Social Networks 13 (1991) 105–140
North-Holland

A new stochastic path-length tree methodology for constructing communication networks

Jaewun Cho [a] and Wayne S. DeSarbo [b]

[a] *College of Business, Arizona State University, Tempe, AZ 85287,*
and [b] *School of Business Administration, University of Michigan, Ann Arbor, MI 48109-1234, USA*

Network analysis has become a popular method for identifying the communication structure in a system where positional and relational aspects are important. In this paper, a maximum likelihood based methodology is presented that allows for the analysis of binary sociometric data. This methodology provides a network representation via estimated path-length or additive trees that indicate the distance between all pairs of members. The methodology is distinguished from traditional hierarchical clustering based procedures by its direct consideration of the asymmetry in a typical communication process, the simultaneous representation of structural characteristics (e.g., clique membership, clique cohesiveness), and the identification of the specialized communication roles of each member (e.g., opinion leader, liaison). A penalty function algorithm is developed and its performance is investigated via a Monte Carlo analysis with synthetic data. An application examining information flows among managers is presented. Finally, directions for future research are suggested.

1. Introduction

A variety of network models have been developed to describe the sociometric structure of the members of small groups. Historically, network models have been proposed within two alternative analytical approaches. These approaches differ primarily in terms of the frame of

Jaewun Cho is an Assistant Professor of Marketing at the College of Business at Arizona State University of Tempe, Arizona. Wayne S. DeSarbo is the S.S. Kresge Distinguished Professor of Marketing and Statistics at the School of Business Administration at the University of Michigan in Ann Arbor, Michigan.

0378-8733/91/$03.50

reference within which an actor is analyzed. In a *relational approach*, network models describe the direct connection between pairs of actors. Sociograms (Northway 1949; Klovdahl 1981), matrix operations methods (Forsyth and Katz 1946; Hubbell 1965), graph-theoretic methods (Luce and Perry 1949; Harary *et al.* 1965; Luce 1950; Siedman and Foster 1978), and distance methods (Bock and Husain 1950) are typical models utilizing this relational approach. In such relational approaches, actors are aggregated in cliques to the degree that they are connected directly to each other by cohesive bonds.

A *positional approach* provides a different perspective to the analysis of subgroup structure, where actors are aggregated into a jointly occupied position or role to the extent that they have a common set of linkages to the other actors in a system. No requirement is imposed that the actors in the same position have direct ties to each other. Hierarchical clustering based on a measure of dissimilarity derived from interactions with other actors (Burt 1977) and block models (White *et al.* 1975; Brieger *et al.* 1975) are examples of the primary models operationalizing this positional approach.

In this paper, we exploit the rich tradition of subgroup-level network models and provide a new methodology for visually representing social networks that explicitly embeds single actor level analysis in a derived *path-length tree* structure. Our model can be regarded as one employing a relational approach. In the next section, we briefly review existing approaches to subgroup-level network analysis and formally outline our research goals. The technical details of the new proposed methodology and a Monte Carlo simulation study evaluating the performances of the algorithm are then presented. The usefulness of the model is illustrated with an empirical application concerning communication networks among managers within a firm (Krackhardt 1987). We conclude with a discussion of the implications of the methodology and suggest directions for future research.

2. Literature review

In reviewing and contrasting our new methodology with existing approaches to network structure analysis, we use the organizing framework shown in Table 1 that classifies existing models in terms of five dimensions: (1) visual representation, (2) incorporation of asymmetry,

Table 1
A comparative summary of network models

Models	Representation method	Asymmetry	Criteria of clique definition	Clique overlap	Levels of analysis
Sociogram	Visual	Accommodated	Not objective	Not allowed	Subgroup/single
Matrix operations	Non-visual	Not accommodated	Not objective	Not allowed	Subgroup only
Graph-theoretic	Non-visual	Accommodated	Objective	Not allowed	Subgroup only
Distance methods	Non-visual	Not accommodated	Not objective	Not allowed	Subgroup only
Burt's method	Visual	Accommodated	Not objective	Not allowed	Subgroup only
Block models	Non-visual	Accommodated	Not objective	Not allowed	Subgroup only
Proposed method	Visual	Accommodated	Objective	Partially allowed	Subgroup/single

(3) clique detection criteria, (4) incorporation of clique overlap, and (5) capability of performing other levels of analysis. It must be noted that our review is by no means exhaustive. Only subgroup-level models are included in this review since clique detection has been primary interest in network research (for a more extensive review of network models, see Knoke and Kuklinski 1982, and Burt 1980).

2.1. Visual representation

Knoke and Kuklinski (1982) argued that although simple network diagrams suffer from lack of parsimony, nonuniqueness, and interpretational complexity, well-constructed visual displays, especially based on graph theory, can convey an intuitive feel for the structure of a system. Several attempts have been made to provide more parsimonious and meaningful representations through sociograms (e.g., Northway 1949; Klovdahl 1981). However, these models typically lack uniqueness and interpretability. Burt (1980) utilized hierarchical clustering based on a metric measuring relational equivalence. Our new methodology visually represents a social structure via an asymmetric *path-length* or *additive* tree, a particular type of graph which is parsimonious and unique up to a set of known tree indeterminacies.

2.2. Incorporation of asymmetry

Most network models force the sociometric matrix to be reciprocal (symmetric) either to satisfy the particular metric requirements of the model or to make the computation easier. Since social relations between actors are typically bidirectional (actors both initiate and receive contacts), network models that capture this asymmetry should prove more accurate in reflecting respective communication processes. Graph-theoretic methods (e.g., Luce and Perry 1949; Luce 1950; Siedman and Foster 1978) and structural equivalence models (Burt 1980; White *et al.* 1976) explicitly consider asymmetry in their analyses. However, these models fail to depict the asymmetry in the derived representations. In our methodology, asymmetry in communication will be explicitly visualized by representing each individual as two different entities: one as a social contact *initiator* and the other as a social contact *recipient*.

2.3. Clique detection criteria

Most network models (except graph-theoretic methods) do not provide objective criteria for clique detection, suggesting that it is likely that substantially different results can be yielded when the same data and a similar generic model (e.g., cluster analysis methods) are used by different researchers. This problem is especially critical in network models utilizing various types of clustering methodologies (e.g., Burt 1980 utilizing hierarchical clustering, and Brieger *et al.* 1975 utilizing block clustering) where the number of clusters to be retained is decided on an ad hoc basis. We will propose a more objective foundation for clique detections embodied in our methodology.

2.4. Incorporation of clique overlap

Most network models detect cliques by forming exclusive partitions of groups and assign each individual to one and only one clique. However, an individual might easily be a member of multiple cliques. Davis (1967) asserts that disjoint cliques are seldom obtained in personal communication because a real disjoint partition is usually a sign of conflict which prohibits actors from participating in intense communication. Our methodology will partially accommodate this feature of clique overlap by allowing for an actor to be classified to one clique as a contact *initiator* and a possibly different clique as a contact *recipient*.

2.5. Levels of analysis

There are three levels of networks analysis: single actor level, subgroup level, and entire system level. Traditionally, network models have been developed to solve problems at a specific level in a specific context, disregarding issues raised at different levels and/or in other contexts. Subgroup-level network models which attempt to detect subgroup patterns (cliques) completely ignore individual level analyses which describe each individual's positional aspects in a system. Similarly, network models describing individual positions, typically, are not capable of accommodating subgroup level analyses. Network models capable of simultaneously handling multiple levels of analysis, i.e., individuals' positional indices, clique detection, and characteristics of an entire system such as social cohesion, are desirable simply because certain

aspects of a network at a certain level can not be fully understood without considering aspects at other levels. Our methodology *simultaneously* portrays clique formation *and* individual roles, and provides measures for inter- and intra-clique relationships.

In short, the primary goals of this paper are to develop a new methodology that will:

1. derive an "optimal" representation of communication processes estimated from collected sociometric data;
2. provide a parsimonious visual representation that:
   a. depicts how a group of social actors are partitioned (clique detection);
   b. portrays each individual's social positions embedded in the group structure;
3. accommodate the intrinsically asymmetric nature of social interactions by portraying an individual as two different entities;
4. suggest a mathematical criteria to decide the number of cliques retained and the respective clique memberships; and,
5. partially allow for multiple clique memberships.

3. A methodology for estimating bidirectional path-length trees from binary sociometric data

3.1. A brief review of tree-fitting methodologies

As background to the work presented, a brief review of tree-fitting methodologies developed in mathematical psychology is presented here. Further details may be found in Carroll (1976), Sattath and Tversky (1977), Furnas (1980), and De Soete *et al.* (1984). A tree is a connected graph without cycles where each pair of objects is joined by a *unique* path. The terminal (or external) nodes in a tree represent objects, and the distance between two objects is defined as either the height of the most immediate common ancestor (internal node) in an *ultrametric tree* (much better known as hierarchical clustering) or the length of the path that joins them in a *path-length or additive tree*.

A tree has several attractive properties as a representation of social networks. First, it can portray hierarchical groupings and allow for an unambiguous interpretation of clusters. This feature provides the

fundamental motivation for developing a network model based on a tree structure. Second, the distance between objects can be meaningfully interpreted (visually) from a tree because of its *metric* properties. Third, a visual representation of the relations among objects is unique (up to a set of known tree indeterminacies) once the distances among objects obey certain metric conditions. For example, in path-length trees for one-mode, two-way data (e.g., an $N \times N$ matrix of proximity data between a common set of actors), the indeterminacy of the root causes a nonuniqueness of the representation (see De Soete *et al.* 1984 for a discussion of indeterminacies of such tree representations). A discussion of these metric conditions follows.

Ultrametric distances for one-mode, two-way symmetric proximities must obey the one-class ultrametric inequality (Johnson 1967):

$$d_{ik} \leq \max(d_{ij}, d_{jk}), \forall i, j, k, \tag{3.1}$$

where d_{rs} indicates the distance between actors r and s. This ultrametric inequality implies that for any three nodes in a tree, two of the distances are equal and the third does not exceed them, forming an equilateral or isosceles triangle. It also implies that given two disjoint clusters, all intracluster distances are smaller than all intercluster distance, and that all the intercluster distances are equal. Even though hierarchical clustering has been well developed and widely used for representing various types of distances (e.g., similarity, social relations) in a tree structure (see Hartigan 1967), it has been criticized because of such a severe restriction on the data given these ultrametric conditions/constraints.

In an attempt to resolve the limitations of enforcing the ultrametric inequality restrictions (e.g., the restrictions on the inter- and intracluster distances), Sattath and Tversky (1977) and Cunningham (1978) advocate using the more general *path-length* tree representation of proximities data (a *path-length* tree is also called an *additive tree*, free tree, or unrooted tree). A path-length tree is a tree with a metric in which the distance between nodes is equal to the length of the path (i.e., sum of branches) that joins them (Sattath and Tversky, 1977). This metric is particularly attractive for modeling social relations given this intuitive interpretation. The necessary and sufficient condition for representing two-way, one-mode proximities by a path-length tree is

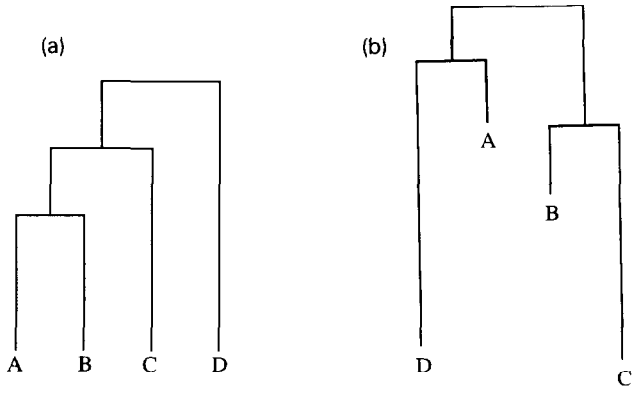

Fig. 1. (a) Ultrametric tree representation vs (b) path-length tree representation.

the four-point inequality (Dobson 1974). That is, for all actors i, j, k, l in the set of actors S,

$$d_{ij} + d_{kl} \leq \max(d_{ik} + d_{jl},\ d_{il} + d_{jk}), \tag{3.2}$$

holds for all quadruples i, j, k, and l. Note that this four-point inequality is less restrictive than the ultrametric inequality in that intracluster distances may exceed intercluster distances and that an actor outside a cluster is no longer equidistant from all objects inside the cluster. In Figure 1, the two representations of proximity data (from Sattath and Tversky, 1977) are compared for the one-mode symmetric case. Consider four actors A, B, C, and D. In the ultrametric tree (Figure 1a), A and B form a single cluster that is sequentially joined by C and D. In the path-length tree (Figure 1b), A and D form one cluster, and B and C form another cluster. Unlike the ultrametric tree, the distance between A and D (intracluster) is allowed to exceed the distance between A and B (intercluster). Moreover, in the path-length tree, the distance between A and B, and the distance between A and C (intercluster distance) are not equal.

Farris (1972) and Carroll (1976) show that it is possible to convert a path-length tree into an ultrametric tree by a simple operation. It is stated that $t_{ik} = d_{ik} - c_i - c_k$ satisfies the ultrametric inequality; d_{ik} is the path-length distance from i to k and c_i (c_k) is an additive constant associated with $i(k)$, which satisfies the positivity condition for distance. Therefore, the path-length distance, d_{ik}, is decomposable into a

t_{ik} that satisfies the ultrametric inequality plus a set of additive constants. Thus, an ultrametric tree is actually a special case (nested) of the more general path-length tree. Because of this nesting, we feel that the path-length tree is even more appealing for depicting the nature of personal relationships given its additional flexibility for portraying different communication network shapes.

The distance of an external node from the root in a path-length tree reflects its average distance from other external nodes when the root is chosen in a manner that minimizes the variance of distances to the external nodes. This heuristic uniquely determines the location of the root. The distance of each communication participant from the root can be interpreted as the degree of involvement in the communication process, either as an information provider or information receiver. Note that this property cannot be represented in an ultrametric tree in which all terminal nodes are equidistant from the root.

Similarly, the distance of an internal node from the root that defines a cluster (clique), which contains all the terminal nodes that follow from it, can be interpreted as a measure of involvement of the clique in the communication process. Thus, the representation of relational data via a path-length tree is a more intuitive way to display the structure of communication networks by simultaneously representing clique structure and, as will be shown, providing the basis for the calculation of various indices for the role of individuals and groups of individuals.

3.2. Tree representation of two-way, two-mode (asymmetric) proximity data

In network analysis, it should be recognized that the social relations between actors are bidirectional. A relation initiated by actor i and terminated by actor k may or may not be the same as a relation initiated by k and terminated by i. In the case of a rectangular proximity matrix (two-mode data where, for example, the row and column actors differ, or where the ik entry $\neq$ to the ki entry), it is not possible to impose the ultrametric inequality condition on such distances since one of the three distances will be missing for every triple. (Note that distances among within-class objects are not defined in such "unfolding" or rectangular representations.)

Following Furnas (1980), for asymmetric matrices with a distance between actors of different classes (e.g., i, j as initiators and k, l as

receivers), the following two-class ultrametric condition is necessary and sufficient for the representation of an ultrametric tree:

$$t_{il} \leq \max(t_{ik}, t_{jk}, t_{jl}) \text{ for all } i \neq j,\ k \neq l. \tag{3.3}$$

In words, for every quadruple of points composed of two from each class, the two largest of the four distances must be equal.

Necessary and sufficient conditions for a two-class, path-length tree can be indirectly derived using the two-class ultrametric condition in (3.3) above and the decomposition of path-length distances into ultrametric distances and additive constants. Such a decomposition holds for rectangular submatrices of distances for a path-length tree (De Soete et al. 1984). Letting d_{ik} be the path-length distance between actors from different classes and r_i (c_k) be i's (k's) additive constant, $t_{ik} = d_{ik} - r_i - c_k$ should satisfy the two-class ultrametric inequality condition. Therefore, the technique used here is to initially derive a two-mode ultrametric tree and then estimate row and column additive constants (equivalent to "star" trees having only one interior node–see De Soete *et al.* 1984). However, unlike De Soete *et al.* (1984), we utilize a stochastic framework with binary (not proximity) data. A technical description of the proposed methodology follows.

3.3. The two-mode path-length tree methodology

Let:

i, j = $1, \ldots, N$ contact initiators;

k, l = $1, \ldots, N$ contact recipients (in most network data, the row and column objects are the same);

m = $1, \ldots, M$ replications (e.g., communication scenarios or settings);

$\mathbf{A}$ = $((a_{ikm}))$ and $(N \times N \times M)$ adjacency matrix, where:

$$a_{ikm} = \begin{cases} 1, & \text{if actor } i \text{ initiates contact with actor } k \\ & \quad \text{in the } m\text{th replication,} \\ 0, & \text{otherwise;} \end{cases}$$

$a_{ikm} \neq a_{kim}$ in general, and

a_{iim} is undefined;

n_{ik} = the number of times actor i initiates contact with actor k in M replications;

$\mathbf{U}$ = $((u_{ikm}))$, where u_{ikm} is a latent, unobservable social distance between actor i and actor k in replication m;

$\mathbf{D}$ $= ((d_{ik}))$, where d_{ik} is a path-length distance between actor i and actor k defined on a two-mode (i.e., representing both contact initiators and recipients as distinct terminal nodes) path-length tree ($u_{ikm} = d_{ik} + e_{ikm}$);

$\mathbf{T}$ $= ((t_{ik}))$ a distance matrix obeying the two-class ultrametric inequality;

$\mathbf{r}$ $= (r_i)$ a vector of row additive constants for initiators;

$\mathbf{c}$ $= (c_k)$ a vector of column additive constants for recipients;

d_{ik} $= t_{ik} + r_i + c_k$, following De Soete *et al.* (1984);

$\mathbf{E}$ $= ((e_{ikm}))$, where e_{ikm} is an error term distribution iid $N(0, \sigma^2)$;

s = a social distance threshold, where we assume that actor i will contact actor k in replication m ($a_{ikm} = 1$) iff $u_{ikm} \leq s$, and will not iff $u_{ikm} > s$.

We assume that a latent social distance exists such that actor i will contact actor k in a specified replication if the members of the pair are sufficiently "close" to one another. This specification utilizing a threshold concept is also used in stochastic spatial choice models (cf., DeSarbo and Cho 1989). This notion of distance in some social region and/or the threshold concept is also employed in several network models (cf., Burt 1980; Hubbell 1965). Here, however, the threshold value (s) is estimated directly and not imputed in an ad hoc manner as in these previous approaches. The threshold level can also be modeled as varying by individuals (e.g., s_i) in this methodology. For now, we assume that the threshold level is common across all respondents.

We assume that the communication process where actor i makes contact with actor k in replication m is Bernoulli distributed with probability of contact of p_{ikm}. Thus;

$$
\begin{aligned}
&p(i \text{ initiates contact with } k \text{ in replication } m) \\
&= p(a_{ikm} = 1) = p(u_{ikm} \leq s) = p_{ikm} \\
&= p(t_{ik} + r_i + c_k + e_{ikm} \leq s) \\
&= p(e_{ikm} \leq s - r_i - c_k - t_{ik}) \\
&= \Phi\left(\frac{s - r_i - c_k - t_{il}}{\sigma}\right) \\
&= \Phi(s - r_i - c_k - t_{ik}), \qquad (3.4)
\end{aligned}
$$

since, without loss of generality, σ can be absorbed in the numerator of the $\Phi(\cdot)$ function (it can thus be assumed $\sigma = 1$ since it can be directly absorbed in the s, r_i, c_k, and t_{ik} terms without loss of any generality), where $\Phi(\cdot)$ is cumulative distribution of the standard normal distribution evaluated at $(\cdot)$. Similarly,

$$
\begin{aligned}
&p(i \text{ does not initiate contact with } k \text{ in replication } m) \\
&= p(a_{ikm} = 0) = p(u_{ikm} > s) = 1 - p_{ikm} \\
&= p(t_{ik} + r_i + c_k + e_{ikm} > s) \\
&= p(e_{ikm} > s - r_i - c_k - t_{ik}) \\
&= 1 - \Phi\left(\frac{s - r_i - c_k - t_{ik}}{\sigma}\right) \\
&= 1 - \Phi(s - r_i - c_k - t_{ik}). \qquad (3.5)
\end{aligned}
$$

Assuming independence over all m, i, and k indices, one can obtain the likelihood function:

$$
\begin{aligned}
L &= \prod_{m}^{M} \prod_{i \neq k}^{N} \prod^{N} \Phi(\cdot)^{a_{ikm}} (1 - \Phi(\cdot))^{(1 - a_{ikm})} \\
&= \prod_{i \neq k}^{N} \prod^{N} \Phi(\cdot)^{n_{ik}} (1 - \Phi(\cdot))^{(M - n_{ik})}, \qquad (3.6)
\end{aligned}
$$

and the corresponding log-likelihood function:

$$
\Lambda = \ln L = \sum_{i \neq k}^{N} \sum^{N} \left[n_{ik} \ln \Phi(\cdot) + (M - n_{ik}) \ln(1 - \Phi(\cdot)) \right], \qquad (3.7)
$$

where $\Phi(\cdot) = \Phi(s - r_i - c_k - t_{ik})$. We can estimate s, r_i, c_k, and t_{ik} by maximizing Λ subject to t_{ik} satisfying the two-class ultrametric inequality. Thus, the communication network for a given set of actors is represented by two different sets of terminal nodes in a two-mode, path-length tree: one set as contact *initiators* and one set as contact

Table 2
An example of a binary sociomatrix

	A	B	C	D	E	F
a	–	1	0	1	0	0
b	1	–	0	0	0	0
c	0	1	–	1	0	0
d	0	1	1	–	0	1
e	0	0	1	1	–	0
f	1	1	0	1	0	–

recipients. Note that the flexibility of setting $r_i = c_i = 0$ and estimating a two-mode ultrametric tree is also allowed for in this methodology.

To motivate how the resulting two-mode path-length tree can describe various relational aspects, let us consider a simple hypothetical illustration: a communication network among a group of physicians concerning who they seek information from concerning new pharmaceutical products. Table 2 presents a binary sociomatrix among six hypothetical physicians where $a_{ik} = 1$ indicates the existence of contact initiated by i and received by k (only one replication), and $a_{ik} = 0$ indicates no contact. Figure 2 depicts a communication network (a two-mode path-length tree) estimated from this binary data. Since each individual typically performs both information giving and receiving roles, he/she is positioned in two terminal nodes; A–F labels information givers and a–f labels information receivers. From a managerial perspective, an appropriately depicted communication network among a group of customers provides insight into how to optimize the efficiency of marketing efforts directed to that group. Marketing managers are concerned with locating the key informant(s) in a group whose opinion substantially impacts others' decisions. The illustrative two-mode path length tree shows that physicians D and B are more influential individuals (detailed interpretation of an estimated tree for actual communication data is described later in the paper) given their central positions in the derived tree. Marketing managers are also concerned with detecting subgroups (cliques). Customers who maintain cohesive bonds within their groups are more likely to behave similarly because they share information and develop similar preferences. Because of the hierarchical groupings embodied in the tree structure, such subgroups are easily detectable. Two major cliques are

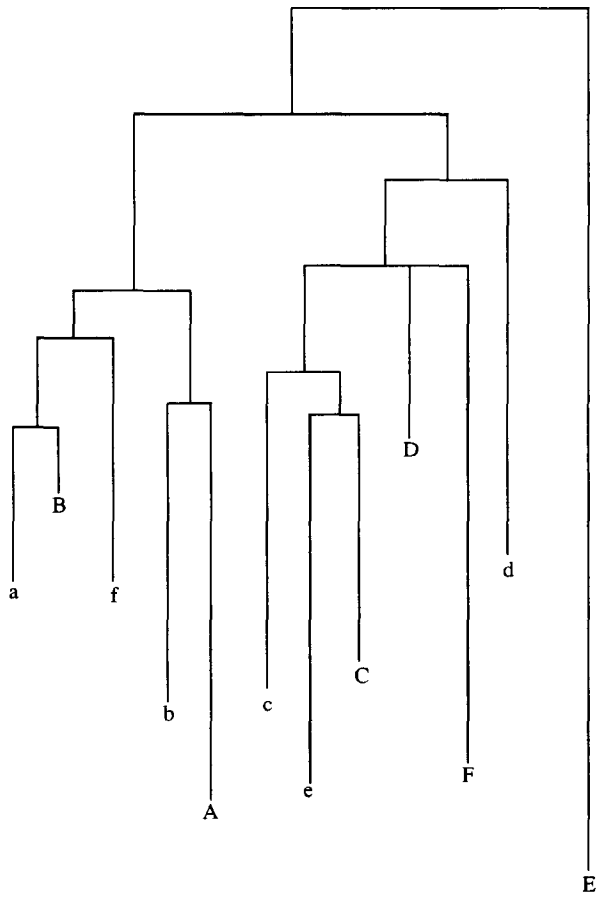

Fig. 2. A hypothetical example of a communication network represented by a two-mode path-length tree.

defined in Figure 2: one consisting of individuals (A, a), (B, b), and f, and the other of (C, c), (D, d), F, and e. Note that individual (F, f) is a member of both subgroups, acting in different roles in each subgroup. Such insight into information flows enables marketing managers to direct efforts toward opinion leaders in the introductory stage of new products. Later, "early adopters" can be targeted within the same cliques. Thus, an efficient, sequential marketing program can be guided by the framework of opinion leadership, clique membership, and other relational measurements (in addition to individual attributes).

3.4. The algorithm

Given a binary, sociometric array **A** we wish to construct a two-mode path-length tree whose structure reflects the communication patterns in **A**. We estimate a matrix **D** satisfying the two-class additive inequality via a maximum likelihood method employed to optimize the log-likelihood function with corresponding metric constraints by using an exterior penalty function approach (Rao 1979).

Penalty methods are procedures for approximating constrained optimization problems by unconstrained problems. The approximation is accomplished by adding to the objective function a term that prescribes a high cost for violating the specified set of constraints. Associated with this method is a penalty parameter (ρ) that determines the severity of the penalty and, consequently, the degree to which the constrained problem approximates the original constrained problem. As the enforcement of the constraints is made more exact by iteratively increasing the penalty parameter, the solution to the unconstrained penalty problem approaches the solution to the original constrained problem. The specific steps of the penalty function algorithm are summarized in the Appendix.

Once the final estimates of t_{ik} $(i \neq k)$, r_i, and c_k are obtained, the two-mode, path-length tree can be constructed. First, a symmetric grand matrix $\mathbf{G}(2N \times 2N)$ satisfying the one-class ultrametric inequality is constructed from the final estimates of **T**. Following Furnas' (1980) procedure, the **T** matrix fills the two $N \times N$ submatrix within **G** for the last (first) N rows and the first (last) N columns. The remaining two submatrices consisting of the first N rows and columns and the last N rows and columns can be estimated as follows:

$$g_{ab} = \begin{cases} t_{(a-N)b} & \text{if } N+1 \leq a \leq 2N \\ & \quad \text{and } 1 \leq b \leq N \\ \min\left[\max(t_{ia},\, t_{ib})\right] & \text{if } 1 \leq a \leq N \\ i = 1, \ldots, N & \quad \text{and } 1 \leq b \leq N \\ \min\left[\max\left(t_{(a-N)k},\, t_{(b-N)k}\right)\right] & \text{if } N+1 \leq a \leq 2N \\ k = 1, \ldots, N & \quad \text{and } N+1 \leq b \leq 2N\big]. \end{cases}$$

Then, **G** is submitted to any standard hierarchical clustering method (e.g., Johnson 1967) which, by definition above, renders a perfect fit to

G. This ultrametric tree is converted to a path-length tree by defining the length of a branch to be the difference in height values of the two nodes (either terminal or internal) connected by that branch (Dobson 1974). Finally, each row and column additive constant is added to the length of the associated branch that connects the terminal nodes to the first internal node. The resulting two-mode path-length tree thus contains two sets of terminal nodes: one as contact initiators and the other as contact recipients.

3.5. Other discrete nonspatial models

As alternative ways of graph-theoretical representation of *proximity* data, more general graph-fitting methodologies have been proposed recently (Hutchinson 1989; Klauer and Carroll 1989) for metric proximity data. While tree structures accomplish parsimonious representation of the data by imposing certain metric restrictions such as the ultrametric condition or path-length condition, general graph methodologies achieve parsimony by eliminating redundant links between nodes without imposing strict metric restrictions on the distances. In general, the distance from one node to another is a function of the lengths of the paths connecting the two nodes. In a complete graph, where all nodes are reciprocally connected, each distance corresponds to the length of the link. Of particular interest is the minimum path-length metric, since this distance function maximizes parsimony by deleting all redundant links without substantially losing goodness-of-fit.

The NETSCAL (Hutchinson 1989) methodology is a two-step procedure that determines which pairs of nodes are directly connected by an arc, and then estimates a pair of lengths for each arc. The link structure is chosen according to the following heuristical principle:

$$\text{If } d_{xy} \leq \min\left\{\max\left\{d_{xz},\ d_{zy}\right\}:\ z \neq x,\ y\right\},$$

then, $(x,\ y)$ is an arc.

In words, a minimum distance between x and y cannot result from an indirect path-length through a third node if, for all third nodes, one of the component distances (d_{xz} or d_{zy}) is greater than d_{xy} (Hutchinson

1989). This provides a sufficient, but not necessary, condition for the presence of an arc; thus, graphs that are not connected may arise (Klauer and Carroll 1989). While a Monte Carlo simulation and applications to various data sets demonstrate the practical utility of the algorithm (Hutchinson 1989), serious problems with locally optimal solutions have been uncovered recently by Klauer and Carroll (1990).

More recently, Klauer and Carroll (1989) proposed a mathematical programming approach to fitting general network graphs to interval scale proximity data. This method removes L links (L is user-specified and fixed) if there are L different triples that satisfy the triangle inequality as an equality. (In NETSCAL, the link structure and the link weights are determined in separate steps.) In the Klauer and Carroll approach, the goodness-of-fit is maximized while the number of links is fixed. Tree structures can be regarded as a special case in these general graph structures where the number of links is equal to the number of objects minus one. Unfortunately, the user is assumed to know L, the total number of connecting links in the network, or cycle through several analyses varying L. Even with the cycling, there are no robust statistical tests known for selecting the appropriate L.

In sum, methodologies utilizing a minimum path-length metric are more general than tree structures in representing various kinds of *proximity* data. However, as discussed above, several potential problems exist with the usage of such models. In addition, neither NETSCAL nor the Klauer and Carroll methods operate on binary relational data. Sociometric communication networks can be more usefully analyzed via tree structures since one of the major concerns of investigating such structures is to define hierarchically organized patterns.

4. A Monte Carlo simulation study

4.1. Overview of the procedure

To examine the performance of the proposed methodology, a Monte Carlo analysis was performed where five independent factors relating to data, error, and algorithm parameters were experimentally manipulated to create synthetic data for testing. These independent factors,

Table 3
Independent factors hypothesized to affect methodology performance

Factor	Level	Code
A. No. of row/column elements (N)	$N = 6$	1
	$N = 8$	2
	$N = 10$	3
B. Number of replications (M)	$M = 1$	1
	$M = 3$	2
	$M = 5$	3
C. Error (σ^2)	$\sigma^2 = 0$	1
	$\sigma^2 = 10$	2
	$\sigma^2 = 20$	3
D. Starting values	Random	1
	0	2
	Data *	3
E. Penalty increase (ρ)	5	1
	10	2
	100	3

* We set the starting value of t_{ik} to $(1 - n_{ik}/M)(t_{\max} - t_{\min})$, where $t_{\max}$ ($t_{\min}$) is upper (lower) limit value of the estimates. Thus, if i and k are tied in every relationship ($n_{ik} = M$), t_{ik} is set to 0, and if i and k are not tied at all in any relationship ($n_{ik} = 0$), t_{ik} is set to 100 (we set 100 for $t_{\max}$ and 0 for $t_{\min}$).

hypothesized to affect the estimation of path-length structures, were:

(i) the number of actors ($N = 6$, 8, and 10),
(ii) the number of replications ($M = 1$, 3, and 5),
(iii) the amount of error ($N(0, \sigma^2)$ added to the data (no error, $\sigma^2 = 10$, and $\sigma^2 = 20$),
(iv) starting values (random, 0, and the data), and
(v) the rate of penalty parameter increase (by a factor of 5, 10, and 100).

Table 3 provides a description of these five independent factors.

Two overall areas of methodological performance were measured: the overall goodness-of-fit and the amount of computational effort required. Overall goodness-of-fit was operationalized in terms of:

(i) the normalized (for number of observations) log-likelihood value, and
(ii) the simple matching coefficient between the actual input binary data ($\mathbf{A}$) and the model predicted values ($\hat{\mathbf{A}}$).

Computer usage for each run was operationalized in terms of:

(i) CPU time (in seconds on IBM Mainframe model 3090), and
(ii) the number of major iterations required for convergence.

Note, the first two independent factors determine the size of the problem and the degrees of freedom in the estimation, and are expected to increase the amount of computational effort as they increase. Also, better goodness-of-fit is expected as the number of replications is increased because of the gain in degrees of freedom. Increasing error variance may negatively affect both the goodness-of-fit and computational effort. The fourth and the fifth factors gauge the sensitivity of the estimation procedure to various user options. A rapid increase in the penalty parameter in the fifth factor (e.g., increase by a factor of 100) may reduce computational effort, but is expected to negatively affect goodness-of-fit, since the likelihood of speeding to a locally optimal solution is greater. A summary of the anticipated effects of independent factors on methodological performance is provided in Table 4.

These five factors were combined via an asymmetric fractional factorial 3^5 design (Addleman 1962) for main-effects-only estimation (as in DeSarbo and Carroll 1985). Sixteen experimental trials were designed and are listed in Table 5. Note, this modest Monte Carlo analysis is not presented as a definitive test of the methodology, but only as a preliminary indication of the performance of the procedure. Clearly, a full factorial design with replications and perhaps additional factors would have been preferable if computer expense were not a limiting aspect. We leave this for future research.

Table 4
Summary of anticipated effects of independent factors on dependent measures

Factor	Goodness-of-fit	Computational effort
A. Increase in N	No effect	Negative
B. Increase in M	Positive	Negative
C. Increase in σ^2	Negative	Negative
D. Starting values (Data option)	Positive	Positive
E. Rapid increase in ρ	Negative	Positive

Table 5
3^5 Fractional factorial experimental design

Trial:	Factors:				
	A	B	C	D	E
1	1	1	1	1	1
2	3	2	2	1	2
3	2	3	2	1	3
4	2	2	3	1	2
5	2	2	2	2	1
6	2	1	3	2	2
7	3	2	1	2	3
8	1	3	2	2	2
9	3	3	3	3	1
10	1	2	2	3	2
11	2	1	2	3	3
12	2	2	1	3	2
13	2	2	2	2	1
14	2	3	1	2	2
15	1	2	3	2	3
16	3	1	2	2	2

For each experimental trial, **D** were generated from exact two-mode, path-length trees that were randomly constructed for each trial. A value of s was then randomly generated. Error was then generated from a Normal $(0, \sigma^2)$ distribution and added to **D** to obtain **D*** (error-perturbed distances). Finally, **A** (error-contained binary data) was created from **D*** using the threshold rule, and was utilized as the input data for each trial.

4.2. Results and analysis

The average matching coefficient measuring goodness-of-fit was exceptionally high (0.977). The average CPU time used was 90.8 seconds requiring an average of 14.5 major iterations. The four dependent measures were analyzed via multiple regression (as in conjoint analysis), where the experimental design was converted to dummy variables. Results for the dependent measures are shown in Table 6. The coefficient displayed next to each factor level represents the regression coefficient for that level. The intercept term represents the combined effect of level 1 of all factors. Note that a logit transformation

Table 6
Multiple regression results of the Monte Carlo simulation

Factor/Level	Dependent variable			
	Y_1	Y_2	Y_3	Y_4
A. $N = 8$	−6.31	−1.50	0.87	46.54
A. $N = 10$	−68.52	−0.67	3.50	126.00 **
B. $M = 3$	−26.70	0.05	−3.30	−1.83
B. $M = 5$	−29.94	−2.23	−0.25	−21.07
C. $\sigma^2 = 10$	−30.54	−7.35 **	−2.85	−1.91
C. $\sigma^2 = 20$	−47.57	−8.52 **	−3.30	−24.72
D. Start = 0	55.99	−0.76	−5.45	−37.83
D. Start = data	27.43	0.16	−5.05	−16.70
E. $\rho \times 10$	0.62	−1.15	−4.60	18.24
E. $\rho \times 10$	24.04	0.15	−6.50	53.41
Intercept	−0.03	13.80	25.03	27.54
S.E.	42.70	2.79	5.32	61.14
R^2	0.76	0.85	0.65	0.86
Adj R^2	0.32	0.54	0.25	0.60
F	1.64	2.72	0.94	3.26

** $P \leq 0.01$.
Y_1: Log-likelihood value. Y_2: Matching coefficient.
Y_3: Number of major iterations. Y_4: CPU time in seconds.

[log $Y/(1 - Y)$] was applied to the matching coefficients (a transformation for normality assumptions to be more tractable) since they are, by definition, restricted between the values of 0 and 1 (Pendyck and Dubinfield 1981).

Concerning the normalized log-likelihood dependent variable, Y_1, no independent factor level is significant, indicating consistent fitting over all independent factor levels. For Y_2, the two higher error levels appear to significantly detract from the matching coefficient, although the entire regression equation is not significant. While there are no significant independent factor levels affecting Y_3, the number of major iterations required for convergence, larger numbers of replications do significantly affect CPU time. Thus, larger data sets appear to require somewhat more extensive computational effort as might be expected. Also, as one adds more error to the data, goodness-of-fit appears to suffer somewhat. The results of this modest simulation analysis demonstrate the somewhat robust performance of the methodology. Again, these results should be considered preliminary given the small scope of this analysis.

5. Application

5.1. Data description

Krackhardt (1987) collected data from a small high-tech manufacturing organization on the (U.S.) west coast. In his research, all the network members indicated "perceived" relationships among all dyads. Twenty-one management-level employees (supervisors through president) were asked: "Who would person X go to for help or advice at work?" Below the question, twenty managers were listed, resulting in three-way ($21 \times 21 \times 21$), binary, sociometric data. The elements of the matrix can be represented as a_{ikm} where i is the "sender" of the relationship, k is the the "receiver" of the relationship, and m is the "perceiver" of the relationship between i and k. Thus, $a_{3,12,8} = 1$ would be interpreted to mean that person 8 thinks person 3 approaches 12 for help and advice. For the purpose of demonstrating the methodology here, six employees were excluded since they were perceived to have very few contacts with others.

5.2. The stochastic path-length tree analysis

5.2.1. Clique detection

The greatest challenge in network analysis is how to formally define "cliques" (see Lankford 1974). Even though several formal definitions are proposed in graph-theoretic approaches, such as the 'maximal complete subgroup" (Luce and Perry 1949), the "maximal strong component" (Harary *et al.* 1965), the "n-clique" (Luce 1950), and the "k-plex" (Siedman and Foster 1978), there is no formal clique definition proposed in *distance approaches*. This problem is present in most clustering methodologies where the number of clusters to be retained is typically selected in an ad hoc manner. The most generally accepted clustering rule in traditional hierarchical clustering, for example, is to choose some arbitrary distance value on the derived tree, and define clusters or cliques in terms of groups of actors who join below this value. At some point, an appropriate value of this criteria is selected that most strikingly shows the pattern of clustering (see Burt 1980 for such a procedure). This ad hoc criterion, however, does not always render the "optimal" clique detection, especially when a strikingly clear pattern of clustering is not present. Here, we propose a new method to

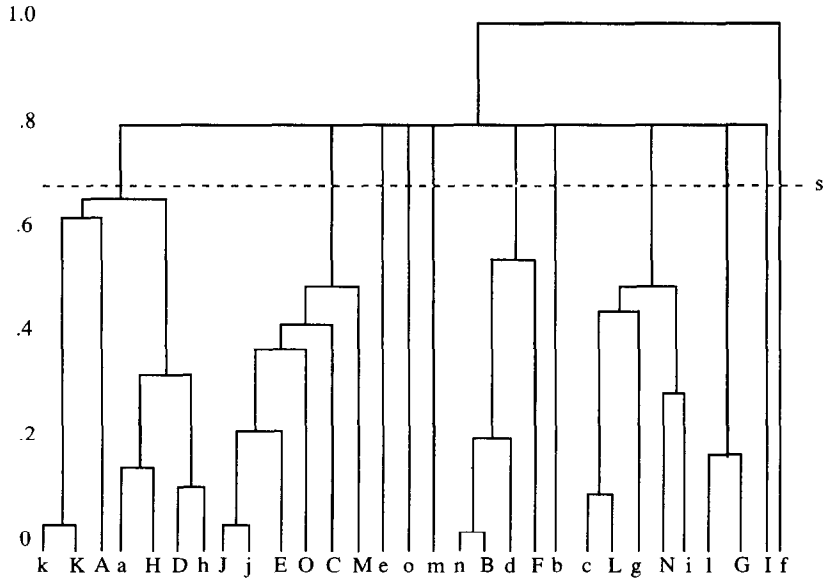

Fig. 3. The estimated two-mode ultrametric tree for the network data. Estimated threshold level(s) = 0.683.

estimate the threshold value (s) embodied in the estimation of the tree structure.

In our estimated two-mode ultrametric trees, the location of s is defined since the distance from all the terminal nodes are equidistant from the root. However, the location of s is not uniquely defined in additive trees because it varies for each dyad or pair of actors. Thus, the approach we propose here is to define cliques utilizing the threshold value estimated in the two-mode ultrametric tree (holding $r_i = c_k = 0$). The resulting two-mode ultrametric tree is shown in Figure 3. The estimated threshold value (s) in this two-mode ultrametric tree(s) equals 0.683 and determines five clusters/cliques and six isolates. We use the letters A–O to represent the individuals as information givers (or contact recipients) and letters a–o to represent information receivers (or contact initiators). Reading from left to right in Figure 3, clique 1 consists of actors (K, k), (A, a), (H, h) and D; clique 2 of actors (J, j), E, O, and M; clique 3 of actors B, F, d, and n; clique 4 of

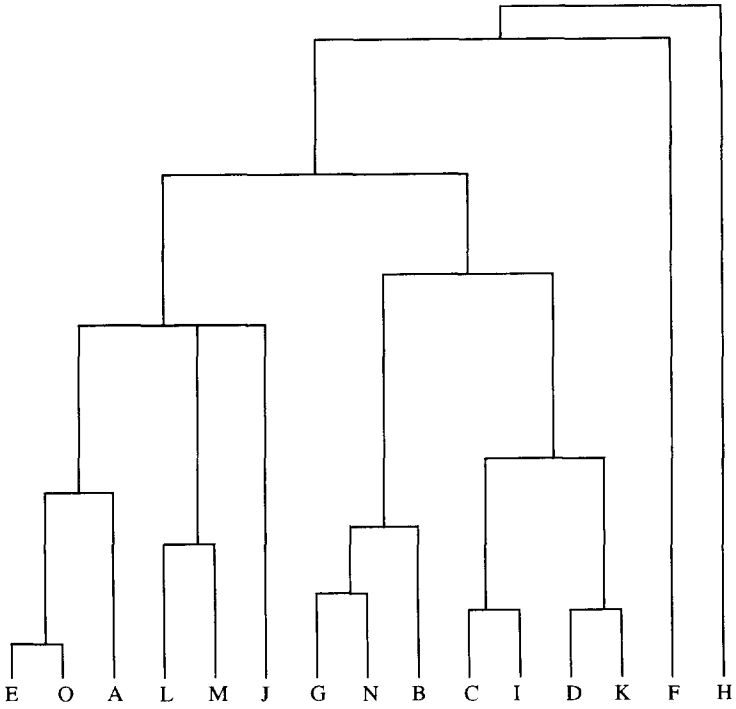

Fig. 4. Hierarchical clustering of the network data utilizing Burt's (1977) model.

actors g, L, i, c, and N; and, clique 5 of actors G and 1. Actors e, o, m, b, I, and f are depicted as isolates.

Note that Figure 3 also depicts overlapping clique structure. Individual (D, d) is a member of both cliques 1 and 3, acting as an information giver in clique 1 and as an information receiver in clique 3. Similarly, individual (G, g) participates in clique 5 as an information giver and in clique 4 as an information receiver. Individual (N, n) acts as an information giver in clique 4 and an information receiver in clique 3.

Figures 4 and 5 show the comparative representations of the aggregated or pooled data using a derived dissimilarity measure when it is submitted to hierarchical clustering (Burt 1977), and to block clustering (Brieger *et al.* 1975), respectively. One can easily see that there are substantial difference in these three results (Figures 3, 4, and 5). This is due to the differences in the data requirements and assumptions of these methodologies (recall that Burt's procedure and block clustering

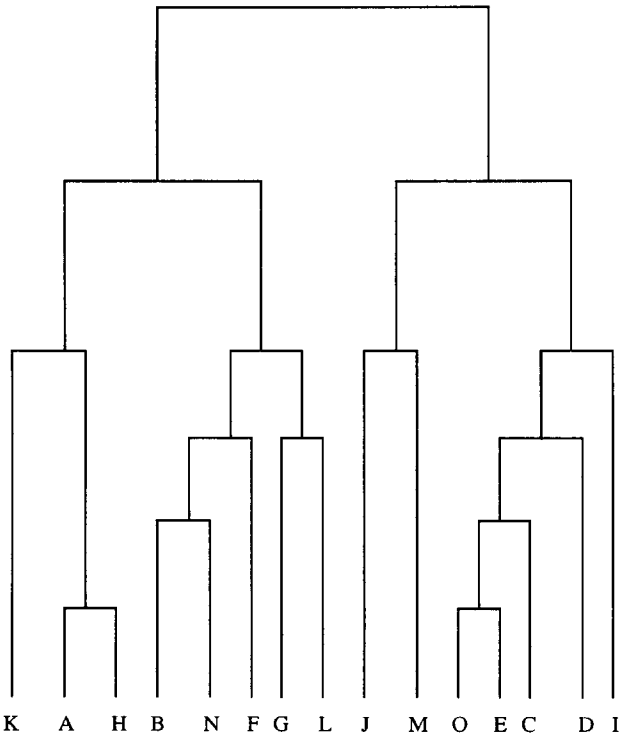

Fig. 5. Block clustering results for the network data using the Brieger *et al.* (1975) method.

are positional approaches, while the proposed methodology takes a relational approach). In addition, the later two approaches do not depict the asymmetry in **A**.

5.2.2. Construction of sociometric indices

After detecting cliques using the two-mode ultrametric tree, we proceed to estimate the additive constants (r_i's and c_k's). A two-mode, path-length tree portraying the communication network among these fifteen managers is shown in Figure 6. The log-likelihood value is -38.2 and the simple matching coefficient is 0.913. Again, the capital letters A–O represent individuals as information givers (or contact recipients) and the small letters a–o represent information receivers (or contact initiators), thus portraying the asymmetric (two-mode) nature of communication. The distance between two actors (d_{ik}) is the sum of the lengths of vertical bars that link them.

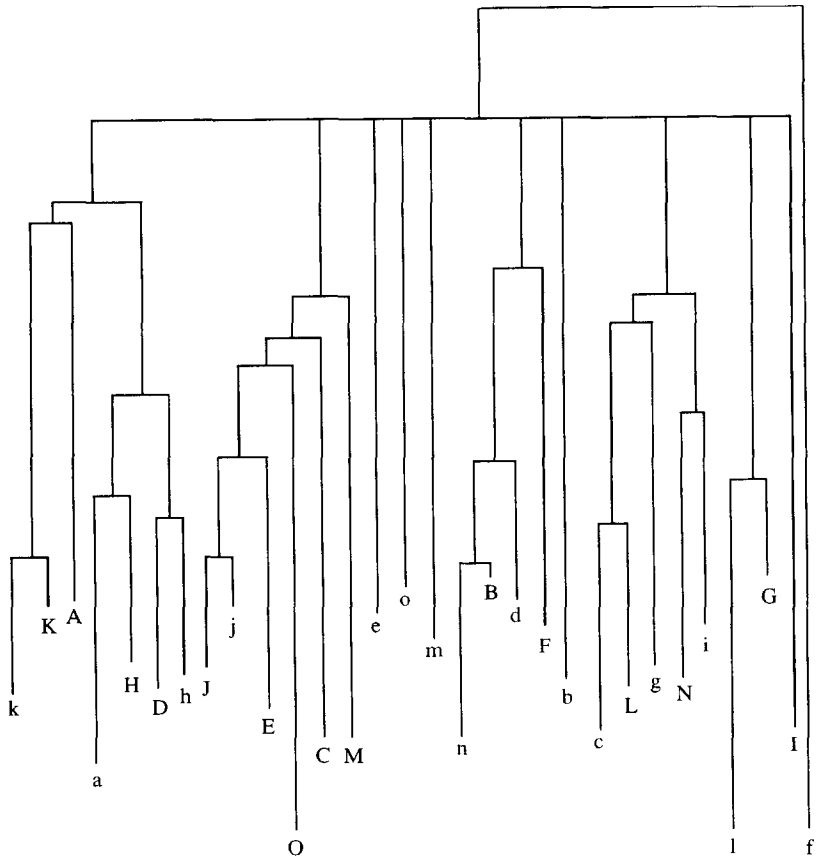

Fig. 6. The estimated two-mode path-length tree for the network data.

A variety of indices can be devised to summarize various characteristics of individual actors, subgroups, and the entire network. In addition to clique detection, one can derive a variety of sociometric measures from this methodology which describe a number of different aspects of the communication process. In the next sub-section, existing methods for constructing sociometric indices are briefly reviewed and new measures for these indices derived from the resulting path-length tree are discussed.

5.2.2.1. Indices for groups of individuals

(i) Clique cohesiveness Much research related to the construction of group indices are concerned with the cohesiveness of a group. Proctor

and Loomis (1951) use the number of mutual choices in a binary sociometric matrix of direct contacts divided by the maximal possible number of such choices as an index of group cohesiveness. The following measure for clique cohesiveness based on the estimated path length tree is proposed:

$$H_c = \frac{\sum_{i \in c} \sum_{k \in c} d_{ik}}{N_{c1} \times N_{c2}}, \tag{5.1}$$

where:

i : indexes an individual as a contact recipient in clique c,
k : indexes an individual as a contact initiator in clique c,
N_{c1} = the number of i's in clique c,
N_{c2} = the number of k's in clique c, and
d_{ij} = the path-length distance between actors i and k.

This measure can be regarded as the average social distance among members within a clique. Subgroup cohesiveness measures computed for the cliques defined via (5.1) are shown in Table 7. In this application, it appears that cliques 2, 3, and 5 are the most cohesive groups, whereas clique 4 is the least cohesive.

(ii) Interclique relations Asymmetric interclique relationships can be analyzed in terms of the *strength* and *direction* of these relationship. A measure indicating the strength of the relationship between cliques is posited as an average distance between all possible dyads, one as a

Table 7
Clique cohesive measures

Clique	Proposed measure *
1	0.21
2	0.16
3	0.14
4	0.37
5	0.13

* The range of this measure has been standardized to sum to 1.00.

Table 8
Average path-length between cliques

Recipients	Initiators					AVG
	1	2	3	4	5	
1	–	93.0	111.5	123.3	141.0	117.2
2	107.0	–	118.5	113.3	131.0	117.5
3	112.0	88.0	–	88.3	106.0	98.6
4	128.7	105.0	93.5	–	101.0	107.1
5	127.0	103.0	91.5	81.3	–	100.7
AVG	118.7	97.2	103.8	101.6	119.8	

contact recipient from clique a and the other as a contact initiator from clique b (IC_{ab}):

$$IC_{ab} = \frac{\sum_{i \in a} \sum_{l \in b} d_{il}}{N_a N_b}, \quad a \neq b, \tag{5.2}$$

where:
i : indexes actors as contact initiators in clique a,
l : indexes actors as contact recipients in clique b,
N_a, N_b =number of i's in clique a, number of l's in clique b.

Table 8 lists the IC_{ab}s of all possible combinations of the five cliques. The table indicates the communication flows (directions of interclique relationship) of clique 2 → clique 1, clique 2 → clique 3, clique 3 → clique 4, clique 3 → clique 5, clique 4 → clique 3, and clique 4 → clique 5. The column/row average of Table 8 can be computed to represent each clique's opinion leadership/followership. It appears that clique 3 plays an opinion-leading role as a group, while clique 2 is most active in acquiring information.

5.2.2.2. Indices for individuals

Individual indices are not only functions of a single individual, but also refer implicitly to some set of other persons with whom the individual is related. An example is the number of contacts an individual receives from a group as a whole or from cliques to which the individual belongs. One individual index is his/her row or column total in the sociomatrix, i.e., the number of choices he/she gives and receives. An

actor is "isolated" on the periphery of a system if he/she has no relations with others in the system (see also Bavelas 1950; Freeman 1979; Lin 1976; Niemien 1974 for measures of individual prestige, centrality, and liason). As discussed before, most analysis methods for determining individual actors' social position are performed without considering subgroup level and/or system level aspects. We propose some important measures for individual roles which we can derive directly from the tree in Figure 6.

(i) Opinion leadership / followership In Figure 6, the path length from the terminal node (representing an actor as a contact recipient) to the root of the tree can be considered as an indirect measure of the degree to which an actor is involved in a communication activity as an information source, i.e., opinion leadership. This is because the closer an actor is to other individuals as an information giver, the higher the corresponding node is located. Although the root in the path-length tree is, in general, arbitrarily determined, the procedure mentioned above (construct ultrametric tree first using t_{ik}, then add constants r_i and c_k to each appropriate terminal node) uniquely defines the root. For instance, actor A's and B's opinion leadership measures are 0.58 and 0.50, respectively, implying that B is contacted by more actors than A. Similarly, the height from the terminal nodes of a–o to the root shows how actively an actor seeks advice from peers (opinion followership). Note that these indices are *visually* embedded in the social structure, measured relative to all participants in a system. Opinion leadership/followership measures computed by path-length distances, along with average column/row totals of 1's and the centrality measure proposed by Bavelas (1950), are listed in Table 9. Individual B apparently receives the most requests for advice, while individual o apparently initiates contacts with nearly everybody. Other participants appear to maintain a moderate amount of contacts with some variations. It must be mentioned that opinion followership is not necessarily negatively correlated with opinion leadership; for instance, an absolute isolate neither provides nor seeks information.

Pearson correlation coefficients between the proposed measure and the column/row totals of 1's were computed to test the validity of this measure in comparison with the traditional individual index. High correlations (-0.86 of opinion leadership measure and -0.83 of opinion followership measure) support the construct validity of the newly proposed measure.

Table 9
Opinion leadership/followership measures

Individual	Leadership		Followership		Bavelas' centrality
	X_1	X_2	Y_1	Y_2	
A	0.58	8	0.88	3	0.13
B	0.50	12	0.73	2	0.16
C	0.82	2	0.58	11	0.15
D	0.73	5	0.64	7	0.16
E	0.77	3	0.61	9	0.14
F	0.63	8	1.00	0	0.09
G	0.54	9	0.70	5	0.16
H	0.68	7	0.71	6	0.15
I	0.81	3	0.62	9	0.14
J	0.70	6	0.54	9	0.17
K	0.57	8	0.75	3	0.13
L	0.74	5	1.00	1	0.07
M	0.82	3	0.72	5	0.09
N	0.72	5	0.83	2	0.09
O	1.00	1	0.50	14	0.17

X_1, Y_1: Proposed opinion leadership/followership.
X_2, Y_2: Average column/row total of 1's.
Pearson correlation between X_1 and $X_2 = -0.86$.
Pearson correlation between Y_1 and $Y_2 = -0.85$.

(ii) Liaison identification Figure 6 also illustrates *visual* detection of liaisons. In our methodology, an actor belonging to two different cliques is defined as a liaison. As discussed above, actors (G, g), (D, d), and (N, n) perform a liaison role. It should be emphasized again that the procedure not only identifies liaisons, but also shows the direction of information flows across cliques.

6. Conclusion

A stochastic, path-length methodology for analyzing two-mode, binary, sociometric data is presented and successfully tested. A penalty-function-based methodology is developed to estimate a tree structure and the respective parameters. Its robustness to various data and algorithm factors are investigated using factorially designed synthetic data. An application study of the methodology to actual sociometric data is

presented. The interpretation of the social structure is discussed and various sociometric measures are derived.

While the results of the methodology are promising, several areas of future research can be identified. Concerning methodological issues, the preliminary simulation results demonstrate that the algorithm performs reasonably well. However, the behavior of the algorithm needs to be investigated further where a number of additional independent factors (e.g., misspecification of the distribution of e_{ikm}) are also experimentally varied. As previously mentioned, more ambitious Monte Carlo analyses need to be conducted using more complex experimental designs (e.g., full factorial designs with replications per cell). In addition, the procedure must be tested with respect to violations in the independence assumptions inherent in the construction of the likelihood function, although such tests with similarly constructed spatial multidimensional scaling models (see DeSarbo and Cho 1989; DeSarbo and Hoffman 1986) demonstrated robustness to several such violations to these independence assumptions.

A number of promising research opportunities are suggested to extend this methodology. This methodology can be easily modified to accommodate various types of *metric* network data that specify the strength or number of contacts among dyads. Multiple path-length trees can also be estimated for a particular data set. As discussed in Carroll and Pruzansky (1980), when there are multiple hierarchies in a data set, two or more trees can be fitted to represent these separate hierarchies (this idea can be thought of as a multidimensional generalization of the single-tree structure). Finally, comparisons with three-way multidimensional scaling representations derived from such binary data (e.g., Jedidi and DeSarbo 1991) for selected applications would prove of value.

On the substantive side, several research directions using this path-length methodology can be speculated. The proposed methodology has wide applicability because of the rich information provided concerning sociometric relations. Whenever an explicit understanding is needed of the personal or interorganizational interactions at a dyadic level, a subgroup level, or an entire group level, the proposed method of network analysis can be gainfully employed.

Appendix

A penalty function algorithm for estimation

Phase I: Starting value generation

Starting estimates of t_{ik} can be set using one of the following two methods:

(i) randomly generate t_{ik} from a Uniform distribution, or

(ii) use the input data, i.e., set the starting value of t_{ik} to $(1 - n_{ik}/M)(t_{\max} - t_{\min})$, where $t_{\max}(t_{\min})$ is the upper (lower) limit value of the estimates. Thus, if i and k are tied in every relationship ($n_{ij} = M$), t_{ik} is set to 0; if i and k are not tied in any relationship ($n_{ik} = 0$), t_{ik} is set to 100 (we set 100 for $t_{\max}$ and 0 for $t_{\min}$).

Initialize the major iteration index (MI) = 0 and set $\rho = 0$.

Phase II: Estimate $\mathbf{T}$, $\mathbf{r}$, $\mathbf{c}$, *and* s

The estimation problem can be stated as:

$$\text{Maximize } \Lambda = \ln L = \sum_{i \neq k}^{N} \sum^{N} \left[n_{ik} \ln \Phi(\cdot) + (M - n_{ik}) \ln(1 - \Phi(\cdot)) \right], \tag{A.1}$$

subject to the condition that $\mathbf{T}$ satisfies the two-class ultrametric inequality. Using an exterior penalty function approach, this constrained optimization problem is solved by sequentially maximizing the unconstrained function:

$$Z(\Omega, \rho) = \Lambda(\Omega) - \rho P(\mathbf{T}) \text{ with } \rho > 0, \tag{A.2}$$

for an increasing sequence of ρ values, where Ω denotes a stacked vector of all the parameters to be estimated (i.e., t_{ik}, r_i, c_k, and s). The first component in (A.2) is given in (A.1), and the second component, $P(\mathbf{T})$, is a penalty function that expresses how strongly $\mathbf{T}$ deviates from

the two-class ultrametric inequality condition. The penalty function is defined as (De Soete *et al.* 1984):

$$P(\mathbf{T}) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} \sum_{k=2}^{N} \sum_{l=1}^{k-1} (u_{ijkl} - v_{ijkl})^2, \tag{A.3}$$

where $u_{ijkl} = \max(t_{il}, t_{ik}, t_{jk}, t_{jl})$; and (A.4)

$$v_{ijkl} = \begin{cases} \max(t_{il}, t_{jk}, t_{jl}) & \text{if } u_{ijkl} = t_{ik}, \\ \max(t_{ik}, t_{jk}, t_{jl}) & \text{if } u_{ijkl} = t_{il}, \\ \max(t_{il}, t_{ik}, t_{jl}) & \text{if } u_{ijkl} = t_{jk}, \\ \max(t_{il}, t_{ik}, t_{jk}) & \text{if } u_{ijkl} = t_{jl}. \end{cases} \tag{A.5}$$

With this framework, we now estimate **T**, **r**, **c**, and s. Estimates of these parameters are sought to maximize the augmented log-likelihood function in (A.2) using a quasi-Newton gradient search method, where the partial derivatives of $Z(\Omega, \rho)$ with respect to t_{ik}, r_i, c_k, and s are:

$$\frac{\partial Z(\Omega, \rho)}{\partial t_{ik}} = \frac{\partial \Lambda(\Omega)}{\partial t_{ik}} - \rho \frac{\partial Z(\mathbf{T})}{\partial t_{ik}}, \tag{A.6}$$

where:

$$\frac{\partial \Lambda(\Omega)}{\partial t_{ik}} = \phi(\cdot)\left(\frac{(M - n_{ik})}{(1 - \Phi(\cdot))} - \frac{n_{ik}}{\Phi(\cdot)} \right), \tag{A.7}$$

$$\frac{\partial P(\mathbf{T})}{\partial t_{ik}} = 2 \sum_{i=1}^{N} \sum_{j=1}^{i-1} \sum_{k=1}^{N} \sum_{l=1}^{k-1} (u_{ijkl} - v_{ijkl})(e_{ijkl}^{ab} - f_{ijkl}^{ab}), \tag{A.8}$$

$$e_{ijkl}^{ab} = \begin{cases} 1 & \text{if } u_{ijkl} = t_{ab} \text{ and } a = i \text{ or } j, \text{ while } b = k \text{ or } l, \\ 0 & \text{otherwise,} \end{cases} \tag{A.9}$$

$$f_{ijkl}^{ab} = \begin{cases} 1 & \text{if } v_{ijkl} = t_{ab} \text{ and } a = i \text{ or } j, \text{ while } b = k \text{ or } l, \\ 0 & \text{otherwise,} \end{cases} \tag{A.10}$$

$$\frac{\partial Z(\Omega, \rho)}{\partial r_i} = \sum_{k=1}^{N} \phi(\cdot)\left(\frac{(M - n_{ik})}{(1 - \Phi(\cdot))} - \frac{n_{ik}}{\Phi(\cdot)} \right), \tag{A.11}$$

$$\frac{\partial Z(\Omega, \rho)}{\partial c_k} = \sum_{i=1}^{N} \phi(\cdot)\left(\frac{(M - n_{ik})}{(1 - \Phi(\cdot))} - \frac{n_{ik}}{\Phi(\cdot)} \right), \tag{A.12}$$

$$\frac{\partial Z(\Omega, \rho)}{\partial s} = \sum_{i \neq k}^{N} \sum^{N} \phi(\cdot)\left(\frac{(M - n_{ik})}{(1 - \Phi(\cdot))} - \frac{n_{ik}}{\Phi(\cdot)} \right), \tag{A.13}$$

where $\phi(\cdot)$ is the standard normal density and $\Phi(\cdot)$ is the associated standard normal cdf, each evaluated at $(\cdot)$. Iterations of the quasi-Newton gradient search procedures occur until no subsequent improvement in the objective function is realized.

Phase III: Test for convergence

If $[\sum_{i \neq k}^{N} \sum^{N} (t_{ik}^{(MI)} - t_{ik}^{(MI-1)})^2]^{1/2} < \epsilon$ (a small constant ~ 0.001), stop; otherwise, go to Phase IV.

Phase IV: Update ρ

Set $MI = MI + 1$ and $\rho^{[MI+1]} = R \times \rho^{[MI]}$ (default value of $R = 10$). Go to Phase II.

References

Addleman, S.
1962 "Orthogonal main-effect plans for asymmetrical factorial experiments". *Technometrics 4*: 21–46.

Bavelas, A.
1950 "Communication patterns in task oriented groups". *Journal of Acoustic Society of America 22*: 271–281.

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

Brieger, R.L., S.A. Boorman and P. Arabie
1975 "An algorithm for clustering relational data with application to social network analysis and comparison with multidimensional scaling". *Journal of Mathematical Psychology 12*: 328–383.

Burt, R.S.
1977 "Positions in multiple network system. Part One: A general conception of stratification and prestige in a system of actors cast as a social typology". *Social Forces 57*: 106–131.
Burt, R.S.
1980 "Models of network structure". *Annual Review of Sociology 6*: 79–141.
Carroll, J.D.
1976 "Spatial, non-spatial, and hybrid models for scaling". *Psychometrika 41*: 439–463.
Carroll, J.D. and S. Pruzansky
1980 "Discrete and hybrid scaling models". In E.D. Lantermann and H. Feger (eds.), *Similarity and Choice*. Bern: Hans Huber.
Cunningham, J.P.
1978 "Free trees and bidirectional trees as representations of psychological distance". *Journal of Mathematical Psychology 17*: 165–188.
Davis, J.A.
1967 "Clustering and balance in graph". *Human Relations 20*: 181–187.
DeSarbo, W.S. and J.D. Carroll
1985 "Three-way metric unfolding via weighted alternating least squares". *Psychometrika 50*: 275–300.
DeSarbo, W.S. and D. Hoffman
1986 "Simple and weighted unfolding MDS threshold models for the spatial analysis of binary data". *Applied Psychological Measurement 10*: 247–264.
DeSarbo, W.S. and J. Cho
1989 "A stochastic multidimensional scaling vector threshold model for the spatial representation of 'Pick Any/N' data". *Psychometrika 54*: 105–129.
De Soete, G., W.S. DeSarbo, G.W. Furnas and J.D. Carroll
1984 "The estimation of ultrametric and path-length trees from rectangular proximity data". *Psychometrika 49*: 289–310.
Dobson, A.J.
1974 "Unrooted trees for numerical taxonomy". *Journal of Applied Probability 11*: 32–42.
Farris, J.S.
1972 "Estimating phylogenetic trees from distance matrices". *American Naturalist 106*: 645–668.
Forsyth, E. and L. Katz
1946 "A matrix approach to the analysis of sociometric data: Preliminary report". *Sociometry 9*: 340–347.
Freeman, L.C.
1979 "Centrality in social networks: Conceptual clarification". *Social Networks 1*: 215–239.
Furnas, G.W.
1980 "Objects and their features: the metric representation of two class data". Unpublished Doctoral Dissertation, Stanford University, Stanford.
Harary, F., R. Norman and D. Cartright
1965 *Structural Models*. New York: Wiley.
Hartigan, J.A.
1967 "Representation of similarity matrices by trees". *Journal of the American Statistical Association 62*: 1140–1158.
Hubbell, C.H.
1965 "An input-output approach to clique identification". *Sociometry 28*: 377–99.
Hutchinson, J.W.
1989 "Netscal: A network scaling algorithm for nonsymmetric proximity data". *Psychometrika 54*: 25–51.

Jedidi, K. and W.S. DeSarbo
1991 "A stochastic multidimensional scaling methodology for the spatial representation of three-mode, three-way binary data". *Psychometrika*, forthcoming.
Johnson, S.C.
1967 "Hierarchical clustering schemes". *Psychometrika 32*: 241–254.
Klauer, K.C. and J.D. Carroll
1989 "A mathematical programming approach to fitting general graphs". *Journal of Classification 6*: 247–270.
Klauer, K.C. and J.D. Carroll
1990 "A comparison of two approaches to fitting directed graphs to nonsymmetric proximity measures". Working Paper, Bell Laboratories, Murray Hill, NJ.
Klovdahl, A.S.
1981 "A note on image of networks". *Social Networks 3*: 197–214.
Knoke, D. and J.H. Kuklinski
1982 *Network Analysis*. Beverly Hills: Sage Publications, Inc.
Krackhardt, D.
1987 "Cognitive social structures". *Social Networks 9*: 108–33.
Lankford, P.M.
1974 "Comparative analysis of clique identification methods". *Sociometry 37*: 287–305.
Lin, N.
1976 *Foundations of Social Research*. New York: McGraw-Hill.
Luce, R.D.
1950 "Connectivity and generalized cliques in sociometric group structure". *Psychometrika 15*: 169–190.
Luce, R.D. and A.D. Perry
1949 "A method of matrix analysis of group structure". *Psychometrika 14*: 95–117.
Niemien, J.
1974 "On the centrality in a directed graph". *Social Science Research 2*: 371–378.
Northway, M.L.
1949 "A method for depicting social relationships obtained by sociometric testing". *Sociometry 3*: 144–150.
Pendyck, R.S. and D.L. Rubinfeld
1981 *Econometric Models and Economic Models*. New York: McGraw Hill.
Proctor, C.H. and C.P. Loomis
1951 "Analysis of sociometric data". In M. Jahoda et al. (eds.), *Research Methods in Social Relations*. New York: Dryden.
Rao, S.S.
1979 *Optimization Theory and Applications*. New York: Wiley.
Sattath, S. and A. Tversky
1977 "Additive similarity tree". *Psychometrika 42*: 319–345.
Siedman, S.B. and B.L. Foster
1978 "A graph-theoretic generalization of the clique concept". *Journal of Mathematical Sociology 6*: 139–54.
White, H.C., S.A. Boorman and R.L. Brieger
1975 "Social structure from multiple networks: Blockmodels of roles and positions". *American Journal of Sociology 81*: 730–80.