Derivation of Subgroups From Dyadic Interactions

Dawn Iacobucci
Department of Marketing, Kellogg Graduate School of Management
Northwestern University

This article presents a simulation study that compares several methods for deriving empirical subgroups from sociometric data. The Monte Carlo study was used to investigate how well the methods recovered the subgroup structure that had been built into the actors' and partners' modes. Forty-eight sociomatrices were generated using a $2^4 \times 3$ factorial design. The factors included the number of individuals in the network, the true number of subgroups into which the individuals were classified, the proportion of individuals falling into each of the subgroups, the structure of the dyadic interactions, and the clarity of the subgroup structure. On the basis of the simulation study's results, subgroups were derived for two real data sets. The first data set described the relations in a monastery (Sampson, 1968). The second data set described the referral network of a service provider (Reingen & Kernan, 1986).

Dyadic social interactions are measured on the set of interdependent behaviors that are observed between two individuals. There are numerous relations that may be observed between the members of a dyad. For example, one person might *like, have respect for,* or *criticize* another person.

More generally, the set of relations may exist between any two "entities," where an entity might be an individual or a group of individuals. For example, relations might exist between corporations (e.g., if one corporation receives payment from another) or relations might exist between nations (e.g., if one nation asks another for assistance in achieving some goal).

If the relations are interpersonal, they may be referred to as *sociometric data,* or *dyadic social interactions.* Relations such as *makes payment to* might be referred to by a more general term, such as *dyadic interactions.*

Dyadic interactions are of interest to researchers working in many substantive areas. For example, clinical psychologists interested in discriminating between behavioral patterns of distressed and nondistressed marital dyads could model the couple's sequence of interactions (Budescu, 1984; Dillon, Madden, & Kumar, 1983; Dumas, 1984; Feick & Novak, 1985; Gottman, 1979; Gottman & Bakeman, 1979; Gottman & Notarius, 1978; Margolin & Wampold, 1981). Developmental psychologists interested in studying the sociometric friendship choices measured in a classroom of children might model the pairs of friendships existent at a single point in time (Allison & Liker, 1982; Kenny, 1981; Kenny & LaVoie, 1984; Mendoza & Graziano, 1982). Dyadic interactions may also interest researchers who work at a more "macro" level, such as a sociologist studying the transactions that exist between a set of corporations and nonprofit organizations (Galaskiewicz & Wasserman, 1987), a researcher in marketing studying distribution channel interactions, or referral networks in the service industry.

In particular, *Psychological Bulletin* has published a series of articles in the field of dyadic interactions. Some of the articles that introduced this research include Iacobucci and Wasserman (1987), Kenny and Judd (1986), and Kraemer and Jacklin (1979). Some of the articles that address particular concerns for modeling social network data include Dumas (1986), Faraone and Dorfman (1987), Iacobucci and Wasserman (1988), and Mendoza and Graziano (1982). The article by Noma and Smith (1985) is especially relevant, because it addresses the issue of clustering individuals into groups on the basis of their patterns of interactions with others.

Data representing dyadic interactions are usually tabulated as elements of a *sociomatrix.* A standard sociomatrix, **X**, is a square, two-way array with rows corresponding to actors and columns corresponding to partners. An entry in the sociomatrix (x_{ij}) depicts a relation originating with actor, or row, i, and going to partner, or column, j. For example, actor i nominates partner j as a friend, or department i communicates with department j.

Many variations of this standard sociomatrix are possible. First, the matrix is two-dimensional to reflect the dyadic nature of the interactions. If instead one were interested in modeling triads (or larger groups), the matrices could be extended to three (or higher) dimensions.

Second, the sociomatrix may be square or rectangular. In a square sociomatrix, the sets of actors and partners consist of the same individuals. (The term *square* is used to represent the characteristic of the set of actors being equal to the set of partners, rather than in a reference to the number of actors being equal to the number of partners. One might also refer to square matrices as "one-mode" matrices.) The diagonal of this matrix

I thank Stanley Wasserman, Joe Schwartz, Larry Hedges, and the anonymous reviewers for their very helpful comments on this research and article.

Correspondence concerning this article should be addressed to Dawn Iacobucci, Department of Marketing, Kellogg Graduate School of Management, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208.

Psychological Bulletin, 1990, Vol. 107, No. 1, 114-132
 0033-2909/90/$00.75

represents the usually unobserved "reflexive" relation. In a rectangular sociomatrix, the set of actors is different from the set of partners (i.e., a two-mode matrix). One set (the actors or partners) might be a subset of the other, or the two sets might consist of entirely different entities.

Third, the relation the matrix describes may be unidirectional or bidirectional. For example, donations flow from corporations to nonprofit organizations, but information could flow both ways.

Fourth, the dyads that interact may be constrained by some design. An example of a sparse social network is the set of interactions between married couples, because each husband interacts with only one wife, and vice versa. The dyadic design with the most possible interactions is the round-robin design, where each actor interacts with all partners. Round-robin designs and intermediate designs have been studied extensively by Kenny (e.g., Kenny & LaVoie, 1984).

Finally, these data are usually discrete. In friendship nominations, friends get scores of one and those not nominated get scores of zero, or friends could be ranked. However, it is easy to imagine network data that are continuous, such as donations measured in dollars. These data could be translated and described in qualitative terms such as *high, medium,* and *low* amounts. This article focuses on two-dimensional, square sociomatrices, where the relations are discrete and bidirectional and may exist between any dyad. However, many of the methods described later may be applied to other types of sociomatrices.

In studying these sociomatrices, one might focus on one of two things: First, one might be interested in estimating effects for the expansiveness of actors (e.g., in friendship choices made), the popularity of partners (e.g., in friendship choices received), and the reciprocity unique to the particular dyadic relationship (e.g., mutual friendship choices). Second, one might be interested in discovering cliques, or clusters of actors who interact with their partners in similar ways. A method for estimating actor, partner, and relationship effects is described, and then the issue of subgroups is discussed.

Simple Effects: Actor, Partner, Relationship

Linear models for estimating actor, partner, and relationship effects for continuous network data have been proposed by Kraemer and Jacklin (1979) and Kenny (1981). The models used in the current study were designed especially for discrete network data. These log-linear models have been discussed in detail in several sources, so they are only briefly introduced here (for example, see Fienberg, Meyer, & Wasserman, 1985; Fienberg & Wasserman, 1981; Holland & Lienhardt, 1981; Iacobucci & Wasserman, 1987; Wasserman, 1987; Wasserman & Galaskiewicz, 1984; Wasserman & Iacobucci, 1986).

By way of notation, actor i relates to partner j at level k, and j relates to i at level m, where k and m come from the set $\{1, 2, 3, \cdots C\}$. The dyad consisting of individuals i and j, then, is described by the values $(x_{ij} = k)$ and $(x_{ji} = m)$. To allow log-linear model fitting by using the widely available statistical computing packages, Fienberg and Wasserman (1981) defined a Y array:

$$Y_{ijkm} = 1 \text{ if } (x_{ij} = k \text{ and } x_{ji} = m), \text{ and}$$
$$= 0 \text{ otherwise.}$$

For g actors and partners, this Y array is $g \times g \times C \times C$. The probability that the dyad takes on the values $(x_{ij} = k)$ and $(x_{ji} = m)$ is denoted π_{ijkl}, and may be modeled

$$\ln \pi_{ijkm} = \lambda_{ij} + \theta_k + \theta_m + \alpha_{i(k)} + \alpha_{j(m)} + \beta_{j(k)} + \beta_{i(m)} + \rho_{(km)}.$$

The λ_{ij}s are parameters that ensure the probabilities sum to one for each dyad, and the θs are grand mean effects for the level of the relational variable. The parameters that are of greater interest are the alphas, betas, and rho. For the example of actors choosing partners as friends, the alphas may be interpreted as effects for the expansiveness of actors, the betas represent effects for the popularity of partners, and rho represents an effect for the interaction, or reciprocity relationship within the dyads. These parameters may be useful in forming subgroups of similar individuals and form the basis of several of the methods investigated in this article.

The parameters in the model (αs, βs, and ρ) are effects in the data. A model that is proposed to describe the data can contain any subset of these parameters. The parameters of interest may be estimated by specifying the marginal totals of the Y array that must be fit in a log-linear model.

Frequently, researchers are not interested in the behavior of particular individuals or dyads, but in the general behavior of subgroups of individuals. For example, clinical psychologists might be interested in comparing the relations that exist within distressed couples with those observed for nondistressed couples. Developmental psychologists might be interested in comparing the social behavior of boys and girls or of older and younger children. Sociologists and economists might be interested in comparing the donations made by small corporations with those made by larger corporations.

Subgroups based on these attributes can be formed using a priori, theoretical grounds. The attribute variables in these examples were status of distress, sex, age, and size. Alternatively, these subgroups could be empirically derived by post hoc methods, which are described shortly.

The grouping of individuals has been incorporated into the log-linear modeling approach by aggregating the Y array over individuals within identical subgroups to form a W array. This aggregation is possible because the individuals within subgroups are assumed to behave similarly; parameters for individuals within subgroups are equated by this aggregation.

The Y array then reflects the behavior of individuals, and the W array reflects the grouped behavior of the actors and partners, or the behavior of the subgroups. For example, in a class of children, the Y array would describe the relations existing between all individual children, and the W array would describe the relations between the subgroups defined by, say, sex.

Researchers interested in individual differences would model the Y array, and researchers interested in group differences would model the W array. One of the methods that is used in deriving subgroups includes fitting models to the Y array in order to estimate effects for individuals, and then clustering individuals on the basis of the similarities of these parameter estimates.

Before leaving the discussion of the log-linear modeling of network data, it should also be briefly noted that, for users of the log-linear models, the *W* array has the advantage that asymptotic tests are achievable. That is, increasing sample size by adding a new individual or dyad necessarily extends the dimensions of the *Y* array and increases the number of parameters one must estimate. However, when a fixed number of subgroups are defined, adding a new individual or dyad does not increase the number of estimated parameters.

Note, however, that although this is true for the use of a priori subgroups, such as those defined by a theoretically important variable like sex, this is not necessarily true of subgroups formed by the post hoc empirical methods. By these post hoc methods, the addition of a new dyad certainly holds the possibility that a dyad is not similar to any existing group, requiring the formation of a new group.

The theoretical basis for forming subgroups is discussed next, followed by a description of the simulation study. In the presentation of the simulation study, several empirical methods are described and evaluated.

Deriving Subgroups

A second concern in the analysis of sociometric data is to find homogeneous clusters of individuals. Researchers have defined two individuals to be "structurally equivalent" if they relate to all others in the same way and in turn are related to by all others in the same way (Knoke & Kuklinski, 1983; Lorrain & White, 1971).

Structural equivalence is guided by the general observation that some groups of people behave similarly. The members of such a group might share an easily identifiable demographic characteristic, such as a common age or sex. The attribute these members share might not be easily identifiable; it might be a personality trait such as extraversion or a set of characteristics taken on with some role, such as "teacher." These attributes together may be hypothesized to create an implicit approach to viewing the world and to behaving with others that makes someone's behavior similar to that of the other members in the group. The particular attributes chosen to describe the actors and partners in the models would presumably be selected because of their relevance to the relational behavior of the group. These expectations about their relevance may then be tested empirically.

If clusters of individuals are formed before further network analyses based on known attributes of the network members, such as sex, subsequent analyses would model not the behavior of a particular dyad, but the aggregate behavior of many dyads. In this subsequent modeling, all women would be assumed to relate to, and be related to, in the same way; and the same is true of the men. For a priori groups, structural equivalence serves as an assumption for the relational behavior of the members with the same set of attributes.

The vast majority of articles in this area focus on post hoc methods of grouping individuals on the basis of their observed relational data (Breiger, Boorman, & Arabie, 1975; Faust, 1985; Noma & Smith, 1985; Wasserman & Anderson, 1987; White, Boorman, & Breiger, 1976). For these post hoc groups, structural equivalence serves as a guideline, or a goal, for a clustering algorithm.

The clusters of individuals are commonly referred to as *blocks,* because sociomatrices consisting of groups of structurally equivalent individuals can be permuted to a blocked pattern of similar data. Most often these relational data are binary, and researchers refer to "zero" or "one" blocks: blocks containing all zeros, indicating no choices were sent, or all ones, indicating all actors in the row block sent choices to all partners in the column block. It is generally recognized that true structural equivalence is probably too restrictive an assumption for real data. For example, a block of data might not contain only entries of one, but the block may primarily consist of entries of one.

An assumption that may be more realistic than structural equivalence, at least for the purpose of modeling real data, might be that of *stochastic equivalence* (Wasserman & Weaver, 1985). The behaviors of two stochastically equivalent individuals are defined to be probabilistically identical, and the individuals share a common set of model parameters. Predictions for stochastically equivalent individuals would be identical, but their observed data might differ.

One method that attempts to derive block models that is popular with network researchers is an algorithm called CONCOR (Breiger et al., 1975; Faust, 1985; Noma & Smith, 1985; White et al., 1976). CONCOR starts with the correlations between rows and columns of a sociomatrix and proceeds iteratively to compute the correlations among the columns of correlations, until the matrix converges to values of plus or minus one. Researchers usually use this algorithm with the goal that the resulting block models will contain roughly structurally equivalent individuals.

Because the definitions of equivalence require identical ties originating from and directed toward individuals, the transpose of the sociomatrix is generally "stacked" on top of the sociomatrix. The initial iteration of correlations, then, is computed over the sociomatrix, stacked on top of its transpose. These correlations are effectively computed over both the rows of the original sociomatrix, which describe the behavior of the actors (and are included in the stack as the transpose), and the columns of the sociomatrix, which describe the behavior of the partners.

CONCOR has been criticized for its entirely empirical nature (cf. Faust, 1985; Schwartz, 1977). For example, although the iterated correlations generally converge to plus and minus one, there have been no proofs of its convergence, and other mathematical properties are not clear. Iterated correlations have also been studied by McQuitty and Clark (Clark & McQuitty, 1970; McQuitty, 1968; McQuitty & Clark, 1968).

Schwartz (1977) has shown that the block model that results from the first split into two groups by CONCOR is equivalent to grouping individuals on the basis of the sign of their loadings on the first eigenvector of the sociomatrix. That is, the first eigenvector associated with the covariance matrix of individuals would generate two blocks, one consisting of people with positive loadings and one consisting of people with negative loadings. Presumably, individuals with loadings near zero are not central figures to their subgroups. Successive subgroup splits would follow similarly for subsequent eigenvectors.

Given the body of theory and statistical knowledge support-

ing eigensolutions, it seems it would be more useful to compute the eigensolution than to use CONCOR. Furthermore, the loadings on the eigenvectors are continuous and the information obtained by CONCOR is only dichotomous (i.e., correlations of plus or minus one). These continuous values might indicate the strength of group membership. Finally, the second, third, and remaining eigenvectors may also be informative. Frequently in data, the information contained in the first eigenvector is not sufficient in describing the data. For the same reasons, one might expect that the formation of block models would also be improved by the use of additional eigenvectors.

In the simulation study that follows, the subgrouping methods include CONCOR, the eigenvector solutions, and approaches that make use of log-linear model parameters.

The factors used to define subgroup structure on the actors and partners are described, followed by a description of the methods used to attempt to recover these structures. The results of this study are then reported, and the article finishes with a description of the derivation of subgroups in two real data sets.

Simulation Study—Subgroup Recovery

The first factor in the simulation study was whether the number of actors and partners was 8 or 16. One of the real data sets consisted of 18 actors, so 16 is on the order that is realistic for data, and 8 is a simple fraction. A fraction of 16 was taken rather than a multiple because of restrictions on one of the subgroup recovery methods. Because of the size of the resulting Y arrays, the log-linear model approach to a network of 16 individuals could not be taken, but these methods could be applied to a network of 8 individuals.

A second factor in the generation of these simulated data was whether there were two or four subgroups imposed on the 8 or 16 individuals. Recovery was expected to be better when there were two subgroups, rather than four. In the same sense that a mean is better estimated when based on more observations, it was expected that having individuals divided into only two groups (i.e., 4 or 8 persons per group) would be better estimated, or recovered, than having those individuals divided into four groups (i.e., only 2 or 4 persons per group). The idea is that the more persons per group, the better any method should be able to identify that group.

Similarly, it is widely recognized that in factor analysis, those factors that are overidentified (i.e., many variables load highly on those factors) are better recovered. If a factor is analogous to a contrast between two subgroups (i.e., those with positive vs. negative loadings on the eigenvector) and the variables are analogous to the individuals, the subgroups should be recovered better if they contain more individuals.

A third factor that seemed important to understanding subgroup recovery was whether the number of individuals within the subgroups was equal or unequal. For 8 actors, the number of individuals in two subgroups would be 4 and 4 or 2 and 6, and the number of individuals in four subgroups would be 2, 2, 2, and 2 or 1, 2, 2, and 3. For 16 actors, the number of individuals in two subgroups would be 8 and 8 or 4 and 12, and the number of individuals in four subgroups would be 4, 4, 4, and 4 or 2, 4, 4, and 6. The question here was whether those subgroups with more individuals would dominate, in the sense that they would be recovered better.

A fourth factor included in this study was whether the majority of the relational ties were mutual or null (i.e., reciprocal), or whether the ties were mostly asymmetric. *Symmetric* means that a choice is reciprocated, or the dyad is null. Thus, the entries in the sociomatrix x_{ij} and x_{ji} would be equal (both equal zero or one), leading to a symmetric matrix. It was hoped that subgroup recovery would be indifferent to the structure of the relational ties. That is, it would be ideal if there were some method that recovered subgroup structure very well, without regard to the overall distribution of dyadic interactions. If the methods were sensitive to whether the ties were mostly mutual and null, or asymmetric, then the researcher would have two preliminary steps in the analysis of network data. First, it would be necessary to determine whether the given network was characterized mostly by mutual ties or asymmetric ties. Then, given that information, the researcher would choose a method that recovers subgroups well for that type of relational structure.

The final factor in this Monte Carlo study was the clarity of the subgroup structure in the given sociomatrix. This factor had three levels, clear, medium, and not clear, and was created as follows: The clearly defined subgrouped sociomatrices were created by hand to contain the structure defined by the factorial design on the other four factors. The medium and not-clear subgrouped sociomatrices were created by perturbing the clear sociomatrices by error. An explanation of "error" follows.

The clear sociomatrix served as input, and for every entry of a 1, the output, perturbed matrix also contained a 1 with some predetermined probability. Every entry of 0 was similarly treated, except that all elements down the main diagonal remained at zero. The sociomatrices were binary to start simply.

The probability distribution was Bernoulli, and the probabilities used were .8 for the medium clarity subgroup structure and .6 for the not-clear subgroup structure (to be complete, the probability was 1.0 for the clear subgroup).[1] For example, for a probability of .8, a one in the clear subgroup matrix was equal to one in the medium-clear matrix with a probability of .8, and equal to zero with a probability of .2. The zeros in the clear subgrouped matrix were equal to zero in the output matrix with a probability of .8, and equal to one with a probability of .2.

These probabilities are conservative, in the sense that the noisiest data possible, or the toughest situation for recovery, are being set up. If one of the probabilities had been .5, the data would be totally random noise, with no subgroup structure. Probabilities less than .5 would merely be the reflections of the subgroups with probabilities greater than .5. For example, the subgroups disturbed by a probability of .1 would be as clear as those disturbed by a probability of .9, but the ones in the data would become zeros, and the zeros would become ones. One

[1] These values were derived from a uniformly distributed random variable that ranged from 0 to 1, generated by the RANF function in the CDC FORTRAN Common Library Mathematical Routines. The modulus for this function is 2**48, and the multiplier is octal 2000120726427173056S. The six digits comprising real time in hours, minutes, and seconds served as the input argument, or seed, to this function.

Table 1
True Subgroup Structure

Subgroup membership	Cells
(1, 2, 3, 4)(5, 6, 7, 8)	a-8, b-8, c-8, d-8, e-8, f-8
(1, 2)(3, 4, 5, 6, 7, 8)	g-8, h-8, i-8, j-8, k-8, l-8
(1, 2)(3, 4)(5, 6)(7, 8)	m-8, n-8, o-8, p-8, q-8, r-8
(1)(2, 3)(4, 5)(6, 7, 8)	s-8, t-8, u-8, v-8, w-8, x-8
(1–8)(9–16)	a-16, b-16, c-16, d-16, e-16, f-16
(1, 2, 3, 4)(5–16)	g-16, h-16, i-16, j-16, k-16, l-16
(1–4)(5–8)(9–12)(13–16)	m-16, n-16, o-16, p-16, q-16, r-16
(1, 2)(3–6)(7–10)(11–16)	s-16, t-16, u-16, v-16, w-16, x-16

other study that was somewhat similar examined the effect of less noise; the probabilities were .9, .85, and .8 (Faust, 1985).

Because the clarity manipulation was probabilistic, the actual degree of disturbed structure varied around .6 and .8. A subsequent step in verifying the results that follow would be to generate several replicates in each cell of the design, rather than just a single sociomatrix in each cell. Notice that for several cells, namely those with perfectly clear structure, replicates would be trivial and the within-cell variances would be zero. Because of this complication, the analysis proceeded with only a single matrix per cell; there are no replications in the present study.

This five-factor factorial design is summarized in the beginning of the appendix. The 48 sociomatrices that this design generated follow. Note that in the matrices with clear structure, effects such as reciprocity are assumed to be perfect. This may not be a realistic portrayal of real data, but for the purposes of the simulation study it serves as a clean baseline. Similarly, the subgroup structure is also, perhaps unrealistically, perfect, in the sense that within a given subgroup there are no differential actor or partner effects: The structural equivalence is exact.

Finally, also note that error was added to the data in a tie-by-tie basis rather than a dyad-by-dyad basis. That is, in modifying a tie, a new relation could change to 0 or 1. If one were to modify a dyad, a pair of relations could change to (0, 0), (0, 1), (1, 0), or (1, 1). These two cases are identical, as the probabilities of the four dyadic states are uniform. However, one might refine the dyadic approach so that sociomatrices in the symmetric condition would take on only the values (0, 0) or (1, 1), preserving the null and mutual symmetric structure. Similarly, the asymmetric sociomatrices might have been constrained to take on dyadic states of (0, 1) or (1, 0) only. This type of restriction seemed more complicated than a first simulation study warranted, so error was added in a tie-by-tie basis rather than at the dyad level. This choice effectively introduced greater asymmetry in the medium-clear and not-clear symmetric cells of the design (and in fact probably contributed to a slight—although insignificant—Clarity × Symmetry interaction that is described later).

The subgroup structure that each of several methods (which are described shortly) was attempting to recover was that structure defined in the clear subgroup cells. For example, the individuals were partitioned into two groups for cell a-8: The first group consisted of individuals 1, 2, 3, and 4, and the second group consisted of individuals 5, 6, 7, and 8. (The subgroups for the actors were the same as for the partners in order to keep things simple.) This partition is denoted (1, 2, 3, 4), (5, 6, 7, 8) and was also the subgroup structure that should be recovered for cells b-8, c-8, d-8, e-8, and f-8. The subgroup structure that should be recovered for each cell is listed in Table 1.

The criterion measure of subgroup recovery was a regression-based measure of partition similarity discussed in (and improved on by) Hubert and Arabie (1985). Each of the observed partitions (from each of the methods) was used to try to predict the true subgroup structure.

The measure is computed as follows. The true partition will form the columns (j) of a matrix, and the observed partition will form the rows (i). For example, if the true partition had been (1, 2, 3, 4), (5, 6, 7, 8), there would be two columns, one for the group of individuals 1, 2, 3, and 4, and one column for the group of individuals 5, 6, 7, and 8. If the observed partition had been (1), (2, 3, 4), (5, 6, 7, 8), there would be three rows, one for individual 1, one for individuals 2, 3, and 4, and one row for individuals 5, 6, 7, and 8.

The entries in this matrix (n_{ij}) would be the number of individuals in the row subgroup that are common to the column subgroup. The matrix for this example is given in Table 2. The row sums (n_{i+}), the column sums (n_{+j}), and the overall total (n_{++}) are computed. The regression coefficient is a function of the cell entries and these marginal totals. The equation follows and is demonstrated for the example in Table 2.

$$b_{\text{true}(j),\text{obs}(i)} = \frac{\sum_{i,j}[n_{ij}(n_{ij}-1)]/2 - \dfrac{\sum_i\{[n_{i+}(n_{i+}-1)]/2\} * \sum_j\{[n_{+j}(n_{+j}-1)]/2\}}{[n_{++}(n_{++}-1)]/2}}{\sum_i\{[n_{i+}(n_{i+}-1)]/2\} - \dfrac{\sum_i\{[n_{i+}(n_{i+}-1)]/2\} * \sum_i\{[n_{i+}(n_{i+}-1)]/2\}}{[n_{++}(n_{++}-1)]/2}}$$

This predictive measure was the criterion along which the methods of deriving subgroups were compared. The statistic is

Table 2
An Example of Computing the Regression Measure of Partition Similarity

The goal in this example is to predict the true partition (1, 2, 3, 4), (5, 6, 7, 8) from the observed partition (1), (2, 3, 4), (5, 6, 7, 8).

	true		
observed	(1, 2, 3, 4)	(5, 6, 7, 8)	row sums
(1)	1	0	1
(2, 3, 4)	3	0	3
(5, 6, 7, 8)	0	4	4
column sums	4	4	8

$$b_{\text{true,obs}} = \frac{(0+0+3+0+0+6) - [(0+3+6)(6+6)/28]}{(0+3+6)[1-(0+3+6)/28]}$$

$$= \frac{9-(9*12/28)}{9[1-(9/28)]} = \frac{9-3.857}{9(1-.321)} = \frac{5.143}{6.111}$$

$$= .842.$$

bounded by one, but can be less than zero, when the cell counts are small.

There were 16 methods used on each of the 48 sociomatrices to try to recover the known subgroups. Before describing these 16 methods, it should be noted that several subjective methods were included. They were included for comparison, but the primary intent was to be able to recommend the best-performing objective methods, so that the subjectivity, which depends on the ability of the particular researcher, would be minimized. It should also be noted that these subjective judgments were of course made "blind," in the sense that subgroups were determined without the information of which cell in the design the sociomatrix belonged to, which would have indicated the true subgroup structure.

The first method is CONCOR. Despite the criticisms of the method mentioned earlier, it is recognized that this method is extremely popular because it is so easy to use and is widely available. For comparison, then, it is important to include this method.

The partition that results from the finest split provided by CONCOR was the subgroup structure derived for this first method, "CONCOR full split." It is not the case that CONCOR took eight actors and divided them into two groups of four, then four groups of two, then eight groups of one. In fact, in some of the clear subgroup cases, the algorithm split the eight individuals into the two appropriate groups and then would not split the groups further. In other words, taking the most divisive split provided by CONCOR did not result in the trivial grouping of eight actors each in their own subgroup.

The next three methods were based on fitting the log-linear model to the Y arrays. (These three methods were not applied to the 16×16 sociomatrices.) These methods used the set of alpha parameter estimates along with the set of beta parameter estimates in order to find a subgroup structure that incorporated both the actor and partner characteristics. Method 2 was a subjective method, and Methods 3 and 4 were objective.

In Method 2, the alpha estimates were plotted against the beta estimates, and subgroups were determined by examining the plots. In Methods 3 and 4, the vectors of alpha and beta estimates were used as coordinates in a two-dimensional space, and interpoint distances were computed. That is, each individual was associated with the coordinates consisting of some alpha and some beta, and the interactor distances were computed over these two vectors. This matrix of distances served as the input to a hierarchical clustering algorithm. Method 3 determined the subgroups on the basis of the single-link procedure, and Method 4 determined the subgroups on the basis of the complete-link procedure.

For both Methods 3 and 4, the partition used was that midway between the two extreme sets of clusters (i.e., the partition where all individuals are in separate groups, and the partition where all individuals are in the same group). The number of clusters to use in an analysis must be determined by the researchers for the data at hand. This can be even more problematic than the determination of the number of factors in factor analysis, or the number of dimensions in multidimensional scaling, because there is no counterpart to an eigenvalue plot. It was expected that the middle partition would be a good guess at the modal choice of researchers.

Both the single-link and complete-link hierarchical clustering schemes were used. Either of these algorithms might describe the social psychology of subgroup formation. In complete-link clustering, a new object, or individual, does not join an existing cluster, or subgroup, until that object is sufficiently close to all objects currently in the cluster. This algorithm might be descriptive of behavior observed in close-knit groups; if one member of the existing group does not like the new member, the new member would not be accepted.

In single-link clustering, a new object, or individual, can join an existing cluster, or subgroup, when that object is sufficiently close to at least one object currently in the cluster. This might also describe the behavior of some groups accepting a new member. Perhaps that new member became friends with an existing member, and the rest of the group also accepted that person because they respected the opinion of the current member.

Most of the remaining methods are based on information obtained through an eigensolution of the (row-centered) sociomatrix. The decomposition of $\mathbf{X'X} = \mathbf{VPV'}$ yielded eigenvectors as columns of $\mathbf{V}$ that described the partner mode, and the decomposition $\mathbf{XX'} = \mathbf{UAU'}$ yielded eigenvectors as columns of $\mathbf{U}$ that described the actor mode ($\mathbf{VV'} = \mathbf{V'V} = \mathbf{UU'} = \mathbf{U'U} = \mathbf{I}$; $\mathbf{P}$, $\mathbf{A}$, diagonal).[2] The row means were subtracted out of the matrix before the decomposition, rather than the column means or both the row and column means. Sociometric data are often collected in such a way that the number of choices per actor is constrained (e.g., "Pick your three best friends"). If such a constraint had been imposed by the data collection method, row centering should not disturb the structure of the matrix to the extent that other centerings might. The unscaled eigenvectors were used rather than the principal components (eigenvectors multiplied by the square root of the corresponding eigenvalue). If subgroups are defined by splitting the actors according to whether their loadings are positive or negative, then the magnitudes of the loadings (scaling the variance to equal one or the eigenvalue) should be irrelevant. The unscaled eigenvectors should contain the fundamental information on the structure of the subgroups.

Methods 5–7 proceeded with the eigensolution as if the true number of eigenvectors to use were unknown, because this is the situation in which researchers would find themselves. Eigenvalues for the actors' and partners' modes were plotted, and the relatively largest q_i actor vectors and q_j partner vectors were retained. In Method 5, the pairs of q_i actor vectors were plotted and studied for a subjective judgment of subgroups. In Method 6, the pairs of q_j partner vectors were plotted and studied, and in Method 7, the pairs of q_i and q_j actor and partner vectors were plotted together and studied.

Methods 8–12 are the objective counterparts to Methods 5–7. Method 8 takes the partition that results from using the plus and minus loadings rule on the q_i actor vectors. That is, the first

[2] This might have been approached as a singular value decomposition of $\mathbf{X}$ ($\mathbf{X} = \mathbf{USV'}$), but treating the modes separately and analogously allows for simpler generalization to more modes. For example, researchers might observe a network on more than one relation and/or at more than one time, and these data might be modeled by way of three- or higher-mode eigensolution-based models (as in Kroonenberg, 1983; Lastovicka, 1981; Tucker, 1966).

split into two groups is formed by taking the first actor vector and putting those individuals with negative loadings into one group and those individuals with positive loadings into another. These splits were made finer by continuing to split by the positive and negative loadings on vectors 2, 3, . . . , q_i. Method 9 takes the same plus or minus approach using the q_j partner vectors, and Method 10 used the plus or minus rule using the first eigenvectors for the actors and for the partners modes.

It became obvious during the analyses that using the plus or minus rule for q_i actor vectors or q_j partner vectors (in Methods 8 and 9) led to too many subgroups. That is, it was often nearly the case that all individuals were placed into separate subgroups. Method 11, then, was the same as Method 8, except that the plus or minus rule was applied only to the first two actors' eigenvectors. (If q_i was one, then Method 11 was applied only to the first eigenvector: the minimum of $[2, q_i]$.) Method 12 was the same as Method 9 except, again, the plus or minus rule was applied to only the first, or the first and second, partners' eigenvectors (min $[2, q_j]$).

In Methods 13 and 14, the q_i actor vectors and the q_j partner vectors were used as the coordinates in a $q_i + q_j$-dimensional space, and the interpoint distances were computed. This matrix of distances was input to hierarchical clustering. Method 13 used the single-link algorithm, and Method 14 used the complete-link algorithm; and again, the partition taken was that intermediate to the cluster that included all individuals, and to the partition where each individual was in a separate cluster.[3]

In Method 15, the actors' and partners' component scores were plotted, and these plots were judged subjectively for subgroups. This method was investigated for the 8×8 sociomatrices, but not for the 16×16 sociomatrices. Because the eigenvector-related approaches were sufficiently complex, it was decided that the eigenvectors should be well understood before any more functions of those eigenvectors were investigated. Also, the information provided by these component scores was very similar to the eigenvectors in the sense that the same subgroups usually resulted.

The final method was again CONCOR, but instead of taking the last, finest split, the partition retained for this method was the intermediate partition. This partition might be more comparable to the partitions derived through the hierarchical clustering methods than the first method, which was the complete split on CONCOR. All 16 methods are listed in Table 3.

One method that has been investigated by other researchers that has not been included here is multidimensional scaling (Breiger et al., 1975; Faust, 1985). A matrix of interperson distances had been computed from the eigenvectors and from the log-linear model parameters, and it would be just as straightforward to model these distances by a scaling solution as by the clustering solutions. However, the scaling solution would only provide a plot of individuals based on these distances, and the judgments of subgroup membership would still need to be subjective. This method could be made more objective if the scaling-model-derived distances were input to a clustering procedure, but then the scaling seems to be an extraneous step.

This list of methods is not exhaustive; other methods that have not been incorporated into this study include those based on roles and positions or feedback cycles. Instead, the 16 methods chosen are variants of block modeling and algorithms expected to detect structural equivalence. Thus, the patterns expected in the data are several. First, the relations are not reflexive (i.e., the diagonal of the sociomatrix is ignored or assumed to be zero). Second, although the patterns of cliques, or maximal complete subgraphs (where all subgroup members interact), are perceived by some researchers as ideal, the definitions of structural equivalence and stochastic equivalence are not as restrictive. The patterns of structure sought are simply described by similarities between subgroup members of ties made and received. This does not necessarily imply that all intrasubgroup relations exist, and it does not preclude the existence of intersubgroup relations. Although the label *partition* might suggest the latter, in this article it is meant only to represent the classification of individuals into subgroups, such as a level of clusters in a hierarchical clustering solution.

In summary, three primary types of group-detection algorithms were focused on. Two of these types of algorithms were statistical in nature: those methods using the parameter estimates from log-linear models and those methods based on an eigenvalue decomposition. The remaining methods were not statistical, but were based on CONCOR, a method popular among network researchers, and one expected to be related to the eigen-based procedures.

---

[3] There are three related issues: First, none of these eigensolution-based methods is exactly identical to that described in Schwartz (1977); however, the differences should be negligible. His pooling the actors' and partners' covariance matrices should be most analogous to the methods in this study that use both vectors of the actors' mode and vectors of the partners' mode.

Second, Methods 5–14 rely on the eigenvectors associated with the first few large eigenvalues. After this Monte Carlo had been completed, Phipps Arabie brought an article by Chang (1983) to my attention as being potentially relevant. Very briefly, this article suggested that "information" about the distance between two multivariate normal populations might not be contained in only the earlier components, but that the later, usually discarded, components might also be informative.

Several notes can be made. First, one does not have multivariate normality in the current context. However, even if the findings in Chang (1983) were directly applicable, then retaining more (or all) vectors might provide better estimation (better subgroup recovery in this context) than that found using the present methods. If retaining more vectors would allow for better recovery, than the methods investigated here are simply more conservative. In addition, the usual argument holds that if all vectors had been kept, the dimensionality of the problem would not have been simplified. In fact, in the present application, using all vectors might simply have led to a too-finely divided partition where each individual was in his or her own subgroup. Finally, it should be noted that retaining only the first few vectors performed well, with respect to subgroup recovery, in this empirical investigation. Nevertheless, the potential applicability of Chang is worth investigation.

Third, in several of the methods, the eigenvectors were used to determine distance and, considering ellipsoids in space, it might be expected that variances should affect measures of distance. This would suggest that only the first few vectors would be necessary, and would further suggest that the vectors be scaled to reflect the different standard deviations (i.e., compute principal components). Chang's (1983) result suggests it may be inappropriate to place greater emphasis on (or scale with greater weights) the earlier vectors, thus, perhaps, strengthening the decision to proceed with unscaled eigenvectors rather than with principal components.

Table 3
List of Methods Used to Recover Subgroup Structure

Method		Description
1.	O	CONCOR full split
2.[a]	S	Fit log linear model to *Y* arrays, plot alpha vs. beta, determine groups by eye
3.[a]	O	Fit log linear model to *Y* arrays, use alpha and beta to compute distances, single-link clustering
4.[a]	O	Fit log linear model to *Y* arrays, use alpha and beta to compute distances, complete-link clustering
5.	S	Eigensolution, q_i actors' plots by eye
6.	S	Eigensolution, q_j partners' plots by eye
7.	S	Eigensolution, $q_i \times q_j$ actor and partner plots eye
8.	O	Eigensolution, q_i actors, + or − rule
9.	O	Eigensolution, q_j partners, + or − rule
10.	O	Eigensolution, $\text{actor}_1 \times \text{partner}_1$, + or − rule
11.	O	Eigensolution, $\min(q_i, 2)$ actors, + or − rule
12.	O	Eigensolution, $\min(q_j, 2)$ partners, + or − rule
13.	O	Eigensolution, q_i actors and q_j partners eigenvectors used to compute distances, single-link clustering
14.	O	Eigensolution, q_i actors and q_j partners eigenvectors used to compute distances, complete-link clustering
15.[a]	S	Eigensolution, actors and partners component scores, plotted, examined by eye
16.	O	CONCOR middle split

Note. S = Subjective method, O = Objective method. Diagonal entries were skipped in the CONCOR algorithm (Methods 1 and 16) and treated as zeros in the remaining methods.
[a] These methods were not fit to the 16 × 16 sociomatrices.

Simulation Study Results

These results were analyzed by using a simple analysis of variance (ANOVA). The dependent variable was the measure of partition similarity described earlier, predicting the true subgroup structure from the observed subgroup structure. This measure is bounded by one, so analyses that were conducted on the raw measure were repeated on the arcsin, square-root transformed measure. However, this transformation did not qualitatively alter the results, so the results reported here are those based on the raw measure.

A second transformation was also used. The measure of similarity is bounded above by one, but it is not bounded below. Hubert and Arabie (1985) described negative measures as uninterpretable. Examining the formula for the measure's computation, it is clear that a negative score will result when the cell entries are all small. These entries will be small when the table is large (i.e., many subgroups are derived for a given number of individuals) and there is poor recovery, so the frequencies are scattered over the table instead of collected in only a few cells. So although negative scores are hard to interpret, they do occur. The second transformation, then, was to simply set all negative scores equal to zero. This transformation also did not appreciably alter the results, so once again the results reported here are those based on the raw measure.

There were six explanatory variables for the ANOVA. The first five factors were those used in generating the data: number of actors (8 or 16), number of subgroups (2 or 4), equal or unequal number of individuals in each subgroup, the dyadic relationship structure (mostly mutual and null ties, or mostly asymmetric ties), and the clarity of subgroup structure (clear, medium, not clear). The final factor was simply which method (1–16) was used in attempting to recover the true subgroups. This is a six factor 2 × 2 × 2 × 2 × 3 × 16 ANOVA, with one observation per cell.

There were too many factors in this ANOVA to estimate all the interactions (a SAS constraint, not a theoretical constraint). When the methods were sorted into the 5 that were subjective and the 11 that were objective, all three-factor and lower-order interactions could be estimated. However, it was desirable to analyze all methods together. There was only 1 three-factor interaction that was significant in both of the separate analyses. (There were no interactions that were significant in one analysis but not in the other, which would have been suggestive of a higher-order interaction.)

The ANOVA reported here is based on all 16 methods, where the model includes main effects, all two-factor interactions, and the one large three-factor interaction. Table 4 lists the results for this ANOVA.

Figures 1–4 include the plots of all significant main effects and interactions. The significance level chosen was .05/22 = .002; the Bonferroni correction to alpha was .05 for the 22 tests listed in Table 4. Each significant effect is described. The first plot in Figure 1 is the plot of the means for the 16 methods. Several observations can be made, keeping in mind there are qualifying interactions among these factors. First, it should be noted that overall, no method performs very well. A score of 1.0 would represent perfect recovery, and these means are nearer .5. Note, however, that this is a main effect that aggregates over all levels of clarity of structure.

Second, the performance of both CONCOR methods (1 and 16) is average. It is not the best method and it is not the worst. Third, most of the methods that best recovered the subgroups were the subjective methods (e.g., Methods 5, 6, 7, and 15). (The best method was significantly different from the worst method, but no other significant differences were found. These statements are meant to describe the results and plots in detail. They are not intended to be overinterpreted as if some of the methods performed quite well and some did quite poorly.)

Fourth, the log-linear model approach did not recover subgroups very well, overall (see Methods 2, 3, and 4). This is partly due to the fact that the model estimates effects for actors and partners, as well as rho, the reciprocity effect. For the sociomatrices with many mutual and null ties, these rho parameters were highly significant (in the log-linear modeling), and the alpha and beta parameters were inconsequential. So the log-linear model is sensitive to the effects in the data, but a plot of alpha versus beta gives no information about rho, and recovery of subgroups suffered. If reciprocity were allowed to differ across dyads or subgroups, in a slightly more complicated log-linear model, the set of rho parameters could be plotted against alpha and beta; and then the log-linear methods would likely perform very well.

The best objective method was Method 9, the use of the plus

Table 4
Analysis of Variance Table—Dependent Variable Is Measure of Subgroup Recovery

Source	*df*	SS(III)[a]	*MS*	*F*	*p*
method	15	2.734	.182	4.31	.000*
subg24	1	2.591	2.591	61.27	.000*
equaln	1	.008	.008	.19	.666
mnora	1	.013	.013	.30	.586
clarity	2	62.614	31.307	740.34	.000*
nact	1	2.570	2.570	60.78	.000*
method*subg24	15	1.662	.111	2.62	.000*
method*equaln	15	.816	.054	1.29	.206
method*mnora	15	2.115	.141	3.33	.000*
method*clarity	30	5.169	.172	4.07	.000*
subg24*equaln	1	.080	.080	1.89	.170
subg24*mnora	1	.094	.094	2.21	.137
subg24*clarity	2	.982	.491	11.61	.000*
equaln*mnora	1	.000	.000	.00	.954
equaln*clarity	2	.042	.021	.49	.612
mnora*clarity	2	.294	.147	3.48	.031
method*nact	11	.263	.024	.56	.857
subg24*nact	1	.042	.042	.99	.320
equaln*nact	1	.048	.048	1.14	.286
mnora*nact	1	.004	.004	.09	.769
clarity*nact	2	1.314	.657	15.53	.000*
method*subg24*mnora	15	2.418	.161	3.81	.000*
error	535	22.624	.042	(including pooled interactions)	
corrected total	671	122.749			

Note. The variable names represent the following factors: method—which of the 16 methods used to recover the subgroups; subg24—whether there were two or four subgroups; equaln—whether subgroups had an equal or unequal number of individuals; mnora—whether the dyadic relations were mostly mutual and null or asymmetric; clarity—the clarity of the subgroup structure; nact—whether there were 8 or 16 individuals.
*$p < .05/22 = .002$.
[a] The Type III sums of squares (from SAS) were used as there were data missing from a few of the cells. (Methods 2, 3, 4, and 15 were not applied to the sociomatrices with 16 actors.) All mean squares used mean squared error as the divisor in the F statistic.

or minus rule on the q_j partner vectors. The next best objective method was Method 12, the use of the plus or minus rule on the first two partner vectors.

The partners' vectors were more helpful in subgroup recovery than the actor vectors. Compare, for example, Methods 9 and 12 on the partners mode to the analogous methods on the actor modes (Methods 8 and 11), which were two of the methods with poorest recovery. This difference in recovery is due to the way the data were designed. For the sociomatrices with many mutual and null ties, the matrix is symmetric, so the actors' vectors will be (and should be) roughly the same as the partners' vectors.

However, for the asymmetric sociomatrices in this study, the matrices differ mostly in the columns. If there is more variability among the partners, or columns, than among the actors, or rows, then it must be the case that the partner vectors would be more useful in recovering structure. The asymmetric ties would necessarily result in differences between the actor and partner vectors. If instead the sociomatrices had been designed to be asymmetric in the opposite way, then the actor vectors would have had to have been more informative than the partner vectors.

The fact that the partners' vectors perform better than the actors' vectors is not particularly interesting. If the asymmetric sociomatrices had been transposed, then the actors' vectors would have been found to perform better than the partners' vectors. If a researcher was willing to determine which of the actor or partner modes was more variable, then he or she could proceed to determine subgroups by using only the relevant set of eigenvectors. However, it should not be assumed that a researcher would make this extra effort, and the method that is suggested in summary (later) includes both actor and partner vectors.

The next two best performing methods were Methods 12 and 13. These were the methods that used q_i actor vectors and q_j vectors to compute distances, which were then input to hierarchical clustering. Method 12 was single link and Method 13 was complete link. These are the methods that are used in the analyses of the real data sets to derive empirical subgroups.

There are three other significant main effects plotted in Figure 1. Two subgroups are indeed recovered better than four. Clear subgroups were recovered better than medium-clear subgroups, and not-clear subgroups were not recovered (i.e., the mean was near zero). Note that even the clear subgroups were not perfectly recovered. Perhaps this result is due to a "degeneracy" of the sort that there is little variability in these clear matrices. Finally, subgroups were recovered better for sociomatrices with 16 actors than for those with 8 actors.

A brief description of the seven interactions follows. The plot of the Method × Subgroup interaction in Figure 2 shows that although two subgroups are generally recovered better than four, Methods 8 and 11 do not recover well for either. These two methods, once again, were based on actors' vectors, and it should be understood by now that for the particular sociomatrices generated in this study, the actor vectors must be used together with the partner vectors for good recovery on the symmetric sociomatrices, and that the partner vectors should suffice for adequate recovery on the asymmetric sociomatrices.

The plot of the Method × Dyadic relationship structure (mutual and null, or asymmetric) interactions in Figure 2 shows again that subgroups for these asymmetric relations cannot be recovered by using only information on the actors' modes (Methods 8 and 11). It also once again shows that the use of the log-linear model's alpha and beta parameters is most beneficial when there are asymmetric dyadic relationships. Other than these few sets of points, it is true that the methods do approximately as well at recovering subgroups for asymmetric sociomatrices as for symmetric ones. That recovery is equally good without regard to the type of relational data is good news; a researcher does not need to first determine what type of relational data characterize the sociomatrix before trying to choose a method to derive subgroups.

The next interaction is plotted in Figure 3 and is the Method × Clarity of Subgroup Structure interaction. The most striking effects are these: First, even the clear subgroups were not perfectly recovered by many methods. Consistently, the methods that did not do well were Methods 8 and 11 (actor vectors) and Methods 3 and 4 (log-linear model-related methods).

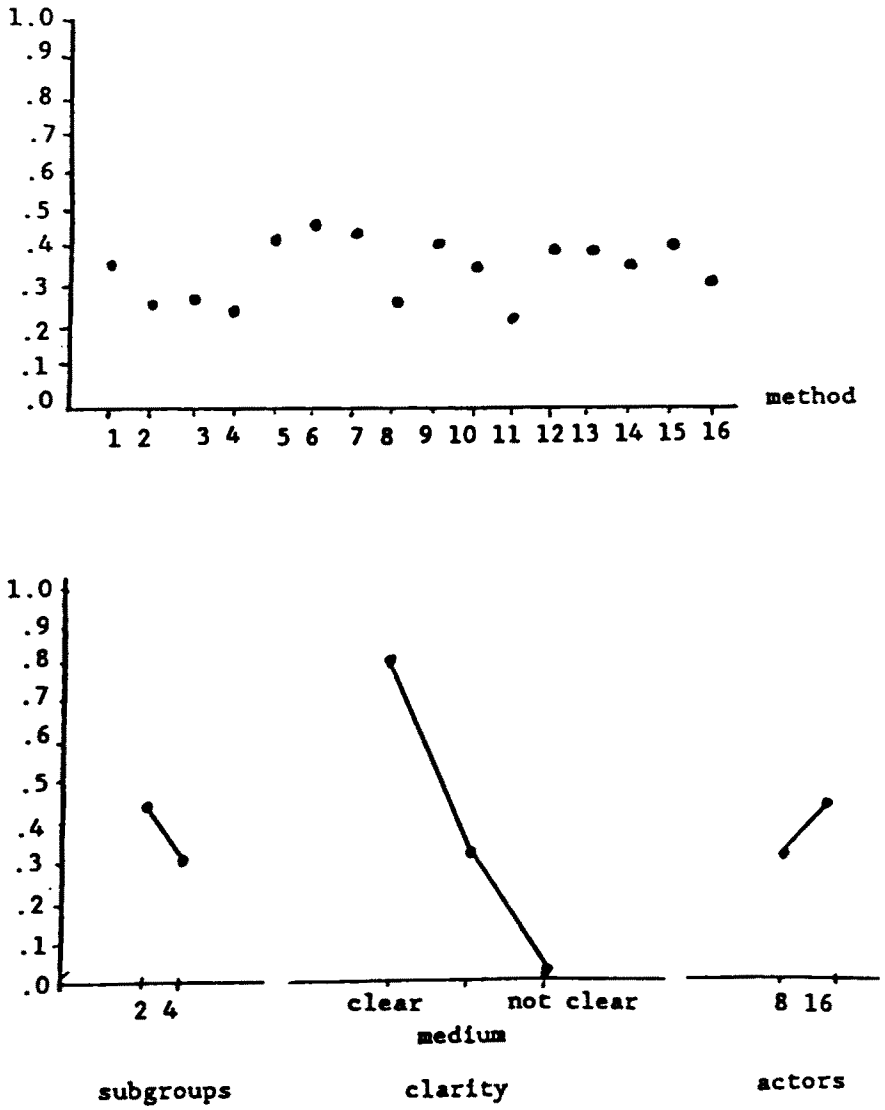

Figure 1. Simulation study—subgroup recovery: Plots of means for significant main effects.

Second, the not-clear subgroup structures were not recovered by any method. The profile for this level of the clarity factor vascillates around zero. What this says is that in the probabilities that defined the clarity factor (1.0, .8, .6), the .6 was too noisy. That is, .6 is only .1 from totally random data, and evidently is still too noisy to hope for recovery of subgroups.

In the same spirit, most data will not show perfectly clear subgroup structure; if they did, the patterns would be discernible by eye, and these methods would not be necessary. The level of error in real data must be somewhere between 1.0 (perfectly clear) and .5 (totally random), so the middle profile might describe the real data situation most accurately. There are very few differences between this middle profile and the plot of the method main effects, suggesting that the summary discussion about the methods that do best overall holds here and might also be relevant for application to real data.

The interaction in Figure 3 is the Number of Subgroups × Clarity of Structure interaction. The effect is slight, but understandable. The two-subgroups structure is recovered better than the four-subgroups structure when the subgroup structure is clear or medium clear. Not surprisingly, when there is no clear subgroup structure, methods cannot recover two groups any better than they can recover four.

The last two-factor interaction is plotted in Figure 4 and is that between the number of actors and the clarity of the subgroup structure. In this plot, one can see that the recovery for clear and not clear improve at the same rate when going from 8 actors to 16 actors. The recovery for medium-clear structure is improved more with the additional actors, which is to say, with real data more observations make the effects clearer.

The final interaction is the three-factor interaction at the bottom of Figure 4. This effect depends on whether there were two or four subgroups, whether the dyadic relations were mutual and null or asymmetric, and which of the 16 methods were used

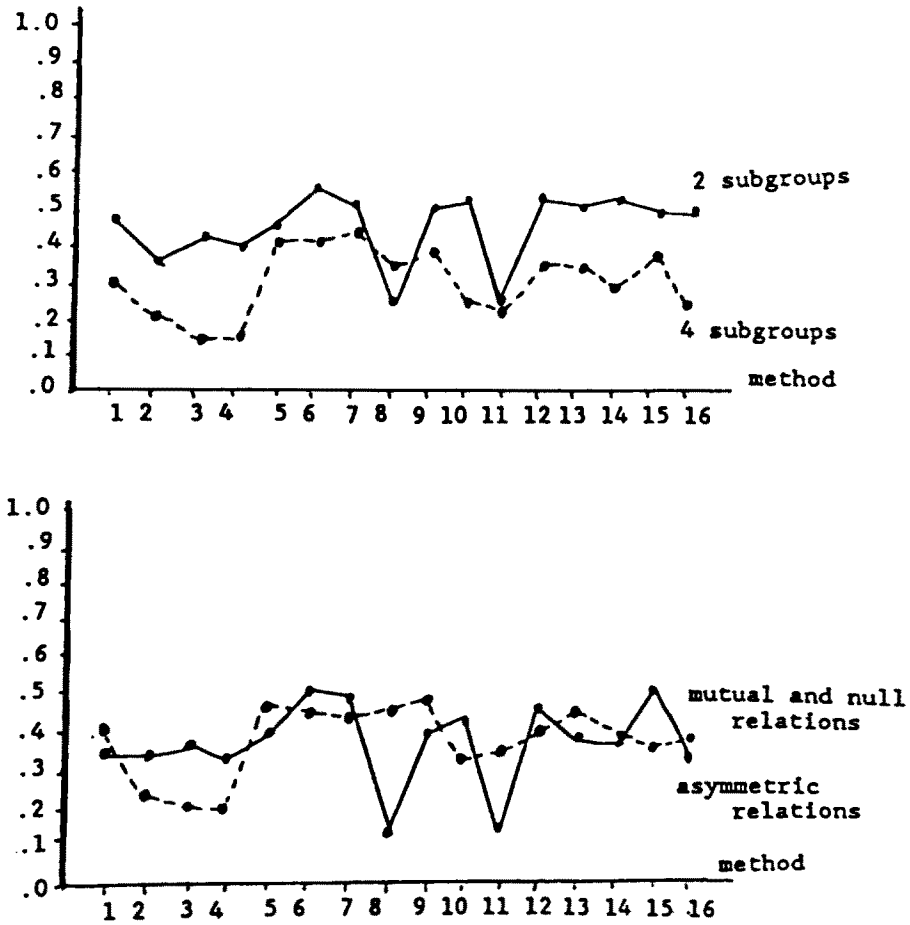

Figure 2. Simulation study—subgroup recovery: Plots of means for interactions, Part 1 of 3.

in recovery. There are some patterns in these figures that make sense. The effects for both CONCOR methods (Methods 1 and 16) appear similar to each other. The effects that use the actor vectors (Methods 8 and 11) appear similar to each other. Several of the plots for the subjective methods appear similar (e.g., Methods 6, 7, and 15). The single-link and complete-link pairs of log-linear-based methods (Methods 3 and 4) and the single-link and complete-link pairs of four-mode vector methods (Methods 13 and 14) appear similar, within the pair.

By way of summarizing these results, a practical recommendation of a method for the derivation of subgroups is offered. The objective methods that performed best were those relying on the partners' eigenvectors using the plus or minus rule, and q_j or $\min(2, q_j)$ vectors. However, the researcher might not know whether there is greater variability among the partners, as is the case in the asymmetric sociomatrices in this study, or among the actors.

Adding the actor vectors would at worst only be redundant with partner vectors, as in the case of a symmetric sociomatrix. In general, if the actors' vectors form the basis of a q_i-dimensional space, and the partners' vectors form the basis of a q_j-dimensional space, the union of the actors' and partners' vectors will describe a space with fewer than or equal to $q_i + q_j$ dimensions. For these reasons, the subgroup structure should be more clearly defined by the inclusion of vectors from both the actors' and partners' modes.

The methods that combined both sets of vectors and performed best were Methods 13 and 14, which used the q_i actor vectors and q_j partner vectors to compute an interperson distance matrix, which was then input to a single-link (Method 13) or complete-link (Method 14) hierarchical clustering procedure. Because the single-link procedure recovered slightly better than the complete-link procedure, the single-link algorithm is the one used to derive subgroups in the real data sets.

Analyses of Two Real Data Sets

Sampson's Monastery

Sampson collected data from monks on eight relations at five points in time (Sampson, 1968). The set of monks who interacted in these networks changed over the five time points, because of philosophical changes in the church. In order to work with a set of individuals that was constant over time, Sampson's Times 2, 3, and 4 were chosen to be modeled. Each of these networks describes the interactions among (the same) 18 monks. The number of actors, and partners, then, is 18.

The brothers ranked their top three choices on the following

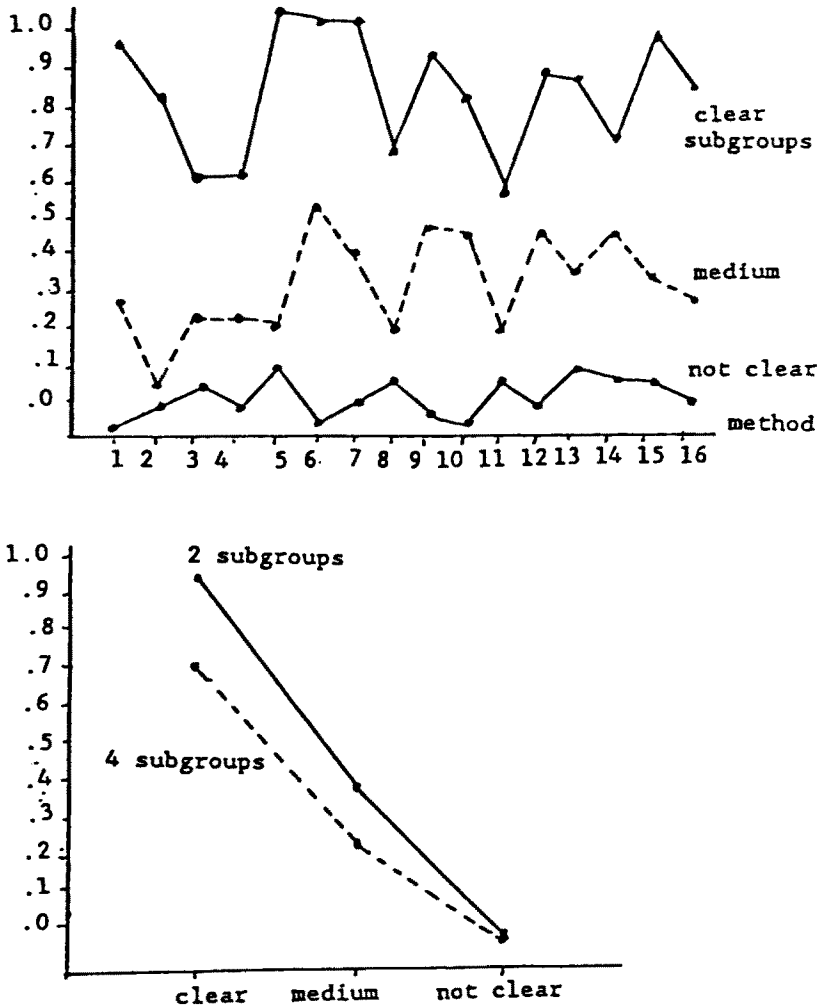

Figure 3. Simulation study—subgroup recovery: Plots of means for interactions, Part 2 of 3.

eight relations: like and dislike (affect), positive and negative esteem, positive and negative influence, and praise and blame (sanction). For the eight relations and three time points, the data take on the value 0 for the brothers not chosen, and the partner nominated as the most liked (for example) got a score of 3, the second most liked got a score of 2, and the third most liked got a score of 1.

Sampson (1968) provided a detailed description of several of the monks, and these characteristics have been used by researchers to substantiate the block models they have derived from these data (Breiger et al., 1975; Faust, 1985; Wasserman & Anderson, 1987). However, there were no complete objective attribute data for these men. For example, there was no list of how old the men were or how long they had been in the seminary. Thus, it was not possible to form a priori subgroups on the basis of psychological characteristics or other theoretically interesting attributes. Clusters of individuals were derived by using the methods that were recommended as a result of the simulation study.

Together, the eight relations and three time points resulted in 24 sociomatrices of the order 18 × 18. This 8 × 3 × 18 × 18 "super-sociomatrix" was modeled by using a four-mode eigensolution. In this study, the concern is not with the vectors describing the time or relations modes; instead, the focus is on the actor and partner modes. For the purpose of deriving subgroups, the obtained eigenvectors on the actors and partner modes from the four-mode model can be thought of as equivalent to aggregating over all eight relations and all three time points. An alternative would have been to choose to model one of the relations at one of the time points, to be closer in spirit to the simulated, single sociomatrices. However, it seemed more interesting and informative to obtain a more global (across relations and times) subgroup structure for these men.

The number of actor vectors retained in the modeling was $q_i = 2$, and the number of partner vectors retained in the modeling was $q_j = 3$, on the basis of the relative sizes of the eigenvalues in their respective modes. (In the decomposition of a standard two-way matrix, the eigenvalues of $\mathbf{X'X}$ equal those of $\mathbf{XX'}$, but $q_i \neq q_j$ here, as the analysis was conducted on a four-mode matrix. That is, the SSCP matrix for the partners was based on the rearranged four-mode matrix, not just a simple transposition of the matrix for the actors.) The five actor and partner vectors

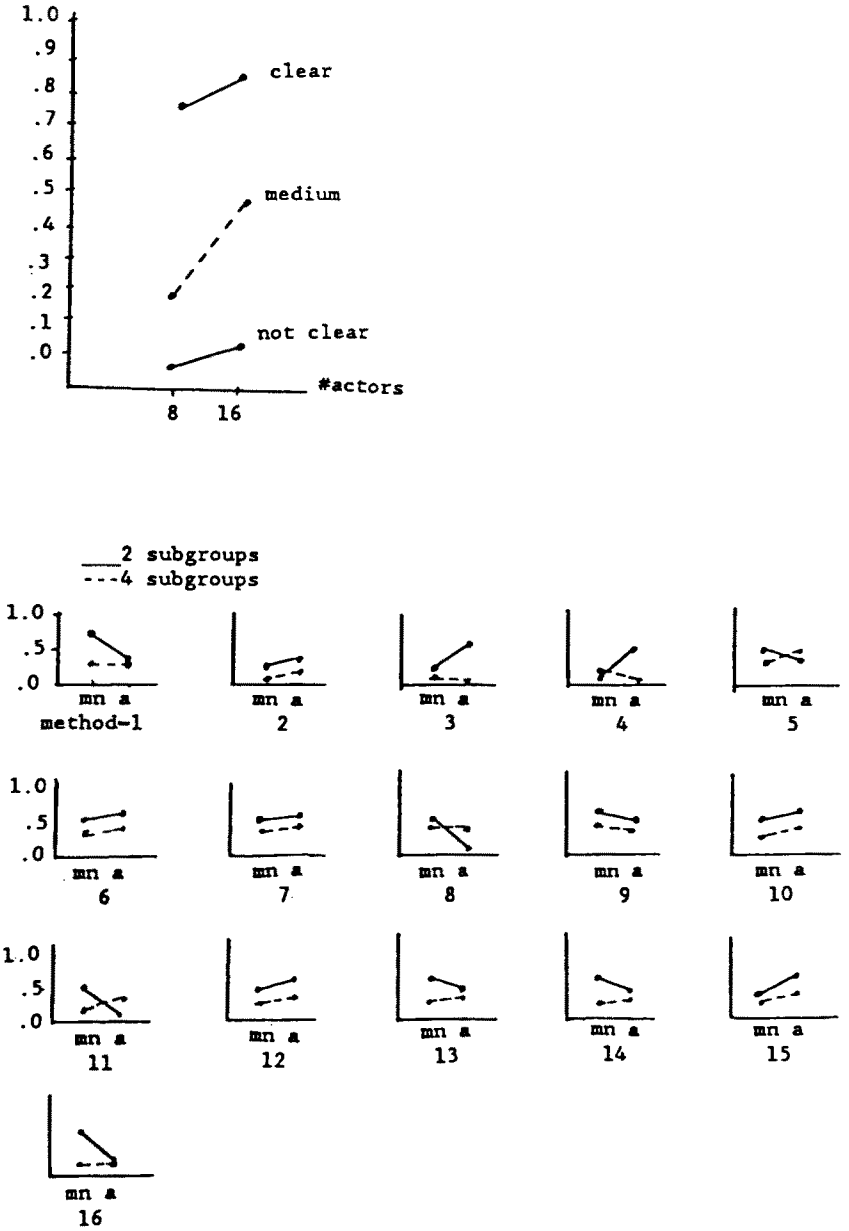

Figure 4. Simulation study—subgroup recovery: Plots of means for interactions, Part 3 of 3 (mn − a = mutual and null or asymmetric).

were used to derive subgroups on the individuals: Interindividual distances were computed over these five vectors, the distance matrix was analyzed using single-link clustering, and the partition taken was that intermediate between the two extreme partitions (i.e., all individuals in one group and all individuals in their own group). The resulting partition of individuals was (1), (2), (3, 17, 18), (4), (5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16), (13).

The obtained partition may be compared with the subgroups other researchers have derived. Breiger et al. (1975) listed several partitions. They used the CONCOR method to partition the individuals into groups as fine as the algorithm would continue to split. Because the partition given was not selected to be the finest split, it should be compared to a partition in the Breiger et al. article that was split to a comparable degree.

In the set of CONCOR partitions, the subgroups that have been split at the same intermediate point were groups consisting of the following individuals: (1, 2, 7, 15), (12, 14, 16), (4, 6, 8), (10, 11), (5, 9, 13), (3, 17, 18). The partitions resulting from the comparable split from single-link clustering and complete-link clustering (that Breiger et al., 1975, applied to the raw data) were, respectively, (5, 6, 8, 9, 10, 11, 4), (12), (7), (1, 14, 2, 15, 16), (13), (3, 17, 18) and (6, 8, 10), (4, 5, 9, 11), (1, 2, 7, 15), (12, 14), (16), (13, 3, 17, 18). Note at least the consistency with which the group of individuals (3, 17, 18) is identified. These

three monks were those characterized by Sampson (1968) as "outcasts."

The subgroups derived in this article seem to be of three types. First, the group of three isolates, or outcasts (3, 17, 18), form one group, because of their similar relations to the others. Second, the subgroup containing 11 of the 18 monks seems to suggest that most of the men behaved in a similar manner, or had high agreement in their choices.

The third type of subgroup is the set of four individuals who were in their own subgroups. They seem to relate to others in ways not common to any other monk. Monk 1 was identified as "arrogant" and Monk 2 was the first to leave the monastery (Reitz, 1982), so some of their atypical behaviors described in Sampson's (1968) observations are reflected in their being in separate subgroups.

The subgroups derived in the current study were not identical to any of the partitions obtained by Breiger et al. (1975). However, different analytical methods were used, and each method casts a slightly different picture of the interactions in the monastery. Furthermore, Breiger et al. analyzed only Time 4, whereas the analyses in this study were based on Times 2, 3, and 4, so the differences in the partitions might be in part because of subtle changes in the network over time. Finally, it should be noted that the three partitions Breiger et al. offered had little in common among themselves.

A final word with respect to comparing this partition to others' partitions is that the method used in this study was based on empirical work where there were known, true subgroup structures, and the method used here performed among the best. The "true" subgroup structure of the monastery cannot be known, but in this study, one is at least armed with a method that performed relatively well in discovering structure. The partitions of Breiger et al. (1975) or others (e.g., Faust, 1985; Wasserman & Anderson, 1987) are standards for comparison in the sense that they are documented, previous attempts at finding interpretable subgroups, but they are not standards in the sense that they are necessarily closer in truth to the monastery structure.

Referral Network of a Piano Tuner

This second example is included for three reasons. First, it is probably new to most network researchers. Second, it is a network in the everyday use of the term (e.g., business referral network, social support network). Third, it is a large, sparse sociomatrix, and it would be informative to know how the eigensolution-based method of subgroup derivation performs. (Although, like the Sampson [1968] data, one cannot know the "true" structure, one can only infer the sensibility of the subgroups from other information.)

The researchers collecting the data used in this example were trying to understand how persons in the service industries, such as piano tuners, get their business (Reingen & Kernan, 1986). The researchers contacted each of a piano tuner's recent clients and asked them how they had come to learn of the piano tuner. Many of the contacts were persons who were recommended the piano tuner after they had contacted a particular music store asking for a recommendation, or after they had purchased a piano from the store. Many of these contacts went on to recommend the piano tuner to a friend of theirs. This part of the network contained many ties that were two units long: the music store recommending the piano tuner to the first person, and the first person referring a second person to the piano tuner. The other part of the network contained the piano tuner's personal network, including five music professionals (i.e., piano teachers) who referred more than one client to the tuner.

The relational ties included in the sociomatrix are the business referrals only, not the friendship ties among some of the players. There were 89 actors and partners, and of the $89(89 - 1)/2 = 3{,}916$ possible ties, there were 88 observed dyads. Note that the actors included such diverse nodes as persons (the central figure of the piano tuner, the key figures of the music professionals, and all the remaining clients) and an institution (the music store). Note also that the relation is unidirectional: Actor i refers the tuner to partner j, and it would not be sensible to think of partner j then "mutually" referring the tuner to actor i, because presumably actor i has already heard of the tuner.

The referral ties are listed in Table 5. The data are not presented in the form of a sociomatrix because the matrix is large (89×89), and the information can be compressed because the sociomatrix is sparse (i.e., there are few ties). The arrows in the table are in the direction of the business referral. For example, the piano tuner contacted person 78 (one of the music professionals), who contacted three people: persons 179, 159, and 80. Of these contacts, person 80 made another referral, to person 81. Note the importance of both the tuner and the music store in making referrals, in terms of the number of ties originating from these nodes. The music professionals are also key players, because they generated more than one referral each.

The majority of the referrals originating from the tuner or the store are two steps (e.g., tuner to 40 to 25, or store to 39 to 4). There are several exceptions. There are two cases where the music store's referral to person j (j = 46 and 189) led to two more consecutive ties. There are also two cases where the store referred the tuner to person j (j = 173 and 21), and j directly referred the tuner to two other persons. A similar case originating with the tuner is the referral flow from the tuner to 17 to 70 and 47. There are also several cases where the referral made from the tuner or the store stops with person j (j = 181, 52, and 85).

The subgroups for the 89 actors were derived by the same method described for Sampson (1968; i.e., eigensolution and clustering), and the resulting partition (with M = music store, PT = piano tuner, and MP = music professional) follows: (M), (PT), (MP5), (MP1, MP2), (MP4), (17), (173, 21), (MP3), (181, 40, 9, 44, 151, 63, 167, 102, 155, 12), (46, 39, 37, 157, 66, 106, 93, 133, 6, 176, 189, 131, 141, 138, 34, 109, 175, 52, 85), and (all others).

These subgroups are easily identified. The music store and the piano tuner are in their own groups, which is appropriate because their volume and pattern of referrals are unlike those of other actors. The music professionals (MP5, MP4, and MP3) are also in isolated clusters. (However, in the next level in the hierarchical clustering, MP3 joins the group that contains "all others." Perhaps this is due to the fact that this MP makes fewer referrals than the other music professionals.) The first and second music professionals are in a group together; their patterns

Table 5
The Referral Ties of a Piano Tuner

Piano tuner (PT) →	78 (MP1) →	179, 159	
	78 (MP1) →	80 →	81
PT →	142 (MP2) →	99, 111	
	142 (MP2) →	177 (MP3) →	100, 154
PT →	126 (MP4) →	171, 130, 18, 82	
PT →	40 →	25	
PT →	9 →	8	
PT →	44 →	165	
PT →	151 →	75	
PT →	63 →	76	
PT →	167 →	92	
PT →	102 →	148 (MP5) →	48, 45
		148 (MP5) →	84 → 112
		148 (MP5) →	60 → 32
PT →	155 →	24	
PT →	12 →	71	
PT →	17 →	70, 47	
PT →	181		
Music store (M) →	46 →	58 →	147
M →	189 →	35 →	159
M →	173 →	172, 89	
M →	21 →	103, 50	
M →	39 →	4	
M →	37 →	158	
M →	157 →	161	
M →	66 →	156	
M →	106 →	180	
M →	93 →	115	
M →	133 →	98	
M →	6 →	2	
M →	176 →	132	
M →	131 →	105	
M →	141 →	139	
M →	138 →	160	
M →	34 →	33	
M →	109 →	94	
M →	175 →	3	
M →	52		
M →	85		

Note. To be consistent with Reingen and Kernan (1986), the actors are represented by the same nominal code as in their original network diagram. The seven special players have additional codes for easier reference. The music store is *M*, and the five music professionals are denoted MP1 through MP5. The piano tuner was called *J* in Reingen and Kernan, but in an Actors × Partners sociomatrix, *j* is often used to refer to the columns, or partners, so in this table the piano tuner is represented by *PT*.

of referrals are fairly similar in that they both refer three persons, one of whom makes additional referrals.

The groups (17) and (173, 21) are those persons who referred the tuner to two others. They are not clustered together because the piano tuner had contacted person 17, whereas the music store made the referral for persons 173 and 21.

The cluster starting with persons 181 and 40 contains all the contacts made directly by the piano tuner, with the exceptions of the music professionals and person 17. The cluster starting with persons 46 and 39 contains all the contacts made directly by the music store, except persons 173 and 21. The remaining group contains the indirect contacts of the tuner, the store, and the other players.

Note that these methods were insensitive to the three-step referrals (M to 46 to 58 to 147, and M to 189 to 35 to 159), in that persons 46 and 189 did not form a unique subgroup. Perhaps these patterns of referrals were not identified because a decomposition of the matrix had been taken at a second power (**XX**) and not some third power. The methods were also unable to identify the single-step referrals (PT to 181, M to 52, and M to 85). Perhaps these latter referrals were simply dominated by the longer, two-step ties and the fuller, multiple referral ties.

However, the majority of the data were well represented by the structure of these subgroups. These results were a pleasant surprise given the sparseness of the matrix. The key players were identified as being different from the other actors, and the larger subgroups were interpretable given the patterns in the relational data.

Summary

This article contained a simulation study that was an investigation of several methods for deriving empirical subgroups for individuals in a network. In generating the sociomatrices, an attempt was made to include many of the factors that might be relevant to network research. The results and discussion of the study are, of course, conditional on the factors chosen and on the particular levels at which each factor was evaluated.

Sociomatrices were simulated to have known subgroup structures, as well as other known properties. Each of several methods was used in attempting to recover the true subgroup partition.

On the basis of the simulation study's results, the method that was used to derive subgroups in the real data sets can be summarized in the following steps. The eigensolutions for the actors' and partners' modes are derived. The eigenvalues from these solutions are plotted to determine the number of vectors to retain for each mode. These q_i actor vectors and q_j partner vectors are used as axes in a $q_i + q_j$-dimensional space. The interindividual distances are computed over these dimensions. This distance matrix is input to a single-link hierarchical clustering algorithm. The partition retained was intermediate to (i.e., roughly one half to two thirds of the way from) the partition where all individuals are in their own cluster and the partition where all individuals are in a single cluster.

The subgroup recovery was also compared along the factors that generated the data. Two subgroups were recovered better than four. Subgroups were recovered better when there were more individuals (16 vs. 8). Subgroups were recovered roughly as well for networks characterized by symmetric or asymmetric relations. Subgroups that contained varying numbers of individuals were recovered roughly as well as subgroups that contained a constant number. Finally, subgroups were recovered more accurately when the data contained less error.

Applying these methods to the Sampson (1968) data led to subgroups that were sensible, although not identical to partitions derived by previous researchers. Applying these methods to the piano tuner's referral network resulted in subgroups that seemed to accurately identify key players and patterns of interaction.

References

Allison, P. D., & Liker, J. K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. *Psychological Bulletin, 91,* 393–403.

Breiger, R. L., Boorman, S. A., & Arabie, P. (1975). An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology, 12,* 328–383.

Budescu, D. V. (1984). Tests of lagged dominance in sequential dyadic interaction. *Psychological Bulletin, 96,* 402–414.

Chang, W. (1983). On using principal components before separating a mixture of two multivariate normal distributions. *Applied Statistics, 32,* 267–275.

Clark, J. A., & McQuitty, L. L. (1970). Some problems and elaborations of iterative, intercolumnar correlational analysis. *Educational and Psychological Measurement, 30,* 773–784.

Dillon, W. R., Madden, T. J., & Kumar, A. (1983). Analyzing sequential categorical data on dyadic interaction: A latent structure approach. *Psychological Bulletin, 94,* 564–583.

Dumas, J. E. (1984). Interactional correlates of treatment outcome in behavioral parent training. *Journal of Consulting and Clinical Psychology, 52,* 946–954.

Dumas, J. E. (1986). Controlling for autocorrelation in social interaction analysis. *Psychological Bulletin, 100,* 125–127.

Faraone, S. V., & Dorfman, D. D. (1987). Lag sequential analysis: Robust statistical methods. *Psychological Bulletin, 101,* 312–323.

Faust, K. (1985). *A comparative evaluation of methods for positional analysis of social networks.* Unpublished doctoral dissertation, University of California, Irvine.

Feick, L. F., & Novak, J. A. (1985). Analyzing sequential categorical data on dyadic interaction: Log-linear models exploiting the order in variables. *Psychological Bulletin, 98,* 600–611.

Fienberg, S. E., Meyer, M. M., & Wasserman, S. S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association, 80,* 51–67.

Fienberg, S. E., & Wasserman, S. S. (1981). Categorical data analysis of single sociometric relations. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 156–192). San Francisco: Jossey-Bass.

Galaskiewicz, J., & Wasserman, S. (1987). *Social action models for the study of change in organizational fields.* Unpublished manuscript.

Gottman, J. M. (1979). *Experimental investigations of marital interactions.* New York: Academic Press.

Gottman, J. M., & Bakeman, R. (1979). The sequential analysis of observational data. In M. E. Lamb, S. J. Suomi, & G. R. Stephenson (Eds.), *Social interaction analysis: Methodological issues* (pp. 185–206). Madison: University of Wisconsin Press.

Gottman, J. M., & Notarius, C. (1978). Sequential analysis of observational data using Markov chains. In T. R. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (pp. 237–285). New York: Academic Press.

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability densities for directed graphs. *Journal of the American Statistical Association, 76,* 33–51.

Hubert, L. J., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2,* 193–218.

Iacobucci, D., & Wasserman, S. (1987). Dyadic social interactions. *Psychological Bulletin, 102,* 293–306.

Iacobucci, D., & Wasserman, S. (1988). A general framework for the statistical analysis of sequential dyadic interaction data. *Psychological Bulletin, 103,* 379–390.

Kenny, D. A. (1981). Interpersonal perception: A multivariate round robin analysis. In M. B. Brewer & B. E. Collins (Eds.), *Knowing and validating in the social sciences: A tribute to Donald T. Campbell* (pp. 288–309). San Francisco: Jossey-Bass.

Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in the analysis of variance. *Psychological Bulletin, 99,* 422–431.

Kenny, D. A., & LaVoie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 141–182). New York: Academic Press.

Knoke, D., & Kuklinski, J. H. (1983). *Network analysis.* Beverly Hills, CA: Sage.

Kraemer, H. C., & Jacklin, C. N. (1979). Statistical analysis of dyadic social behavior. *Psychological Bulletin, 86,* 217–224.

Kroonenberg, P. M. (1983). *Three-mode principal component analysis.* Leiden, The Netherlands: DSWO Press.

Lastovicka, J. L. (1981). The extension of component analysis to four-mode matrices. *Psychometrika, 46,* 47–57.

Lorrain, F. P., & White, H. C. (1971). Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology, 1,* 49–80.

Margolin, G., & Wampold, B. E. (1981). Sequential analysis of conflict and accord in distressed and nondistressed marital partners. *Journal of Consulting and Clinical Psychology, 49,* 554–567.

McQuitty, L. L. (1968). Multiple clusters, types, and dimensions from iterative intercolumnar correlational analysis. *Multivariate Behavioral Research, 3,* 465–476.

McQuitty, L. L., & Clark, J. A. (1968). Clusters from iterative, intercolumnar correlational analysis. *Educational and Psychological Measurement, 28,* 211–238.

Mendoza, J. L., & Graziano, W. G. (1982). The statistical analysis of dyadic social behavior: A multivariate approach. *Psychological Bulletin, 92,* 532–540.

Noma, E., & Smith, D. R. (1985). Benchmarks for the blocking of sociometric data. *Psychological Bulletin, 97,* 583–591.

Reingen, P. H., & Kernan, J. B. (1986). Analysis of referral networks in marketing: Methods and illustration. *Journal of Marketing Research, 23,* 370–378.

Reitz, K. P. (1982). Using log linear analysis with network data: Another look at Sampson's monastery. *Social Networks, 4,* 243–256.

Sampson, S. F. (1968). *A novitiate in a period of change: An experimental and case study of social relationships.* Doctoral dissertation, Department of Sociology, Cornell University.

Schwartz, J. E. (1977). An examination of CONCOR and related methods for blocking sociometric data. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 255–282). San Francisco: Jossey-Bass.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31,* 279–311.

Wasserman, S. (1987). Conformity of two sociometric relations. *Psychometrika, 52,* 3–18.

Wasserman, S., & Anderson, C. (1987). Stochastic a posteriori blockmodels: Construction and assessment. *Social Networks, 9,* 1–36.

Wasserman, S., & Galaskiewicz, J. (1984). Some generalizations of p_1: External constraints, interactions, and non-binary relations. *Social Networks, 6,* 177–192.

Wasserman, S., & Iacobucci, D. (1986). Statistical analysis of discrete relational data. *British Journal of Mathematical and Statistical Psychology, 39,* 41–64.

Wasserman, S., & Weaver, S. O. (1985). Statistical analysis of binary relational data: Parameter estimation. *Journal of Mathematical Psychology, 29,* 406–427.

White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social structure from multiple networks. I: Blockmodels of role and positions. *American Journal of Sociology, 81,* 730–780.

(Appendix follows on next page)

Appendix

Simulated Data for Monte Carlo Study

Table 1-A
Design Used to Generate Simulated Sociomatrices

Number of individuals per subgroup and level of clarity	2 subgroups		4 subgroups	
	m & n	asym	m & n	asym
	8 actors			
Equal				
Clear	a-8	d-8	m-8	p-8
Medium	b-8	e-8	n-8	q-8
Not clear	c-8	f-8	o-8	r-8
Unequal				
Clear	g-8	j-8	s-8	v-8
Medium	h-8	k-8	t-8	w-8
Not clear	i-8	l-8	u-8	x-8
	16 actors			
Equal				
Clear	a-16	d-16	m-16	p-16
Medium	b-16	e-16	n-16	q-16
Not clear	c-16	f-16	o-16	r-16
Unequal				
Clear	g-16	j-16	s-16	v-16
Medium	h-16	k-16	t-16	w-16
Not clear	i-16	l-16	u-16	x-16

The design according to which these 48 sociomatrices were generated appears in Table 1-A. The entry in each cell is a letter *a* through *x*, followed by the number 8 or 16, indicating the number of actors and partners in the network. The sociomatrices that are listed after this design summary are preceded by a header card that includes the letters *a* through *x* and the number 8 or 16 to indicate to which cell in the design the sociomatrix corresponds.

The notation for the factors in the design is as follows: *m & n* represents mostly mutual and null dyadic relational ties, *asym* represents mostly asymmetric ties, *equal* (or *unequal*) represents an equal (or unequal) number of individuals per subgroup, and the levels for the clarity of subgroup structure factor are denoted as *clear, medium,* or *not clear.*

The simulated sociomatrices follow.

a-8)
01110000
10110000
11010000
11100000
00000111
00001011
00001101
00001110

b-8)
01100000
10110000
00010011
11100010
01100110
00010001
00101101
00001110

c-8)
00010001
00010101
11011010
00100100
00100100
00111010
01000100
10110100

d-8)
01110000
10110000
11010000
11100000
11110000
11110000
11110000
11110000

e-8)
01100000
10110000
00010011
11100010
10010001
11101010
11010000
11110000

f-8)
00010001
00010101
11011010
00100100
11010011
11000001
10111001
01001010

g-8)
01000000
10000000
00011111
00101111
00110111
00111011
00111101
00111110

h-8)
01010000
10000000
11011100
00101101
01010110
00100001
00011101
00111110

i-8)
00100001
00100101
00010101
11101011
00010100
00001010
01110100
10000100

j-8)
01000000
10000000
11000000
11000000
11000000
11000000
11000000
11000000

k-8)
01010000
10000000
00000011
11000010
10100001
11011010
11100000
11000000

l-8)
00100001
00100101
11001010
00000100
11100011
11110001
10001001
01111010

m-8)
00110000
00110000
11001100
11001100
00110011
00110011
00001100
00001100

n-8)
00100000
00110000
00001111
11001110
01010010
00101001
00101100
00001100

o-8)
01010001
10010101
11000110
00001000
00010000
00000010
01000101
10110110

p-8)
01110000
10110000
11000000
11000000
11110000
11110000
00111101
00111110

q-8)
01100000
10110000
00000011
11000010
10010001
11101010
00011101
00111110

r-8)
00010001
00010101
11001010
00000100
11010011
11000001
01110100
10000100

s-8)
01111111
10100000
11000000
10000111
10000111
10011011
10011101
10011110

t-8)
01101111
10100000
00000011
10000101
11100110
10000000
10111101
10011010

u-8)
00011110
00000101
11001010
01000011
10100100
10101011
11010100
00100000

v-8)
00000000
00100000
01000000
00000111
00000111
01100000
01100000
01100000

w-8)
00010000
00100000
10000011
00000101
01100110
01111010
01000000
01100000

x-8)
01100001
10000101
01001010
11000011
00100100
01010001
00101001
11011010

A-16)
0111111100000000
1011111100000000
1101111100000000
1110111100000000
1111011100000000
1111101100000000
1111110100000000
1111111000000000
0000000011111111
0000000010111111
0000000011011111
0000000011101111
0000000011110111
0000000011111011
0000000011111101
0000000011111110

B-16)
0111011011010101
0011111000000000
1000010110100000
1110011100001000
1111001100100000
1101100100000001
1011010100110010
1111111010010100
0000010001101011
0000000010111001
0000100001011111
0010101001101111
0000100001000101
0000000001001011
0011000011011101
0010000011110110

I-16)
0110110111010101
0000100100000000
1000101010100010
1010101000001000
0000001101010111
0010100011110110
0101010011001101
0100101001101011
1000101101101011
1001101010111001
1000000100010101
0010000101101111
0001011101000101
1000010101000011
0011010111011101
1110110110100100

J-16)
0111000000000000
1011000000000000
1101000000000000
1110000000000000
1111000000000000
1111000000000000
1111000000000000
1111000000000000
1111000000000000
1111000000000000
1111000000000000
1111000000000000
1111000000000000
1111000000000000
1111000000000000

C-16)
0011001101000111
1010001101000101
0100111111101000
1100100011001100
0110011101011011
1101001110110101
0100110000010000
0101100001001011
0000000000111111
1010111110100100
0100000011011010
0001000010001101
0100111010110111
1110000110100001
1001001011101100
1011100111110000

D-16)
0111111100000000
1011111100000000
1101111100000000
1110111100000000
1111011100000000
1111101100000000
1111111010000000
1111111000000000
1111111100000000
1111111100000000
1111111100000000
1111111100000000
1111111100000000
1111111100000000
1111111100000000

K-16)
0110000010101010
1011101000000011
1001101000001000
0100000100100010
0111010000001101
0111000010000100
0111000000010000
1010001000001100
1010000001000001
1100010000000000
1100000000000000
1111000100100000
1111000000010000
1111000000010000
0011000100100001
1111000100000000

L-16)
0110110111010101
0000100100000000
1000101010100010
1010101000001000
1111010010101000
1101001100001001
1010100100110010
1011010010010100
0111010000010100
0110010100000110
0111111011001010
1101111010000000
1110100010110010
0111101010111000
1100101000100000
0001001001011010

E-16)
0110111110101010
1011010100000011
1001010100001000
0100111000100010
0111001100001101
0111101110000100
0111110100010000
1010110000001100
1010111101000001
1100101100000000
1100111100000000
1111111000100000
1111111100010000
1111111100010000
0011111000100001
1111111000000000

F-16)
0110001011010101
0000011000000000
1000010110100010
1010010100001000
1111001110101000
1101100000001001
1010010000110010
1011101010010100
0111101100010100
0110101000000110
0111000111001010
1101000110000000
1110011111010010
0111010110111000
1100010100100000
0001110101011010

M-16)
0000111100000000
0000111100000000
0000111100000000
0000111100000000
1111000011110000
1111000011110000
1111000011110000
1111000011110000
0000111100001111
0000111100001111
0000111100001111
0000111100001111
0000000011110000
0000000011110000
0000000011110000
0000000011110000

N-16)
0001111110101010
0000010100000011
0100010100001000
1010111000100010
0111010011111101
0111000001110100
0111000011100000
1010001011111100
0101111101001110
0011101100001111
0011111100001111
0000111000101111
0000000011100000
0000000011100000
1100000111010001
0000000111110000

O-16)
0001001011010101
1011011000000000
0101010110100010
0100010100001000
1111010001011000
1101001111111001
1010100111000010
1011010001100100
1000101100011011
1001101000001001
1000000111000101
0010000110001111
0001100001000010
1000101001001000
0011101011010000
1110001010101010

P-16)
0111111100000000
1011111100000000
1101111100000000
1110111100000000
1111000000000000
1111000000000000
1111000000000000
1111000000000000
1111111100000000
1111111100000000
1111111100000000
1111111100000000
0000111111110111
0000111111111011
0000111111111101
0000111111111110

Q-16)
0110111110101010
1011010100000011
1001010100001000
0100111000100010
0111010000001101
0111000010000100
0111000000010000
1010001000001100
1010111101000001
1100101100000000
1100111100000000
1111111000100000
0000111111100111
0000111111101011
1100111011011100
0000111011111110

R-16)
0110001011010101
0000011000000000
1000010110100010
1010010100001000
1111010010101000
1101001100001001
1010100100110010
1011010010010100
0111101100010100
0110101000000110
0111000111001010
1101000110000000
0001011101000101
1000010101000011
0011010111011101
1110110110100100

U-16)
0110001000101010
0000011011111111
1000011010100010
1010011000001000
1111000010101000
1101101100001001
1001100100001101
1000010010101011
0100010000101011
0101010100111001
0100110100010101
1110110101101111
1101101101000101
0100100101000011
1111100111011101
0010000110100100

V-16)
0000000000000000
0000000000000000
0001110000000000
0010110000000000
0011010000000000
0011100000000000
0000000000111111
0000000000111111
0000000000111111
0000000000111111
0011110000000000
0011110000000000
0011110000000000
0011110000000000
0011110000000000

S-16)
0111111111111111
1011111111111111
1101110000000000
1111011000000000
1111010000000000
1111100000000000
1100000000111111
1100000000111111
1100000000111111
1100000000111111
1100001111011111
1100001111101111
1100001111110111
1100001111111011
1100001111111101
1100001111111110

T-16)
0110111101010101
1011010111111100
1001011000001000
0100110100100010
0111000000001101
0111100010000100
0100000000101111
1001001000110011
1001000001111110
1111010000111111
1111001111011111
1100001011001111
1100001111100111
1100001111101011
0000001011011100
1100001011111110

W-16)
0001000010101010
0000101000000011
0101011000001000
1000110100100010
1011000000001101
1011100010000100
1000000000101111
0101001000110011
0101000001111110
0011010000111111
0000110000000000
0011110100100000
0011110000010000
0011110000010000
1111110100100001
0011110100000000

X-16)
0001110111010101
1011100100000000
0100011010100010
0110011000001000
0011000010101000
0001101100001001
0101100100001101
0100010010101011
1000010000101011
1001010100111001
1011001011001010
0001001010000000
0010010010110010
1011011010111000
0000011000100000
1101111001011010

Received October 14, 1987
Revision received February 5, 1989
Accepted April 20, 1989 ■