Progress in Communication Sciences

Volume VI

1985

edited by

Brenda Dervin

University of Washington

Melvin J. Voigt

University of California, San Diego

ABLEX Publishing Corporation
Norwood, New Jersey 07648

---

4 An Overview of Network Analysis Methods and Programs*

Ronald E. Rice
Annenberg School of Communications
University of Southern California
Los Angeles, CA 90089-0281

William D. Richards, Jr.
Department of Communication
Simon Fraser University
Burnaby, British Columbia
Canada, V5A 1S6

---

*We gratefully acknowledge the helpful comments and criticisms of Drs. Russ Bernard, R. Vince Farace, Andy Hardy, Peter Killworth, Ed Parker, Lee Sailer, Barry Wellman, and Rolf Wigand. Dr. Peter Monge was especially helpful. Drs. Gil McCann and Steven Berkowitz provided commentary on blockmodelling, semi-groups, and entailment analysis. We also thank Dr. Eric Wendler and Bill Darst for their assistance. Errors in this text are the result of our ingenuity in introducing them after these people made their improvements. We do note, however, that references are current only as of mid-1983, when this manuscript was submitted.

Dervin B. Voigt M. (eds)

I. ASPECTS OF NETWORK ANALYSIS METHODS

The goal of network analysis is to obtain from low-level or raw relational data higher-level descriptions of the structure of a system. The higher-level descriptions identify various kinds of patterns, or test hypotheses about those patterns, in a set of relationships. These patterns will be based upon the way individuals or objects interrelate in a network, and, to some extent, upon the measurement tools and methods used.

Interest in human networks has flourished since the early work of sociometricians, small group researchers, anthropologists and theorists of social structure. (See, for example, Berkowitz, 1982; Berkowitz & Wellman, 1983; Boissevain, 1974; Boissevain & Mitchell, 1973; Burt & Minor, 1983; Coleman, 1958; Coleman et al., 1966; Galaskiewicz, 1979; Gottlieb, 1981; Knoke & Kuklinski, 1982; Leinhardt, 1977; Lindzey & Byrne, 1968; Marsden & Lin, 1982; Mitchell, 1969; Moreno, 1953, 1960; Rogers & Kincaid, 1981; Shaw, 1971.) An excellent history of early sociometry has been written by Nehnevajsa (1960). Lengthy bibliographies are provided by Feger et al. (1978), Freeman (1976), Klovdahl (1979a), Pitts (1979), and Wellman and Whitaker (1974). The range of these studies includes such substantive interests as corporate directorates, organizational communication, community helping networks, diffusion of innovations, social and kinship links, community power structures, and public service agency interaction.

The recent growth in network research has generated, and has been generated by, new developments in graphical, mathematical, algorithmic, and computerized techniques for processing and analyzing network data. Quite often, in spite of this large literature, network researchers use an approach or method based largely upon accessibility, cost, or the influence of the nearby implementor, and perhaps less on the conceptual foundations or practical validity of the program. This chapter reviews and examines network analysis methods and programs in the belief that the informed user can then make more careful decisions about which methods to apply. Only until network methods and their assumptions are tested, understood, and applied appropriately, will the "new communication paradigm of network analysis" (Rogers & Kincaid, 1981) begin to provide researchers with the insights and benefits possible from this perspective. The analysis of communication interactions has much to gain from network analysis properly applied.

Network analysis methods require data that describe a set of relationships among the members of a system. How these data are collected (diaries, survey questionnaires, rosters, monitored behavior, archival research, psychological tests, controlled experiments, etc.) and the form of these data (binary, scalar, continuous, frequencies, strengths, similarity scores, rankings, scales, evaluations of importance, etc.) are important topics, but not a primary focus of this chapter (except Section I,B,2). Related topics, such as the correspondence between recalled/reported and observed/monitored communication data, while crucial to understanding the validity of network analyses, are beyond the scope of this chapter, and, indeed, are still being argued in the literature (Bernard et al. 1982). It is not clear how important the type and level of measurement are to the final results of any network analysis, since asssmptions about the level of mesurement are seldom tested and often violated in practice.

Network analysis methods vary along a number of dimensions. The first dimension is the *conceptual* basis of the implementation or use of a particular network analysis method; that is, the extent to which the use of a method is motivated by theoretical and algorithmic implications of the method. The second dimension is the *operational* aspects of the particular method; that is, how the method works, its data requirements, its essential approach to handling the analysis. The third dimension concerns *pragmatic* aspects; that is, what the method produces, the level of correspondence between input and output, the levels of analysis. Our choice of dimensions is somewhat arbitrary, but we feel the three dimensions will both motivate a descriptive comparison of network analysis methods and clarify some of the ways in which conceptual perspectives interact with operations performed on network data for pragmatic purposes.

The format of this chapter is as follows. The three very broad dimensions upon which network analysis methods and programs may be said to vary will be introduced. Definitions of common (though not comprehensive) terms are provided under each dimension. The subsequent section then considers several general classes of methods: factor analysis, cluster analysis, multidimensional scaling, blockmodelling and graph-theoretical network methods, log-linear analysis. Each method is discussed in light of the three dimensions, and then one or two major implementations—programs—of the method are described. A subsection then briefly notes a variety of recently developed or otherwise hybrid methods, ending in a mention of some major integrated network analysis computer packages. A short analysis of a famous relational data set, using a variety of methods, is intended to highlight some of the insights that each method provides. A brief conclusion precedes a comprehensive bibliography which is intended to provide those interested in network analysis approaches with a starting place for further reading.

A. Conceptual Aspects

1. Approaches to the Problem

Two distinct approaches to the use of methods in network analysis can be noted. With the first approach, the development and use of methods are determined a priori by the theoretical guidance or needs of the researcher. With the second approach, there is less theoretical perspective to guide operational decisions; rather, the goals of analysis are defined a posteriori in terms of what the method produces.

Important differences between the two approaches can be illustrated by the ways in which the major analytic goals are achieved. Most of the programs examined below identify some kind of "clique." A wide variety of terms, including group, clique, clan, club, maximally complete subgraph, k-plex, catij, bloc, cluster, and semigroup, are used in the literature, and in a very general sense refer to some kind of grouping of nodes. The term "clique" was originally applied in sociometric studies where liking or disliking relations were of interest, but it has since been used in a broader sense. (It is the term that will be used in this chapter to include all of the above-named entities.) A general, early notion involved some type(s) of contraint(s) in the relationships within a group. Indeed, there are many senses in which the terms are used. Nonetheless, two distinctly different approaches to the definition of a clique can be identified:

(a) The definitions that stem from a theoretical basis can be considered to be a priori because they precede and determine the operations used in the identification of cliques. For example, the theoretic position of the investigator may lead to a definition of "clique" such as this: "a set of individuals that have most of their interactions with one another, rather than with individuals not in the set." Analytic procedures would then be designed to locate all sets of individuals that satisfy the terms of this definition. We do not here attempt to summarize (although Burt, 1980b, does) the various theoretic and statistical *models* of sociometric behavior (for example is transitivity, asymmetry, or balance a guiding principle? is triad nonrandomness predictable? is stratification different from clique-formation?), except insofar as a particular method or implementation is explicitly a priori. Few are.

(b) The definitions that stem from an operational basis can be considered to be a posteriori (often even ad hoc) because here the definition of "clique" is determined by the procedures used in the method that is selected. When factor analysis is performed on network data, for example, the analysis essentially begins with the definition of a mathematical procedure. "Cliques" are defined as "what the method produces"—factors. A correspondence between the investigator's conceptualization of the problem and the resulting factors that are obtained is less likely than in the a priori approach—unless the investigator's theoretic position accepts such a definition of cliques.

2. Representations of the Data

A crucial, even fundamental, conceptual aspect of network analysis concerns the nature of the relationship among the initial data, and, by extension, what the results are assumed to portray. At the most general level, two models are used in the representation of network data: *similarity* and *linkage*. The similarity model

involves "position" or "distance," while the linkage model involves "connected relations" or "cohesion" (Burt, 1976, 1978).[1]

Similarity methods use both the presence and absence of a link between two nodes (represented by a nonzero value—usually 1—or a 0, respectively, in the corresponding matrix cell) as data. Raw sociometric data may be transformed into a matrix of "similarities," "correlations," or "distances," which indicate *how members relate to all other members* of the system. The "distances" in this model are rarely actual physical distances, but rather are typically defined in terms of the mathematical operations used to compute them from the raw data. In general, two nodes will be said to be close or *similar* if their corresponding columns (alternatively, rows, and in some approaches, both) in the sociomatrix are highly correlated. Note that a preponderance of zeroes in two columns will indicate great similarity between the two nodes, perhaps a disadvantage when the sparce matrices such as those found in large systems are used. The similarity or distance values of the nodes can be interpreted to indicate the nodes' relative *positions* in the patterns of relations (regardless of whether or not they are linked to one another). This approach leads to analyses of social structure and role stratification. In the extreme interpretation of such distance, Burt (1980a) argues that "When multiple individuals have low distances between them relative to the distances among other actors in a system, the multiple actors jointly occupy a single ideographic position in the system."

The *linkage* approach uses the raw who-to-whom sociomatrix (with binary or scalar cell values), or in a binary form obtained by deleting links with low scalar values or nonreciprocated links (see Section I.B.2.), or by other transformations. Thus, a linkage matrix (sometimes called the *adjacency* matrix) may be either binary or scalar, with higher values referring to stronger or more frequent or more highly rated links. Linkage-based methods emphasize the analysis of connecting *links* or *link paths* between and among particular nodes. The linkage approach is most useful for analyses of relations, between and among specific nodes (individuals, groups, organizations).

Although analyses of small, moderately connected matrices may produce similar results from the two models, the differences are conceptually clear and empirically important. There are other differences between the two models, besides the way in which the final representation of the data is obtained:

(a) The similarity approach requires, at some point in the related algorithms, at least one half of a full symmetrical matrix (in general; but see Carroll & Arabie, 1980). There must be a value for every pair of elements in the system, and the distance from A to B must equal the distance from B to A. A

---

[1] Burt argues that a clique based upon *linkage* can be subsumed as a particular kind of structural *position*, and thus the positional approach is the more general of the two. Some of his recent papers (Burt 1980b,c; 1981) support this claim, for the types of status problems he studies.

linkage-based adjacency matrix, however, may be largely empty, signifying the absence of relationships between many pairs of nodes. Furthermore, relationships are not required to be symmetrical; it is possible for there to be a link from A to B when there is none from B to A.

(b) In a sense, the calculation of the *distance* matrix creates new information—information about pairs of elements not directly related to one another in the raw data.[2] However, it is in general impossible to reconstruct a unique adjacency matrix, given a distance matrix. This is true because, while it is possible to tell how *similar* the sets of links for a particular pair of nodes are, it is not possible to tell whether or not the nodes are linked to any *particular* node in a network. In fact, it is quite possible (in many cases substantively meaningful) for a pair of nodes to be very "close" to one another, or even be "structurally equivalent," and yet have no direct contact.

3. Type of Structural Description

Structural descriptions of networks can be classified into two categories: *discrete* or *continuous*. Discrete descriptions divide members of networks into mutually exclusive categories. Continuous descriptions locate nodes (or cliques) in a relative manner, either as they relate to all other individuals or as they relate to a set of "underlying dimensions." Group member, liaison, follower, and isolate are *discrete* role categories, for example. A *continuous* description is exemplified by factor loadings in factor analysis or multidimensional scaling coordinates which locate individuals in different forms of multidimensional "space." However, continuous descriptions are often transformed into discrete ones by means of a "cutoff" or "threshold" or "fit criterion."

B. Operational Aspects

1. Procedural Approaches for Clique Identification

As most network methods attempt to identify cliques, they must have procedures for doing so. These procedures can be sorted into five different categories, based on the main characteristics of the way they work.

(a) *Division.* The entire network is divided into (typically, two) parts. Each part is then divided into (two) more parts, and so on, until the desired "fineness" is reached, as established by a cutoff or threshold.

(b) *Agglomeration.* Cliques are started with a "seed"—a node which is chosen because of some characteristic such as a large number of links. Nodes are identified and added to the seed by a variety of criteria, until no more nodes can be found that fit the criteria. The result is a clique. Subsequent cliques are built from the remaining nodes.

(c) *Dimensioning.* Underlying dimensions, sometimes representing network attributes, are statistically constructed from relations of variance in the sociomatrix or its correlation matrix. Cliques are identified, by informal or formal rules, based upon relations between nodes and dimensions.

(d) *Searching.* Some methods use operational procedures which search heuristically or combinatorially through the network for patterns satisfying formally or informally defined configurations. Some graph-theoretical programs use search procedures to satisfy formal graph-theoretical definitions.

(e) *Overall pattern recognition.* This is the method used in the analysis of sociograms and in several matrix-based methods. The adjacency matrix (or a similarities matrix) may be reorganized in some way that allows cliques to be identified, either by inspection or by some process that is analogous to "looking" at the whole system. There may or may not be formal criteria for the constitution of cliques.

2. Treatment of Relationships

There are several characteristics of relationships that may or may not be considered in a given network analysis method.

(a) *Reciprocation/confirmation* and *directionality* of links. These are closely related concepts that must be understood separately if they are to be dealt with effectively.

The concept of *direction* may be irrelevant for a particular analysis, in which case the relationship is considered "symmetrical": a link from A to B is considered the same as one from B to A. It is more correct to say that A has a link *with* B; for example, that, A and B communicate with each other. Alternatively, the direction of relationships may be of great importance, such as when the relationship is something like "gets advice from" or "owes." Here, the relationship is called "asymmetrical," and a link from A to B is not considered the same as one from B to A.

In much of the social network literature, the term "reciprocation" is used to describe instances of mutual choice. If A says she talks *with* B, and B says he talks with A, the A–B link is "reciprocated." "Confirmation," on the other hand, refers to situations in which one person's report of a relationship is supported or "confirmed" by the other person's report. For example, C says he *gets advice from* D. D says she *gives advice to* C. If the relationship under investigation is *symmetrical*, reciprocation will be the same as confirmation. But if the relationship is *directed*, this will not be the case. The "advice" example described above shows confirmation of C's report of his link from D by D's report of the complementary action. For the C from D link to be reciprocated, D would have to have reported that she gets advice from C at the same time that C says he gets advice from D. If neither C nor D say they give advice to the other, then the C from D link is reciprocated by the D to C link, but neither would be confirmed.

While directionality refers to the way in which the relationship is concep-

---

[2] Inherent in this process are assumptions about the transitivity of relationships, which research tentatively supports (Holland & Leinhardt, 1973, 1976); but see Killworth (1974).

tualized, confirmation/reciprocation seems to be associated with the process of measurement, although one can argue it depends on the epistemological view of social reality that the researcher adopts.

In the case of undirected relationships, it would be expected that, given perfect measurement and perfect respondent memory, all links would be reciprocated. The sociomatrix would then be symmetrical about the main diagonal. If some links *are* unreciprocated, reciprocation could be forced by (a) either dropping all unreciprocated links, or by (b) adding the "missing halves" of the unreciprocated links. Of course, the adoption of either strategy makes simplifying assumptions about the nature of the measurement error, and either discards or creates information. Hubert and Baker (1979) offer a procedure for estimating a symmetrical proximity matrix from an asymmetrical one. Brown (1979) has developed a program which constructs missing sociomatrix choices (the row values of an absent or unsurveyed member) with Bayesian estimations based upon chooser's responses (the column values of surveyed members).

In the case of directed relationships, it is possible to measure the relationships from the perspective of either one or both of the individuals involved. That is, if the relationship is "owes," the data could come from either the perspective of the loaner or from that of the debtor. If A owes money to B, the data could either say "A reports owing money to B" or "B reports money being owed by A." Dual measurement (i.e., "who do you owe money to?"; "Who owes money to you?") is needed to establish reciprocation in this case, and the data from both sides must be merged in this analysis.

Some computer programs offer flexibility in handling directionality and confirmation/reciprocity. Others force the researchers to accept the methods' and implementers' assumptions. In general, the theory of directed relationships in network analysis is underdeveloped.

(b) *Binary vs. scalar link values*. In the simplest case, links (measures of relationships) may be considered only as "present" or "absent." In contrast to this simplifying binary approach, some methods allow (or assume) the strength or ranking of a link (i.e., frequency, duration, importance, etc.) to be expressed as a variable scalar quantity. The choice of binary or other values involves theories of measurement as well as substantive understanding of the system under analysis. Unfortunately, many network methods demand a particular form (often binary) which may not correspond to the conceptual foundations of the theory or social structure being explored.

(c) *Ability to handle multiplex relationships*. Relationships may be described in terms of their content or mode. For example, a pair of individuals may be connected by a friendship relationship, a communication relationship, or an influence relationship. Relationships may also be differentiated by the channel or medium of communication—two people may communicate by electronic mail, phone, letter, and/or interpersonal channels. When there is more than one type of connection between a pair of individuals the relationship is described as *multiplex*. Some network analysis programs are able to analyze multiplex relationships simultaneously, while others can only deal with one type of relationship in a given computer run; of course, results from separate runs can then be integrated in a single analysis.

(d) ***Restrictions on the number of links per node***. Some network analysis programs restrict the number of links each node may have to a constant number, such as 3, 5, or 7. The apparent simplifying advantage of limiting the number of links each node may report is offset by the high possibility of severe distortion of the potentially measurable social structure (such as deflated density or inadmissable patterns among sets of nodes). Of course, these and other restrictions are often imposed on the data at the time they are collected, creating a situation that leads to the same kinds of problems. (See Bernard & Killworth, 1973; Holland & Leinhardt, 1973; Richards, 1979).

C. Pragmatic Aspects

1. Correspondence Between Data and Results

With some methods, the final results are expressed directly in terms of the original data which represent the behavior of the nodes in the system. With other methods, the relationship between the original data and the results is not nearly so explicit. This is because the mathematical operations performed on the data produce more abstract descriptions which are related to the original in statistical ways. It will then generally be impossible to reconstruct the original data, given the final description. This may be a disadvantage if the investigator is interested in the analysis of the individual, initial relationships in the system. However, such a systemic or structural approach has very important applications in sociological theory.

2. Measures and Indices

Network analysis programs provide a variety of measures of individual, clique, and system network attributes. Barnes (1969), Davis (1977), Edwards and Monge (1977), Felling (1975), Freeman (1978, 1979), Hubert and Baker (1978), Jablin (1980), Katz and Powell (1960), Lin (1976), Moch et al. (1980), Mouton et al. (1960), Rogers and Kincaid (1981), and Tichy et al. (1979) provide more detailed reviews and analyses of network measures. Davis (1977), and Eisenberg et al. (1985) provide detailed empirical analyses of the relative importance of some network measures in predicting structure. For comparisons among different operationalizations of the "same" measure, and comparisons across measures for validity, see Freeman (1979), Freeman et al. (1980), and Friedkin (1981).

(a) *Clique identification*. The identification of cliques may be a principal goal of some, but not all, analyses. Some methods provide explicit identification of cliques and their members, while others do not. When the identification of cliques is not an explicit product of a method, it may be necessary to perform manual, graphical, or additional statistical operations on the results or to make subjective decisions to determine the cliques.

(b1) *Proximity measures: similarity methods.* All similarity methods begin by transforming the descriptions of relationships into matrices of distances or similarities. There are a number of ways of obtaining this information (correlations, scaling, structural equivalence). The exact meaning of the proximity/distance indices produced is determined by the specific computational operations used.

(b2) *Proximity measures: linkage methods.* Some linkage methods, while they do not provide measures of proximity *per se*, do, however, provide measures that are either similar to, or that may be interpreted as, measures of proximity. These measures include: the number of links in the shortest path from each member of a clique to each other clique member; "centrality"/ "peripherality"; "isolation"/"integration"; and overlap in personal networks.

(c) *Measures of connectivity (density) and connectedness.* The density or connectivity of a *clique* or a network is defined as the ratio between the observed number of links in the whole network and the maximum possible number of links for a clique or network of the same number of nodes. The connectedness of an individual is defined as the ratio between the number of actual links and the maximum number of possible links that a person could have. These measures are common in most linkage-based programs.

(d) *Measures of integrativeness.* The integrativeness of an individual is the extent to which the people linked with that individual are linked with one another, e.g., the connectivity of that individual's first-order zone. Group integrativeness (or openness) measures the extent to which clique members are linked with outside cliques (Freeman, 1980).

(e) *Measures of structure.* Structure is defined as organization or deviation from randomness or from some criterion. Some methods provide exact measures of the extent to which networks deviate from a given random model. The greater the deviation, the more highly ordered, stratified, differentiated, or organized is the network. The model itself may be based upon expected dyad- or triad-type frequency (Holland & Leinhardt, 1973; Killworth, 1974; Richards, 1974), clique and system entropy, the number of shared linkages, etc.

(f) *Reciprocation/confirmation.* Some methods provide information about the relative numbers of reciprocated and unreciprocated links in a network.

(g) *Measures of discrepancy in reciprocated links.* If an individual's description of network links is compared with the descriptions provided by the individuals he or she has links with to or from, it is possible to measure the extent to which the two descriptions agree.

(h) *Strength analyses.* Some methods compare and describe the strength of links with respect to other factors such as whether or not the link is reciprocated and whether or not the link is a within-clique link, a between-clique link, a link

with a liaison, or some other type of link. Distance between two nodes or positions may be listed and tested against a criterion.

(i) *Status.* There is a wide variety of status and influence measures, based upon network properties such as in-degree, ranking, structural equivalence, reachability, hierarchization, point bases and contrabases, and articulation-points.

(j) *Multiplexity.* Most methods cannot simultaneously consider several networks, although hand calculations are possible. Some methods may specifically consider and measure multiplexity, although the typical result is averaged across all networks. Some methods may provide measures of *diversity* or *range* in the number and kinds of relationships.

Section II describes six general classes of methods with respect to their conceptual, operational, and pragmatic aspects as outlined in this section.

II. MAJOR NETWORK ANALYSIS METHODS AND IMPLEMENTATIONS

In the discussion that follows, we examine six of the most widely-used methods in network analysis. Each of these methods is associated with particular computer programs: specific implementations of the methods. The purpose of the discussion here is to give the reader an understanding of the conceptual basis for the methods, how they work, what they produce, and how they compare to one another. When there is only a small number of program-implementations of any of the methods, some relevant characteristics of a representative program will be mentioned. Following the description of the different methods, some of the more important packages of analytic routines will be briefly noted.

The early methods of drawing sociograms and manipulating matrices were and are very useful for small-sized networks, and many of these matrix procedures are included in the more sophisticated programs described below. Unfortunately, these methods required many subjective choices and became intractable in the case of even moderate-sized networks. Lindzey and Byrne (1968) and Wigand (1977) review these early methods and their attributes. Only a brief summary of these early methods follows here.

Matrix manipulation (Beum & Brundage, 1950; Borgatta & Stolz, 1963; Coleman & MacRae, 1960; Forsyth & Katz, 1946; Jacobson & Seashore, 1951; Rattinger, 1973; Weiss & Jacobson, 1955; and a variety of articles in *Sociometry* in the early 1950s) is a linkage-based method which simultaneously permutes the rows and columns of a sociomatrix in order to cluster the nonzero entries about the diagonal. Groups are visible by pattern-recognition, as they cluster, and are defined (loosely) as members who interact more with each other than with other nodes.

Matrix multiplication (Festinger, 1949; Guimaraes, 1972; Hubbell, 1965;

Katz, 1947; Luce & Perry, 1949) is a binary linkage-based method which raises the sociomatrix to successively higher powers to identify the direct and indirect paths between nodes. Direct and indirect links are used in measures of system and individual structure. The concept of a clique, the number of cliques, and their membership are not explicit in the simple process of matrix multiplication.

A. Factor Analysis

(Bock & Husain, 1950, 1952; Bonacich, 1972a; Freeman, 1968; Holzinger & Harman, 1941; MacRae, 1960; Tryon & Bailey, 1970; Wright & Evitts, 1961.)

1. Conceptual

Factor analysis methods construct new dimensions corresponding to variance patterns, where successive dimensions or factors account for as much of the remaining variance as is possible, using linear combinations of the variables. Nodes "load" on each factor in proportion to the extent to which they vary in ways similar to the factor as a whole. Each dimension will therefore correspond to a set of individuals who load similarly on the factor. Factor analysis may be used on a who-to-whom matrix, a person-by-activity matrix, or a symmetric matrix of correlations representing similarities among the nodes with respect to other nodes or participation in activities.

The structural description provided by factor analysis is initially continuous. Factors can then be converted to discrete definitions of cliques by choosing a cutoff point to discriminate between members and nonmembers (nodes with high absolute loadings or nodes with low absolute loadings on given factors, respectively). As members may load on several factors, multiple clique membership as well as clique size may depend on the cutoff points used. In addition, the number of factors, and thus of cliques, is, to an extent, subjective. Because of high overall correlations, even a loading cutoff as high as .6 may be insufficient to differentiate between cliques (Killworth & Bernard, 1974). Furthermore, whereas a two-dimensional sociogram may be insufficient to portray subclique structure, factor analysis may provide the opposite difficulty of too much structure: Bock and Husain (1950) had to stop at 11 factors with only 67% of the variance explained in a network of directed links.

Because cliques are defined as factors, this is an a posteriori form of clique definition. (Although factor analysis may be used in a confirmatory manner to test the adequacy of a given clique analysis—see Burt & Bittner, 1981.) Cliques here will be composed of persons having similar patterns of connections to others in the network, regardless of whether or not they are connected to one another (as with blockmodelling—see section II.D.1.). The results produced by factor analysis are thus theoretically more similar to those produced by blockmodelling (see below) than to those produced by other methods.

In some cases, the factors can be interpreted in terms of some aspect of the relationship being studied. For example, using a nonsymmetric matrix, higher loadings on the "chosen" (as opposed to "chooser") factor may indicate that the nodes are leaders (MacRae, 1960). Such "chooser" and "chosen" structures have been interpreted as two batteries of variables (rows and columns) and analyzed by Beaton (1966) by inter-battery factor analysis.

2. Operational

Factor analysis uses a dimensioning approach to identifying cliques. The data for factor analysis may be binary or scalar. If the matrix is not symmetric, factor analysis will provide different results for row-factoring and column-factoring. Typical responses by researchers to this situation are (a) to be explicit about using only one direction (row or column) in factoring, (b) to make all links reciprocal, (c) to interpret both factorizations, or (d) to transpose the matrix and enter it along with the other before factorization. The second solution is the most common; we prefer the first or third solution.

3. Pragmatic

The final results of this method are statistical descriptions of abstract patterns of variance, rather than more concrete descriptions of the original behavior of the members of the system. Actual communication linkage and linkage patterns are not well represented by factor analysis, precisely because the factors are based upon a correlation matrix—similarity measures generated from initial sociometric data. However, Farace and Mabee (1980, p. 376) suggest a transformation of ratio-level symmetric links into similarities which maintain the linkage connotation. If factor scores are obtained, they may also be interpreted as a form of proximity measure: in this case, relating the individual to the clique, rather than to other individuals.

4. Implementations

Factor analysis programs are so widely available that it would add little to this discussion to describe them. Among the best-known and most-tested versions of factor analysis are those in SPSS and SAS. Most implementations are restricted to networks having fewer than 500 members.

B. Cluster Analysis

(Arabie, 1977; Arabie & Shepard, 1973; Coleman, 1970; Davis, 1967, 1970; Delattre & Hansen, 1980; Salzinger, 1982. For extensive reviews, see Bailey, 1974; Blashfield, 1976; Cormack, 1971; Hartigan, 1975; Jardine & Sibson, 1968; Krippendorf, 1980; Kuiper & Fisher, 1975; McPhee & Poole, 1979.) Due to the wide variety and common usage of cluster analysis for purposes other than network analysis, we refer the reader to the above review articles, especially Cormack (1971) and Hartigan (1975), for more detail than provided below.

1. Conceptual

Cluster analysis is basically a similarity method. "Distances" can be defined as (dis)similarities, closeness, etc. among network nodes. The definition of "clique" here will be a function of the procedures used to create the distance matrix, and of the cutoff points that are adopted. General usage of cluster analysis is a posteriori. For any chosen loading cutoff, clusters are discrete, but the choice of a cutoff point will determine how inclusive the clusters will be. (See Arabie, 1977; Arabie et al., 1978; and Carrington & Heil, 1981, on selecting this clustering cutoff level. Davies & Bouldin, 1979, suggest a cluster-separation measure.) These methods produce sets of clusters which are increasingly inclusive, and often hierarchical, although there is a school of thought which argues for the relevance of identifying overlapping social structures. Bonacich (1972a,b) describes an overlap measure which is independent of group size, and presents a related program, OVERLAP. Arabie (1977) and Arabie and Carroll (1983) review and discuss the issue of group overlap and present a clustering method designed to resolve a set of overlap problems.

The particular theoretical concerns of any single type of cluster analysis are usually evident from the operational approaches the method uses.[3] McPhee and Poole (1979) broadly classified these methods either as *difference* clustering, which uses distances between clusters as the basis for partitioning (and includes "single link" and "average link" approaches); or as *heterogeneity* clustering, which attempts to maximize within-cluster homogeneity (and includes "complete link" "centroid," and "minimum variance" approaches).[4]

2. Operational

The data for cluster analysis of an N × N sociomatrix usually consist of similarity values. A nobe-by-activity matrix will not be symmetric. There is no theoretical limit on the number of links each node may have. As noted, most clustering programs do not provide statistically-determined collapsing or cutoff levels. Some cluster methods also have difficulties with ties in link strength (Breiger et al., 1975).

---

[3] Note, however, that Davis (1967, 1970) proposes a specific sociometric structure model which allows for clusterings as found in these methods. He opposes Cartwright and Harary's (1970) more constrained model of structural balance which would imply fragmentation into only two cliques.

[4] Not mentioned by McPhee and Poole is a "min-max" iterative approach based on trajectory problems (Dunn, 1974). Another development is Delattre and Hansen's (1980) bicriterion clustering program, which, using a graph-theoretic algorithm, combines the difference and heterogeneity criteria. Their program presently can handle up to 400 nodes and will also indicate when the data have no significant structure.

3. Pragmatic

The presentation is statistical or graphical, rather than explicitly related to original member behavior. The results of clustering programs may be used with graphical programs (such as multidimensional scaling) to portray two- or three-dimensional clique structure: based upon which nodes are included in which clusters (at specific cutoff values), lines are drawn around those nodes in the graphical portrayal of multidimensional scaling output. Some clustering programs which produce spatial representations have implications consistent with spatial characteristics of social networks (Farace & Mabee, 1980). Baker (1974), Blashfield (1976), and Kuiper and Fisher (1975) have reported on comparisons of different clustering methods.

4. Two Implementations

(a) SOCK (Alba & Guttman, 1974) is a package of routines that works together with COMPLT, a graph-theoretical package (see section II.E.4.(b)). SOCK is a combination of a variety of distance and linkage programs for decomposing and clustering sociomatrices. The program begins with the calculation of a matrix of proximities from the original binary symmetricized adjacency matrix, which is supplied in list form. There are options for the proximity metric that is used, including various measures (for dyads) of the overlapping of social circles or zones, as well as some of the more conventional (correlational) programs. The user may choose from several clustering programs. Two distance forms are supplied—one is a version of Johnson's (1967) hierarchical clustering, and the other is a form of nonmetric multidimensional scaling, in which the clusters are visually identified by the user. Some implementations of SOCK produce sociogram plots; however, the computer-drawn linkages of dense cliques are generally unreadable (see Tichy et al., 1979). The programs provide a number of indicators of proximity, as well as several indices of the goodness-of-fit of the various solutions to the analysis and additional statistical measures about the network's structuring. (See also STRUCTURE, under blockmodelling programs.)

(b) MAPCLUS (Arabie & Carroll, 1980, 1983) allows overlapping clusters which need not be maximal complete subgraphs (see section II.E.1), allows the user to specify the number of resultant subsets, and provides a measure of goodness-of-fit (variance explained) of the solution. This program may thus reveal negative as well as positive-affect groupings, and alternative clustering solutions with similar explained variance. Current size limitations on the input matrix is 30 × 30.

C. Multidimensional Scaling

(Bloombaum, 1970; Carroll, 1973; Carroll & Arabie, 1980; Gillham & Woelfel, 1977; Goldstein et al., 1966; Green & Rao, 1972; Guttman, 1968; Kruskal, 1969; Kruskal & Wish, 1978; Lingoes, 1973; Norton, 1980; Shepard, 1972; Woelfel & Danes, 1980.)

1. Conceptual

Carroll and Arabie (1980), Green and Rao (1972), and Kruskal and Wish (1978) provide the most comprehensive, state-of-the-art reviews and critiques of multidimensional scaling. Arabie and Carroll (1980, 1983) present detailed algorithms for MAPCLUS, the latest and most efficient in a developing series of scaling programs. Due to the explosion in the growth of MDS implementations and applications, we provide only a brief discussion of MDS here and refer the reader to these articles.

In general, multidimensional scaling methods are somewhat similar to factor analysis methods. MDS methods usually begin with matrices of response similarities or proximities rather than of sociometric choices or correlations, although some MDS programs provide considerable pre-processing and transformation routines to transform nonproximity data into proximity data. MDS is thus a *distance-based* method, and provides a posteriori clique-detection.

The goal of these methods is to construct an n-dimensional space and locate the objects (here, nodes) in the space so that (a) the number of dimensions in the space is minimized and (b) the relationships among the proximities of the nodes in the space are in agreement with those among the distances between the objects in the (socio)matrix. The latter goal is accomplished either by maintenance of the *order* in the relationships (nonmetric methods) or by maintenance of the *magnitude* as well as order (metric methods). Kruskal and Wish (1978) claim that two dimensions can contain about as much information about clique structure as can hierarchical clustering of the same data; Barnett and Woelfel (1979) critique the various techniques used to determine sufficient dimensionality and suggest alternatives.

The output of these methods is a set of coordinates and spatial representations of the proximities in a set of n dimensions—usually representing sets of two or three dimensions for paper plotting. As MDS methods do not identify cliques, cluster analysis methods may be applied to the results to give a description of network structure. An MDS "clique" is therefore dependent on the particular clustering method used, when one as used.

Because the data take the form of proximities, the dimensional axes will represent abstract statistical characteristics of cliques, rather than actual behavior patterns (Lankford, 1974). These characteristics may or may not be related to patterns in the behavior in ways that make them useful and interpretable. Indeed, interpretation is a prime task of a researcher using MDS methods,[5] particularly because the axes in many programs are rigidly rotatable from the original (arbitrary) coordinates, for *any* given set of dimensions. And, the choice of the number of dimensions to use is often subjective, or based upon selected cutoffs (Kruskal & Wish, 1978, p. 48; Barnett & Woelfel, 1979).

MDS generally emphasizes larger distances over smaller distances, thus minimizing the portrayal of associations among clique members and of large variance in proximity values. The situation of isolates is particularly problematic as they cannot be adequately represented in any graphical form of dimensional space. The result is often distorted space as portrayed in the printout—particularly for nonmetric programs. One solution is to invert the distance either by taking their reciprocal, or by subtracting each distance from the maximum distance. This procedure locates isolates near the center of the MDS map. However, as Burt (1976) shows, transformations of the data from linkage to similarity form (his terms are relational and positional) may avoid this problem.

Clique presentation is thus continuous in many ways—the final solution depends on many choices, and, even then, cliques are not portrayed discretely. However, the ability to explore and solve for dimensionalities may provide analytical insight and intermediate solutions to network analyses.

2. Operational

Metric MDS requires high measurement reliability. Nonmetric MDS assumes that the data support Euclidean distances, so data which may be non-Euclidean are considered unreliable and are transformed to approximate metric values. Using nonmetric MDS on non-Euclidean but highly reliable data thus actually introduces error into the final results (Woelfel & Danes, 1980). The types of data acceptable for various MDS programs are described by Carroll and Arabie (1980).

3. Pragmatic

Original network behavior cannot be reconstructed from MDS results, although links from the original data may be drawn in (visually recreating physical network aspects) on the spatial distribution of nodes produced by MDS. In such drawings, long lines indicate large residuals in the statistical solution (the fit between proximities and dimensional distances), rather than long network linkages. If there are a number of such long lines, the clustering is likely to be invalid (Kruskal & Wish, 1978, p. 46). Suggestions for graphic display of configurations in more than two dimensions are provided by Friedheim (1982).

A measure of stress or alienation (the discrepancy between the proximities and the dimensional configurations resulting from the rotated spaces in MDS) is usually provided. The measure may incorporate modifications and weightings to handle different data requirements. Unfortunately, a best stress value for a particular computer run may be only a local optimum, and not the actual global optimum. In addition, stress values are very sensitive to the ratio of objects to n-dimensionality (as is factor analysis to the variable-to-subjects ratio), and to

---

[5] As well, indeed, with factor analytic and all network methods.

violations of assumptions such as no missing proximities, few ties, and Euclidean distances (violations which are generally likely in sociometric data).

4. Implementations

One nonmetric implementation discussed here is Samllest Space Analysis (Guttman, 1968; McFarland & Brown, 1973); the metric implementation included here is GALILEO (Woelfel & Danes, 1980). See Carroll and Arabie (1980) and Kruskal and Wish (1978, p. 79) for a list of other MDS programs.

(a) Nonmetric Implementation: SSA

When Smallest Space Analysis is used for network analysis, it usually begins with a proximity matrix, where the *i,j* entry is the number of steps in the shortest path from node *i* to node *j*. A correlation matrix may also be entered. However, SSA assigns ranks to the distances, and analyzes the ranks (Norton, 1980).

Because of its usual dependence on the discrete path-length form of the distance matrix (derived from the adjacency matrix), SSA might be considered a linkage-based method. However, subsequent steps in the analysis and interpretation of the results of SSA clearly fit the distance model better than the linkage model.

The nonmetric goal of SSA is to construct an *n*-dimensional space in which the order relations—rankings—between original data distances are preserved in a spatial model with a low number of dimensions. Cliques are identified by simple visual inspection, or by results from cluster analysis.

SSA emphasizes small distances rather than the large distances usually emphasized by other MDS methods (Kruskal & Wish, 1978).

Network data for SSA are typically binary and symmetric, which SSA converts into path-length rankings. There are no necessary restrictions on the number of contacts each node may have, although high linkage densities across the network may render the method ineffective, in much the same way that datasets with very low variance are not good candidates for fruitful factor analysis. For this reason, investigators using this method often restrict the number of contacts per node to three or five. Since this restriction is done for operational convenience rather than for any theoretical or conceptual reasons, some researchers would be reluctant to use this method, especially when it is known that the number of links per node is higher than five. However, Norton (1980) argues that incomplete rankings can still provide useful insights into data.

The relation between the original communication behavior and the results produced by SSA is statistical, and thus SSA is subject to the same considerations as factor analysis. SSA is unique, however, in its use of the distance matrix as the initial distance representation of the data. This approach leads to a variety of problems, including:

(1) Because the distance matrix is constructed from a (typically binary) adjacency matrix, the addition or deletion of a single link may radically alter the path-lengths and thus the final solution. This is a very real problem, given the relatively crude measurement techniques and potentially unreliable data in sociometric research.

(2) The path-length form of the distance matrix is problematic for other, more fundamental, reasons. Although this form might seem acceptable, given that SSA only attempts to preserve order relationships, this is not the case, because it is not at all clear that the order relationships among the path lengths computed from a symmetricized binary adjacency matrix are meaningful. For example, because link strengths are ignored, it is possible for two nodes connected by a short chain of intermediaries to be operationally much "farther apart" than a pair of nodes connected by a longer chain of intermediaries. In addition, the number of paths from one node to another is ignored in the construction of the distance matrix. Thus, the "distance" from A to B in Figure 1 will be the same as that from B to C, in spite of the obvious differences.

The proximity measure provided by SSA is a monotonic transformation of the path-length form of the distance matrix for the network. It is thus at best an ordinal measure. Stress and goodness-of-fit values, which may be partially improved by additional iterations or by using more dimensions, are also produced by SSA. The Lingoes-Guttman programs for SSA (Lingoes, 1973) are widely available and documented.

(b) Metric Implementation: GALILEO

GALILEO assumes that sociometric data may contain non-Euclidian distance relations which violate the basic triangle inequality assumption of metric scaling methods. GALILEO instead assumes Reimann space, and indeed offers a measure of the discrepancy of the data space from Euclidean space (and thus a surrogate measure of intransitivity). By including the (mathematically) imaginary part of complex numerical representations necessary to perform the scaling, and by providing a theoretical interpretation of it, this program attempts to preserve and extract more of the information that is contained in the data (and that can be reliably reported by respondents) than do other MDS programs. Woelfel and Danes (1980) discuss the theoretical foundations of GALILEO.

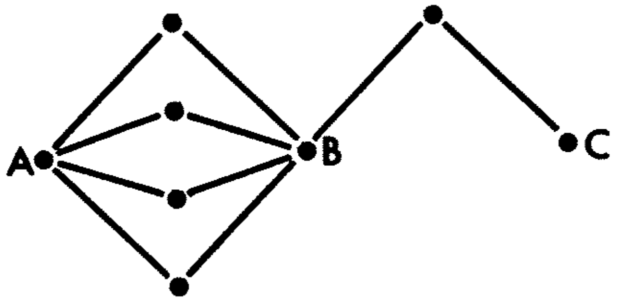

FIG. 1. Number of paths compared to path distance.

Coupled with the provision for non-Euclidean distances, GALILEO's measurement approach provides constant scaling, reliable measurement, and faithful maintenance of dimensional proportions. Rotation of axes is required, but leaves the distances invariant, and useful initial solutions provide a reference frame to which later or other sociomatrices and network structures may be compared. Additionally, GALILEO allows chosen relationships to be held constant while other relationships or attributes (such as time) are varied by rotation. GALILEO explicitly allows analysis of different perspectives in the network, by allowing translations of the origin (such as from the perspective of a leader of a high-status clique rather than from leader of low-status clique). Like INDSCAL (Carroll & Arabie, 1980), this program allows three-way dimensional analysis—for exam- ple, with time intervals as the third dimension. This facility has been used to good purpose by Barnett and Palmer (1983) and Rice and Barnett (1985) for networks involving 14 to 25 time periods. (See Rice, 1981, for a review of longitudinal network analysis and a lengthy bibliography.) GALILEO can analyze up to 45 nodes.

D. Blockmodelling

(Arabie et al., 1978; Breiger, 1976a,b; Breiger et al., 1975; Heil & White, 1976; White, 1974a,b; White & Breiger, 1975; other references in *Connections*, 1978, 2,1)

1. Conceptual

The excellent review and critique by Arabie et al. (1978) and Arabie and Boorman (1982) provide the best introductions to blockmodelling, including the conceptual, mathematical, and application precursors from as far back as 1899. Social structure here is conceived as a network of roles and relations; certain relations are inherent in the roles (Leinhardt, 1977, Introduction). Relations are of (more or less) structurally equivalent sets of expectations and of interactions (Lorrain & White, 1971). Because blockmodels attempt to indicate structurally equivalent positions, not relations, "there is no implication that members of a block cooperate or coordinate with one another. In fact, the individuals in a block need not be connected at all to one another." Thus, "it is clear that blocks need not be cliques in the standard graph-theoretical sense or any of its many so- ciometric generalizations" (Breiger et al., 1975, p. 10). Thus, blockmodelling was initially based on an algebraic theory of categorizing the members of a system on the basis of similar interaction patterns, with the *same* others, rather than on any notions of interconnection. Related to Sailer's criticisms (Sailer, 1978; White & Sailer, 1979) is a more general concept (role equivalence or general structural equivalence) which involves nodes or sets of relations which are connected in the same ways to *structurally similar* (not necessarily the same) others. As Carrington and Heil (1981) write, "The problem in blockmodelling is

to find a mapping of the many data nodes into a few model nodes such that the ties among the data nodes can be satisfactorily represented by the ties among the model notes."

A *block* is both a set of network nodes having similar (structurally equivalent) relations to nodes in other blocks, and an area in a matrix (a submatrix) identified with those nodes. As the results of this method are formally akin to those of factor analysis, trait and activity data may be more appropriate than strict sociometric data (Farace & Mabee, 1980). Indeed, the goal of analyses which consider structural equivalence is to identify, describe, and test roles and categories of roles—not cliques—which are different (e.g., a family is a clique, a mother is a role).

Block structures may remain quite stable, even when individual memberships change, because roles, hierarchies, and patterns of relations in society tend to remain relatively constant (Breiger & Pattison, 1978). The definition and usage of "blocks" in the literature qualifies blockmodelling as an a priori meth- od. Blockmodelling methods may be either similarity-based or linkage-based initially, but result in similarity analyses.

2. Operational

Conceptually, the data for blockmodelling may be either directed or undirected, and there is no restriction on the number of contacts per node, but the input sociomatrix to the most widely used blocking algorithm must be in binary form. Blocking outcomes are reported to be robust over different dichotomizations of ordered sociometric data (Breiger et al., 1975). The method is unique in its ability to handle multiplex relationships simultaneously, although typically the several matrices are averaged and the distinctions between them are lost early in the process (Farace & Mabee, 1980). Input matrices may be "weighted" by repeating them, but this is an inexact and cumbersome approach.

In the more widely used approaches, the initial input matrix of relations is transformed to a correlation, or similarity, matrix. The next step is to iteratively create a new correlation matrix from the prior correlation matrix, in order to produce a matrix of entries equalling approximately either +1.0 or −1.0. This matrix of distance is then clustered into subsets. This initial division or clustering may be repeated until each node is partitioned into its own "block." Thus, this clustering stage of blockmodelling is a divisive hierarchical clustering method.

There are two approaches leading to a final portrayal of the network as a set of blocks. The first is based upon the clustering results directly. The level of clustering—assigning nodes to a given number of blocks—may be chosen on substantive grounds (depending on how many positions or cliques are of interest) or on the basis of the initial density of links within and across blocks. There are three kinds of blocks: those having only zeroes as entries (there are no ties among members of this block), those having some ones (there are some ties among members of this block), or those having only ones. Submatrices with sufficiently high final densities (see "fit" criterion below) are represented in the final blockmodel as cells with a "1" in them, and the other submatrices are represented as blocks containing zeroes. This final blockmodel is called an "image" of the original network. For example, if ten nodes are clustered into two blocks, the iamge may be portrayed as a 2 × 2 matrix. If, after the fit criterion is applied, blocks (1,2) and (2,2) contain ones and the other cells contain zeroes, the nodes grouped together in block 2 interact with each other in the same or similar ways, but do not interact (at the criterion level) with the nodes in block 1. Conversely, the nodes in block 1 interact with the nodes in block 2 in the same or similar ways, but not with each other. This image would look like this:

$$\begin{array}{cc} 0 & 1 \\ 0 & 1 \end{array}$$

Blocking directly from the clustering results requires substantial judgement in the choice of the final number of blocks. This choice depends on the "fit" criterion: the most rigorous fit requires strict structural equivalance of nodes with a block (i.e., link density in a zero block is exactly zero); a more relaxed and more realistic fit requires that the nodes within a block have structural equivalence greater than some specified (*alpha*) density of 1's in the one-block. For example, *alpha* could be the grand mean of all block densities in the network.

The second approach subjects the clustering results to a blocking procedure which treats these groupings as an hypothesis to be tested, again subject to a "fit" criterion (see II.D.4.(a)).

3. Pragmatic

The results of blockmodelling are at an intermediate level of abstraction somewhere between the highly statistical representations produced by factor analysis and the very concrete representations produced by graph-theoretic methods. However, many of the indices and results of Burt's continuous measurement approach to blockmodelling (see section II.D.4.(c)) can be statistically tested (e.g., the difference between structural positions). Graphical representations of results portray the blockmodels and hierarchical clusterings.

4. Implementations

(a) CONCOR (Breiger et al., 1975) is a metric hierarchical clustering routine, based on an iterative correlational procedure which successively partitions the population at any alpha-level into dichotomous subsets of prior subsets. The initial relational matrix may be asymmetric, with valued links. CONCOR provides a hierarchical clustering table based on the final correlation matrix, and a blockmodel based on the alpha value chosen for each level of clustering.

Sailer and D. White (and Sim & Schwartz, 1983) claim that CONCOR'S theoretical definition of structural equivalance is actually different from its algorithmic definition, and that CONCOR is thus misleading. In addition, Sailer states that CONCOR results are identical when 0's and 1's in the data are interchanged, contrary to the implications of CONCOR's most rigorous fit criterion. Reitz (1983) and others question the limitations of performing blockmodelling on columns only, which CONCOR uses, when structural equivalence implies receipt as well as transmission. Further, Sim and Schwartz (1983) tested CONCOR results against simulated but known structures (families, neighbors and pals, and a factory) and conclude that blockmodel partitions produced by CONCOR (for these small and idealized networks) do not uniformly recover social positions—positions as compared to structural components such as coalitions, groups, etc.

Schwartz (1976) claims that CONCOR and related blocking programs which rely on iterative correlation analysis provide no more information than does principal components analysis of the original matrix (because he argues that the diagonal values should be ignored), and may provide fewer insights as it throws away information about relatedness after the first correlation transformation. Arabie et al. (1978) argue that Schwartz's criticisms may not always apply. Schwartz (1976) and Freeman (1977) consider CONCOR to be a special case of principal components analysis, while Carrington and Heil (1981) note that CONCOR's iterations are more time-consuming than those of factor analysis. Schwartz shows how principal components includes isolates in one factor, whereas CONCOR distributes them among several blocks. Nor does CONCOR find leaders, who, although they are *individuals*, still constitute social *roles* (Reitz, 1983). In addition, there are problems when one or more nodes choose, or are chosen by, all others—because there is no variance in the column or row of such nodes, which causes difficulties with the calculations. When multiple networks are used, different connectedness levels in each matrix will differentially weight the networks in the overall analysis.

BLOCKER (Breiger et al., 1975; White et al., 1976) is the original companion program to CONCOR. BLOCKER uses only binary symmetrized inks as input and tests if CONCOR's partitions satisfy the hypothesized blocking under a given fit criterion. It assigns any nodes to blocks in a way that can satisfy the hypothesis.

(b) COBLOC (Carrington & Heil, 1981) and its goodness-of-fit measure are both improvements, corrections and advances upon CONCOR (Sailer, 1978). COBLOC optimizes a goodness-of-fit measure at each successive clustering level, and allows backtracking to other clustering levels. (See also Baker & Hubert, 1981; Bielby, 1981; and Carrington et al., 1980, for a discussion of fit measures.) Thus, "local" maxima of the fit measure can be recognized, and different or better sequences of clusterings can be interactively produced. Like Burt's and Sailer's approaches, COBLOC defines structural equivalence as "being related to structurally equivalent" (but not necessarily the same) nodes.

As an aside concerning goodness-of-fit, Panning (1982) offers a comprehensive approach to this problem; he offers a regression model approach to blockmodelling, based upon attribute data. His algorithm seeks to identify blockmodels with maximal fit to the data. The procedure, unlike CONCOR, is not limited to hierarchical solutions, and operates simultaneously on both columns and rows. It also provides measures of cluster properties and distances between clusters, identifies nodes most typical of their respective clusters, and provides standardized distance measures which can be entered into MDS for scaling and graphic portrayal. The author does not discuss the problem of nonindependence among nodes and among clusters that may lead to biased estimates of the indices and measures described.

(c) STRUCTURE (Burt, 1980c; Project, 1977) uses both clustering and blocking methods to categorize clique membership and node status, and analyze role relations, by a version of structural equivalence. Burt defines structural equivalence as the "proportion of variance in distances to the actors that is described by the first principal component of the variance-covariance matrix for distances to the actors." Alternatively, individuals *i* and *j* are structurally equivalent if the sum of Euclidean distances from each of their relations (in one or multiple networks) are equal. Then *i* and *j* denote a joint position for which structural proximities can replace the original separate Euclidean distances. Thus, node distances are not identical to CONCOR's iterative correlations, so blockings from STRUCTURE will not necessarily be the same.

Operationally, this difference is because STRUCTURE's continuous distance approach not only takes into account row-relations as well as column-relations, but preserves distance relations rather than only providing discrete blockings. Thus, blocks may be ordered in their relations (Knoke & Kuklinski, 1982). Also, STRUCTURE aggregates nodes, instead of partitioning them.

The initial matrix, as with CONCOR, may be scalar and asymmetric. A clustering cutoff is used to determine equivalence, although extensions to the program test statistical criteria for locating clique "indicator" roles. The program provides up to 11 network indices for each node, path distances and role distances among nodes, and derived measures for use in subsequent analyses. It is limited to 150 nodes in up to nine networks, and cannot handle missing data. STRUCTURE, like BLOCKER, requires the user to specify the number and membership of blocks, based upon the initial clustering results, as hypotheses. Options allow for four different a priori conceptualizations of the equilibrium nature of the social system (Burt, 1979).

E. Graph-Theoretical Methods

(Alba, 1973; Alba & Gutman, 1974; Alt & Schofield, 1975; Augustson & Minker, 1970; Burton, 1970; Cartwright & Harary, 1956; Farace & Mabee, 1980; Felling & van der Weegen, 1976; Gengerelli, 1963; Gleason, 1971; Harary and Ross, 1957; Harary et al., 1965; Luce, 1950; Luce & Perry, 1949; Purdy, 1973.)

1. Theoretical

Early approaches to communication network analysis were based in graph theoretic methods used by Bavelas and first defined by Luce and Perry (1949). They required "maximal complete subgraphs," which are subsets of the whole network that are both (a) completely interconnected and (b) not contained in a larger maximal complete subgraph. Mokken (1979) differentiates three less stringent operationalizations, or types of "*n*-cliques." For example, what he calls a "*n*-clique L of a graph G" is a *maximal subgraph* of G where the shortest path in G between L members (*i,j*) is less than or equal to *n*. This subgraph definition involves external reachability, and allows for subgraph disconnectability. For large graphs, Arabie and Carroll (1983) argue that there are far too many complete subgraphs to be useful in clique detection. Alba and Moore have recently (1983) recanted their support for the use of the *n*-clique (particularly in large networks) because of this prevalence of complete graphs.

Luce's early model did not allow discrete partitioning of cliques and determination of their membership, although Harary and Ross (1957) did establish a criterion for distinguishing cliques (Farace & Mabee, 1980). Harary et al. (1965) further explicated graph theory for social scientists and advanced the method to consider system-environment relationships, roles and measures (Cartwright & Harary, 1970). For applications of graph theory to clustering, see Hubert (1974). Q-analysis and hypergraphs are the most recent extensions of graph theory as applied to network analysis (Atkin, 1974; Doreian, 1981; Freeman, 1980; Gould, 1980; Seidman, 1981; Seidman & Foster, 1978).

Graph-theoretic programs can operate either on the raw binary sociomatrix or on a dichotomized version of a distance matrix constructed from the sociomatrix. In the first case, they will be linkage-based, while in the second they will be similarity-based.

2. Operational

Most graph-theoretic methods use agglomerative procedures for clique-detection. They build cliques by searching combinatorially for all sets of nodes that satisfy the criteria for either maximally complete subgraphs or for *n*-cliques (depending on the method and the decisions made by the user). (But see SONET, section II.E.4.(h).) Subgraphs which are unsuitable are rejected, and, when it is appropriate, subgraphs are combined or condensed into large subgraphs. The result will be a specification of the subsets in the system that meet the criteria of the method. These will usually be sets of highly interconnected network elements.

3. Pragmatic

Depending on whether the starting point is the raw sociomatrix or a dichotomized distance or similarities matrix, the results may be either very concrete or expressed in more abstract terms that must be related to the particular method of calculating distances or similarities. There may be a high degree of user control over every step of the analysis, and a large variety of indices may be provided as part of the output.

These indices may include measures of the statistical significance of the result, in the sense that they indicate the probability of getting a configuration similar to the one obtained under particular conditions. Measures are usually based upon separate or summed node linkages, although they may be used with structural definitions to measure system and environment properties (Cartwright & Harary, 1970). Because searching is done combinatorially, graph theoretic programs are generally restricted to systems having no more than a few hundred members.

4. Implementations

(a) NEGOPY (Farace et al., 1977; Jablin, 1980; Rice, 1978, 1979a,b; Richards, 1974, 1975, 1976, 1977; Richards & Lindzey, 1974; Richards & Rice, 1981; Rogers & Agarwala-Rogers, 1976; Rogers & Kincaid, 1981; Weimann, 1982; Wigand, 1977, 1979, 1982).

(1) Conceptual

The goals of this linkage-based pattern-recognition program are to classify, according to explicit definitional criteria, the nodes of the network into several role categories, and to provide individual-, group-, and system-level descriptions of sociometric structure. Structural description with this method is, therefore, discrete.

Role categories include the following: The *group*, which consists of a set of at least three individuals who have more than half of their interaction with other members of the same set, all of which are connected by some path, lying entirely within the group, to each of the other members of the group. There must be no node or link which, if removed, causes any of the conditions not to be met. *Liaisons* are individuals who have most of their interaction with members of groups or with other liaisons, but not with the members of any single group—thus, connecting groups directly or through other liaisons. *Bridges* are group members linked to other groups directly. *Isolates* can have (a) no links, (b) one link to another isolate, or (c) one link to a tree node, group member, or liaison, while not being a member of any other role. A *tree node* is a branching node connected on one end to a tree node or one other role, and on the other end to a tree node or an isolate. These roles reveal the conceptual basis of NEGOPY: primarily communication patterns whereby group members interact more with each other than with out-of-group nodes. Because the analytic procedures are based on a set of conceptual definitions and devised specifically to detect whether nodes satisfy these definitions or not, NEGOPY is an a priori method.

Classification criteria are based upon the amount of interaction (strength, frequency, evaluation, etc.), rather than only on the presence/absence or even number of links.[6] Ratio-level measurement is assumed, although binary data may be used when "amount" is conceptualized as "presence" by the researcher. Reitz (1983) points out that isolates *are* classified solely on the basis of links present or absent, however, because no matter how strong an isolate of type (b) may be linked to the network, it remains an isolate. Further, he notes that there is no *guarantee* of a solution, or of a unique solution. On the other hand, Barnett (1973) did show that NEGOPY does not "find" differentiated results when simulated, random data are used.

(2) Operational

Several approaches (matrix manipulation and multiplication, iterative searches and tests, graph-theory, and heuristics such as suggested by Weiss & Jacobson, 1955) are involved in NEGOPY.

Pattern-recognition routines are combined with logical test routines to determine whether definitional criteria are satisfied. For example, tentative role categorizations are found by an iterative vector-averaging process, in which the links for a node are combined to give a weighted mean value assigned to that node for use in subsequent iterations. First-step link strengths and the number of *n*th-step links supporting the direct links may be combined in the weighting calculations. The results of this iterative process can be transformed into either (a) a virtual representation of the original sociomatrix, in which the rows and columns have been permuted so as to concentrate the nonzero entries about the major diagonal; or (b) something like a closeness continuum, in which nodes are located closer to the nodes they are connected with than to the ones they are not connected with. Pattern-recognition routines scan this continuum to obtain a tentative description of system structure, which is then subjected to testing and modification to meet the definitional criteria.

Reciprocation/confirmation and directionality of links are considered explicitly. Multiplex relationships are not considered explicitly within any one analysis; but links from two or more networks may be added together in one analysis by means of additional data lines or by multiple weightings within one data set. There are no restrictions on the number of links per node other than those due to computer memory and peripheral constraints.

(3) Pragmatic

The final results of this method are expressed directly in terms of the original data from the system members. Nonoverlapping cliques and roles are explicitly identified. Linkage-based measures are provided at the individual, group and system levels.

---

[6] Moch et al. (1980) found increased system differentiation when link weightings, rather than binary values, were used.

NEGOPY can handle networks with up to 4,096 nodes and up to 80,000 links, due to its use of linked-list data structures, rather than full matrix representations of the data. A variety of network descriptions at different levels of analysis is provided as described by Richards and Rice (1981). A set of plotter routines that produces graphical representation of all or part of the network is also available (Lesniak et al., 1977).

The characteristics of NEGOPY and its sensitivity to variations in both the data and the options chosen by the user (who may impose or modify assumptions made available by the program) have been tested and described (Rice, 1978, 1979a,b). These options include a strength formula which allows algebraic and geometric transformation of one or two weighting values for each link, the use of directed links, and different step-weighting functions.

(b) COMPLT (Alba, 1972, 1973; Alba & Guttman, 1974) is part of a comprehensive package of network analysis program (see SOCK, section II.B.4.(a)). COMPLT is a linkage-based method that begins with the identification of all *n*-cliques (a relaxation of the maximal complete subgraph) that meet specific criteria. These are then subjected to a number of tests, in which cliques that are too weakly connected, too small, etc., are removed. Overlapping cliques that meet a specified criterion will be combined.

This program accepts only binary symmetrical data. Strengths of links are not considered, unless a preprocessing step is added to dichotomizing scalar-valued links. The program will not handle multiplex relationships. The number of contacts per node must be a constant number across the network, but the value of this number is not limited by the program itself. The requirement for symmetrical data interacts with the one for a constant number of contacts per node to create some interesting problems. If the measurement technique allows unreciprocated or unconfirmed links, links will need to be either added or deleted when the data matrix is symmetricized. But this will change the number of links for some nodes, which causes problems for the second requirement.

(c) NCLIQUE (Felling, 1975; Felling & van der Weegen, 1976) is actually a suite of subroutines which provide a variety of measures, processed matrices, and clique-detection approaches. Input may be dichotomized, univalent or multivalent, directed or undirected matrices; parameters allow internal preprocessing of any of these matrices. Subroutine SOCMAT computes a wide variety of matrices, in- and out-measures, components and sets of points, and measures of points, status, and centrality. Subroutine HIMAAT computes several hierarchization-measures of a graph. Subroutine STATUS computes "point centralities" based upon distances, statuses, and contrastatuses based upon distances and upon direct and *n*-sequences. Subroutine HUBBEL detects cliques as defined by Hubbell (1965), and Subroutine N-CLIQUE detects all weak and strong *n*-cliques.

(d) Augustson and Minker (1970) describe an iterative program that generates graph-theoretic cliques. This program allows overlapping cliques, something that is uncommon for graph-theoretic approaches. (Indeed, although Luce's and Alba's early formulations did not prohibit overlapping, neither implementer felt that clique overlap was desireable—see Arabie, 1977). Bron and Kerbosch's algorithm (1973) apparently is a faster version.

(e) Doreian's (1969) algorithm for clique-detection allows valued rather than strictly binary matrices. Peay's (1974) revision of the algorithm (UPWARD/DOWNWARD) allows overlapping hierarchical clusters. Peay's (1980) general model for valued networks includes formalizations of some central network concepts. His model allows for qualitative and quantitative structures and values.

(f) SOCPAC-I decomposes a graph into triads and then tests each type of triad to see whether it is more or less frequently observed than would be expected according to a particular model (Holland & Leinhardt, 1973, 1976). See Killworth and Bernard (1979) for use of this program in analyzing measurement problems. Sodeur's (1979) NETZ data management program also performs triad census analysis. Although the analysis of triad counts represents a strong tradition in the study of networks, it appears to have little relevance to communication network analysis. See Fombrun (1982) for possible applications in the study of organizational communication.

(g) Kushner and DeMaio (1977) describe an application of graph theory to analyze voter behavior and note their program to calculate the relational measure of expected value, in terms of being on the winning side of a vote, of *i*'s agreeing with *j*'s vote.

(h) SONET (Seidman, 1978) includes a variety of clique identification procedures based on different proximity measures derived from graph theory and matrix manipulation. This program allows overlapping networks. SONET handles large (1500) matrices, and can include binary attribute data to be analyzed with network measures. Particular networks can be modelled on the basis of attribute data, and compared to actual or other modelled networks.

SONET includes hyper-graph analysis and extensions of Q-analysis. Hypergraph analysis operates at the level of structural relationships. In particular, it identifies and emphasizes overlapping subsets of a graph, such as events connected by people who co-attend *some* but not *all* of those events. Linkage-based analysis at the dyadic level would likely obscure certain subsets of importance. Indeed, nodes are viewed here as a list of events they attend, and may be grouped in such subsets. One such type of subset is the "pseudo-event"—an overlap of events in which each pair of nodes attends some, but not necessarily the same, events. In a political context, a pseudo-event is a (small) set of community activities with a considerable likelihood of political activity, due to the coattendance by community elites. Pseudo-events requiring a small number of nodes to represent all events in the sample are more likely to be sites of political activity than those requiring more nodes.

Everett (1982b) extends the applications of Seidman and Foster's hypergraph approach in his graph-theoretic blocking program called EBLOC. Using symmetric binary matrices, the algorithm "*k*-blocks" nodes which occur on cycles of *k* or less. Because Everett is theoretically interested in cliques which include feedback loops, cycles are seen as significant insofar as they provide opportunities for feedback, as are highly proximal links. Blocks, then, are built up from cycles via shared links and shared nodes, as in Q-analysis. The advantage of EBLOC over CONCOR is that "floaters" and "hangers-on," including bridges *between* blocks, are readily identified rather than obscured in a common block. The user may select cutoff levels for link strengths and cycle length. Reitz (1983) praises this algorithm, but points out that its main utility comes from multiple runs at different settings, and it is unable to handle dense networks.

F. Log-linear Analysis

1. Conceptual

Log-linear analysis is a method increasingly applied to the study of networks. Its fundamental utility and initial motivation—the analysis of multiway contingency tables (Fienberg, 1979)—seems particularly appropriate for sociometric data matrices. The essential approach is to postulate a model of how a given contingency table (network) is structured—that is, which information flows or relational patterns are significant aspects of the network. This model is represented by constraints on the row, column, and grand totals (marginals) of the input matrices. Although the development of this method was not directed toward network analysis, the theoretical and operational as interpreted by network analysts attributes qualify log-linear analysis an an a priori method.

2. Operational

The raw input data can take a variety of forms. For example, in analysing reciprocity of dyadic relations over time, a dyad may be linked or not in each of several different time periods. The dichotomous categories of relations at each time period are pooled in a single transition matrix, and estimates of the log-odds relation of outcome conditions (linked or not), given initial conditions (linked or not), may be used to support or reject hypotheses of conditional reciprocity over time (Fienberg & Wasserman, 1981; Wasserman, 1980a,b.). Alternatively, the rows and columns can represent categories or groups within a larger network; then models of cross-group and within-group relations may be tested (see Marsden, 1981; Rice, 1982). In each case, the data are typically row- and column-standardized by iterative proportional fitting of the data to the modelled marginals.

3. Pragmatic

The fit of the data to a given model is measured by a goodness-of-fit index, whose distribution is the chi-square. (This assumes the data are multinomially distributed, which may or may not be true for most network data; however, this approach overcomes the usual problems of nonindependence between and within rows and columns of network data that plague other tests.) In addition, subsets of the most general model can be tested for significant improvements in the initial model. Other outputs include parameter estimates for "attractiveness," "expansiveness," reciprocity, within-group relations, and similarity to other within-group relations, depending on the nature of the network data. Combinations of such measures can be used to role-categorize nodes and groups (see Marsden, 1981; Rice, 1982). Similarities between sets of nodes may be tested in some programs by methods similar to analysis-of-variance in terms of "blocking" categories together in a given constrained model.

4. Implementations

(a) Fienberg and Wasserman (1981) provide algorithmic introductions to the application of log-linear analysis to network data, while Reitz (1982, 1983) makes this introduction more accessible. In particular, he shows how the BMDP statistical package can yield network parameter estimates. His ITR program (Reitz, 1983) doubly standardizes a sociomatrix in order to integrate the interactions of the entire matrix into each node-specific cell value. However, ITR requires at least one in-link and one out-link for each node. Input data may be an adjacency matrix or a list of links for each node. These links may be scalar, perhaps representing the number of shared multiplex relations. When each row's doubly-standardized cell values are then divided by the total number of nodes, the resultant value from 0 to 1 represents the probability of a link from *i* to *j*, given *N* nodes and the total linkage within the network. A related clustering program uses these average expected linkage probabilities as logical cutoffs in deciding upon clustering results.

(b) Marsden's (1981, 1983) GENMRG3D program fits log-linear models of which information attributes characterize a multigroup system to an I × I × 2 (outflow by inflow by internal group communication flows for a system with I groups) contingency table with varying constraints in the fitting of the three-variable interaction term. This program allows for the detection and testing of group attributes: inflow, outflow, internal relations, and selected cross-group relations. It has several distinct advantages over other network analysis methods: (1) Unlike network role characterizations as described by Burt (1978), Harary et al. (1965), or Richards and Rice (1981), Marsden's method allows for role distinctions based on all three dimensions of information flow. (2) Participation levels are defined probabilistically, relative to average information flow levels, and thus take into account the communication volume and size of both the system and its constituent groups. This facilitates comparisons across different networks or time periods. (3) The program can take into consideration the total amount of links possible (i.e., two matrices are usually entered as data: one of actual communication links, and one of linkages absent; this last matrix is the difference between potential linkage amount within and across groups, and actual linkage amount). (4) Perhaps the method's most rigorous attribute is that roles defined are subject to statistical testing. That is, roles imposed by the method can be tested against the data for their adequacy rather than accepted arbitrarily. This ability is particularly necessary in light of Burt and Bittner's (1981) article, which emphasizes (1) how few studies have tested the fit of the roles analyzed, and (2) how analyses can be misleading when, with complex data, several alternative role categorizations are possible but not tested against each other.

G. Other Methods, Programs, and Developments

The reader might wish to refer to Sonquist (1980, Table 1) for a list of some computer programs with respect to network analysis techniques they perform (such as clique plotting, triad census, and centrality measures).

1. MEA

A method has been developed by Douglas White (White & McCann, 1983) which can be used to examine statistically the set-subset relations among nodes in a network. It is called *material entailment analysis* (MEA).[7]

(a) Conceptual

This technique was originally developed to examine the pattern of *entailments* in sets of attributes using *material* (empirical) data; that is, statements of the form "If a person or object has attribute *X*, then that person also has attribute *Y*." If we think of a network, however, as made up of sets of people (or other objects) who are linked with sets of other people, then we can examine the statement: "If a person has a link with a person in group *X*, then that person also has a link with a person in group *Y*." If *X* and *Y* represent *sets* of people, there is no necessary implication that a specific person in *X* is linked with a specific person in *Y*. Furthermore, the method is designed to examine chains of such if-then connections (entailments), testing for transitivity. Thus, the method is designed to detect logical or set relations among the data objects. As such, it is an a priori method. Further, the method has a statistical component, because with real data there may be a few "exceptions" to the set inclusion relations, which we wish to attribute to "error" or random fluctuations.

(b) Operational

MEA requires data in the form of a node-by-node matrix, which need not be square. The data may either be binary or "fuzzy," a probability between 0 and 1. The program treats the *columns* as variables and computes all pairwise correlations (using phi). Positive correlatons reflect set inclusion and negative

correlations represent set exclusion ("If *X*, then not *Y*"). The pairs are then ordered by "strength": those with the fewest "exceptions" are ranked highest, and within each exception level, those with the highest correlations are ranked highest. A Monte Carlo technique is used to eliminate relations that are likely by chance. The remaining pairs are "potential" entailments. Starting with the highest-ranked pair, pairs are included in the final structure as long as they do not violate the transitivity criterion, ie., if *X* entails *Y* and *Y* entails *Z* (*X* and *Y* are positively correlated with few exceptions, and likewise for *Y* and *Z*), both entailments are kept only if *X* entails *Z* within the imposed criteria levels. The user may specify the amount of exception constituting the criterion level. As a further criterion, in a transitive entailment (*A–B; B–C;* therefore *A–C*), the partial correlation *AC.B* must be greater than zero.

(c) Pragmatic

The final structure is a series of connected or unconnected "chains" of entailments of the form *X* entails *Y* entails *Z* entails *V* . . . reflecting data in which those nodes connected to *X* are also connected to *Y* and to *Z* and to *V*, while at each step the sets increase in size. Based on analysis of attribute data, the method appears to produce results compatible with hierarchical clustering, but with more detail: each of the chains fall within a cluster, revealing structure *within* clusters, and connections between chains show connections *among* clusters (see White & McCann, 1983, for an example).

(d) Implementation

The method is implemented through a program, ENTAIL, written for DEC computers in relatively straightforward FORTRAN. The program can be adapted for IBM machines without too much difficulty.

2. Semigroups and Other Algebraic Approaches

Among the most abstract and high-level approaches to social structure are the algebraic (Berkowitz, 1982, pp. 91–103, 133–141). The primary thrust of this attack has come from the examination of kinship structures (Boyd, 1969; White, 1963), as well as some analysis of organizations (Bonacich, 1978), and most recently it has been used to extend blockmodeling and related graph-analytic approaches (Bonacich, 1980; Bonacich & McConaghy, 1980; Boorman & White, 1976; Lorrain & White, 1971; McConaghy, 1981; Sailer, 1978; White, 1981; White & Reitz, 1983). We present only brief comments on this very recent extension of network analysis.

(a) Theoretical

The most general (i.e., least restrictive) and most generally used algebraic structure is a *semigroup*, which is a set (a collection of any kind of object, such as people, corporations, nodes of a graph, or even other sets, such as relations between nodes) with an *associative binary operation* defined on it, such as addition, the compounding of relations or blockmodel images, or the union or intersection of sets. The concatenation or composition of relations is a "natural"

---

[7] Drs. Gil McCann and Stephen Berkowitz kindly provided the basis for this description of material entailment analysis, and of semi-groups in the following section.

operation for social scientists: for example, the relation "mother of" can be composed (followed, juxtaposed, concatenated, compounded) with "brother of" to produce "uncle of" (mother's brother), or "squared" (followed by itself) to produce "grandmother" (mother's mother). Furthermore, since relations are important for mathematics as well as social science, semigroups of the mathematical theory of relations are well-studied.

The means for connecting or "reducing" one algebraic structure or model to another is the *homomorphism*—a "many-to-one" mapping of elements in one set to elements in another which *preserves structure*. There are homomorphisms which apply to semigroups, to groups, and to graphs. The idea is to find a semigroup or other algebraic structure which represents or models the underlying data. Different structures found in different networks, societies, etc., can then be compared in the hopes of relating the differences to other social phenomena.

There seem to be three major areas of algebraic analyses in network studies: (1) kinship, utilizing mathematical group theory and the theory of semigroups; (2) extensions of graph theory utilizing lattices or other algebraic structures; and (3) extensions of blockmodeling and alternatives to blockmodeling to get at the notion of a social "role" based on types of relationships or connections (Bonacich, 1980; Boorman & White, 1976; Everett, 1982a; Sailer, 1978; White, 1981). This third approach extends blockmodelling (itself a method for partitioning graphs by means of homomorphisms) by looking at the image blocks as relations, and manipulating the semigroups which result from compounding these relations. The basic thrust of this work is to model the theoretical notion of a social *role*; i.e., to partition the data into sets of people or other objects which play the same role in the structure (Boorman & White, 1976). At a higher level of analysis, two structures can be reduced to a common structure through the use of appropriate homomorphisms to examine closeness of structure or elements of structural similarity (Bonacich, 1980; Bonacich & McConaghy, 1980; Boorman & White, 1976). Issues still being debated center around the definition of *role*, and the appropriate semigroup or homomorphism to use.

(b) Operational

The algebraic approaches, using some form of homomorphism, have been applied at all levels of data from raw linkage data (graphs or matrices), to aggregates of data (blockmodels) to abstract relations (kin terminology). Small semigroups can be calculated by hand and their structure presented in a "multiplication table" (Berkowitz, 1982, pp. 91–103; Boorman & White, 1976). Mathematical theory can be used to describe many semigroups arising from composition of relations.

The data may take the form of a conventional matrix of binary links (of one or more types) or some more abstract form as in kinship studies. Homomorphisms are numerous, and the investigator must choose one with some set of theoretical or mathematical criteria in mind. One can also rely on a given algorithm or program to find the appropriate one (such as a blocking algorithm for raw data).

(c) Pragmatic

The results of these algebraic approaches are presented at a very high level. Typically, the result is that a particular system is found to be modelled by a mathematical object (such as a semigroup or a semilattice. We may find, as did Boorman and White (1976), that two relations or two networks produce the *same* abstract structure (e.g., the same semigroup). Ideally, one of the results is to find what roles exist or what role-structure is present, that is, the investigator finds a partition of the objects of interest (people, etc.) so those within each "group" (those who get mapped together) play the same role(s). Unfortunately, few studies have been done on empirical data to see if we can discern roles that are recognizable. Further, most of the mathematical objects produced, such as a particular semigroup, do not, at the moment, have intuitive interpretations. However, if we note that blockmodelling involves homomorphic reductions of relations in raw data, then blockmodel studies, many of which are cited above, can be considered applications of (semigroup) algebra.

(d) Implementations

There seem to be a few scattered programs around. White and his associates (Boorman & White, 1976) have GENSG to generate semigroups and JNTHOM, which finds a joint homomorphic reduction to two semigroups to determine structure they have in common. Bonacich (1980) provides an algorithm for his "common structure semigroup." Doug White at University of California, Irvine, has programs to find various kinds of "equivalences" (i.e., roles—partitions of nodes or individuals who are linked to similar others). However, much of the work in this area appears to have been done by hand.

3. CATIJ

CATIJ, formerly named KBPAK, (Killworth & Bernard, 1974) uses nonsymmetric sociometric rank data. If *i* ranks *j* *n*th, the initial distance matrix contains (*n*-1) in cell (*i,j*). Then a minimal path is found for each "level," or increasing number of intermediaries between *i* and *j*. Finally, a category matrix is constructed which contains in each cell a value equal to one plus the number of intermediaries used to connect *i* and *j* through the shortest path. The rows are then factored to produce clusters, using a loading cutoff of .6. Rows are used to control for problems associated with nonsquare matrices, and to emphasize the *receipt* of linkages. This approach allows the identification of idiosyncratic, perhaps internally unlinked, and single-node, cliques, particularly as represented in respondents's cognitive maps, because subjects provide the *rankings* of others. Results from CATIJ may be used to assign nodes to discrete roles quite similar to those of NEGOPY. However, clique members are assumed to view their relations to the larger network similarly. The hierarchies of steps in networks are assumed to indicate social structuring. Multiple content relations or network matrices may be compared. The authors argue that their program is quite robust to error in the data. The development and use of CATIJ qualifies it as an a priori method.

4.

If the notion of a network has any substantive justification, then one can think of an "effect" of the dependencies among the nodes on a dependent variable. Several approaches to detecting this effect have appeared recently, all classified as *spatial* (or *network*) *autocorrelation* methods.

These methods have arisen initially as solutions to problems of spatial influence, such as how the diffusion of tribal languages influenced the role and status of women in a particular African village. The solutions differ primarily in two ways: (a) by the algorithmic strategy used to derive parameter estimates and significance tests, and (b) whether the "effect" is located in the error term or is tested as a separate effect coefficient. The cost of not testing for, or not identifying, the network effect, if it exists, is to overestimate the significance of some or all of the independent variables.

Doreian (1980, 1982) uses a maximum likelihood (MLE) approach which unfortunately requires an expensive and lengthy search for parameter estimates, and thus limits network size to less than 90. However, he discusses the theoretical and substantive issues in, and provides algorithms for, locating the effect in either or both the disturbance and coefficient form. Dow et al. (1982) and White et al. (1981) test only for the disturbance effect, but provide a straightforward BASIC program which is inexpensive. They also compare the various methods across extant and simulated data. Except for large networks, for which the MLE approach is as yet too expensive, the method these authors describe produces results equal or superior to those of MLE. The basic conclusion of the spatial autocorrelation approach is that ordinary least squares regression is *not* sufficient when a network effect is significantly present. Rogers and Kincaid (1981), Burt (1981), and many others would argue that this is the predominant situation.

5.

McQuitty's procedure (1957) for nonoverlapping clustering is a distance-based agglomerative program which, however, uses linkage information; starting with reciprocal pairs having highest similarity, it sequentially adds unreciprocated relations to find cliques defined a posteriori.

6.

Hubbell (1965) developed an Input/Output program to raise matrices to the *p*th power, sum all *p* matrices, and use cutoffs determined by percentiles of total linkages. Matrix cells also indicate *j*'s contribution to *i*'s status, and the dynamics of this status can be evaluated.

7.

Gleason and Cartwright (1967) and Hansen and Delattre (1978) describe algorithms that use the mathematical theory of map coloring to determine the "color-ability"— or something like clique distinguishability—of a network.

8.

Phillips and Conviser (1972) describe the mathematics of a program based on information theory. This program derives clusters, using nominal data, based on the "minimization of entropy." That is, the increase in uncertainty caused by mislocating boundaries is measured and boundaries are the iteratively located so as to minimize the measure of uncertainty. Structure is defined as the interdependence of parts, the ability to predict the communication behavior of *j* (the dependent variable), given knowledge of *i* (the independent variable). This type of method tends to produce equal-sized clusters, according to the authors. See also Krippendorf (1981), Horan (1975), and Shaw (1983) for the use of information-theoretic measures in network analysis.

9. MACRONET

MACRONET (Klovdahl, 1979b) includes a variety of matrix manipulations and extensions and allow tests of stochastic matrix series.

10. SPIN

SPIN (Curnow et al. 1976) is an interactive program which accepts interaction matrices, directed graphs, and causal-loop diagrams, and provides feedback loop analysis, pulse analysis, matrix powering, and clustering for structural modelling.

11. Network Packages

Flexible, integrated program packages such as N-CLIQUE, SOCK/COMPLT, and STRUCTURE really indicate the orientation of future network analysis systems—packages offering a variety of methods, measures, and options. In fact, several such larger packages or libraries are also already in existence: Berkowitz et al. (1979) document a large number of batch programs developed to analyze corporate information; and Anthonisse and Lagewey (1980) describe their GRAPHLIB, a very comprehensive library of programs. It seems clear that, in the U.S., the greatest progress is being made at the University of California at Irvine, where an extensive set of routines called NETWORKER is being implemented as part of the project in Structural Analysis. The Structural Analysis project at Irvine is currently implementing and comparing many different network programs; most have been implemented on personal computers.

Perhaps the most integrated and comprehensive network analysis package other than NETWORKER is GRADAP (Stokman, 1983; Stokman & van Veen, 1981). Graph theory is the foundation of the package, and the documentation serves as a fine primer in theory and algorithms. GRADAP provides most graph theory analyses and measures, while allowing numerous manipulations, extensions, aggregations, and subsetting of the initial data. A particular advance in data storage is GRADAP's indexing and retrieval of both node-oriented and link-oriented information. This facility enables analysis of overlapping subsets. Analysis routines are available for simple graphs, digraphs, and valued graphs; graph sampling; testing of triad censuses; identification of equivalence classes; simulated graphs or networks; and plotting. GRADAP uses SPSS-style statement language and task sequencing, allows storage and later manipulation of interim analyses, and interfaces with SPSS for further analyses. The 1983 version handles 6000 nodes, but is compatible only with CDC computers.[8]

One day soon, a truly accessible and validated Network Package for the Social Sciences (NPSS) will enable network researcher's to rise above the peculiarities of locatable programs and facilitate the use of theoretical and conceptual considerations in deciding which programs to use. New programs are generally noted in the International Mathematics Science Library listings or in journals such as *Behavioral Science*; the *Newsletter of the Special Interest Group on Social Science Computing* of the Association for Computing Machinery; *Connections*, the bulletin for the International Network for Social Network Analysis; and *Social Networks*.

III. AN EXAMPLE FROM A MONASTERY

It might be helpful to show how several of the methods described in this chapter can be used to indicate relational structure. This section is not intended to provide a detailed theoretical, analytical, or algorithmic comparison among methods and programs; our only intent is to portray some of the forms of output and some of the insights that one might gain from network analysis.

We use a quite famous and well-analyzed set of relational data, that of 18 novices in a monastery (Sampson, 1969). We borrow data, descriptions and results from Arabie and Carroll (1983), Breiger et al. (1975), Reitz (1983) and White et al. (1976). Sampson collected data from the novices as they recalled their relations at several time periods. He asked the novices to rank the top three and bottom three monks on four relations measured on a +3 to −3 scale (constituting like/antagonism, esteem/disesteem, influence/negative influence, and praise/blame). Because of tension among contemporary and liberal stances toward Catholicism after Vatican II in the mid-1960s, and because of the recent arrival of more highly educated and liberal novices, several alliances or groupings became evident to Sampson, who provided extensive qualitative data in addition to the relational data. The earlier novices were labelled the Loyal Opposition (LO); the new novices were the Young Turks (YT); there were three unaffiliated Outcasts (O); and the text indicates three Waivers (W) who in other analyses and interpretations are classified at YTs and LOs. The group classification and node numbers are presented in Table 1.

---

[8] An IBM-compatible version should be available as of this chapter's publication date. Write: Computer Center of the University of Nijmegen, 6252, Ed Nijmegen, Denmark.

TABLE 1. Classification of Novices from Sampson Monastery Data

Group	Abbreviation	Novices: Data Identification Number and Role
Young Turks	(YT)	1 (L), 2 (L), 7 (F), 12 (L), 14 (F), 15 (F)
Loyal Opposition	(LO)	4 (L), 5 (F), 6 (L), 9 (L), 11 (L)
Outcasts	(O)	3, 17, 18
Waiverers	(W)	8 (or LOL), 10 (LOF), 13 (O, LOF)

NOTE: L = leader; F = follower; LOL = loyal opposition leader; LOF = loyal opposition follower. The waiverers were assigned to loyal opposition by most analysts, although Reitz (1982) makes explicit some ambivalence by Sampson (1969) as to their alliances. Reitz also indicates that some of the LOL's were more followers than leaders.

The fourth time interval the novices were asked to recall was just before the expulsion of novice 2 and the Outcasts by the monastery superiors for not acting in conformity with the Order's spirit; specifically, monk 2 was expelled for independence and arrogance, and the others for immaturity. Shortly after the explusion, the Waiverers and all of the Young Turks except novice 12 left voluntarily. How have network analysis methods and interpretation shed light on this relational structure?

Breiger et al. analysed a correlation matrix based upon the four +3 to −3 matrices, entered as one 72 × 18 matrix. The divisions that CONCOR produced are shown in Table 2 (a). The first partition discriminates between the LO/W group and the YT/O group; the second partition splits the Os from the YTs, but further splits the LOs, and not clearly along the W/LO line. The authors used Johnson's (1967) hierarchical clustering program, HCLUS, to find quite similar results with the diameter method (except that 13 is included with the outcasts), while the connectedness method (shown in Table 2 (b)) is similar at the second division but diverges rapidly from there. Knoke and Kuklinski (1982), along with other network analysts, prefer the diameter (complete link) method in general, because it does not create "chains" of clusters as does the connectedness (single link) method. White et al. (1976) used eight separate matrices (0 to +3 for each of two poles of the four relations), and CONCOR produced the same partitioning as the diameter clustering method. Everett (1982a) used 12 various settings of his EBLOC program on these data, to suggest a summary clustering of the YTs, the Outcasts, the LOs with Waiverers 10 and 8, and one residual novice, 13.

Figure 2 shows the matrix of rankings for the esteem relation (positive relation), permuted in accord with CONCOR's partitioning. After several analyses, White et al. (1976) then used half the average block density as the Fit criterion in BLOCKER to produce the image matrix show in Table 3.

The blockmodel indicates that each of the three "blocks" or groupings esteems itself, while the Outcasts also esteem the Young Turks. A more detailed blockmodel of esteem relations, based upon a 5-partition blockmodel, reveals the

TABLE 2(a). CONCOR Division of Data.*

```
                  Novices' Identification Number
Split
Number        1 1 1 1   1 1       1 1     1
        1 7 2 5 2 4 6 3 7 8 4 6 8 0 1 5 9 3

4       XXX XXX XXX         XXX X
3       XXXXXXX XXXXX X XXX XXXXX XXX XXX X
2       XXXXXXXXXXXXX XXXXX XXXXXXXXX XXXXX
1       XXXXXXXXXXXXXXXXXXX XXXXXXXXXXXXXXX
        XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

TABLE 2(b). HCLUS Division of Data.**

```
                  Novices' Identification Number
Split
Number        1     1 1     1   1 1 1   1 1
        5 8 6 0 9 4 1 2 7 1 4 2 5 6 3 3 7 8

4       XXXXXXXXXXXXX   XXXXXXXXXXX   XXXXX
3       XXXXXXXXXXXXX   XXXXXXXXXXX XXXXXXX
2       XXXXXXXXXXXXX XXXXXXXXXXXXX XXXXXXX
1       XXXXXXXXXXXXX XXXXXXXXXXXXXXXXXXXXX
        XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

Source: Breiger et al. (1975).

* Adapted from Figure 2, p. 352.

** Adapted from Table 9(a), p. 357. Connectedness method; only first four of 15 clusterings used for comparability with CONCOR divisions.

deference relations among subblocks: LOs 4, 6, 8, 11 and YTs 1, 2, and 12 receiving most deference (White et al., 1976, p. 752).

Table 4 shows tie densities when all positive, and all negative, relations are pooled into two separate density matrices. In general, positive relations are primarily within groups, with few between groups, whereas the opposite is true for negative relations. Notice that the YTs actually receive the lowest average negative relations; the greatest negative affect is directed toward the Outcasts.

Breiger et al. (1975, p. 366) used MDSCAL to provide a two-dimensional mapping of the novices' relationships, using the initial correlation matrix.

Figure 3 (a) shows this dimensional mapping with the hierarchical clustering results of Table 2 (a) drawn in. Here, you can see the ambiguity in novice 13's position; he is *clustered* with the LOs, but is much *closer* to the Outcasts. Novices 8, 10, and 13, categorized variously as Waiverers or LOs, form the interface between the YTs and the LOs, but are clustered with the LOs. Arabie and Carroll's (1983) use of MAPCLUS provides overlapping clustering of the same MDSCAL mapping, as shown in Figure 3 (b). This eight-subset result

Novice ID.	10	5	9	6	4	11	8	12	1	2	14	15	7	16	13	3	17	18
10		1	1		3										2			
5			1		3	2												
9		1					3	2										
6		1	2		3													
4		3		1		2												
11		2					3				1							
8			1	2	3													
12									3	2			1					
1								1			2					3		
2								2	3				1					
14								1	3			2						
15								2		3			1					
7								1		3				2				
16										3		1	2					
13		2											1					3
3									3						2		1	2
17										1						2		3
18										1						2	3	

FIG. 2. Positive Esteem Relational Matrix.
Source: White et al. (1976, p. 750).
Note: Multiple or tied rankings as in source.

TABLE 3. Esteem Blockmodel

	LO	YT	O
LO	1	0	0
YT	0	1	0
O	0	1	1

NOTE: LO = 4,5,6,9,11,8,10
YT = 1,2,7,12,14,15,16
O = 3,17,18,13

TABLE 4. Average Tie Densities, Positive and Negative Relations

Group	Positive Relations			Negative Relations		
	LO	YT	O	LO	YT	O
LO	.38	.09	.04	.04	.16	.63
YT	.04	.43	.05	.29	.07	.43
O	.04	.10	.83	.42	.05	.00

Source: Breiger et al. (1975, p. 355). Values rounded from text. Each set of four relational matrices based upon top or bottom two rankings only, then combined by Boolean union; tie densities based upon resultant two matrices.

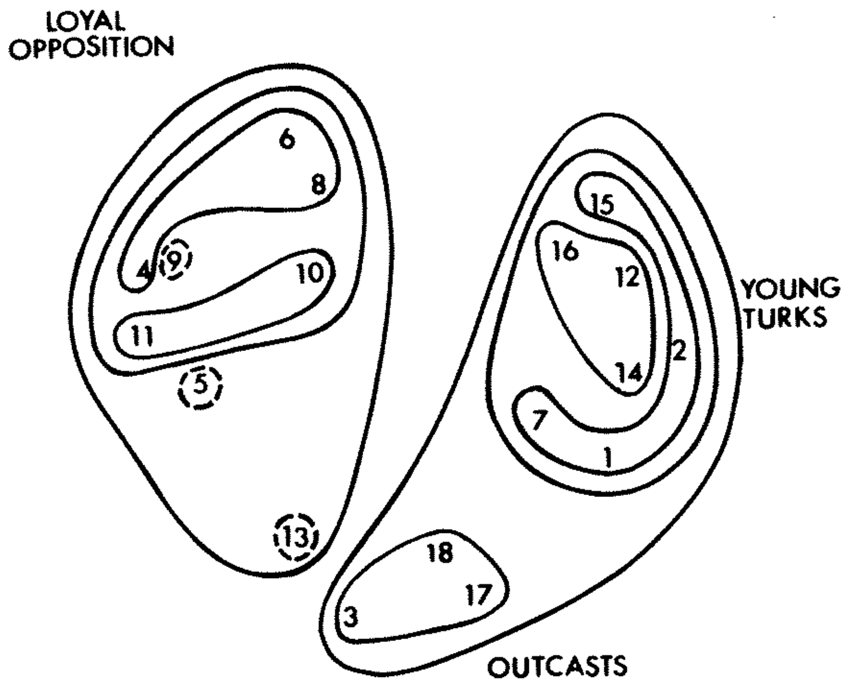

FIG. 3(a). MDSCAL Mapping With Hierarchical Clustering.
Source: Breiger et al. (1975, p. 366).
Note: Three novices circled by dotted lines are clustered together at one stage in CONCOR'S results.

explains 62.4% of the variance. Clusters are labelled alphabetically in terms of *decreasing* weights in the final solution. Clusters A, B, and E repeat the clusterings described above, except that novice 13 is considered part of the Outcasts. Arabie and Carroll point out that cluster F shows the Outcasts' sympathies with the YTs, as they are grouped at this level with the YT leaders. The two LOs who left voluntarily, 8 and 10, are also shown in the two weakest clusterings (G and H) with YTs. Conversely, novice 12, who stayed, is vaguely associated with the LO (A) cluster. Arabie and Carroll suggest that such overlapping clustering "is precisely the type of social structure that a partition or hierarchical clustering *cannot* represent."

Applying his ITR, HRCHY, and CLSTR programs to the Sampson data, Reitz (1983) was able to produce a different kind of clustering and graphic display of the novices' relational structure. He transformed the data by dichotomizing the 0 to +3 values of the positive relation matrices into values of zero and one (where 3 = 2 = 1), then added across the matrices to produce a single matrix with values from zero to four. He argued that this approach maintained the initial levels of measurement while creating an interval-level data matrix required by HRCY. (He notes that the prior methods all took liberties with the original data.) Figure 4 shows the results.

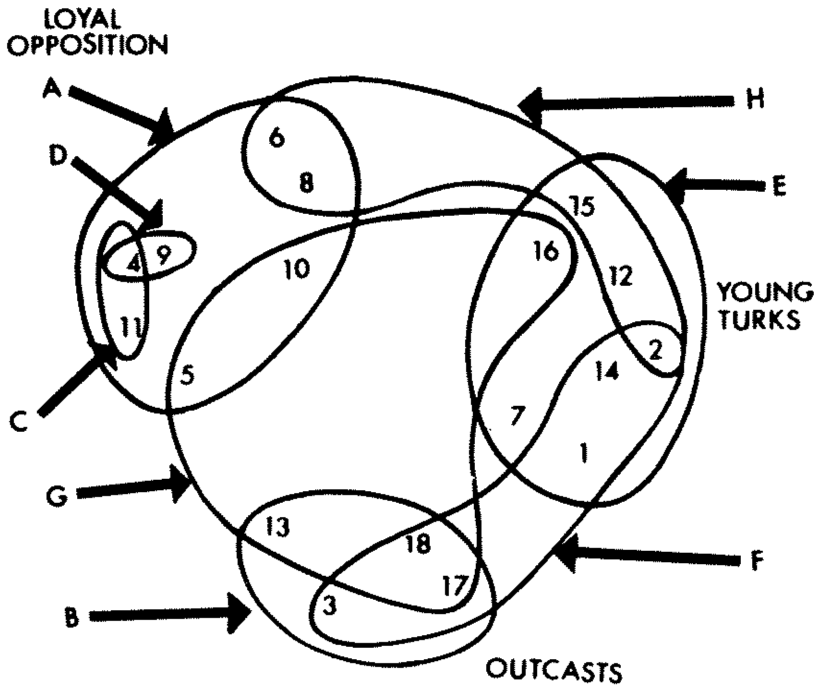

FIG. 3(b). MDSCAL Mapping With MAPCLUS Overlapping Clustering.
Source: Arabie and Carroll (1983, Fig. 2).

The novices are grouped into the familiar three clusters, but each cluster "has internal structure which can be interpreted in terms of popularity as well as closeness" (Reitz, 1983, p. 81). This clustering result is determined by the program, not the researcher, and the intercluster connectedness values Reitz reports confirm that each cluster is significantly distinct from each other. Leaders are also clearly identified: 4 and 5 in the LO, and 2 in the YT. Novice 1 is highly popular, but not really a close neighbor to the others.

The Waiverers are again grouped with the LOs, but note that they are at the bottom of each hierarchy. Novice 2's position at the top of the YT hierarchy provides evidence of the source of the superiors' fears of his popularity, interpreted as arrogance and independence.

Additional insight is provided by Reitz' (1982) log-linear analysis of the data. He hypothesized a significant expansiveness, attractiveness, and reciprocity parameter for the dyadic relationships among the novices. That model pro-

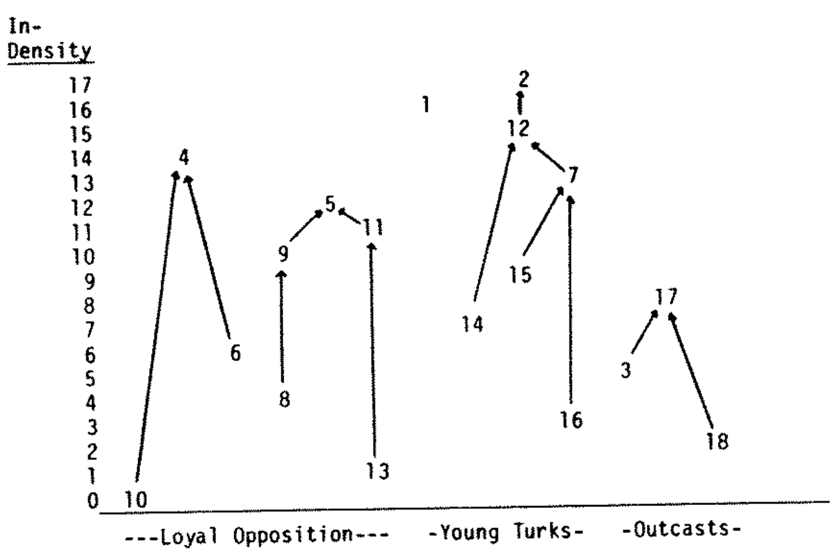

FIG. 4. Clustering and Mapping Based Upon Log-Linear Densities and Linkages.
Source: Reitz (1983, p. 81).
Note: In-Densities are rankings, with 17 highest.

vided an excellent fit to the data. Table 5 provides a few of the most extreme parameter estimates produced by his analysis.

Note that novice 2 has the highest positive expansiveness and attractiveness values, again potentially leading to an evaluation of arrogance. Novice 3 has negative estimates for both parameters, clearly the traits of an Outcast. Novice 10, a Waiverer, is "expansive," but not popular (as shown in Figure 4). And novice 12, the YT who did not follow the other YTs, receives higher-than-average nominations (particularly from those who were chary with their nominations, according to Reitz' definition of attractiveness), but was not very "socia-

TABLE 5. Parameter Estimates for Two Traits, for Four Selected Novices

Novice	Expansiveness	Attractiveness
2	1.10	5.07
3	−1.32	−.13
10	1.12	−2.96
12	−.54	1.45

Source: Reitz (1983, p. 74).

ble" himself. Perhaps this is the reason he did not leave; his was not a "proactive" personality, and perhaps he preferred the isolation of the monastery.

Because factor analysis is a similarity approach like blockmodelling, and a dimensioning approach like clustering, we would hope that the use of this method might provide similar or at worst tangential results. The use of factor analysis, an a posteriori method which is accessible and most likely understandable to researchers, highlights one of the basic issues of network analysis. That is, without a qualitative and theoretical insight into the data, there is little basis in choosing from among the possible results. We compare two sets of results to the preceding analyses.

The symmetric correlation matrix provided by Breiger et al. (1975, p. 351), was based upon the eight matrices (of positive and of negative relations, with values in each from 0 to +3), stacked together. The correlations themselves provide some insights. For example, the highest positive correlation ($R = .56$) occurs between novices 17 and 18, whom we know were rejected together as Outcasts. The strongest negative correlation ($R = {}^{-}.56$) is between novices 6 and 18. Note that novice 6 was a LO follower (see Figure 4) but also clustered at a low level with some of the primary YTs (see Figure 3(b)); thus the high negative value mirrors the extreme social distance between the two.

A straightforward application of factor analysis (rotated orthogonal principal components, not varimax) to this matrix leads to the solution portrayed in Table 6. Seven factors explain 71.6% of the variance; eigenvalues and variance for each factor are shown in the table. If we knew nothing else about the data, and used this "default" factor result, we certainly would not arrive at the same understanding of the network. But then again, each of the earlier analyses also depended on assumptions built into the data, the algorithms, the cutoffs, and the interpretations. Analytical precedence generates status for certain results. Some correspondences among the various analyses do obtain, however, and they tend to be the more detailed aspects. For example, novices 9 and 4 are a separate factor, as mirrored by cluster 4 in Figure 3(b). Novice 16, a low-order YT (see Figure 4), is his own man, so to speak. The Outcasts 13, 17 and 18 do represent one extreme of a separate dimension, while the opposition of 14 to 10 and 11 is also portrayed in factor two as well as in Figure 3(b). Due to measurement error in the recall data, sparse links in the eight relational matrices, forced symmetrization in the correlation matrix, and general instability in factors (particularly with few cases), one might well hesitate to use a default solution consisting of all the factors with eigenvalues greater than 1.0.

Knowing what Sampson had to say about the novices, we might choose instead to look for (or even test) a more parsimonious and perhaps more intuitive social structure; after all, that is the approach most of the analyses described so far have taken. So we solve for three factors. Results appear in Table 7. Things look "better." For each factor, the qualitatively-identified cliques are separated quite strongly by the signs of the loadings, although they are not separated across

TABLE 6. Seven-Factor Solution to Correlation Input

	Factors						
Novice	1	2	3	4	5	6	7
1			.53				
2			.48				
7	−.63						
12							−.54
14		−.73					
15			.82				
16						−.90	
3				−.74			
13	.39						
17	.72						
18	.86						
4	−.53				−.50*		
5							.82
6	−.63						
8				.73			
9					−.75		
10		.81					
11		.67					
E**	3.68	2.63	1.66	1.43	1.39	1.06	1.01
% Var.	20.4	14.6	9.2	8.0	7.7	5.9	5.6

* Second highest loading. Loadings less than .47 not significant if two-tailed alpha (.025) for N=18 is used as criterion for correlation significance.

** E=Eigenvalue; % Var.=Percentage of variance explained by factor.

factors. The Outcasts (including novice 13) load together, on the opposite end of the dimension of novices 6 (a low-level LO) and 8 (a waiverer)—this may be considered the vertical dimension of Figures 3(a) and (b). Similar dimensioning occurs for the loading splits on factors two and three. In fact, although we do not get a result which identifies a particular "clique" with a specific factor, we do get a fairly good picture of the three poles along which the novices differ. However, MDS seems to be more appropriate if that is the explicit goal of the analysis. It is clear, though, that this high-level and statistical representation of the network obscures or (intentionally) ignores other more pragmatic descriptions of and insights into the network. Further, it is very unclear how a researcher would know how many factors to solve for (without being more familiar with the data) if default settings are not utilized. But this was true for the other analyses too.

TABLE 7. Three-Factor Solution to Correlation Input*

	Factors		
Novice	1	2	3
1			−.37
2			−.64
7		.59	
12			−.41
14			−.69
15		.62	
16		.48	
3	.60		
13	.44		
17	.73		
18	.77		
4		−.70	
5		−.38	
6	−.69		
8	−.51		
9		−.65	
10			.73
11			.72

* The following four novices (with loading signs) comprised the third of four factors in the four-factor solution: #1(+), #4(−), #7(−) and #9(−). Oblique solutions were very similar to the orthogonal solutions for all analyses.

Taking a concrete, linkage-based approach, we use NEGOPY on a matrix consisting of the sum of the positive relation values portrayed in White et al. (1976, p. 751). That is, if novice *i* gives a 3, 2, 0, and 2 to novice *j* in each of the four positive relations, the value in the input matrix for cell (*i,j*) is 7. The maximum value possible "linkage" is thus 16, although the actual maximum is 12. (We also used the more conservative values used by Reitz, above, in another analysis resulting in the assignment of the outcasts to the Young Turks. All other results were quite similar.) NEGOPY prefers large networks, because of its emphasis on linkage instead of simple links (or even binary presence/absence), so that it may take advantage of greater variance in the list of linkages for each node. Thus, a network of 18 does not provide a large amount of material for NEGOPY to sink its graph-theoretical teeth into. Therefore, we set several parameters to attempt to detect as fine a structure as possible (directedness = 2; min split = 5; drop split = 5).

Even with this explicit attempt to partition the network, NEGOPY finds just two groups: the Young Turks (group 1) and everyone else (group 2) with no isolates, liaisons, or tree nodes. The density of the 18-node network is .41, with a mean strength of the relations which receive rankings of 8.63. There were 23 two-way (reciprocal) links and 39 one-way links. Both groups had around 85% of the ranking strength located in within-group relations, but the within-group links themselves constituted between 70% and 77%. Thus, not surprisingly, there is a tendency for within-group links to be stronger. The Young Turks had high density (.91), however, compared to the others (.49). This is largely an artifact of the inclusion of Outcasts and Waiverers in Group 2, and the necessary decline in density for larger groups (see Friedkin, 1981). The Outcasts and Waiverers 8 comprise 50% of all directed bridge links and 70% of the unreciprocated bridge links, as befits their labelling. Each group had a novice with high integrativeness—novice 3 (.83) and novice 15 (.87), and with low integrativness—novice 13 (.24) and novice 1 (.38).

Novice 15, a solid follower in the YTs (see Figure 4 and 3(b)), had all six of his links within the group, while novice 12 (the one who remained) did have one bridge link to novice 9. Of novice 12's links, four were reciprocated, while three were "received" but not reciprocated by 12, and the discrepancy in strength in those four reciprocated links had a mean of 1.25, nearly the lowest in the network. This implies that he was generally more attractive than sociable, but there was general convergence on the relative standing with those for whom he did have positive affect. Contrast that with novice 13. *None* of his seven links (four given, three received) was reciprocated—this represents total discrepancy, or a complete hierarchy, in novice 13's social structure. It is this discrepancy which is mirrored in how different researchers categorize this novice—as an Outcast or a Loyal Opposition member. Novice 2, the expelled, attractive, expansive, and arrogant member, had the most links of any novice—16. Four links were reciprocated, seven were received but not given, and only *one* was given but not received—and that to novice 8, a Waiverer.

From the more particularistic linkage analysis of NEGOPY, we see two basic divisions, bridged by the group of Outcasts and Waiverers. Novice 2 is clearly the most active relationally, and indeed receives far more esteem, praise, liking, and influence than the other novices. The actual strengths and linkages of the two groups are similar, and this contributes to NEGOPY's lack of finer partitioning; on the other hand, less partitioning is precisely the result that should occur under less system differentiation. And there is less differentiation partly because the *absence* of linkages, which is taken into account in the similarity methods, is not taken into account by NEGOPY. On the other hand, we can see how specific linkages provide understanding of the mechanisms whereby novices *are* outcasts, bridges, leaders, and the like.

We hope that the preceding very brief analyses of the Sampson data have provided some understanding of the biases, strengths and weakness, and appropriate applications of several of the leading network analysis methods. In this way, the network researcher may move with a bit more caution, but a bit more purpose, in analyzing social structure and the process of communication.

IV. CONCLUSION

Rogers and Kincaid (1981, Chapter 4) provide a useful chart describing summary characteristics of several of the programs included in this chapter (i.e., maximum node capacity, availability, and approximate costs). They also summarize some of the tradeoffs in choice—accessibility, cost, graphical presentation, size of network, multiplexity of networks, measures provided, and portrayal of roles. We agree with their main point: if different programs are available, use them in conjunction, to provide multiple validation and to provide different network aspects. Thus, a program with large capacity may be used to break the system into identifiable cliques which may be presented graphically by another program, or which may be analyzed separately for specific network measures. Multiple networks can be analyzed by some programs, particular roles and statuses by others. However, the theoretical bases of the researcher's analysis must be matched with those of the method and the assumptions about the data must be consistent and appropriate across programs.

Finally, perhaps the most significant assumption left unexamined by network analyses is the fundamental concept of clique at the base of each method—and some methods were not developed explicitly to find groups at all! The researcher should clearly understand the meaning of group, clique, or cluster as implied or defined by the method involved. When comparing linkage-based versus similarity-based results, for example, revealed groups may, with great justification, be very different even when based upon identical data.

We hope that the paradigm of network analysis will percolate through social science research, in appropriate ways. By that we mean that network analysis *may* provide useful insights into data, opportunities for hypothesis testing, and a general concern for the relational, processual nature of human communication. However, like any methodological tool kit—or ideological stance—it will be fruitful in the long run only if its assumptions are well understood; its data are reliable, valid, and appropriate in form and content for the method; its limitations are known and noted in the research; and it is motivated by an honest theoretical background. Network methods are intriguing and produce "interesting" results—but they must be subservient to the researcher's needs, and not dominate them.

REFERENCES

Alba, R. D. (1972). "COMPLT: A Program for Analyzing Sociometric Data and Clustering Similarity Matrices." *Behavioral Science 17*, 566–567.

Alba, R. D. (1973). "A Graph-Theoretic Definition of a Sociometric Clique." *Journal of Mathematical Sociology 3*, 113–126.

Alba, R. D., and Guttman, M. P. (1974). "SOCK: A Sociometric Analysis System." New York: Columbia University Report.

Alba, R. D., and Moore, G. (1983). "Elite Social Circles." *In* R. S. Burt and M. J. Minor (Eds.), *Applied Network Analysis*, pp. 245–261. Beverly Hills, CA: Sage.

Alt, J. E., and Schofield, N. (1975). "CLIQUE: A Suite of Programs for Extracting Cliques from a Symmetric Graph." *Behavioral Science 20*, 134–135.

Anthonisse, J. M., and Lagewey, B. J. (1980). *Graphlib, A Library of Computer Programs for Analysis of Graphs & Networks*, rev. version. Amsterdam: Mathematisch Centrim.

Arabie, P. (1977). "Clustering Representations of Group Overlap." *Journal of Mathematical Sociology 5*, 113–128.

Arabie, P., and Boorman, S. A. (1982). "Blockmodels: Developments and Prospects." *In* H. Hudson (Ed.), *Classifying Social Data*, pp. 177–198. San Francisco, CA: Jossey-Bass.

Arabie, P., and Carroll, J. D. (1980). "MAPCLUS: A Mathematical Programming Approach to Fitting the ADCLUS Model." *Psychometrika 45*, 211–235.

Arabie, P., and Carroll, J. D. (1983). "Conceptions of Overlap in Social Structure." *In* L. Freeman, A. K. Romney, and D. R. White (Eds.), *Methods of Social Network Analysis*. Berkeley, CA: University of California Press.

Arabie, P., and Shepard, R. N. (1973). "Representation of Similarities as Additive Combinations of Discrete, Overlapping Properties." Paper presented to Mathematical Psychology Meetings, Montreal.

Arabie, P., Boorman, S. A., and Levitt, P. R. (1978). "Constructing Blockmodels: How and Why." *Journal of Mathematical Psychology 17*, 21–63.

Atkin, R. H. (1974). *Mathematical Structure in Human Affairs*. New York: Crane, Rusak.

Augustson, J. G., and Minker, J. (1970). "An Analysis of Some Graph Theoretical Cluster Techniques." *Journal of the Association for Computing Machinery 17*, 571–588.

Bailey, K. D. (1974). "Cluster Analysis." *In* D. R. Heise (Ed.), *Sociological Methodology*, pp. 59–128. San Francisco, CA: Jossey-Bass.

Baker, F. B. (1974). "Stability of Two Hierarchical Grouping Techniques, Case I: Sensitivity to Data Errors." *Journal of the American Statistical Association 69*, 440–445.

Baker, F. B., and Hubert, L. J. (1981). "The Analysis of Social Interaction Data: A Non-Parametric Technique." *Sociological Methods and Research 9*, 339–361.

Barnes, J. A. (1969). "Graph Theory & Social Networks: A Technical Comment on Connectedness and Connectivity." *Sociology 3*, 215–232.

Barnett, G. A. (1973). "The Nature of Random Structures." East Lansing, MI: Michigan State University, Department of Communication.

Barnett, G. A., and Palmer, M. T. (1983). "Longitudinal Network Analysis Using Multidimensional Scaling." Paper presented to International Communication Association, Dallas.

Barnett, G. A., and Woelfel, J. (1979). "Dimensionality of Psychological Processes." *Quality and Quantity 13*, 215–232.

Beaton, A. E. (1966). "An Interbattery Factor-Analytic Approach to Clique Analysis." *Sociometry 29*, 135–145.

Berkowtiz, S. D. (1982). *An Introduction to Structural Analysis*. Toronto, Ontario: Butterworths.

Berkowitz, S. D., and Wellman, B. (Eds.) (1983). *Structural Sociology*. Cambridge, England: Cambridge University Press.

Berkowitz, S. D., Carrington, P. J., Corman, J. S., and Waverman, L. (1979). "A Design for a Large Scale Data Analysis for Corporate Information." *Social Networks 2*, 85–90.

Bernard, H. R., and Killworth, P. D. (1973). "On the Social Structure of an Ocean-Going Research Vessel and Other Important Things." *Social Science Reesearch 2*, 145–184.

Bernard, H. R., Killworth, P. D., and Sailer, L. (1982). "Informant Accuracy in Social Network Data V. An Experimental Attempt to Predict Actual Communication from Recall Data." *Social Science Research 11*, 30–66.

Beum, C. O., Jr., and Brundage, E. G. (1950). "A Method for Analyzing the Sociomatrix." *Sociometry 13*, 141–145.

Bielby, W. T. (1981). "Neighborhood Effects: A LISREL Model for Clustered Samples." *Sociological Methods and Research 10*, 82–91.

Blashfield, R. K. (1976). "Mixture Model Tests of Cluster Analysis: Accuracy of Four Agglomerative Hierarchical Methods." *Psychological Bulletin 83*, 377–388.

Bloombaum, M. (1970). "Doing Smallest Space Analysis." *Journal of Conflict Resolution 14*, 409–416.

Bock, R. D., and Husain, S. Z. (1950). "An Adaption of Holzinger's B-Coefficient for the Analysis of Sociometric Data." *Sociometry 13*, 146–153.

Bock, R. D., and Husain, S. Z. (1952). "Factors of the Tele: A Preliminary Report." *Sociometry 15*, 206–219.

Boissevain, J. (1974). *Friends of Friends: Networks, Manipulators and Coalitions*. Oxford, England: Basil Blackwell.

Boissevain, J., and Mitchell, J. C. (1973). *Network Analysis: Studies in Human Interaction*. The Hague, Amsterdam: Mouton.

Bonacich, P. (1972a). "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology 2*, 112–120.

Bonacich, P. (1972b). "A Technique for Analyzing Overlapping Group Memberships." *In* H. L. Costner (Ed.), *Sociological Methodology*, pp. 176–185. San Francisco, CA: Jossey-Bass.

Bonacich, P. (1978). "Using Boolean Algebra to Analyze Overlapping Memberships." *In* K. F. Schuessler (Ed.), *Sociological Methodolgy 1978*, pp. 101–115. San Francisco, CA: Jossey Bass.

Bonacich, P. (1980). "The 'Common Structure Semigroup,' a Replacement for the Boorman and White 'Joint Reduction.'" *American Journal of Sociology 86*, 159–166.

Bonacich, P., and McConaghy, M. J. (1980). "The Algebra of Blockmodeling." *In* K. F. Schuessler (Ed.), *Sociological Methodology 1980*, pp. 489–532. San Francisco, CA: Jossey-Bass.

Boorman, S. A., and White, H. C. (1976). "Social Structure from Multiple Networks II: Role Structures." *American Journal of Sociology 81*, 1384–1446.

Borgatta, E. F., and Stolz, W. (1963). "A Note on the Computer Program for Rearrangement of Matrices." *Sociometry 26*, 391–392.

Boyd, J. P. (1969). "The Algebra of Group Kinship." *Journal of Mathematical Psychology 6*, 139–167.

Breiger, R. L. (1976a). "Community at the Village Level: A Case of Sparse Networks." Unpublished Paper, Harvard University, Department of Sociology, Cambridge, MA.

Breiger, R. L. (1976b). "Career Attributes and Network Structure: A Blockmodel Study of a Biomedical Research Speciality." *American Sociological Review 41*, 117–135.

Breiger, R. L., and Pattison, P. E. (1978). "The Joint Role Structure of Two Communities' Elites." *Sociological Methods and Research 7*, 213–226.

Breiger, R. L., Boorman, S. A., and Arabie, P. (1975). "An Algorithm for Clustering Relational Data with Applications to Social Network Analysis and Comparison with Metric Dimensional Scaling." *Journal of Mathematical Psychology 12*, 328–383.

Bron, C., and Kerbosh, J. (1973). "Finding All Cliques of an Undirected Graph." *Communications of the ACM 1*, 575–577.

Brown, P. B. (1979). "BEMSD: A General-Purpose Program for Estimating Missing Data in a Sociomatrix." *Behavior Research Methods and Instrumentation 11*, 465–466.

Burt, R. S. (1976). "Positions in Networks." *Social Forces 55*, 106–131.

Burt, R. S. (1978). "Cohesion Versus Structural Equivalence as a Basis for Network Subgroups." *Sociological Methods and Research 7*, 189–212.

Burt, R. S. (1979). "Relational Equilibrium in a Social Topology." *Journal of Mathematical Sociology 6*, 211–252.

Burt, R. S. (1980a). "Actor Interests in a Social Topology: Foundations for a Structural Theory of Action." *Sociological Inquiry 49*, 107–132.

Burt, R. S. (1980b). "Models of Network Structure." *Annual Review of Sociology 6*, 79–141.

Burt, R. S. (1980c). "Innovation as a Structural Interest: Rethinking the Impact of Network Position on the Diffusion of Innovation." *Social Networks 2*, 327–356.

Burt, R. S. (1981). *Toward A Structural Theory of Actions: Network Models of Stratification, Perception and Action in a System of Actors*. New York: Academic Press.

Burt, R. S., and Bittner, W. M. (1981). "A Note on Inferences Regarding Network Subgroups." *Social Networks 3*, 71–88.

Burt, R. S., and Minor, M. J. (1983). *Applied Network Analysis*. Beverly Hills, CA: Sage.

Burton, R. (1970). "KLICOR: A CDC 6400 Program for Finding Subgraphs of Sociometric Nets." *Behavioral Science 15*, 536.

Carrington, P. J., and Heil, G. H. (1981). "COBLOC: A Hierarchical Method for Blocking Network Data." *Journal of Mathematical Sociology 8*, 103–132.

Carrington, P. J., Heil, G. H., and Berkowitz, S. D. (1980). "A Goodness of Fit Index for Blockmodels." *Social Networks 2*, 219–234.

Carroll, J. D. (1973). "Models and Algorithms for Multidimensional Scaling, Conjoint Measurement and Related Techniques." *In* P. E. Green, and Y. Wind, *Multi-Attribute Decisions in Marketing*, Appendix B. New York: Holt, Rinehart, Winston.

Carroll, J. D., and Arabie, P. (1980). "Multidimensional Scaling." *Annual Review of Psychology 31*, 607–649.

Cartwright, D., and Harary, F. (1956). "Structural Balance: A Generalization of Heider's Theory." *Psychological Review 63*, 277–292.

Cartwright, D., and Harry, F. (1970). "Ambivalence and Indifference in Generalizations of Structural Balance." *Behavioral Science 15*, 497–513.

Coleman, J. S. (1958). "Relational Analysis: A Study of Social Organization with Survey Methods." *Human Organization 17*, 28–36.

Coleman, J. S. (1970). "Clustering with N Dimensions by Use of a System of Forces." *Journal of Mathematical Sociology 1*, 1–47.

Coleman, J. S., and MacRae, D., Jr. (1960). "Electronic Data Processing of Sociometric Data for Groups up to 1000 in Size." *American Sociological Review 25*, 722–727.

Coleman, J. S., Katz, E., and Menzel, H. (1966). *Medical Innovation: A Diffusion Study*. New York: Bobbs-Merrill.

Cormack, R. M. (1971). "A Review of Classification." *Journal of the Royal Statistical Society (A) 134*, 321–367.

Curnow, R. C., McLean, M., and Shepherd, P. (1976). "Techniques for the Analysis of Structure." Sussex, England: University of Sussex. (SPRU Occasional Paper #1.)

Davies, D. L., and Bouldin, D. W. (1979). "A Cluster-Separation Technique." *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1*, 224–227.

Davis, J. A. (1967). "Clustering and Structural Balance in Graphs." *Human Relations 20*, 181–187.

Davis, J. A. (1970). "Clustering and Hierarchy in Interpersonal Relations: Testing Two Graph Theoretical Models on 742 Sociograms." *American Sociological Review 35*, 843–851.

Davis, J. A. (1977). "Sociometric Triads as Multi-Variate Systems." *Journal of Mathematical Sociology 5*, 41–60.

Delattre, M., and Hansen, P. (1980). "Bicriterion Cluster Analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-2*, 277–291.

Doreian, P. (1969). "A Note on the Detection of Changes in Valued Graphs." *Sociometry 32*, 237–242.

Doreian, P. (1980). "Linear Models with Spatially Distributed Data: Spatial Disturbances or Spatial Effects?" *Sociological Methods and Research 9*, 29–66.

Doreian, P. (1981). "Polyhedral Dynamics and Conflict Mobilization in Social Networks." *Social Networks 3*, 107–116.

Doreian, P. (1982). "Maximum Likelihood Methods for Linear Models: Spatial Effect and Spatial Disturbance Terms." *Sociological Methods and Research 10*, 243–270.

Dow, M. M., Burton, M. L., and White, D. R. (1982). "Network Auto-Correlation: A Simulation Study of a Foundational Problem in Regression and Survey Research." *Social Networks 4*, 169–200.

Dunn, C. J. (1974). "A Graph Theoretic Analysis of Pattern Classification via Tamura's Fuzzy Relation." *IEEE Transactions on Systems, Man, and Cybernetics SMC-4*, 310–313.

Edwards, J. A., and Monge, P. R. (1977). "The Validation of Mathematical Indices of Communication Structure." *Communication Yearbook 1*, 183–194.

Eisenberg, E. M., Farace, R. V., Monge, P. R., Bettinghaus, E. P., Kurchner-Hawkins, R., Miller, K. I., and Rothman, L. (1985). "Communication Linkages in Interorganizational Systems: Review and Synthesis." *Progress in Communications Sciences 6*, 236–261.

Everett, M. (1982a). "A Graph-Theoretic Blocking Procedure for Social Networks." *Social Networks 4*, 147–168.

Everett, M. (1982b). "Graph-Theoretic Blockings of K-Plexes and K-Cutpoints." *Journal of Mathematical Sociology 7*, 91–111.

Farace, R. V., and Mabee, T. (1980). "Communication Network Analysis Methods." *In* P. R. Monge and J. N. Cappella (Eds.), *Multivariate Techniques for Human Communication Research*, pp. 365–392. New York: Academic Press.

Farace, R. V., Monge, P. R., and Russell, H. M. (1977). *Communicating & Organizing*. Menlo Park, CA: Addison-Wesley.

Feger, H., Hummell, H. J., Pappi, F., Sodeur, W., and Ziegler, R. (1978). *Bibliographie zum Projeckt Analyse Sozialer Netwerke*. Wuppertal, W. Germany: Gesamthochschule Wuppertal.

Felling, A. J. A. (1975). "A Graph-Theoretical Approach to the Structure of Local Elites." *Zeitschrift für Soziologie 4*, 221–233.

Felling, A. J. A., and van der Weegen, T. (1976). "Programmer's Notes for Main Program NCLIQUE." Nijmegen, Netherlands: Mathematical Sociology and Research Technical Department, University of Nijmegen.

Festinger, L. (1949). "The Analysis of Sociograms Using Matrix Algebra." *Human Relations 2*, 153–158.

Fienberg, S. E. (1979). *The Analysis of Cross-Classified Categorical Data*. Cambridge, MA: MIT Press.

Fienberg, S. E., and Wasserman, S. (1981). "Categorical Data Analysis of Single Sociometric Relations." *Sociological Methodology 1981*, 156–192.

Fombrun, C. J. (1982). "Strategies for Network Research in Organizations." *Academy of Management Review 7*, 280–291.

Forsyth, E., and Katz, L. B. (1946). "A Matrix Approach to the Analysis of Sociometric Data: Preliminary Report." *Sociometry 9*, 340–347.

Freeman, L. C. (1968). *Patterns of Local Community Leadership*. Indianapolis, IN: Bobbs-Merrill.

Freeman, L. C. (1976). *A Bibliography of Social Networks*. Monticello, IL: Council of Planning Librarians, Exchange Bibliography.

Freeman, L. C. (1977). (Personal communication to W. D. Richards, Jr.)

Freeman, L. C. (1978). "On Measuring Systematic Integration." *Connections 2* (No. 1), 13–14.

Freeman, L. C. (1979). "Centrality in Social Networks: Conceptual Clarification." *Social Networks 1*, 215–239.

Freeman, L. C. (1980). "Q-Analysis and the Structure of Friendship Networks." *International Journal of Man-Machine Studies 12*, 367–378.

Freeman, L. C., Roeder, D., and Mulholland, R. R. (1980). "Centrality in Social Networks II: Experimental Results." *Social Networks 2*, 119–142.

Freeman, L. C., and Romney, A. K., and White, D. R. (Eds.) (1983). *Methods of Social Network Analysis*. Berkeley, CA: University of California Press.

Friedheim, E. A. (1982). "Graphic Display for Multidimensions: A Note." *Sociological Methods and Research 10*, 485–493.

Friedkin, N. E. (1981). "The Development of Structure in Random Networks: An Analysis of the Effects of Increasing Network Density of Five Measures of Structure." *Social Network 3*, 41–52.

Galaskiewicz, J. (1979). *Exchange Networks and Community Politics*. Beverly Hills, CA: Sage.

Gengerelli, J. A. (1963). "A Method for Detecting Subgraphs in a Population and Specifying their Membership." *Journal of Psychology 55*, 457–468.

Gillham, J., and Woelfel, J. (1977). "The Galileo System of Measurement." *Human Communication Research 3*, 222–234.

Gleason, T. C. (1971). "DIP: A Directed-Graph Processor." Ann Arbor, MI: University of Michigan, Institute for Social Research.

Gleason, T. C., and Cartwright, D. (1967). "A Note on a Matrix Criterion for Unique Colorability of a Signed Graph." *Psychological Biometrika 32*, 291–296.

Goldstein, K. M., Blackman, S., and Collins, D. J. (1966). "Relationship between Sociometric and Multidimensional Scaling Measures." *Perceptual and Motor Skills 23*, 639–643.

Gottlieb, B. H. (Ed.) (1981). *Social Networks and Social Support*. Beverly Hills, CA: Sage.

Gould, P. (1980). "Q-Analysis, or a Language of Structure: An Introduction for Social Scientists, Geographers and Planners." *International Journal of Man-Machine Studies 13*, 169–199.

Green, P. E., and Rao, V. R. (1972). *Applied Multidimensional Scaling: A Comparison of Approaches and Algorithms*. Hinsdale, IL: Dryden Press.

Guimaraes, L. L. (1972). "Communication Integration in Modern and Traditional Social Systems." Unpublished dissertation, Michigan State University.

Guttman, L. (1968). "A General Nonmetric Technique for Finding the Smallest Coordinate Space for a Configuration of Points." *Psychometrika 33*, 469–508.

Hansen, P., and Delattre, M. (1978). "Complete-Link Cluster Analysis by Graph Coloring." *Journal of the American Statistical Association 73*, 397–403.

Harary, F., and Ross, I. C. (1957). "A Procedure for Clique Detection Using the Group Matrix." *Sociometry 20*, 205–215.

Harary, F., Norman, R. Z., and Cartwright, D. (1965). *Structural Models: An Introduction to the Theory of Directed Graphs*. New York: Wiley.

Hartigan, J. A. (1975). *Clustering Techniques*. New York: Wiley.

Heil, G. H., and White, H. C. (1976). "An Algorithm for Finding Simultaneous Homomorphic Correspondences Between Graphs and Their Image Graphs." *Behavioral Science 21*, 26–35.

Holland, P. W., and Leinhardt, S. (1973). "The Structural Implications of Measurement Error in Sociometry." *Journal of Mathematical Sociology 3*, 85–111.

Holland, P. W., and Leinhardt, S. (1976). "The Statistical Analysis of Local Structure in Social Networks." *In* D. R. Heise (Ed.), *Sociological Methoaology 1975*, pp. 1–45. San Francisco, CA: Jossey-Bass.

Holzinger, K. J., and Harman, H. H. (1941). *Factor Analysis*. Chicago, IL: University of Chicago Press.

Horan, P. M. (1975). "Information-Theoretic Measures and the Analysis of Social Structures." *Sociological Methods and Research 3*, 321–340.

Hubbell, C. H. (1965). "An Input-Output Approach to Clique Identification." *Sociometry 28*, 377–399.

Hubert, L. J. (1974). "Some Applications of Graph Theory to Clustering." *Psychometrika 39*, 283–309.

Hubert, L. J., and Baker, F. B. (1978). "Evaluating the Conformity of Sociometric Measurements." *Psychometrika 43*, 31–41.

Hubert, L. J., and Baker, F. B. (1979). "Evaluating the Symmetry of a Proximity Matrix." *Quality and Quantity 13*, 77–84.

Jablin, F. M. (1980). "Organization Communication Theory and Research: An Overview of Communication Climate and Network Research." *In* D. Nimmo (Ed.), *Communication Yearbook 4*, pp. 327–347. Beverly Hills, CA: Sage.

Jacobson, E., and Seashore, S. (1951). "Communication Practices in Complex Organizations." *Journal of Social Issues 7*, 28–40.

Jardine, N., and Sibson, R. (1968). "The Construction of Hierarchic and Non-Hierarchic Classifications." *Computer Journal 11*, 177–184.

Johnson, S. C. (1967). "Hierarchical Clustering Schemes." *Psychometrika 32*, 241–254.

Katz, L. B. (1947). "On the Matrix Analysis of Sociometric Data." *Sociometry 10*, 233–241.

Katz, L. B., and Powell, J. H. (1960). "A Proposed Index of the Conformity of One Sociometric Measurement to Another." *In* J. L. Moreno (Ed.), *The Sociometry Reader*, pp. 298–306. Glencoe, IL: Free Press.

Killworth, P. D. (1974). "Intransitivity in the Structure of Small Closed Groups." *Social Science Research 3*, 1–23.

Killworth, P. D., and Bernard, H. R. (1974). "CATIJ: A New Sociometric and Its Application to a Prison Living Unit." *Human Organization 33*, 335–350.

Killworth, P. D., and Bernard, H. R. (1979). "Informant Accuracy in Social Network Data III: A Comparison of Triadic Structure in Behavioral and Cognitive Data." *Social Networks 2*, 19–46.

Klovdahl, A. S. (1979a). *Social Networks: Selected References for Course Design and Research Planning*. Monticello, IL: Vance Bibliographies.

Klovdahl, A. S. (1979b). "MACRONET." *Connections 2* (No. 2), 110–111.

Knoke, D., and Kuklinski, J. H. (1982). *Network Analysis*. Beverly Hills, CA: Sage.

Krippendorf, K. (1980). "Clustering." *In* P. R. Monge and J. N. Cappella (Eds.), *Multivariate Techniques for Human Communication Research*, pp. 259–308. New York: Academic Press.

Krippendorf, K. (1981). "An Algorithm for Identifying Structural Models of Multi-Variate Data." *International Journal of General Systems 7*, 1–17.

Kruskal, J. B. (1969). "Nonmetric Multidimensional Scaling: A Numerical Method." *Psychometrika 29*, 115–159.

Kruskal, J. B., and Wish, M. (1978). *Multidimensional Scaling*. Beverly Hills, CA: Sage.

Kuiper, F. K., and Fisher, L. (1975). "A Monte Carlo Comparison of Six Clustering Processes." *Biometrics 31*, 777–783.

Kushner, H. W., and Demaio, G. (1977). "Using Digraphs to Determine the Crucial Actors in a Voting Body." *Sociometry 40*, 361–364.

Lankford, P. M. (1974). "Comparative Analysis of Clique Identification Methods." *Sociometry 37*, 287–305.

Leinhardt, S. (Ed.) (1977). *Social Networks*. New York: Academic Press.

Lesniak, R., Yates, M. P., Goldhaber, G. M., and Richards, W. D., Jr. (1977). "NETPLOT: An Original Computer Program for Interpreting NEGOPY." Paper presented to the International Communication Association, Berlin.

Lin, N. (1976). *Foundations of Social Research*. New York: McGraw-Hill.

Lindzey, G., and Byrne, D. (1968). "Measurement of Social Choices and Interpersonal Attractiveness." *In* G. Lindzey and E. Aronson (Eds.), *The Handbook of Social Psychology*, Vol. 2, pp. 452–525. Reading, MA: Addison-Wesley.

Lingoes, J. (1973). "The Guttman-Lingoes Nonmetric Program Series. Ann Arbor, MI: Mathesis Press.

Lorrain, F. P., and White, H. C. (1971). "Structural Equivalence of Individuals in Social Networks." *Journal of Mathematical Sociology 1*, 49–80.

Luce, R. D. (1950). "Connectivity and Generalized Cliques in Sociometric Group Structure." *Psychometrika 15*, 169–190.

Luce, R. D., and Perry, A. D. (1949). "A Method of Matrix Analysis of Group Structure." *Psychometrika 14*, 94–116.

MacRae, D., Jr. (1960). "Direct Factor Analysis of Sociometric Data." *Sociometry 23*, 360–371.

Marsden, P. V. (1981). "Models and Methods for Characterizing the Structural Parameters of Groups." *Social Networks 3*, 1–27.

Marsden, P. V. (1983). "Methods for the Characterization of Role Structures in Network Analysis." *In* L. Freeman, A. K. Romney, and D. R. White (Eds.), *Methods of Social Network Analysis*. Berkeley, CA: University of California Press.

Marsden, P. V., and Lin, N. (Eds.) (1982). *Social Structure and Network Analysis*. Beverly Hills, CA: Sage.

McConaghy, M. J. (1981). "The Common Role Structure: Improved Blockmodelling Methods Applied to Two Communities' Elites." *Sociological Methods and Research 9*, 267–286.

McFarland, D. D., and Brown, D. J. (1973). "Social Distance as Metric: A Systematic Introduction to Smallest Space Analysis." *In* E. O. Laumann (Ed.), *Bonds of Pluralism: The Form and Substance of Urban Social Networks*, pp. 213–253. New York: Wiley.

McPhee, R. D., and Poole, M. S. (1979). *Workshop on Developmental Methodology*. Urbana, IL: University of Illinois-Urbana.

McQuitty, L. L. (1957). "Elementary Linkage Analysis for Isolating Orthogonal and Oblique Types and Typal Relevancies." *Educational and Psychological Measurement 17*, 207–229.

Mitchell, J. C. (1969). *Social Networks in Urban Situations*. Manchester, England: University Press.

Moch, M., Feather, J. N., and Fitzgibbons, D. (1980). "Conceptualizing and Measuring the Relational Structure in Organizations." *In* S. E. Seashore (Ed.), *Assessing Organizations: A Guide to Practice*. New York: Wiley Interscience.

Mokken, R. J. (1979). "Cliques, Clubs and Clans." *Quality and Quantity 13*, 161–173.

Monge, P. R., and Cappella, J. N. (Eds.) (1980). *Multivariate Techniques for Human Communication Research*. New York: Academic Press.

Moreno, J. L. (1953). *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. New York: Beacon House. (Originally published, 1934.)

Moreno, J. L. (Ed.) (1960). *The Sociometry Reader*. Glencoe, IL: Free Press.

Mouton, J. S., Blake, R. R., and Fruchter, B. (1960). "The Reliability of Sociometric Measures." *In* J. L. Moreno (Ed.), *The Sociometry Reader*, pp. 320–361. Glencoe, IL: Free Press.

Nehnevajsa, J. (1960). "Sociometry: Decades of Growth." *In* J. L. Moreno (Ed.), *The Sociometry Reader*, pp. 707–762. Glencoe, IL: Free Press.

Norton, R. W. (1980). "Nonmetric Multidimensional Scaling in Communication Research: Smallest Space Analysis." *In* P. R. Monge and J. N. Cappella (Eds.), *Multivariate Techniques for Human Communication Research*, pp. 308–331. New York: Academic Press.

Panning, H. W. (1982). "Fitting Blockmodels to Data." *Social Networks 4*, 81–101.

Peay, E. R. (1974). "Hierarchical Clique Structures." *Sociometry 37*, 54–65.

Peay, E. R. (1980). "Connectedness in a General Model for Valued Networks." *Social Networks 2*, 385–410.

Phillips, D. P., and Conviser, R. H. (1972). "Measuring the Structure and Boundary Properties of Graphs: Some Uses of Information Theory." *Sociometry 35*, 235–254.

Pitts, F. (1979). "A Bibliography of Social Networks." Paper presented to Seminar on Communication Network Analysis. East-West Communication Center, Honolulu.

Project in Structural Analysis. (1977). Structure: A Computer Program Providing Basic Data for the Analysis of Empirical Positions in System of Actors. Berkeley, CA: University of California Survey Research Center. (Working Paper 5.)

Purdy, G. (1973). "Finding the Cliques of a Graph." Urbana IL: University of Illinois Center for Advanced Computation. (Technical Memorandum.)

Rattinger, H. (1973). "Eine Einfache Methode und ein FORTRAN-Programm zur Ermittlung van Cliquen." *Zeitschrift für Sozialpsychologie 4*, 5–14.

Reitz, K. P. (1982). "Using Log Linear Analysis with Network Data: Another Look at Sampson's Monastery." *Social Networks 4*, 243–256.

Reitz, K. P. (1983). "Social Groups, A Network Approach." Unpublished dissertation, University of California, Irvine.

Rice, R. E. (1978). *Negopy User's Manual*. Palo Alto, CA: Stanford University, Institute for Communication Research.

Rice, R. E. (1979a). "Investigations into the Validity and Reliability of NEGOPY: A Computer Program for Network Analysis." Paper presented to International Communication Association, Philadelphia.

Rice, R. E. (1979b). "NEGOPY: Parameter Effects." Paper presented to Seminar on Communication Network Analysis, East-West Communication Center, Honolulu.

Rice, R. E. (1981). "Resources for Longitudinal Network Analysis." *Connections 4* (No. 2), 10–22.

Rice, R. E. (1982). "Communication Networking in Computer Conferencing Systems: A Longitudinal Study of Group Roles and System Structure." *In* M. Burgoon (Ed.), *Communication Yearbook 6*, pp. 925–944. Beverly Hills, CA: Sage.

Rice, R. E., and Barnett, G. (1986). "Group Communication Networking in an Information Environment." *In* M. McLaughlin (Ed.), *Communication Yearbook 9*. Beverly Hills, CA: Sage.

Richards, W. D., Jr. (1974). "Network Analysis in Large Complex Organizations: Theoretical Basis; The Nature of Structure; Techniques and Methods—Tools." Four papers presented to the International Communication Association, New Orleans.

Richards, W. D., Jr. (1975). *A Manual for Network Analysis (Using NEGOPY Analysis Program)*. Stanford, CA: Stanford University, Institute for Communication Research.

Richards, W. D., Jr. (1976). "A Coherent Systems Methodology for the Analysis of Human Communication Systems." Unpublished dissertation, Stanford University.

Richards, W. D., Jr. (1977). "Network Analysis Methods: Conceptual and Operational Approaches." Paper presented to Colloquium on Social Networks, University of Hawaii, Manoa.

Richards, W. D., Jr. (1979). "Measurement of Problems in Network Analysis I: Reciprocity and Directed Relationships." Paper presented to International Communication Association, Philadelphia.

Richards, W. D., Jr. and Lindzey, G. (1974). "Social Network Analysis: An Overview of Recent Developments." Paper presented to Social Processes, Conference of the American Cybernetic Society, Philadelphia.

Richards, W. D., Jr., and Rice, R. E. (1981). "NEGOPY Network Analysis Program." *Social Networks 3*, 215–223.

Rogers, E. M., and Agarwala-Rogers, R. (1976). *Communication in Organizations*. New York: Free Press.

Rogers, E. M., and Kincaid, D. L. (1981). *Communication Networks: Toward a New Paradigm for Research*. New York: Free Press.

Sailer, L. D. (1978). "Structural Equivalence: Meaning and Definition, Computation and Application." *Social Networks 1*, 73–90.

Salzinger, L. L. (1982). "The Ties that Bind: The Effect of Clustering on Dyadic Relationships." *Social Networks 4*, 117–145.

Sampson, S. F. (1969). "Crisis in a Cloister." Unpublished dissertation, Cornell University.

Schwartz, J. E. (1976). "An Examination of CONCOR and Related Methods for Blocking Sociometric Data." *In* D. R. Heise (Ed.), *Sociological Methodology*, pp. 255–282. San Francisco, CA: Jossey-Bass.

Seidman, S. B. (1978). "SONET-I." *Connections 2* (No. 1), 46–48. (Also published in S. Seidman and B. Foster, "SONET-I." *Social Networks 2*, 85–90 (1979).)

Seidman, S. B. (1981). "Structures Induced by Collections of Subsets: A Hypergraph Approach." *Mathematical Social Sciences 1*, 381–396.

Seidman, S. B., and Foster, B. L. (1978). "A Graph-Theoretic Generalization of the Clique Concept." *Journal of Mathematical Sociology 6*, 139–154.

Shaw, M. E. (1971). *Group Dynamics: The Psychology of Small Group Behavior*. New York: McGraw-Hill.

Shaw, W. W., Jr. (1983). "Statistical Disorder and the Analysis of a Communication-Graph." *Journal of the American Society for Information Science 34*, 146–149.

Shepard, R. N. (1972). *Multidimensional Scaling*. New York: Seminar Press.

Sim, F., and Schwartz, M. (1983). "Does CONCOR Find Positions?" Paper presented to Social Networks Conference, San Diego.

Sodeur, W. (1979). "NETZ—Unterprogramme zur Analyse Sozialier Netze." Wuppertal, W. Germany: Gesamthochschule Wuppertal. (mimeo.)

Sonquist, J. A. (1980). "Concepts and Tactics in Analyzing Social Network Data." *Connections 3* (No. 1), 33–56.

Stokman, F. N. (1983). "GRAPH Definition and Analysis Package (GRADAP)." *Connections 6* (No. 1), 57–60.

Stokman, F. N., and van Veen, F. (1981). *GRADAP, Graph Definition and User's Manual.* Vols. 1 and 2. Amsterdam, Netherlands: University of Amsterdam-Groningen-Nijmegen, interuniversity Project Group GRADAP.

Tichy, N. M., Tushman, M. L., and Fombrun, C. (1979). "Social Network Analysis for Organizations." *Academy of Management Review 4*, 507–520.

Tryon, R. C., and Bailey, D. E. (1970). *Cluster Analysis.* New York: McGraw-Hill.

Wasserman, S. (1980a). "A Stochastic Model for Directed Graphs with Transition Rates Determined by Reciprocity." *In* K. Scheussler (Ed.), *Sociological Methodology 1980*, pp. 392–412. San Francisco, CA: Jossey-Bass.

Wasserman, S. (1980b). "Analyzing Social Networks as Stochastic Processes." *Journal of the American Statistical Association 75*, 280–294.

Weimann, G. (1982). "On the Importance of Marginality: One More Step into the Two-Step Flow of Communication." *American Sociological Review 47*, 764–773.

Weiss, R. S., and Jacobson, E. (1955). "A Method for the Analysis of the Structure of Complex Organizations." *American Sociological Review 20*, 661–668.

Wellman, B., and Whitaker, M. (Eds.) (1974). "Community Network Communication: An Annotated Bibliography." Toronto, Ontario: University of Toronto, Centre for Urban and Community Studies. (Report #4.)

White, D. R. (1981). "Role Equivalence in Networks." Irvine, CA: University of California, Irvine: School of Social Sciences, University of California, Irvine. (Unpublished manuscript.)

White, D. R., and McCann, G. (1983). "Material Entailment Illustrated by an Analysis of Citation Patterns of 18th Century Chemistry." *In* S. D. Berkowitz and B. Wellman (Eds.). *Structural Sociology.* Cambridge, England: Cambridge University Press.

White, D. R., and Reitz, K. (1983). "Graph and Semigroup Homomorphisms on Networks of Social Relations." *Social Networks 5*, 193–234.

White, D. R., and Sailer, L. (1979). "Conference Comments on Structural Equivalence." CC246–252, in Conference 35, EIES System.

White, D. R., Burton, M. L., and Dow, M. M. (1981). "Sexual Division of Labor in African Agriculture: A Network Autocorrelation Analysis." *American Anthropologist 83*, 824–849.

White, H. C. (1963). *An Anatomy of Kinship.* Englewood Cliffs, NJ: Prentice-Hall.

White, H. C. (1974a). "Models for Interrelated Roles from Multiple Networks in Small Populations." *In* P. J. Knopp and G. H. Meyer (Eds.), *Proceedings of the Conference on the Application of Undergraduate Mathematics in the Engineering, Life, Managerial and Social Sciences.* Atlanta, GA: Georgia Institute of Technology.

White, H. C. (1974b). "Multiple Networks in Small Populations, II: Compound Relations and Equations." Cambridge, MA: Harvard University, Department of Sociology. (Unpublished paper.)

White, H. C., and Breiger, R. L. (1975). "Patterns Across Networks." *Society 12*, 68–73.

White, H. C., Boorman, S. A., and Breiger, R. L. (1976). "Social Structure from Multiple Networks I. Blockmodels of Roles and Positions." *American Journal of Sociology 81*, 730–780.

Wigand, R. T. (1977). "Some Recent Developments in Organizational Communication: Network Analysis—A Systemic Representation of Communication Relationships." *International Journal of Communication Research 2*, 181–200.

Wigand, R. T. (1979). "Communication Network Analysis: A Computerized Tool in the Analysis of Large Social Systems with Applications for Organizations." *In* G. Klir et al. (Eds.), *Basic and Applied General Systems Research.* New York: Elsevier North Holland.

Wigand, R. T. (1982). "The Communication Industry in Economic Integration: The Case of West Germany." *Social Networks 4*, 47–79.

Woelfel, J., and Danes, J. (1980). "Multidimensional Scaling Methods for Communication Research." *In* P. R. Monge and J. N. Cappella (Eds.), *Multivariate Techniques in Human Communication Research*, pp. 333–364. New York: Academic Press.

Wright, B., and Evitts, M. S. (1961). "Direct Factor Analysis in Sociometry." *Sociometry 24*, 82–98.