattern of relationships among the collection of actors. The concept of a network emphasizes the fact that each individual has ties to other individuals, each of whom in turn is tied to a few, some, or many others, and so on. The phrase "social network" refers to the set of actors and the ties among them. The network analyst would seek to model these relationships to depict the structure of a group. One could then study the impact of this structure on the functioning of the group and/or the influence of this structure on individuals within the group.

In the example of trade among nations, information on the imports and exports among nations in the world reflects the global economic system. Here the world economic system is evidenced in the observable transactions (for example, trade, loans, foreign investment, or, perhaps, diplomatic exchange) among nations. The social network analyst could then attempt to describe regularities or patterns in the world economic system and to understand economic features of individual nations (such as rate of economic development) in terms of the nation's location in the world economic system.

Network analysis can also be used to study the process of change within a group over time. Thus, the network perspective also extends longitudinally. For example, economic transactions between nations could certainly be measured at several points in time, thereby allowing a researcher to use the network prespective to study changes in the world economic system.

The social network perspective thus has a distinctive orientation in which structures, their impact, and their evolution become the primary focus. Since structures may be behavioral, social, political, or economic, social network analysis thus allows a flexible set of concepts and methods with broad interdisciplinary appeal.

1.2 Historical and Theoretical Foundations

Social network analysis is inherently an interdisciplinary endeavor. The concepts of social network analysis developed out of a propitious meeting of social theory and application, with formal mathematical, statistical, and computing methodology. As Freeman (1984) and Marsden and Laumann (1984) have documented, both the social sciences, and mathematics and statistics have been left richer from the collaborative efforts of researchers working across disciplines.

Further, and more importantly, the central concepts of relation, network, and structure arose almost independently in several social and behavioral science disciplines. The pioneers of social network analysis came from sociology and social psychology (for example, Moreno, Cartwright, Newcomb, Bavelas) and anthropology (Barnes, Mitchell). In fact, many people attribute the first use of the term "social network" to Barnes (1954). The notion of a network of relations linking social entities, or of webs or ties among social units emanating through society, has found wide expression throughout the social sciences. Furthermore, many of the structural principles of network analysis developed as researchers tried to solve empirical and/or theoretical research puzzles. The fact that so many researchers, from such different disciplines, almost simultaneously discovered the network perspective is not surprising. Its utility is great, and the problems that can be answered with it are numerous, spanning a broad range of disciplines.

In this section we briefly comment on the historical, empirical, and theoretical bases of social network methodology. Some authors have

seen network analysis as a collection of analytic procedures that are somewhat divorced from the main theoretical and empirical concerns of social research. Perhaps a particular network method may appear to lack theoretical focus because it can be applied to such a wide range of substantive problems from many different contexts. In contrast, we argue that much network methodology arose as social scientists in a range of disciplines struggled to make sense of empirical data and grappled with theoretical issues. Therefore, network analysis, rather than being an unrelated collection of methods, is grounded in important social phenomena and theoretical concepts.

Social network analysis also provides a formal, conceptual means for thinking about the social world. As Freeman (1984) has so convincingly argued, the methods of social network analysis provide formal statements about social properties and processes. Further, these concepts must be defined in precise and consistent ways. Once these concepts have been defined precisely, one can reason logically about the social world. Freeman cites *group* and *social role* as two central ideas which, until they were given formal definitions in network terms, could only serve as "sensitizing concepts." The payoff of mathematical statements of social concepts is the development of testable process models and explanatory theories. We are in full agreement with Leinhardt's statement that "it is not possible to build effective explanatory theories using metaphors" (Leinhardt 1977, page xiv). We expand on this argument in the next section.

1.2.1 Empirical Motivations

It is rare that a methodological technique is referred to as an "invention" but that is how Moreno described his early 1930's invention, the *sociogram* (Moreno 1953). This innovation, developed by Moreno along with Jennings, marked the beginning of *sociometry* (the precursor to social network analysis and much of social psychology). Starting at this time point, this book summarizes over a half-century of work in network analysis. There is wide agreement among social scientists that Moreno was the founder of the field of sociometry — the measurement of interpersonal relations in small groups — and the inspiration for the first two decades of research into the structure of small groups. Driven by an interest in understanding human social and psychological behavior, especially group dynamics, Moreno was led to invent a means for depicting the interpersonal structure of groups: the sociogram. A sociogram is a picture in which people (or more generally, any social units) are represented as points in two-dimensional space, and relationships among pairs of people are represented by lines linking the corresponding points. Moreno claimed that "before the advent of sociometry no one knew what the interpersonal structure of a group 'precisely' looked like" (1953, page lvi). This invention was revealed to the public in April 1933 at a convention of medical scholars, and was found to be so intriguing that the story was immediately picked up by *The New York Times* (April 3, 1933, page 17), and carried in newspapers across the United States. Moreno's interest went far beyond mere depiction. It was this need to model important social phenomena that led to two of the mainstays of social network analysis: a visual display of group structure, and a probabilistic model of structural outcomes.

Visual displays including sociograms and two or higher dimensional representations continue to be widely used by network analysts (see Klovdahl 1986; Woelfel, Fink, Serota, Barnett, Holmes, Cody, Saltiel, Marlier, and Gillham 1977). Two and sometimes three-dimensional spatial representations (using multidimensional scaling) have proved quite useful for presenting structures of influence among community elites (Laumann and Pappi 1976; Laumann and Knoke 1987), corporate interlocks (Levine 1972), role structures in groups (Breiger, Boorman, and Arabie 1975; Burt 1976, 1982), and interaction patterns in small groups (Romney and Faust 1982; Freeman, Freeman, and Michaelson 1989).

Recognition that sociograms could be used to study social structure led to a rapid introduction of analytic techniques. The history of this development is nicely reviewed by Harary, Norman, and Cartwright (1965), who themselves helped pioneer this development. At the same time, methodologists discovered that matrices could be used to represent social network data. These recognitions and discoveries brought the power of mathematics to the study of social systems. Forsyth and Katz (1946), Katz (1947), Luce and Perry (1949), Bock and Husain (1950, 1952), and Harary and Norman (1953) were the first to use matrices in novel methods for the study of social networks.

Other researchers also found inspiration for network ideas in the course of empirical research. In the mid-1950's, anthropologists studying urbanization (especially British anthropologists — such as Mitchell and Barnes) found that the traditional approach of describing social organization in terms of institutions (economics, religion, politics, kinship, etc.) was not sufficient for understanding the behavior of individuals in complex societies (Barnes 1954; Bott 1957; Mitchell 1969; Boissevain 1968;

Kapferer 1969). Furthermore, as anthropologists turned their attention to "complex" societies, they found that new concepts were necessary in order to understand the fluid social interactions they observed in the course of ethnographic field work (for example, see Barnes 1954, 1969a; Boissevain 1968; also Mitchell 1969; and Boissevain and Mitchell 1973, and papers therein). Barnes (1972), Whitten and Wolfe (1973), Mitchell (1974), Wolfe (1978), Foster (1978/79), and others provide excellent reviews of the history of social network ideas in anthropology. Many of the current formal concepts in social network analysis, for example, density (Bott 1957), span (Thurman 1980), connectedness, clusterability, and multiplexity (Kapferer 1969), were introduced in the 1950's and 1960's as ways to describe properties of social structures and individual social environments. Network analysis provided both a departure in theoretical perspective and a way of talking about social phenomena which were not easily defined using then current terminology.

Many social psychologists of the 1940's and 1950's found experimental structures useful for studying group processes (Leavitt 1949, 1951; Bavelas 1948, 1950; Smith 1950; and many others; see Freeman, Roeder, and Mulholland 1980, for a review). The experimentally designed communication structures employed by these researchers lent themselves naturally to graphical representations using points to depict actors and lines to depict channels of communication. Key insights from this research program indicated that there were both important properties of group structures and properties of individual positions within these structures. The theory of the impact of structural arrangement on group problem solving and individual performance required formal statements of the structural properties of these experimental arrangements. Structural properties found by these researchers include the notions of actor *centrality* and group *centralization*.

Clearly, important empirical tendencies led to important new, network methods. Very important findings of tendencies toward reciprocity or *mutuality* of positive affect, structural balance, and transitivity, discovered early in network analysis, have had a profound impact on the study of social structure. Bronfenbenner (1943) and Moreno and Jennings (1945) were the first to study such tendencies quantitatively.

1.2.2 Theoretical Motivations

Theoretical notions have also provided impetus for development of network methods. Here, we explore some of the theoretical concepts that have motivated the development of specific network analysis methods. Among the important examples are: social group, isolate, popularity, liaison, prestige, balance, transitivity, clique, subgroup, social cohesion, social position, social role, reciprocity, mutuality, exchange, influence, dominance, conformity. We briefly introduce some of these ideas below, and discuss them all in more detail as they arise in later chapters.

Conceptions of social group have led to several related lines of methodological development. Sociologists have used the phrase "social group" in numerous and imprecise ways. Social network researchers have taken specific aspects of the theoretical idea of social group to develop more precise social network definitions. Among the more influential network group ideas are: the graph theoretic entity of a *clique* and its generalizations (Luce and Perry 1949; Alba 1973; Seidman and Foster 1978a; Mokken 1979; and Freeman 1988); the notion of an interacting community (see Sailer and Gaulin 1984); and social circles and structures of affiliation (Kadushin 1966; Feld 1981; Breiger 1974; Levine 1972; McPherson 1982). The range and number of mathematical definitions of "group" highlights the usefulness of using network concepts to specify exact properties of theoretical concepts.

Another important theoretical concept, *structural balance*, was postulated by Heider during the 1940's (Heider 1946), and later Newcomb (1953). Balanced relations were quite common in empirical work; consequently, theorists were quick to pose theories about why such things occurred so frequently. This concept led to a very active thirty-year period of empirical, theoretical, and quantitative research on triples of individuals.

Balance theory was quantified by mathematicians using graph theoretical concepts (Harary 1953, 1955b). Balance theory also influenced the development of a large number of structural theories, including *transitivity*, another theory postulated at the level of a triple of individuals.

The related notions of social *role*, social *status*, and social *position* have spawned a wide range of network analysis methods. Lorrain and White were among the first social network analysts to express in social network terms the notion of social role (Lorrain and White 1971). Their foundational work on the mathematical property of *structural equivalence* (individuals who have identical ties to and from all others in a network) expressed the social concept of role in a formal mathematical procedure. Much of the subsequent work on this topic has centered on appropriate conceptualizations of notions of position (Burt 1976; Faust 1988; Borgatti and Everett 1992a) or role (White and Reitz 1983, 1989;

Winship and Mandel 1983; Breiger and Pattison 1986) in social network terms.

1.2.3 Mathematical Motivations

Early in the theoretical development of social network analysis, researchers found use for mathematical models. Beginning in the 1940's with attempts to quantify tendencies toward *reciprocity*, social network analysts have been frequent users and strong proponents of quantitative analytical approaches. The three major mathematical foundations of network methods are graph theory, statistical and probability theory, and algebraic models. Early sociometricians discovered graph theory and distributions for random graphs (for example, the work of Moreno, Jennings, Criswell, Harary, and Cartwright). Mathematicians had long been interested in graphs and distributions for graphs (see Erdös and Renyi 1960, and references therein), and the more mathematical social network analysts were quick to pick up models and methods from the mathematicians. Graph theory provides both an appropriate representation of a social network and a set of concepts that can be used to study formal properties of social networks.

Statistical theory became quite important as people began to study reciprocity, mutuality, balance, and transitivity. Other researchers, particularly Katz and Powell (1955), proposed indices to measure tendencies toward reciprocation.

Interest in reciprocity, and pairs of interacting individuals, led to a focus on threesomes. Empirical and theoretical work on balance theory and transitivity motivated a variety of mathematicians and statisticians to formulate mathematical models for behavior of triples of actors. Cartwright and Harary (1956) were the first to quantify structural balance propositions, and along with Davis (1967), discussed which types of triads (triples of actors and all observed relational linkages among the actors) should and should not arise in empirical research. Davis, Holland, and Leinhardt, in a series of papers written in the 1970's, introduced a wide variety of random directed graph distributions into social network analysis, in order to test hypotheses about various structural tendencies.

During the 1980's, research on statistical models for social networks heightened. Models now exist for analyzing a wide variety of social network data. Simple log linear models of dyadic interactions are now commonly used in practice. These models are often based on Holland and Leinhardt's (1981) p_1 probability distribution for relational data.

This model can be extended to dyadic interactions that are measured on a nominal or an ordinal scale. Additional generalizations allow one to simultaneously model multivariate relational networks. Network interactions on different relations may be associated, and the interactions of one relation with others allow one to study how associated the relational variables are. In the mid-1970's, there was much interest in models for the study of networks over time. Mathematical models, both deterministic and stochastic, are now quite abundant for such study.

Statistical models are used to test theoretical propositions about networks. These models allow the processes (which generate the data) to show some error, or lack of fit, to proposed structural theories. One can then compare data to the predictions generated by the theories to determine whether or not the theories should be rejected.

Algebraic models have been widely used to study multirelational networks. These models use algebraic operations to study combinations of relations (for example, "is a friend of," "goes to for advice," and "is a friend of a friend") and have been used to study kinship systems (White 1963; Boyd 1969) and network role structures (Boorman and White 1976; Breiger and Pattison 1986; Boyd 1990; and Pattison 1993).

Social network analysis attempts to solve analytical problems that are non-standard. The data analyzed by network methods are quite different from the data typically encountered in social and behavioral sciences. In the traditional data analytic framework one assumes that one has a set of measurements taken on a set of independent units or cases; thus giving rise to the familiar "cases by variables" data array. The assumption of sampling independence of observations on individual units allows the considerable machinery of statistical analysis to be applied to a range of research questions. However, social network analysis is explicitly interested in the interrelatedness of social units. The dependencies among the units are measured with structural variables. Theories that incorporate network ideas are distinguished by propositions about the relations among social units. Such theories argue that units are not acting independently from one another, but rather influence each other. Focusing on such structural variables opens up a different range of possibilities for, and constraints on, data analysis and model building.

1.2.4 *In Summary*

The historical examination of empirical, theoretical, and mathematical developments in network research should convince the reader that social network analysis is far more than an intuitively appealing vocabulary, metaphor, or set of images for discussing social, behavioral, political, or economic relationships. Social network analysis provides a precise way to define important social concepts, a theoretical alternative to the assumption of independent social actors, and a framework for testing theories about structured social relationships.

The methods of network analysis provide explicit formal statements and measures of social structural properties that might otherwise be defined only in metaphorical terms. Such phrases as webs of relationships, closely knit networks of relations, social role, social position, group, clique, popularity, isolation, prestige, prominence, and so on are given mathematical definitions by social network analysis. Explicit mathematical statements of structural properties, with agreed upon formal definitions, force researchers to provide clear definitions of social concepts, and facilitate development of testable models. Furthermore, network analysis allows measurement of structures and systems which would be almost impossible to describe without relational concepts, and provides tests of hypotheses about these structural properties.

1.3 Fundamental Concepts in Network Analysis

There are several key concepts at the heart of network analysis that are fundamental to the discussion of social networks. These concepts are: actor, relational tie, dyad, triad, subgroup, group, relation, and network. In this section, we define some of these key concepts and discuss the different levels of analysis in social networks.

Actor. As we have stated above, social network analysis is concerned with understanding the linkages among social entities and the implications of these linkages. The social entities are referred to as *actors.* Actors are discrete individual, corporate, or collective social units. Examples of actors are people in a group, departments within a corporation, public service agencies in a city, or nation-states in the world system. Our use of the term "actor" does not imply that these entities necessarily have volition or the ability to "act." Further, most social network applications focus on collections of actors that are all of the same type (for example, people in a work group). We call such collections *one-mode networks.* However, some methods allow one to look at actors of conceptually different types or levels, or from different sets. For example, Galaskiewicz (1985) and Galaskiewicz and Wasserman (1989) analyzed monetary donations made from corporations to nonprofit agencies in the Minneapolis/St. Paul area. Doreian and Woodard (1990) and Woodard and Doreian (1990) studied community members' contacts with public service agencies.

Relational Tie. Actors are linked to one another by social *ties*. As we will see in the examples discussed throughout this book, the range and type of ties can be quite extensive. The defining feature of a tie is that it establishes a linkage between a pair of actors. Some of the more common examples of ties employed in network analysis are:

- Evaluation of one person by another (for example expressed friendship, liking, or respect)
- Transfers of material resources (for example business transactions, lending or borrowing things)
- Association or affiliation (for example jointly attending a social event, or belonging to the same social club)
- Behavioral interaction (talking together, sending messages)
- Movement between places or statuses (migration, social or physical mobility)
- Physical connection (a road, river, or bridge connecting two points)
- Formal relations (for example authority)
- Biological relationship (kinship or descent)

We will expand on these applications and provide concrete examples of different kinds of ties in the discussion of network applications and data in Chapter 2.

Dyad. At the most basic level, a linkage or relationship establishes a tie between two actors. The tie is inherently a property of the pair and therefore is not thought of as pertaining simply to an individual actor. Many kinds of network analysis are concerned with understanding ties among pairs. All of these approaches take the *dyad* as the unit of analysis. A dyad consists of a pair of actors and the (possible) tie(s) between them. Dyadic analyses focus on the properties of pairwise relationships, such as whether ties are reciprocated or not, or whether specific types of multiple relationships tend to occur together. Dyads are discussed in detail in Chapter 13, while dyadic statistical models are discussed in Chapters 15 and 16. As we will see, the dyad is frequently the basic unit for the statistical analysis of social networks.

Triad. Relationships among larger subsets of actors may also be studied. Many important social network methods and models focus on the *triad*; a subset of three actors and the (possible) tie(s) among them. The analytical shift from pairs of individuals to triads (which consist of three potential pairings) was a crucial one for the theorist Simmel, who wrote in 1908 that

> ...the fact that two elements [in a triad] are each connected not only by a straight line – the shortest – but also by a broken line, as it were, is an enrichment from a formal-sociological standpoint. (page 135)

Balance theory has informed and motivated many triadic analyses. Of particular interest are whether the triad is transitive (if actor i "likes" actor j, and actor j in turn "likes" actor k, then actor i will also "like" actor k), and whether the triad is balanced (if actors i and j like each other, then i and j should be similar in their evaluation of a third actor, k, and if i and j dislike each other, then they should differ in their evaluation of a third actor, k).

Subgroup. Dyads are pairs of actors and associated ties, triads are triples of actors and associated ties. It follows that we can define a *subgroup* of actors as any subset of actors, and all ties among them. Locating and studying subgroups using specific criteria has been an important concern in social network analysis.

Group. Network analysis is not simply concerned with collections of dyads, or triads, or subgroups. To a large extent, the power of network analysis lies in the ability to model the relationships among systems of actors. A system consists of ties among members of some (more or less bounded) group. The notion of group has been given a wide range of definitions by social scientists. For our purposes, a *group* is the collection of all actors on which ties are to be measured. One must be able to argue by theoretical, empirical, or conceptual criteria that the actors in the group belong together in a more or less bounded set. Indeed, once one decides to gather data on a group, a more concrete meaning of the term is necessary. A group, then, consists of a finite set of actors who for conceptual, theoretical, or empirical reasons are treated as a finite set of individuals on which network measurements are made.

The restriction to a *finite* set or sets of actors is an analytic requirement. Though one could conceive of ties extending among actors in a nearly infinite group of actors, one would have great difficulty analyzing data on such a network. Modeling finite groups presents some of the more problematic issues in network analysis, including the specification of network boundaries, sampling, and the definition of group. Network sampling and boundary specification are important issues.

Early network researchers clearly recognized extensive ties among individuals (de Sola Pool and Kochen 1978; see Kochen 1989 for recent work on this topic). Indeed, some early social network research looked at the "small world" phenomenon: webs and chains of connections emanating to and from an individual, extending throughout the larger society (Milgram 1967; Killworth and Bernard 1978).

However, in research applications we are usually forced to look at finite collections of actors and ties between them. This necessitates drawing some boundaries or limits for inclusion. Most network applications are limited to a single (more or less bounded) group; however, we could study two or more groups.

Throughout the book, we will refer to the entire collection of actors on which we take measurements as the *actor set*. A network can contain many groups of actors, but only one (if it is a one-mode network) actor set.

Relation. The collection of ties of a specific kind among members of a group is called a *relation*. For example, the set of friendships among pairs of children in a classroom, or the set of formal diplomatic ties maintained by pairs of nations in the world, are ties that define relations. For any group of actors, we might measure several different relations (for example, in addition to formal diplomatic ties among nations, we might also record the dollar amount of trade in a given year). It is important to note that a relation refers to the collection of ties of a given kind measured on pairs of actors from a specified actor set. The ties themselves only exist between specific pairs of actors.

Social Network. Having defined actor, group, and relation we can now give a more explicit definition of social network. A *social network* consists of a finite set or sets of actors and the relation or relations defined on them. The presence of relational information is a critical and defining feature of a social network. A much more mathematical definition of a social network, but consistent with the simple notion given here, can be found at the end of Chapter 3.

In Summary. These terms provide a core working vocabulary for discussing social networks and social network data. We can see that social network analysis not only requires a specialized vocabulary, but also deals with conceptual entities and research problems that are quite difficult to pursue using a more traditional statistical and data analytic framework.

We now turn to some of the distinctive features of network analysis.

1.4 Distinctive Features of Network Theory and Measurement

It is quite important to note the key features that distinguish network theory, and consequently network measurement, from the more usual data analytic framework common in the social and behavioral sciences. Such features provide the necessary motivation for the topics discussed in this book.

The most basic feature of network measurement, distinctive from other perspectives, is the use of structural or relational information to study or test theories. Many network analysis methods provide formal definitions and descriptions of structural properties of actors, subgroups of actors, or groups. These methods translate core concepts in social and behavioral theories into formal definitions expressed in relational terms. All of these concepts are quantified by considering the relations measured among the actors in a network.

Because network measurements give rise to data that are unlike other social and behavioral science data, an entire body of methods has been developed for their analysis. Social network data require measurements on ties among social units (or actors); however, attributes of the actors may also be collected. Such data sets need social network methods for analysis. One cannot use multiple regression, *t*-tests, canonical correlations, structural equation models, and so forth, to study social network data or to test network theories. This book exists to organize, present, critique, and demonstrate the large body of methods for social network analysis.

Social network analysis may be viewed as a broadening or generalization of standard data analytic techniques and applied statistics which usually focus on observational units and their characteristics. A social network analysis must consider data on ties among the units. However, attributes of the actors may also be included.

Measurements on actors will be referred to as network *composition*. Complex network data sets may contain information about the characteristics of the actors (such as the gender of people in a group, or the GNP of nations in the world), as well as structural variables. Thus, the sort of data most often analyzed in the social and behavioral sciences (cases and variables) may also be incorporated into network models. But the fact that one has not only structural, but also compositional, variables can lead to very complicated data sets that can be approached only with sophisticated graph theoretic, algebraic, and/or statistical methods.

Social network theories require specification in terms of patterns of relations, characterizing a group or social system as a whole. Given appropriate network measurements, these theories may be stated as propositions about group relational structure. Network analysis then provides a collection of descriptive procedures to determine how the system behaves, and statistical methods to test the appropriateness of the propositions. In contrast, approaches that do not include network measurements are unable to study and/or test such theories about structural properties.

Network theories can pertain to units at different levels of aggregation: individual actors, dyads, triads, subgroups, and groups. Network analysis provides methods to study structural properties and to test theories stated at all of these levels. The network perspective, the theories, and the measurements they spawn are thus quite wide-ranging. This is quite unique in the social and behavioral sciences. Rarely does a standard theory lead to theoretical statements and hence measurements at more than a single level.

1.5 Organization of the Book and How to Read It

The question now is how to make sense of the more than 700 pages sitting in front of you. First, find a comfortable chair with good reading light (shoo the cats, dogs, and children away, if necessary). Next, make sure your cup of coffee (or glass of scotch, depending on the time of day) is close at hand, put a nice jazz recording on the stereo, and have a pencil or highlighting pen available (there are *many* interesting points throughout the book, and we are *sure* you will want to make note of them).

This book is organized to highlight several themes in network analysis, and to be accessible to readers with different interests and sophistication in social network analysis. We have mentioned these themes throughout this chapter, and now describe how these themes help to organize the methods discussed in this book. These themes are:

- The complexity of the methods
- Descriptive versus statistical methods

- The theoretical motivation for the methods
- The chronological development of the methods
- The level of analysis to which the methods are appropriate

Since social network analysis is a broad, diverse, and theoretically varied field, with a long and rich history, it is impossible to reflect all of these possible thematic organizations simultaneously. However, insofar as is practical and useful, we have tried to use these themes in the organization of the book.

1.5.1 Complexity

First, the material progresses from simple to complex. The remainder of Part I reviews applications of network analysis, gives an overview of network analysis methods in a general way, and then presents notation to be used throughout the book. Part II presents graph theory, develops the vocabulary and concepts that are widely used in network analysis, and relies heavily on examples. It also discusses simple actor and group properties. Parts II, III, and IV require familiarity with algebra, and a willingness to learn some graph theory (presented in Chapter 4). Parts V and VI require some knowledge of statistical theory. Log linear models for dyadic probabilities provide the basis for many of the techniques presented later in these chapters.

1.5.2 Descriptive and Statistical Methods

Network methods can be dichotomized into those that are descriptive versus those that are based on probabilistic assumptions. This dichotomy is an important organizational categorization of the methods that we discuss. Parts II, III, and IV of the book are based on the former. The methods presented in these three parts of the book assume specific descriptive models for the structure of a network, and primarily present descriptive techniques for network analysis which translate theoretical concepts into formal measures.

Parts V and VI are primarily concerned with methods for testing network theories and with statistical models of structural properties. In contrast to a descriptive approach, we can also begin with stochastic assumptions about actor behavior. Such models assume that there is some probabilistic mechanism (even as simple as flipping a coin) that underlies observed, network data. For example, one can focus on dyadic interactions, and test whether an observed network has a specified amount of reciprocity in the ties among the actors. Such a test uses standard statistical theory, and thus one can formally propose a null hypothesis which can then be rejected or not. Much of Chapter 13 is devoted to a description of these mechanisms, which are then used throughout Chapters 14, 15, and 16.

1.5.3 Theory Driven Methods

As we have discussed here, many social network methods were developed by researchers in the course of empirical investigation and the development of theories. This categorization is one of the most important of the book.

Part III covers approaches to groups and subgroups, notably cliques and their generalizations. Sociological tendencies such as cohesion and influence, which can cause actors to be "clustered" into subgroups, are among the topics of Chapters 7 and 8. Part IV discusses approaches related to the sociological notions of social role, status and position, and the mathematical property of structural equivalence and its generalizations. The later sections of the book present statistical methods for the analysis of social networks, many of which are motivated by theoretical concerns. Part V covers models for dyadic and triadic structure, early sociometry and social psychology of affective relations (dyadic analyses of Chapter 13), and structural balance and transitivity (triadic analyses of Chapters 6 and 14).

1.5.4 Chronology

It happens that the chapters in this book are approximately chronological. The important empirical investigations of social networks began over sixty years ago, starting with the sociometry of Moreno. This research led to the introduction of graph theory (Chapter 4) to study structural properties in the late 1940's and 1950's, and methods for subgroups and cliques (Chapter 7), as well as structural balance and transitivity (Chapters 6 and 14). More recently, H. White and his collaborators, using the sociological ideas of formal role analysis (Nadel and Lorrain), introduced structural equivalence (Chapter 9), and an assortment of related methods, in the 1970's, which in the 1980's, led to a collection of algebraic network methods (Chapters 11 and 12).

As can be seen from our table of contents, we have mostly followed this chronological order. We start with graph theory in Chapter 4, and discuss descriptive methods in Parts III and IV before moving on to the more recent statistical developments covered in Parts V and VI. However, because of our interest in grouping together methods with similar substantive and theoretical concerns, a few topics are out of historical sequence (structural balance and triads in Chapters 6 and 14 for example). Thus, Part V (Dyadic and Triadic Methods) follows Part IV (Roles and Positions). This reversal was made to place dyadic and triadic methods next to the other statistical methods discussed in the book (Part VI), since the methods for studying dyads and triads were among the first statistical methods for networks.

1.5.5 Levels of Analysis

Network methods are usually appropriate for concepts at certain levels of analysis. For example, there are properties and associated methods pertaining just to the actors themselves. Examples include how "prominent" an actor is within a group, as quantified by measures such as centrality and prestige (Chapter 5), actor-level expansiveness and popularity parameters embedded in stochastic models (Chapters 15 and 16), and measures for individual roles, such as isolates, liaisons, bridges, and so forth (Chapter 12). Then there are methods applicable to pairs of actors and the ties between them, such as those from graph theory that measure actor distance and reachability (Chapter 4), structural and other notions of equivalence (Chapters 9 and 12), dyadic analyses that postulate statistical models for the various states of a dyad (Chapter 13), and stochastic tendencies toward reciprocity (Chapter 15). Triadic methods are almost always based on theoretical statements about balance and transitivity (Chapter 6), and postulate certain behaviors for triples of actors and the ties among them (Chapter 14).

Many methods allow a researcher to find and study subsets of actors that are homogeneous with respect to some network properties. Examples of such applications include: cliques and other cohesive subgroups that contain actors who are "close" to each other (Chapter 7), positions of actors that arise via positional analysis (Chapters 9 and 10), and subgroups of actors that are assumed to behave similarly with respect to certain model parameters arising from stochastic models (Chapter 16). Lastly, there are measures and methods that focus on entire groups and all ties. Graph theoretic measures such as connectedness and diameter

(Chapter 4), group-level measures of centralization, density, and prestige (Chapter 5), as well as blockmodels and role algebras (Chapters 9, 10, and 11) are examples of group-level methods.

1.5.6 Chapter Prerequisites

Finally, it is important to note that some chapters are prerequisites for others, while a number of chapters may be read without reading all intervening chapters. This ordering of chapters is presented in Figure 1.1. A line in this figure connects two chapters if the earlier chapter contains material that is necessary in order to read the later chapter. Chapters 1, 2, 3, and 4 contain the introductory material, and should be read before all other chapters. These chapters discuss social network data, notation, and graph theory.

From Chapter 4 there are five possible branches: Chapter 5 (centrality); Chapter 6 (balance, clusterability, and transitivity); Chapter 7 (cohesive subgroups); Chapter 9 (structural equivalence); or Chapter 13 (dyads). Chapter 8 (affiliation networks) follows Chapter 7; Chapters 10 (blockmodels), 11 (relational algebras), and 12 (network role and position) follow, in order, from Chapter 9; Chapter 15 (statistical analysis) follows Chapter 13. Chapter 14 requires both Chapters 13 and 6. Chapter 16 (stochastic blockmodels and goodness-of-fit) requires both Chapters 15 and 10. Lastly, Chapter 17 concludes the book (and is an epilogue to all branches).

A good overview of social network analysis (with an emphasis on descriptive approaches including graph theory, centrality, balance and clusterability, cohesive subgroups, structural equivalence, and dyadic models) could include Chapters 1 through 10 plus Chapter 13. This material could be covered in a one semester graduate course. Alternatively, one could omit Chapter 8 and include Chapters 15 and 16, for a greater emphasis on statistical approaches.

One additional comment — throughout the book, you will encounter two symbols used to label sections: ○ and ⊗. The symbol ○ implies that the text that follows is tangential to the rest of the chapter, and can be omitted (except by the curious). The symbol ⊗ implies that the text that follows requires more thought and perhaps more mathematical and/or statistical knowledge than the other parts of the chapter, and should be omitted (except by the brave).

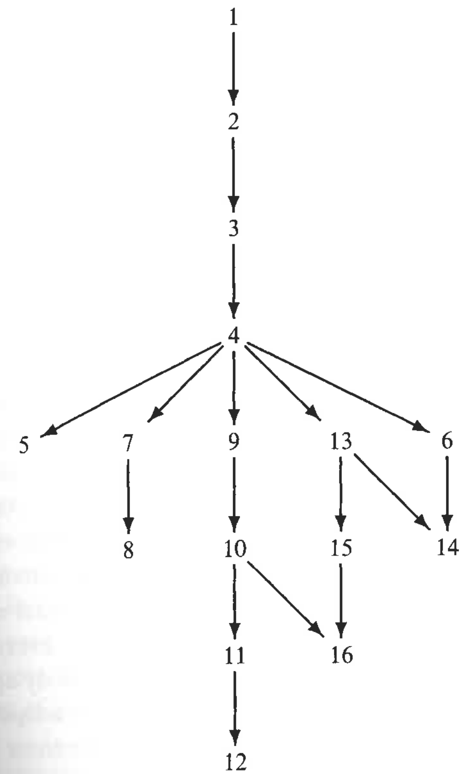

Fig. 1.1. How to read this book

1.6 Summary

We have just described the history and motivations for social network analysis. Network theories and empirical findings have been the primary reasons for the development of much of the methodology described in this book.

A complete reading of this book, beginning here and continuing on to the discussion of network data in Chapter 2, then notation in Chapter 3, and so forth, should provide the reader with a knowledge of network methods, theories, and histories. So without further ado, let us begin....