Social Networks 12 (1990) 337–357
North-Holland

LS SETS, LAMBDA SETS AND OTHER COHESIVE SUBSETS *

Stephen P. BORGATTI
University of South Carolina

Martin G. EVERETT
Thames Polytechnic

Paul R. SHIREY
University of California, Irvine

Seidman (1983a) has suggested that the engineering concept of LS sets provides a good formalization of the intuitive network notion of a cohesive subset. Some desirable features that LS sets exhibit are that they are difficult to disconnect by removing edges, they are relatively dense within and isolated without, they have limited diameter, and individual members have more direct links to other members than to non-members. Unfortunately, this plethora of features means that LS sets occur only rarely in real data. It also means that they do not make good independent variables for structural analyses in which greater-than-expected in-group homogeneity is hypothesized with respect to some substantive dependent variable, because it is unclear which aspect of the LS set was responsible for the observed homogeneity. We discuss a variety of generalizations and relations of LS sets based on just a few of the properties possessed by LS sets. Some of these simpler models are drawn from the literature while others are introduced in this paper. One of the generalizations we introduce, called a *lambda set*, is based on the property that members of the set have greater edge connectivity with other members than with non-members. This property is shared by LS sets. Edge connectivity satisfies the axioms of an ultrametric similarity measure, and so LS sets and lambda sets are shown to correspond to a particular hierarchical clustering of the nodes in a network. Lambda sets are straightforward to compute, and we have made use of this fact to introduce a new algorithm for computing LS sets which runs an order of magnitude faster than the previous alternative.

1. Introduction

Since the introduction of the sociogram (Moreno 1934), and the sociomatrix (Forsyth and Katz 1946), one of the main preoccupations

* The authors are grateful for helpful comments by Linton Freeman, Steve Seidman, and especially Katherine Faust. This research was supported in part by grant number R000231292 of the Economic and Social Research Council awarded to Martin Everett.

of what is now called network analysis has been the detection of cohesive subsets in social networks. The first formal description of a cohesive subset was by Luce and Perry (1949), who formalized the *clique* as a maximal set of actors each of whom named the other as a friend in a sociometric interview. Since then a myriad of cohesive subset definitions have appeared in the literature.

In this paper, we take as our point of departure the suggestion by Seidman (1983a) that the electrical engineering concept of LS sets (Lawler 1973; Luccio and Sami 1969) provides a useful formalization of the social networks notion of a cohesive subset. One reason why LS sets are appealing in this context is that they possess many of the characteristics that we intuitively associate with the notion of cohesive subset. In fact, they may have too many features, in the sense that few LS sets are found in observed social networks. This leads to practical problems with using LS sets to analyze empirically derived datasets. The many properties possessed by LS sets can also lead to certain problems of interpretation, which we discuss in the final section. The main objectives of this paper are to explore some of the key properties of LS sets, and to suggest generalizations based on these properties that may be more useful in some applications. In the process, we relate LS sets to a number of alternative models of cohesive subsets. We also introduce a new algorithm for computing LS sets which is an order of magnitude faster and easier to comprehend than previously published methods.

2. Notation, terminology and scope

In this paper, we consider only networks represented as connected, undirected irreflexive graphs. We use $G(V, E)$ to denote a graph with vertex set V and edge set E. We use the terms "vertex", "node", "point", and "actor" synonymously. Similarly, we use "edge", "line", and "link", synonymously. The number of edges linking vertex sets A and B, where A and B are disjoint subsets of V, is represented by $\alpha(A, B)$. When $B = V - A$, we may write $\alpha(A)$. By a slight abuse of notation we also use $\alpha(a, B)$ to denote the number of edges linking a vertex $\boldsymbol{a}$ to a set of vertices B. The subgraph induced by a subset of nodes S is denoted G_S. The minimum degree of a graph or subgraph G is denoted $\delta(G)$. The edge connectivity of a pair of vertices $\boldsymbol{a}$, $\boldsymbol{b}$ is

given by the minimum number of edges that must be removed in order to disconnect them, and is denoted $\lambda(a, b)$. Technically, $\lambda(a, a)$ is infinite, but for the purposes of data analysis we conventionally assign $\lambda(a, a) = \text{MAX}(\{\lambda(i, j): i, j \in V \text{ and } i \neq j\})$. The edge connectivity of a graph G is denoted $\lambda(G)$ and is equal to $\text{MIN}((\lambda(a, b): a, b \in V\})$. The edge connectivity of a subset S is written $\lambda(S) = \text{MIN}(\{\lambda(a, b): a, b \in S))$. The quantity $\lambda(S)$ should not be confused with $\lambda(G_S)$, which is the edge connectivity of the subgraph induced by S.

3. LS sets

The notion of LS sets was first introduced by Luccio and Sami (1969), who termed them *minimal groups*. Lawler (1973) renamed them LS sets, after their authors. The definition is as follows:

Definition 1. Let $G(V, E)$ be a graph. Then a subset H of V is termed an *LS set* if for any proper subset K of H, $\alpha(K, V - K) > \alpha(H, V - H)$.

The essence of the idea is that an LS set may be thought of as the union of its subsets, and this union is "better" than any subset because it has fewer connections to the outside. In the context in which LS sets were introduced, graphs were used to represent circuits of electronic components mounted on silicon chips. The design objective was to group components onto physical chips in such a way as to minimize the number of connections across chips. The LS set definition guarantees that if any subset contained in an LS set were mounted on a chip by itself, it would require more connections to the outside than the LS set

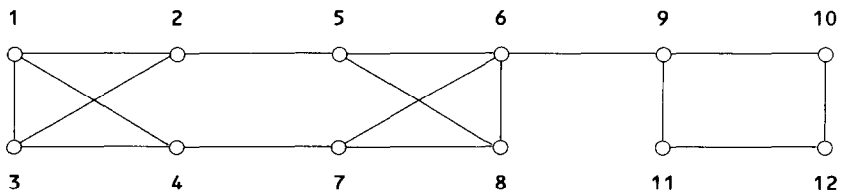

Fig. 1.

would. Hence, each subset is better off joining the LS set than going out on its own, so to speak.

Note that the set of all nodes V is an LS set, and that every singleton subset $\{v\}$, $v \in V$, is also an LS set. We refer to these as "trivial" LS sets. Connected graphs have $|V|+1$ trivial LS sets.

As a network model of a cohesive subset, an LS set has many appealing characteristics. One important property is that each member of an LS set is required to have more connections with other members than with non-members. For example, the graph in Figure 1 contains the following 16 LS sets:

{1} {2} {3} {4} {5} {6} {7} {8} {9} {10} {11} {12}
{1 2 3 4}
{1 2 3 4 5 6 7 8} {9 10 11 12}
{1 2 3 4 5 6 7 8 9 10 11 12}.

For each one (except singletons), it can be seen that individual members of the LS set have more connections to other members than to outsiders. As others have pointed out (Phillips and Conviser 1972; Alba 1973; Sailer and Gaulin 1984), the requirement that members have more ties within-set than without is an important part of what we mean by a cohesive subset.

However, the definition of an LS set requires more than just individual members having more internal than external links. Rather, every proper subset of an LS set must have more ties, as a set, to the rest of the LS set, than to outsiders. Hence, in Figure 1, the LS set {1 2 3 4} contains 14 proper subsets, each of which has more connections to the remaining members than to non-members (see Table 1). Seidman (1983a,b) has proved the following proposition:

Proposition 1. Let $G(V, E)$ be a graph. A subset H of V is an LS set if and only if for any proper subset K of H, $\alpha(K, H-K) > \alpha(K, V-H)$.

The proposition states that LS sets are subsets of actors in which every proper subset K of the subset H has more connections to the remaining members of the subset, $H-K$ than to all outsiders, $V-H$. Since the proposition is an "if and only if", it can be regarded as an alternative to the original definition given by Luccio and Sami.

Table 1
All subsets of LS set $H = \{1\ 2\ 3\ 4\}$ from Figure 1. Note that $V - H = \{5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\}$

Subset K	$H - K$	$\alpha(K, H - K)$	$\alpha(K, V - H)$
{1}	{2, 3, 4}	3	0
{2}	{1 3 4}	2	1
{3}	{1 2 4}	3	0
{4}	{1 2 3}	2	1
{1 2}	{3 4}	3	1
{1 3}	{2 4}	4	0
{1 4}	{2 3}	3	1
{2 3}	{1 4}	3	1
{2 4}	{1 3}	4	2
{3 4}	{1 2}	3	1
{1 2 3}	{4}	2	1
{1 2 4}	{3}	3	2
{1 3 4}	{2}	2	1
{2 3 4}	{1}	3	2

In substantive terms, if we equate the sociological notion of a "group" with the network notion of a cohesive subset, Proposition 1 guarantees that LS sets do not contain splinter groups with more "allegiance" to outsiders than to the rest of the group. Although it is outside the scope of this paper to put forth a theory of group stability and fission (see, for example Tutżauer 1985 and Zachary 1984), we can easily imagine that the existence of such splinter groups would increase the probability of a future schism. Since LS sets cannot contain factions of this kind, we can expect LS sets to be relatively stable over time.

It is a general property of LS sets is that they do not, in the precise sense detailed below, contain minimum weight cutsets. Lawler (1973) has proved the following proposition:

Proposition 2. Let $G(V, E)$ be a graph. Let S be any subset of V, let H and T be proper subsets of S such that H is an LS set and $\alpha(T)$ is minimal. Then H is a subset of T or $S - T$.

The sense of Proposition 2 is as follows. Begin by letting the set S equal V, the set of all nodes. Find a minimum weight cutset C. In the case of ordinary graphs whose edges are not valued, a cutset is a collection of edges whose removal disconnects the graph. A minimum

weight cutset is a cutset with no more edges than any other. Let the two sets of nodes separated by C be called T and $V-T$. Note that $\alpha(T) = |C|$ is minimal, because C is a minimal weight cutset. If there are any LS sets in the graph, aside from V itself, they will be wholly contained in (or equal to) either T or $V-T$. In other words, they cannot "straddle" a minimum weight cutset. This is the simplest case. The more complicated case allows S to be any subset of V. For example, we could let S equal the set T found above in the simple case. Within this new S we find a new T such that $\alpha(T)$ is minimal. Again, this amounts to finding a minimum weight cutset separating some members of S (to be called T) from all other nodes (to be called $V-T$). The proposition says that any LS sets wholly contained in this S will be found in T or $S-T$.

Substantively, Proposition 2 reinforces Seidman's claim that LS sets are highly cohesive. As Zachary (1977) has suggested in a slightly different context [1], minimum weight cutsets may be thought of as fault lines in a network, along which rifts or breakups are likely to occur. LS sets do not, in the relative sense specified by the proposition, contain these fault lines.

A consequence of Proposition 2 is that LS sets are relatively robust in the face of random removal of edges. In order to disconnect an LS set (*i.e.*, isolate one or more members from the others), a relatively large number of edges must be removed. For example, a minimum of 3 edges must be removed from the graph in Figure 1 in order to disconnect members of the LS set {1 2 3 4}, yet only 2 edges need to be removed to disconnect any member of this set from any non-members (and only 1 edge to disconnect from some non-members). The reason LS sets are so hard to disconnect is that every pair of points within the set is connected by a relatively large number of independent paths. Independent paths are edge-disjoint, which means they have no edges in common. Removing an edge from one path destroys only that path, so that other paths connecting the points are not disturbed. In other words, members of LS sets are joined to other members by more edge-disjoint paths than they are to non-members. Consequently, graphs subject to random deletion of edges will tend to retain connections within LS sets, while losing connections between them. The general

[1] Zachary's primary interest was in the interpretation of minimum weight cutsets separating pairs of actors who were identified, a priori, as leaders of opposing factions.

principle, given by Proposition 3, is that the edge connectivity between members of an LS set is greater than the connectivity between members and non-members.

Proposition 3. Let H be an LS set of a graph $G(V, E)$. Then for all u, v, $w \in H$ and $x \in V - H$, $\lambda(u, v) > \lambda(w, x)$.

Proof. We first note that if e is an edge not in H then H is an LS set in $G - e$. Deleting edges outside H can lower the value of $\lambda(H)$ as well as $\lambda(w, x)$, but $\lambda(H)$ is bounded below by the edge connectivity of the induced subgraph G_H. If $\lambda(H)$ is less than or equal to $\lambda(w, x)$ then consider the induced subgraph G' formed by H and all (w, x) disjoint paths. H is still an LS set in G' with $\lambda'(H) \leq \lambda(H)$ and $\lambda(w, x)$ is unchanged. It follows that H must contain a minimal cutset, contradicting Lawler's result. □

The characterization of LS sets in terms of edge connectivities has some practical benefits. Although we have ignored the point so far, the reader will have noticed that LS sets can contain each other, but they cannot partially overlap. In other words if H and L are LS sets, then either $H \cap L = \varnothing$ or $H \cap L = H$ or $H \cap L = L$ (Luccio and Sami 1969). Consequently, we can represent the set of all LS sets in a graph as series of nested partitions of the set of vertices in the graph. For example, the LS sets in Figure 1 form the following partitions:

{{1} {2} {3} {4} {5} {6} {7} {8} {9} {10} {11} {12}}
{{1 2 3 4} {5} {6} {7} {8} {9} {10} {11} {12}}
{{1 2 3 4 5 6 7 8} {9 10 11 12}}
{{1 2 3 4 5 6 7 8 9 10 11 12}}

We can represent these nested partitions as a dendrogram (see Figure 2) such as produced by hierarchical clustering programs. One question

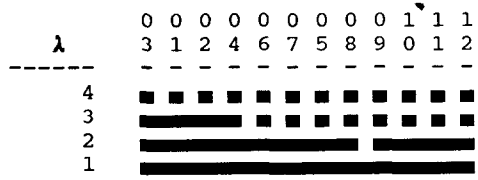

Fig. 2. Dendogram representing the LS sets of the graph in Figure 1.

that comes to mind, is in what important way do the smaller, less inclusive LS sets differ from the larger, more inclusive LS sets? The answer is edge connectivity. In Figure 2, the first column gives the edge connectivity of all LS sets found at that level in the partition. Members of the highest, most exclusive LS sets have high edge connectivity with each other and are therefore highly cohesive. Members of the lower, less exclusive LS sets are less well connected. Thus, edge connectivity provides a framework for organizing the LS sets of a graph.

4. Lambda sets

The fact that LS sets form nested partitions, where the level of nesting corresponds to the internal cohesiveness of the subsets, can be explored further. Suppose we compute the edge connectivity between every pair of points in a graph, and represent the information in a matrix as shown in Figure 3. Notice that all points whose maximum edge connectivity is 3 have identical rows in the matrix, aside from the cells corresponding to each other. In other words if points i and j have the same maximum connectivity, then for all $k \neq i \neq j$, then $\lambda(i, k) = \lambda(j, k)$. For example, the rows associated with nodes 1 and 2 are identical except for the values in columns 1 and 2. More generally, for every triple i, j, $k \in V$, at least two of following must be equal: $\{\lambda(i, j), \lambda(j, k), \lambda(i, k)\}$. This occurs because edge-connectivity

4	3	3	3	2	2	2	2	1	1	1	1
3	4	3	3	2	2	2	2	1	1	1	1
3	3	4	3	2	2	2	2	1	1	1	1
3	3	3	4	2	2	2	2	1	1	1	1
2	2	2	2	4	3	3	3	1	1	1	1
2	2	2	2	3	4	3	3	1	1	1	1
2	2	2	2	3	3	4	3	1	1	1	1
2	2	2	2	3	3	3	4	1	1	1	1
1	1	1	1	1	1	1	1	4	2	2	2
1	1	1	1	1	1	1	1	2	4	2	2
1	1	1	1	1	1	1	1	2	2	4	2
1	1	1	1	1	1	1	1	2	2	2	4

Fig. 3. Matrix of edge-connectivities between pairs of vertices for the graph in Figure 1.

```
      0 0 0 0 0 0 0 0 0 1 1 1
  λ   1 2 3 4 5 6 7 8 9 0 1 2
----- - - - - - - - - - - - -
  4   . . . . . . . . . . . .
  3   XXXXXXX XXXXXXX . . . .
  2   XXXXXXXXXXXXXXX XXXXXXX
  1   XXXXXXXXXXXXXXXXXXXXXXX
```

Fig. 4. Hierarchical clustering of edge connectivities given in Fig. 3. Each cluster is a lambda set with connectivity λ.

satisfies the triangle inequality condition of an ultrametric, namely that for all $i, j, k \in V$, $\lambda(i, k) \geq \text{MIN}(\lambda(i, j), \lambda(j, k))$.

The fact that edge connectivity satisfies the axioms of an ultrametric [2] implies (Johnson 1967) that there exists a corresponding hierarchical clustering of the nodes of the graph. Thus, the matrix of edge connectivities given in Figure 3 corresponds to the hierarchical clustering given in Figure 4. Note that all the LS sets of the graph are among the clusters, but there are other ones as well. This result suggests the possibility of an interesting generalization of LS sets, which we will call *lambda sets*. The definition is as follows:

Definition 2. Given a graph $G(V, E)$, a *lambda set* S is a subset of V such that for all $a, b, c \in S$ and $d \in V - S$, $\lambda(a, b) > \lambda(c, d)$.

A lambda set is a maximal subset of actors who have more edge-independent paths connecting them to each other than to outsiders. Since we can characterize a lambda set S by the minimum connectivity among its members, we can also refer to lambda sets as *lambda-**k*** sets, where $\boldsymbol{k} = \lambda(S)$. In the dendrogram in Figure 4, every cluster is a lambda set (Proposition 3), and the connectivity within each set is given by the level of clustering.

A convenient feature of lambda sets is that they are more numerous in real data than LS sets, and therefore more practical as an analytic tool. Yet since every LS set is a lambda set (Proposition 3), any analysis based on lambda sets will not contradict one based on LS sets. Whenever it makes sense to break down the structure of a graph into

[2] The other two axioms of an ultrametric are easily satisfied since $\lambda(i, j) = \lambda(j, i)$ and λ achieves its maximal value only for $\lambda(i, i)$.

LS sets, it also makes sense to examine the lambda sets, since they can only add additional information to the analysis.

The fact that lambda-sets are disjoint at a given λ-level, rather than partially overlapping, is convenient for data analysis. From a measurement point of view, the lambda set an actor falls in may be seen as a categorical attribute of that actor, scarcely different in this context from his or her ethnic group, race, party affiliation, birthplace, and so on. Thus, the assignment of actors to lambda sets may be used as a categorical variable in further analysis. For example, it might be used to predict amount of knowledge about the other actors possessed by the actor, or to estimate the probability of adoption of an innovation.

The fact that lambda sets generate a series of groupings that are nested hierarchically within each other means that the data analyst is able to choose the level of detail to analyze. All lambda set groupings describe the same cohesive subset structure of a network, but they vary according to the coarseness or fineness of the description.

Lambda sets are easy to interpret. Actors form lambda sets if there are more distinct paths linking them to each other than to others. Actors in lambda sets with connectivity λ have a minimum of λ independent paths linking any one to any other. When λ is large, a lambda set describes a subset that is relatively difficult to disconnect by means of edge removals. For example, in a graph where nodes represent cities and edges represent roads, a lambda set is a collection of cities which, in the event of a natural disaster are unlikely to be completely disconnected from each other.

Lambda sets may be viewed as pockets of a network which are less vulnerable to disruption than other areas. We can imagine that if what flows across the edges of the network are information or goods or even genes, then over time we can expect members of the same lambda set to be more homogeneous with respect to the goods, information or genes they possess. For example, we expect animals of a given species living in geographic areas accessible via multiple independent paths should share more genetic traits, all else being equal, than animals nearly isolated from each other. Similarly, human groups that maintain multiple, independent, paths of communication are more likely to share cultural traits than those with few paths, if there is a high likelihood of edge destruction.

Even in affective networks, we might expect that subsets that are lambda sets are much more likely to survive spats between individual

members than other subsets, because bad feeling between any pair of members will not cut the subset in half: all the other members will continue to be connected to each other despite the missing link. Thus, problems remain local rather than necessarily escalating into group fission. In this sense, we can expect lambda sets to demonstrate greater stability or persistence over time than subsets with less edge connectivity.

5. Computation of lambda sets and LS sets

Lambda sets may be computed via a simple two-step algorithm. In the first step, we compute a matrix C of edge connectivities such that $C_{ij} = \lambda(i, j)$ for $i, j \in V$. Ford and Fulkerson's (1956) well known "max-flow, min-cut" theorem states that the capacity of a minimum weight cutset separating two vertices (*i.e.*, their edge connectivity) is equal to the maximum flow between them. Ford and Fulkerson's constructive proof yields an algorithm for computing the maximum flow between any two vertices. The basic algorithm is reproduced in many standard texts of graph theory (such as Bondy and Murty 1976) and combinatorics (such as Nijenhuis and Wilf 1975). For non-valued graphs with n vertices and m edges the standard algorithm runs in $O(nm)$ time for each pair of vertices, or approximately $O(n^3m) \leq O(n^5)$ time for the entire matrix of edge connectivities. However, since we know in advance that we will need to compute flows between all pairs of vertices, some efficiencies are possible, thanks to an algorithm by Gomory and Hu (1964) which runs in $O(n^4)$ time.

In the second step, we perform the equivalent of a hierarchical clustering on C. To do this, we successively partition V such that $P^k(i) = P^k(j)$ only if $\lambda(i, j) \geq k$, where $0 \leq k \leq \text{Max}[\lambda(i, j) \mid i, j \in V]$. For example, if we implement P as an integer matrix with $\text{Max}[\lambda(i, j) \mid i, j \in V] + 1$ rows and $|V|$ columns, then we should have the following code fragment:

```
for k := 0 to MaxC do begin
  for j := 1 to n do P[k, j] := j;
  for i := 2 to n do for j := 1 to i - 1 do if C[i, j] > = k
    then P[k, i] := P[k, j];
end;
```

On output, each row of P stores a different, hierarchically nested, clustering of nodes. Two nodes i and j are in the same lambda set at level k if and only if $P[k, i] = P[k, j]$.

LS sets may be computed from the set of all lambda sets by exploiting a theorem by Luccio and Sami (1969). The theorem states that the union of two or more LS sets S_i is itself an LS set if $\alpha(S_1 \cup S_2 \cup \ldots \cup S_m) < \text{MIN}(\alpha(S_1), \alpha(S_2), \ldots, \alpha(S_m))$. In other words, the union of the LS sets is an LS set if its outdegree is less than that of any of its LS set constituents. The importance of the theorem is that we do not need to consider all possible subsets in evaluating whether a given set is an LS set. Rather, we need only consider those subsets which are themselves LS sets. Further, our own Proposition 3 guarantees that all LS sets are lambda sets, so we need not test any but the set of lambda sets for the LS condition.

This suggests the following procedure. Begin with the smallest lambda set S that is not a single vertex. Since S is the union of smaller LS sets (singletons), we know that S is an LS set if for all $s_i \in S$, $\alpha(S, V - S) < \text{MIN}(\alpha(s_1, V), \alpha(s_2, V), \ldots)$. Repeat for every other lambda set whose elements do not include other lambda sets. Now take the smallest lambda set obtainable as a union of the LS sets created in the first round. Test whether it is an LS set by determining whether its outdegree is less than that of any of the LS sets within it. Repeat for every other lambda set. The process continues until the only lambda set left is the one containing all vertices in the graph. Since this process obviously executes in less than $O(n^4)$ time, the entire procedure for extracting LS sets including the lambda set step runs in $O(n^4)$ time. This compares favorably with Luccio and Sami's non-polynomial algorithm and Lawler's $O(n^5)$ algorithm. [3]

6. Other generalizations and relaxations of LS sets

Lambda sets capture one aspect (connectivity) of the notion of a cohesive subset very nicely. Unlike LS sets, however, they completely

[3] Actually, Lawler (1973: 282) claims only a modest $O(m^2n^4) \leq O(n^8)$ time for his algorithm. However, the maximum flow algorithm he uses is designed for valued hypergraphs, not simple graphs. Substituting a fast $O(n^3)$ flow algorithm such as Dinic's (1970) brings the overall procedure down to $O(n^5)$. The advantage of our algorithm over Lawler's, besides simplicity, is that ours can make use of Gomory and Hu's efficient algorithm for computing flows between all pairs of vertices whereas Lawler's cannot.

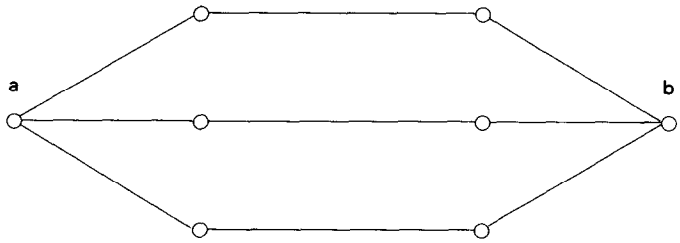

Fig. 5. Nodes a and b for a lambda-3 set together, despite being only distantly connected.

fail to capture many other key aspects. For example, while members of lambda sets are required to have more connections with insiders than outsiders, these connections need not be direct. For example, in the graph in Figure 5, members of lambda set $\{a, b\}$ have no direct links with each other. Even worse, the intermediaries that link them are outsiders. Despite their usefulness in some settings, lambda sets are a very different kind of cohesive subset than we normally think of, and cannot be thought of as a general purpose model.

LS sets on the other hand, capture many of the key aspects underlying the intuitive notion of cohesive subset. For example, it has been shown here or in Seidman (1983a,b) that LS sets are in certain ways difficult to disconnect, relatively dense within and isolated without, of limited diameter, and so forth. Unfortunately, LS sets are also relatively rare in empirical datasets, perhaps because they do impose so many conditions on membership. Consequently, they can be of limited value in real data analysis situations. There is also a problem with interpretability. Because LS sets have a number of important characteristics, structural studies that use the LS clusterings as independent variables face difficulties interpreting the results. For example, if it happens that the LS sets of a graph do show statistically significant homogeneity or homophily with respect to a substantive variable of interest (e.g. members of the same set share more attitudes than non-members), we cannot determine what structural property is the cause. Is it the large number of face-to-face contacts or the multitude of independent indirect paths? LS sets have both. For structural investigations such as these, it would be useful to define simpler, purer subsets which generalize LS sets along just one or two key properties.

For example, one basic property of LS sets that is very appealing is the condition that individual members have more links to other mem-

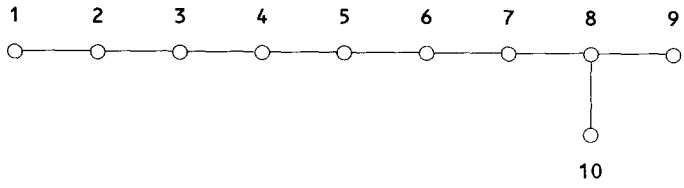

Fig. 6. Sets {1...9} and {8 9 10} are alpha sets, among others.

bers than to non-members. This is not a property possessed by lambda sets. We define a generalization of LS sets called *alpha sets* based exclusively on this principle:

Definition 3. Let $G(V, E)$ be a graph. An *alpha set* S is a subset of V such that for all $s \in S$, $\alpha(s, S) > \alpha(s, V - S)$.

An unfortunate characteristic of alpha sets is that they can be quite numerous, and they can partially overlap. For example, the graph in Figure 1 contains both alpha set $A = \{1\ 2\ 3\}$ and alpha set $B = \{1\ 3\ 4\}$. Another unfortunate characteristic is that since the quantifying condition is relative degree rather than absolute degree and it is expressed at the level of the individual member, there is little we can say about the connectivity or density of the subset as a whole. For example, an alpha set can consist primarily of hangers-on or pendants, as shown in Figure 6. Such nodes are more connected to the subset than to others simply because they are not connected to any other points.

Of course, the notion of alpha set is not as pure and simple as possible, since it implicitly requires attention to both inner and outer connections. We could achieve greater clarity by taking each of these types of connections separately. For example, we could define a structure whose only qualification is that members of the subset have no fewer than k connections with other members, regardless of their connections with non-members:

Definition 4. Let $G(V, E)$ be a graph. A *k-degree set* is a subset of V such that for all $s \in S$, $\alpha(s, S) \geq k$.

A k-degree set S induces a subgraph G_S in which the minimum degree $\delta(G_S) = k$. Like alpha sets, k-degree sets are numerous and overlap-

ping. Seidman (1983a,b) has shown that maximal k-degree sets, which he terms k-cores, may be regarded as "seedbeds, within which we can expect highly cohesive subsets to be found". Unlike ordinary degree sets, k-cores do not partially overlap. Rather, for any given level of $\boldsymbol{k}$, k-cores form a set of disjoint sets of vertices of the graph. Like lambda sets and LS sets, k-cores form hierarchical clusterings of the nodes of a graph.

All LS sets are k-degree sets where $k = \alpha(H)/2$, but they are not necessarily k-cores. They must be contained in any k-core in which $k < \alpha(H)/2$.

Another well-known structure which is generalized by the notion of a k-degree set is the clique (Luce and Perry 1949). A clique S of a graph $G(V, E)$ is a maximal subset of V such that for all $s \in S$, $\alpha(s, S) = |S| - 1$. Thus, a clique is a maximal set of actors each of whom is directly connected to every other actor in the set. Cliques may also be described as maximal $(|S| - 1)$-degree sets. Like LS sets, cliques are not k-cores but are wholly contained by them, provided $k < |S| - 1$.

Seidman and Foster (1978) have introduced a generalization of the clique called a k-plex. A k-plex S of a graph $G(V, E)$ is a subset of V such that for all $s \in S$, $\alpha(s, S) \geq |S| - k$. Thus a clique may be described as a maximal k-plex. A k-plex of size $|S|$ will be wholly contained in a p-core C with $p = |C| - k$.

Building on the work of Luce (1950) and Alba (1973), Mokken (1979) has introduced a family of three generalizations of cliques, namely n-clans, n-clubs, and n-cliques. An n-clique S of a graph G is a maximal subset of V such that for all $u, v \in S$, $d(u, v) <= n$. An n-club S of G is a maximal subset of V such that the diameter of the subgraph induced by S is less than or equal to n. An n-clan S is an n-clique of G such that the diameter of the subgraph induced by S is less than or equal to n. While n-clans are both n-cliques and n-clubs, n-clubs are not necessarily n-cliques, though they are always wholly contained by an n-clique. Like all generalizations of cliques, including k-degree sets and k-plexes, n-cliques n-clubs and n-clans are concerned only with connections within subsets. They differ from k-degree sets and k-plexes, however, in that they generalize cliques by relaxing distances among members rather than the number of links between members.

Since structures defined entirely in terms of within-set connections

have proven useful, it is reasonable to consider structures defined only in terms of out-set connections. For example, a structure analogous to k-degree sets is as follows:

Definition 5. A k-outdegree set S of a graph G is a subset of V such that for all $s \in S$, $\alpha(s, V - S) \leq k$.

Essentially, k-outdegree sets are subsets whose members have no more than $\boldsymbol{k}$ connections with non-members. An entire graph is a 0-outdegree set. In themselves, k-outdegree sets are not particularly useful, since in unconnected graphs, even collections of isolates would qualify. However, like k-degree sets, they form subregions of graphs which contain important subsets like lambda sets and LS sets.

One type of k-outdegree set is the minimal set described by Lawler (1973). A minimal set S of a graph $G(V, E)$ is a subset of V such that $\alpha(S, V - S)$ is as small as possible. In other words, S is a minimal set if no other subset of V can be found which has fewer connections with outsiders. Note that whereas the qualifying condition of k-outdegree sets is phrased in terms of individual actors, the qualifying condition of minimal sets is phrased in terms of the subset as whole. Minimal sets may be viewed as the result of minimum weight cutsets. Removing a minimum weight cutset from a graph divides the graph into two components, each of which is a minimal set. Because minimal sets are created by splitting graphs at their weakest points (*i.e.*, where the fewest cuts need to be made), they may be viewed as regions of the graph which are relatively cohesive or connected. LS sets have already been shown to be wholly contained by minimal sets (Proposition 2), and lambda sets are also clearly contained by them.

7. Discussion

Two kinds of relations between different models of subsets have been discussed. One is the generalization relation that records which models are "a type of' which others. For example, LS sets are a type of lambda set, just as cliques are a type of k-plex. The other relation is the subset relation that records which models "are contained in" which others, in the special sense used in Proposition 2. For example, k-plexes are contained in k-cores while lambda sets are not.

The various models of subsets can also be related by the extent to which they share specific features or properties, such as restrictions on the minimum number of internal links, the maximum number of external links, the maximum distance among members, the diameter of the induced subgraph, etc. Subset models can viewed as unique combinations of features, and some models can be "derived" from others by adding or subtracting features. For example, n-clans may be "derived" from n-cliques by adding a restriction on the diameter of the induced subgraph. The features found in the set of subset models described in the network literature to date can be viewed as a set of generative elements which could be combined in various ways to form a very large set of models, most of which are yet to be named or described in the literature.

One question that arises is whether any of these models, known or unknown, is better than any other. We suggest that if the purpose of these various subset definitions were to formalize the intuitive sociological notion of a group, then the winner should be the model which captures as many of the characteristics underlying the intuitive notion as possible. Among the models reviewed in this paper, this would probably be the LS set. The fact that LS sets are empirically rare should, in this context, be regarded as a problem and a challenge to develop a means of quantifying the extent to which a collection of actors departs from the conditions of an LS set, or alternatively, of measuring the extent to which individuals depart from belonging to a given LS set.

However, we must also ask whether formalizing the sociological notion of group is a sensible goal, given the fact that, at present, no precise theory of social groups exists. Sociological usage of the term "group" is nearly as vague as lay usage. In the absence of a theory of groups, the sociologist's intuitive understanding of the word "group" is scarcely different from that of any native speaker of English. Consequently, attempts to formalize the intuitive notion of group unfortunately reduce to an exercise in mathematical ethnography, which presumably is not what most network researchers intend.

Rather, network researchers try to use these mathematical models as explanatory variables. For example, a study of cultural consensus might hope to find that members of the same cohesive subset share significantly more cultural traits than members of different subsets. For another example, members of the same cohesive subset might be found

to frequently suffer the same diseases at similar times. It is important to note, however, that the cohesive subsets do not in themselves explain the substantive outcome. Rather, they "stand in" for the structural variables that are related to the dependent variables. Since different cohesive subset definitions embody different features or properties, they stand in for different structural independent variables.

For example, if the probability of an actor contracting a certain disease were a function of the number of direct links between the individual and carriers of the disease, then we would expect dense regions of the network to show greater homogeneity with respect to infection of the disease than less dense regions. Since cliques are precisely maximally dense regions of graphs, we could reasonably expect common membership in cliques to predict homogeneity with respect to disease infection. Hence a cohesive subset such as a clique can serve as a convenient surrogate for the key underlying structural variable, which in this case would be the number of links with disease carriers.

Similarly, suppose that the sharing of cultural traits depends upon the existence (and perhaps quantity) of direct or indirect avenues of communication between actors. Then in the presence of disruptive factors such as natural disasters which could sever these pathways, we would expect that, over time, sets of actors connected by a larger number of wholly edge-independent paths would show a greater proportion of shared traits than actors connected by few independent paths. Since lambda sets are precisely subsets of actors with more edge-disjoint paths to each other than to outsiders, we would expect common lambda set membership to be a good predictor of cultural consensus. Again, the cohesive subset (here, the lambda set) provides a convenient categorical independent variable that stands in for the key structural variable, which in this case is the number of independent paths linking pairs of actors.

If we take this approach to using models of cohesive subsets, it becomes clear that, contrary to our previous conclusion, complex models like LS sets are not as useful as simpler models such as lambda sets, because the complex models stand in for several important variables at the same time. That is, in complex definitions like LS sets, such variables as the maximum distance between members, the number of disjoint paths linking members, the ratio of direct to indirect links to other group members from each member, and so on, all coincide

perfectly, so that each acts as a mask for the other. We could achieve greater explanatory power by directly correlating separate measures of each of these variables with the dependent variable than by collapsing them together into a complex subset definition.

For example, if we were investigating the social determinants of a monastery breaking in two and it happened that common clique membership failed to predict the schism while common lambda set membership succeeded, we could tentatively conclude that having many direct ties among actors was not as important in maintaining cohesion as having many independent paths linking them together. Thus, in this approach to using cohesive subsets, we try to apply many different definitions of subsets to reveal different aspects of the structure, rather than to choose the one subset definition that best captures the analyst's intuitive understanding of the group concept. Hence the need for relatively simple cohesive subset models such as lambda sets.

In summary, we have taken seriously Seidman's suggestion that the engineering concept of LS sets provides a useful formalization of the notion of a cohesive subset. LS sets capture many of the key aspects underlying the intuitive notion of a cohesive subset (and, incidentally, a social group). For example, they are difficult to disconnect by removing edges, they are relatively dense within and isolated without, they have limited diameter, members have more direct links to other members than to non-members, and so on. Unfortunately, this plethora of features means that LS sets occur only rarely in real data. It also means that they do not make good independent variables for structural analyses in which greater-than-expected in-group homogeneity is hypothesized with respect to some substantive dependent variable, because it is unclear which aspect of the LS set was responsible for the observed homogeneity. We have discussed a variety of generalizations and relaxations of LS sets based on just a few of the properties possessed by LS sets. Among these is the lambda set, based on the LS property that members of the set have greater edge connectivity with other members than with non-members. Lambda sets may prove especially useful in the study and design of networks which are subject to disruption by removal or decay of edges.

References

Alba, R.D.
1973 "A graph-theoretic definition of a sociometric clique". *Journal of Mathematical Sociology* 3: 113–126.

Bondy, J.A. and U.S.R. Murty
1976 *Graph Theory with Applications*. Amsterdam: North-Holland.
Dinic, E.A.
1970 "Algorithms for the solution of a problem of maximum flow in a network with power estimation". *Doklady Nauk. S.S.S.R. 11*: 1277–1280.
Edmonds, J. and R.M. Karp
1972 "Theoretical improvements in algorithmic efficiency for network flow problems". *Journal of the Association for Computing Machinery 19*: 248–264.
Elias, P., A. Feinstein and C.E. Shannon
1956 "A note on the maximum flow through a network". *IRE Transactions on Information Theory IT-2*: 117–119.
Ford, L.R. Jr. and D.R. Fulkerson
1956 "Maximum flow through a network". *Canadian Journal of Mathematics 8*: 399–404.
Forsyth, E. and L. Katz
1946 "A matrix approach to the analysis of sociometric data: Preliminary report". *Sociometry 9*: 340–347.
Gomory, R.E. and T.C. Hu
1964 "Synthesis of a communication network". *Journal of SIAM (Appl. Math.) 12*: 348.
Johnson, S.C.
1967 "Hierarchical clustering schemes". *Psychometrika 32*: 241–254.
Lawler, E.L.
1973 "Cutsets and partitions of hypergraphs". *Networks 3*: 275–285.
Luccio, F. and M. Sami
1969 "On the decomposition of networks into minimally interconnected networks". *IEEE Transactions on Circuit Theory CT-16*: 184–188.
Luce, R.D.
1950 "Connectivity and generalized cliques in sociometric group structure". *Psychometrika 15*: 169–190.
Luce, R.D. and A. Perry
1949 "A method of matrix analysis of group structure". *Psychometrika 14*: 94–116.
Menger, K.
1927 "Zur allgemeinen Kurventheorie". *Fundamenta Mathematica 10*: 96–115.
Mokken, R.J.
1979 "Cliques, clubs and clans". *Quality and Quantity 13*: 161–173.
Moreno, J.L.
1934 *Who Shall Survive? A New Approach to the Problem of Human Interrelations*. New York: Beacon House.
Nijenhuis, A. and H.S. Wilf
1975 *Combinatorial Algorithms*. Academic Press.
Phillips, D.P. and R.H. Conviser
1972 "Measuring the structure and boundary properties of groups: Some uses of information theory". *Sociometry 35*: 235–254.
Sailer, L.D. and S.J.C. Gaulin
1984 "Proximity, sociality, and observation: The definition of social groups". *American Anthropologist 86*: 91–98.
Seidman, S.B.
1983a"LS sets and cohesive subsets of graphs and hypergraphs". *Social Networks 5*: 92–96.
Seidman, S.B.
1983b"Internal cohesion of LS sets in graphs". *Social Networks 5*: 97–107.
Seidman, S.B. and B.L. Foster
1978 "A graph-theoretic generalization of the clique concept". *Journal of Mathematics Sociology 6*: 139–154.

Tutzauer, F.
1985 "Toward a theory of disintegration in communication networks". *Social Networks* 7: 263–285.
Whitney, H.
1932 "Congruent graphs and the connectivity of graphs". *American Journal of Mathematics 54*: 150–168.
Zachary, W.W.
1977 "An information flow model for conflict and fission in small groups". *Journal of Anthropological Research 33*: 452–473.
Zachary, W.W.
1984 "Modelling social network processes using constrained flow representations". *Social Networks 6*: 259–292.