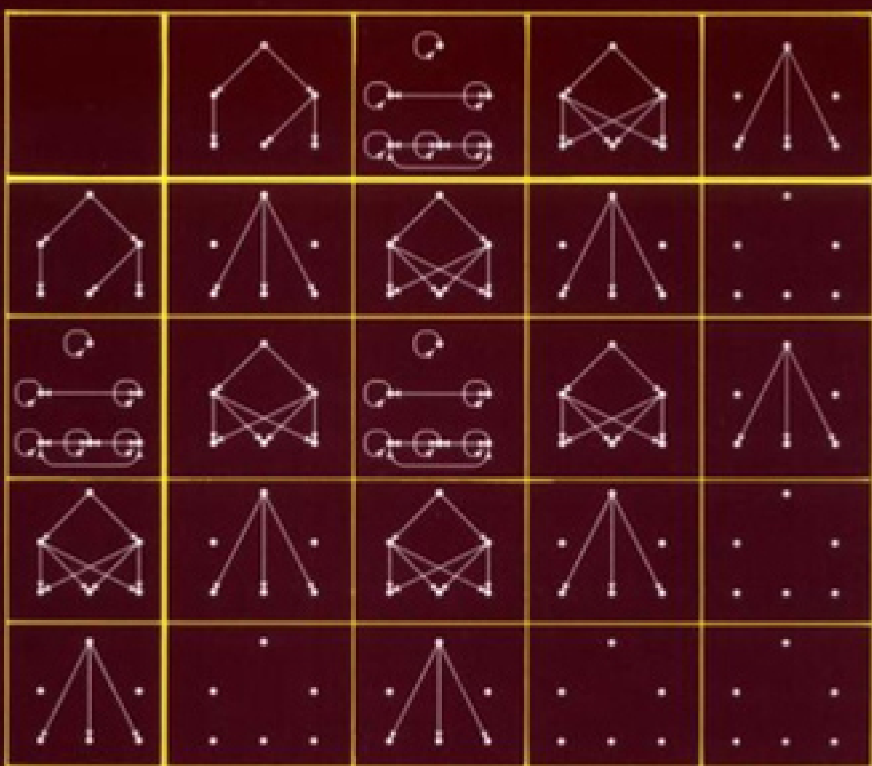
STRUCTURAL ANALYSIS IN THE SOCIAL SCIENCES 8

Social Network Analysis

Methods and Applications

Stanley Wasserman and Katherine Faust



CAMBRIDGE

more information - www.cambridge.org/9780521387071

Social network analysis is used widely in the social and behavioral sciences, as well as in economics, marketing, and industrial engineering. The social network perspective focuses on relationships among social entities; examples include communications among members of a group, economic transactions between corporations, and trade or treaties among nations. The focus on relationships is an important addition to standard social and behavioral research, which is primarily concerned with attributes of the social units.

Social Network Analysis: Methods and Applications reviews and discusses methods for the analysis of social networks with a focus on applications of these methods to many substantive examples. The book is organized into six parts. The introductory chapters give an overview of the social network perspective and describe different kinds of social network data. The second part discusses formal representations for social networks, including notations, graph theory, and matrix operations. The third part covers structural and locational properties of social networks, including centrality, prestige, prominence, structural balance, clusterability, cohesive subgroups, and affiliation networks. The fourth part examines methods for social network roles and positions and includes discussions of structural equivalence, blockmodels, and relational algebras. The properties of dyads and triads are covered in the fifth part of the book, and the final part discusses statistical methods for social networks.

Social Network Analysis: Methods and Applications is a reference book that can be used by those who want a comprehensive review of network methods, or by researchers who have gathered network data and want to find the most appropriate method by which to analyze them. It is also intended for use as a textbook, as it is the first book to provide comprehensive coverage of the methodology and applications of the field.

SOCIAL NETWORK ANALYSIS

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

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521387071

First published 1994
Reprinted 1995
Reprinted with corrections 1997
19th printing 2009

Printed in the United States of America

A catalog record for this publication is available from the British Library.

ISBN 978-0-521-38269-4 Hardback
ISBN 978-0-521-38707-1 Paperback

To Sarah

and

To Don and Margaret Faust

Contents

List of Tables

List of Illustrations

Preface

Our goal for this book is to present a review of network analysis methods, a reference work for researchers interested in analyzing relational data, and a text for novice social networkers looking for an overview of the field. Our hope is that this book will help researchers to become aware of the very wide range of social network methods, to understand the theoretical motivations behind these approaches, to appreciate the wealth of social network applications, and to find some guidance in selecting the most appropriate methods for a given research application.

The last decade has seen the publication of several books and edited volumes dealing with aspects of social network theory, application, and method. However, none of these books presents a comprehensive discussion of social network methodology. We hope that this book will fill this gap. The theoretical basis for the network perspective has been extensively outlined in books by Berkowitz (1982) and Burt (1982). Because these provide good theoretical overviews, we will not dwell on theoretical advances in social network research, except as they pertain directly to network methods. In addition, there are several collections of papers that apply network ideas to substantive research problems (Leinhardt 1977; Holland and Leinhardt 1979; Marsden and Lin 1982; Wellman and Berkowitz 1988; Breiger 1990a; Hiramatsu 1990; Weesie and Flap 1990; Wasserman and Galaskiewicz 1994). These collections include foundational works in network analysis and examples of applications from a range of disciplines.

Finally, some books have presented collections of readings on special topics in network methods (for example, Burt and Minor 1983), papers on current methodological advances (for example, Freeman, White and Romney 1989), or elementary discussions of basic topics in network analysis (for example, Knoke and Kuklinski 1982; Scott 1992). And there are a number of monographs and articles reviewing network methodology (Northway 1952; Lindzey and Borgatta 1954; Mitchell 1974; Roistacher 1974; Freeman 1976; Burt 1978b; Feger, Hummell, Pappi, Sodeur, and Ziegler 1978; Klovdahl 1979; Niesmoller and Schijf 1980; Burt 1980; Alba 1981; Frank 1981; Wellman 1983; Rice and Richards 1985; Scott 1988; Wellman 1988a; Wellman and Berkowitz 1988; Marsden 1990b). Very recently, a number of books have begun to appear, discussing advanced methodological topics. Hage and Harary (1983) is a good example from this genre; Boyd (1990), Breiger (1991), and Pattison (1993) introduce the reader to other specialized topics.

However, the researcher seeking to understand network analysis is left with a void between the elementary discussions and sophisticated analytic presentations since none of these books provides a unified discussion of network methodology. As mentioned, we intend this book to fill that void by presenting a broad, comprehensive, and, we hope, complete discussion of network analysis methodology.

There are many people to thank for their help in making this book a reality. Mark Granovetter, the editor of this series for Cambridge University Press, was a source of encouragement throughout the many years that we spent revising the manuscript. Lin Freeman, Ron Breiger, and Peter Marsden reviewed earlier versions of the book for Cambridge, and made many, many suggestions for improvement. Alaina Michaelson deserves much gratitude for actually reading the *entire* manuscript during the 1990–1991 academic year. Sue Freeman, Joe Galaskiewicz, Nigel Hopkins, Larry Hubert, Pip Pattison, Kim Romney, and Tom Snijders read various chapters, and had many helpful comments. Colleagues at the University of South Carolina Department of Sociology (John Skvoretz, Pat Nolan, Dave Willer, Shelley Smith, Jimy Sanders, Lala Steelman, and Steve Borgatti) were a source of inspiration, as were Phipps Arabie, Frank Romo, and Harrison White. Dave Krackhardt, John Padgett, Russ Bernard, Lin Freeman, and Joe Galaskiewicz shared data with us. Our students Carolyn Anderson, Mike Walker, Diane Payne, Laura Koehly, Shannon Morrison, and Melissa Abboushi were wonderful assistants. Jill Grace provided library assistance. We also thank the authors of the computer programs we used to help analyze the data in the book — Karel Sprenger and Frans Stokman (*GRADAP*), Ron Breiger (*ROLE*), Noah Friedkin (*SNAPS*), Ron Burt (*STRUCTURE*), and Lin Freeman, Steve Borgatti, and Martin Everett (*UCINET*). And, of course, we are extremely grateful to Allison, Drew, Eliot, Keith, Ross, and Sarah for their notoriety!

Emily Loose, our first editor at Cambridge, was always helpful in finding ways to speed up the process of getting this book into print. Elizabeth Neal and Pauline Ireland at Cambridge helped us during the last stages of production. Hank Heitowit, of the Interuniversity Consortium for Political and Social Research at the University of Michigan (Ann Arbor) made it possible for us to teach a course, Social Network Analysis, for the last seven years in their Summer Program in Quantitative Methods. The students at ICPSR, as well as the many students at the University of Illinois at Urbana-Champaign, the University of South Carolina, American University, and various workshops we have given deserve special recognition. And lastly, we thank Murray Aborn, Jim Blackman, Sally Nerlove, and Cheryl Eavey at the National Science Foundation for financial support over the years (most recently, via NSF Grant #SBR93-10184 to the University of Illinois).

We dedicate this book to Sarah Wasserman, and to Don Faust and Margaret Faust, without whom it would not have been possible.

Stanley Wasserman
Grand Rivers, Kentucky

Katherine Faust
Shaver Lake, California

August, 1993

Note to Revised Printing: We thank our readers for finding the errors (both typographical and otherwise) in the original that have been corrected in this revised printing.

September, 1996

Part I

Networks, Relations, and Structure

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).

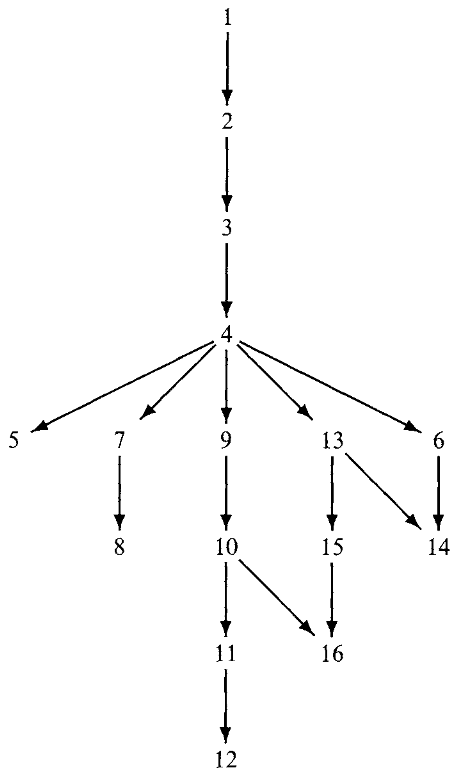


Fig. 1.1. How to read this book

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....

2

Social Network Data: Collection and Applications

This chapter discusses characteristics of social network data, with an emphasis on how to collect such data sets. We categorize network data in a variety of ways, and illustrate these categories with examples. We also describe the data sets that we use throughout the book. As noted in Chapter 1, the most important difference between social network data and standard social and behavioral science data is that network data include measurements on the relationships between social entities. Most of the standard data collection procedures known to every social scientist are appropriate for collecting network data (if properly applied), but there are a few techniques that are specific to the investigation of social networks. We highlight these similarities and differences in this chapter.

2.1 Introduction: What Are Network Data?

Social network data consist of at least one structural variable measured on a set of actors. The substantive concerns and theories motivating a specific network study usually determine which variables to measure, and often which techniques are most appropriate for their measurement. For example, if one is studying economic transactions between countries, one cannot (easily) rely on observational techniques; one would probably use archival records to obtain information on such transactions. On the other hand, friendships among people are most likely studied using questionnaires or interviews, rather than using archival or historical records. In addition, the nature of the study determines whether the entire set of actors can be surveyed or whether a sample of the actors must be taken.

The nature of the structural variables also determines which analytic methods are appropriate for their study. Thus, it is crucial to understand the nature of these variables. The data collection techniques described here determine, to some degree, the characteristics of the relations.

2.1.1 Structural and Composition Variables

There are two types of variables that can be included in a network data set: *structural* and *composition.* Structural variables are measured on pairs of actors (subsets of actors of size 2) and are the cornerstone of social network data sets. Structural variables measure ties of a specific kind between pairs of actors. For example, structural variables can measure business transactions between corporations, friendships between people, or trade between nations. Actors comprising these pairs usually belong to a single set of actors.

Composition variables are measurements of actor attributes. Composition variables, or *actor attribute* variables, are of the standard social and behavioral science variety, and are defined at the level of individual actors. For example, we might record gender, race, or ethnicity for people, or geographical location, after-tax profits, or number of employees for corporations. Some of the methods we discuss allow for simultaneous analyses of structural and composition variables.

2.1.2 Modes

We will use the term "mode" to refer to a distinct set of entities on which the structural variables are measured (Tucker 1963, 1964, 1966; Kroonenberg 1983; Arabie, Carroll, and DeSarbo 1987). Structural variables measured on a single set of actors (for example, friendships among residents of a neighborhood) give rise to one-mode networks. The most common type of network is a *one-mode* network, since all actors come from one set.

There are types of structural variables that are measured on two (or even more) sets of entities. For example, we might study actors from two different sets, one set consisting of corporations and a second set consisting of non-profit organizations. We could then measure the flows of financial support flows from corporations to non-profit actors. A network data set containing two sets of actors is referred to as a *two-mode* network, to reflect the fact that there are two sets of actors. A two-mode network data set contains measurements on which actors from one of the sets have ties to actors in the other set. Usually, not all actors can initiate ties. Actors in one of the sets are "senders," while those in the other are "receivers" (although the relation itself need not be directional). We will consider one-mode and two-mode, and even mention higher-mode, social networks in this book.

2.1.3 *Affiliation Variables*

A special type of two-mode network that arises in social network studies is an *affiliation* network. Affiliation networks are two-mode, but have only one set of actors. The second mode in an affiliation network is a set of *events* (such as clubs or voluntary organizations) to which the actors belong. Thus, in affiliation network data the two modes are the actors and the events. In such data, the events are defined not on pairs of actors, but on subsets of actors. These subsets can be of any size. A subset of actors affiliated with an affiliation variable is that collection of actors who participate in a specific event, belong to a given club, and so forth. Each affiliation variable is defined on a specific subset of actors.

For example, consider a set of actors, and three elite clubs in some city. We can define an affiliation variable for each of these three clubs. Each of these variables gives us a subset of actors — those actors belonging to one of the clubs.

The collections of individuals affiliated with the events can be found in a number of ways, depending on the substantive application. When events are clubs, boards of directors of corporations, or committees, the membership lists or rosters give the actors affiliated with each event. Often events are informal social occasions, such as parties or other gatherings, and observations or attendance or interactions among people provide the affiliations of the actors (Bernard, Killworth, and Sailer 1980, 1982; Freeman and Romney 1987). One of the earliest, and now classic, examples of an empirical application is the study of Davis, Gardner, and Gardner (1941) of the cohesive subgroups apparent in the social activities of women in a Southern city. Using newspaper records and interviews, they recorded the attendance of eighteen women at fourteen social events.

2.2 Boundary Specification and Sampling

A number of concerns arise in network studies that must be addressed prior to gathering any network data. Typically, a researcher must first identify the population to be studied, and if sampling is necessary, worry about how to sample actors and relations. These issues are considered here.

2.2.1 What Is Your Population?

A very important concern in a social network study is which actors to include. That is, who are the relevant actors? Which actors are in the population? In the case of small, closed sets of actors (such as all employees at a service station, faculty in an academic department, or corporations headquartered in a major metropolitan area), this issue is relatively easy to deal with. For other studies, the *boundary* of the set of actors may be difficult (if not impossible) to determine. The boundary of a set of actors allows a researcher to describe and identify the population under study.

Actors may come and go, may be many in number and hard to enumerate, or it may be difficult even to determine whether a specific actor belongs in a set of actors. For example, consider the study of elites in a community. The boundary of the set, including all, and only, the elites within the community, may be difficult, or impossible, to determine. However, frequently there will be a clear "external" definition of the boundary of the set which enables the researcher to determine which actors belong in it.

In some instances it is quite plausible to argue that a set of actors is relatively bounded, as for example, when there is a fairly complete membership roster. In such a case, the entire set of members can make up the actor set. However, there are other instances when drawing boundaries around a set is somewhat arbitrary. In practice, while network researchers recognize that the social world consists of many (perhaps infinite) links of connection, they also find that effective and reasonable limits can be placed on inclusion. Network researchers often define actor set boundaries based on the relative frequency of interaction, or intensity of ties among members as contrasted with non-members.

Laumann, Marsden, and Prensky (1989) describe two different approaches to boundary specification in social network studies. The first way, which they refer to as the *realist* approach, focuses on actor set boundaries and membership as perceived by the actors themselves. For example, a street-corner gang is acknowledged as a social entity by its members (it may even have a name — "Jets" or "Sharks") and the membership of the gang is the collection of people the members acknowledge as belonging to the gang. The second way of specifying network boundaries, which Laumann, Marsden, and Prensky refer to as the *nominalist* approach, is based on the theoretical concerns of the researcher. For example, a researcher might be interested in studying the flow of computer messages among researchers in a scientific specialty. In such a study, the list of actors might be the collection of people who published papers on the topic in the previous five years. This list is constructed for the analytical purposes of the researcher, even though the scientists themselves might not perceive the list of people as constituting a distinctive social entity. Both of these approaches to boundary specification have been used in social network studies.

Consider now two specific examples of how researchers have defined network boundaries. The first example illustrating the problem of identifying the relevant population of actors comes from a study of how information or new ideas diffuse through a community. Coleman, Katz, and Menzel (1957) studied how a new drug was adopted by physicians. Their solution to the problem of boundary identification is as follows:

> It was decided to include in the sample, as nearly as possible, *all* the local doctors in whose specialities the new drug was of major potential significance. This assured that the "others" named by each doctor in answer to the sociometric questions were included in the sample. (page 254)

The second example comes from the study of community leaders by Laumann and Pappi (1973). They asked community leaders to define the boundary by identifying the elite actors in the community of Altneustadt. These leaders were asked to

> ... name all persons [who] are now in general very influential in Altneustadt.

From these lists, each of which can be considered a sample of the relevant actors in the elite network, the actor set was enumerated.

Many naturally occurring groups of actors do not have well-defined boundaries. However, all methods must be applied to a specific set of data which assumes not only finite actor set size(s), but also enumerable set(s) of actors. Somehow, in order to study the network, we must enumerate a finite set of actors to study.

For our purposes, the set of actors consists of all social units on which we have measurements (either structural variables, or structural and compositional variables). Social network analysis begins with measurements on a set of actors. Researchers using methods described here must be able to make such an assumption. We assume, prior to any data gathering, that we can obtain relevant information on all substantively important actors; such actors will be included in the actor set. However, some actors may be left out unintentionally or for other reasons. Thus, the constitution of the actor set (that is, its size and composition) depends on both practical and theoretical concerns. The reason for the assumption that the actor set consists of all social units on which we have measurements is quite simple — the methods we discuss here cannot handle amorphous set boundaries. We will always start our analyses with a set (or sets) of actors, and we must be able to enumerate (or label) all members.

Many network studies focus on small collectivities, such as classrooms, offices, social clubs, villages, and even, occasionally, artificially created and manipulated laboratory groups. All of these examples have clearly defined actor set boundaries; however, recent network studies of actors such as elite business leaders in a community (Laumann and Pappi 1976), interorganizational networks in a community (Galaskiewicz 1979, 1985; Knoke 1983; Knoke and Wood 1981; Knoke and Kuklinski 1982), and interorganizational networks across an entire nation (Levine 1972) have less well-defined boundaries.

In several applications, when the boundary is unknown, special sampling techniques such as *snowball sampling* (Goodman 1949, 1961; Erickson 1978) and *random nets* (first proposed by Rapoport 1949a, 1949b, 1950, and especially 1963; recently resurrected by Fararo 1981, 1983, and Fararo and Skvoretz 1984) can be used to define actor set boundaries. Examples of social network studies using snowball sampling include: Johnson (1990) and Johnson, Boster, and Holbert (1989) on commercial fishermen; Moore (1979) and Alba and Moore (1978) on elite networks. Such sampling techniques are discussed in the next section.

2.2.2 Sampling

Sometimes, it may not be possible to take measurements on all the actors in the relevant actor set. In such situations, a sample of actors may be taken from the set, and inferences made about the "population" of actors from the sample. Typically, the sampling mechanism is known, and the sample is a good, probability sample (with known selection probabilities).

We will not assume in this book that the actors in the actor set(s) are samples from some population. Most network studies focus on well-defined, completely enumerated sets, rather than on samples of actors from larger populations. Methodology for the latter situation is considerably different from methods for the former. With a sample, one usually views the sample as representative of the larger, theoretically interesting population (which must have a well-defined boundary and hence, a known size), and uses the sampled actors and data to make inferences about the population. For example, in a study of major corporate actors in a national economy, a sample of corporations may be taken in order to keep the size of the problem manageable; that is, it might take too much time and/or too many resources actually to take a census of this quite large population.

There is a large literature on network sampling, both applied and theoretical. The primary focus of this literature is on the estimation of network properties, such as the average number of ties per actor (see Chapter 4), the degree of reciprocity present (see Chapter 13), the level of transitivity (see Chapters 6 and 14), the density of the relation under study (see Chapter 5), or the frequencies of ties between subgroups of actors (see Chapter 7) based on the sampled units.

Frank (1977a, 1977b, 1977c, 1978b, 1979a, 1979b, 1980, 1985) is the most widely known and most important researcher of sampling for social networks. His classic work (Frank 1971) and more recent review papers (Frank 1981, 1988) present the basic solutions to the problems that arise when the entire actor set is not sampled. Erickson and Nosanchuk (1983) review the problems that can arise with network sampling based on a large-scale application of the standard procedures to a network of over 700 actors. Various other sampling models are discussed by Hayashi (1958), Goodman (1961), Bloemena (1964), Proctor (1967, 1969, 1979), Capobianco (1970), Sheardon (1970), and Cabobianco and Frank (1982).

One very clever network sampling idea originated with Goodman (1961). A snowball network sample begins when the actors in a set of sampled respondents report on the actors to whom they have ties of a specific kind. All of these nominated actors constitute the "first-order" zone of the network. The researcher then will sample all the actors in this zone, and gather all the additional actors (those nominated by the actors in the first-order zone who are not among the original respondents or those in this zone). These additional actors constitute the "second-order" zone. This snowballing proceeds through several zones. Erickson (1978) and Frank (1979b) review snowball sampling, with the goal of understanding how other "chain methods" (methods designed to trace ties through a network from a source to an end; see, for example, Granovetter 1974, and Useem 1973, for applications) can be used in practice. Chain methods include snowball sampling and the small world technique discussed below. Erickson also discusses at length the differences between standard network sampling and chain methods.

In some network sampling situations, it is not clear what the relevant sampling unit should be. Should one sample actors, pairs of actors, triples of actors, or perhaps even subsets of actors? Granovetter (1977a, 1977b) and Morgan and Rytina (1977) have sensitized the network community to these issues (see also Erickson, Nosanchuk, and Lee 1981, and Erickson and Nosanchuk 1983). In other situations, one might sample actors, and have them report on their ties and the ties that might exist among the actors they choose or nominate. Such samples give rise to "ego-centered" networks (defined later in this chapter). With a sample of ego-centered networks, one usually wants to make inferences about the entire population of such networks (see for example, the epidemiological networks discussed by Klovdahl 1985; Laumann, Gagnon, Michaels, Michael, and Coleman 1989; and Morris 1989, 1990). Statistically, sampling dyads or ego-centered networks leads to sampling designs which are not simple; the sampling is actually clustered, and one must adjust the standard statistical summaries to allow for possible biases (Reitz and Dow 1989).

2.3 Types of Networks

There are many different types of social networks that can be studied. We will categorize networks by the nature of the sets of actors and the properties of the ties among them. As mentioned earlier in this chapter, we define the *mode* of a network as the number of sets of entities on which structural variables are measured. One-mode networks, the predominate type of network, study just a single set of actors, while two-mode networks focus on two sets of actors, or one set of actors and one set of events. One could even consider three- (and higher) mode networks, but rarely have social network methods been designed for such complicated data structures. Our discussion in this section is organized by the number of modes in the network. We will first discuss one-mode networks (with a single set of actors), then discuss two-mode networks, first with two sets of actors and then with one set of actors and one set of events. Applications of these three types of networks are the focus for methods presented in this book.

The number of modes in a network refers to the number of distinct kinds of social entities in the network. This usage is slightly different from the use of the term "mode" in the psychometric literature (Tucker 1964;

Carroll and Arabie 1980). In that literature, mode refers to a "particular class of entities" (Carroll and Arabie 1980, page 610). Thus, a study in which subjects respond to a set of stimuli (such as questionnaire items) gives rise to two modes: the subjects and the stimulus items. In the standard sociometric data design, a number of actors are presented with a list of the names of other people in the actor set, and asked to rate each other person in terms of how much they "like" that person. In a non-network context one could view these data as two-mode: the people as respondents are the first mode, and the names of the people as stimulus (questionnaire) items are the second mode. However, as a social network, these data contain only a single set of actors, and thus, in our terminology, it is a one-mode network in which the relation of friendship is measured on a single set of people. One might very well be interested in studying the set of respondents making evaluations of the other people, in addition to studying the people as the "stimuli" that are being evaluated. In that case one would consider respondents and stimuli as two different modes (Feger and Bien 1982; Noma 1982b; Kumbasar, Romney, and Batchelder n.d.).

We first categorize networks by how many modes the network has (one or two), and by whether affiliational variables are measured. There are, however, other kinds of relational data that are not one of these types. One example is data arising from an ego-centered network design. Data on such networks are gathered using special sampling strategies that allow the researcher to focus on a specific set of respondents, and the ties that these respondents have to particular others. We briefly describe special ego-centered networks and special dyadic designs at the end of this section.

We turn now to a discussion of one-mode, two-mode, and then affiliational, and egocentric and special networks.

2.3.1 One-Mode Networks

Suppose the network under study is one-mode, and thus involves measurements on just a single set of actors. Consider first the nature of the actors involved in such networks.

Actors. The actors themselves can be of a variety of types. Specifically, the actors may be

- People

- Subgroups
- Organizations
- Collectives/Aggregates:
  - Communities
  - Nation-states

Note that subgroups usually consist of people, organizations usually consist of subgroups of people, while communities and nation-states are larger entities, containing many organizations and subgroups. Thus, there is a natural progression of types of actors from sets of people, to collections or aggregates. Throughout this book, we will illustrate methodology with examples consisting of social network data on different types of actors.

Relations. The relations measured on the single set of actors in a one-mode network are usually viewed as representing specific substantive connections, or "relational contents" (Knoke and Kuklinski 1982). These connections, measured at the level of pairs of actors, can be of many types. Barnes (1972) distinguishes, quite generally, between attitudes, roles, and transactions. Knoke and Kuklinski (1982) give a more extensive list of general kinds of relations. Specifically, the kinds of relations that we might study include:

- Individual evaluations: friendship, liking, respect, and so forth
- Transactions or transfer of material resources: lending or borrowing; buying or selling
- Transfer of non-material resources: communications, sending/receiving information
- Interactions
- Movement: physical (migration from place-to-place), social (movement between occupations or statuses)
- Formal roles
- Kinship: marriage, descent

One or more of these types of relations might be measured for a single set of actors.

Individual evaluations are usually measurements of positive or negative affect of one person for another. Sometimes, these relations are labeled *sentiment*, and classically were the focus of the early sociometricians (see Moreno 1934; Davis 1970; Davis and Leinhardt 1972). Without question, such relations historically have been the most studied.

Transactions, or transfers of material resources, include business transactions, imports and exports of goods, specific forms of social support, such as lending and borrowing, contacts made by one actor of another in order to secure valuable resources, and transfer of goods. Such relations include exchange of gifts, borrowing or lending items, and sales or purchases (Galaskiewicz and Marsden 1978; Galaskiewicz 1979; Laumann, Galaskiewicz, and Marsden 1978). Social support ties are also examples of transactions (Wellman 1992b).

Transfers of non-material resources are frequently communications between actors, where ties represent messages transmitted or information received. These ties involve sending or receiving messages, giving or receiving advice, passing on gossip, and providing novel information (Lin 1975; Rogers and Kincaid 1981; Granovetter 1974). Information about innovations is frequently diffused over such communication channels (Coleman, Katz, and Menzel 1966; Rogers 1979; Michaelson 1990).

Interactions involve the physical interaction of actors or their presence in the same place at the same time. Examples of interactions include: sitting next to each other, attending the same party, visting a person's home, hitting, hugging, disciplining, conversing, and so on.

Movement can also be studied using network data and processes. Individuals moving between communities can be counted, as well as workers changing jobs or people changing statuses (see, for example, Breiger 1981c).

Formal roles, such as those dictated by power and authority, are also relational. Ties can represent authority of one actor over others, especially in a management setting (White 1961). Example of formal roles include boss/employee, teacher/student, doctor/patient, and so on.

Lastly, kinship relations have been studied using network methods for many years. Ties can be based on marriage or descent relationships and marriage or family relationships can be described using social network methods (for example, see White 1963; Boyd 1969).

Actor Attributes. In addition to relational information, social network data sets can contain measurements on the characteristics of the actors. Such measurements of actor attribute variables constitute the composition of the social network.

These variables have the same nature as those measured in non-network studies. People can be queried about their age, gender, race, socioeconomic status, place of residence, grade in school, and so on. For corporate actors, one can measure their profitability, revenues, geographical location, purpose of business, and so on. The "size, shape, and flavor" of the actors constituting the network can be measured in many ways.

2.3.2 *Two-Mode Networks*

Suppose now that the network under study is two-mode, and thus involves measurements on two sets of actors, or on a set of actors and a set of events. We will first consider the case in which relations are measured on pairs of actors from two different actor sets. We will then discuss a special kind of two-mode network in which measurements are taken on subsets of actors.

Two Sets of Actors. Relations in a two-mode network measure ties between the actors in one set and actors in a second set. We call such networks *dyadic* two-mode networks, since these relations are functions of dyads in which the first actor and the second actor in the dyad are from different sets.

With respect to the different types of actors, the types of relations, and the types of actor attribute variables, all of our discussion about one-mode networks is relevant. Note, however, that there can be multiple types of actors, and we can have a unique collection of attribute variables for each set of actors.

Actors. In a two-mode network that contains two sets of actors, these actors can be of the general types as described for one-mode networks. However, the two sets of actors may be of different types.

Relations. In a two-mode network with two sets of actors, at least one relation is measured between actors in the two sets. In a more extensive two-mode network data set, relations can also be defined on actors within a set. However, for the network to be truly two-mode with two sets of actors, at least one relation must be defined between the two sets of actors.

An example of such a network can be found in Galaskiewicz and Wasserman (1989). The data analyzed there consisted of two sets of actors: a collection of corporations headquartered in the Minneapolis/St. Paul metropolitan area, and the non-profit organizations (such as the Red Cross, United Way, public radio and television stations) which rely on contributions from the public sector for their operating budgets. The primary relation was the flow of donations from the corporations to the non-profit organizations, clearly a two-mode relation. Also, it is important to note that this relation is *unidirectional* since it flows from actors in one set to actors in the other set, but not the reverse. In addition, the analysis by Galaskiewicz and Wasserman considered a number of relations defined just for the corporations (such as shared country club memberships among the chief executive officers) and several just for the non-profits (such as interlocking boards of directors). A part of this data set will be discussed in more detail later in this chapter.

One Set of Actors and One Set of Events. The next type of two-mode social network, which we refer to as an *affiliation network*, arises when one set of actors is measured with respect to attendance at, or affiliation with, a set of events or activities. The first mode in an affiliation network is a set of actors, and the second is a set of events which affiliates the actors.

An example comes from Davis, Gardner, and Gardner (1941), as described and analyzed by Homans (1950) and Breiger (1974). A set of women attended a variety of social functions, and this attendance was recorded over a period of several months. Each social function can be viewed as a variable, and a binary measurement made as to whether a specific actor attended the specific function. These variables are termed *affiliational*. Such data and networks are called *affiliation* networks, or sometimes, *membership* networks. And since the affiliations are measured on subsets of actors, such networks are non-dyadic, two-mode networks.

Actors. In an affiliation network, we have a first set of actors, and a second set of events or activities to which the actors in the first set attend or belong. The types of actors in affiliation networks can be exactly the same as those in one-mode and two-mode networks. The only requirement is that the actors must be affiliated with one or more events.

Events. In affiliation networks, actors (the first mode) are related to each other through their joint affiliation with events (the second mode). The events are often defined on the basis of membership in clubs or voluntary organizations (McPherson 1982), attendance at social events (Davis, Gardner, and Gardner 1941), sitting on a board of directors, or socializing in a small group (Bernard, Killworth, and Sailer 1980, 1982; Wilson 1982).

The nature of the events, which affiliate the actors, depends on the type of actors involved. People may attend social functions or belong to athletic clubs, subgroups of people may attend various committee meetings (for example, departments at a major university send representatives to college committee meetings), organizations may be represented on various boards of directors in a community, or countries might belong to treaty organizations, and so on.

Attributes. We can have actor attribute variables that are of the same types as those for one-mode and two-mode networks. In addition, the events themselves may have characteristics associated with them which can be measured and included in the network data set. For example, clubs will be of a particular size or located in a specific geographical area. Events usually occur at discrete points in time, as well as in particular geographical places. Thus, there can be two sets of attribute variables in an affiliation network data set: attributes of the actors, and attributes of the events.

Methods for analyzing affiliation network data are described in Chapter 8, and are applied to a network data set giving the memberships of a set of chief executive officers of major corporations in Minneapolis/St. Paul in a set of exclusive clubs.

2.3.3 Ego-centered and Special Dyadic Networks

Not all structural data give rise to standard social network data sets. With standard network data (regardless of how many modes the network has), one enumerates not only the actors, but the relevant pairs as well. All actors (theoretically) can relate to each other in one-mode networks. In two-mode networks with two sets of actors, all actors in the first mode can (theoretically) relate to all in the second. However, some data collection designs gather structural information on some pairs but not others. An example of such data arises in studies of couples. Each partner in the couple can interact with the other but with no other person during counseling sessions. Interactions during these sessions are then recorded. When interest centers on a collection of pairs (husband-wife, father-son, and so forth), one frequently samples from a large population of such pairs. We will refer to these non-network relational data as *special dyadic* designs.

An actor may also relate to a limited number of "special" other actors. For example, one might observe mothers interacting with their own children in an experimental situation. In this case, mothers only interact with their own children, and children only interact with their own mother. Thus, the partners for one person (either mother or child) are different from the partners for another. In this situation, the design of the experiment constrains the interactions among the set of people so that all people cannot, theoretically, interact with all others.

Another related design is an *ego-centered* network. An ego-centered network consists of a focal actor, termed *ego*, as set of alters who have ties to ego, and measurements on the ties among these alters. For example, when studying people, one samples respondents, and each respondent reports on a set of alters to whom they are tied, and on the ties among these alters. Such data are often referred to as *personal network* data. Clearly these data are relational, but limited, since ties from each actor are measured only to some (usually only a few) alters. For example, in 1985 the General Social Survey conducted by the National Opinion Research Center (see Burt 1984, 1985) asked respondents:

> Looking back over the last six months — who are the people with whom you discussed matters important to you? (1984, page 119)

Respondents also reported on the ties between the people they listed. Bernard, Johnsen, Killworth, McCarty, Shelley, and Robinson (1990), Killworth, Johnsen, Bernard, Shelley, and McCarty (1990), Huang and Tausig (1990), Burt (1984, 1985), Marsden (1987, 1990b), Wellman (1993), as well as Campbell, Marsden, and Hurlbert (1986) discuss measurement of such personal, ego-centered networks.

Ego-centered networks have been widely used by anthropologists to study the social environment surrounding individuals (Boissevain 1973) or families (Bott 1957). Ego-centered networks are also used quite often in the study of social support. The term "social support" has been used to refer to social relationships that aid the health or well-being of an individual. The emphasis on relationships has allowed researchers to study support using social networks. Such networks are of great interest in clinical and community psychology, as well as in sociology. A variety of hypotheses (see Hammer 1983; Cohen and Syme 1985) have been offered to explain how personal relationships, as reflected by such ego-centered networks, can affect the emotional and physical well-being of an individual.

The methods described in this book assume that there are no theoretical limitations on interactions among actors. A social network arises when all actors can, theoretically, have ties to all relevant actors. The primary object of study for methods discussed in this book is this complete collection of actors (one or more sets) and the ties among them.

2.4 Network Data, Measurement and Collection

We now turn to issues concerning the measurement and collection of network data, the accuracy, validity, and error associated with these data, and particular design considerations that can arise in network studies.

2.4.1 Measurement

Social network data differ from standard social and behavioral science data in a number of important ways. Most importantly, social network data consist of one (or more) relations measured among a set of actors. The presence of relations has implications for a number of measurement issues, including the unit of observation (actor, pair of actors, relational tie, or event), the modeling unit (the actor, dyad, triad, subset of actors, or network), and the quantification of the relations (directional vs. nondirectional; dichotomous vs. valued). We will discuss each of these issues in turn.

Social network data can be studied at a number of different levels: the individual actor, the pair of actors or dyad, the triple of actors or triad, a subset of actors, or the network as a whole. We will refer to the level at which network data are studied as the *modeling unit*. However, social network data often are gathered at a level that is different from the level at which they are modeled. We discuss the unit of observation and the modeling unit in the next two sections.

Unit of Observation. The unit of observation is the entity on which measurements are taken. Most often social network data are collected by observing, interviewing, or questioning individual actors about the ties from these actors to other actors in the set. Thus, the unit of observation is an actor, from whom we elicit information about ties. The dyad is the unit of observation when one measures ties among pairs of actors directly. For example, one could record instances of aggression among pairs of children on a playground. When affiliation network data are collected, the unit of observation is often the event. The researcher selects events or social occasions, and for each event, records the actors who are affiliated with it.

Modeling Unit. Just as social network data can be observed at a number of levels, there are several levels at which network data can be modeled or summarized. These levels are the:

- Actor
- Dyad
- Triad
- Subgroup
- Set of actors or network

In categorizing network methods, it is useful to consider the level to which a model or network property applies. Some network properties pertain to actors (for example the number of "choices" that an individual actor receives from others in the network). Other properties pertain to pairs of actors (for example, if one person "chooses" another as a friend, is the "choice" returned by the second person?). Models at the level of the triad consider triples of actors and the ties among them. Many methods pertain to subgroups of actors; for example, one could study whether there are subsets of actors in the network who interact frequently with each other. Finally many properties pertain to the network as a whole, for example, the proportion of ties that are present in the network.

Relational Quantification. There are two properties of relations that are important for understanding their measurement, and for categorizing the methods described here: whether the relation is *directional* or *nondirectional*, and whether it is *dichotomous* or *valued*. In a directional relation, the relational tie between a pair of actors has an origin and a destination; that is, the tie is directed from one actor in the pair to the other actor in a pair. For example, one country exports manufactured goods to a second country; the first country is the source of the manufactured goods, and the second country is the destination. In a nondirectional relation the tie between a pair of actors does not have a direction. For example, we could define a tie as present between two countries if they share a border.

A second important property of a relation is whether it is dichotomous or valued. Dichotomous relations are coded as either present or absent, for each pair of actors. For example one could record whether one country sends an ambassador to a second country; thus giving rise to a dichotomous relation that can only take on two values: "send" or "not send." On the other hand, valued relations can take on a range of values, indicating the strength, intensity, or frequency of the tie between each pair of actors. For example, we could record the dollar value of manufactured goods that are exported from one country to a second country, thus giving rise to a valued relation.

2.4.2 Collection

There are a variety of ways in which social network data can be gathered. These techniques are:

- Questionnaires
- Interviews
- Observations
- Archival records
- Experiments
- Other techniques, including ego-centered, small world, and diaries

Each of these techniques will be discussed and illustrated with examples.

Questionnaire. This data collection method is the most commonly used (especially when actors are people). The questionnaire usually contains questions about the respondent's ties to the other actors. Questionnaires are most useful when the actors are people, and the relation(s) that are being studied are ones that the respondent can report on. For example, people can report on who they like, respect, or go to for advice. Questionnaires can also be used when the actor in a study is a collective entity, such as a corporation, but an individual person representing the collective reports on the collective's ties. For example, Galaskiewicz (1985) asked officers in charge of corporate giving whether or not the corporation had made a donation to a non- profit agency.

There are three different question formats that can be used in a questionnaire:

- Roster vs. free recall
- Free vs. fixed choice
- Ratings vs. complete rankings

In the following sections we will discuss each of these formats and describe examples of their use.

Roster vs. Free Recall. One issue in the design of a questionnaire to gather network data is whether each actor should be presented with a complete list, or *roster*, of the other actors in the actor set. Rosters can be constructed only when the researcher knows the members in the set prior to data gathering. For example, Krackhardt and Stern (1988) collected information on friendships among members of a university class as part of their study of "simulated" corporations. They had each person rate their friendship with every member of the class on a five point scale:

> Everyone in the class completed a questionnaire which asked them to rate every other person in the class as to how close a friend he or she was. The directions for this questionnaire included the following: "Please place a check in the space that best describes your relationship with each person on the list." The names of everyone participating in the game were listed below, with five categories from which the respondent could choose: "trust as a friend", "know well", "acquaintance", "associate name with face", and "do not know". (page 131)

For some network designs, the researcher does not present a complete list of the actors in the network to the respondent on the questionnaire. In such instances, it is common simply to ask respondents to "name those people with whom you (*fill in specific tie*)". Such a format, where respondents generate the list of names, is called *free recall*. For example, Rapoport and Horvath (1961) studied friendships in two junior high schools. Students were asked to list their best friends, but were not presented with a roster. Specifically,

> Each pupil in both schools was asked to write his name, age, grade, and home room number on a card and to fill in the blanks in the statements:
>
> - "My best friend in (name of school) Junior High School is ..."
> - "My second best friend is ..."
> - ...
> - "My eighth best friend is. ..." (page 281)

Note here how the network membership is known beforehand (all students in a school are the set of actors) but students listed their friends using free recall.

In some settings, the researcher might not even have a list; that is, the actors within the actor set might not even be known in advance. In this situation, sampling or enumeration techniques are necessary (as we have discussed earlier in this chapter). For example, in studies of community elites (Friedkin 1984; Moore 1979; Alba and Moore 1978), selected actors are asked to name other actors they believe to be influential in the community.

Free vs. Fixed Choice. If actors are told how many other actors to nominate on a questionnaire (for example, to name a specific number of "best friends"), then each person has a fixed number of "choices" to make. Such designs are termed *fixed choice.* In a fixed choice design each actor has a fixed maximum number of ties to the other actors in the set of actors. For example, Coleman, Katz, and Menzel (1957), in a study of diffusion of a medical innovation among physicians, interviewed all physicians in a community. Specifically,

> Each doctor interviewed was asked three sociometric questions:
>
> (i) "To whom did he most often turn for advice and information?"
> (ii) "With whom did he most often discuss his cases in the course of an ordinary week?"
> (iii) "Who were the friends, among his colleagues, whom he saw most often socially?"
>
> In response to each of these questions, the names of three doctors were requested. (page 254)

In this study, each person was constrained to have no more than three ties for each of the three relations.

On the other hand, if actors are not given any such constraints on how many nominations to make, the data are *free choice.* For example, Carley and Wendt (1988) studied the ties among people in an "invisible college" of users of a computer program at a variety of universities.

> Each individual was asked to denote for each member of the user group whether or not they:
>
> - Had an office next to each other
> - Attended the same school at the same time
> - Shared an office
> - Lived in the same living group or apartment
> - Were at the same school at the same time
> - Were in the same academic department at the same time

Note that there is no constraint on the number of people that an individual respondent can choose on these six relations.

The study of a university class by Krackhardt and Stern (1988) was a free choice design, since respondents were not limited in the number of friends they could choose. The Rapoport and Horvath design allowed each student to make eight choices; however, as Rapoport and Horvath note, students did not always fill in all of the 8 choices. Similarly, in a study of 384 sociograms that were collected using a fixed choice procedure, Holland and Leinhardt (1973) found that in fewer than 20 percent of the data sets did all respondents conform to the fixed number of choices.

Later in this chapter, we discuss limitations of social network data collected using fixed choice designs.

Ratings vs. Complete Ranking. In some network designs, actors are asked to rate or rank order all the other actors in the set for each measured relation. Such measurements reflect the intensity of strength of ties. Ratings require each respondent to assign a value or rating to each tie. Complete rankings require each respondent to rank their ties to all other actors.

An example of a complete rank order design is the study by Bernard, Killworth, and Sailer (1980). They asked each of forty members of a social science research office to report the amount of communication with each other member of the office using the following procedure:

> ... each participant was given the familiar deck of cards containing the names of the other participants. They arranged (that is, ranked) the cards from most to least on how often they talked to others in the office during a normal working day. (page 194)

Such data are *complete rankings* or *complete rank orders.* This questionnaire design is quite different from that employing *ratings* of the ties.

Alternatively, one can gather ratings from each actor about their ties to other members on every relation. These ratings can be dichotomous, as in the Carley and Wendt (1988) study (ties are either present or absent), or valued, as in the Krackhardt and Stern (1988) study where ratings were made by choosing one of five possible categories for the strength of each tie.

Full rank-orders and rating scales with multiple response categories produce *valued* relations. Response formats where respondents either nominate a person or not on a given relation produce *dichotomous* relations. In either case, when "choices" are directed from respondents to the people they name, the resulting relations are *directional.*

Interview. Interviews, either face-to-face or over the telephone, are occasionally used to gather network data in instances where questionnaires are not feasible. For example, Galaskiewicz (1985) interviewed the chief executive officers of the largest corporations in the Minneapolis/St. Paul metropolitan area. Chief executive officers were much more willing to participate in face-to-face interviews than via an impersonal questionnaire.

Interviews have been used to gather data from respondents in ego-centered networks, such as the 1985 NORC General Social Survey (Burt 1984, 1985), Wellman's study of social support in East York, Ontario (Wellman 1979; Wellman, Carrington, and Hall 1988; Wellman and Wortley 1990, and references therein), and Fischer's study of friendships in a community in Northern California (Fischer 1982).

Observation. Observing interactions among actors is another way to collect network data. This method has been widely used in field research to study relatively small groups of people who have face-to-face interactions (Roethlisberger and Dickson 1961; Kapferer 1969; Hammer, Polgar, and Salzinger 1969; Thurman 1980; Bernard and Killworth 1977; Killworth and Bernard 1976; Bernard, Killworth, and Sailer 1980, 1982; Freeman and Romney 1987; Freeman, Romney, and Freeman 1987; Freeman, Freeman, and Michaelson 1988, 1989). For example, Freeman, Freeman, and Michaelson (1988, 1989) observed a collection of fifty-four windsurfers on a beach in Southern California.

> Observations on the subjects' interaction patterns were made for two half-hour periods on each day of 31 consecutive days. (Freeman, Freeman, and Michaelson 1989, page 234)

The information recorded was the number of minutes of interaction between pairs of people.

Observational methods have been used extensively in the studies of Bernard, Killworth, and Sailer (Bernard and Killworth 1977; Killworth and Bernard 1976; Bernard, Killworth, and Sailer 1980, 1982). These researchers systematically observed interactions among people in a variety of social settings, such as a social science research office, faculty, staff, and graduate students in a university department, and members of a college fraternity. Their research focused on the relationship between these observed interactions and actors' recollections of their own interactions. Since data are collected by observing interactions, without requiring verbal responses from the people, this method is quite useful with people who are not able to respond to questionnaires or interviews.

Observational methods are widely used in the study of interactions among non-human primates (Dunbar and Dunbar 1975; Sade 1965). For instance, Wolfe (see MacEvoy and Freeman n.d.) observed a colony of monkeys, and recorded which monkeys visited a river together. Sailer and Gaulin (1984) present data collected on interactions among members of a colony of mantled howler monkeys.

Observational methods are also useful for collecting affiliation network data. The researcher can record who attends each of a number of social events. For example, Freeman, Romney, and Freeman (1987) recorded which faculty members and graduate students attended a weekly departmental colloquium over the course of a semester. Each colloquium is an event in this affiliation network.

In some studies, the researcher observes a set of actors for an extended period of time, and then summarizes his or her impressions of the ties among all pairs of actors in the set (Roethlisberger and Dickson 1961; Kapferer 1969; Thurman 1980). The ties are based on the researcher's impressions.

Archival Records. Some network researchers measure ties by examining measurements taken from records of interactions. Such records can take many forms, such as measurements on past political interactions among nations, previously published citations of one scholar by another, and so on. Burt and Lin (1977) discuss how social networks can be obtained from archival data, such as journal articles, newspapers, court records, minutes of executive meetings, and the like. Frequently, as noted by Burt and Lin, such data give rise to longitudinal relations and can be used to reconstruct ties that existed in the past. For example, Burt (1975, 1983) obtained information on interactions among corporate actors from the front pages of previously published issues of *The New York Times.*

Rosenthal, Fingrutd, Ethier, Karant, and McDonald (1985) used biographical records to study the organizational affiliations of women reformers in the 19th century in New York. These researchers were interested in the overlaps among the organizations. The list of women and their affiliations was compiled from biographical dictionaries which included information about organizational affiliations of 202 women, and 1015 organizations. These data are thus affiliation data compiled from archival sources.

Galaskiewicz (1985) obtained information on memberships of the chief executive officers of corporations in Minneapolis/St. Paul in elite country clubs by examining the membership rosters of the clubs. Other researchers have conducted similar elite studies by looking at volumes such as *Who's Who*, and social registers.

Another common use of archival records is for the study of sociology of science, specifically, patterns of citations among scholars. One can examine "who cites whom" in order to understand diffusion of a scientific innovation (Burt, 1978/1979a; Breiger 1976; McCann 1978; Noma 1982a, 1982b; Doreian and Fararo 1985; White and McCann 1988; Michaelson 1991; Carley and Hummon 1993). In these studies, the unit of observation is a citation, but since a given article usually contains many citations, the actor can be the article containing the citation, or the journal containing the article, or even the authors of the cited articles.

All of the data collection methods discussed above attempt to measure the ties among all the actors in the set. Many network studies employ a variety of data collection methods for recording ties, in addition to gathering actor attribute information. These data collection methods (questionnaires, observations, interviews, experiments, and so forth) are common social and behavioral science procedures.

Other. Here, we focus on other designs for collecting relational data. These include the cognitive social structure design (which is an extension of sociometric data to include actor perceptions of the network), experimental studies (in which network data are collected under controlled situations), and studies in which information is collected on ties among just some actors. Often these studies are used to estimate the size (de Sola Pool and Kochen 1978; Freeman and Thompson 1989; Bernard, Johnsen, Killworth, and Robinson 1989; Wellman 1992b) or composition (Verbrugge 1977; Wellman 1979; Marsden 1988; Wellman and Wortley 1990, and references therein) of an individual's ego-centered network. Perhaps only a few actors are chosen as respondents. Or, the actors might not even be members of a well-defined set of actors. Clearly in these instances, we are not studying a network with a boundary. We refer to such studies as special network designs.

In the next paragraphs, we discuss data collection procedures for cognitive social structure designs, experimental, ego-centered networks, and small- and reverse small-world techniques.

Cognitive Social Structure. In a standard sociometric questionnaire, one asks respondents about their own ties. A variation of this design is to ask respondents to give information on their perceptions of other actors' network ties. Such designs are called *cognitive social structures* because they measure perceived relations (Krackhardt 1987a; Kumbasar, Romney, and Batchelder n.d.).

As an example, Krackhardt and Porter (1985) studied turnover in several fast food restaurants. They were interested in the employees' perceptions of friendships among all other employees in the restaurant. Thus, they had to gather information from each person not only about their own friendships, but also about their perceptions of the friendships among all other pairs of employees. They collected network data at two points in time.

Their procedure is described as follows:

> In the first questionnaire, each person in the work group was asked to record who they perceived to be a friend of whom. While simple on the surface, this substantial task required that employees consider all possible pairs of friends in the restaurant. To accomplish this, the respondent was told to check the names of all those listed whom he or she thought would be considered a friend by employee # 1 (for example, "Henry"). Then, the same list was repeated on the next page, and the respondent was asked to check all names of those whom he or she thought would be considered a friend of employee # 2 ("Rita"). This process was repeated a total of N times (for N employees). In this way, we could assess each person's perception of everyone's friends, their own as well as their coworkers. (page 250)

Alternatively, one can ask respondents to report subgroups of people who form relatively tightly knit subgroups within the larger collection of people (Freeman, Freeman, and Michaelson 1988, 1989).

Data collected using a cognitive social structure design gives considerably more information than the usual sociometric design, since actors report not only on their own ties, but also on their perceptions of ties among all pairs of actors.

Experimental. Social network data can be collected using experimental designs. There are (at least) two basic ways to conduct such experiments. First, one can choose a set of actors and observe their interactions in an experimentally controlled situation. The researcher then records interactions or communications between pairs of actors. Ties may be observed between all pairs of actors. Second, one can not only choose actors but also specify which pairs of actors are permitted to communicate with each other during the experiment. One only records the frequency or content of communications between those pairs of actors who are permitted to interact.

Group problem-solving experiments (Bavelas 1950; Leavitt 1949, 1951) in which actors are assigned to positions within the network defined by the experimenter and allowed to communicate only with specific others are an example of the second type of experiment. The experimenter manipulates both group members and their ties. Power and exchange experiments are also of the second type (Cook, Emerson, Gilmore, and Yamagishi 1983; Bonacich 1987; Markovsky, Willer, and Patton 1988; and Friedkin and Cook 1990). The experimenter assigns actors to positions, and allows certain pairs of actors to negotiate the exchange of resources.

Ego-centered. An ego-centered, or *local,* network consists of a focal person or respondent (ego), a set of alters who have ties to ego, and measurements on the ties from ego to alters and on the ties between alters. One begins by asking a collection of respondents about their ties to other people to elicit the set of alters. In 1985 the NORC General Social Survey (see Burt 1984, 1985) asked a sample of 1531 people

> From time to time, most people discuss important matters with other people. Looking back over the past six months, who are the people with whom you discussed matters important to you? (page 119)

One also asks respondents information about the ties among the people that the respondent has named. The 1985 General Social Survey contained a question about the ties among all pairs of people named by the respondent. If we label two of the people named by a particular respondent "Alter 1" and "Alter 2," then the question can be worded

> Please think about the relations between the people you just mentioned. Some of them may be total strangers, in the sense that they would not recognize each other if they bumped into each other on the street. Others might be especially close, as close to each other as they are to you. First think about [Alter 1] and [Alter 2]. Are these people total strangers? (Burt 1985, page 120)

Such measurements give rise to ego-centered networks.

Small World. Special network designs are also used in small world and reverse small world studies. A small world study is an attempt to determine how many actors a respondent is removed from a target individual based on acquaintanceship. Of primary interest is not only how long these "chains" are, but also the characteristics of the intermediate actors in the chain. This data collection design was pioneered by Milgram (Milgram 1967; Travers and Milgram 1969). Korte and Milgram (1970) describe the typical small world study as follows:

> The small world method consists of presenting each of the persons in a "starting population" with the description of a given "target person" — his name, address, occupation, and other selected information. The task of a starter is to advance a booklet toward the target person by sending