ELSEVIER

Social Networks 21 (1999) 397–407

SOCIAL NETWORKS

www.elsevier.com/locate/socnet

Peripheries of cohesive subsets

Martin G. Everett [a,*], Stephen P. Borgatti [b,1]

[a] University of Greenwich School of Computing and Mathematical Sciences, 30 Park Row, London SE10 9LS UK

[b] Department of Organization Studies, Carroll School of Management, Boston College, Chestnut Hill, MA 02467 USA

Abstract

Network analysts have developed a number of techniques for identifying cohesive subgroups in networks. In general, however, no consideration is given to actors that do *not* belong to a given group. In this paper, we explore ways of identifying actors that are not members of a given cohesive subgroup, but who are sufficiently well tied to the group to be considered peripheral members. We then use this information to explore the structure of the network as a whole.

1. Introduction

In a companion piece, Borgatti and Everett (1999) have proposed a formal model of core/periphery structures. In their model, [2] a network has a core/periphery structure if the network can be partitioned into two sets: a core whose members are densely tied to each other, and a periphery whose members have more ties to core members than to each other. Implicit in their discussion is the conception of a core as a cohesive subgroup, and in particular as a clique (Luce and Perry, 1949). In the cohesive subgroups literature, however, many other formal definitions of subgroups have been proposed [for reviews, see Scott (1991) and Wasserman and Faust (1994)]. Theoretically, any of these could have been used instead of cliques to define network cores.

Furthermore, Borgatti and Everett do not seriously consider the possibility of multiple cores. Their concern is with detecting whether the network as a whole forms a

* Corresponding author. Tel.: +44-020-8331-8716; fax: +44-020-8331-8665; e-mail: m.g.everett@gre.ac.uk

[1] Tel.: +1-617-552-0452, fax: +1-617-552-4230, e-mail: borgatts@bc.edu.

[2] Borgatti and Everett present two formal models, one discrete and the other continuous. In this paper we are concerned only with the discrete model.

0378-8733/99/$ - see front matter
PII: S0378-8733(99)00020-9

core/periphery structure. Logically, however, any cohesive subgroup can be regarded as the core of a highly localized region of the network. From this perspective, any node can be classified as a member of a local core, or as a member of the periphery of that core, or as unrelated to either one.

Thus, whereas Borgatti and Everett essentially take a network and seek to identify the subgraph that serves as its core, this paper takes a subgraph which has been declared to be a core and seeks to identify the network region for which it serves as core. That region is the core's periphery.

2. Defining peripheries

We take as a starting point the identification of a specific cohesive subgroup via any of the methods and definitions found in the literature. These would include such groups as a clique (Luce and Perry, 1949), an n-clique (Luce, 1950; Alba, 1973), an n-club or n-clan (Mokken, 1979), or a k-plex (Seidman and Foster, 1978). Once chosen, we declare the subgroup to be a core, and then seek to identify its periphery.

An obvious approach is to define the periphery as the set of all vertices not in the core that are adjacent to at least one member of the core. Restricting the periphery in this way to 'hangers-on' is appropriate if we think of the periphery as actors that are clearly associated with the core (and perhaps would like to move into the core). However, for other purposes we may prefer to include other nodes — not directly connected to the core — as part of the periphery as well. That is, we could conceive the periphery as simply all outsiders. These are two extremes of a continuum of possible definitions and it would make sense to have a general definition of periphery that could encompass all of these possibilities. We therefore propose the following:

Definition 1. Let G be any graph, and let C be a cohesive subgroup, called a *core* of G. Then the periphery P is $G - C$. If $v \in P$ then we say v is in the k-periphery if v is a distance less than or equal to k from C.

The measure of distance between a node v and a group C is left deliberately undefined, so as to keep the definition completely general. Both the method of detecting the cohesive subgroup and the distance measure can be determined by the researcher and can therefore be related to the type of data to be analyzed. An obvious choice of distance would be the 'nearest neighbor'. That is, the distance from a node to the core is defined as the shortest graph-theoretic distance from the node to any member of the core. By analogy to the hierarchical clustering of Johnson (1967), an obvious alternative choice of distance would be the 'farthest neighbor' or maximum distance. Other choices would include the average distance, the median distance, and so on. It should also be noted that distance to the core need not be defined in terms of graph-theoretic distance at all. For example, a metric based on the extent to which the node is structurally equivalent to the members of the core would do just as well.

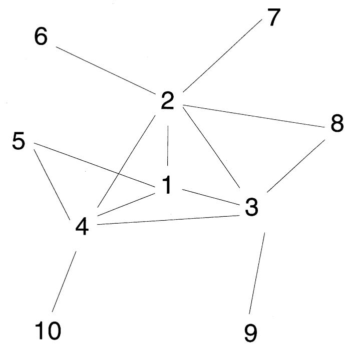

Fig. 1.

Let us assume that we take average geodesic distance as the definition of distance between a node and the core. In this case, the ''k'' of the k-periphery is not necessarily an integer, and provides a sensitive measure of the ''coreness'' of each peripheral point. (In fact, it can also be used to measure the coreness of each member of the core as well.)

If instead we take the minimum geodesic distance as the definition of node-to-core distance, then k does not provide as sensitive a measure of coreness. For example, consider the network in Fig. 1, taking the clique {1,2,3,4} as the core. All other points in the graph comprise the 1-periphery, even though some of these nodes are better connected to the core than others. Since even in graphs with 2-peripheries and 3-peripheries the 1-periphery is likely to be a focus of interest, it makes sense to try to develop a secondary measure of coreness that distinguishes among members of the 1-periphery. Our measure will be based loosely on the idea that members of the 1-periphery have as a primary goal the desire to become members of the core. [3] The measure, $CP(v)$, we propose is as follows:

Definition 2. For $v \in V$, if $v \in C$ then $CP(v) = 1$, otherwise let q be the minimum number of edges incident with v that are required to make v part of C and let r be the number of those edges that are already incident with v. Then $CP(v) = r/q$. We call $CP(v)$ the coreness of vertex v.

The CP measure varies from 0 to 1. All core actors have a value of 1. Given the choice of nearest neighbor distance in Definition 1, all 1-periphery actors will have a non-zero value for cores defined by any of the standard cohesive subgraph models. Note that while k-peripheries are defined in terms of distance, the CP measure is defined in terms of volume of ties. It is well known that there exist two key dimensions of cohesion

[3] As an anonymous review has pointed out, such a desire could be seen as consistent with maximizing the player's structural holes (Burt, 1992) if the network has only one core, but not if the network has multiple cores.

in networks: distance and density of ties. For example, there are many relaxations of the clique concept. Some, like *n*-clique, are based on relaxing the distances within the group. Others, like *k*-plex, are based on relaxing the density of ties. The relaxation principles are complementary and should be used together. Hence, it is very appropriate that we use distance to define the 1-periphery, and then use density of ties to subdivide the 1-periphery. [4]

Given that we are using the Luce and Perry clique as the basis for identifying the core in Fig. 1, to enlarge the core we must supply a vertex that is adjacent to all four of the existing core members. This would mean that $q = 4$. It follows that in Fig. 1 the vertices with degree 1 have $CP(v) = 0.25$, the vertices with degree 2 have $CP(v) = 0.5$ and all other vertices (those in the core) have $CP(v) = 1$.

An empirical example is provided by the Zachary Karate Club data (Zachary, 1977). As an alternative to cliques we shall use *k*-plexes. [5] We select as a core the only 2-plex of size 6, which consists of actors {1,2,3,4,8,14}. By definition, each of these actors is adjacent to four others in that group. The 1-periphery of this core consists of actors {5,6,7,9,10,11,12,13,15,18,20,22,28,29,31,32,33,34}. That is, each of these actors is adjacent to 1 or more actors in the 2-plex. To enlarge the 2-plex to seven actors, each actor must be connected to five others. Each member of the 1-periphery has either one or two of the five edges required to place them in the 2-plex. Those with two edges present (numbers 9, 13, 18, 20 and 22) therefore have $CP = 0.4$ and the remainder have $CP = 0.2$. We note that none of the periphery are very close to becoming members of the core, which suggests a structure in which a tight inner group has distanced itself from those on the outside.

It should be noted that working with *k*-plexes introduces an additional complexity that is not found with cliques. Consider, for example, the graph in Fig. 2. The set {1,2,3,4,5} is a 2-plex, as each node is connected to at least $5 - 2 = 3$ others in the set. To add node 6 to the set (bringing the set to six members) would require adding two ties from node 6 to the set, so that it would have $6 - 2 = 4$ ties to other group members. But the ties have to be carefully chosen as otherwise the resulting group will not be a 2-plex. In particular, the additional ties must be to nodes 5 and 3, since otherwise each of them will be connected to only three members of the group, which would mean the group was no longer a 2-plex. An interesting thing occurs if node 6 was to ignore our prescription and become adjacent to nodes 4 and 3. Then a new 2-plex would form consisting of {1, 2, 3, 4, 6}, which would exclude node 5.

So far we have not taken account of ties among the peripheral nodes. The use of a cohesive subgraph to define a core means that ties among the core will normally be numerous and their density is likely to be greater than core-to-periphery, but periphery-

[4] An alternative approach would have been to define *k*-peripheries in terms of the number of ties needed to enter the core, and distance to the core to refine those bands of closeness. However, because the number of ties needed to enter a core varies with the size of the core, in practice this approach is much more difficult to understand.

[5] We assume that *k*-plexes must be maximal, although the original definition by Seidman and Foster (1978) did not include that requirement.

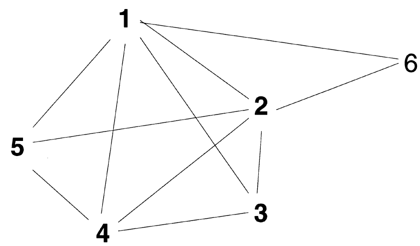

Fig. 2.

to-periphery interaction could still be quite high. We can measure this for any node by the *Peripheral Degree* index, P_D, where

$$P_D(v) = \frac{\text{Number of peripheral actors connected to } v}{\text{Total number of peripheral actors}}.$$

For the idealized network in Fig. 1, all peripheral actors have $P_D(v) = 0$. In the Zachary dataset (again choosing the 2-plex of size 6 as the core), some actors have $P_D(v) = 0$, such as actors 12 and 13, but others have much higher values. For example, actor 34 has $P_D = 0.89$. This suggests that actor 34, in contrast to 12 and 13, has rather more interest in the other peripheral actors and may not view the core as a group to get more involved with. Rather than look at the peripheral degree of individual actors, we may be interested in the density of the peripheral actors as a group. In this instance we are more concerned with the overall structural properties of the network as a whole. This is in the spirit of the partition-based models described in the next section. If C is a core with periphery P, then the peripheral density is simply the density of the subgraph induced by P. A low peripheral density for a 1-periphery could indicate an elite structure in which the periphery members only wish to have contacts with the elite core. How close peripheral individuals are getting to the elite is given by the coreness measure.

The periphery for the Zachary data has a density of 0.12, which is similar to density of the network as a whole (0.14). This suggests that members of this periphery do not actively shun one another, as one might expect in a status-based core/periphery structure. In such a structure, both members of the core and the periphery prefer to interact with the core, so we expect the density of the periphery to be especially low. Interestingly, if we remove actor 34 (whose $P_D = 0.89$) from the periphery, the peripheral density drops down to 0.07, which is only half the density of the network as a whole and which, contrary to our earlier conclusion, is suggestive of a status-based core/periphery structure. One possible explanation could be that actor 34 is part of another cohesive subgroup which is in competition for members with the identified core we have been considering.

A clique analysis of the data reveals two basic groups, the 2-plex that we have been examining, {1,2,3,4,8,14}, and a second group {9,24,30,31,33,34}. This second group contains four members of the 1-periphery of the first group. From the perspective of looking for a status-based core/periphery structure, these four do not really belong in the 1-periphery: at the very least, they are playing a rather different structural role than the other members of the 1-periphery. If we remove them and recalculate the density of the 1-periphery, we find it drops down to 0.05. It should be noted that the justification for removing these nodes is not to decrease this density but that being a member of a different core is simply not consistent with the status-based core/periphery notion we are trying to explore.

3. An alternative approach

One of the problems with the 1-periphery is that it can be quite large. The 2-plex core of the Zachary data consisted of five actors but had a 1-periphery consisting of 18 actors. We can use the values of *CP* and P_D to restrict the size of the periphery, but these values do not give any insight or restriction on the structure of the periphery (i.e., the ties among peripheral members). An alternative approach is to look at the generalizations of cliques that contain parameters that can be adjusted so as to relax the conditions of membership. Such generalizations produce hierarchies of clique-like structures, and these can be viewed as cores and peripheries. For example we can take a 1-clique as a core and define the periphery as members of the inclusive 2-clique which are not in the original clique.

As an example, consider the graph in Fig. 3. The single 1-clique is {c,d,e}. The two 2-cliques that include {c,d,e} are {b,c,d,e} and {c,d,e,f}. Therefore, we consider the 2-clique periphery to be {b,f}, which are the two nodes in the inclusive 2-cliques that are not in the 1-clique.

As a practical example, consider the Taro data reported in Hage and Harary (1983). This dataset looks at the relation of gift-giving among 22 households in a Papuan village. Table 1 lists the 2-plexes with four or more actors. We shall use these as the cores. Each core is simply considered in turn and is the focus of an individual analysis. For each core we find the associated 4-plexes (i.e., the 4-plexes which contains the 2-plex as a subset) and list the vertices contained in the 4-plexes which are not in the 2-plex. This we label as the 4-plex periphery. The third column gives the 1-periphery of the core as defined previously.

We can see immediately by looking at the second two columns that this method has the desired effect of reducing the size of the corresponding peripheries. This is principally because we are only identifying those actors that are close to becoming core members, whereas the 1-periphery includes highly peripheral actors. Note that the elements in the 4-plex periphery could not have been identified by using the *CP* or P_D

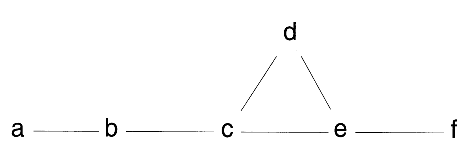

Fig. 3.

Table 1
Taro cores and peripheries

2-plexes	4-plex-periphery	1-periphery
1 2 3 17	18 22	4 5 16 18 22
4 5 6 7	16 19 20 21	1 3 8 16 19 20 21
5 11 20 21	12 18	1 4 6 10 12 18
12 13 14 15	16 19	9 11 16 19

values as filters. This is because these measures are calculated independently for each actor. In contrast, the technique discussed here takes account of combinations of actors which together satisfy the conditions of being members of the periphery but which separately would not.

Related to the approach we have just discussed is the idea of *k*-cores first proposed for use in social network research by Seidman (1983). A *k*-core is a connected maximal induced subgraph which has minimum degree greater than or equal to *k*. Loosely speaking, these are regions of the network that contain pockets of densely connected actors. They may not be cohesive themselves, but any cohesive structures within the network must be contained within them. Seidman describes them as seedbeds for cliques or other cohesive structures. Having first identified the *k*-cores, it would be possible to remove the cliques or clique-like structures from them to leave a periphery. This is applying the same technique as we have just described but instead of using a more relaxed version of the cohesive structure we are using the *k*-cores. Since *k*-cores do not overlap, we simply find the largest value of *k* in which our cohesive subgraph is contained. (If we wished to extend our periphery, we could simply take smaller values of *k*.)

For an empirical example we return to the 2-plex of size 6 found in the Zachary data. The largest value of *k* that has a *k*-core containing the 2-plex is 4, the 4-core is {1,2,3,4,8,9,14,31,33,34} which yields a periphery consisting of {9,31,33,34}. This immediately demonstrates the problem of using *k*-cores, since this group is actually another clique that is located close to the core 2-plex. This problem also occurred when we examined the *CP* and P_D values in the 1-periphery. The 3-core periphery consists of {5, 6, 7, 9, 11, 20, <u>24</u>, <u>25</u>, <u>26</u>, 28, 29, <u>30</u>, <u>31</u>, <u>32</u>, <u>33</u>, <u>34</u>}. All the underlined actors are members of cliques contained wholly within the 3-core periphery. If we consider removing all 2-plexes with four or more vertices contained completely within the periphery (since we used this as the defining property of the core), we end up with a periphery consisting of just 28 and 29. Clearly the *k*-core model is useful in identifying peripheries, but care must be taken as the *k*-cores contain both cores and peripheries and it may be difficult to separate out individual actors into appropriate groups.

4. Completing the picture

An individual actor can be in more than one core and more than one periphery. Peripheral actors who belong to more than one periphery may be well placed to exploit

Table 2
CP matrix for cliques of Taro data

Actor	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
R1	0.66	1.00	0.33	0.33	0.00	0.33	0.00	0.00	0.00	0.00
R2	1.00	1.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R3	1.00	0.66	0.33	0.33	0.33	0.00	0.00	0.00	0.00	0.00
R4	0.33	0.00	0.00	1.00	1.00	0.33	0.00	0.00	0.00	0.00
R5	0.00	0.33	0.00	1.00	0.66	1.00	0.00	0.66	0.00	0.00
R6	0.00	0.00	0.00	1.00	1.00	0.33	0.00	0.00	0.00	0.00
R7	0.00	0.00	0.00	0.66	1.00	0.00	0.33	0.00	0.00	0.00
R8	0.00	0.00	0.00	0.00	0.33	0.00	1.00	0.00	0.00	0.00
R9	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.33	0.00
R10	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.33	0.00	0.00
R11	0.00	0.00	0.33	0.00	0.00	0.66	0.33	1.00	0.33	0.33
R12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	1.00	1.00
R13	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	1.00	0.66
R14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00
R15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.66	1.00
R16	0.33	0.33	0.33	0.00	0.33	0.00	0.00	0.00	0.00	0.33
R17	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
R18	0.33	0.33	1.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00
R19	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.33
R20	0.00	0.00	0.00	0.33	0.00	1.00	0.00	1.00	0.00	0.00
R21	0.00	0.00	0.00	0.33	0.00	1.00	0.00	1.00	0.00	0.00
R22	0.33	0.33	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

their position to advantage, in the same way that actors who are between other actors have a high betweenness centrality score. [6] Clearly cores that overlap sufficiently could be merged into single cores. These observations suggest that a 2-mode analysis of the data based upon membership in the cores and peripheries may help provide a better picture of the overall pattern of connections in a social network.

We therefore define the *CP matrix*: a 2-mode matrix in which the rows are actors, the columns are cores, and the cells indicate the relationship (the coreness) of the row actor with respect to the column core. For example, a clique analysis of the Taro data yields 10 cliques each containing three actors. Since each clique contains exactly three actors the *CP* scores for each actor with relation to each clique are either 0, 0.33, 0.66 or 1.0, as shown in Table 2. The *CP* matrix was then submitted to the correspondence analysis procedure in UCINET 5 for Windows (Borgatti et al., 1999). The resulting plot is given in Fig. 4.

The points labeled with an R correspond to the actors, the points labeled with a C correspond to the core/periphery structures. Three or four basic groups are visible, suggesting that some of the core/periphery structures are essentially the same. The map clearly indicates which actors are associated with which aggregated core/periphery complexes.

[6] We are grateful to an anonymous referee for suggesting this line of reasoning.

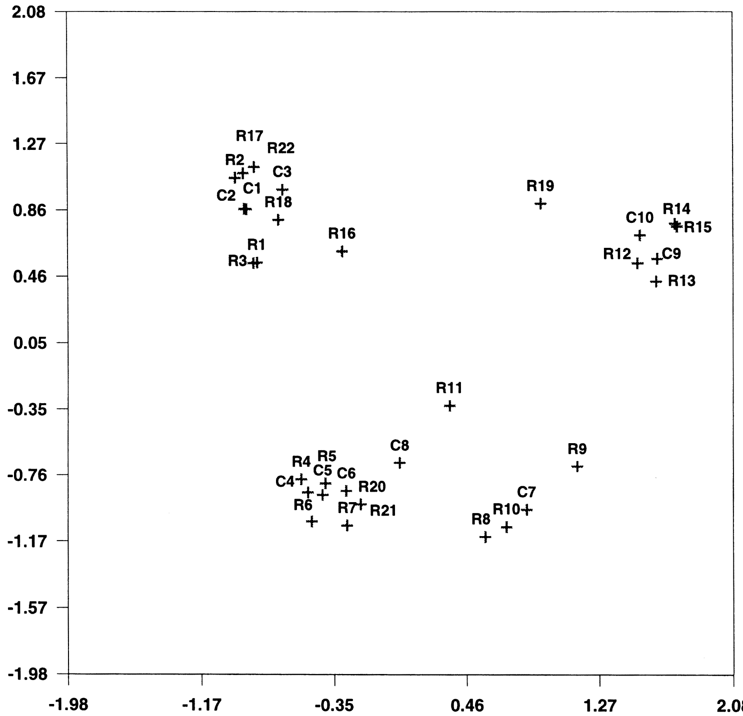

Fig. 4. Correspondence analysis of Taro *CP* matrix.

The results are roughly consistent with a routine cluster analysis of the clique co-membership matrix. Such a cluster analysis finds four basic groups, broken out as follows: (1,2,3,17,18,22), (4,5,6,7,11,20,21), (8,9,10) and (12,13,14,15). Note that two actors, 16 and 19, belong to no cliques and are not classified by the cluster analysis. In constrast, the correspondence analysis of the *CP* matrix clearly shows which groups they are primarily associated with, and also provides clues about their secondary attachments. For example, actor 16 belongs to the periphery of core/periphery structures 5 (lower left) and 10 (upper right) and so is pulled towards toward the center. Actor 19 is attached to peripheries in the top left group and is visibly drawn in that direction.

The most striking feature of Fig. 4 is the position of actor 11, which the correspondence analysis places in the center of the space, but which is not singled out in any way by the cluster analysis. This actor has connections into all four groups and so is very central to the whole network structure. A betweenness centrality analysis identifies actor 11 as the most central but not clearly distinguishable from other actors, such as actor 7 and actor 17, which have fairly similar scores. What the correspondence analysis of the *CP* matrix reveals is that the centrality of these latter actors is derived from their ties with actors belonging to the same larger groupings, whereas 11 is strategically placed between the aggregate groups. This is an important structural feature that the traditional methods of centrality analysis and clique analysis do not detect.

5. Directed graphs

Although we have developed these methods with undirected graphs in mind, the concepts can be extended fairly easily to the case of directed graphs. First, we can think of a periphery as the union of two peripheries: an out-periphery and an in-periphery. The out-periphery consists of all the actors that receive arcs from the core while the in-periphery consists of all the actors that send arcs to the core. We can similarly define a k-in-periphery and a k-out-periphery in the obvious way. And, of course, if we are particularly interested in status-based core/periphery structures, we would probably want to pay the most attention to the k-in-periphery.

With directed data we can also define the concept of a *strong periphery*, which is the intersection of the in-periphery and out-periphery. Similarly, concepts such as peripheral density can be extended to in-periphery density and out-periphery density.

6. Conclusion

We have explored ways of defining the peripheries of cohesive subsets. Depending on the parameters chosen, the peripheries can range from small, exclusive sets closely tied to just one subgroup to large inclusive sets that include every node in the network that is not in the cohesive subset. We have also discussed attributes of peripheries, such as their density. For example, we have speculated that when the density of a periphery is much lower than the background density of the network as a whole, it could be indicative of a status-based core/periphery structure in which all actors seek ties with core members and avoid ties with periphery members. We have then shown how this information can be combined to provide a deeper analysis of the pattern of connections in the network as a whole. This technique provides the analyst with extra tools which help them understand the complex structures concealed within network data.

Implicit in this paper is the notion that once we identify a cohesive subgroup, we induce a partition of all nodes in the network into three classes: the members of the subgroup, the periphery ‘‘belonging to’’ that subgroup, and the rest of the nodes in the network. This contrasts with most of the literature on cohesive subgroups, which usually divides nodes into two classes: ingroup and outgroup. We believe that thinking in terms of the tripartite division may make it easier to design algorithms to detect cohesive subgroups, at least when the procedure for constructing the periphery is independent of the procedure for constructing the subgroup, as we present in the earlier part of this paper. Otherwise, the algorithm for detecting cohesive subgroups must trade-off cohesiveness or other key criteria with coverage (i.e., obtaining just a few but fairly large subgroups).

References

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

Borgatti, S.P., Everett, M.G., 1999. Models of core/periphery structures. Social Networks, submitted for publication.

Borgatti, S.P., Everett, M.G., Freeman, L.C., 1999. UCINET 5 for Windows: Software for Social Network Analysis. Analytic Technologies, Natick.

Burt, R.S., 1992. Structural Holes, Harvard University Press, Cambridge.

Hage, P., Harary, F., 1983. Structural Models in Anthropology. Cambridge Univ. Press, Cambridge.

Johnson, S., 1967. Hierarchical clustering schemes. Psychometrika 38, 241–254.

Luce, R.D., 1950. Connectivity and generalized cliques in a sociometric group structure. Psychometrika 15, 159–190.

Luce, R.D., Perry, A.D., 1949. A method of matrix analysis of a group structure. Psychometrika 14, 95–116.

Mokken, R., 1979. Cliques, clubs and clans. Quality and Quantity 13, 161–173.

Scott, J., 1991. Social Network Analysis: A Handbook. Sage Publications, Newbury Park.

Seidman, S., 1983. Internal cohesion of LS sets in graphs. Social Networks 5, 97–107.

Seidman, S., Foster, B., 1978. A graph-theoretic generalization of the clique concept. Journal of Mathematical Sociology 6, 139–154.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge Univ. Press, Cambridge.

Zachary, W., 1977. An information flow model for conflict and fission in small groups. Journal of Anthropological Research 33, 452–473.