Models for
Social Networks
With Statistical Applications

Suraj Bandyopadhyay | A R. Rao | Bikas K. Sinha

ABOUT THE AUTHORS

Suraj Bandyopadhyay (PhD, Sociology, McGill University, Montreal, Canada) was, before his retirement, professor of sociology and head of the Sociological Research Unit, Indian Statistical Institute, Kolkata. After retirement, for sometime, he was also affiliated with the Sociological Research Unit and Statistics-Mathematics Unit, Indian Statistical Institute, Kolkata, as an honorary visiting scientist. For his academic performance at McGill University he was awarded Bobbs-Merrill Award in Sociology in 1968. He has received a number of academic invitations and Fellowships from different international institutions such as the Canadian International Development Agency (CIDA, Ottawa), Centre for Developing-Area Studies (McGill University), International Data Library and Reference Service of the Survey Research Centre at the University of California (Berkeley, USA), and Overseas Development Group at the University of East Anglia (Norwich, UK). He has served as expert member in academic committees of different institutions and published more than thirty research papers.

A R. Rao (PhD, Indian Statistical Institute) was a professor in the Division of Theoretical Statistics and Mathematics of the Indian Statistical Institute at Kolkata. He was a visiting assistant professor at the University of Minnesota for one year. He published more than 25 research papers in graph theory, is a joint author of a book on linear algebra at the honors level, and edited the proceedings of three conferences on graph theory. His major interests were graph theory and its applications to social sciences and linear algebra.

Bikas K. Sinha (PhD, Statistics, Calcutta University) is currently a professor of statistics in the Applied Statistics Division of the Indian Statistical Institute, Kolkata. He was a recipient of PCMahalanobis Medal in 1980. He has served as an expert on mission in survey methodology for the United Nations and has also served as a consultant for the U.S. Environmental Protection Agency (EPA).

He has traveled extensively and visited a host of universities in the United States, Canada, Germany, Finland, Poland and other countries as a visiting faculty/research collaborator. He has authored more than 110 research articles in refereed journals, one graduate-level textbook (Wiley) and two research monographs (Springer-Verlag Lecture Notes Series in Statistics Publications). He is an elected member of the International Statistical Institute. His range of expertise includes survey theory and methods, design of experiments, statistical modeling, and statistical inference.

PREFACE

———✦•✦———

This book is dedicated to the hallowed memory of our very dear colleague, the late Professor A R. Rao, who suddenly left for his heavenly abode while we were in the middle of this project. His thorough and disciplined specialized training in graph theory and his intrinsic interest in graph-theoretic and statistical aspects of social networks brought him close to one of us (SB) almost 30 years ago. His insightful critical comments have been instrumental in shaping the course of our thoughts in writing this book. He had very enthusiastically prepared the draft of the first few chapters, and we have tried to keep them almost intact. We have endeavored to complete the book in our own way—without having the tension to receive his critical comments. We fondly hope this has met his expectations.

The study leading to this book is rooted in experiences of a real-life situation of villagers in a typical rice-producing zone in Md. Bazar Community Development Block of Birbhum District in West Bengal. This was accrued in course of the project. "The Conditions of Rural Progress in India," undertaken by Suraj Bandyopadhyay (Indian Statistical Institute [ISI], Kolkata) and Donald Von Eschen (McGill University, Montreal, Canada). The project was funded by the Canadian International Development Agency (CIDA, Ottawa, Canada).

The study initially covered, during the years 1971 to 1974, 2,697 households in 21 villages. The final report of the study was submitted to CIDA in 1981, and an extended summary of its findings was published later (Bandyopadhyay & von Eschen, 1991 [see Chapter 1 for detailed reference]). The study report is available from the CIDA (see Chapter 1 for exact reference). Subsequently, some of the 21 villages were resurveyed after 25 to 30 years during 1998–2004 (under a project sponsored by the Indian Statistical Institute, Kolkata), and the data were analyzed by a team, including the present three authors.

The villages were marked by steep stratification by caste, class, and power. Again, the three systems of stratification were confounded with one another. This made the villagers, and the lower rung in particular, highly vulnerable to various crises and urgencies of life and living. The situation was further aggravated by the fact that a major source of livelihood of the village elite was appropriation, not production. Distrust of others and hopelessness about the conditions changing for the better in the near future characterized the worldview prevalent in the villages at that time.

Under such circumstances, it was sought to examine the commonly held view that the ethos of rural society, unlike that of urban society, is most likely to be conducive to promote mutual help and cooperation, which would enable the villagers, particularly the lower rung, to withstand at least partially the negative effects of steep stratification. Data on requests for help and cooperation provided from one household to another at the time of an emergency or crisis were collected in detail in each of the above 21 villages within a specified reference period. Articulation of the social relation of "help and cooperation" brought out "social networks" of help and cooperation among the households in each of these villages. Briefly speaking, the findings of social network analysis (SNA) belied expectations about rural society. The incidence of ties of help, especially the ties of mutual help, varied widely from village to village. In the lower rung of the society, the level of mutual help was found to be quite low. Even in the case of matters such as an emergency in daily family life, most of the households were dependent on the local elite. Asymmetrical relations dominated the social networks on the whole. The rural elite enjoyed not only high social position and economic strength, besides being bestowed with positions of power, but also were found to be more integrated among themselves by ties of mutual help and cooperation (Bandyopadhyay & von Eschen, 1988, 1995 [see Chapter 1 for detailed references]).

Recent official records claim that during the past few decades, a number of socioeconomic and administrative measures have been implemented in rural areas of West Bengal. Some examples include: land reforms, registration of rights of share croppers and distribution of surplus land among the landless, a three-tier system of Panchayats, an extension of minor irrigation and road transport facilities, a total literacy campaign, and others. Whereas the impact of implementing such measures for changes in socioeconomic conditions of the villagers may be explicitly manifested, there are hardly any studies of the impact, if any, on social relations among the villagers! One of the objectives of the

resurvey was to ascertain whether the implementation of these measures in rural areas has made any perceptible impact on the pattern of social relations among the villagers. We intended to study this facet of societal change by an analysis of social networks of help and cooperation in the course of daily life.

The focus of the earlier study was to explore the methodology of the quantification of data relating to the flow of help and cooperation among the villagers and to measure the extent of reciprocity within the village community. Subsequently, it was expanded to include changing patterns of some important structural parameters of a group or community—namely, fragmentation, hierarchy, reachability (how far who can reach whom), power, potential sphere of influence, and so on. For our purposes, appropriate measures had to be derived since these were not readily available in the literature and, even if available in a few cases, suffered from limitations. Thus, we stepped into the study of characteristics of global (whole) social network. In this context, we gratefully acknowledge the interest shown by Professor Barry Wellman when he visited ISI, Kolkata, almost two decades ago. At that time, he discussed with us about the potentialities and constraints of whole social network analysis. He welcomed our endeavor to analyze a whole social network as a unit and to subsequently share our experiences with the community of social network researchers. We particularly mention that whole social network analysis, whether of a village or of an ethnic community, immensely contributes to deeper comprehension of various nuances of social reality. As for the present study, supplemented by ground data, it has revealed alterations in the distribution of social power. This has also indicated shifts in power centers, traditionally rooted in socioeconomically dominant strata in the community to its lower rungs (Chapter 1). We have shown how qualitative concepts of social characteristics can be measured by whole social network analysis and made amenable to quantitative analysis (Chapter 5) and how they may be applied to examine structural patterns of social networks (Chapter 6). Optimist social scientists have surmised that "community" has not been lost but transformed into social networks (Introduction to Part II on community in Wellman & Berkowitz, 1988, p. 125; see Wellman, 1988, at the end of Chapter 1 for detailed reference). We, however, find that a community remains a community while the distribution of ties in a network changes whether longitudinally over a period of time or horizontally across space. Ties of direct reciprocity that closely bind dyads or moderately sized groups of actors are transformed to indirectly bind together quite a large segment of the network (Chapter 6). In this

context, the possibility of applying multivariate statistical methods to measure the structural similarity of social networks has also been demonstrated. Concern for methodology of whole network analysis has also prompted us to discuss elementary global social network models, problems of sampling and inference in the case of whole social network analysis, and model validation as well as application of methodology of computer simulation to estimate complex measures of whole social networks (Chapters 4, 6, and 7).

Evidently, such a venture could not be successful without interdisciplinary collective teamwork. It started with a sociologist's (SB's) query, which was taken up by a graph theorist (ARR) once a social network was recognized as a kind of digraph (courtesy of the late Professor B. P. Adhikary, then director of ISI, who had introduced SB to ARR), and later, a statistician (BKS) joined the endeavor. So far, we have a number of publications to our credit, covering some graph-theoretical and statistical aspects of the social networks. The references are cited in different chapters of this book. We express our grateful thanks to Professor J. K. Ghosh, the world-famous statistician, currently professor emeritus at the Indian Statistical Institute and professor at Purdue University, who has been kind enough to encourage us in our venture over all these years.

We must also acknowledge the encouragement and support we have received from Professor Donald Von Eschen of McGill University, Canada, for this study.

SNA deals with the pattern of flow/articulation of a relationship such as friendship or help/cooperation among individuals or households (HHs)—that is, societal units whatsoever, in a reference population. Although SNA deals with individual units (such as HHs), unlike socioeconomic surveys on individual HHs, SNA dwells on a study of dyads formed by pairs of HHs from a population. SNA is concerned with a dyadic relationship, marked by the "direction of flow" (involving various features of the community such as help and cooperation). Hence, the proper setting for studying social networks is what is called a directed graph, abbreviated as *digraph*.

Chapters in the book are arranged as follows.

In Chapter 1, the concepts and a few basic parameters related to SNA are discussed at length, along with the highlights of their interpretations. A comparative study of the social networks of a village at two time periods is also presented.

Chapter 2 is devoted exclusively to an introduction to graphs. In this chapter, we present in a systematic manner all relevant basic concepts involving

both directed and undirected graphs and their extensions and generalizations. Also, various parameters of graphs are introduced, and their properties are discussed. This chapter forms a basis for exploring many workable graph-theoretic features of SNA. We have been able to concentrate only on a few of them in terms of proper exploration of their graphical and statistical aspects in this book. In Chapter 3, graph-theoretic (deterministic) and statistical (stochastic) models are introduced and discussed at length. This chapter constitutes the backbone of the book. The difficulty level in examining theoretical properties of many of the parameters related to the models is highlighted, and their extraction via simulation is indicated. The simulated distributions of various statistics under some of the probabilistic models are included to give the reader a good idea of what to expect from the models. Stochastic models for dyadic interactions are also discussed at the end of the chapter.

The validity of the statistical models introduced in Chapter 3 needs to be examined with reference to observed networks in terms of tests of goodness of fit and other valid statistical tests based on relevant data arising out of the networks.

Chapter 4 is devoted to an extensive discussion of model fitting and model validation. In Chapter 5, we discuss graph-theoretic and statistical measures at length. We broadly divide the measures in terms of (a) local measures of ego-centric (i.e., related to an individual actor as the unit) characteristics, (b) local-cum-global measures of egocentric and global characteristics, and (c) global characteristics. Collectively, reciprocity, cohesion (density), expansiveness (out-degrees), popularity and power (in-degrees), connectedness and fragmentation (strong and weak components), reachability, cliques, centrality and hierarchy, and other such measures are discussed at length. Several of these measures have been derived by the authors and are appearing for the first time in a book. Although detailed theoretical derivations/proofs are omitted, the measures are presented along with the underlying assumptions and can be used without any advanced knowledge of either graph theory or statistics.

Chapter 6 contains two case studies based on actual field investigation and extraction of real-life network data. In the first case study, we have compared longitudinal changes in social structure with reference to the networks of the same village (Kabilpur) as observed during 1971–1972 and also during 1997–1998. We do this by SNA, obtaining the values of graph-theoretic measures of some global characteristics of social networks at two time periods. The second case study examines cross-sectional variation in the pattern of circulation of social ties of

a community across villages. We have chosen the case of a minority religious group for this purpose in a few villages situated in the same region.

In addition, in Chapter 6, a section on the computation of a reciprocity measure based on a "weighted" digraph (i.e., weighted network) and its interpretation is also added. Last, fitting of some dyadic choice models to actual but small networks is also illustrated, and necessary interpretations are provided.

Finite population sampling and related inference is yet another very important topic. It has to be understood that for large populations, it is not at all an easy task to enumerate all the HHs and compile data on networks (with due attention to any nonresponse or any reporting errors whatsoever!).

Sampling of some of the HHs for collecting necessary data seems to be a viable alternative as it can be conducted competently and more cautiously to avoid any misreporting or nonresponse. By doing so, we do not make any attempt to create a prototype of a population network.

Creating a prototype is simply not possible. However, for some of the population parameters, we may attempt to provide "reasonably accurate" estimates based on a sample network, which is precisely the goal of a sampling expert. We present a reasonably complete account of this exercise in Chapter 7.

It is worthwhile to mention that Professor Vladimir Batagelj of the Department of Mathematics and Physics, University of Ljubljana, Slovenia, was amazed to know that the social network diagram of Kabilpur of 1970 was drawn by hand in 1972 by one of the authors (SB). He wondered how it was planned to handle that "enormous volume of ties." He commented highly about the diagram and said that Pajek came out much later.

The social network data of Kabilpur with brief notes of 1971–1972 and 1997–1998 are available on the following Web site: http://pajek.imfm.si/doku .php?id=topics:kabilpur.

This book is not a textbook or a handbook on SNA; rather, is expected to serve as a reference and text for demonstration of applying global SNA. A great deal of emphasis has been given to generate interest in this direction among students, teachers, and researchers in social sciences as well as for others who are interested in studying social networks. We believe this book has new material of interest to offer new ideas to graph theorists and statisticians equally, besides social science researchers. Last but not least, we derive much pleasure in thanking (1) an anonymous referee of the first draft of the manuscript for critically going through it and offering numerous constructive suggestions to bring our thoughts in proper perspectives; (2) the director of the

ISI for providing excellent research atmosphere throughout the tenure of this project; (3) the head of the Sociological Research Unit and head of the Stat-Math Unit, ISI, Kolkata, for providing adequate administrative support; (4) Dr. Prabir Ghosh Dastidar (a scientist in the Ministry of Earth Sciences, Government of India) for complying with our numerous requests for drawing the networks and the cover page of the book with extreme competence and courtesy; (5) Dr. Rabindranath Jana of ISI for complying with our requests, without losing patience, regarding technical assistance of various types as well as in keeping track of the manuscript pages while we were running back and forth among the many versions of the chapters; (6) Dr. Anil K. Choudhuri of the ISI, who provided us rich observations from the field on various angularities of articulating ties of help and cooperation among the villagers; (7) Mrs. Bandyopadhyay for hoping against hope that someday the project would eventually be over; (8) Professor S. B. Rao, our former colleague at ISI, Kolkata, and currently director of C R Rao Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS), Hyderabad, for giving us constant encouragement at very critical times, following the passing away of A. R. Rao; (9) Ms. Suchismita Roy for her assistance with our literature search; and (10) Ms. Lisa Cuevas and Ms. Vicki Night for reading our series of e-mails and responding to each one with extreme patience and encouragement. On behalf of Sage Publications, initially Ms. Cuevas and later Ms. Night did *not* lose hope on the book project, and here we deliver the manuscript, believing that we have accomplished the task to the satisfaction of all concerned.

—Suraj Bandyopadhyay
—Bikas K. Sinha
Indian Statistical Institute, Kolkata
August 28, 2009

TABLE OF CONTENTS

INTRODUCTION TO SOCIAL NETWORK ANALYSIS

―――●◆●―――

1.1 INTRODUCTION

The study of social networks is a new but quickly widening multidisciplinary area involving social, mathematical, statistical, and computer sciences (see Burt, Minor, & Associates, 1983, for application in diverse social environments; in the latter sciences, see Wassermann & Faust, 1994, and especially for the field of economics, see Dutta & Jackson, 2003). It has its own parameters and methodological tools. In this book, we intend to show how graph-theoretic and statistical techniques can be used to study some important parameters of global social networks and illustrate their use in social science studies with some examples of real-life survey data. We hope our illustrations will provide ideas to researchers in various other fields as well.

1.2 CONCEPT OF A SOCIAL NETWORK

The term *social network* refers to the articulation of a social relationship, ascribed or achieved, among individuals, families, households, villages, communities, regions, and so on. Each of them can play dual roles, acting both as a unit or node of a social network as well as a social actor (cf. Laumann

& Pappi, 1976). Kinship is a very common example of an ascribed relationship, while some common examples of an achieved relationship are those that are established in the course of regular interaction in the processes of daily life and living, cultural activities, and so on, such as one household requesting help, support, or advice from another; ties of friendship or choice of individuals to spend leisure time together; and preferences in marriage. Incidentally, a relationship can also be *negative*—for instance, hostility or conflict as opposed to friendship or alliance and alienation versus mutuality or integration. In this book, we will focus on *positive relationships*. Again, much of what we will discuss is based on sociological data, but it can also be used to study demographic and economic processes such as migration from one region to another, value of any type of economic (e.g., postal money order or trade) exchange between regions, volume of flow of goods between countries, flow of traffic between different places, and so on.

Thus, the units of a "social network" can be different, no doubt, such as individuals, families, households, and rural or urban areas, according to the relationship under consideration. But there is a common feature—namely, whatever the type of units we study, a specific dyadic relationship exists or does not exist between the members of any pair of them.

Furthermore, if the relationship exists between a pair of units, it is also quite pertinent to ask whether it flows in both directions or only in one direction and, in the latter case, from which direction to the other, because a social relationship is not necessarily symmetric. Asymmetric relations, such as the following examples, are as common as symmetric ones. For instance, A prefers B, A invites B to a household festival, or A goes to B for help or advice. But B may or may not prefer, invite, or approach A.

We should mention, however, that only because of the presence of such pairwise ties, a *social network* should not be equated with *social group*. There are two concepts of a social group: realist and nominalist. The realist concept is most commonly used in sociological parlance. According to this concept, it is an entity consisting of social actors such as individuals, families, and so on and is set apart from the rest. A social group retains a multidimensional system of somewhat durable contacts or interactions within the group: psychic, emotional, verbal, and behavioral. Thus, there is an element of a feeling of awareness or consciousness shared by its members. Besides, a social group generates its own boundary within which its members obey certain rules, norms, and functional roles toward each other as well as toward its common

goal. (For a detailed discussion of different characteristics of social groups, see Homans et al., 1968.) However, moving outside the realist concept of social group, a researcher also enjoys the option to impose his or her own definition of the boundary of group membership to identify a group for a study. This is the nominalist concept of a social group. For example, compare the Marxian concept of class as a "class for itself," a realist concept, and a class as "class in itself," the nominalist view (Laumann, Marsden, & Prensky, 1983). Wasserman and Faust (1994) have followed the nominalist concept of a social group for an illustration of methods. Thus, while a social group can be both realist and nominalist, a social network cannot be a realist one. A social network is a category of actors bound by a process of interaction among themselves. It is thus a nominalist category. However, a social network or its parts are endowed with the potential of being transformed into a social group in a realist sense provided that there is enough interaction.

For analytical purposes, a social network is conceptualized as a *digraph* (or a *graph* if the relationship has no direction). Digraph diagrams may be drawn to instantly provide direct mapping of ties showing their clustering as well as scatteredness. In a digraph, we call a unit—whether an individual, a family, a household, or a village—a *vertex* or *node*. A tie between two nodes indicates the presence of the relationship connecting them. Absence of a tie indicates absence of the relationship. A tie with a direction is called an *arc*, and a tie without direction is called an *edge*. One could also note the value or volume of flow as the weight of a tie and thus obtain a network that would then be a *weighted digraph*. More precise definitions of the graph-theoretic terms will be given in Chapter 2. Since the structure of the same network can be visually perceived differently depending on the manner in which a diagram is drawn, it is necessary to eliminate the bias in visual perception in order to draw an inference about the structure of a network from a digraph diagram (McGrath, Blythe, & Krackhardt, 1997). This visual bias is eliminated if we take recourse to numerically measure some of the selected important characteristics of a network and draw inference from there (see Chapter 6 for illustration).

For the sake of simplicity, we will concentrate on social networks showing only the presence (1) *or absence* (0) *of the relationship.* We also assume that ties have directions. Later, in Chapter 6, we will indicate, citing reciprocity as an illustration, how social network analysis can be extended to the case when the 0–1 restriction is dropped and there are nonnegative weights associated with the ties.

Networks are usually represented by diagrams where the vertices are represented by points, arcs by lines with arrowheads, and edges by lines without arrowheads. When two nodes are connected by ties in both directions, we often represent the two ties together by an edge, omitting the arrowheads.

1.3 SOCIAL NETWORK ANALYSIS

Search for a Theoretical Base in Sociological Theories of Generalized Social Exchange Behavior—A Brief Interlude Social network analysis (SNA) means analyzing various characteristics of the pattern of distribution of relational ties as mentioned above and drawing inferences about the network as a whole or about those belonging to it considered individually or in groups.

Beginning its journey as a descriptive metaphor, social network, in the course of the past few decades, has, as a parallel to the theories of market exchange, carved out a position for itself in the realm of theories and methodology for the study of society (Collins, 1988). Although its theoretical premise seems to be very close to market theories, it does not consist of looking for a best bargain in the case of an utilitarian exchange of goods and services. Rather, as a matter of generalized social exchange, it conceptualizes exchange not only in terms of economic interest but also of reciprocal role expectations as well as value orientations, social norms, and obligations. (See Homans, 1961, and Blau, 1964, for explication of the basic ideas relevant to understanding the rationale of the workings of social network, and see Turner, 1987, for a comprehensive discussion of social exchange and exchange network theories.) These attributes have made network theory more comprehensive and flexible enough to accommodate both asymmetric and symmetric relations as its natural elements. Hence, while social network theory does not deny the role of traditionally used a priori structural-functional concepts and categories in social research such as family, kinship, caste and ethnic groups, status groups, class, strata, and organization, it sees the actors and their roles and positions in a real-life situation rather in the light of the crystallization of patterns of interactions among individuals (Laumann, 1966; Wellman, 1988). This has also been discussed in detail in the context of Indian society (Srinivas & Beteille, 1964).

Social network theories do not consider individuals as forming a mechanical aggregate but as an organic whole where the constituent elements are connected among themselves as well as with the others through a mosaic of

ties based on interactions, directly or indirectly, at various domains such as social, economic, political, and the like. This enables a social network to be quite flexible to include the ties of the relationship of a social actor, which exist in ground reality even if those fall outside the boundaries of traditional social categories and derive appropriate ways to incorporate them in theoretical and methodological structure befitting the dynamics in social reality. Thus, SNA, unlike conventional social science methodologies, is rooted in the fact that the social universe does not consist of an aggregate of mutually independent social actors. On the contrary, they exist in a system of interlinkages and interdependence, creating and structuring ties among themselves. (Incidentally, Berkowitz, 1982, has encapsulated the concept of SNA in the wider perspective of structural analysis.)

The methodology of SNA has also contributed to the formulation of precise quantitative measures of many qualitative concepts that have long been in use in the study of society but have remained vague often due to the degree of separation between the concept and the measure (Adhikari, 1960). Power, cohesion, fragmentation, reciprocity, hierarchy, cliques, and alliances are some examples. Dissatisfaction with prevalent macro theories of society, such as those of structural functionalism, therefore has led to an alternative in social network theories, particularly with respect to studying the lack of social cohesion and conflicting situations, steep asymmetry, and fragmentation in society.

Furthermore, SNA also serves as a powerful tool for the identification of changes in a pattern of group structure, whether it is the case of data obtained on participation in a small group; survey research with large communities such as villages, towns, and so on; or flows of population, trade, traffic, and so on among different regions.

The mainstream of social network studies in the past has been the study of personal networks, even though the central thrust is oriented toward "looking for community," "to discover it," in various areas of life and living of individuals (Wellman, 1997). The aim of this book is to supplement this by showing how one may obtain the social network of a community by undertaking a survey, how one may derive measures of various sociological parameters from SNA, and how combined with contextual data it can provide deep insight into the changing pattern of a society, its dynamicity. We will illustrate these at the end of this chapter with a quick comparison of the social networks of a large village community before and after various official measures aimed at