Models and Methods in Social Network Analysis

Edited by
Peter J. Carrington
John Scott
Stanley Wasserman

Models and Methods in Social Network Analysis

Edited by

PETER J. CARRINGTON
University of Waterloo

JOHN SCOTT
University of Essex

STANLEY WASSERMAN
Indiana University

CAMBRIDGE
UNIVERSITY PRESS

Acknowledgments

Contributors

Vladimir Batagelj is a professor of discrete and computational mathematics at the University of Ljubljana and is chair of the Department of Theoretical Computer Science at IMFM, Ljubljana. He is a member of the editorial boards of *Informatica* and the *Journal of Social Structure*. He was visiting professor at the University of Pittsburgh in 1990/1991 and at University of Konstanz (Germany) in 2002. His main research interests are in graph theory, algorithms on graphs and networks, combinatorial optimization, data analysis, and applications of information technology in education. He is co-author (with Andrej Mrvar) of Pajek, a program for analysis and visualization of large networks.

Steve P. Borgatti is Associate Professor of Organization Studies at Boston College. His research interests include social networks, cultural domain analysis, and organizational learning. He is co-author of the UCINET software package and a past President of INSNA, the professional association for social network researchers.

Peter J. Carrington is Professor of Sociology at the University of Waterloo and Editor of the *Canadian Journal of Criminology and Criminal Justice*. His main teaching and research interests are in the criminal and juvenile justice systems, social networks, and research methods and statistics. He has published articles in the *Canadian Journal of Criminology and Criminal Justice, American Journal of Psychiatry, Journal of Mathematical Sociology*, and *Social Networks*. He is currently doing research on police discretion, criminal and delinquent careers and networks, and the impact of the Youth Criminal Justice Act on the youth justice system in Canada.

Patrick Doreian is a professor of sociology and statistics at the University of Pittsburgh, where he also chairs the Department of Sociology. He edits the *Journal of Mathematical Sociology* and is a member of the editorial board of *Social Networks*. His research and teaching interests include social networks, social movements, and mathematical sociology.

Marijtje A. J. van Duijn is an assistant professor in the Department of Sociology of the University of Groningen. Her research interests are in applied statistics and statistical methods for discrete and/or longitudinal data, including multilevel modeling and social network analysis. She teaches courses on multivariate statistical methods and on item response theory.

Martin Everett has a masters degree in mathematics and a doctorate in social networks from Oxford University. He has been active in social network research for more than

25 years. During a sabbatical at the University of California, Irvine, in 1987, he teamed up with Stephen Borgatti, and they have collaborated ever since. Currently he is a Provost at the University of Westminster, London.

Katherine Faust is Associate Professor in the Sociology Department at the University of California, Irvine, and is affiliated with the Institute for Mathematical Behavioral Sciences at UCI. She is co-author (with Stanley Wasserman) of *Social Network Analysis: Methods and Applications* and numerous articles on social network analysis. Her current research focuses on methods for comparing global structural properties among diverse social networks; the relationship between social networks and demographic processes; and spatial aspects of social networks.

Anuška Ferligoj is a professor of statistics at the University of Ljubljana and is dean of the Faculty of Social Sciences. She has been editor of the series *Metodoloski zvezki* since 1987 and is a member of the editorial boards of the *Journal of Mathematical Sociology*, the *Journal of Classification*, *Social Networks*, and *Statistics in Transition*. She was a Fulbright Scholar in 1990 and Visiting Professor at the University of Pittsburgh. She was awarded the title of Ambassador of Science of the Republic of Slovenia in 1997. Her interests include multivariate analysis (constrained and multicriteria clustering), social networks (measurement quality and blockmodeling), and survey methodology (reliability and validity of measurement).

Ove Frank was professor of statistics at Lund University, Sweden, 1974–1984, and at Stockholm University from 1984, where he recently became emeritus. He is one of the pioneers in network sampling and has published papers on network methodology, snowball sampling, Markov graphs, clustering, and information theory. Jointly with David Strauss he introduced Markov graphs in 1986 and explained how sufficient network statistics can be deduced from explicit assumptions about the dependencies in a network.

Linton C. Freeman is a research professor in the Department of Sociology and in the Institute for Mathematical Behavioral Sciences at the University of California, Irvine. He began working in social network analysis in 1958 when he directed a structural study of community decision making in Syracuse New York. Freeman was an early computer user and taught information and/or computer science at Syracuse and at the universities of Hawaii and Pittsburgh. In 1978 he founded the journal *Social Networks*. Beginning in the 1950s, and continuing to the present time, one of his continuing areas of interest has been the graphical display of network structure.

Mark Huisman is an assistant professor in the Department of Psychology of the University of Groningen. He teaches courses on statistics and multivariate statistical methods. His current research interests focus on statistical modeling of social networks, methods for nonresponse and missing data, and software for statistical data analysis.

Laura M. Koehly is an assistant professor in the Department of Psychology at Texas A&M University. She completed her Ph.D. in quantitative psychology at the University of Illinois–Urbana/Champaign, after which she completed postdoctoral training at the

University of Texas M.D. Anderson Cancer Center. Her methodological research interests focus on the development of stochastic models for three-way social network data and ego-centered network data. Her substantive research focuses on the application of social network methods in the health domain, specifically in the areas of hereditary cancers. She has recently developed a research program in organizational psychology that focuses on socialization processes within organizational settings, consensus and accuracy in perceptions of social structure, and the evolution of leadership within teams.

Peter V. Marsden is Professor of Sociology at Harvard University. His academic interests include social organization, social networks, and social science methodology. With James A. Davis and Tom W. Smith, Marsden is a co-Principal Investigator of the General Social Survey and has been a lead investigator for three National Organizations Studies conducted between 1991 and 2003.

Philippa Pattison is a professor in the Department of Psychology at the University of Melbourne. Her current research is focused on the development of dynamic network-based models for social processes and on applications of these models to a diverse range of phenomena, including mental health, organizational design, the emergence of markets, and disease transmission.

Garry Robins teaches quantitative methods in the Department of Psychology at the University of Melbourne, Australia. His research is centered on methodologies for social network analysis, particularly on exponential random graph (p^*) models. He has a wide range of collaborations arising from empirical research related to social networks.

John Scott is Professor of Sociology at the University of Essex. An active member of the British Sociological Association, he served as its president from 2001 until 2003. He has written more than fifteen books, including *Corporate Business and Capitalist Classes* (1997), *Social Network Analysis* (1991 and 2000), *Sociological Theory* (1995), and *Power* (2001). With James Fulcher, he is the author of the leading introductory textbook *Sociology* (1999 and 2003). He is a member of the Editorial Board of the *British Journal of Sociology* and is an Academician of the Academy of Learned Societies in the Social Sciences

Tom A. B. Snijders is professor of Methodology and Statistics in the Department of Sociology of the University of Groningen, The Netherlands, and Scientific Director of the Research and Graduate School ICS (Interuniversity Center for Social Science Theory and Methodology). His main research interests are social network analysis and multilevel analysis.

Thomas W. Valente is an associate professor in the Department of Preventive Medicine, Keck School of Medicine, and Director of the Master of Public Health Program at the University of Southern California. He is author of *Evaluating Health Promotion Programs* (2002, Oxford University Press); *Network Models of the Diffusion of Innovations* (1995, Hampton Press); and numerous articles on social network analysis, health communication, and mathematical models of the diffusion of innovations.

Stanley Wasserman is Rudy Professor of Sociology, Psychology, and Statistics at Indiana University. He has done research on methodology for social networks for 30 years. He has co-authored with Katherine Faust *Social Network Analysis: Methods and Applications*, published in 1994 in this series by Cambridge University Press, and has co-edited with Joseph Galaskiewicz *Social Network Analysis: Research in Social and Behavioral Sciences* (1994). He has also been a very active consultant and is currently Chief Scientist of Visible Path, an organizational network research firm.

Contents

1

Introduction

Stanley Wasserman, John Scott, and Peter J. Carrington

Interest in social network analysis has grown massively in recent years. This growth has been matched by an increasing sophistication in the technical tools available to users. *Models and Methods in Social Network Analysis (MMSNA)* presents the most important of those developments in quantitative models and methods for analyzing social network data that have appeared during the 1990s. It is a collection of original chapters by leading methodologists, commissioned by the three editors to review recent advances in their particular areas of network methods.

As is well-known, social network analysis has been used since the mid-1930s to advance research in the social and behavioral sciences, but progressed slowly and linearly, until the end of the century. Sociometry (sociograms, sociomatrices), graph theory, dyads, triads, subgroups, and blockmodels – reflecting substantive concerns such as reciprocity, structural balance, transitivity, clusterability, and structural equivalence – all made their appearances and were quickly adopted by the relatively small number of "network analysts." It was easy to trace the evolution of network theories and ideas from professors to students, from one generation to the next. The field of network analysis was even analyzed as a network (see, for example, Mullins 1973, as well as analyses by Burt in 1978, and Hummon and Carley in 1993). Many users eventually became analysts, and some even methodologists. A conference of methodologists, held at Dartmouth College in the mid-1970s, consisted of about thirty researchers (see Holland and Leinhardt 1979) and really did constitute a "who's who" of the field – an auspicious, but rather small gathering. Developments at this time were also summarized in such volumes as the methodological collection edited by Linton Freeman and his colleagues (1989), which presented a collection of papers given at a conference in Laguna Beach, California, in the early 1980s, and the collection edited by Barry Wellman and the late Stephen Berkowitz (2003 [1988]). Much of this early research has been brought together in a recent compilation, together with some later contributions (Scott 2002).

However, something occurred in about 1990. It is not completely clear to us what caused it. Interest in social networks and use of the wide-ranging collection of social network methodology began to grow at a much more rapid (maybe even increasing) rate. There was a realization in much of behavioral science that the "social contexts" of actions matter. Epidemiologists realized that epidemics do not progress uniformly through populations (which are almost never homogeneous). The slightly controversial view that sex research had to consider sexual networks, even if such networks are just dyads, took hold. Organizational studies were recognized as being at the heart

of management research (roughly one-third of the presentations at the Academy of Management annual meetings now have a network perspective). Physicists latched onto the web and metabolic systems, developing applications of the paradigm that a few social and behavioral scientists had been working on for many, many years. This came as a surprise to many of these physicists, and some of them did not even seem to be aware of the earlier work – although their maniacal focus on the small world problem (Watts 1999, 2003; Buchanan 2002) has made most of their research rather routine and unimaginative (see Barabasi, 2002, for a lower-level overview). Researchers in the telecommunications industry have started to look at individual telephone networks to detect user fraud. In addition, there is the media attention given to terrorist networks, spawning a number of methodologists to dabble in the area – see *Connections* 24(3) (2001): a special issue on terrorist networks, as well as the proceedings from a recent conference (Breiger, Carley, and Pattison 2003) on this topic. Perhaps the ultimate occurred more recently when *Business 2.0* (November 2003) named social network applications the "Hottest New Technology of 2003." All in all, an incredible diversity of new applications for what is now a rather established paradigm.

Sales of network analysis textbooks have increased: an almost unheard-of occurrence for academic texts (whose sales tend to hit zero several years after publication). It has been 10 years since the publication of the leading text in the area – *Social Network Analysis: Methods and Applications* (Wasserman and Faust 1994) – and almost 15 years since work on it began. It is remarkable not only that is it still in print, but also that increasing numbers of people are buying it, maybe even looking at parts of it. Yet, much has happened in social network analysis since the mid-1990s. Some general introductory texts have since appeared (Degenne and Forsé 1999; Scott 2000), but clearly, there is a need for an update to the methodological material discussed in Wasserman and Faust's standard reference.

Consequently, we intend *MMSNA* to be a sequel to *Social Network Analysis: Methods and Applications*. Although our view of the *important* research during the 1990s is somewhat subjective, we do believe (as do our contributors) that we have covered the field with *MMSNA*, including chapters on all the topics in the quantitative analysis of social networks in which sufficient important work has been recently published. The presentations of methodological advances found in these pages are illustrated with substantive applications, reflecting the belief that it is usually problems arising from empirical research that motivate methodological innovation. The contributions review only already published work: they avoid reference to work that is still "in progress."

Currently, no volume completely reviews the state of the art in social network analysis, nor does any volume present the most recent developments in the field. *MMSNA* is a complement, a supplement, not a competitor, to Wasserman and Faust (1994). We expect that anyone who has trained in network methods using Wasserman and Faust or who uses it as a reference will want to update his or her knowledge of network methods with the material found herein. As mentioned, the range of topics in this volume is somewhat selective, so its coverage of the entire field of network methods is not nearly as comprehensive as that of Wasserman and Faust. Nevertheless, the individually authored chapters of *MMSNA* are more in-depth, definitely more up-to-date, and more advanced in places than presentations in that book.

We turn now to the individual chapters in *MMNSA*. Peter Marsden's "Recent Developments in Network Measurement" is a significant scene-setting chapter for this whole volume. He explores the central issues in the measurement of social relations that underpin the other techniques examined in the book. His particular concern is not with measuring network structures themselves, but in the acquisition of relevant and reliable data. To this end, he looks specifically at the design of network studies and the collection of source data on social relations.

Marsden's starting point is the recognition that whole network and egocentric approaches can be complementary viewpoints on the same data. Whole network studies are concerned with the structural properties of networks at the global level, whereas egocentric studies focus on the network as it appears from the standpoint of those situated at particular locations within it. Despite this complementarity, however, issues of sampling and data selection mean that it is rarely possible to move with any ease from the "structure" to the "agent," or vice versa. Marsden examines, in particular, the implications of the identification of network boundaries on the basis of positional, event-based, and relational measures, showing how recent developments have moved beyond the conventional, and often inadequate, approaches to boundary setting.

Data collection for network analysis, in whatever kind of study, has most typically involved survey and questionnaire methods, and Marsden reviews the work of recent authors on the specific response formats for collecting factual and judgmental data on social relations. He considers in particular depth the problems of recall and recognition in egocentric approaches, especially with the use of name-generator methods, and he gives focused attention to studies that aim to collect data on subjective images and perceptions of networks rather than merely reporting actual connections. A key issue in both types of research is the meaning given to the relations by the actors – most particularly, the meaning of such apparently obvious terms as "friend." Marsden shows that a number of issues in this area are significantly related to the position that the respondent occupies in the network on which he or she is reporting. The chapter concludes with some briefer remarks on archival and observational methods where the researcher has less direct control (if any at all) over the nature of the raw data.

Marsden's remarks on the sampling problem are further considered in Ove Frank's chapter, "Network Sampling and Model Fitting." Frank has been the leading contributor to work on network sampling for many years, and here he begins from a consideration of the general issues in sampling methodology that he sees as central to the analysis of multivariate network data. A common method in network analysis has been implicit or explicit snowball sampling, and Frank looks at the use of this method in relation to line (edge) sampling as well as point (vertex) sampling, and he shows that the limitations of this method can be partly countered through the use of probabilistic network models (i.e., basing the sampling on population model assumptions). These are examined through the method of random graphs, especially the uniform and Bernoulli models, and the more interesting models such as Holland-Leinhardt's p_1, p^*, and Markov random graphs.

Frank gives greatest attention, however, to dyad-dependence models that explicitly address the issue of how points and lines are related. These are models in which network structure is determined by the latent individual preferences for local linkages, and Frank

suggests that these can be seen as generalizations of the Holland-Leinhardt p_1 model and that they are equally useful for Bayesian models. He examines log-linear and clustering approaches to choosing such models, arguing that the most effective practical solution may be to combine the two. These general conclusions are illustrated through actual studies of drug abuse, the spread of AIDS, participation in crime, and social capital.

The next group of chapters turns from issues of data design and collection to structural measurement and analysis. Centrality has been one of the most important areas of investigation in substantive studies of social networks. Not surprisingly, many measures of centrality have been proposed. The chapter by Martin Everett and Stephen Borgatti, "Extending Centrality," notes that these measures have been limited to individual actors and one-mode data. Their concern is with the development of novel measures that would enlarge the scope of centrality analysis, seeking to generalize the three primary concepts of centrality (degree, closeness, and betweenness) and Freeman's notion of centralization. They first show that it is possible to analyze the centrality of groups, whether these are defined by some external attribute such as ethnicity, sex, or political affiliation, or by structural network criteria (as cliques or blocks). A more complex procedure is to shift the measurement of centrality from one-mode to two-mode data, such as, for example, both individuals and the events in which they are involved. Although such measures are more difficult to interpret substantively, Everett and Borgatti note that they involve less loss of the original data and do not require any arbitrary dichotomizing of adjacency matrices. Finally, they look at a core-periphery approach to centrality, which identifies those sub-graphs that share common structural locations within networks.

Patrick Doreian, Vladimir Batagelj, and Anuška Ferligoj, in "Positional Analyses of Sociometric Data," examine blockmodeling procedures, reviewing both structural equivalence and regular equivalence approaches. Noting that few empirical examples of exact partitioning exist, they argue that the lack of fit between model and reality can be measured and used as a way of comparing the adequacy of different models. Most importantly, they combine this with a generalization of the blockmodeling method that permits many types of models to be constructed and compared. Sets of "permitted" ideal blocks are constructed, and the model that shows minimum inconsistency is sought. In an interesting convergence with the themes raised by Everett and Borgatti, they use their method on Little League data and discover evidence for the existence of a center-periphery structure. They go on to explore the implications of imposing pre-specified models (such as a center-periphery model) on empirical data, allowing the assessment of the extent to which actual data exhibit particular structural characteristics. They argue that this hypothesis-testing approach is to be preferred to the purely inductive approach that is usually employed to find positions in a network.

Thomas Valente's "Network Models and Methods for Studying the Diffusion of Innovations" turns to the implications of network structure for the flow of information through a network. In this case, the flow considered is information about innovations, and Valente reviews existing studies in search of evidence for diffusion processes. His particular concern is for the speed of diffusion in different networks and the implications of this for rates of innovation. A highly illuminating comparison of available mathematical models with existing empirical studies in public health using event history

analysis shows that network influences are important, but that the available data prevent more definitive conclusions from being drawn. Valente argues for the collection of more adequate data, combining evidence on both information and network structure, and the construction of more adequately theorized models of the diffusion process.

Katherine Faust's "Using Correspondence Analysis for Joint Displays of Affiliation Networks" convincingly shows the need for formal and strict representational models of the joint space of actors and relational ties. Correspondence analysis (a scaling method), she argues, allows a high level of precision in this task. Having specified the nature of the method and its relevance for social network data, rather than the more typical "actors x variable" data with which it is often used, Faust presents a novel analysis of a global trading network, consisting of international organizations and their member countries. This discloses a clear regional structure in which the first dimension separates South American from Central American countries and organizations, whereas the second dimension separates North American and North Atlantic countries from all others.

The exponential family of random graphs, p^*, has received a lot of attention in recent years, and in "An Introduction to Random Graphs, Dependence Graphs, and p^*," Stanley Wasserman joins with Garry Robins to review this recent work. Wasserman and Robins made the important generalization of the model from Markov random graphs to a larger family of models. In this chapter, however, they begin with dependence graphs to further clarify the models. They see the great value of p^* models as making possible an effective and informed move from local, micro phenomena to overall, macro phenomena. Using maximum likelihood and pseudolikelihood (based on logit models) estimation techniques, they show that the often-noted tendency towards model degeneracy (the production of trivial or uninteresting results) can be offset by using more complex models in which 3- or 4-star configuration counts are used. That is, the model incorporates the first three or four moments of the degree distribution to produce more realistic models. Evidence from simulation studies confirms the power of this approach. Indeed, degenerate models may not always be trivial, but may point to regions where stochastic processes have broken down. In making this point, they make important connections with recent developments in small world networks.

Although analyses of two-mode, affiliation networks involve one significant move away from the conventional one-mode analysis of relational, adjacency data, analyses of multiple networks involves a complementary broadening of approach. Laura Koehly and Philippa Pattison ("Random Graph Models for Social Networks: Multiple Relations or Multiple Raters?") turn to this issue of multiple networks, arguing that most real networks are of this kind. Building on simpler, univariate p^* models, they make a generalization to random graph models for multiple networks using dependence graphs. They examine both actual relations and cognitive perceptions of these relations among managers in high-technology industries, showing that the multiple network methods lead to conclusions that simply would not be apparent in a conventional single network approach. Their work is the first step toward richer models of generalized relational structures.

The idea of dependence graphs was central to the chapters of Wasserman and Robins and of Koehly and Pattison. Garry Robins and Philippa Pattison join forces to explore this key idea in "Interdependencies and Social Processes: Dependence Graphs and

Generalized Dependence Structures." They make the Durkheimian point that depen-dence must be seen as central to the very idea of sociality and use this to reconstruct the idea of social space. As they correctly point out, the element or unit in social space is not the individual but the ties that connect them, and they hold that the exploration of dependence models allows the grasping of the variety of ties that enter into the con-struction of social spaces. From this point of view, dependence graphs are to be seen as representations of proximity in social space, and network analysts are engaged in social geometry.

The analysis of social networks over time has long been recognized as something of a Holy Grail for network researchers, and Tom Snijders reviews this quest in "Models for Longitudinal Network Data." In particular, he examines ideas of network evolution, in which change in network structure is seen as an endogenous product of micro-level network dynamics. Exploring what he terms the independent arcs model, the reciprocity model, the popularity model, and the more encompassing actor-oriented model, Snijders concludes that the latter offers the best potential. In this model, actors are seen as changing their outgoing ties (choices), each change aiming at increasing the value derived from a particular network configuration. Such changes are "myopic," concerned only with the immediate consequences. A series of such rational choices means that small, incremental changes accumulate to the point at which substantial macro-level transformations of structure occur. He concludes with the intriguing suggestion that such techniques can usefully be allied with multiple network methods such as those discussed by Koehly and Pattison.

The final two chapters in the book are reviews of available software sources for visualization and analysis of social networks. The visualization of networks began with Moreno and the early sociograms, but the use of social network analysis for larger social networks has made the task of visualization more difficult. For some time, Linton Freeman has been concerned with the development of techniques, and in "Graphical Techniques for Exploring Social Network Data," he presents the latest and most up-to-date overview. The two families of approaches that he considers are those based on some form of multidimensional scaling (MDS) and those that involve an algebraic procedure. In MDS, points are optimally located in a specified, hopefully small, number of dimensions, using metric or non-metric approaches to proximity. In the algebraic methods of correspondence analysis and principal component analysis, points are located in relation to dimensions identified through procedures akin to the analysis of variance. Using data on beachgoers, Freeman shows that the two techniques produce consistent results, but an algebraic method produces a more dramatic visualization of the structure. Importantly, he also notes that wherever a network is plotted as a disc or sphere, it has few interesting structural properties. Freeman goes on to examine the use of specific algorithms for displaying and manipulating network images, focusing on MAGE, which allows points to be coded for demographic variables such as gender, age, and ethnicity. The use of this method is illustrated from a number of data sets. The longitudinal issues addressed by Snijders are also relevant to the visualization issue, and Freeman considers the use of MOVIEMOL as an animation device for representing small-scale and short-term changes in network structure. He shows the

descriptive power of this technique for uncovering social change, but also shows how it can be used in more analytical ways to begin to uncover some of the processes at work.

The final chapter turns to the issue of the software available for different kinds of network analysis. Mark Huisman and Marijtje van Duijn, in "Software for Social Network Analysis," present what is the most up-to-date review of a continually changing field. A total of twenty-seven packages are considered, excluding the visualization software considered by Freeman. Detailed attention is given to six major packages: UCINET, Pajek, MultiNet, NetMiner, STRUCTURE, and StOCNET. Wherever possible, the packages are compared using the same data set (Freeman's EIES network). This is a true road test, with interesting and somewhat surprising results. The authors conclude that there is no single "best buy" and that the package of choice depends very much on the particular questions that are of interest to the analyst.

References

Barabasi, A.-L. 2002. *Linked: The New Science of Networks*. Cambridge, Mass.: Perseus.

Breiger, R. L., Carley, K., and Pattison, P. 2003. *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*. National Academy of Sciences/National Research Council, Committee on Human Factors. Washington, DC: National Academies Press.

Buchanan, M. 2002. *Nexus: Small Worlds and the Groundbreaking Science of Networks*. New York: Norton.

Burt, R. S. 1978. "Stratification and Prestige Among Elite Experts in Methodological and Mathematical Sociology *Circa* 1975." *Social Networks*, 1, 105–8.

Degenne, A., and Forsé, M. 1999. *Introducing Social Networks*. London: Sage.

Freeman, L. C., White, D. R., and Romney, A. K. 1989. *Research Methods in Social Network Analysis*. New Brunswick: Transaction Books.

Holland, P., and Leinhardt, S. (eds.) 1979. *Perspectives on Social Networks*. New York: Academic.

Hummon, N., and Carley, K. 1993. "Social Networks as Normal Science," *Social Networks*, 15, 71–106.

Mullins, N. C. 1973. *Theories and Theory Groups in American Sociology*. New York: Harper and Row.

Scott, J. 2000. *Social Network Analysis*, 2nd ed. London: Sage. (Originally published in 1992).

Scott, J. (ed.) 2002. *Social Networks: Critical Concepts in Sociology*, 4 vols. London: Routledge.

Wasserman, S., and Faust, K. 1994. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Watts, D. 1999. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton, N.J.: Princeton University Press.

Watts, D. 2003. *Six Degrees: The Science of a Connected Age*. New York: Norton.

Wellman, B., and Berkowitz, S. (eds.) 2003 [1988]. *Social Structures: A Network Approach*. Toronto: Canadian Scholars' Press. (Originally published in 1988 by Cambridge University Press.)

2

Recent Developments in Network Measurement

Peter V. Marsden

Harvard University

This chapter considers study design and data collection methods for social network studies, emphasizing methodological research and applications that have appeared since an earlier review (Marsden 1990). It concentrates on methods and instruments for measuring social relationships linking actors or objects. Many analytical techniques discussed in other chapters identify patterns and regularities that measure structural properties of networks (such as centralization or global density), and/or relational properties of particular objects/actors within them (such as centrality or local density). The focus here is on acquiring the elementary data elements themselves.

Beginning with common designs for studying social networks, the chapter then covers methods for setting network boundaries. A discussion of data collection techniques follows. Survey and questionnaire methods receive primary attention: they are widely used, and much methodological research has focused on them. More recent work emphasizes methods for measuring egocentric networks and variations in network perceptions; questions of informant accuracy or competence in reporting on networks remain highly salient. The chapter closes with a brief discussion of network data from informants, archives, and observations, and issues in obtaining them.

2.1 Network Study Designs

The broad majority of social network studies use either "whole-network" or "egocentric" designs. Whole-network studies examine sets of interrelated objects or actors that are regarded for analytical purposes as bounded social collectives, although in practice network boundaries are often permeable and/or ambiguous. Egocentric studies focus on a focal actor or object and the relationships in its locality.

Freeman (1989) formally defined forms of whole-network data in set-theoretic, graph-theoretic, and matrix terms. The minimal network database consists of one set of objects (also known as *actors* or *nodes*) linked by one set of relationships observed at one occasion; the cross-sectional study of women's friendships in voluntary associations given by Valente (Figure 6.1.1, Chapter 6, this volume) is one example. The matrix representation of this common form of network data is known as a "who to whom" matrix or a "sociomatrix." Wasserman and Faust (1994) termed this form a *one-mode* data set because of its single set of objects.

Elaborations of the minimal design consider more than one set of relationships, measure relationships at multiple occasions, and/or allow multiple sets of objects (which

may change over occasions). Data sets with two sets of objects – termed *two-mode* by Wasserman and Faust (1994) – are common; Table 7.4.1 of Chapter 7 in this volume gives an example, a network of national memberships in trade and treaty organizations. Many studies also measure multiple relations, as in Lazega's (1999) study of collaboration, advising, and friendships among attorneys. As Snijders (Chapter 11, this volume) indicates, interest in longitudinal questions about social networks is rising; most extant data sets remain single occasion, however. In addition to relationships, almost all network data sets measure attributes (either time constant or time varying) of objects, but this chapter does not consider issues of measurement for these.

A further variation known as a *cognitive social structure* (CSS) design (Krackhardt 1987) obtains measurements of the relationship(s) under study from multiple sources or observers. Chapter 9 in this volume presents models for such data. The CSS design is widely used to study informant variations in the social perception of networks. In applications to date, observers have been actors in the networks under study, but in principle the sets of actors and observers could be disjoint.

Egocentric network designs assemble data on relationships involving a focal object (*ego*) and the objects (*alters*) to which it is linked. Focal objects are often sampled from a larger population. The egocentric network data in the 1985 General Social Survey (GSS; see Marsden 1987), for example, include information on up to five alters with whom each survey respondent "discusses important matters."

Egocentric and whole-network designs are usually distinguished sharply from one another, but they are interrelated. A whole network contains an egocentric network for each object within it (Marsden 2002). Conversely, if egos are sampled "densely," whole networks may be constructed using egocentric network data. Kirke (1996), for instance, elicited egocentric networks for almost all youth in a particular district, and later used them in a whole-network analysis identifying within-district clusters. Egocentric designs in which respondents report on the relationships among alters in their egocentric networks may be seen as restricted CSS designs – in which informants report on clusters of proximate relationships, rather than on all linkages.

Aside from egocentric designs and one-mode (single-relation or multirelational), two-mode, and CSS designs for whole networks, some studies sample portions of networks. Frank discusses network sampling in depth in Chapter 3 (this volume). One sampling design observes relationships for a random sample of nodes (Granovetter 1976). Another, known as the "random walk" design (Klovdahl et al. 1977; McGrady et al. 1995), samples chains of nodes, yielding insight into indirect connectedness in large, open populations.

2.2 Setting Network Boundaries

Deciding on the set(s) of objects that lie within a network is a difficult problem for whole-network studies. Laumann, Marsden, and Prensky (1989) outlined three generic boundary specification strategies: a positional approach based on characteristics of objects or formal membership criteria, an event-based approach resting on participation in some class of activities, and a relational approach based on social connectedness.