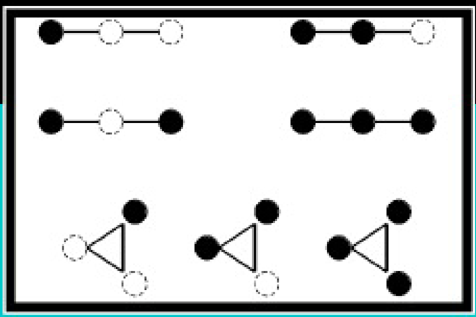
STRUCTURAL ANALYSIS IN THE SOCIAL SCIENCES 27

Models and Methods in Social Network Analysis



Edited by

Peter J. Carrington
John Scott
Stanley Wasserman

CAMBRIDGE
more information - www.cambridge.org/9780521809597

This page intentionally left blank

Models and Methods in Social Network Analysis

Models and Methods in Social Network Analysis presents the most important developments in quantitative models and methods for analyzing social network data that have appeared during the 1990s. Intended as a complement to Wasserman and Faust's *Social Network Analysis: Methods and Applications*, it is a collection of original articles by leading methodologists reviewing recent advances in their particular areas of network methods. Reviewed are advances in network measurement, network sampling, the analysis of centrality, positional analysis or blockmodeling, the analysis of diffusion through networks, the analysis of affiliation or "two-mode" networks, the theory of random graphs, dependence graphs, exponential families of random graphs, the analysis of longitudinal network data, graphic techniques for exploring network data, and software for the analysis of social networks.

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.

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.

Stanley Wasserman is Rudy Professor of Sociology, Psychology, and Statistics at Indiana University. He has done research on methodology for social networks for thirty 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 the Social and Behavioral Sciences* (1994). His work is recognized by statisticians, as well as social and behavioral scientists, worldwide. He is currently Book Review Editor of *Chance* and an Associate Editor of the *Journal of the American Statistical Association* and *Psychometrika*. He has also been a very active consultant and is currently Chief Scientist of Visible Path, an organizational network software firm.

Structural Analysis in the Social Sciences   27

Mark Granovetter, General editor

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 determination; and (d) explanation 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 that 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 reduction or the determination of ideas, technology, or material conditions. Although 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, this series hopes to encourage the use of this very fruitful approach.

Other books in the series:

1. Mark S. Mizruchi and Michael Schwartz, eds., *Intercorporate Relations: The Structural Analysis of Business*
2. Barry Wellman and S. D. Berkowitz, eds., *Social Structures: A Network Approach*
3. Ronald L. Brieger, ed., *Social Mobility and Social Structure*
4. David Knoke, *Political Networks: The Structural Perspective*
5. John L. Campbell, J. Rogers Hollingsworth, and Leon N. Lindberg, eds., *Governance of the American Economy*
6. Kyriakos Kontopoulos, *The Logics of Social Structure*
7. Philippa Pattison, *Algebraic Models for Social Structure*
8. Stanley Wasserman and Katherine Faust, *Social Network Analysis: Methods and Applications*
9. Gary Herrigel, *Industrial Constructions: The Sources of German Industrial Power*
10. Philippe Bourgois, *In Search of Respect: Selling Crack in El Barrio*
11. Per Hage and Frank Harary, *Island Networks: Communication, Kinship, and Classification Structures in Oceana*
12. Thomas Schweizer and Douglas R. White, eds., *Kinship, Networks, and Exchange*
13. Noah E. Friedkin, *A Structural Theory of Social Influence*
14. David Wank, *Commodifying Communism: Business, Trust, and Politics in a Chinese City*
15. Rebecca Adams and Graham Allan, *Placing Friendship in Context*
16. Robert L. Nelson and William P. Bridges, *Legalizing Gender Inequality: Courts, Markets and Unequal Pay for Women in America*

Continued after the Index

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

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

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521809597

First published in print format 2005

ISBN-13 978-0-511-11004-7 eBook (Adobe Reader)
ISBN-10 0-511-11004-9 eBook (Adobe Reader)

ISBN-13 978-0-521-80959-7 hardback
ISBN-10 0-521-80959-2 hardback

ISBN-13 978-0-521-60097-2 paperback
ISBN-10 0-521-60097-9 paperback

Contents

Acknowledgments

The editors want to thank Mary Child and Ed Parsons of Cambridge University Press and Mark Granovetter, the general editor of the series, for their support and patience during the long genesis of this volume. We are also grateful to Anthony Matarazzo, who prepared the index, and who created and maintained the Web site that served as a virtual workplace and meeting place for everyone who contributed to the book.

We are, of course, very grateful to our contributors. Their expertise and hard work have made this an easy project for us. Thanks go to all of them.

Preparation of the book was supported by Social Sciences and Humanities Research Council of Canada Standard Research Grants No. 410-2000-0361 and 410-2004-2136 and U.S. Office of Naval Research Grant No. N00014-02-1-0877.

We dedicate this volume to social network analysts everywhere, in the hope that they will find these chapters useful in their research.

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.

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 dependence 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 construction 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.

Employment by an organization (e.g., Krackhardt 1990) is one positional criterion. The "regulars" at a beach depicted by Freeman (Figure 12.2.3, Chapter 12, this volume; see also Freeman and Webster 1994) were identified via an event-based approach; regulars were defined as persons observed 3 or more days during the study period.

Doreian and Woodard (1992) outlined a specific version of the relational approach called *expanding selection*. Beginning with a provisional "fixed" list of objects deemed to be in a network, it then adds objects linked to those on the initial list. This approach is closely related to the snowball sampling design discussed by Frank in Chapter 3, this volume; Doreian and Woodard, however, added a new object only after finding that it had several links (not just one) to elements on the fixed list. They review logistical issues in implementing expanding selection, and compare it with the fixed-list approach in a study of social services networks. More than one-half of the agencies located via expanding selection were not on the fixed list. Added agencies were closely linked to one another, although the fixed-list agencies were relatively central within the expanded network. The fixed-list approach presumes substantial prior investigator knowledge of network boundaries, whereas expanding selection draws on participant knowledge about them.

Elsewhere, Doreian and Woodard (1994) suggested methods for identifying a "reasonably complete" network within a larger network data set. They used expanding selection to identify a large set of candidate objects, and then selected a dense segment of this for study. They adopted Seidman's (1983) "k-core" concept (a subset of objects, each linked to at least k others within the subset) as a criterion for setting network boundaries. By varying k, investigators can set more and less restrictive criteria for including objects.

Egocentric network studies typically set boundaries during data collection. The "name generator" questions discussed in this chapter accomplish this.

2.3 Survey and Questionnaire Methods

Network studies draw extensively on survey and questionnaire data. Surveys allow investigators to decide on relationships to measure and on actors/objects to be approached for data. In the absence of archival records, surveys are often the most practical alternative: they make much more modest demands on participants than do diary methods or observation, for example. Surveys do introduce artificiality, however, and findings rest heavily on the presumed validity of self-reports.

Both whole-network and egocentric network studies use survey methods, but the designs typically differ in how they obtain network data and in what they ask of respondents. A whole-network study usually compiles a roster of actors before data collection begins. Survey and questionnaire instruments incorporate the roster, allowing respondents to recognize rather than recall their relationships. Egocentric studies, however, are often conducted in large, open populations. The alters in a respondent's network are not known beforehand, so setting network boundaries must rely on respondent recall.

Whole-network studies ordinarily seek interviews with all actors in the population, and ask respondents to report only on their direct relationships. (The CSS studies discussed later are an exception; they ask for much more data.) In egocentric studies, however, practical and resource considerations usually preclude interviewing a respondent's alters. Such studies ask respondents for data on their own relationships to alters, and also often ask for information on linkages between alters; moreover, they commonly request proxy reports about alters.

Surveys and questionnaires in whole-network studies use several response formats to obtain network data: binary judgments (often termed *sociometric choices*) about whether respondents have a specified relationship with each actor on the roster, ordinal ratings of tie strength, or rankings. Binary judgments are least difficult for respondents; ranking tasks are most demanding. Eudey, Johnson, and Schade (1994) found that a large majority of respondents preferred rating over ranking tasks. Ferligoj and Hlebec (1999) reported the reliability of ratings to be somewhat higher than that of binary judgments.

Batchelder (1989) considered network data of different scale types (dichotomous, ordinal, interval, ratio, absolute) and the inferences about network-level properties (e.g., reciprocation, presence of cliques) that can be drawn meaningfully from them. Among other things, Batchelder showed that findings may be affected if respondents have differing thresholds for claiming a given type of tie when making dichotomous judgments; Feld and Carter (2002) referred to this as *expansiveness bias* (see also Kashy and Kenny 1990). Likewise, implicit respondent-specific scale and location constants for rating relationship strength can complicate inferences. Eudey et al. (1994), however, used both ratings and rankings in studying a small group, and found quite high correlations between measures based on the two response formats.

Surveys sometimes include "global" items asking respondents about the size, density, or composition of their egocentric networks. Such questions pose extensive cognitive demands. To answer a global network density question, for instance, respondents must decide who their alters are, ascertain relationships among alters, and aggregate (Burt 1987). Sudman (1985) measured network size using both a global item and a recognition instrument; the measures had similar means, but the global item had a far greater variance. Instead of global items, contemporary studies usually measure egocentric networks using multiple-item instruments that ask respondents for only one datum at a time.

(A) *Name Generator Instruments for Egocentric Networks*

Surveys have long collected data on a respondent's social contacts and relationships (Coleman 1958). Such egocentric network instruments typically include two types of questions (Burt 1984): *name generators* that identify the respondent's alters, and *name interpreters* that obtain information on the alters and their relationships. Name generators are free-recall questions that delineate network boundaries. Name interpreters elicit data about alters and both ego–alter and alter–alter relationships. Many indices of network form and composition are based on such data.

Instruments for egocentric networks use both single and multiple name generators. A single-generator instrument focusing on alters with whom respondents "discuss important matters" first appeared in the 1985 GSS, and later in several other studies (Bailey and Marsden 1999). It tends to elicit small networks of "core" ties; Marsden (1987) reported a mean network size of 3.0 for U.S. adults in 1985, whereas Ruan et al. (1997) reported a mean of 3.4 for adults in a Chinese city in 1993. Hirsch's (1980) Social Network List (SNL) for social support networks is another one-generator instrument. Respondents list up to twenty persons they regard as "significant" and have seen during the prior 4 to 6 weeks.

Any given name-generating relationship elicits only a fraction of a respondent's social contacts. Moreover, many conceptual understandings of networks extend beyond "core" ties to include more mundane forms of social support. Fischer (1982a), for example, used name generators for instrumental aid and socializing, as well as confiding. Fischer and Shavit's (1995) U.S.–Israel support network comparison used a multiple-generator instrument. Another example is the Social Support Questionnaire (SSQ; Sarason et al. 1983), a twenty-seven-generator instrument eliciting persons to whom respondents can turn and on whom they can rely in differing circumstances.

The first consideration in choosing between single and multiple name generator instruments must be a study's conceptualization of a network. Single-generator methods may be sufficient for core networks, but more broadly defined support networks almost certainly require multiple name generators. A practical issue is the availability of interview time. Multiple-generator instruments that elicit many alters can be quite long, and measuring egocentric networks must be a central focus of studies including them.

More extensive definitions of "a network" include alters and relationships that do not provide even minor social support. McCarty et al. (1997) sought to measure features of "total personal networks," including all alters "known" by a respondent, those who "would recognize the respondent by sight or by name" (p. 305). Networks thus defined are too large to enumerate fully. McCarty et al. sampled total network alters by selecting a series of first names and asking if respondents know anyone by those names; they posed name interpreter questions about the sampled alters. The authors acknowledge that age, gender, and race/ethnic differences in naming practices may limit the representativeness of their samples. Nonetheless, their sampled total networks are less dense and less kin centered than are core or support networks, as one would anticipate. Further investigation of this technique as a means of measuring extensively defined egocentric networks seems warranted.

Because name generator instruments are complex by comparison with conventional survey items (Van Tilburg 1998), they often are administered in person so interviewers can assist respondents who need help completing them. Such instruments have, however, appeared in both paper-and-pencil (Burt 1997) and computerized questionnaires (Bernard et al. 1990; Podolny and Baron 1997). Little research has examined differences in data quality by data collection mode.

Methodological research on name generator instruments rarely addresses questions of validity because criterion data from other sources are unavailable. Some test–retest studies of instrument reliability are reviewed subsequently. Most research, however, examines the in-practice performance of instruments: how name generators differ, how respondents handle sometimes challenging tasks that instruments pose, and how key terms are understood. Much of this research reflects attention to cognitive and communicative processes involved in answering survey questions (Sudman, Bradburn, and Schwarz 1996).

Comparing Name Generators

Several studies systematically compare properties of name generators. Campbell and Lee (1991), Milardo (1992), and Van der Poel (1993) highlighted conceptual differences between generators in criteria for including alters. Some refer to specific social exchanges, such as discussing important matters or borrowing household items; others use affective criteria ("closeness"); others specify particular role relations such as kinship or neighboring; and still others measure frequent interaction. Also, some generators specify temporal (e.g., contact within the prior 6 months) or spatial/organizational restrictions on eligible alters (Campbell and Lee 1991).

Varying name generator content influences egocentric network size, among other features. Campbell and Lee (1991) and Milardo (1992) showed that intimate name generators – whether affective or exchange based – elicit smaller networks than those specifying less intense thresholds for naming alters. Mean network sizes reported in seven intimate generator studies (all in North American settings) range between three and seven. Multiple-generator exchange-based instruments produce appreciably larger networks; across seven studies using such instruments, mean network size ranged between ten and twenty-two. Studies using exchange-based name generators tended to produce networks having smaller fractions of family members than did those using intimate generators.

Bernard et al. (1990) administered the GSS name generator and an eleven-generator social support instrument within a single study. The GSS instrument elicited smaller networks than did the social support instrument. These were core contacts: about 90% of GSS alters were also named for the social support instrument.

Instruments with many name generators impose appreciable respondent burden. Three studies suggest small sets of name generators for measuring support networks. Van der Poel (1993) identified subsets of name generators that best predict the size and composition of networks elicited using a ten-generator instrument. A three-generator subset consists of items on discussing a major life change, aid with household tasks, and monthly visiting; a five-generator version adds borrowing household items and going out socially. Bernard et al. (1990) isolated questions about social activities, hobbies, personal problems, advice about important decisions, and closeness as a "natural group" of name generators. Burt (1997) used a construct validity criterion – the association between network constraint and achievement – in an organizational setting. He concluded that a minimal module of name generators should measure both intimacy and activity; it might consist of the GSS "important matters" item, socializing, and discussion of a job change.

Recall, Recognition, and Forgetting

Brewer (2000) reviewed nine studies that asked respondents first to freely recall lists of persons, and then to supplement their lists after consulting an inventory listing all eligible persons. For instance, Brewer and Webster (1999) asked dormitory residents to recall their best friends, close friends, and other friends; the respondents then reviewed a dormitory roster and could add to each list of friends. Friends recognized on the roster were deemed to have been "forgotten" in the recall task.

Across studies, Brewer reported an appreciable level of forgetting, although it varied substantially across groups and relationships. In the dormitory study, one-fifth of all friends were not named in the recall task. As in several other studies Brewer reviewed, the likelihood of forgetting alters varied inversely with tie strength: students forgot only 3% of best friends and 9% of close friends, but added 26% of other friends after inspecting the dormitory listing.

Brewer's review makes it clear that name generators elicit only a fraction of those persons having a criterion relationship to a respondent, and that intimate name generators enumerate a larger fraction of eligible alters than do weaker ones. Implications of these findings depend on the purposes for which network data are used. If one seeks to describe a network precisely or to contact alters (e.g., partner notification concerning an infectious disease; Brewer, Garrett, and Kulasingam 1999), then any shortfall in the enumeration of alters is an obvious drawback. If instead a study seeks indices contrasting the structure and composition of networks, then forgetting is more serious to the extent that indices based on the recalled and recalled/forgotten sets of alters diverge. Brewer and Webster (1999), for example, reported relatively high correlations between measures of centrality, egocentric network size, and local density based on recalled alters only, and the same measures based on recalled and recognized alters. They found appreciable differences in some network-level properties, however.

Brewer (2000) suggested several steps toward reducing the level of forgetting. These include the use of recognition rather than recall when possible and, if using recall methods, nonspecific probes for additional alters. Using multiple name generators may limit forgetting because persons forgotten for one generator are often named in response to others.

Test–Retest Studies

Brewer (2000) also reviewed eight test–retest studies. These used a variety of affective, support, and exchange name generators. Most test–retest intervals were 1 month or less. In all but one study, more than 75% of first-occasion alters were also cited at the second occasion. Brewer suggested that respondents may have forgotten the uncited alters.

Two studies examine over time stability in network size for social support instruments. Rapkin and Stein (1989) measured networks over a 2-month interval using both closeness and "importance" criteria. Between-occasion correlations of network size were 0.72 and 0.56, respectively. Size declined over time for both criteria, however, suggesting that respondents were unenthusiastic about repeating the task on the second occasion. Bass and Stein (1997) found higher 4-week stability in network size for the support-based SSQ (Sarason et al. 1983) than for the affective SNL (Hirsch 1980).

Morgan, Neal, and Carder (1997) conducted a seven-wave panel study of widows, using an importance criterion to elicit networks every 2 months. Core networks were very stable – 22% of alters were named on all seven occasions. These were often family members. There was also much flux at the periphery because 24% of alters were named only once. Morgan et al. found network properties to be more stable across occasions than were alters. They suggest that between-occasion differences in alters mix unreliability (or forgetting) and genuine turnover.

Patterns in the Free Recall of Persons

Several studies of social cognition have examined the free recall of persons under different conditions. Their findings suggest strongly that social relationships organize memories for persons. Understanding these principles of memory organization can improve instruments such as name generators that seek to tap into such memories.

Bond, Jones, and Weintraub (1985) asked subjects to name acquaintances ("people you know") and recorded the order in which acquaintances were named. Successive nominations tended to be clustered by affiliations with social groups, rather than by similarity in physical or personality characteristics. Moreover, the time intervals separating names within a given group tended to be short; subjects paused for longer periods between names of persons in different groups. Social relations thus appear to be an important basis for remembering persons: Bond et al. concluded that "the person cognizer is more a sociologist than an intuitive psychologist" (p. 336). Fiske (1995) reported results for two similar studies; clusters of persons named by his subjects were grouped much more strongly by relationships than by similarity of individual features such as gender, race, or age.

Brewer (1995) conducted three studies asking subjects to name all persons within a graduate program, a religious fellowship, and a small division of a university. He too found that memory for persons reflects social relational structures: names of graduate students, for example, tended to be clustered by entering cohort, and shorter time intervals intervened between the naming of persons within a cohort than those in different cohorts. More generally, perceived social proximity appears to govern recall of persons. Brewer also found that subjects tended to name persons in order of salience. Those in groups proximate to the subject tended to be named first, as were persons of high social status and those frequently present in a setting.

These studies suggest that respondents recall alters in social clusters when answering name generators. The basis for clustering likely varies across situations, but it is plausible that foci of activity such as families, neighborhoods, workplaces, or associations (Feld 1981) offer a framework for remembering others. Aiding respondent recall with reminders of such foci might encourage more complete delineation of alters. Brewer's studies also indicate that respondents tend to order their nominations of alters by tie strength (see Burt 1986).

The Meaning and Interpretation of Name Generators

Name generators always refer to a specific type of social tie, and researchers assume that respondents share their understanding of this criterion. Fischer (1982b) questioned this assumption for "friends" (see Kirke 1996, however). He and others suggested that meanings are more apt to be shared for specific exchanges than for role labels or affective criteria. This calls for studies of the meanings attributed to exchange name generators.

Because it has been widely used, several studies have examined the GSS "important matters" name generator. Respondents decide what matters are "important" while answering, so the content of the specific exchanges it measures may vary. Ruan (1998) investigated the intersection between the sets of alters named for the GSS name generator and those for several subsequently administered exchange name generators. In her Chinese urban sample, the GSS name generator elicited social companions and persons with whom private issues are discussed, but not alters providing instrumental aid.

Bailey and Marsden (1999) used concurrent think-aloud probes to investigate how respondents interpret the GSS name generator. Their convenience sample of U.S. adults offered a variety of interpretations: some respondents referred to specific matters, but others translated the question into one about intimacy, frequent contact, or role labels. When probed about the matters regarded as "important," most respondents referred to personal relationships; health, work, and politics were other often-mentioned categories. Differences in interpretive framework or definitions of important matters were not strongly associated with the types of relationships elicited, however.

Straits (2000) conducted an experiment: one-half of his student sample answered the GSS name generator, whereas the other half answered a generator about "people especially significant in your life." The two question wordings produced virtually identical numbers of alters. Only modest compositional differences were observed: women named a somewhat greater number of male alters for the "significant people" question than for the "important matters" question. Overall, however, Straits concluded that the "important matters" criterion also elicits "significant people."

McCarty (1995) investigated respondent judgments of how well they "know" others. Indicators of tie strength – closeness, duration, friendship, kinship – were associated with knowing alters well. Frequent contact was linked to knowing others moderately well. Low levels of knowing were distinguished by awareness of factual (but not personal) information and acquaintanceship.

Interview Context Effects

When name generators contain terms requiring interpretation, respondents may look to the preceding substantive content of an interview for cues about their meaning. A context experiment was embedded in the Bailey and Marsden (1999) study. One-half of the respondents answered a series of questions about politics before the "important matters" name generator; the other half began with questions about family. When subsequently debriefed about what types of matters were "important," family-context respondents were considerably more likely to mention family matters than were political-context respondents. Because this study is based on a small sample, these findings only suggest the prospect that context influences the interpretation of a name generator.

Interviewer Effects

Three nonexperimental studies document sizable interviewer differences in the size of egocentric networks elicited by name generator methods. Van Tilburg (1998) studied a seven-generator instrument with an elderly Dutch sample, reporting a within-interviewer correlation of network size of more than 0.2. This fell only modestly after controls for respondent and interviewer characteristics. Marsden (2003) studied a single-generator instrument eliciting "good friends" administered in the 1998 GSS, finding a somewhat smaller (0.15) intraclass correlation than Van Tilburg's. Straits (2000) reported a similar figure (0.17) for the GSS "important matters" name generator administered by his student interviewers.

These interviewer differences are much larger than typical for survey items (Groves and Magilavy 1986). Large interviewer effects are, however, common for questions like name generators that ask respondents to list a number of entities. One conjecture is that interviewer differences reflect variations in the extent of probing. The findings highlight the need for careful interviewer training to ensure standardized administration of name generators. They also suggest the potential value of computer-assisted methods for obtaining network data, which operate without interviewers.

Name Interpreters

Although name generators have attracted much methodological interest, name interpreter items provide much of the data on which measures of egocentric network form and composition rest. Once alters are enumerated, most instruments follow up with questions about each alter and about pairs of alters.

The survey research literature on proxy reporting (e.g., Moore 1988) includes many studies comparing self-reports with proxy reports. In most of these, proxy respondents report on others in their households, so findings may not apply directly to reports about alters in an egocentric network. Sudman et al. (1994) observed that memories about others (especially distant others) are less elaborate, less experientially based, and less concerned with self-presentation than are memories of the self. This implies that self- and proxy reporters use different tactics to answer questions. Proxy respondents are prone, for example, to anchor answers on their own behavior, rather than retrieving answers directly from memory (Blair, Menon, and Bickart 1991). Sudman et al. (1994) hypothesized that the quality of proxy reports rises with respondent–alter interaction, and offered supportive data from a study of spouses.

Studies in the network literature establish that survey respondents can report on many characteristics of their alters with reasonable accuracy (Marsden 1990). White and Watkins (2000) found that Kenyan village women could report observable data on their alters – such as number of children or household possessions – relatively well. Ego–alter agreement was much lower for use of contraception, something often kept secret. Respondents often projected their own contraceptive behavior onto alters.

Shelley et al. (1995) studied networks of HIV^+ informants. Most sought to limit knowledge of their HIV status to certain alters; only one-half of the relatives in these networks were said to know the informant's HIV status. Nonetheless, informants reported that this was a better-known datum than several others, including political party affiliation and blood type. Such findings call for caution in formulating name interpreters because respondents may often lack certain information about their alters.

In addition to proxy reports, important name interpreters refer to ego–alter and alter–alter ties. Studies of network perception discussed subsequently are relevant to understanding answers to such questions.

Providing name interpreter data about a series of alters can be a repetitive, tedious task. White and Watkins (2000) noted that their respondents quickly became bored when answering such questions, and they therefore asked about no more than four alters. A useful step toward limiting respondent burden is to ask some or all name interpreter items only about a subset of alters (or dyads), as in Fischer (1982a) and McCarty et al. (1997). Acceptably reliable measures of network density and composition are often available from data on only three to five alters (Marsden 1993).

(B) *Additional Instruments for Egocentric Networks*

Many name generator instruments do not elicit weak ties that are crucial in extending network range. In addition, even single-generator instruments require substantial interview time and pose notable respondent burdens. This section reviews alternative instruments developed to address such limitations.

Instruments for Measuring Extensive Network Size

Estimating the size of extensive egocentric networks, including all alters someone "knows," is difficult in large, open populations. Several survey instruments have been developed for network size. The "summation" method (McCarty et al. 2001) uses global network questions to estimate the numbers of persons with whom respondents have sixteen relationships (e.g., family, friendship, neighboring), taking the sum of a respondent's answers as total network size. Two U.S. surveys using this method estimate that mean network size lies between 280 and 290.

Killworth et al. (1998b) developed "scale-up" methods that estimate extensive network size using data on the known size of subpopulations, such as people named "Michael" or people who are postal workers. These methods rest on the proposition that egocentric network composition resembles population composition, that is,

$$\frac{m}{c} = \frac{e}{t},$$

where m is the number of alters from some subpopulation in an egocentric network, c is network size, e is subpopulation size, and t is population size. Survey data on m, together with data on e and t from official statistics or other archives, lead to scale-up estimates of network size c.

The previous proposition will not, of course, hold precisely for all persons and subpopulations. Implementations of the scale-up approach estimate c using data on m and e for several subpopulations. Studies using the approach yield a range of values for mean network size. Killworth et al. (1990) obtained a mean of around 1,700 for U.S. informants, and one of about 570 for Mexico City informants; these estimates assume a broad definition of "knowing" ("ever known during one's lifetime"). Killworth et al. (1998a) reported the mean size of "active networks" (involving mutual recognition and contact within the prior 2 years) to be about 108 for Floridians; Killworth et al. (1998a) obtained a mean active network size of 286 from a U.S. survey. The authors note that scale-up methods depend heavily on a respondent's abilities to report accurately on the numbers of persons known within subpopulations.

The *reverse small world* (RSW) method (see, e.g., Killworth et al. 1990) is still another approach to measuring extensive networks. It presents respondents with many (often 500) "target" persons described by occupation and location, asking for an alter more likely than the respondent to know each target. RSW identifies alters who could be instrumentally useful; it omits those who are known, but not judged to be useful. Bernard et al. (1990) reported mean RSW network sizes of 129 for Jacksonville, Florida, informants, and 77 for Mexico City informants.

Position Generators

Rather than identifying particular alters and later ascertaining their social locations using name interpreters, the "position generator" measures linkages to specific locations directly. It asks respondents whether they have relationships with persons in each of a set of social positions. For example, Lin, Fu, and Hsung (2001) asked respondents if they have any relatives, friends, or acquaintances who hold fifteen different occupations. Follow-up questions may ascertain the strength of links to locations. Position generator data allow construction of indices of network range (e.g., number of occupations contacted) and composition (e.g., most prestigious occupation contacted).

Several empirical studies (e.g., Erickson 1996) use the position generator effectively. It identifies weak and strong contacts, if the threshold for contact with locations is of low intimacy; Erickson, for example, asked respondents to "count anyone you know well enough to talk to even if you are not close to them" (1996: p. 227). Because position generators do not ask about individual alters, they require less interview time than do many name generator instruments. However, position generators measure network range and composition only with respect to the social positions presented. Most applications focus on class or occupational positions; thus, the resulting data do not reflect racial or ethnoreligious network diversity, for example.

Smith (2002) experimentally compared measures of interracial friendship based on a one-item position generator, a name generator instrument, and a global approach in the 1998 GSS. His global items asked for a respondent's number of "good friends" and the number who are of a different race. Percentages of respondents claiming interracial good friends were highest for the position generator (whites, 42%; blacks, 62%), intermediate for the global approach (whites, 24%; blacks, 45%), and lowest for the name generator instrument (whites, 6%; blacks, 15%). Smith suggested that the name generator approach provides the most valid figures because it enumerates friends first, and later determines their race. The other approaches focus attention on the particular social location (race) of interest, encouraging respondents to inventory their memories for anyone who might meet the "good friend" criterion. Respondents seeking to present themselves favorably might alter their definition of "good friend" so they can report an interracial friend. Smith's findings may or may not apply to position generators measuring contact with occupational positions. Further instrument comparisons like this are needed.

The Resource Generator

Very recently, Van der Gaag and Snijders (2004) proposed the "resource generator" as an instrument for measuring individual-level social capital, which they defined as "resources owned by the members of an individual's personal social network, which may become available to the individual" (p. 200). Their instrument focuses on whether a survey respondent is in personal contact with anyone having specific possessions or capacities, such as the ability to repair vehicles, knowledge of literature, or high income. The resource generator does not enumerate specific social ties: in its most elementary