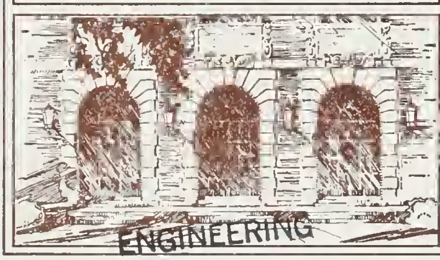
ENGINEERING

# Center for Advanced Computation

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, ILLINOIS 61801

CAC Document No. 129

## A REVIEW OF MATHEMATICAL METHODS IN SOCIOMETRY

by
Richard C. Roistacher

August 17, 1974

CAC Document No. 129

# A Review of Mathematical Methods in Sociometry

By

Richard C. Roistacher

Center for Advanced Computation
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

August 17, 1974

This work was supported in part by the Milton Olive Corporation.

A Review of Mathematical Methods in sociometry

ABSTRACT

This paper reviews current mathematical and computer methods for the analysis of group and organizational structures. Sociometric data collection methods, the detection of cliques and subgroups, sociometric index construction, and the use of computers in sociometry are discussed. Clique detection methods are classified as either linkage methods, in which sociometric data are treated as linear graphs, or as distance methods, in which data are treated as a configuration of points in a space. Several linkage and distance analysis methods are discussed and compared. It is suggested that some of the methods of numerical taxonomy could be applied to the analysis of group and organizational structures.

# A Review of Mathematical Methods in Sociometry

Richard C. Roistacher
Center for Advanced Computation
University of Illinois
Urbana, IL 61801

Over the last thirty-eight years sociometry has become one of the standard procedures for the analysis of group structures of communication, influence, and affiliation. Substantive sociometric work is being done in settings ranging from industrial plants and hospitals to nursery schools. In addition to investigators who use sociometric techniques in their research on groups and organizations, there is a steadily increasing number of workers who are primarily interested in the invention and refinement of techniques for the measurement of group structures.

This report surveys sociometric methodology rather than sociometric investigation. Research in substantive areas has been cited to illustrate methodological points rather than to demonstrate what sociometric research has revealed about groups and organizations. Lindzey and Byrne (1968) provide the best overall review of sociometric research, while Glanzer and Glaser (1959) review mathematical techniques of sociometric analysis in detail. Guimaraes (1968) provides an excellent introduction to the mathematical concepts of sociometric analysis and review of matrix multiplication in sociometry.

This paper begins by surveying present definitions of sociometry and by suggesting some extensions to these definitions. Its major sections are organized around two main tasks of the sociometric investigator: collecting data; viewing the group as composed of many, possibly overlapping subgroups, and computing individual and group indices. The section on data collection considers questionnaires and rating instruments, but not direct observational techniques. Limited- versus unlimited-choice instruments are compared, and ranking techniques are compared with direct choice and rating schemes.

The section on the definition of sociometric subgroups is divided into two major parts. The first deals with what can be called linkage analysis techniques, in which group members are considered as points linked by lines. The second deals with what can be called distance analysis techniques, in which group members are considered as points distributed through a geometric space. The paper concludes with a short section on the construction of sociometric indices and a review of computing techniques in sociometry.

<u>Why sociometry.</u> Sociometry is a technique deeply

grounded in common experience. People are constantly evaluating and being evaluated by their acquaintances. People and their acquaintances are bound together by patterns of friendship and association. Groups are perceived as having centers, peripheries, and boundaries; individuals as isolates--people marginal to our own or other groups, or as central to a pattern of association. Most of the metaphors used to describe networks of associations have spatial or geometric implications. Moreno has often commented on the sociogram's immediate impact on members of the group it represents; evidently a representation of a group in a two-dimensional Euclidian space speaks very strongly to people's feelings about themselves and their associates.

If we are so taken by these immediate geometric and spatial perceptions of social structure, then our motivation for investigating the spatial and geometric aspects of group structures in a more rigorous form is easy to explain. Although the world is experienced as a three-dimensional Euclidian space, there is no reason why the structure of acquaintance patterns should be restricted to this form.

The rubric of sociometry now covers a broad range of investigative techniques and a multitude of sins. The reviewer's task is not so much to provide a single definition of the technique as to assemble a set of definitions which characterize the tasks of sociometry and which will cover an ever-broadening set of techniques.

Definitions on the basis of data content. The first use of sociometry was by J. L. Moreno in the 1920s and was popularized in his book, *Who Shall Survive?* (1934). Moreno's criteria for a sociometric instrument include: that it be given to a limited group of respondents, each of whom is asked to indicate his preference for engaging in an activity or activities with other members of the group; that there be an unlimited number of choices and rejections allowed within the group; that the choices be made privately and anonymously; and that the questions fit the group's level of understanding. These criteria still describe most of the sociometric research done today.

Moreno also suggested that valid results would not be obtained unless the group were aware that some consequence would result from the sociometric study. However, Mouton, Blake, and Fruchter (1960a, 1960b) in their reviews of the reliability and validity of sociometric instrumentation, observe that it is not possible to obtain an equivalent retest in a group which has been restructured as a result of a previous sociometric study.

Most attempts to define the sociometric method have concentrated on the objects of measurement and the content of the measuring instruments. Bain (1943), in an attempt to synthesize the concepts of sociometry, concluded that sociometry "will remain a generic term to describe all measurement of societal and interpersonal data." More recently Bjerstedt (1956, pp. 15-23) identified thirteen definitions of sociometry which he submitted to a panel of 269 experts for rating. The modal response among the 131 returns was that sociometry is the quantitative treatment of "preferential inter-human relations."

This definition is based on the content of sociometric data sets, most of which contain information on person i's evaluation of person j. However, it is also possible to define the sociometric research process in terms of the structure of sociometric data. The structure of sociometric data matrices uniquely distinguishes them from other types of social science data. Coleman (1958) gave a brief, but far-sighted analysis of the relation among opinion polling, survey, and sociometric data. The same characteristics, however, relate sociometric data to the data of other scientific disciplines, most particularly to numerical taxonomy in the biological sciences. The natural structure of sociometric data is a matrix. After a short digression on the structural properties of data matrices, the relation of data structure to analytical goals in sociometry and other areas will be considered.

Matrix representations of data. It is possible to represent almost all forms of psychological data as matrices. We may think of an investigation as dealing with two sets of entities, which can be called subjects and objects. Each cell provides a place for representing a relation between one subject and another, one object and another, or between a subject and an object.

If the relation described by the matrix M is symmetric then the matrix will be symmetric, i.e., the entry in cell m(i,j) will be equal to that in cell m(j,i). If the relation is not symmetric, then following Coombs' (1964) usage the matrix is often called conditional, for the interpretation of the difference between two entries in a row depends on the particular row being considered.

For instance, suppose that a matrix contained the distances between the houses of the subjects, or between the houses of the subjects and some set of objects. Since the distance from A to B is the same as the distance from B to A, the matrix would be symmetric. In addition, it would be possible to use the matrix to trace the length of a path among a set of houses, i. e., to add entries across rows of the matrix. However, if the i-th row of the matrix contained numbers indicating how much subject i liked each of the other subjects, it would not be possible to make comparisons among individuals (between rows), nor would the matrix necessarily be symmetric.

A matrix may be binary, consisting of only ones and zeros, or it may contain real numbers as entries. A binary matrix indicates only that some relation either does or does not hold between two entities, while a real matrix may indicate the strength or type of relation between them. Where repeated or multiple measurements are made, the data may be in three- or higher-dimensional form.

These properties of data matrices serve to link otherwise disparate fields of inquiry. The structure of a data matrix largely defines the types of analysis appropriate to the matrix, and to some extent the sorts of questions it can answer. Similar data matrices thus serve to unite otherwise dissimilar fields. (Coombs, 1964, and Shepard, 1972, give fuller expositions of the properties of data matrices.)

Sociometry and data structure. Sociometric data naturally falls into what Coombs (1964) calls an intact matrix form, wherein the set of subjects is the same as the set of objects. A sociometric data matrix may be binary or real, but it is almost always conditional in its raw form. These characteristics of intactness and conditionality extend the definition of sociometric data and relate it to other studies of proximity, especially those involving taxonomic analysis.

So far, most of the efforts to link sociometry to other areas of research have been based on shared subjects of investigation, rather than on similarity of research methods. Links between sociometry and psychometrics, for instance, have come about as investigators have tried to find personal correlates of sociometric choice status, i. e., number of choices received from others. The motivation for the linkages is that both techniques are being used to investigate the same group of respondents. Thus, sociometric studies have been combined with studies of reputation, acceptance, status, physical environment, etc. Cronbach and Gleser (1953) review methods of profile analysis, another way of viewing the sociometric problem where multiple-scale instruments are used.

Sociometry can be categorized in three ways. To say that sociometry is the study of preferential interhuman relations indicates the content of many sociometric data sets, but not the data of studies of functional association, such as in studies of invisible colleges, informal working groups of scientists, etc. To define sociometry in terms of conditional intact data matrices establishes links among sociometry and other scientific fields.

There are, however, several studies of social structure which do not use an intact matrix, but use an off-diagonal matrix, in which groups of subjects are formed on the basis of sharing common properties. Phillips and Conviser (1973) clustered a set of people into subgroups using a proximity measure based on their shared activities. A reasonable way to capture this aspect of sociometric investigation is to say that sociometricians are often interested in forming clusters of people on the basis of some relation among them.

## Sociometric Data Collection Methods

In the years since Moreno and his associates first developed the original sociometric technique, many new methods for collecting and analyzing sociometric data have been developed. Sociometric questionnaires, although often concerned with interpersonal preferences, have also been used to investigate working habits, sources of information, exchange of help, and other group activities.

Sociometric questionnaires usually ask for direct choices, for rankings, or for ratings of others in the group. Each type of questionnaire can be further classified into those with a limited number of choices and those which allow an unlimited number of choices. Respondents may be asked to list the names of people they would like to associate with or to share some activity with, or may be presented with a roster of group members and asked to rate or rank some or all of them according to one or more criteria.

Direct choice. The respondent's original task was was to list those group members with whom he preferred to engage in some activity. The activity was generally chosen so that choice of partner would reflect the chooser's overall evaluation of the chosen person. One of the major variations in sociometric technique is in the imposition of a limit on the number of choices the respondent may make. Sometimes the limitation of choices is extreme, as where only the most- and/or least-preferred associate may be listed (Fishbein, 1965). More commonly, a respondent is asked to rank all of the group members according to a criterion.

There has been much debate over the relative merits of limited versus unlimited choices and of forced rankings as investigative techniques. Eng and French (1948) found that a rank order derived from paired comparisons of group members correlated more closely with rankings when there were unlimited choices rather than when choices were limited. Lindzey and Byrne (1968), in their extensive review of the sociometric literature, concluded that unlimited choices reveal more information about the status of any particular member, and about the total group structure than do limited choices. However, they also conclude the exigencies of time, statistical analysis, and research design often make it preferable to use a limited choice technique. Holland and Leinhart (1973) argue convincingly that limiting the number of choices a respondent is allowed to make distorts the group structure as revealed by the data beyond the ability of structural models to recover it. They suggest that only specialized unlimited-choice procedures will recover unbiased sociometric data.

Despite the prevalence of limited choice instruments in past research, some strong arguments can be made for the use of an unlimited number of choices and or roster instruments which present each respondent with a list of the entire group.

Rating scales. One of the major reasons for limiting the number of choices is that otherwise the ranking task becomes unduly difficult and time consuming. In large groups, especially those in which an individual may have only the most limited information about many of the other members, it is not possible for a member to make a meaningful rank ordering of the entire group. However, there is no reason why so difficult a task should be imposed upon the respondent. In many cases, the sorting of the group into a limited number of ordered classes will yield a satisfactory quality of data, while at the same time making the respondent's task easier.

One of the most common ways of imposing a weak ordering on a group is the use of rating scales. To rate each member on a seven-point scale is exactly the same as to order the group into seven classes, each consisting of all group members who receive a given rating. The same is true for such sorting techniques as the Q-sort. When a weak ordering is desired, a roster instrument has a number of advantages over a questionnaire which requires the respondent to write in a stated number of choices. A roster allows each respondent to be presented with an opportunity to evaluate every other member of the group. If the respondent is free to make as many choices as he desires, each entry on the roster carries information, even where the respondent has chosen not to rate a particular individual. It is safer to interpret missing data in a roster instrument as representing individuals whom the respondent did not know well enough to rate, especially where an unlimited number of choices is involved. This may be especially true toward the middle of a rank order where it becomes difficult to assign discrete ranks to those who are simply not very salient on the dimension being evaluated.

A roster questionnaire was used by Roistacher (1971) in a sociometric study of high school boys. Roistacher found that the eighth grade boys in his study responded enthusiastically, and provided data of high quality. Singleton (personal communication) found that third grade students were able to read the names on a roster instrument and also responded well to a request to rate classmates.

Although a number of scaling techniques are beginning to be used in sociometric research, by far the most common method for generating numbers is to use indices derived from simple ordinal rating scales. Scales on sociometric instruments have had from two to as many as one hundred points.

In addition, scores for sociometric nominations have been constructed from sums or differences between scales from the same instrument. There have been some attempts to use versions of the semantic differential (Davis and Warnath, 1957; Lerner, 1965; Fishbein, 1963, 1965), but the procedure of summing ratings on the various scales such as good-bad must be done with care, since it is easy to obtain "artifacts" when rankings and rating scale values are treated as if they were real numbers.

Other methods of scaling have included adjective check lists, (Davis and Warnath, 1957; Lerner, 1965; and Newcomb, 1961). A Q-Sort was employed by Peterson, Komorita, and Quay (1964) in which a series of sociometric statements were placed in a seven-point forced normal distribution in terms of their appropriateness in describing each individual.

There have been few successful attempts to go beyond the use of rankings and simple ordinal scales in the collection of sociometric data. Gardner and Thompson (1956) constructed a set of scales based on forced-choice comparisons of group members with respect to their capacity to satisfy certain of Murray's needs. Analysis showed that the scales so produced were fairly reliable and had approximately equal interval scores. Pepinski, Siegel, and Vanatta (1952) used an approximation of Thurstone's method of equal intervals to measure participation in a group activity. After much work they produced a group of twenty-four items which represented eight scale positions with each of three items.

## Finding Sociometric Cliques and Subgroups

From the beginning, sociometric investigators have been interested in discovering and then representing the structure of a group. Group structures are defined in terms of individuals' being connected by stronger or weaker links, or being in some sense "closer to" or "more distant from" each other. A major research task has been to divide the population into cliques, groups of individuals who have significantly more relation to each other than they do to other members of the group.

The original relationship was individual choice, a clique being defined as a group of individuals each of whom prefers the company of at least one other member of the group in a specified activity more than the company of anyone outside the group. Festinger, Schachter, and Back (1950) defined a clique as a subsystem of three or more elements in mutual interaction with each other. Luce and Perry (1949) formalized the graph theoretical notion of a clique as a maximal complete subgraph, a set of completely linked points not contained in a larger completely linked set. Luce (1950) generalized his definition of a clique to the n-clique, in which individuals within n links of each other were treated as directly connected, and cliques then extracted from this n-graph.

A major trend in sociometric analysis has been the generalization of group membership criteria from mutual choice to other relationships. In general, cliques may be defined on the basis of any relation among, or attribute of the members of the group being studied. A relation is defined between two members, while an attribute is defined on a single individual.

For example, a group of scientists could be clustered into cliques on the basis of how frequently they associate. Alternatively, they might be clustered according to some index of similarity in the work they were doing. Although the two sets of cliques might have quite similar memberships, the former set of cliques is defined on the basis of a bilateral relation while the latter set is formed on the basis of a shared attribute. The former example is more sociometric, while the latter example is traditionally a problem in taxonomy.

Most clique detection methods attempt to form a partition of the sociometric group i. e., to assign individuals to cliques such that each individual is a member of one and only one clique. Partitions are seldom obtained without doing some violence to the data, since a group seldom divides into completely disjoint subgroups. Davis (1967) found that groups did not partition into two disjoint cliques, and he developed structural indicators of a group's tendency to develop cliques. Davis (personal communication) later suggested that partitioning of sociometric groups is a false goal, since partitioned groups are as rare and as suspicious as correlations of 1.00. A real partition is usually a sign of conflict, the most obvious case being the partitioning of people into sides in a war.

Sociograms. The original method of analyzing sociometric data was the construction of a sociogram, a diagram of points and lines representing individuals in the sociometric group and their choices of other group members. The sociogram has proved to be a compelling way of representing group structure, especially to the group's own members. The sociogram's major defect is that its assumptions are not explicit and that it seems to convey more information than it actually carries. It is used as a two-dimensional picture of a group's structure. The relative position of points represent individuals' "positions" with respect to each other and the lines represent their choices of each other. Unfortunately, there is no a priori reason why a group's structure should be representable in the two-dimensional Euclidian space provided by a sheet of paper. In general, if there are more than three people in the group, it is impossible to represent the majority of the mutual relations possible to them as distances in two dimensions. Sociograms suffer from not being mathematically "well defined," i. e., there is no one-to-one relation between a set of sociometric choices and a given sociogram. In fact, from a given set of responses an infinite number of diagrams may be constructed which could convey widely diverging impressions of the group's structure.

A number of workers have suggested schemes for the construction of more invariant representations of sociometric structure. Northway (1940) described a "target sociogram" in which a set of concentric circles is used as a ground to indicate the choice status of individuals. Overchosen individuals are placed toward the center of the target, while underchosen or low-rated individuals are distributed around its periphery. The target sociogram is extensively used by elementary school teachers, who seem quite pleased with the representation of choice structures which it gives. Borgatta (1951) suggested that sociograms be drawn so as to minimize the number of crossed lines. Techniques such as these may prove useful when adequate methods have been developed for representing group structures in graphic form. At present, there seems to be more payoff in developing better algorithms for representing social structures than in trying to draw sociograms by intuition. Although the sociogram is universally regarded as the best final representation of group structure, it is a poor tool for determining that structure.

Linkages and distances. The techniques used to determine group structures from sociometric data can be divided into methods of linkage analysis and methods of distance analysis. The model underlying distance analyses is that the set of people can be regarded as a set of points in a geometric space, some of which are "closer" to each other than others. The analytical process consists of determining the dimensionality of the space, finding (and naming) the principal dimensions, and determining the configuration of the points in the space. The task of finding structures and subgroups consists of constructing a rule for determining closeness, and determining which of the points (individuals) are mutually close enough to be regarded as cliques. Factor analysis, multidimensional scaling, and the various cluster analysis and taxonomic methods are all examples of distance analyses.

The model underlying linkage analyses is that the sociometric group can be regarded as a set of points, some of which are joined by lines representing a relation. The major advantage of linkage analysis is its freedom from assumptions about unknown or undefined distances. In a distance analysis, there is a distance defined between any pair of points, whether that distance is known or not. In a linkage analysis, points either are or are not joined by a line, and the model is not necessarily incomplete because some points are not connected. In a distance analysis, two unrelated points may be described as infinitely distant from each other, but often this is "not as satisfactory" as simply being able to say that the points have no relation defined between them. Linkage methods, which are not as common or as numerous as distance methods, include the graphic analysis of sociograms, the statistical analysis of frequencies, and graph theoretic procedures.

Linkage Methods

Matrix permutation. Since the work of Forsyth and Katz (1946), almost all work on the analysis of group structure involves the matrix representation of sociometric data. Like the sociogram, the sociomatrix is not mathematically "well defined" since members of the group may be arbitrarily assigned to rows and columns. However, the number of matrix representations of a sociometric data set is at least countable, which is more than can be said for the number of possible sociogram representations of the data. The earliest work on sociomatrices, by Forsyth and Katz (1946), was on ways of rearranging the matrix to bring mutual choices closer to each other and closer to the main diagonal of the matrix. Beum and Brundage (1950) developed an algorithm for bringing the matrix to this canonical form and Borgatta and Stolz (1963) wrote the algorithm into a computer program.

Spilerman (1966) constructed the most sophisticated of the matrix permutation techniques, which eliminated the octopus-like arms produced by earlier methods when cliques overlapped. Spilerman duplicated individuals so that they appear in simple chains with everyone to whom they are linked. The duplication makes the matrix look as if it is composed of nonoverlapping cliques. Thus all cliques consist of adjacent individuals, so that the area of the main diagonal is tightly packed. Spilerman's strategy was not to make a priori definitions of cliques, but to rearrange the sociogram so as to allow the investigator to make better a posteriori partitions of the sociometric group. Spilerman's method is in marked contrast to the more usual strategy of attempting to determine the power and generality of a priori rules for clique detection.

<u>Matrix multiplication.</u> The most common matrix operation in sociometric analysis is matrix multiplication. If A and B are matrices, and C=AB, then c(i,j), an entry in C, is defined as
c(i,j)=SUM(k=1,n) a(i,k) X b(k,j).

Luce and Perry (1949) proved several theorems which related properties of matrix powers to properties of group structures. When a binary sociomatrix, C, consisting only of 1s and 0s, is multiplied by itself, c(i,j) contains the number of sequences of length 2 between members i and j. Entries in the diagonal show the number of reciprocated choices made by a member of the group. Similarly, up to k=3, the k-th power of the matrix has entries containing sequences of length k between members; in particular, cells on the diagonal contain cycles of length k, i.e., sequences which both begin and end with the same person. For powers greater than three, the values in the matrix become periodic, because the sequence a, b, c of length 2 shows up in the fourth power of the matrix as the sequence a,b,a,b,c of length 4.

These clique detection methods work only for nonoverlapping cliques. Luce and Perry (1949), and Harary and Ross (1957) outlined some iterative methods for finding multiple cliques in sociomatrices, but these methods have been superseded by computer algorithms (Augustson and Minker, 1970; Purdy, 1973) for generating all of the cliques in a graph, even where cliques overlap.

Lingwood (1969) has developed a technique which is in the spirit of matrix multiplication but requires considerably less computation. The technique consists of beginning with a single member of the group and tracing all direct connections to other members. A "1" is then placed in the vector of group members for each directly connected member. The direct connections of each member in the vector are traced and a "2" is entered into the vector for each of these connections. The process is continued until all those who are connected to the starting individual have been reached, and the length of the path added to the vector.

The most elegant matrix multiplication scheme is Hubbell's (1965), which uses Leontief's (1941) model as an analogue for communication inputs and outputs by group members. Leontief's original model considers N factories, each of which uses the products of itself and the other N-1 factories to produce a product. The model assumes that for any given level of inputs a factory will produce a given level of output, and the N factories are represented as a series of N linear equations. Among the results derivable from the model is that given a set of initial conditions, the N factories will reach an equilibrium rate of production, in which the total amount of each good produced will equal the total amount of that good consumed by the n factories. In addition, the model allows the effects of changing the initial or boundary conditions to be predicted.

Hubbell treated the giving and the receipt of sociometric choices as analogous to the inputs and outputs of the factories in Leontief's model. In Hubbell's model, inputs to a given member need not, as in the Leontief model, equal the output from that member, and choices may have both positive and negative weights. Hubbell stated that "where negative lengths are present...[c]lique-mates can ... be allied in part through a shared antipathy for a third party."

Hubbell assumes that status is the equivalent of an output in the Leontief model. Inputs are sociometric choices, weighted by the strength of the choice and the status of the chooser. Inverting the choice matrix, the equivalent of summing all powers of the matrix, yields a matrix whose cells are contributions of member j to the status of member i and whose row totals are statuses for each member of the group. The inverse of the matrix is more than a set of status weights. The i,j-th cell of the matrix contains the effect of a unit change in the status of individual j on the status of individual i given the present group structure. Thus, the inverse of the matrix can be multiplied by a vector of statuses for each member to investigate the dynamics of status in the group. Partitioning of the group into cliques is done by using as the measure of association between i and j the minimum of the contribution of one to the status of the other, i.e., min[a(i,j), a(j,i)].

Hubbell points out that these measures of association can be rank ordered and a cutting point parameter chosen at a level which will eliminate any specified proportion of the links in the group. Thus, cliques of varying degrees of density may be formed and examined. In addition, various boundary conditions may be introduced and the effect of boundary condition changes on clique structure may be investigated.

<u>Matrices and graphs.</u> The matrix techniques discussed above represent a major attempt to formalize a powerful, but largely intuitive, method for describing social structures. The classical way of using mathematics in science is to associate things in the real world with mathematical concepts, to operate on the concepts using the content-free rules of mathematics, and then to investigate the relation between the results of the mathematical operations and the real world. In the case of sociometry, the "things" were links between individuals, and the mathematical concept was the matrix. There has been some degree of success with the operation of matrix multiplication, but overall, linkage analyses have not had

much relation to the body of mathematics in which matrix theory is imbedded.

Graph Theoretic Methods. Graph theory is a branch of mathematics concerned with abstract configurations of points and lines. Harary, Norman, and Cartwright (1965) define a directed graph in terms of four primitives and four axioms. The primitives defining a graph are:

P1: A set V of elements called "points."

P2: A set X of elements called "lines."

P3: A function f whose domain is X and whose range is contained in V.

P4: A function s whose domain is X and whose range is contained in V.

The axioms defining a graph are:

A1: The set V is finite and not empty.

A2: The set X is finite.

A3: No two distinct lines are parallel.

A4: There are no loops.

A graph looks like a sociogram, and it has been suggested that theorems derived from graph theory may be useful in the analysis of sociometric data. Sociograms which differ only in the physical placement of their points on paper are often regarded as different, while graph theory explicitly regards any representations of equal numbers of points connected by similar arrangements of lines as equivalent.

Bavelas (1948, 1950) first applied the notion of abstract graphs to the analysis of communication nets in small groups. Since then, there has been an increasing amount of work on the application of graph theory to group structure and to the theory of cognitive balance. Workers in mathematics, psychology, and operations research have all made contributions in applying graph theory to the analysis of social structures. Berge (1963) and Flament (1963) provide formal introductions to the mathematical theory of graphs, with Flament's treatment being more slanted toward social science applications. Ore (1963) provides a general, but less formal, introduction to the subject. Harary, Norman, and Cartwright (1965) provide the best introduction to the theory of directed graphs for the social scientist. (Applications of graph theory can also be found in the literature of operations research, where it is used in describing flows in networks; and in computer science, where it is used in describing data structures.) Graph theory has developed many useful definitions of configurations of social structures. Algorithms have been developed for locating the cliques of a graph (where cliques are defined in several alternative ways), and for finding the degree of connectivity or fragmentation of given parts of a graph.

Some graph theoretical models use numerical information about links between objects to generate networks, which are graphs whose lines have numerical values. Networks may be isomorphic to sociometric structures (cf. Ford and Fulkerson, 1962; Harary, Norman, and Cartwright, 1965), but as Gleason (1969) has pointed out, network models generally require that the quantity assigned to a path be some combination of the quantity assigned to its constituent lines. To perform mathematical operations such as addition on these numbers often requires stronger metric assumptions than can safely be made with sociometric data. In fact, even to take the maximum or minimum value of a set of lines is risky where it requires comparison of ratings across subjects. Without additional measurement, it is impossible to say that a likes b more than c likes d from a set of paired comparison data or rank orders.

The two major applications of graph theory in sociometry have been in defining cliques, and in defining relations among individuals and among cliques in terms of structural balance. The first graph theoretic definition of a sociometric clique was suggested by Luce and Perry (1949), who defined a clique as a maximal complete subgraph, which is a set of points completely connected to each other and not part of a larger, completely connected set of points. The maximal complete subgraph definition is rigorous and useful, but has proved too strict for many purposes. Alba (1972) used Luce's (1950) invention of the n-graph to provide a more tractable graph theoretic definition for cliques. An n-clique, according to Luce, is one in which each member is connected to every other member by a path of n or fewer lines. Alba introduced the additional condition that the longest path, the clique's diameter, must be of length n, for otherwise, members of an n-clique can be connected through individuals who are not clique members.

This definition is more inclusive of structures which "look like" cliques, but do not meet the strict requirements of Luce and Perry's (1949) original definition. It should be noted that none of this family of definitions necessarily partitions a group into cliques.

Most of the graph theoretical work on sociometric structures has involved extensions of Heider's (1946) concept of structural balance. Balance theory concerns signed graphs representing relations between people and objects. The lines of the graph can take on the values "+" and "-," representing positive and negative relations between the pairs of points. A positive relation might be "likes," or "trusts," while a negative relation might be "dislikes," or "mistrusts." Heider defined a relation on a triad of three points connected by three lines as balanced if there were an even number of lines with negative signs.

Much work in graph theory has been concerned with extensions of the concept of structural balance to structures of more than three points. Harary (1955) defined a balanced graph as one in which all cycles were positive. Flament (1963) proposed an alternative definition of a balanced graph as one in which all triads were balanced, and proved that, for a complete graph, his definition was equivalent to Harary's. Since balance is an extreme condition for a graph, especially one of a real group or organization, measures of degrees of balance were developed. Cartwright and Harary suggested that a cycle, a set of lines beginning and ending at the same point, be called balanced if it had an even number of negative lines. The relative balance of a graph was the ratio of balanced cycles to the total number of cycles in the graph. It seemed reasonable that the balance of longer cycles should not be as heavily counted as that of shorter cycles in determining the balance of the graph as a whole. Harary (1955) suggested the measure of N-order balance, which is that all cycles of N or fewer lines be balanced. Cartwright and Harary later refined this into the concept of relative N-balance, the ratio of balanced cycles of length N or less, to the total number of cycles of length N or less.

Norman and Roberts (1972) derived a measure of relative balance from a set of axioms used in measurement theory. Their measure of the balance of a graph, G, is

SUM(m=3,inf) [a(m) f(m)] / SUM(m=3,inf) [b(m) f(m)]

where a(m) is the number of balanced cycles of length m, b(m) is the number of unbalanced cycles of length m, and f(m) is a monotone decreasing, real-valued function. Notice that where f(m) is identically 1, the measure is the same as that of Cartwright and Harary. Norman and Roberts show that their measure gives a balance ordering among graphs which is unique up to a multiple of f(m), and that their measure has many of the desirable properties found in empirically derived measures of balance. Cartwright and Harary (1956) extended the theory of balance to a graph with an arbitrary number of points, and showed that a definition of balance implied that the graph would be partitioned into exactly two opposing cliques.

Davis (1967) pointed out that a balanced graph is an ideal type, since groups consisting of only two opposed factions are seldom found, and are hardly in a desirable state of affairs. Davis proved a set of theorems giving conditions for the partitioning of a graph into two or more clusters. A cluster is a set of points connected to each other by positive lines, and to points in other clusters only by negative lines. A graph which can be split into clusters is known as clusterable, and the partition into clusters is known as a clustering. Davis' clustering theorems are based on the condition that every cycle--undirected path beginning and ending at the same point--in a graph must contain exactly two negative lines if the graph is clusterable. If the graph is complete, a unique clustering is obtained; otherwise several clusterings may be found. Davis gives a classification of tendencies, which may have analogues in social relations, toward clusterability in incomplete graphs.

Cartwright and Harary (1968) related the partitioning of points in a signed graph to the classical mathematical problem of coloring graphs. They defined a graph as "colorable" if it was possible to partition its points into "color sets" in a fashion similar to Davis' clustering. Gleason and Cartwright (1967) developed an alogrithm which determines the (possibly unique) colorability of a graph from its adjacency matrix.

Peay (1970) has proposed a set of criteria which would allow specification of degrees of colorability and balance in graphs rather then having to use all-or-none criteria which could be violated by a single instance. A triad is balanced when two individuals connected by a positive path (of length 2) are not connected by a negative line. Peay's criterion for a graph's tendency toward colorability is that points connected by positive paths of less than some critical length not be connected by a negative line. A colorable graph is one for which the property holds for a path of any length. Peay devised a set of techniques for determining tendencies toward colorability in binary and scalar matrices, and proposed a method of statistical evaluation for degrees of balance.

The graph theoretic work described above has been concerned with structures, such as cliques and clusters, which extend over several pairs of individuals.

Even though the definitions of cliques and balanced structures have been extended to include n-cliques and structures with tendencies toward balance, the probability of colorability, clusterability, or balance being violated in an empirically obtained graph is considerable, especially where there may be missing data.

The difficulty of finding extended structures in empirically derived graphs has led to an alternative approach, which is to view a graph as composed of all its many triads of individuals, rather than to aggregate pairs of individuals into more extended structures. A directed graph of mutual and asymmetric choices can be composed of sixteen types of triads. Most of the triads are distinguished by having different combinations of lines connecting their members, while some with asymmetric choices are distinguished by the directions of the asymmetric choices. Holland and Leinhart (1970) reported a theorem for computing the expected numbers, variances and covariances of the sixteen types of triads in a graph of known size, given the observed total distribution of mutual, asymmetric, and null choices for all pairs. This distribution provides a means of statistically evaluating the tendency of a graph to deviate from randomness. Holland and Leinhart used their results to evaluate the transitivity of relations in several graphs, a property which includes balanced and hierarchical relations as special cases.

A choice relation on individuals a, b, and c, is transitive if, when a chooses b, and b chooses c, then a chooses c. The arithmetic relations denoted by "=" and "<" are other examples of transitive relations. Seven of the sixteen possible triads in a directed graph are intransitive while the other nine are either transitive or vacuously transitive. A triad is vacuously transitive when it does not fulfill the "if" condition for transitivity. In particular, a triad in which fewer than two members make choices is vacuously transitive. If a relation of no choice between a pair is defined as equivalent to a negative choice, then a colorable graph will have no intransitive triads.

There are logical relationships between the structure of a graph (macrostructure) and the types of triads (microstructure) which will be found in it. Flament (1963) proved that a balanced graph will have only balanced triads in it. Other models of social structure, such as colorability, and Davis and Leinhart's (1970) ranked clusters model, imply that particular triads will be absent from a graph. In a perfect case of conformity to a structural type, no forbidden triads would be present. The degree to which the prescribed and forbidden triads are present indicates the degree to which the graph conforms to the structural model. Davis and Leinhart (1971) constructed an index showing the extent to which nonpermissable triads were present in a graph, but had no way of testing the statistical significance of the index. The theorem of Holland and Leinhart (1970) allows a statistical test of a graph's conformity to a structural model.

Davis and Leinhart proposed a ranked clusters model of social structure, in which cliques are ranked into several status levels. Individuals are connected to fellow clique-members by mutual choices, and to members of other cliques at the same status level by null choices. Members of lower status cliques are connected to members of higher status cliques by unreciprocated choices. In a graph with a ranked clusters structure, all triads will be transitive. In addition, one vacuously transitive triad should be absent, one with a single nonreciprocated choice. In a triad, a,b,c, the fact that a does not choose b shows that a and b are members of different cliques at the same status level. The same is true of b and c, who are also connected by a null choice. However, the choice of c by a implies that c is at a higher status level than a, thus contradicting the effect of the other two (null) choices.

Holland and Leinhart (1970) proposed a more general model of transitivity in social structure, of which the Davis and Leinhart model is a special case. The ranked clusters model assumes a single status hierarchy. Where there are multiple status hierarchies, the triad with a single asymmetric choice is not inconsistent. A common situation in which there are multiple status hierarchies is in classrooms, where there is a "sex cleavage" between boys and girls. Holland and Leinhart showed that there was a higher than expected number of these triads in sexually heterogeneous groups, and a lower than expected number in sexually homogeneous groups.

Davis (1970) investigated the contribution of families of triads to structural balance and hierarchy in graphs. Davis suggested that asymmetric choices might indicate relations which were positive, but not so positive as mutual choices. In Davis and Leinhart's work on the ranked clusters model, certain forbidden triads were far more prevalent than other equally forbidden triads. Davis investigated the effect of classifying asymmetric relations as if they were positive, and as if they were negative. He found that the classification of mutual choices as positive and mutual null choices as negative was consistent with clusterability in all groups. In small groups, a better fit to the clusterability model was found by classifying asymmetric choices as negative, while in large groups a better fit was found by classifying them as positive. Davis suggested that the collection of data in a small group was more likely to be one of unrestricted choice, while in a large group, restriction of the number of choices would make a null choice less of a rejection.

Davis also found that where triads consisted of three asymmetric choices, the number of cyclic triads was less than would be found in a random graph, indicating that where there is a hierarchy of choice, choices tend to be arranged in transitive patterns up the hierarchy.

Linkage Frequency Methods. Some of the earliest work on the partitioning of sociometric data has been with relative frequencies of choice within and among potential cliques. An excellent introduction to the statistical analysis of linkage patterns may be found in Moreno et al.'s Sociometry Reader. After a false start, in which the probability of one person's choosing another was viewed as an independent binomial event, more realistic characterizations of the statistical nature of the sociogram were developed.

Katz and Powell (1960) established a basis for the statistical investigation of directed graphs by elaborating the sample space of a directed graph. They stated that the sample space of a graph is not based on the presence or absence of a single line between a given pair of points, but is rather composed of sets of graphs with the same total number of points and lines, or of graphs with the same number of lines arriving at and/or leaving a given point. This work makes it possible to evaluate such things as the degree to which a group contains more or fewer isolates than would be expected by chance.

Most of the statistical work on graphs has involved comparison with a totally random graph, a somewhat rare creature in real life. However, it seems likely that statistical methods worked out with random comparisons can also be used to evaluate differences between nonrandom graphs, once good graph metrics are worked out.

Proctor (1967) gives a formula and procedure for estimating the variance of the linkage density in a sample of group members. Criswell (1949; cited in Proctor and Loomis, 1951) constructed a chi-square test for determining the relative degree of ingroup choice of a subset of a sociometric group.

The distributional characteristics of sociometric networks were investigated in a series of three papers by three overlapping sets of workers at the University of Michigan's Mental Health Research Institute. The data for all three papers is a set of up to eight ordered sociometric choices made by each of 956 students in a junior high school. Rapoport and Horvath (1961) derived a distribution function for the number of votes an individual received as n-th (where n ranged from one to eight) friend. A good fit to the distribution of the number of choices as n-th friend was given by a negative binomial distribution function which the authors derived in two ways. If friendship choices are considered independent events, then the negative binomial distribution is derivable from a distribution of Greenwood and Yule (1920) which compounds a Poisson distribution for the number of choices received, and a Pearson Type III distribution for the expected number of choices for an individual. However, the same negative binomial distribution function is derivable from a Polya process, a time dependent model of contagion, which assumes that the probability of an event occurring in a small present interval of time increases linearly with the number of events which have occurred up to now, but decreases overall with the passage of time. Thus, the same distribution of votes to an individual occurs when votes are seen as independent events or as stochastically dependent. Rapoport and Horvath showed that each person had a characteristic "popularity intensity" which influenced the number of votes he received as a friend at any level of choice. There was no such thing as a person who was characteristically a "second best," or "n-th best" friend, rather than a generally popular or unpopular person.

Work begun by Rapoport and Horvath on tracing the sociometric net was continued by Foster, Rapoport, and Orwant (1963). The tracing process consisted of randomly selecting a small set of "starters," deciding on the number of choices to be traced from each person, and adding the names of previously unchosen persons to the set as they were reached by successive generations of choices. The authors developed a recursion equation which gave the cumulative number of people who should be chosen if choices were made at random. The authors made several empirical tracings of their large sociogram, which they compared with the data derived from a hypothesis of a random network.

The empirical tracings showed a significantly smaller rate of addition of new people than did the "random net" tracing. Foster, Rapoport, and Orwant showed that the bias in the empirical tracings could be assigned to two sources, which they called a parent and a sibling bias. A parent of individual A was defined as someone who chose A. Two individuals were defined as siblings if they had a common parent. The bias in the tracings through the network were found to be attributable to tendencies to choose parents and siblings at a higher than chance rate. Since parents and siblings are already members of the tracing, choices to them do not result in the addition of new members to the tracing. The authors computed the parent and sibling bias probabilities in their empirical data and developed a new recursion equation which included the two probabilities as parameters. A good fit to the data was obtained without having to include a parameter for bias resulting from unreciprocated choices of generally popular individuals.

Foster and Horvath (1971) investigated the distribution of reciprocated choices and suggested the probability of a reciprocated choice as a measure of social distance. They developed a model which treated the reciprocation of a choice as a sequential process in which the best friend is chosen first, then the second best, etc. The authors developed a linear equation for the probability of a mutual choice between an i-th and a j-th best friend.

The probability of a choice's being reciprocated decreases as each person gets farther away from his choice of best friend. The two empirically derived constants in the linear equation were related to the asymptotic probability of being chosen at all, and to the width of individuals' circles of friendship.

These contributions link the study of sociometric networks to the study of generalized stochastic networks and graphs, work which is also pursued in physics, communication theory, and information theory. The treatment of sociometric choicees through a network as a stochastic process links the study of sociograms to the study of self-propagating processes, such as rumors and epidemics.

Lingwood (1969) used a chi-square statistic to test the equivalence between cliques determined on the basis of information gathering, and cliques determined on the basis of research methodology, in a population of communication researchers. Lingwood's hypothesis was that the frequency of choice within and between cliques determined by one method would predict the relative frequency of choice within and between cliques determined by the other. Lingwood concluded that the chi-square method was successful once the cliques had been discovered by other methods. Until the advent of computers, it was far easier to test the significance of a partition into cliques than it was to make the partition. Trial-and-error was used to split the group so as to maximize the ratio of within-clique to between-clique linkages. There now exists a whole array of techniques for "hunting" through a set of data and subsetting it in a way which maximizes some measure of optimality. (E. g., Sonquist, 1970)

Work on the distributional characteristics of networks and graphs is leading to the development of measures of difference between sociometric structures and of measures of the conformity of a graph to a structural model. Proctor (1960) discusses the probabilistics of finding a partition for a group which best fits the group's set of choices. Boorman and Olivier (1973) derived several metrics on a space of finite trees, which may be used for such things as determining the relative degree of similarity among a set of hierarchical social structures. The tree metrics are closely related to partition metrics, which yield measures of similarity between partitions of a set.

## Distance Methods

Distance methods of analysis are characterized by the assumption that the sociometric group can be represented as a set of points distributed through a geometric space. The idea of relations among individuals and groups being expressed in terms of social distance is a plausible one. In addition, there is a large and well-known body of mathematics dealing with points and spaces, which finds wide use in the sciences. The main philosophical problem in distance analyses concerns the validity of the spatial metaphor for human relations, rather than the validity of mathematical operations once the spacial analogy has been made. Almost all statistical methods are examples of distance analyses.

Some properties of spaces. A space in its most basic form is a set of objects called points, and is defined by specifying the points in it. For instance, if we are observing the height, weight, and sex of a group of subjects, our observations define a three-dimensional space. Each point in this observation space is a triple (h,w,s). Among the properties of this observation space is that it has no negative regions, since no one has a negative height, weight, or sex. Other properties are that two of its dimensions are expressible as positive real numbers, while the third dimension is expressible as two integers, e. g., 0 and 1.

The most important forms of spaces for measurement purposes are metric spaces, on which a metric, a measure of distance, has been defined. A metric must satisfy three properties:

> it must be either 0, or greater than 0, and if the distance between two points is 0 then they are the same;
>
> a metric must be symmetric, i. e., the distance from A to B must be the same as the distance from B to A;
>
> a metric must satisfy the triangle inequality, i. e., the distance from A to B plus the distance from B to C must be greater than or equal to the distance from A to C.

Any function of two points in the space satisfying these three criteria will serve as a metric on the space. Metric spaces are important because they are civilized. Not only do they behave in ways similar to our perceptual space, but they are mathematically tractable. Locations and distances in metric spaces tend to behave as we expect; a large set of measurement techniques are available for application to a set of points in a metric space. Most statistical techniques are valid only for data which define a metric space. One of the most important tasks in dealing with nonmetric data is to find some way of converting the data to metric form.

Consideration of some of the partitioning and index construction methods reveals that many measures of association will not satisfy all of the requirements of a metric. In most instances, the raw data is in the form of rankings or other ordinal measures. In the majority of sociometric data c(i,j) and c(j,i) of a choice matrix, C, are not equal, and thus violate the symmetry property of a metric. In some cases a function of the two cells is used to replace the values, thus symmetrizing the matrix.

The Euclidean space is one whose metric is the familiar hypotenuse rule of plane geometry. Where a and b are points the distance between them, d(a,b), is: $d(a,b) = (a^2+b^2)^{1/2}$.

The Euclidean metric underlies most commonly used statistical methods, but workers in the areas of scaling and measurement (Coombs, 1964; Torgerson, 1958) have used metric spaces of a variety of types. One commonly used generalization of the Euclidian metric is the generalized Minkowski metric in which $d(a,b) = (a^r+b^r)^{1/r}$. The properties of the Minkowski metric can be altered by adjusting the value of r. When r is set equal to 2, then the Minkowski metric is a Euclidian metric. When r is set to 1, then the Minkowski metric becomes a "city block" metric, in which points equally distant from an origin form not a circle, but a square standing on one corner.

Many techniques for partitioning sociometric groups into cliques assume that group members may be represented as points distributed in a space of known type and metric, but of unknown dimensionality. Individuals are points, and the entries in the sociomatrix are perceived proximities of the individual points to each other. The problem is to determine the configuration of the individual points in the space and to decide which individuals are sufficiently close to each other to be regarded as a clique.

Until recently, factor analysis was the most commonly used distance analysis technique. Factor analysis has been augmented by nometric techniques such as multidimensional scaling and cluster analysis, which use rank orders rather than raw scores as input. A promising area for finding sociometric techniques is numerical taxonomy, a field concerned with generating structures from matrices of pairwise association.

Factor analysis. The original use of factor analysis was by Bock and Husain (1950), who converted the original choice matrix into a matrix of correlations and performed a centroid factor analysis. The rationale behind the use of factor analysis is that a clique may be viewed as a set of individuals who tend to make similar choices, or, conversely, that a clique is a set of individuals who tend to be chosen by similar sets of others. When a sociomatrix is factor analyzed, each clique is represented as a factor, and the dimensionality of the space is equal to the number of cliques. An individual's membership in cliques may therefore be expressed as a set of factor loadings which define the strength of his association with each clique.

As computing techniques evolved, the more powerful varimax factor analysis was used by MacRae (1960) and by Wright and Evitts (1961) in the analysis of choice matrices. The results of these analyses are somewhat mixed, as there are separate sets of chooser and chosen cliques which must be reconciled. Beaton (1966) extended MacRae's model by treating the chooser and chosen structures as two batteries of variables and by performing an interbattery factor analysis. Nosanchuk (1963) compared direct factor analysis with several matrix manipulations and sociogram techniques for partitioning sociometric data sets. He found that factor analysis produced a better recovery of a preconstructed sociogram than did any of the other techniques. There have not been any published reports comparing the nonmetric techniques to factor analysis in recovering preconstructed sociograms. However, nonmetric techniques have proved successful in reconstructing other sorts of configurations of points, especially where the original configuration was subjected to a nonlinear transformation before being recovered.

<u>Nonmetric scaling.</u> One method of analysis which overcomes many objections to the metric assumptions of factor analysis is nonmetric multidimensional scaling. This technique, like factor analysis, requires as input a matrix of proximities, e. g., a correlation matrix. In nonmetric scaling, only the row-wise rank orders of the values in the matrix are used in the analysis. An iterative trial-and-error procedure is used to "jiggle" a set of points in a Euclidean space to satisfy the rank orders of proximities. The result of this procedure is that a set of points in a space of unknown metric and dimension is given an analogue of points in a Euclidean space of given dimensionality. Any of a number of powerful clustering techniques may then be applied. (An introduction to the field of nonmetric scaling may be found in Green and Carmone, 1970. A somewhat more advanced treatment may be found in Shepard, Romney, and Nerlove, 1972. Both of these works contain extremely complete bibliographies.)

Gleason (1969) used nonmetric multidimensional scaling to evaluate data obtained by Newcomb (1961). Gleason converted the preference data of a group of college men into a geometric representation with a nonmetric scaling program. The points thus obtained were then subjected to a Euclidean space hierarchical cluster analysis to determine cliques. The problem of satisfying the symmetry requirement for a metric was solved, as in factor analytic methods, by representing individuals as separate chooser and chosen points, since the choices were not reciprocal. Gleason suggested that an analysis of the difference between an individual's location as chooser and chosen might prove useful, but at present this disparity has not yet been related to other characteristics of individuals and groups.

Multiple discriminant analysis. Another technique which makes strong assumptions on the space is the use of multiple measurements and multiple discriminant analysis. Riffenburgh (1966) suggested that if k measurements are made on the individuals in each of p groups, each individual may be regarded as a point in a space of k dimensions. The centroid of each group in the space may be computed and distances between groups determined. By making some reasonable assumptions about the space, confidence intervals for the intergroup distances can be computed.

Riffenburgh computed a set of sociometric distances between five racial groups in Hawaii with a high degree of success. He pointed out that, in general, p+1 sociograms can be drawn for a set of p groups. Each group will provide a complete set of distances for the sociogram as well as contributing to an overall set of distances. Riffenburgh's approach takes its strength primarily from its reliance on multiple measurements, which allow the stronger metric assumptions which justify multiple discriminant analysis; his procedure seems well-suited to situations in which multiple measurements can be taken and has one of the most elegant measurement bases of any sociometric method.

Cluster analysis. To the extent that the problem of sociometric analysis is seen as one of locating and clustering points in a metric space, sociometry becomes a specialized application of numerical taxonomy. Lorr (1968), in a review of the literature of classification and typological procedures, describes four major types of clustering methods. (Most of the references in this section are from Lorr's 1968 paper, which is not generally available.)

What Lorr calls "complete linkage analysis" or "successive cluster build-up" uses a similarity cutting point criterion. The criterion is used here to specify that each member of the clique is closer to every other member than he is to anyone outside of it. Alternatively, the centroid of the clique may be used as an origin for the cutting point. The criterion for single linkage analysis is a rigorous expression of the original definition of a clique as a group of people who preferred or were closer to each other than to others. Single linkage analysis thus seems appropriate as a way of partitioning a group into disjoint cliques.

Single linkage analysis uses an arbitrary cutting point of similarity to define clique membership. Any members of the data set whose distance is less than the cutting point value are regarded as being in the same clique. Thus, every member of the clique is more similar to at least one other member of that clique than he is to anyone outside of it. Single linkage analysis is useful for tracing chains of individuals and produces extended, serpentine clusters. Single linkage analysis seems appropriate for analyzing networks which connect cliques once some idea of the clique structure has been obtained by a complete linkage analysis. (Fuller treatments of single linkage analysis may be found in McQuitty, 1957; Sokal and Sneath, 1963; Needham, 1961; Parber-Rhodes 1961; Cattell and Coulter, 1966.)

Successive cluster build-up occurs when the cutting point is allowed to increase and new members are added to the clique in what might be thought of as successive layers. (See McQuitty, 1963; Michener and Sokal, 1957; Sawrey, Keller, and Conger, 1960; Saunders and Schucman, 1962; Lorr and Radhakrishnan, 1967; Bonner, 1964.) Many of the Euclidean space hierarchical clustering programs are of this sort. They begin with a matrix of Euclidean distances between points, and construct centroids by taking root-mean-square distances between individuals already included in a cluster.

A third approach is to partition the space into subspaces containing subsets of points which are as tightly packed as possible. The algorithm to achieve compactness attempts to minimize the mean within-cluster variance of the points, shifting points between clusters to do so. (See Thorndike, 1953; Edwards and Cavalli-Sforza, 1965; Ball and Hall, 1965; McQuitty and Clark, 1968.)

Cormack (1971) gives a very complete and concise review of classification methods, including consideration of the clustering techniques discussed by Lorr as part of a more comprehensive and technical discussion. The reader is referred to Cormack's extensive bibliography for further reading on the subject.

Norman (1967) introduced a different approach to the minimum-variance clustering problem, using a technique which partitions the entire set of individuals into sets which minimize within-cluster variance while maximizing between-cluster variance. The partitioning process is iterated to yield a hierarchical set of increasingly smaller clusters. Norman's procedure is especially interesting because it permutes the matrix of distances in order to operate on submatrices, and thus combines aspects of both graphic matrix manipulation and computational methods in the partitioning of data sets.

Networks and covers.

A partition of a set requires that each member of the set be located in one and only one subset. For the sociometric investigator the partition has been a useful goal because it seemed the only way to prevent cliques from blending with each other into a hopelessly confused mass. Spilerman (1966) used a matrix manipulation procedure in which individuals were duplicated in order to show their simultaneous membership in a number of cliques. Spilerman's procedure illustrates the use of covers rather than partitions of sociometric sets. A cover of a set S is defined as a family of subsets such that every member of S is a member of at least one subset. The conditions for covering a set are considerably less rigorous than those required for a partition, which is a family of subsets of S such that each member of S is a member of one and only one subset. The difficulties of using a cover show up when a linkage frequency analysis is used. In general, linkage frequency methods require transitivity of the linkage in order to form cliques. If C(1) and C(2) are two cliques, and i is linked to members of each of them strongly enough to be a member of each, then transitivity will usually contradict the assertion that C(1) and C(2) are disjoint.

Guimearaes (1968) suggested that qualitative networks could provide a means of representing sociometric structures. A network is usually defined as a graph whose lines have numerical values, which may represent such variables as frequency, channel capacity, or intensity of linkage. However, a network might be defined as a graph with functional as well as structural information. A sociometric group could be connected by a variety of networks, i.e., an information network, a friendship network, etc. It might be possible to use a clique detection method to partition a single network in a very strong fashion. The union of a set of such partitioned networks would form a cover, since networks of different kinds would overlap, even where cliques of a single network were constructed so as not to overlap. It seems appropriate to use logical combinations of multiple measures before using arithmetic combinations, since logical combinations make far weaker metric assumptions of the data. (For reviews of network theory see Van Valkenburg, 1964.)

Abstract Algebras. Boyle (1969) used concepts from modern algebra to investigate clique structures without reference to metric assumptions about the data. He defines sociometric choices as mappings from the chooser to the chosen which form a transitive relation on the set of individuals. By making some assumptions about choices in hierarchies, e.g., that choices will be made only laterally or to an individual one level higher, he was able to make inferences about the structures of two hierarchical organizations.

## Construction of Sociometric Indices

After analyzing cliques and clusters, a second analytical task for the sociometric investigator is to construct scores for individuals or groups of individuals. We may think of scores as being of two general types: a set, I, of individual scores, and a set G of group scores. The mappings comprising these scores may be specified as

$I: S \times PS \rightarrow R$

$G: PS \times PS \rightarrow R$

where S is the set of individuals in the sociometric group, and PS is the power set of S, the set of all subsets of S. The individual indices are functions of individuals and of subsets of the sociometric group. Individual scores refer explicitly to a single individual, but also refer implicitly to some set of other people to whom the individual is related. An example would be the number of choices an individual receives from the group as a whole, or from some subset of the group. Group indices are functions of a single subset of the group and of one or more other subsets. An example would be the proportion of group members in a given group, or the number of other cliques in direct contact with a given group. The construction of an index is dependent on specifying not only the individual for whom the index is to be constructed, but also the set of individuals in which he will be embedded. (Proctor (1967) gives an example of some techniques required for hypothesis testing when data are derived from pairs of individuals as well as from single individuals.)

The most obvious index to compute for an individual is his row or column total in the sociomatrix, the number of choices he gives and receives. Several investigators (e.g., Jennings, 1950; Dunnington, 1957), have separated groups into relatively highly chosen versus relatively unchosen individuals, using a variety of criteria to construct a cutting point. Early work on the construction of group indices led to the construction of such indices as Moreno's "emotional expansiveness," the ratio of total choices made to the total number of choices possible in the group. Proctor and Loomis (1951) used as an index of group cohesiveness the number of mutual choices in the group divided by the maximum number of such choices in the group. The same authors provide a large variety of other arithmetic indices. Many similar arithmetic indices have been constructed by a number of workers, e.g., Norman (1953), McKinney (1948), Smucker (1949), and Beauchamp (1965).

Later work has taken into account not merely the direct choices between individuals, but also the indirect linkages. Katz (1953) used the sum of successive powers of the matrix, together with an attenuation factor for each power, which made an individual status a weighted sum of choices made at various distances. Taylor (1969) modified Katz's index by normalizing it according to the number of choices that the person gave, as well as received. The most elegant form of this type of status index is in Hubbell's (1965) model which considered choices given, choices received, and exogenous variables.

Lin (1970) has presented a number of individual indices which are easily computed but which make strong assumptions about the data's level of measurement. Lin begins by establishing the "influence domain" of an individual, defined as the total number of persons who directly or indirectly chose that person. The individual's centrality is defined as the mean distance between him and all individuals in his influence domain, thus treating linkages as equal intervals. Finally, Lin defines the prestige of an individual as:

$P(i)=I(i)/(C(i)*(n-1))$

if $C(i)\neq 0$; $P(i)=0$ otherwise. $I(i)$ is the size of i's influence domain and $C(i)$ is i's centrality index. The prestige of person i is thus defined as a function of the number of persons in the group, the number of persons either directly or indirectly choosing i, and the average distance of the choosing chains. The prestige index varies from 0 to 1, equaling 0 when no other person chooses i, and 1 when every other person chooses i directly.

Indices are an important product of sociometric research, especially where they relate to optimality of structure or functioning in a social organization. Unfortunately, most indices are ad hoc, with no accompanying proofs of their properties. Since few sociometric investigators have the inclination for producing mathematical proofs, and since few mathematicians are acquainted with problems of data collection, it appears that better indices must arise from better interdisciplinary partnerships.

## Computers and Sociometry

The computer provides a medium for organizing and storing sociometric data, for editing and correcting sociometric data sets, and for performing sociometric analyses. As larger sociometric data sets are collected and libraries of sociometric data are assembled, the need for explicit standards and procedures for managing sociometric data has grown. Small data sets may easily be kept on punched cards and managed by hand, but large matrices, especially ones involving multiple measures are best kept on tape or disk files. A problem still to be resolved is the adoption of a standard format for sociometric files, since it is often necessary to go through extensive reformatting of sociometric data files in the course of using analysis programs written by different authors. Similar problems of file standardization are presently being discussed by the developers of statistical systems.

Data editing and formatting. Most sociometric data sets begin as scalar matrices which are transcribed from their original format to punched cards. One of the most common reasons for limiting the size of a sociomatrix to less than eighty is so that an individual's choices can be stored on a single card.

Such matrices can be manipulated on the computer through the use of standard mathematical packages and subsystems which allow a data matrix to be read in, symmetrized, augmented, totaled, etc. These systems have no "sociometric" options as such, but provide tools which the user may use to write his own sociometric procedures. Probably the most common of such systems is BASIC, an interactive language for mathematical calculation found in the libraries of most timesharing computer systems. APL\360, distributed by IBM (1970) has a more powerful set of matrix manipulation procedures than does BASIC, but requires a special keyboard and type face on the computer terminal. Several variants of APL have been written for both IBM and other computers, which do not require a special keyboard. Speakeasy, a widely available system developed by Cohen and Vincent (1968), has many of the features of APL, but is somewhat easier for the new user to learn and uses a standard keyboard. Interactive systems such as these are useful for performing relatively simple analyses of small sociometric matrices, and for editing and checking small matrices. Analyses which require operations more complicated than the computation and statistical testing of marginal frequencies require specially written programs, most of which are interactive.

Linkage analysis programs. The most complete sociometric program available for distribution, and the only one approximating a sociometric utility system, is Alba's (1972) SOCK, a set of routines for subsetting and clustering sociomatrices. The system reads in a set of sociometric choices and edits them into a matrix in standard form, assigning new identification numbers to individuals. The matrix is then decomposed into disconnected groups, each of which can be analyzed separately. SOCK contains routines for generating several measures of proximity and pairwise association. The user may then use Johnson's (1967) hierarchical clustering algorithm for either a single or a complete linkage analysis. Kruskal's (1964a, 1964b) MDSCAL is incorporated into SOCK for distance analyses of proximity matrices.

COMPLT, described in Alba (1972), is a program for forming the adjacency matrix of a graph and extracting cliques and n-cliques from the matrix. Where the overlap between a pair of extracted cliques is greater than a specified threshold, the program combines them into a single clique. For each subgraph (clique) thus constructed, the ratio of actual lines present to the total number of lines possible in the clique is computed. The ratio of the number of lines joining points in the clique to lines leading outside of the subgraph is also computed. Both ratios are tested against a random graph using a hypergeometric test statistic described in Alba (1972).

COMPLT is one of several programs which extract maximal complete subgraphs by using an algorithm of Bierstone's programmed by (Augustson and Minker, 1970). However, Mulligan and Corneil (1972) showed that Augustson and Minker's algorithm was wrong. Purdy (1973) developed an algorithm for finding the cliques of a graph which runs twice as fast as any previous implementation of the Bierstone algorithm, and presented a formal proof of his algorithm's correctness. Any programs using the Bierstone algorithm should therefore be modified to use the improved Purdy algorithm.

DIP (Directed Graph Processor) developed by Gleason (1971) is a system for manipulating graphs in matrix form. DIP performs all of the usual arithmetic and logical operations on matrices which are found in other packages, i. e., it adds and multiplies matrices, and extracts and augments rows and columns. In addition, DIP computes distance and connectivity matrices for graphs. DIP is related to SOCK and COMPLT in that it is a system for processing graphs in matrix form, but it is more general in its application than either of the other two programs. In SOCK, the user may invoke specific programs, such as the routine for generating proximities, and may specify program options, such as which proximity measure is to be generated. DIP is a language rather than a set of programs. The user writes his own program for an analysis using the DIP language. DIP is superior to other computing languages because graph operations are built into it. For instance, the user may obtain the distance matrix of a graph, A, merely by writing, "COMPUTE DISTANCES FOR A," rather than by specifying the long sequences of arithmetic operations involved.

SOCPAC-I, developed by Leinhart (1971) decomposes a directed graph into its constituent triads and tests the frequencies of various types of triads against a random graph with the same number of null, asymmetric,and mutual links. SOCPAC-I tests the statistical significance of triad distributions by using the distribution theorem described in Holland and Leinhart (1971).

<u>Distance analysis programs.</u> While linkage analysis of social structures requires the use of specially written programs, distance analyses can often be made by

using sociomatrices as input to general purpose distance analysis routines. SOCK uses Kruskal's MDSCAL-IV, a widely used program, for performing distance analyses on matrices edited by other routines. A review of factor analysis, multidimensional scaling, and taxonomic and clustering programs is beyond the scope of this paper, both because there are so many programs available, and because the rate of development Information on new programs for sociometric analysis are published in the program abstracts section of <u>Behavioral Science</u>, and in the Newsletter of SIGSOC, the Special Interest Group on Social Science Computing of the Association for Computing Machinery.

<u>Index construction and statistical analysis.</u> Once the cliques and nets in a sociomatrix have been determined, it is often desirable to analyze characteristics of cliques, either at the individual or the group level. Such analyses are best performed using standard statistical systems such as OSIRIS (Interuniversity Consortium for Political Research, 1971), SPSS (Nie, Bent, and Hull, 1971), or DATA-TEXT (Armor and Couch, 1971). All of these systems use data in the form of a rectangular matrix each row of which contains the observations for one case. Each column of the matrix represents observations of one variable over all cases.

Data obtained from specialized sociometric analysis programs can be added to standard statistical files of data on individuals in the group. For instance, a specialized sociometric system could be used to obtain a list of cliques and their members. The cliques can be assigned identification numbers and each individual in the group tagged with the identification number of each clique of which he or she is a member. The clique identification numbers can then be used as control variables in standard statistical analyses, such as testing the hypothesis that the mean socioeconomic status of individuals varies significantly between two cliques.

These statistical systems all have programs for aggregating individual records into records at the clique level, allowing for the testing of hypotheses concerning clique indices, such as size or connectivity. These statistical systems can also accept sociomatrices in either triangular (symmetric) or square form for input to their distance analysis programs.

## Some Unsolved Problems

There remain in sociometric analysis several methodological and philosophical problems for which there are at present no satisfactory solutions. Sociometry began as a geometric problem with the creation of a sociogram, but one whose basic mathematical attributes were not in doubt. The sociogram is a geometrical construction in a two dimensional Euclidean space, i.e., physical lines drawn on a physical sheet of paper. The matrix representation at first seemed to overcome the geometrical problems in representing sociometric data. It becomes more and more clear that the majority of successful representations of sociometric data require the assumption that a sociometric data set consists of points distributed throughout some kind of space. Most of the powerful statistical operations such as factor analysis and multiple discriminant analysis require a Euclidean metric space, a very difficult commodity to obtain in most psychological measurement. Sociometric analysis has made a great deal of progress since Moreno and seems to be maintaining a unity of ends reached through a variety of means. There is strong agreement that the graphic representation of sociometric information is the most powerful and the most easily understood, but at present the matrix representation of sociometric data provides the only viable instrument for objective analysis. Progress is being made both on the problems of partitioning and covering a sociometric data set and in the construction of individual and and group indices. At present the field seems to be divided between those who assume that almost any arithmetic operation may be performed on sociometric data with impunity, and those who will make no assumption beyond those allowed by a weak order relation on the data. The former workers are probably making mistakes in their construction of indices, but it is also likely that the "weak order" people are not making best use of the available data.

Another problem is that of interpersonal comparison of utilities. If a and b are two objects, and i and j two persons, we are able to tell if i prefers a to b, but we are unable to tell whether i likes a more than j likes b. Such interpersonal comparisons underlie many of the arithmetic procedures used to partition groups. It is possible that a solution to this problem might be approximated, but the experience in price theory does not indicate that we should be optimistic.

The combination of sophistication in mathematical theory with increasingly skillful use of computing power should provide a viable mixture of analytic subtlety and brute force for the analysis of group and organizaional structure. In addition, the perception of sociometry as closely allied to taxonomic and classification theory should result in the finding of new allies in other disciplines. It would be a shame to have psychologists, biologists, and mathematicians laboriously duplicating each others' work and thereby missing the chance for faster progress and intellectual partnership.

## REFERENCES

Alba, R. D.
1971 "COMPLT - A program for analyzing sociometric data and clustering similarity matrices." New York: Columbia University, Bureau of Applied Social Research, mimeo.

1972 "A graph-theoretic definition of a sociometric clique." Journal of Mathematical Sociology, 3:113-126.

Alba, R. D., and Gutmann, M. P.
1971 "SOCK: a sociometric analysis system." New York: Columbia University, Bureau of Applied Social Research, mimeo.

Armor, D. J., and Couch, A. S.
1971 The Data-Text Primer: An Introduction to Computerized Social Data Analysis Using the Data-Text System. New York: Free Press.

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.

Bain, R.
1943 "Sociometry and social measurement." Sociometry 6:206-213.

Ball, G. H., and Hall, D. J.
1965 ISODATA, A Novel Method of Data Analysis and Pattern Classification. Menlo Park, CA: Stanford Research Institute.

Bavelas, A.
1948 "A mathematical model for group structures." Applied Anthropology 7:16-30.

1950 "Communication patterns in task-oriented groups." Journal of the Acoustical Society of America 57:271-282.

Beaton, A. E.
1966 "An interbattery factor analytic approach to clique analysis." Sociometry 29:135-145.

Beauchamp, M. A.
1965 "An improved index of centrality." Behavioral Science 10:161-163.

Berge, C.
1962 The Theory of Graphs and its Applications. New York: John Wiley & Sons, Inc.

Beum, C. O., and Brundage, E. G.
1950 "A method for analyzing the sociomatrix." Sociometry 13:141-145.

Bjerstedt, A.
1956 Interpretations of Sociometric Choice Status. Copenhagen: Munksgaard.

Bock, R. D., and Husain, S. Z.
1950 "An adaptation of Holzinger's B-coefficients for the analysis of sociometric data." Sociometry 13:146-153.

Bonner, R. E.
1964 "On some clustering techniques." I.B.M. Journal of Research and Development 22-33.

Boorman, S. A. and Olivier, C. C.
1973 "Metrics on spaces of finite trees." Journal of Mathematical Psychology 10:26-59.

Borgatta, E. F.
1951 "A diagnostic note on the construction of sociograms and action diagrams." Group Psychotherapy 3:300-308.

Borgatta, E. F., and Stolz, W.
1963 "A note on a computer program for rearrangement of matrices." Sociometry 26:391-392.

Boyle, R. P.
1969 "Algebraic systems for normal and hierarchical sociograms." Sociometry 32:99-119.

Cartwright, D. P., and Harary, F.
1956 "Structural balance: A generalization of Heider's theory." Psychological Review 63:277-292.

1968 "On the coloring of signed graphs." Elemente der Mathematik 23:85-89.

Cattell, R. B., and Coulter, M. A.
1966 "Principles of behavioural taxonomy and the mathematical basis of the taxonome computer program." British Journal of Mathematical and Statistical Psychology 19:237-269.

Cohen, S., and Vincent, C. M.
1968 "An introduction to SPEAKEASY." Argonne, IL: Argonne National Laboratory.

Coleman, J. S.
1958 "Relational analysis: The study of social organizations with survey methods." Human Organization 17:28-36.

Coombs, C. H.
1964 A Theory of Data. New York: John Wiley & Sons, Inc.

Cormack, R. M.
1971 "A review of classification." Journal of the Royal Statistical Society 134:321-367.

Criswell, J. H.
1949 "Sociometric concepts in personnel administration." Sociometry 12:287-[illegible].

Cronbach, L. J., and Gleser, G. C.
1953 "Assessing similarity between profiles." Psychological Bulletin 50:456-473.

Davis, J. A.
1967 "Clustering and structural balance in [illegible]." Human Relations 20:181-187.

1970 "Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices." American Sociological Review 35:841-851.

Davis, J. A., and Leinhart, S.
1971 "The structure of positive interpersonal relations." In J. Berger (Ed.) Sociological Theories in Progress. Vol. II. Boston, MA: Houghton Mifflin Co.

Davis, J. A., and Warnath, C. F.
1957 "Reliability, validity, and stability of a sociometric rating scale." Journal of Social Psychology 45:111-121.

Dunnington, M. J.
1957 "Investigation of areas of disagreement in sociometric measurement of preschool children." Child Development 28:93-102.

Edwards, A. F., and Cavalli-Sforza, L. L.
1965 "A method for cluster analysis." Biometrics 21:362-375.

Eng, E., and French, R. L.
1948 "The determination of sociometric status." Sociometry 11:368-371.

Festinger, L., Schachter, S., and Back, K
1950 Social Pressures in Informal Groups: A Study of Human Factors in Housing. Stanford, CA: Stanford University Press.

Fishbein, M.
1963 An investigation of the relationships between beliefs about an object and the attitude toward that object." Human Relations 16:223-239.

1965 "Prediction of interpersonal preferences and group member satisfaction from estimated attitudes." Journal of Personality and Social Psychology 1:663-667.

Flament, C.
1963 Applications of Graph Theory to Group Structure. Englewood Cliffs, NJ: Prentice-Hall.

Ford, L. R., and Fulkerson, D. R.
1962 Flows in Networks. Princeton, NJ: Princeton University Press.

Forsyth, E., and Katz, L.
1946 "A matrix approach to the analysis of sociometric data: Preliminary report." Sociometry 9:340-347.

Foster, C. C., and Horvath, W. J.
1971 "A study of a large sociogram III: Reciprocal choice probabilities as a measure of social distance." Behavioral Science 16:429-435.

Foster, C. C., Rapoport, A., and Orwant, C. J.
1963 "A study of a large sociogram II. Elimination of free parameters." Behavioral Science 8:56-65.

Gardner, E. F., and Thompson, G. G.
1956 Social Relations and Morale in Small Groups. New York: Appleton-Century-Crofts.

Glanzer, M., and Glaser, R.
1959 "Techniques for the study of group structure and behavior: I Analysis of structure." Psychological Bulletin 56:317-332.

Gleason, T. C.
1969 Multidimensional Scaling of Sociometric Data. Ann Arbor, MI: Institute for Social Research.

1971 "DIP: A directed graph processor." Ann Arbor, MI: Institute for Social Research.

Gleason, T. C., and Cartwright, D.
1967 "A note on a matrix criterion for unique colorability of a signed graph." Psychometrika 32:291-296.

Green, P. E., and Carmone, F. J.
1970 Multidimensional Scaling and Related Techniques in Marketing Analysis. Boston, MA: Allyn and Bacon.

Greenwood, M., and Yule, G. U.
1920 "An inquiry into the nature of frequency distributions representative of multiple happenings." Journal of the Royal Statistical Society 83:255-279.

Guimaraes, L. L.
1968 "Matrix multiplication in the study of interpersonal communication." Unpublished master's thesis, Michigan State University.

Harary, F.
1954 "On the notion of balance of a signed graph." Michigan Mathematical Journal 2:143-146.

1969 Graph Theory. Reading, MA: Addison-Wesley.

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

Harary, F., and Ross, I. C.
1957 "A procedure for clique detection using the group matrix." Sociometry 20:205-215.

Heider, F.
1946 "Attitudes and cognitive organization." Journal of Psychology 21:107-112.

Holland, P. W., and Leinhart, S.
1970 "A method for detecting structure in sociometric data." American Journal of Sociology 76:492-513.

1973 "The structural implications of measurement error in sociometry." Journal of Mathematical Sociology 3:85-111.

Hubbell, C. H.
1965 "An input-output approach to clique identification." Sociometry 28:377-399.

International Business Machines Corporation
1970 "APL\360 User's manual." White Plains, NY: Author.

Inter-University Consortium for Political Research
1971 OSIRIS II User's Manual. Ann Arbor, MI:Institute for Social Research.

Jennings, H. H.
1950 Leadership and Isolation. New York: Longmans, Green.

Johnson, S. C.
1967 "Hierarchical clustering schemes." Psychometrika 32:241-254.

Katz, L.
1953 "A new status index derived from sociometric analysis." Psychometrika 18:39-43.

Katz, L., and Powell, J. H.
1960 "Probability distributions of random variables associated with a structure of the sample space of sociometric investigations." In J. L. Moreno (Ed.) The Sociometry Reader Glencoe, IL: Free Press, 309-317.

Kruskal, J. B.
1964 "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis." Psychometrika 29:1-27.

1964b "Nonmetric multidimensional scaling: A numerical method." Psychometrika 29:115-130.

Leinhart, S.
1971 "SOCPAC I: A FORTRAN IV program for structural analysis of sociometric data." Behavioral Science 16:515.

Leontief, W. W.
1941 The Structure of the American Economy, 1919-1929. Cambridge, MA: Harvard University Press.

Lerner, M. J.
1965 "The effect of responsibility and choice on a partner's attractiveness following failure." Journal of Personality. 33:178-187.

Lin, N.
1970 "Measuring prestige as a property of group structure." Baltimore, MD: The Johns Hopkins University, mimeo.

Lindzey, G., and Byrne, D.
1968 "Measurement of social choice and interpersonal attractiveness." Pp. 452-525 in Lindzey, G. and Aronson, E. (Eds.) The Handbook of Social Psychology (2nd ed.) Vol 2. Reading, MA: Addison-Wesley.

Lingwood, D. A.
1969 "Interpersonal communication, research productivity, and invisible colleges." Unpublished doctoral dissertation, Stanford University.

Lorr, M.
1968 "A review and classification of typological procedures." Paper read at the American Psychological Association meeting, San Francisco, CA.

Lorr, M., and Radhakrishnan, B. K.
1967 "A comparison of two methods of cluster analysis." Educational and Psychological Measurement 27:47-53.

Luce, R. D.
1950 "Connectivity and generalized cliques in sociometric group structure." Psychometrika 14:169-190.

Luce, R. D., and Perry, A. D.
1949 "A method of matrix analysis of group structure." Psychometrika 14:95-116.

MacRae, D., Jr.
1960 "Direct factor analysis of sociometric data." Sociometry 23:360-371.

McKinney, J. C.
1948 "An educational application of a two-dimensional sociometric test." Sociometry 11:356-367.

McQuitty, L. L.
1957 "Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies." Educational and Psychological Measurement 17:207-229.

1963 "Rank order typal analysis." Educational and Psychological Measurement 23.

McQuitty, L. L., and Clark, J. A.
1968 "Clusters from iterative, intercolumnar correlational analysis." Educational and Psychological Measurement 28:211-238.

Michener, C. D., and Sokal, R. R.
1957 "A quantitative approach to a problem in classification." Evolution. 11:130-162.

Moreno, J. L.
1934 Who Shall Survive? Washington, DC: Nervous and Mental Disease Publishing Company.

Moreno, J. L., Jennings, H. H., Criswell, J. H., Katz, L., Blake, R. R., Mouton, J. S., Bonney, M. E., Northway, M. L., Loomis, C. P., Proctor, C., Tagiuri, R., and Nehnevajsna, J.
1960 The Sociometry Reader. Glencoe, IL: Free Press.

Mouton, J. S., Blake, R. R., and Fruchter, B.
1960a "The reliability of sociometric measures." In J. L. Moreno et al. (Eds.) The Sociometry Reader. Glencoe, IL: Free Press. Pp. 320-361.

1960b "The validity of sociometric responses." In J. L. Moreno et al. (Eds.) The Sociometry Reader. Glencoe, IL: Free Press. Pp. 362-387.

Mulligan, G. D., and Corneil, D. G.
1972 "Corrections to Bierstone's algorithm for generating cliques." Journal of the Association for Computing Machinery 19:244-247.

Needham, R. M.
1961 "The theory of clumps, II." Report M. L. 139, Cambridge Language Research Unit, Cambridge, England.

Newcomb, T. M.
1961 The Acquaintance Process. New York: Holt, Rinehart, and Winston.

Nie, N. H., Bent, D. H, and Hull, C. H.
1970 SPSS: Statistical Package for the Social Sciences. New York: McGraw-Hill.

Norman, R. Z., and Roberts, F. S.
1972 "A derivation of a measure of relative balance for social structures and a characterization of extensive ratio systems." Journal of Mathematical Psychology 9:66-91.

Norman, W. T.
1967 "Cluster analysis: A diametric rationale and procedure." Ann Arbor, MI: University of Michigan, mimeo.

Northway, M. L.
1940 "A method for depicting social relationships obtained by sociometric testing." Sociometry 3:144-150.

Nosanchuk, T. A.
1963 "A comparison of several sociometric partitioning techniques." Sociometry 26:112-124.

Ore, O.
1963 Graphs and Their Uses. New York: Random House.

Parker-Rhodes, A. F.
1961 "Contributions to the theory of clumps, I." Report M. L. 138, Cambridge, England: Cambridge Language Research Unit.

Peay, E. R., Jr.
1970 Extensions of clusterability to quantitative data with an application to the cognition of political attitudes. Unpublished doctoral dissertation, University of Michigan.

Pepinsky, H. B., Siegel, L., and Vanatta, E. L.
1952 "The criterion in counseling: A group participation scale." Journal of Abnormal and Social Psychology 47:415-419.

Peterson, R. J., Komorita, S. S., and Quay, H. C.
1964 "Determinants of sociometric choices." Journal of Social Psychology 62:65-75.

Phillips, D. P., and Conviser, R. H.
1972 "Measuring the structure and boundary properties of groups: Some uses of information theory." Sociometry 35:235-254.

Proctor, C. H.
1960 "A sociometric method for distinguishing social structures." In R. N. Adams and J. J. Preiss (Eds.) Human Organization Research Homewood, IL:Dorsey Press, pp. 308-315.

1967 "The variance of an estimate of linkage density from a simple random sample of graph nodes." American Statistical Association. Proceedings of the social statistics section 342-343.

1969 "Analyzing pair and point data on social relationships, attitudes and background characteristics of Costa Rican census bureau employees." American Statistical Association. Proceedings of the social statistics section 457-465.

Proctor, C. H., and Loomis, C. P.
1951 "Analysis of sociometric data." Pp. 561-585 in Marie Jahoda et al. (Eds.) Research Methods in Social Relations, with Especial Reference to Prejudice. New York: Dryden.

Purdy, G.
1973 "Finding the cliques of a graph." Technical Memorandum, Center for Advanced Computation. Urbana, IL: University of Illinois.

Rapoport, A., and Horvath, W. J.
1961 "A study of a large sociogram." Behavioral Science 6:279-291.

Riffenburgh, R. H.
1966 "A method of sociometric identification on the basis of multiple measurements." Sociometry 29:280-290.

Roistacher, R. C.
1971 "Peer nominations of exploratory behavior." Buffalo: Studies in Psychotherapy and Behavioral Change, No. 2. Buffalo, NY: State University of New York at Buffalo.

1972 Peer nominations, clique structures, and exploratory behavior in boys at four junior high schools. Unpublished doctoral dissertation, University of Michigan.

Saunders, D. R., and Schucman, H.
1962 "Syndrome analysis: An efficient procedure for isolating meaningful subgroups in a nonrandom sample of a population." Paper read at the third annual Psychonomic Society meeting, St. Louis, MO.

Sawrey, W. L., Keller, L., and Conger, J. J.
1960 "An objective method of grouping profiles by distance functions and its relation to factor analysis." Educational and Psychological Measurment. 20:651-674.

Shepard, R. N.
1972 A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, and M. B. Nerlove. (Eds.) Multidimensional Scaling: Theory and Applications in the Behavioral Sciences. Vol I. Theory. New York: Seminar Press.

Shepard, R. N., Romney, A. K., and Nerlove, S. B.
1972 Multidimensional Scaling: Theory and Applications in the Behavioral Sciences. Vol. I. Theory. New York: Seminar Press.

Smucker, O.
1949 "Near-sociometric analysis as a basis for guidance." Sociometry 12:326-340.

Sokal, R. R., and Sneath, P. H. A.
1963 Principles of Numerical Taxonomy. San Francisco, CA: Freeman.

Sonquist, J. A.
1970 Multivariate Model Building: The Validation of a Search Strategy. Ann Arbor, MI: Institute for Social Research.

Spilerman, S.
1966 "Structural analysis and the generation of sociograms." Behavioral Science 11:312-318.

Taylor, M.
1969 "Influence structures." Sociometry 32:490-502.

Thorndyke, R. L.
1953 "Who belongs in the family?" Psychometrika 18:267-276.

Torgerson, W. S.
1958 Theories and Methods of Scaling. New York: Addison-Wesley.

Van Valkenberg, M. E.
1964 Network Analysis. Englewood Cliffs, NJ: Prentice-Hall.

Wright, B., and Evitts, M. S.
1961 "Direct factor analysis in sociometry." Sociometry 24:82-98.