Cohesive Subgroups and Core-Periphery Structure

Xiao

Hyo-shin
Cohesive Subgroups:

- Theoretical motivation
- Assess
- Directional relations and valued relations
- Interpretation of Cohesive subgroups
- Alternative (multidimensional scaling and factor analysis)
Wasserman and Faust

• Definition:

“Cohesive subgroups are subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties.”

Pp249
Wasserman and Faust

- Motivations:
  - direct and indirect positive relationships and the emergence of consensus
  - Clique, Homogeneity
  - Social forces operation through direct contact among subgroup members, through indirect conduct transmitted via intermediaries, or through the relative cohesion within as compared to outside the subgroup.
Wasserman and Faust

• Types of subgroups
  
  - The concentration of ties within the subgroup
  
  - Comparison of strength or frequency of ties within the subgroup to the strength or frequency of ties outside the subgroup.
Wasserman and Faust

- The concentration of ties within the subgroup
  - subgroups based on complete mutuality
  - subgroups based on reachability and diameter
  - subgroups based on nodal degree
Wasserman and Faust

- subgroups based on complete mutuality
  - A clique
  - Consideration about clique:
    . The definition is very strict
    . Over lapping, not quite informative
Wasserman and Faust

- subgroups based on reachability and diameter
  - n-cliques
  - n-clans
  - n-clubs
Wasserman and Faust

• N-clique

Figure 6.11  n-cliques of size 4
Wasserman and Faust

• N-clique – Considerations:

- 2 is often a useful cutoff value, 2-cliques are subgraphs in which all members need not be adjacent, but all members are reachable through at most one intermediary.

- Considerations: Because it does not require that the paths connecting the nodes are all within the subgroup, there are problems: 1) a subgraph may have a diameter greater than n, 2) an n-clique might be disconnected.
Wasserman and Faust

- N-cliques-consideration:

Figure 6.12  Sub-graphs and 2-cliques
Wasserman and Faust

• N-clans:

An n-clan is an n-clique in which the geodesic distance, $d(i,j)$, between all nodes in the subgraph is no greater than n for paths within the subgraph.
An $n$-club is defined as a maximal subgraph of diameter. That is, an $n$-club is a subgraph in which the distance between all nodes within the subgraph is less than or equal to $n$; further, no nodes can be added that also have geodesic distance $n$ or less from all members of the subgraph.
Wasserman and Faust

• The social theory behind the reachability definitions:

1) Erickson (1988) The cohesive subgroup definitions based on reachability are important for understanding “processes that operate through intermediaries, such as the diffusion of clear cut and widely salient information”.

2) Hubbell (1965) ties between actors are “channels for the transmission of influence”. Pp263
Wasserman and Faust

- subgroups based on nodal degree
  - in another word, these definitions require that all subgroup members be adjacent to some minimum number of other subgroup members
Wasserman and Faust

• Motivations of subgroups based on nodal degree

1) “understanding processes that operate primarily through direct contacts among subgroup members.”

2) “vulnerability” of n-cliques.

Pp264
Wasserman and Faust

- **K-plexes**: a k-plex is a maximal subgraph containing $g_s$ nodes in which each node is adjacent to no fewer than $g_s - k$ nodes in the subgraph.
- In other words, each node in the subgraph may be lacking ties to no more than $k$ subgraph members. (pp265).
Wasserman and Faust

• An important property of a k-plex is that the diameter of a k-plex is constrained by the value of a k.

• Consideration of k-plexes: researcher should restrict the size of a k-plex so that it is not too small relative to the number of ties that are allowed to be missing. Pp266
Wasserman and Faust

- **K-cores**: k-core is a subgraph in which each node is adjacent to at least a minimum number, k, of the other nodes in the subgraph.

Figure 6.6  A 3k-core
Wasserman and Faust

• a comparison of strength or frequency of ties within the subgroup to the strength or frequency of ties outside the subgroup.

- An ideal subgroup: complete component (or strong alliance by Freeman): consists of ties between all pairs of members within the subgroup, and no ties from subgroup members to actors not in the subgroup.
Wasserman and Faust

• a comparison of strength or frequency of ties within the subgroup to the strength or frequency of ties outside the subgroup.

- LS sets:
- Lambda Sets:
LS sets

We will define a subset of nodes taken from Ns as L, so that $L \subset Ns$. The set of nodes, Ns, is an LS set if any proper subset $L \subset Ns$ has more lines to the nodes in $Ns - L$ (other nodes in the subset) than to $N-Ns$ (nodes outside the subset). Pp269
Wasserman and Faust

• Summary of LS sets:
  1) relatively robust, and do not contain “splinter” groups
  2) in a given graph, any two LS sets either are disjoint or one LS set contains the other.
  3) within a graph there is a hierarchical series of LS sets.
Wasserman and Faust

• Lambda Sets

- The idea is that a cohesive subset should be relatively robust in terms of its *connectivity*.

- The line connectivity of nodes $i$ and $j$, denoted $\lambda(i,j)$, is equal to the minimum number of lines that must be removed from the graph in order to leave no path between the two nodes.

- **Lambda Sets**: A lambda set is a subset of nodes, $Ns \subset N$, such that for all $i,j,k \in Ns$, and $l \in N-Ns$, $\lambda(i,j) > \lambda(k,l)$
**Summary of Lambda Sets:**

1) no overlapping only containing

2) more general than LS sets. Any LS set in a network will be contained within a lambda set.

3) A given network is more likely to contain lambda sets than it is to contain LS sets. Pp270

4) no need for members of a lambda set to be adjacent, they may be quite distant from one another in the graph.
Wasserman and Faust

• How to assess the cohesiveness of network subgroups

the ratio of the strength of ties within the subgroup to ties between subgroups does not decrease appreciably with the addition of new members.
• How to assess the cohesiveness of network subgroups

\[
\frac{\sum_{i \in N_s} \sum_{j \in N_s} x_{ij}}{\sum_{i \in N_s} \sum_{j \notin N_s} x_{ij}} \frac{g_s \left(g_s - 1\right)}{g_s \left(g - g_s\right)}
\]

The numerator of this ratio is the average strength of ties within the subgroup and the denominator is the average strength of the ties that are from subgroup members to outsiders. The ratio =1, >1 or <1.
The probability of observing $q$ or more lines in a subgraph of size $g_s$ form a graph with $L$ lines is:

$$P(L_s \geq q) = \sum_{k=q}^{\min(L, \frac{g_s(g_s-1)}{2})} \binom{L}{k} \left( \frac{g(g-1)}{2} - L \right) \left( \frac{g_s(g_s-1)}{2} - K \right)$$

If the probability in the equation is small, then the observed frequency of lines within the subgraph is greater than would be expected by chance, given the frequency of lines in the graph as a whole. Thus, this probability can be interpreted as a p-value for the null hypothesis that there is no difference between the density of the subgraph and the density of the graph as a whole.
Wasserman and Faust

- **Cliques based on reciprocated ties**

  Redefine the directional graph:

  \[ x_{ij}^{\text{min}} = x_{ji}^{\text{min}} = \begin{cases} 
  1 & \text{if } x_{ij} = x_{ji} = 1, \\
  0 & \text{otherwise},
\end{cases} \]

The relation \( x_{ij}^{\text{min}} \) can then be analyzed using methods for finding cliques or other cohesive subgroups in a nondirectional relation.
• N-cliques in directional relations

  - Connectivity in directional relations:

    Four types of connectivity: A pair of nodes, i, j, is:

(i) Weakly n-connected if they are joined by a semipath of length n or less.

(ii) Unilaterally n-connected if they are joined by a path of length n or less from i to j, or a path of length n or less from j to i.

(iii) strongly n-connected if there is a path of length n or less from i to j, and a path of length n or less from j to I; the path from i to j may contain different nodes and arcs than the path from j to i.

(iv) Recursively n-connected if they are strongly n-connected, and the path from i to j uses the same nodes and arcs as the path from j to I, in reverse order.
Wasserman and Faust

- Four different kinds of cohesive subgroups based on the four types of connectivity. Pp267

- A weakly connected n-clique is a subgraph in which all nodes are weakly n-connected, and there are no additional nodes that are also weakly n-connected, to all nodes in the subgraph.

- A unilaterally connected n-clique is a subgraph in which all nodes are unilaterally n-connected, and there are no additional nodes that are also unilaterally n-connected to all nodes in the subgraph.

- A strongly connected n-clique is a subgraph in which all nodes are strongly n-connected, and there are no additional nodes that are also strongly n-connected to all nodes in the subgraph.

- A recursively connected n-clique is a subgraph in which all nodes are recursively n-connected, and there are no additional nodes that are also recursively n-connected to all nodes in the subgraph.
Wasserman and Faust

- Four different kinds of cohesive subgroups based on the four types of connectivity. Pp267

  Note: Finding weakly connected n-cliques and recursively connected n-cliques requires symmetrizing the relation using the appropriate rule, and then using a standard n-clique algorithm.
• Valued Relations

**Tie strength:**

• Clique at level $c$: a subgraph with actor set is a clique at level $c$ if for all actors, and there is no actor $k$, such that for all (Peay 1975a)

• A path of level $c$: A path at level $c$ in a valued graph is a path in which all lines have values of $c$ or greater.
Wasserman and Faust

• Finding Cliques, n-cliques, and k-plexes:

Define $\chi_{ij}^{(c)} = \begin{cases} 1 & \text{if } x_{ij} \geq c \\ 0 & \text{otherwise} \end{cases}$
Wasserman and Faust

• Interpretation of Cohesive subgroups

(i) Individual actor

Members and non-members of cohesive subset. What are the differences between members and non-members. And non-members can also occupy critical locations between groups.

(ii) Subset of actors

If the network data set contains information on attributes of the actors, then one can use these attributes to describe the subsets.

(iii) Whole group

The numbers of actors in the subgroups and the degree to which these subgroups overlap can be used to describe the structure of the network as a whole. Pp284
Wasserman and Faust

- Alternative approaches for studying cohesiveness in networks
  - Matrix permutation approach
  - Multidimensional scaling
  - Factor analysis
The idea is that if actors who “choose” each other occupy rows and columns that are close in the sociomatrix, then ties that are present will be concentrated on the main diagonal of the sociomatrix, and ties that are absent will be concentrated far from the main diagonal of the sociomatrix.

For an entire matrix a summary measure of how close large values of are to the main diagonal is given by:

$$\sum_{i=1}^{g} \sum_{j=1}^{g} x_{ij} (i - j)^2 \text{ for } i \neq j$$
Wasserman and Faust

• Matrix permutation approach

- A matrix permutation analysis does not locate discrete subgroups.

- It can be quite informative to present the sociomatrix with rows and columns permuted to suggest the subgroup structure.
Multidimensional scaling

Fore revealing which actors are “close” to each other, and for presenting possible cleavages between subgroups.

Multidimensional scaling: 1) It is a very general data analysis technique. 2) It seeks to represent proximities among a set of entities in low-dimensional space so that entities that are more proximate to each other in the input data are closer in the space, and entities that are less proximate to each other are farther apart in the space.
Wasserman and Faust

• Factor analysis
  - Direct factor analysis
  - Factor analysis of a correlation or covariance matrix derived from the rows (or columns) of a sociomatrix

- Core/periphery structure

- Formalize the intuitive notion of core/periphery structure

- Suggests algorithms for detecting the structure, and statistical tests for testing a priori hypotheses

• Intuitive concept:
• A group or network cannot be subdivided into exclusive cohesive subgroups or factions, although some actors may be much better connected than others. The network consists of just one group to which all actors belong to a greater or lesser extent.

• Intuitive concept
• In the terminology of blockmodeling, the core is seen as a 1-block, and the periphery is seen as a 0-block. This is the sense in which Breiger_1981. uses the terms. The blocks representing ties between the core and periphery can be either 1-blocks or 0-blocks.

- Intuitive concept:
- Given a map of the space, such as provided by multidimensional scaling, nodes that occur near the center of the picture are those that are proximate not only to each other but to all nodes in the network, while nodes that are on the outskirts are relatively close only to the center.

Fig. 1. A network with a core/periphery structure.

• Two Models:
  - Discrete Model
  - Continuous model

- Discrete Model

The core/periphery model consists of two classes of nodes, a cohesive subgraph, and a class of actors that are more loosely connected to the cohesive subgraph but lack any maximal cohesion with the core

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|  1  |  2  |  3  |  4  |  5  |  6  |  7  |  8  |  9  | 10  |
|-----+-----+-----+-----+-----+-----+-----+-----+-----+-----|
| 1   | 1 1  | 1 1  | 1 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  |
| 2   | 1 1  | 1 1  | 0 1  | 1 1  | 1 0  | 0 0  | 0 0  | 0 0  | 0 0  |
| 3   | 1 1  | 1 1  | 0 0  | 0 0  | 1 1  | 0 0  | 0 0  | 0 0  | 0 0  |
| 4   | 1 1  | 1 1  | 1 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  |
| 5   | 1 0  | 0 1  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  |
| 6   | 0 1  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  |
| 7   | 0 1  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  |
| 8   | 0 1  | 1 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  |
| 9   | 0 0  | 1 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  |
|10   | 0 0  | 0 1  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  | 0 0  |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
```

- Discrete Model

An idealized version that corresponds to a core/periphery structure of the adjacency matrix:

```
  1  2  3  4  5  6  7  8  9  10
1  1  1  1  1  1  1  1  1  1  1
2  1  1  1  1  1  1  1  1  1  1
3  1  1  1  1  1  1  1  1  1  1
4  1  1  1  1  1  1  1  1  1  1
5  1  1  1  1  1  1  1  1  1  1
6  1  1  1  1  1  1  1  1  1  1
7  1  1  1  1  1  1  1  1  1  1
8  1  1  1  1  1  1  1  1  1  1
9  1  1  1  1  1  1  1  1  1  1
10 1  1  1  1  1  1  1  1  1  1
```

- Discrete Model

An idealized version that corresponds to a core/periphery structure of the adjacency matrix:

Fig. 2. Freeman’s star.

Fig. 3. Core/periphery structure.

How well the real structure approximates the ideal:

\[ \rho = \sum_{i,j} a_{ij} \delta_{ij} \quad (1) \]

\[ \delta_{ij} = \begin{cases} 1 & \text{if } c_i = \text{CORE or } c_j = \text{CORE} \\ 0 & \text{otherwise} \end{cases} \quad (2) \]

- Testing a priori partitions:
  A troop of monkeys, hypothesis male core/female periphery:
• Testing a priori partitions:

A troop of monkeys, hypothesis male core/female periphery:

The correlation between these two matrices is 0.26 which according to the QAP permutation test is not significant (p>0.1)

QAP test is a permutation test for the independence of two proximity matrices.

- Detecting core/periphery structures in data

- Additional pattern matrices

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- Additional pattern matrices

\[
\delta_{ij} = \begin{cases} 
1 & \text{if } c_i = \text{CORE} \text{ and } c_j = \text{CORE} \\
0 & \text{if } c_i = \text{PERIPHERY} \text{ and } c_j = \text{PERIPHERY} \\
\text{otherwise} & \end{cases}
\]

- **Additional pattern matrices**
  - Can handle asymmetric data
  - Handle valued data

Asymmetric core/periphery model

Correlation: 0.826
• Continuous model
  - In continuous model, each node is assigned a “coreness”
  - Coreness: would correspond to distance from the centroid of a single point cloud.

\[ \delta_{ij} = c_i c_j \]
• Estimating coreness empirically

The objective is to obtain values of C so as to maximize the correlation between the data matrix and the pattern matrix associated with

\[ \delta_{ij} = c_i c_j \]
• Coreness and centerality
- All coreness measures are centrality measures, but the converse is not necessarily true.
- Research subject: State supreme courts network
- Research objective: Identify, describe and explain the bases of discrete networks among the state supreme courts.
- Research question: Do interpretable blocks of state supreme courts emerge? If so, what binds these sets of appellate courts? Do constellations of leaders and followers develop? Do networks go beyond particular regions?

- Research technique: clustering techniques, discriminante analysis
- Data: the data come from a count of all references each state supreme court as well as the Court of Appeals for the District of Columbia made to every other state court of last resort during 1975.

- Definitions:
  - Structure equivalence:
  - Discriminant analysis

- Findings
  - Social cohesion: Which courts tend to use each others’ legal precedents in a reciprocal fashion
  - SOCK software

Fig. 1. Dendrogram showing hierarchical clustering of cliques form SOCK.
Fig. 2. Multidimensional plot of distances among the state supreme courts.

What forces bring the state supreme courts into distinct sets?

Table 3
Discriminant functions for cliques in a social network

<table>
<thead>
<tr>
<th>Variable</th>
<th>Standardized coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>Legal professionalism</td>
<td>-0.59</td>
</tr>
<tr>
<td>Caseload</td>
<td>1.02</td>
</tr>
<tr>
<td>Innovation</td>
<td>0.52</td>
</tr>
<tr>
<td>Population</td>
<td>0.09</td>
</tr>
<tr>
<td>Wealth</td>
<td>1.05</td>
</tr>
<tr>
<td>Prestige</td>
<td>-0.65</td>
</tr>
<tr>
<td>Progressivism</td>
<td>0.42</td>
</tr>
<tr>
<td>Legal capital</td>
<td>0.64</td>
</tr>
<tr>
<td>Cumulative percentage of variance explained</td>
<td>44.3</td>
</tr>
<tr>
<td>Proportion classified correctly</td>
<td></td>
</tr>
</tbody>
</table>

- Structural equivalence

Fig. 3. Dendrogram showing hierarchical clustering of distances from STRUCTURE.

Fig. 4. Map of structurally equivalent positions in the communication of legal precedent.
Discriminant functions for seven structurally equivalent positions

<table>
<thead>
<tr>
<th></th>
<th>Standardized coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>Legal professionalism</td>
<td>0.39</td>
</tr>
<tr>
<td>Caseload</td>
<td>0.30</td>
</tr>
<tr>
<td>Innovation</td>
<td>0.22</td>
</tr>
<tr>
<td>Population</td>
<td>0.05</td>
</tr>
<tr>
<td>Wealth</td>
<td>0.33</td>
</tr>
<tr>
<td>Prestige</td>
<td>0.49</td>
</tr>
<tr>
<td>Progressivism</td>
<td>-0.01</td>
</tr>
<tr>
<td>Legal capital</td>
<td>0.20</td>
</tr>
<tr>
<td>Cumulative percentage of variance explained</td>
<td>67.6</td>
</tr>
<tr>
<td>Proportion classified correctly</td>
<td>68.1</td>
</tr>
</tbody>
</table>
Table 5
Densities and images for seven structurally equivalent positions

<table>
<thead>
<tr>
<th></th>
<th>Social position</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td>8.2</td>
<td>2.6</td>
<td>2.2</td>
<td>1.0</td>
<td>1.1</td>
<td>0.6</td>
<td>1.1</td>
<td>2.0</td>
</tr>
<tr>
<td>II</td>
<td></td>
<td>5.5</td>
<td>3.4</td>
<td>1.6</td>
<td>0.9</td>
<td>0.9</td>
<td>1.2</td>
<td>1.4</td>
<td>2.0</td>
</tr>
<tr>
<td>III</td>
<td></td>
<td>6.2</td>
<td>2.9</td>
<td>2.4</td>
<td>0.9</td>
<td>1.2</td>
<td>0.8</td>
<td>1.1</td>
<td>2.0</td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td>7.1</td>
<td>2.1</td>
<td>2.7</td>
<td>2.7</td>
<td>0.8</td>
<td>0.9</td>
<td>0.9</td>
<td>2.0</td>
</tr>
<tr>
<td>V</td>
<td></td>
<td>4.4</td>
<td>3.1</td>
<td>2.1</td>
<td>0.6</td>
<td>1.7</td>
<td>1.4</td>
<td>1.1</td>
<td>2.0</td>
</tr>
<tr>
<td>VI</td>
<td></td>
<td>6.1</td>
<td>3.2</td>
<td>1.2</td>
<td>0.5</td>
<td>0.7</td>
<td>1.5</td>
<td>1.9</td>
<td>2.0</td>
</tr>
<tr>
<td>VII</td>
<td></td>
<td>4.5</td>
<td>3.5</td>
<td>1.3</td>
<td>0.8</td>
<td>0.9</td>
<td>1.5</td>
<td>1.7</td>
<td>2.0</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>5.5</td>
<td>3.1</td>
<td>1.8</td>
<td>0.9</td>
<td>1.1</td>
<td>1.2</td>
<td>1.4</td>
<td>2.0</td>
</tr>
</tbody>
</table>

B. Image matrix

<table>
<thead>
<tr>
<th></th>
<th>Block (a = 2.0)</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>III</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IV</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>V</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VI</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VII</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Definition of Peripheries

• Two extremes of a continuum of possible definitions of peripheries

A: Periphery as actors clearly associated with the core
B: Periphery as simply all outsiders

• A general definition of periphery should be able to encompass all of these possibilities.
First Definition

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

- The measure of distance between a node 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. (ex. shortest distance (nearest neighbor), maximum distance (farthest neighbor), average distance, medium distance, and so on.

- It is not necessary that distance to the core should be defined in terms of graph-theoretic distance at all. For instance, a metric based on the extent to which the node is structurally equivalent to the members of the core would do just as well.
Second Definition

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 $CP(v) = \frac{r}{q}$. 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.
- While K-peripheries are defined in terms of distance, the $CP$ measure is defined in terms of volume of ties.
Peripheral Degree Index

- $P_D$: Peripheral Degree Index meaning the density of periphery-to-periphery interaction

\[
P_D(u) = \frac{\text{Number of peripheral actors connected to } u}{\text{Total number of peripheral actors}}
\]

- What does higher Peripheral Degree mean?
An Example: Figure 3

- Definition 1
- Definition 2
- Peripheral Degree Index
- Definition 1: Consider the network in Figure 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.

- Definition 2: Given that we are using cliques as the basis for identifying the core in Figure 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 Figure 1 the vertices with degree 1 \((6,7,9,10)\) have \(CP(v) = 0.25\), the vertices with degree 2 \((5,8)\) have \(CP(v) = 0.5\) and all other vertices (those in the core) have \(CP(v) = 1\).

- 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.

- In Figure 1, all peripheral actors have \(P_D(v) = 0\). If some peripheral actors have higher values than others in a given dataset, they may not view the core as a group to get more involved with. Concerning the density of the peripheral actors as a group, 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.
Two Alternative Approaches

- One of the problems with the 1-periphery is that it can be quite large. 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.

- An Alternative Approach 1: The generalization of cliques that contain parameters that can be adjusted so as to relax the conditions of membership.

An Conceptual Example: Figure 4 (pp. 402)

- 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\}. To reduce the size of the corresponding peripheries, 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.
An Alternative Approach 2: The idea of $k$-cores

- Definition: a $k$-core is a connected maximal induced subgraph which has minimum degree greater than or equal to $k$. (Seidman, 1983)
- 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.

An Empirical Example: 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.
- 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.
Completing the picture

• An individual actor can be in more than one core and more than one periphery.

  ➢ 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.

  ➢ 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.

  ➢ Table 2: CP matrix for cliques of Taro data (pp.407)
<table>
<thead>
<tr>
<th>Actor</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>C9</th>
<th>C10</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>0.66</td>
<td>1.00</td>
<td>0.33</td>
<td>0.33</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R2</td>
<td>1.00</td>
<td>1.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R3</td>
<td>1.00</td>
<td>0.66</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R4</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R5</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>1.00</td>
<td>0.66</td>
<td>1.00</td>
<td>0.00</td>
<td>0.66</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R6</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>1.00</td>
<td>0.66</td>
<td>1.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R7</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R9</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.66</td>
<td>1.00</td>
<td>0.33</td>
<td>0.33</td>
<td>0.00</td>
</tr>
<tr>
<td>R12</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>R13</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>1.00</td>
<td>0.66</td>
</tr>
<tr>
<td>R14</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>R15</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.66</td>
<td>1.00</td>
</tr>
<tr>
<td>R16</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
</tr>
<tr>
<td>R17</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R18</td>
<td>0.33</td>
<td>0.33</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>R20</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R21</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>R22</td>
<td>0.33</td>
<td>0.33</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
The CP matrix was then submitted to the correspondence analysis procedure in UCINET 5 for Windows. The resulting plot is given in Figure 5 Correspondence analysis of Taro CP matrix (pp.405).
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.

• 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.

• 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.
Identifying Cohesive Subgroups

Frank (1995)
Main Topics

• This article applies recent advances in the statistical modeling of social network data to the task of identifying cohesive subgroups from social network data.

• Through simulated data, the author describes a process for obtaining the probability that a given sample of data could have been obtained from a network in which actors were no more likely to engage in interaction with subgroup members than with members of other subgroups.

• By obtaining the probability for a specific data set, through further simulations, the article develops a model which can be applied to future data sets.

• Through simulated data, the author characterizes the extent to which a simple hill-climbing algorithm recovers known subgroup memberships.

• The author applies the algorithm to data indicating the extent of professional discussion among teachers in a high school, and he shows the relationship between membership in cohesive subgroups and teachers' orientations towards teaching.
Introduction

- Cohesive subgroups have been a crucial link between individuals and organizations; individuals are most strongly influenced by the members of their primary groups; and primary groups are integral to understanding people within the contexts of their communities.
- Since actors are directly influenced by their interpersonal interactions, subgroups based on the pattern of interaction are more likely to be related to the sentiments and actions of actors than subgroups based on formal positions.
- Methodologists have developed and employed various techniques for identifying cohesive subgroups from data indicating the extent of interactions between each pair of actors.
- The author contends that probably the most promising approaches for identifying non-overlapping cohesive subgroups are those which utilize goodness of fit, or statistical criteria, associated with the fitting of subgroups to social network data.
Cohesive Subgroups and Stochastic Blockmodels

• The statistical models developed for sociometric data (or directed graphs) have been applied to the task of identifying blocks of stochastically equivalent actors.

• This approach is important because it can be applied to identification of cohesive subgroups.
This application can start by the $p_1$ model (Holland and Leinhardt, 1981). For $n$ actors and $n \times (n - 1)$ possible pairs of actors $i$ and $i'$, define:

$X_{i'i} = 1$ if actor $i$ chooses (in a sociometric sense), or initiates an exchange with, $i'$

$0$ otherwise.

$$ p_1 = P(X = x) = K \exp\{ pm + \theta x_{++} + \sum_i \alpha_i x_{i+} + \sum_{i'} \beta_{i'} x_{+i'} \} $$ (1)

$$ p_1 (x|\lambda) = K \exp\{ pm + \theta x_{++} + \sum_i \alpha_i x_{i+} + \sum_{i'} \beta_{i'} x_{+i'} + \sum_{g,h} \lambda_{gh} x_{++}(B_g \times B_h) \} $$ (2)

$$ p_1 (x|\lambda) = K \exp\{ pm + \theta x_{++} + \sum_i \alpha_i x_{i+} + \sum_{i'} \beta_{i'} x_{+i'} + \lambda_c x_{++}(B_g \times B_g) \} $$ (3)

$$ p_1 (x|\lambda) = K \exp\{ \theta x_{++} + \gamma (B_g \times B_g) + \lambda_c x_{++}(B_g \times B_g) \} $$ (4)
This log-linear model includes a main effect for the overall extent of exchange, associated with $x_{++}$, a main effect for the number of pairs contained by subgroup boundaries, associated with $(B_g \times B_g)$, and an effect based on the interaction of $x_{++}(B_g \times B_g)$. Therefore, the model can be restated at the level of the pair of actors $i$ and $i'$ in terms of the logit model (Agresti 1984):

$$\log \left( \frac{p[X_{ii'}=1|X_{ii'}^c]}{p[X_{ii'}=0|X_{ii'}^c]} \right) = \theta_0 + \theta_1 \text{samegroup}_{ii'} \quad (5)$$

where $\text{samegroup}_{ii'} = 1$ if actors $i$ and $i'$ are members of the same subgroup, 0 otherwise.

Model (5) can be equated with $p_1^*$ model by defining four new quantities. The $p_1^*$ models can be defined as

$$\log \left( \frac{P[X_{ii'}=1|X_{ii'}^c]}{P[X_{ii'}=0|X_{ii'}^c]} \right) = \Theta' \left[ z(X_{ii'}^+) - z(X_{ii'}^-) \right] \quad (6)$$
By multiplying $\Theta' [z(X_{ii'}^+) - z(X_{ii'}^-)]$, the model (6) can be re-expressed as

$$\log \left( \frac{P[X_{ii'} = 1 | X_{ii'}^c]}{P[X_{ii'} = 0 | X_{ii'}^c]} \right) = \theta_0 + \theta_1 \text{ (if } i \text{ and } i' \text{ are in the same subgroup)} \quad (7)$$

This right-hand side is precisely the same as in (5). Strauss and Ikeda (1990) show that maximizing the likelihood for model (7) is equivalent to maximizing the pseudo-likelihood function

$$PL(\theta) = \prod_{i' \neq i} \Pr(X_{ii'} = 1 | X_{ii'}^c) X_{ii'}^c \Pr(X_{ii'} = 1 | X_{ii'}^c)^{(1 - X_{ii'})} \quad (8)$$

which expresses the likelihood conditioning on $X_{ii'}^c$ and therefore does not assume independence of pairs in $X$. But one can maximize the likelihood in model (5) to obtain estimates of $\theta_0$ and $\theta_1$ without assuming independence of pairs.
When defining $G^2$ to be the likelihood ratio statistic which measures the adequacy of the fit of a given log-linear or logit model to sociometric data, one natural approach to identifying cohesive subgroups is to maximize the change in $G^2$ between a model which includes $\theta_1$ (e.g. model 5) and an identical model except for the removal of the term involving $\theta_1$, such as

$$\log\left(\frac{p[X_{ii'}=1]}{1-p[X_{ii'}=1]}\right) = \theta_0$$ \hspace{1cm} (9)

Alternatively, since the focus is on $\theta_1$, one can identify cohesive subgroups by assigning actors so as to maximize $\theta_1$. Unlike $G^2$, $\theta_1$ is invariant with respect to network size, has a range from $-\infty$ to $\infty$ and has a readily understandable interpretation.
• Typically, the subgroup memberships are not known. When defining $\Omega$ to be the lower triangular portion of an $n \times n$ matrix, with $\omega_{ii'} = 1$ if actors $i$ and $i'$ are in the same subgroup, 0 otherwise, and the restriction that if $\omega_{ij} = 1$ and $\omega_{jk} = 1$ then $\omega_{ik} = 1$.

• In effect, $\Omega$ is a block-diagonal 'target' matrix indicating subgroup membership, and $\omega_{ii'}$ is considered as entities which a clustering algorithm estimates. By estimating $\Omega$, latent subgroups can be identified so as to maximize $\hat{\theta}_1 \mid \tilde{\Omega} \cdot X$. In this sense, the subgroup memberships are 'estimated' and so is $\theta_1$, which is referred to as $\tilde{\theta}_1$, where $\tilde{\theta}_1$ is half the log-odds of the cells in Table 1 based on $X$ and $\tilde{\Omega}$. 
• The hill-climbing algorithm

- Given a criterion for evaluating the extent to which subgroups are cohesive, \( \widehat{\theta_1} \), a simple hill-climbing, or iterative partitioning, algorithm have been employed to identify cohesive subgroups.

- Everitt (1986) showed that this type of algorithm will converge on a local maximum in the criterion, and, in Appendix A, the author restates this procedure in terms of the E-M (expectation-maximization) algorithm, which would ensure convergence to maximum likelihood estimates of \( \theta_1 \) and \( \Omega \).
Testing Whether There Are Underlying Subgroup Processes: Evaluating the Internal Validity of $\hat{\theta}_1$

- Cohesive subgroups could be salient even if there is reciprocity beyond that within subgroup boundaries.
- Therefore, instead of testing whether $\theta_1$ is different from zero or whether there is reciprocity not captured by subgroup boundaries, it may be more sensible to test whether $\theta_1$ is greater than $\theta_{1base}$; where $\theta_{1base}$ is the value of $\theta_1$ associated with subgroups identified by the algorithm when applied to data which are generated without regard for subgroup memberships.
• This can be tested through the change in $G^2$ between the following two models compared to a $\chi^2$ distribution on one degree of freedom:

$$\log\left(\frac{p[X_{ii'}=1]}{1-p[X_{ii'}=1]}\right) = \theta_0 + \theta_{1\text{base}} \text{samegroup}_{ii'} \quad (10)$$

and

$$\log\left(\frac{p[X_{ii'}=1]}{1-p[X_{ii'}=1]}\right) = \theta_0 + \theta_{1\text{base}} \text{samegroup}_{ii'} + \theta_{1\text{subgroup processes}} \text{samegroup}_{ii'} \quad (11)$$

The null hypothesis is that $\theta_{1\text{subgroup processes}} = 0$

• Specific case: Simulation customized to each new data set

• General case: Predicting $\hat{\theta}_{1\text{base}}$
Performance of the Algorithm: Evaluating $\hat{\Omega}$

- The iterative partitioning algorithm proceeds linearly until it identifies a local maximum in $\hat{\theta}_1$.
- The iterative partitioning algorithm could be extended through optimal annealing procedures by introducing the possibility of random re-assignment of actors as $\hat{\theta}_1$ approaches a local maximum (Hajek 1988).
- In order to address whether or not the algorithm recovers the true subgroup memberships (the set of assignments associated with the global maximum), the author simulates sociometric data sets ($X$) based on varying values of the concentration of exchanges within subgroups ($\theta_{1\text{subgroup processes}}$), network size ($n$), maximum subgroup size ($\text{Max}(n_g)$), and maximum number of exchanges any actor initiated ($C_{\text{max}}$).
Interpretation of Subgroup Memberships

- The author has applied the methods of identifying the cohesive subgroups to a dataset of teachers.
- By going this, he has confirmed that
  - teachers engage in exchanges within identifiable subgroup boundaries
  - the boundaries are likely to be similar to those associated with the global maximum in $\hat{\theta}$. 
Table 1 Characteristics of teachers by subgroup

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Race</th>
<th>Gender</th>
<th>First subject field</th>
<th>Moral agency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subgroups = A/1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Black</td>
<td>Female</td>
<td>English</td>
<td>0.61764</td>
</tr>
<tr>
<td>7</td>
<td>Black</td>
<td>Female</td>
<td>Business</td>
<td>0.30719</td>
</tr>
<tr>
<td>9</td>
<td>Black</td>
<td>Female</td>
<td>Business</td>
<td>0.28530</td>
</tr>
<tr>
<td>1</td>
<td>Black</td>
<td>Female</td>
<td>Social Science</td>
<td>0.81606</td>
</tr>
<tr>
<td>Subgroups = B/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Black</td>
<td>Male</td>
<td>Maths</td>
<td>0.16649</td>
</tr>
<tr>
<td>14</td>
<td>White</td>
<td>Male</td>
<td>Music</td>
<td>0.17634</td>
</tr>
<tr>
<td>11</td>
<td>White</td>
<td>Male</td>
<td>Physical</td>
<td>0.21391</td>
</tr>
<tr>
<td>18</td>
<td>Black</td>
<td>Male</td>
<td>Physical</td>
<td>-0.01690</td>
</tr>
<tr>
<td>**</td>
<td>White</td>
<td>Male</td>
<td>Physical</td>
<td></td>
</tr>
<tr>
<td>Subgroups = C/3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>White</td>
<td>Female</td>
<td>Other</td>
<td>0.19880</td>
</tr>
<tr>
<td>22</td>
<td>White</td>
<td>Female</td>
<td>English</td>
<td>-0.42454</td>
</tr>
<tr>
<td>21</td>
<td>White</td>
<td>Male</td>
<td>Other</td>
<td>-0.32503</td>
</tr>
<tr>
<td>12</td>
<td>White</td>
<td>Male</td>
<td>Maths</td>
<td>0.20916</td>
</tr>
<tr>
<td>Subgroups = D/4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Other</td>
<td>Male</td>
<td>English</td>
<td>0.06467</td>
</tr>
<tr>
<td>20</td>
<td>White</td>
<td>Female</td>
<td>Foreign</td>
<td>-0.08439</td>
</tr>
<tr>
<td>6</td>
<td>White</td>
<td>Female</td>
<td>Special Education</td>
<td>0.31899</td>
</tr>
<tr>
<td>8</td>
<td>White</td>
<td>Female</td>
<td>Science</td>
<td>0.29029</td>
</tr>
<tr>
<td>23</td>
<td>Other</td>
<td>Female</td>
<td>Special Education</td>
<td>-0.51569</td>
</tr>
<tr>
<td>Subgroups = E/5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>White</td>
<td>Male</td>
<td>Social Science</td>
<td>0.58795</td>
</tr>
<tr>
<td>16</td>
<td>White</td>
<td>Male</td>
<td>Science</td>
<td>0.15222</td>
</tr>
<tr>
<td>19</td>
<td>White</td>
<td>Male</td>
<td>Maths</td>
<td>-0.02972</td>
</tr>
<tr>
<td>5</td>
<td>White</td>
<td>Male</td>
<td>Science</td>
<td>0.34734</td>
</tr>
<tr>
<td>10</td>
<td>White</td>
<td>Male</td>
<td>Maths</td>
<td>0.21541</td>
</tr>
<tr>
<td>4</td>
<td>White</td>
<td>Male</td>
<td>Other</td>
<td>0.35663</td>
</tr>
</tbody>
</table>
• In Table I, it is clearly observed that teachers are more likely to be members of a subgroup with others of the same race (a test of independence and Fisher's exact test produced $p < 0.01$).

• Teachers also are aligned according to gender (again, a test of independence and Fisher's exacts test produced $p < 0.01$).

• Moral agency indicates each teacher’s tendency toward being a moral agent.

» A Kruskal-Wallis test revealed that the probability of achieving the differences in moral agency among the subgroups by chance alone was 0.06. Using Tukey's honestly significant differences, the members of subgroup A were significantly higher on moral agency than were the members of subgroup C.

» This is reflected in the fact that, when interviewed, the teachers of Subgroup A perceived their role to be more that of 'saving' students, than did the teachers in subgroup C.
The Structure of a National Elite Network
Moore (1979)
Main Topics

- This paper addresses a long-disputed issue: the degree of integrating among political elites in the United States.
- This issue is examined through an investigation of the structure of an elite interaction network as revealed by recently developed procedures for network analysis.
- The data, taken from the American Leadership Study conducted by the Bureau of Applied Social Research in 1971 and 1972, consist of interviews with 545 leaders of major political, economic and social institutions.
- The study’s wide institutional representation, sociometric data, and focus on major issues of the early 1970s make it virtually unique for examining elite integration.
Introduction

• Ruling class and power elite theorists’ such as Mills and Domhoff find a considerable amount of integration, with various bases, in the national power structure.

• Pluralists find little integration among elites in diverse sectors. Elites are seen as fragmented rather than integrated since each is involved primarily with its own relatively narrow concerns and constituencies.

• This study assesses the extent of integration in a network of political elites in the United States. The concept of political elite integration has several dimensions, at least, including social homogeneity, value consensus and personal interaction.

• Sociometric ties are seen as fostering integration, cohesiveness and perhaps consensus within the business community. Thus, this paper has examined sociometric ties among elites for major groups at the national level.
Research Design

Data Collection
• The data used are taken from the American Leadership Study, a survey of 545 top position holders in key institutions in American society conducted in 1971-72 by the Bureau of Applied Social Research, Columbia University.
• Through personal interviews, information was gathered on respondents' policy influencing and policy making activities on major national issues; the overall completion rate for the interviews is just over 70%.
• Extensive attitude and social background data also were collected.
• The study's wide institutional representation, collection of sociometric data and focus on major issues of the time make it well-suited for evaluating elite integration.
• The sample is drawn from persons in the top positions in ten institutional sectors assumed to exercise power in American society.
• Most persons in the snowball or reputational sample were in one of the ten positional sectors, especially Congress and the media; a few were not; these include academics, White House staff, governors and mayors.
• The interpersonal network in these data, resulting from interaction related to a specific issue for each respondent, generally reflects informal discussions or day-to-day interaction on these issues.
Research Design

Data Analysis

• The analysis of these sociometric data utilizes a procedure developed by Alba (1972; 1973) which is well-suited for evaluating network integration since it identifies the more cohesive parts of networks.

• The identification of social circles among political elites in the U.S. seems an appropriate base for studying the extent of their integration.

• If social circles can be identified in a network, the nature of their memberships and their relations to each other and the rest of the network are critical for assessing the extent of the network’s integration.
Findings

Network Structure
• 32 social circles and cliques are located in the connected network from the American Leadership data.
• This connected part contains 876 persons and the remaining 65 individuals are all isolates, connected neither to each other nor to anyone in the network.
• Of the 32 circles and cliques, all but four are unified around concern with a common issue or through common sector membership.
• With one prominent exception, these are narrow groups of persons with similar issue concerns. Some are further specified by ideology.
Findings

Attributes of Central Circle Members and Nonmembers

A. Central circle members

• Central circle members can be expected to be more influential than other persons in similar positions.

• Three items are used as indicators of policy influence activities; 1) number of federal advisory committee memberships; 2) number of times an individual has testified before Congressional committees; and 3) number of memberships in major policy planning organizations such as Council on Foreign Relations, Committee for Economic Development, Business Council.

• Visibility is measured by level of communications output, an index which includes recent interviews by the press, number of magazine articles and books written, and presentation of lectures. The number of reputation nominations received from other sample members is used as an indicator of influence among other leaders.
Findings

- Because activities, influence and visibility vary considerably from one sector to another, comparisons on these items are made within a given sector. Thus comparisons of characteristics of those who are and are not members of the central circle are always made among individuals in similar elite positions.

- Overall, the most striking differences are found for business elites, with central circle members scoring significantly higher than others in this sector on all of these indicators of influence.

- More generally, in most sectors this central circle includes in its membership the more active, visible and influential persons in elite positions.
Findings

B. Social Origins

- Social origins are measured by four items: parents' socioeconomic status, status of ethno-religious origins, prestige level of high school attended, and the quality of college attended.

- High status social origins are at most a very small advantage to those in elite positions in becoming connected to a central leadership circle. Similarly, affiliations with elite private sector organizations are generally unrelated to central circle membership.

- It should be noted that social club membership is important for business leaders.
Conclusions

- No fragmentation of elites in different institutions or issue areas was found.
- On the contrary, the evidence indicates that considerable integration exists among elites in all major sectors of American society. The existence of a central elite circle facilitates communication and interaction both within that large, diverse group and between its members and those in more specialized elite circles and cliques.
- The structure of this network and especially the central circle indicate at least a potential for unity among these elite individuals, almost all of whom have a high potential for control through their incumbency in high-level positions.
- Thus the structure exists through which elites in various institutions in the U.S. could become unified in the pursuit of common interests. Yet, in another way, the structure of the elite network analyzed here does not support a critical aspect of the contention that American elites form cohesive ruling group acting in concert to further common interests.
Conclusions

• The integrated network found in this study is not based on similarities in the social origins and affiliations of its members, to say nothing of upper class origins.

• While central circle members are more influential than others in similar formal positions, they differ little from the latter in social origins or connections to major private sector organizations.

• The most reasonable interpretation of the network analyzed here is to see it as one involving day-to-day discussions of major issues that have appeared within the public arena.