The Cohesiveness of Blocks in Social Networks:

Connectivity and Conditional Density

Abstract. The social cohesion of a group, as measured by patterns of network ties, increases with the level of redundancy of interconnections of its members. We will see that the minimal number k of independent paths that connect every pair of actors in a group, the higher the cohesion. The cohesiveness of a group is also measured by the extent to which it is not disconnected by removal of 1, 2, 3,..., n actors. Menger's Theorem proves that these two measures are equivalent. Within this conceptual framework, we evaluate the family of concepts of cohesion and establish the validity of a pair of related measures: *Connectivity* – the minimum number k of its actors whose removal would not allow the group to remain connected or would reduce the group to but a single member – measures the social cohesion of a group at a general level;

Conditional Density measures cohesion on a finer scale, that of surplus density beyond that implied by connectivity.

For each k, these two measures may be combined into an aggregate measure of social cohesion, suitable for both small- and large-scale network studies. Using these measures within a new methodology of *cohesive blockin*g we test hypotheses about the consequences of cohesion for social groups and their members, and demonstrate with empirical examples the significance and theoretical relevance of network cohesion as measured by connectivity in a variety of substantively important applications in sociology.

Keywords: Graph theory, social networks, algorithmic detection, cohesive groups, social boundaries.

The Cohesiveness of Blocks in Social Networks: Node Connectivity and Conditional Density

Douglas R. White and Frank Harary[1]
SM-wh7a/f.doc
June, 2000

Submitted to *Sociological Methodology 2001*

Solidarity is a generic concept encompassing multiple ways that individuals coalesce into groups. Logically, we can distinguish three kinds of bonds that contribute to solidarity: *members to group*; *members to leaders*; and *members to members*. We can conceive of these bonds as having *forms* such as moral rules, norms, incentives, or contexts, and *contents* such as various types of relationships. Form and content, social "facts" of relationships versus norms that might govern them, and other oppositions are not removed from one another. Rather, they are performed, enacted and understood in dynamical networks of interactions, which can often be visualized as graphs.

What we call *attachment* of members to groups often involves complex interactions among psychological, dispositional, moral, normative, and contextual concerns, and are often for this reason difficult to measure and to depict as graphs. To simplify measurement, researchers often try to elicit from individuals indicators of their attachments to groups.

The same is true of what we call *adherence* or adhesion to leadership. Common research questions in this domain are: What are the attractive or charismatic qualities of leaders (or attractions *to* their followers) that create weaker or stronger many-to-one ties or commitments?

What we call *cohesion* are the many-to-many ties among individuals, as they form into clusters. One of the fundamental questions of sociology is: How and when do groups, norms, leaders and commitments emerge out of cohesive clusters? Alternatively, we can ask: How and when does the formation of groups and the emergence of leaders lead to the transformation of cohesive clusters? Because cohesion often spills over the boundaries of formal groups, dynamic reconfiguration of groups and alliances can be studied in the interplay between these two questions. Alternately, as different groups emerge and overlap, and groups interact at another level of organization, dynamic reconfigurations of cohesiveness can be studied, oscillating between different levels involving questions about top down and bottom up effects. In heterarchic systems, such as a government that derives its legitimacy from "We the people" to guarantee empowerment against intrusions at intermediate levels (Morowitz 2000:11-2), multiple relations contend for and oscillate in their salience for regulatory processes. Such oscillations include centralized systems that are hierarchically organized from upper to lower levels. Alternatives include decentralized systems that are emergent, often hierarchically, from lower to upper levels.

The *content* of ties is also important to how we view the dynamics of cohesion and group transformations. At the simplest level (Harary 1953), negative ties tend to repel and positive ties attract. Balance theory is concerned with the statics and dynamics (Harary 1961) of segregation and clustering.

Here we are principally concerned with positive dyadic bonds and the concept and measurement of cohesion as one component of social solidarity. Ours is a companion piece to Moody and White (n.d.), who begin with a preliminary and intuitive definition of cohesion as follows:

> *1. A group is cohesive to the extent that the social relations of its members hold it together.*

As in Moody and White, we explicate a sliding series of definitions that are increasingly precise and additive in the measurement and application of the concept of cohesion. We explore different facets of these definitions, and further develop the measurement concepts in the social relations approach to cohesion. We begin with a graph theoretic clarification of formal concepts, and then examine a case study of a factional dispute in a karate club where these concepts apply. The case study highlights why we need to further elaborate our formal concepts of cohesion, and it illustrates how these concepts may be used along with other concepts that engage static and dynamical aspects of cohesion, and the importance of linking the various concepts under the rubric of social solidarity.

Two intuitive aspects of the definition of cohesion can be distinguished: [2]

1.1. *A group is cohesive to the extent that the social relations of its members are resistant to the group being pulled apart.*

1.2. *A group is cohesive to the extent that the multiple social relations of its members pull it together.*

For clarity of presentation in formalizing the intuitions in definitions 1.1 and 1.2, a social relation hypothesized or assumed to contribute to cohesion is considered as a graph. Specifically, a *graph* G=(V,E) consists of a set V of n *nodes* or *vertices* and a set E of m *edges* each joining a pair of nodes. We say G has *order* n and *size* m. The two nodes in each unordered pair (u,v) in E are said to be *adjacent* and constitute an edge that is *incident* with nodes u and v.[3] The *complete graph* K_n of order n has every pair of nodes adjacent. A *path* in G is an alternating sequence of distinct nodes and edges, beginning and ending with nodes, in which each edge is incident with its preceding and following node. Two paths are *disjoint* or *independent* if they have no nodes in common other than their endnodes. A graph is *connected* if there is a path between every pair of nodes. The *distance* between two nodes in G is the minimum size of a path of G that connects them.

We regard a social group (with symmetric interpersonal relations) as part of a social network, i.e., as a subgraph of a larger graph. A *subgraph* of a graph G is a graph having all of its nodes and edges in G. A set S is *maximal* (*minimal*) with respect to some property if no proper superset (subset) of S, containing more (fewer) elements than S, has the property but S does. A *component* of G is a maximal connected subgraph. A *clique* of a graph G is a maximal complete subgraph of G of order at least 3 (a maximal subgraph

K_n of G, $n \geq 3$). Figure 1 shows a disconnected graph with two components and three cliques, each a K_3.

(Insert Figure 1 about here)

In what follows, we first define cohesion in terms of "node connectivity," compare it to "edge connectivity," and use Menger's theorems to identify the equality between its two fundamental properties: resistance to being pulled apart (1.1), and stick-togetherness (1.2). We show the sociological significance of defining the latter using the levels of redundancy of multiple disjoint paths by which actors are connected, and discuss the sociological implications of these two aspects of cohesion. Our hypotheses state some of the expected correlates and consequences of these measures of cohesion across a wide array of potential network studies. We then examine a case study to evaluate our measures of both adhesion and cohesion, and show how it is necessary as well to take into account relative density in the context of connectivity subsets. We then define conditional density and a scalable measure of cohesion that combines connectivity and conditional density within nested patterns of subgroup cohesion and subgroup inhomogeneity.

I. Connectivity, and resistance to being pulled apart by removal of nodes.

Simmel (1908 [1950: 123]) noted the fundamental difference between a solitary dyad and a triad:

> "The social structure [of the dyad] rests immediately on the one and on the other of the two, and the secession of either would destroy the whole" … "as soon,

however, as there is a sociation [clique] of three, a group continues to exist even in case one of the members drops out"

Cohesion begins with simple connectedness, but group cohesiveness entails something more, namely, a generalization of Simmel's intuition that the removal of some one (two, three…) actor(s) should not disconnect a cohesive group.

1.1.1. *A group is cohesive to the extent that social relations distributed among different pairs of members prevent the group from being pulled apart by the removal of subsets of members.*

Minimal cohesion would be the case in a social network in which a central node of a tree (Figure 2) might be a group leader or popular figure. Even with a charismatic leader, the cohesion of the group is low if the leader's removal would disconnect the group. We use the concept of "adherents" of a social group to indicate the many-to-one commitments of individuals to the group itself or to its leadership. What holds the group together where this is the major factor in group solidarity is the strength of adhesion of members to the leader, not the cohesiveness of group members in terms of social ties amongst themselves. The model of "adhesion" rather than cohesion might apply to the case of a purely vertical bureaucracy where there are no lateral ties.

(Insert Figure 2 about here)

Harary et al. (1965) anticipated the approach of utilizing connectivity as a measure of cohesiveness. Later, White (1998), Moody and White (n.d.), and White et al. (2000) developed the idea further. We implement this approach here.

1.1.2. *A group's cohesion is equal to the minimum number of actors who, if removed from the group, would disconnect the group.*

This is the minimum number k of its actors whose removal would not allow the group to remain connected or would reduce the group to a single member.[4] The *trivial graph* K_1 of one node and no edges (n=1 and m=0), or a disconnected graph, has cohesion 0. A solitary dyad has cohesiveness 1, a triad has 2, and a 4-clique has 3.

In the terminology of graph theory, 1.1.2 corresponds precisely to the (node) connectivity of a graph. Two primary references on the node and edge connectivities of G, denoted by κ (kappa) and κ', respectively, are Harary (1969, Ch. 5) and Tutte (1966). The *removal of a node* v from G leaves the subgraph G - v that does not contain v or any of its incident edges. The *(*node-*) connectivity* κ(G) is defined as the smallest number of nodes that when removed from a graph G leave a disconnected subgraph or a trivial subgraph.[5] The connectivity of a disconnected graph is zero as no nodes need to be removed; it is already not connected.

A *cutnode* of a connected graph G is one whose removal results in a disconnected graph. A *node cut set* is a set of nodes whose removal results in a disconnected graph. An *endnode* of a connected graph G is one with a single incident edge. Its removal does not result in a disconnected graph. A *cycle* C_n is obtained from a path P_n with $n \geq 3$ by adding an edge joining its two endnodes. A cycle containing nodes u,v entails the existence of two disjoint paths between u and v. A *tree* is a connected graph with no cycles (Figure 4). It is easy to see that each node in a tree is either an endnode or a cutnode.

A connected graph has connectivity 1 if it has a cutnode, and conversely. Thus, a tree has connectivity 1 but a cycle does not; it has $\kappa = 2$. In Figure 5, which shows the eleven graphs of order 4, the first five graphs are disconnected (the first graph is *totally disconnected*), while the remaining six are connected and thus consist of a single component. The first two connected graphs are the trees of order four. The last is the complete graph K_4. The second of size 4 is the cycle C_4. The graph before C_4 has a cutnode and hence connectivity 1.

(Insert Figure 5 about here)

A maximal connected subgraph of G with connectivity k > 0 is called a *k-component* of G, with synonyms *component* for 1-component, *bicomponent* for 2-component (called a *cyclic component* by Scott 1991:108) and *tricomponent* for 3-component (called a *brick* by Harary and Kodama 1964). In Figure 5 graphs 5 and 8 have bicomponents of order 3, namely, triangles; the three graphs from C_4 to K_4 have bicomponents of order 4; and the complete graph K_4 is itself a tricomponent.

A *block* of G is a maximal connected subgraph with no cutnodes (Harary 1969, Even 1979, Gibbons 1985).[6] The blocks of a graph give a partition of its edges. In Figure 5 there are three graphs that are single blocks: C_4 and the last two graphs. Graphs 2 and 5 contain a single block plus isolated nodes. There are two K_2 blocks in graphs 3 and 4 and two blocks in the graph before C_4. Three blocks are contained in each of the two trees of order 7, since each edge is a block. A block may contain a solitary dyad (not contained in a cycle) whereas a bicomponent is a block in which there are 3 or more nodes.

We define here a *cohesive block* of a graph G (a term that will be useful for sociological analyses of cohesion) as any k-component of G, where the associated value of connectivity defines the cohesion of the block. Within the cohesive blocks of connectivity 1 will be nested other cohesive blocks, if any, of higher connectivity.[7] We use the term *cohesive groups* to refer to substantive contexts where this concept has been applied to identify social groups on the basis of their network connectivities. We use *cohesiveness* to refer to cohesive blocks of $\kappa = 2$ and above. We use the term *cohesive subsets* to refer to subgraphs of a graph G that may be cohesive in some respects but that do not necessarily correspond to cohesive blocks defined by connectivities of subgraphs.

II. Edge connectivity, and resistance to being pulled apart by removal of edges.

As distinct from node removal, the *removal of an edge* e from G leaves the subgraph G - e that contains all the nodes of G. Edge removal presents a lesser vulnerability to a graph being pulled apart than node removal, which removes all incident edges. An edge of a connected graph whose removal results in a disconnected graph is called a *bridge*.[8] The *removal of a set* of edges in G is the successive removal of each edge e in the set. An *edge-cutset* of G is a set of edges whose removal results in a disconnected graph. The *edge connectivity* $\kappa'(G)$ of a connected graph is the smallest number of edges in an edge-cutset. Thus a disconnected graph has $\kappa' = 0$.

Parallel to 1.1, we can define the adhesion of a group by reference to the specific ties between pairs of individuals, e.g., member to leader, member to specific other member:

2.1. *A group is* ***adhesive*** *to the extent that the social relations of its members are* ***pairwise-resistant*** *to being pulled apart*.

We can then make this definition more precise by correspondence to the concept of edge-connectivity:

2.1.1. *A group is adhesive to the extent that social relations distributed among different* ***pairs of members*** *prevent the group from being pulled apart by the removal of* ***subsets of edges****.*

2.1.2. *A group's adhesion is equal to the* ***minimum number of edges*** *between group members that, if removed, would disconnect the group.*

Edge connectivity does not differ from node connectivity for the graphs in Figure 5, where both types of connectivity are equal: zero for the first five graphs; one for the next three; two for the ninth; and three for K_4. Only at order 5 do the connectivity and edge connectivity of graphs begin to diverge,[9] as exemplified in Figure 6, where the edge connectivity is 2 but the connectivity is 1. Nodes 1 and 2 have just one node-independent path, but two edge-independent paths.

(Insert Figure 6 about here)

Figure 8 displays another illustration of the difference between node- and edge-connectivity. Here, between nodes 1 and 2 there are two node-independent and three edge-independent paths. Three of the nodes are separable by removal of two edges. Node and edge connectivity both equal 2 for the total graph.

(Insert Figure 8 about here)

III. Degree, Volume and Density: The Inadequacy of Egocentric, Dyadic and Density Criteria as Measures of Cohesion

The *degree* of a node u, denoted deg u, is the number of nodes to which u is adjacent. The *minimum degree* $\delta(G)$ is the smallest degree of a node in G.

Seidman (1983) defines a *k-core* as a connected maximal subgraph of order n with $\delta \geq k$. Doreian and Woodward (1994) prove that k-cores form hierarchical series, i.e., for k' > k, a k'-core is a (possibly empty) subgraph of a k-core. The *k-core* is no guarantee of cohesion. The bowtie graph in Figure 6 is a graph with $\kappa = \delta = 2$, hence a 2-core, but is minimally cohesive because it has a cutnode and does not have connectivity 2. Such a graph has a cohesion or connectivity value of 1, but no *cohesiveness* (where $\kappa \geq 2$) because of the cutnode. For larger k the same observation holds. Thus there is no necessary concomitant increase in cohesion.

The *density* $\rho(G)$ is the ratio of the difference between m and the maximum number m_1 of edges of a graph G of order n. As $m_1 = m(K_n) = n(n-1)/2$ we have

$$\rho(G) = m / m_1 = 2m/n(n-1). \qquad (1)$$

Increases in size or density for a fixed n do not necessarily increase connectivity, and connectivity can vary independently of them. For example, the sixth graph in Figure 5 has n = 7 and m = 8 and connectivity 0. Other graphs with the same order and size have connectivity 1 or 2. There are, however, some dependencies between connectivity, degree, volume and density. We will make use of these dependencies later, in defining conditional density.

Whitney's theorem states the inclusion relations between connectivity $\kappa(G)$ at the stronger end of a scale of cohesiveness, edge connectivity $\kappa'(G)$ at the middle, and minimum degree $\delta(G)$ at the weaker end (more inclusive, but less or at best equally cohesive):

Theorem (Whitney 1932; cf. Harary 1969: 43): For any graph G,

$$\kappa(G) \leq \kappa'(G) \leq \delta(G). \quad (2)$$

From Whitney's theorem it follows that every k-component is nested in a k-edge-component that is contained in a k-core, but not conversely. In 1736, Euler proved the first theorem in graph theory, that the sum of the degrees of the nodes of any graph G is 2m, twice the size of the graph (see Harary 1969: 14). Letting $\underline{d}$ denote the average degree, this shows that $\underline{d} = 2m/n$. Since the smallest degree cannot be bigger than the average degree, i.e., $\delta \leq \underline{d}$, we have $\delta \leq 2m/n$ so $2m \geq n\delta$. Given the values $\delta \geq \kappa' \geq \kappa$ of G, Whitney's theorem implies that $2m \geq n\delta \geq n\kappa' \geq n\kappa$. Recall that the density of a graph G is $\rho(G) = 2m/n(n-1)$. It follows that the minimum density $\rho(G)$ of a graph with minimum degree δ, substituting the inequality $m \geq n\delta/2$, and canceling, is $\delta/(n-1)$. Similarly, it follows that $\kappa'/(n-1)$ and $\kappa/(n-1)$ are the minimum densities, respectively, given connectivities κ' and κ of a graph G.

Seidman and Foster (1978) define a *k-plex* as a maximal connected subgraph of order n where every node has degree n - k or greater. Not every k-plex is an (n - k)-component.

Figure 6 contains a 3-plex of order 5 (hence n - k = 2) which is not cohesive as a 2-component because it contains a cutnode. For increases in k $\geq$ 2, a k-plex may still have a cutnode and thus there is no necessary concomitant increase in cohesion. In general, k-plexes and k-cores do not entail either respective n - k or k connectivity or edge-connectivity. Figure 9 shows a graph of order 8 with a bridge between nodes 1 and 2. This graph is a 3-core and 5-plex that lacks both connectivity 3 and edge connectivity 3. Because of the bridge, the graph has edge (and node) connectivity 1. A connected graph with *no bridges* (e.g., graphs 9-11 in Figure 5) has κ' at least 2. We shall not refer to a k-core or a k-plex further, as neither lend themselves to useful theorems or measures relating to cohesion in groups.

(Insert Figure 9 about here)

IV. Connectivity, Multiple Independent Paths, and Node-flow as Cohesion : Menger's Equality

Harary, Norman and Cartwright (1965) were the first to propose the connectivity of a graph (for the digraph case) as the primary measure of cohesiveness. It is their definition of cohesion that we gave in 1.1.2. By extension:

> 1.1.3. Social groups can be regarded as multiply nested in terms of connectivity values in the following sense: a connected graph (1-connected) can contain several maximal 2-connected subgraphs, each of which can contain 3-connected subgraphs, and so forth.

While the k-components of a graph are determined by maximal sets of nodes with no k-node cut sets (1.1.1), the degree of cohesion in a group also depends on the amount of redundancy determined by multiple independent paths by which pairs of actors are connected (1.2).[10]

Sociologically,

> 1.2.1. A group is cohesive to the extent that the multiplicity of (node-) independent paths amongst its members pull it together.

One of the most useful results in all of graph theory is the formulation and characterization of k-connected graphs due to Karl Menger (1927).[11] A graph G is *k-connected* if its connectivity is at least k. It is *k-edge-connected* when its edge connectivity is at least k. Menger proved the equality of k-connectedness and the number of node-independent paths between every pair of nodes (the two most salient attributes of cohesion).[12] The *local connectivity* of two nonadjacent nodes u,v of a graph G is written $\kappa(u,v)$ and is defined as the minimum number of nodes needed to disconnect u,v from one another. When u and v are adjacent they cannot be disconnected by the removal of any number of nodes. Therefore local connectivity $\kappa(u,v)$ is not defined when u and v are adjacent. In particular, a complete graph, in which every pair of nodes are adjacent, does not have any local connectivities, and its connectivity is defined as n-1. But when G is not complete, the (global) connectivity $\kappa(G)$ is the minimum value of the local connectivity taken over all nonadjacent pairs of nodes. Local and global edge connectivities are defined similarly, except that there is no exception in the case of adjacent nodes.

Local Theorem A (Menger's Theorem à la Whitney 1932). Let nodes u and v in G be nodes that are not adjacent. The minimum number of nodes κ(u,v) whose removal disconnects the pair u,v equals the maximum number of independent u-v paths.[13]

Global Theorem A (Menger's Theorem 1927). A graph G is k-connected if and only if between any two nodes u,v of G, there are at least k u-v paths that have no common nodes except for u and v.

Restating the previous formulation of cohesion in terms of groups, we see

1.2.2. A group's cohesion is equal to the minimum number of node-independent paths taken over all pairs of members.

For sociology, Menger's theorem states the equivalence of our two parallel series of definitions of cohesion (now combined): connectivity and number of independent paths. We now begin to expand on these two aspects of cohesion and to give them a fuller sociological interpretation.

A *multigraph* M is obtained from a graph G when some of the edges are converted to two or more edges. Thus, every graph is a multigraph, but not conversely. An (integer valued) *network* is obtained from a graph G by assigning natural numbers, called *weights* or *values* or *capacities*, to the edges of G. Therefore, when each edge with value t in a network is replaced by t edges joining u and v, we have a multigraph. The local (*edge-*) *flow* from u to v in a multigraph M is precisely κ' (u,v), the minimum number of edges that must be deleted from M to disconnect u from v. We now consider only node-independent u-v paths: The local *node-flow* (i.e., node-independent flow) along node-independent u-v paths of a graph G , $\kappa(u,v)$, is the minimum number of nodes that must be deleted to disconnect u from v. By Local Theorem A, this is the maximum number of paths from u to v that do not pass through any of the same intermediate nodes.

To capture the idea of node-flow for a multigraph (equivalently, for a network) we must define a new concept of connectivity, since we are considering flow through edges and not the possible limiting capacities of nodes. The local *node-flow* $\kappa''(u,v)$ in a multigraph M is the smallest number of edges in a set of node-independent u-v paths whose removal disconnects u and v.[14] The *node-flow connectivity* $\kappa''(M)$ is the smallest number of edges in a set of node-independent paths connecting any pair of nodes, whose removal disconnects M. Figure 7 shows a multigraph where $\kappa(u,v) = \kappa(M) = 1$ but $\kappa''(u,v) = \kappa''(M) = 2$. For a graph, $\kappa''(G) = \kappa(G)$, but not necessarily for a multigraph.

(Insert Figure 7 about here)

For any u-v path in M, there exist p *parallel u-v paths* (including the original path) if and only if between each pair of adjacent nodes in the path, there are at least p independent edges. By restatement of Menger's Local and Global Theorems[15],

> $\kappa''(u,v)$ equals the maximum number of parallel node-independent paths in M that connect u to v.
>
> $\kappa''(M)$ equals the minimum of the maximum number of parallel node-independent paths in M joining any pair of nodes.

From these restatements of Menger´s Theorems for node-independent paths, we can derive a new version of the celebrated Ford-Fulkerson Theorem (1956):

> The maximum node-flow between any pair of nodes in a multigraph M equals the minimum number of edges in node-independent paths whose removal disconnects M.

Because removing a node v of a connected G removes deg v edges and deg v is 1 or more, we obtain

$$\kappa(u,v) \leq \kappa'(u,v) \leq \min(\deg(u), \deg(v)) \qquad (3)$$

The concept of node-flow, as defined for the first time here, is of importance to sociology because the maximum node-flow in a multigraph derives from both of its fundamental properties of social cohesion, 1.1.2 ("resistance to being pulled apart"), and 1.2.2 (“stick togetherness”). Multiple independent paths are especially important in considering influences or effects as they spread through a network, and in compensating for distance decay. Since higher redundancy -- i.e., node independence -- in flow compensates for transmission decay at larger distances, blocks of actors that are multiply connected through independent paths may act as amplification systems for boosting the coherent signals transmitted in social interactions. Moody and White (n.d.) provide substantive examples.

The concept of node-flow helps to explain how social cohesion can operate in social groups when there are many independent paths between nodes but the average distance between nodes is somewhat high. Further

> 1.2.3. Social groups are multiply nested in terms of their node-flow-connectivity values in the sense that a node-flow connected multigraph (1-node-flow-connected) can contain several maximal 2-node-flow -connected subgraphs, each of which can contain 3-node-flow-connected subgraphs, and so forth.

High connectivity (or node-flow) connectivity gives a social group resistance to being taken apart by removal of nodes (or edges in node-independent u-v paths). Thus, a subgraph (or subgroup) with connectivity k is k times more resistant to being pulled apart than a subgraph with connectivity 1. This cohesive resistance operates independently of the distances or shortest paths that characterize the cohesive block. The analogy here is with a metal chain: its strength is that of its weakest link, and is unaffected by its length.

V. Edge-Connectivity, Edge-Independent Paths and Flows as Adhesion: Menger's Equality

Parallel to the definition of adhesion in terms of resistance to disconnection through edge removal (2.1), we can also define adhesion in terms of the strength of ties, including edges in a multigraph, indirect ties and flows.

> 2.2. *A group is* ***adhesive*** *to the extent that the social relations of its members pull them together* ***pairwise****.*
>
> 2.3. Social groups can be regarded as multiply nested in terms of edge-connectivity values (adhesion) in the following sense: an edge-connected graph (1-edge connected) can contain several maximal 2-edge connected subgraphs, each of which can contain 3-edge connected subgraphs, and so forth.

The edge version of Menger's Global Theorem A follows.

> **Global Theorem B** (Menger). A graph G is k-edge-connected if and only if between every pair of nodes u,v there are at least k edge-disjoint u-v paths.

We will not give here the local version of Theorem B.

The maximum flow-minimum cut theorem of Ford and Fulkerson (1956) is one of the most useful results in all of operations research.[16]

> The maximum flow between any pair of nodes in a multigraph M (equivalently, a network) equals the minimum number of edges whose removal disconnects M.

As a sociological method for identifying network bases of solidarity, however, maximum flow-minimum cut is the weaker of the two methods we have described for measuring network-based solidarity (adhesion being the weaker and cohesion based on node connectivity the stronger). A network may have a very high capacity for flow between all pairs of nodes, and yet have a cutnode whose removal reduces potential flows between two or more parts of the graph to zero.[17] We compare the two methods (adhesive edge-connectivity and cohesive node-connectivity) in the case study following our hypotheses.

VI. Hypotheses

Differences in cohesiveness should have predictive consequences for social groups and their members across many different social contexts. Only in the long run and over many studies can we evaluate whether connectivity-based cohesion is a primary basis for the numerous social consequences at the group and individual level that are thought to follow from cohesion. Our general hypothesis is that connectivity-based cohesion will have such consequences for social groups of many different orders.

Why should the cohesiveness measure of cohesion should be predictive? There are some relations that serve as conduits for information, gossip, favors, transmission of disease, exchange of goods, etc., and for which connectivity-based measures of cohesion are ideal for describing pockets in which (node-independent) redundancies of transmission circulate (the internet is a good example).

Cohesion as resistance to a group's being pulled apart, however, is independent of distance decay, since there are still bonding effects to consider. As noted, transmission decay aside, it is k times more difficult to break apart two nodes if they have k node-independent chains of connections than it is to break them apart if they have a single chain of connections. Hence higher connectivity is an indicator of cohesion even in the complete absence of transmission effects.

Grannis (1998), for example, found that the best predictor of the contiguous zones of homogeneity in urban neighborhoods was not closeness of ties or walking or driving distance, but chains of neighboring relations along residential streets, bounded by freeways, commercial districts or natural barriers. The bonding chains did not imply that members of the homogeneous sets had a high density of neighbor relations, but that they had multiple independent or sometimes even single-stranded chains of neighboring by which members of the homogeneous (and we would argue, "cohesive") group were neighbors of neighbors of neighbors etc. without constraint on path length. Nor was a high degree of transmission of information necessary along these indirect paths. These were simply local neighborhood bonds that implied higher local homogeneity relative to non-neighbors in different graph theoretic zones of cohesion. The concatenation of these local bonds implied only the relative transitivity of homogeneity, independently of transmission or influence involving effects indirect paths of neighborhood.

Our general hypothesis is that connectivity must be considered a sine-qua-non of social cohesion, although there are other factors to consider as well. The choice of adhesion, cohesion and other measures such as density or distance on shortest paths will be partly guided by type of social relations being studied.

In the next section we analyze the statics and dynamics of cohesion in a social network case study of moderate scale.

VII. An Empirical Test Case: the Karate Club

Zachary's (1975, 1977) two-year ethnographic network study of 34 members of a karate club is a good proving ground on which to examine hypotheses involving the concepts and measures we have discussed as aspects of social solidarity. It provides an opportunity to test the predictiveness of the boundaries of nested cohesive sets against the outcome variable of "sides" taken in a factional dispute that ultimately split the club into two. The disputants were the karate teacher (T, #1, Mr. Hi) and the club administrator (A, #34, John) and the dispute was about T's desire to raise fees and A's insistence to hold down costs. This was the cause for their competition in calling meetings at which the alternating conveners hoped to gain majority votes to pass self-serving resolutions by the club members attending. The formation of factions was visible to the ethnographer and evident in meeting attendance, which varied in factional proportions according to the convener. Ultimately Mr. Hi (T) was fired, set up a separate club, and the factional split became the basis for peoples' choices of which club they would follow.

Figure 8 shows the network of (affective) friendships among the 34 members. Zachary used minimum edge cuts (the Ford-Fulkerson maximum flow-minimum cut theorem) to separate two subgroups with respect to the two disputing leaders. He did so by weighting each friendship edge by the number of contexts (karate and other classes, tournaments, bars and hangouts) in which they met. The minimum weighted-edge cut between A and T gave him a near-perfect prediction of actual faction membership, except for three persons who did not take sides in the factions.

(Insert Figure 8 about here)

Using our definitions, the adhesive subsets in Figure 8 are labeled according to their edge connectivity. These sets form concentric rings of 1-, 2-, 3- and 4-edge-connectivity. They give some sense of how friendships and adhesions unified club members before the factional dispute emerged into an actual division of the club into two separate units. Among the sets shown in Table 1, the most adhesive set consisted of 10 members, including both A and T. The table gives the members and number in each set, the edge connectivity κ' and the concentric nesting of the sets. Zachary himself did not utilize criteria for subgroup cohesion.

(Insert Table 1 about here)

The dynamics of the dispute give us the opportunity to examine adhesion, which is especially relevant to T's role as a charismatic teacher with two separate groups of adherents, as can be seen from Figure 8, and to examine the role of cohesive blocks before the split and in mobilizing action after the split. The network as a whole is unified around its adhesive leaders and their friendship circles, but is not cohesively unified. The nested 1-, 2- and 3-edge-connected subsets contain a cutnode (T), and so are not cohesive beyond simple connectedness.

The minimum of edges that need to be removed before T and A are disconnected is 10. Since there are many different 10-cuts which disconnect A and T, there is no clear prediction of factional memberships from unweighted edge cuts.

In Figure 9, five cohesive blocks (connectivity 2 or greater) are circled, and all five have one node (T) in common. Two cohesive blocks (a 3-component within a 2- component) exclude node A and three (a 4- component within a 3- within a 2-component) include node A. Table 2 shows the five cohesive blocks, their members, the number of nodes in each block, value of connectivity, and hierarchical nesting (the conditional density in the last column will be explained later). In addition, T (the charismatic teacher) is a member of the most cohesive subset shown within the dotted oval (not, however, a maximal cohesive block), if we take into account both its connectivity of 4 and its density. This is a subgroup of 6 persons, but it is part of a 10 person 4-connected group. The need to introduce density in addition to connectivity in measuring cohesion will lead to our defining density conditional to connectivity in the next section.

(Insert Figure 9 about here)

(Insert Table 2 about here)

If we set the problem of determining factional divisions in the context of the opposition between leaders, as Zachary did, the decision-making situation of the few club members who had friendships with both leaders leads to a prediction. These in-between persons – 9, 14, 20, and 32 – had to make up their minds which leader to follow as the club split. It is support for the cohesion hypothesis that while each of them had stronger or at least (for #14) equal strength ties with 34, the choices they made corresponded not with the number of contexts in which they have friendships with T (Mr. Hi) or A (John) but by the "pull" of cohesive ties with others. Students 14 and 20 had more independent connecting paths with T than with A,[18] and they aligned with Mr. Hi's faction in terms of meeting attendance. If we take these choices into account, and remove from the graph the line 14,34 (or also 20,34), then the 4-connected set of 10 persons breaks up leaving only one 4-connected subset {1,2,3,4,8,14}, including T (1), nested within a larger 3-connected subset, as shown in Figure 10. The 4-connected block of six persons all align with T, as predicted from cohesion. If we allocate the remaining nodes according to their cohesion with T versus A using node cut set and independent path criteria, we see that member 20 goes with T and the remainder with A. Continuing to use these criteria to allocate the remaining people, only person 10 is indeterminate, and was one of the three not factionally aligned. The final prediction of faction membership from the cohesion criteria (r=.969, p<.000001) is shown in Table 3a, where the columns indicate whether cohesion is greater, equal or less with Mr. Hi (T) than with John (A), and is nearly identical to Zachary's prediction using the Ford-Fulkerson maximum flow-minimum cut algorithm on weighted edges (Table 3b, (r=.955).

(Insert Figure 10 about here)

(Insert Tables 3a and 3b about here)

Zachary's prediction, however, contained an implicit attribute-bias in that A and T and their closest associates had the highest edge-weightings because as leaders of the club they were involved in the widest variety of different contexts. Zachary lacked two things that would have made his analysis convincing: one was a justification for the relevance of the mincut/maxflow algorithms, and second, a processual explanation and model of how and why the factions divided along the edges of a weighted minimum cut. Hence, our prediction from cohesion criteria alone is a test of the predictive aspects of solidarity that derive from network relationships alone (i.e., edges alone and not their existence in different contexts that Zachary used to define edge capacities)[19], and it gives a processual step-by-step account of the emergence of the factional lines of division that were not so apparent at the outset. Another reason that Zachary's capacitated flow argument does so well is that flow, in the unweighted case, is highly correlated with number of independent paths. Zachary has the right result, but for the wrong reasons!

Although both models make near-perfect predictions, the cohesion argument "wins" over the capacitated flow and possibly other arguments in the karate study on the grounds of parsimony, a clear explanatory principle, and a process model of efficient cause.[20] Using the same data as in Table 3a, Table 3c allows us to evaluate goodness-of-fit in further detail as to the correlation between cohesion measured by k-connectivity and faction membership (r=.929). What is most interesting in this table are the sixteen individuals in the lower rightmost cells, which show a reversal of the hypothesis of co-linearity between k-connectivity values and membership in John's faction. The cause of this departure is evident from the fact that the cohesiveness of the high-connectivity members of John's faction is eroded by their high connectivity as well (k=3 and k=4) with Mr. Hi's faction. Three of John's high connectivity allies are only weak faction members, and one of these (person 9) defects at the later stage where Mr. Hi sets up his own school. It is the more peripheral of John's supporters (with connectivity 2, linking only with him and his closest ally, 33), who are more consistently his strong faction members. This closer examination of the data supports the cohesiveness arguments as to faction membership in far more detail than explanations that follow the maximum flow-minimum cut arguments.

The karate club study also shows that while cohesion, measured by connectivity, is a measure that makes correct prediction as to expected effects, there is also a need for identifying localized high-density subgroups within cohesive blocks. This need will be evident in most empirical studies of cohesion, and motivates our development of a still more precise measure of cohesion that combines connectivity with relative density within a k-connected group. This measure of density is intended not only for this but a host of other studies.

VIII. Conditional Density

To define precisely the conditional density of a graph G with respect to some structural property, we need some preliminary definitions. Let P be a generic property of graphs, such as connected, or bipartite, etc. We always denote the order of G by n (nodes) and its size by m (edges). Let $m_0(G:P)$ be the minimum size of a graph G of order n that enjoys property P, and let m_1(G:P) be the maximum size. Then the *conditional P-density*, ρ(G:P), is defined by

$$\rho(G:P) = (m-m_0) / (m_1-m_0).$$

There is a limit on the maximum number of edges for which a graph of order n which is **not complete** can retain property P. This *upper size limit*, m_2, equals m_1+1. Thus, m_1 is the maximum m in an (n,m) graph with a given κ, and m_2 is the smallest m that forces that graph to surpass connectivity κ.

.

The *conditional P-density*, ρ_2(G:P), which is always less than one, is defined within the lower and upper size limits m_0 and m_2:[21]

$$\rho_2(G:P) = (m-m_0)/(m_2 - m_0) < 1 \qquad (4)$$

If P is omitted from either of these formulas, so that $m_0 = 0$ and we let $m_1 = m_2 = m(K_n) = n(n-1)/2$, both reduce to the usual graph density formula:

$$\rho(G) = \rho_2(G) = m / m_1 = m/(n(n-1)/2) = 2m/n(n-1).$$

For graph 8 in Figure 3, for example, the ordinary density $\rho(G) = .67$, and where the property P is that of connectedness ($\kappa = 1$), graph 8 has a surplus of one edge beyond those needed for the property $\kappa = 1$, while the maximum such surplus is two edges for a connected graph with 4 nodes, hence the conditional P-density of this graph is ρ_2 $(G{:}\kappa = 1) = 0.5$. Two surplus edges beyond those required for $\kappa = 1$ are needed to force a graph with 4 nodes to have $\kappa = 2$. The conditional P-density $\rho_2(G{:}\kappa = 1)$ is thus .33. In general, conditional P-density ρ_2 is the ratio of surplus edges, beyond those minimally needed for a graph of order n to have property P to the minimum number of surplus edges at which a graph with order n *cannot* still retain property P.[22]

Connectivity and Conditional Density: A unified approach to measuring cohesion[23]

Graph connectivity and density are two aspects of cohesion that are tightly bound together. We take advantage of their interdependence to combine and unify them into a single measure of cohesion.

To apply conditional density to the property of connectivity[24] requires the values of $m_0(G{:}\kappa)$ and $m_1(G{:}\kappa)$ or $m_2(G{:}\kappa)$ for a graph G of size m and order n with connectivity κ. These are known from extremal graph theory.[25] Let $\lceil x \rceil$ be the fraction x rounded up to the nearest integer, and for conciseness, let $m^* = m(K_{n-1})$, the size of the complete graph of order $n - 1$. The limiting size $m_2(G{:}\kappa)$ of a graph of order n with connectivity κ, where $0 \leq \kappa < n - 1$, is $1 + \kappa + m^*$. For $\kappa = n - 1$ we define $m_2(G{:}\kappa) = n(n-1)/2$, the maximum size of a graph, giving maximum conditional density of 1 only for K_n. The minimum numbers of edges m_0 of G with connectivity $\kappa = 0$, 1, and >1 are 0, $n - 1$, and $\lceil n\kappa/2 \rceil$, respectively. In general, $m_0 = \lceil n\kappa/2 \rceil$ rises linearly with n, while each of m_2 or m_2 rises quadratically, so that conditional densities are more tightly limited when there are fewer nodes and higher values of κ.

When $\rho_2(G{:}\kappa)$ is close to zero, the connectivity structure is fragile, in that the removal of a randomly chosen edge is likely to reduce the connectivity of G. The minimum size of a graph G of order n for $\kappa = 2$, for example, is n, realized only by the cycle C_n. The removal of any one of these n edges reduces the connectivity to $\kappa = 1$. If G contains one surplus edge, the chance that random removal of an edge will reduce the connectivity of G to 1 is n/(n+1). As the surplus density $\rho_2(G{:}P = \kappa)$ increases, more nodes will have extra edges, and the graph becomes less vulnerable to a lowered connectivity with the removal of a random edge. As conditional density approaches 1, the connectivity structure is more robust: Many randomly chosen edges can be removed with less chance of reducing connectivity.

Cohesion: An Aggregate Measure (connectivity & κ-density)

The sum of the connectivity κ and the conditional density $\rho_2(G{:}\kappa)$ of a graph G is not the only possible measure of its cohesion but is the best of the single measures currently available (see Moody and White, n.d.). Connectivity and conditional density each contribute independently to cohesion, according to the two objective criteria reviewed earlier. We now consider how density plays into these criteria.

For the first criterion of cohesion, namely, that a cohesive block stays together, the value k of κ is the guarantee that a graph cannot be disconnected without removal of at least k nodes. In addition, higher values of conditional density reduce the likelihood that removal of a random edge will diminish the value of κ.

For the second criterion of cohesion, that the nodes of a cohesive block should be strongly tied, the value of κ is also the guarantee, by Menger's Theorem A, that every pair of nodes in a graph has κ or more independent paths connecting them. In addition, the higher the value of conditional density, $\rho_2(G{:}\kappa)$, the less the likelihood that the removal of a random edge will diminish the minimum number κ of independent paths between every pair of nodes.

There are, of course, additional indicators of cohesion besides connectivity and density, among which are the distances between pairs of nodes (average distance, for example, along first and second shortest paths). However, to establish whether the distribution of distances between nodes in a graph with connectivity κ and conditional density $\rho_2(G{:}\kappa)$ is less than, equal to, or greater than expected in a "random" distribution, we are better off using simulation (and different probabilistic models of what is meant by "random") rather than direct measurement (Watts 1999a). Approximation methods have not yet been developed to answer these questions within random graph theory (Palmer 1985, Kolchin 1999).

A well-constructed measure of cohesion

A well-constructed measure requires a demonstration of the *unit* of measurement that gives a monotone increase in magnitude of the quantity measured. The unit for which the aggregate measure of cohesion is monotone is the addition of an edge within a graph of order n, which should be an objective correlative for an increase in cohesion. It can be verified that the aggregate cohesion measure behaves correctly by this measurement criterion: it is successively increased in any sequence of graphs of a given order in which edges are successively added. The sequences of graphs of order 4 shown in Figure 11, for example, satisfy this criterion. These are the same eleven graphs as in Figure 3, along with directed arrows showing which graphs are transformed to another by addition of an edge; their aggregate cohesion measures are given in the fourth row of Table 4: for each successive graph under edge addition, there is an *increase* in aggregate cohesion.

(Insert Figure 11 about here)

(Insert Table 4 about here)

Only certain sequences of adding edges will give maximum increases in aggregate cohesion. Cohesion is increased maximally in the sequences in Figure 11, for example, that lead from the totally disconnected graph (1) to the graph with all nodes are connected by a single path (6), to one with all nodes connected by a single cycle (9), to the graph consisting of a single clique (11). The principle of attaining successive maximum cohesion by adding edges in a graph of any order n is always to build first a single connecting path, then a cycle (connectivity 2), then to place the minimum edges needed build connectivity 3, then connectivity 4, and so forth.

Adding edges so as to build cliques of maximum order as a subgraph, of course, does not maximize overall connectivity or aggregate cohesion, although it does increase subgroup density and heterogeneity. Nor does adding edges to a graph increase cohesion if it also adds one or two new nodes incident to the edge. Unless cohesion was initially zero, this will decrease the cohesion of the graph.

Table 5, for various values of n and κ given in column 1, shows in columns 2 and 3 the values m_0 and m_2 for graphs of order $n > 1$ with $\kappa = 0$ or 1, and for graphs with $4 \leq n \leq 7$ and $n = 40$ nodes for various values of $\kappa > 1$. Illustrative graphs from Figures 1 and 2 are referenced in the table. The bowtie graph ($n = 5$, $\kappa = 1$) in Figure 4, for example, has a conditional density of (6-4)/(9-4) =0.4 and an ordinary density of 0.5.

(Insert Table 5 about here)

As noted above, the denominator of $\rho_2(G{:}\kappa)$, conditional density, insures that it cannot reach 1.0 for a given κ. This allows the aggregate measure of cohesion – as a connectivity integer plus a conditional density decimal (<1) – to correctly distinguish between the case of maximum density at connectivity κ and minimum density at connectivity $\kappa+1$, where the aggregate cohesiveness of the former is always less than that of the latter.[26]

Subgroup Cohesion

The boundaries and measures of each of the k-components of a graph provide a convenient way to study the structure of social cohesion. One of the major problems in the measure of social cohesion is the fact that different cohesive subsets overlap. The measure of cohesion based on the sum of the connectivity $\kappa(S)$ of a subgraph S and its conditional density $\rho_2(S;\kappa)$ provides a solution to the problem of overlapping cohesive subgroups in that each nested k-component of a graph has an associated measure of cohesion.

To illustrate measurement of subgroup cohesion, relationships among the cohesion measures for each k-component of the eleven 4-node graphs in Figure 3 (or Figure 11) are shown in Table 4. The first three rows of the table show the connectivity κ, conditional density $\rho_2(G{:}\kappa)$, and aggregate cohesion $\kappa+\rho_2(G{:}\kappa)$ for each of the graphs in their entirety. The second, third and fourth sets of four rows each show these values for the largest component, bicomponent and tricomponent of each of the 11 graphs.

Subgroup Inhomogeneities

Social groups with networks of high connectivity have high cohesion, but they may be highly inhomogeneous if they have high conditional densities as well. Groups with low conditional densities have relatively fewer surplus edges with which to create local subgroup inhomogeneities. Thus, some of the problems in the study of nested subgroups, their relative homogeneities and inhomogeneities, and the relation between cohesion and social solidarity (Markovsky and Lawler 1994, Markovsky and Chaffee 1995, Markovsky 1998) can be studied by means of connectivity and conditional density.

IX. Testing predictiveness of cohesion measures on a larger scale

The evidence of the karate data is useful in establishing that connectivity is a measure of cohesion that makes correct predictions about the consequences of cohesion for individual behavior and the emergence or division of social groups, but it does so on a relatively small scale in which different measures and approaches to cohesion also give similar results. The analysis of overlapping cliques in this case gives similar results for these data (Everett ms.), but large high-connectivity groups will not in general be constructed of overlapping cliques. We do not presume that connectivity is the only measure of cohesion, only that it the most fundamental component of interpersonal cohesion in social groups, large or small.

In our computation of factional groups for the karate club, we used the measure of the number of independent paths between nodes to compute pairwise cohesion. Pairwise maximum flow measures the number of edge-independent paths, which is always the same or greater that the pairwise maximum node-flow or number of node-independent paths between nodes. White and Johansen (2000) successfully used the maximum flow measure, equivalent to an unweighted version of Zachary's method, in predicting large-scale political factions in a nomadic society. The use of pairwise measures of cohesive strength (both flow and node-flow) may be widely useful in studying patterns of social cohesion and adhesion, but it is still an open research question which measures make better predictions, and why.

Moody and White (n.d.) characterize socially cohesive friendship groups in twelve American high schools and financially cohesive business groups by computing the k-components of the respective graphs. They find that embeddedness in k-connected groups is a strong predictor of school attachment, and is the only predictor of attachment among diverse network variables that replicates significantly across all schools. Using data from a study of business unity by Mizruchi (1992), Moody and White also show that cohesion or level of connectivity applied to the network of corporate interlocks among 57 firms, controlling for other network measures, predicts similarity in business behaviors. They argue for and illustrate the fundamental importance of connectivity and its hierarchical embedding[27] to a wide range of applications in sociology. Moody's algorithm successively removes sets of nodes with the lowest connectivity. By combining several algorithms of low complexity, Moody has made computation feasible for relatively large graphs. Studies using connectivity to measure social cohesion (such as that by Grannis, 1998, discussed above) are still, however, quite rare.[28]

Brudner and White (1997) and White et al. (2000), for example, identified sociologically important cohesive blocks in two large (n = 2332 & 1458 respectively), sparse networks using the concept of cohesion measured by connectivity. The first of these studies showed that membership in a cohesive block, defined by marital ties among households in an Austrian farming village, was correlated with stratified class membership, defined by single-heir succession to ownership of the productive resources of farmsteads and farmlands. In the second study, they found that the cohesive block defined by marital ties of Mexican villagers was restricted to a bicomponent that included families with several generations of residence and excluded recent immigrants and families in adjacent villages. The cohesive block defined by *compadrazgo*,[29] on the other hand, crosscut this village nucleus and integrated recent immigrants. In contrast to the first study, the Mexican case established a network basis for the observed cross-village egalitarian class structure.

X. Summary and Conclusion

In arguing for a distributed and scalable definition of cohesion, our purpose is to provide a theoretical foundation for asking some empirical questions about social cohesion that lie at the heart of sociology and social anthropology. It is clear that cohesion is an important concept beyond social psychology and is fundamental to defining social groups and their boundaries as emergent phenomena. Watts' (1999a,1999b) work on the small world problem tied network structure to an important global characteristic of networks, their “connectivity,” but like Wasserman and Faust (1994) he does not utilize the actual graph theoretic measurement of connectivity. Small scale "social psychological" cohesion based on the model of cliques and attachment-to-group makes a difference because it affects the strength of group norms, how much individuals are willing to sacrifice for the groups, and so forth. But in a large group what difference does cohesion defined by connectivity make at the macro-level? Suppose you had two large groups both equal in the density of choices, both connected and similar in other structural features as well, but one being more cohesive than the other in the sense defined here. What would be the concomitant differences in the behavior of the groups?

This paper establishes a solid theoretic and measurement basis by which macro-level questions of this sort can be addressed in real examples of large-scale social networks. It is the second and foundational paper for a series of empirical studies currently being addressed.[28] The previous paper in this series (Moody and White n.d.) contains literature reviews and empirical analyses of case studies that are addressed to questions of the consequences of cohesion defined as distributed connectivities in large-scale social groups.

We have shown how high connectivity plus a modest amount of extra-local “randomness” in social ties, which reduces average path length, is capable of generating large-scale group cohesion, coordinated social action, homogeneity, and the emergence of group norms at the macro level. The identification of bounded connectivity subgroups in a network is an ideal means of finding the boundaries of cohesive small-worlds and measuring the degree of cohesiveness in each of their embedded subgroups.[30]

Sociologists need to reconsider the concept of cohesion as a potentially “distributed” phenomena.[31] A potential bias in favor of defining social cohesion as existing “only” in face-to-face groups may well have accumulated from the past century of small-group research. We have shown logically, step by step, within a graph theoretic framework, why such a bias is mistaken in the context of measurement theory. Using the data of social networks we examined cohesion as the network component of a more inclusive concept of solidarity, which includes individual psychological attachments to a group. We offered an alternative hypothesis that network cohesion may exist in large-scale groups as easily as in small.

Our precise and scalable method for measuring cohesion in networks and subgroups of any size has the advantage of detecting boundaries of subgroups, finding hierarchies of embedded subgroups, and measuring cohesiveness at each level of embedding. No other method of measuring cohesiveness has these advantages (most give an overwhelming welter of overlapping subgroups that introduce a second intervening level of analysis and interpretation), and our review of graph theoretic concepts (White 1998, Moody and White, n.d.) shows no other method to possess equal validity in terms of the construct criteria that cohesive groups are “resistant to breaking apart” and are weaker or stronger in proportion to the “multiplicity of bonds” that hold them together. When combined with conditional density, the connectivity-based measure of cohesion has measurement validity in that our measure of cohesion increases with each additional edge added to a graph of fixed size. In the karate example we showed how cohesive blocks contribute to a process of group division, but our measure of cohesion would be not be appropriate if the relation under study was that of conflict or antagonism.

The fundamental intuition involved in the concept of social cohesion, we argue, must be consistent with the idea that the greater the minimum number of actors whose removal disconnects a group, the greater the cohesion. Equivalently, as demonstrated by Karl Menger, the greater the number of multiple independent paths the higher the cohesion. The level of cohesion is higher when members of a group are connected as opposed to disconnected, and further, when the group and its actors are not only connected but also have redundancies in their interconnections. Overlapping circles of friends increase social cohesion, for example, although the idea here is not the same as the "intersecting social circles" concept of Simmel.[32] The higher the redundancies of independent connections between pairs of nodes, the higher the cohesion, and the more social circles in which any pair of persons is contained.[33]

Given that the cohesiveness of a group is greater when there are higher redundancies of interconnections by multiple independent paths, the cohesion of a graph or subgraph is measured by its connectivity and, on a finer scale, by surplus density conditional on connectivity. To give a brief synopsis of the argument, measurements of social cohesion by connectivity and conditional density are constructed by the following two sets of definitions:

(1) The *connectivity* $\kappa(G)$ of a graph G is the smallest number k of nodes whose removal disconnects any component of G or reduces the order of any component to a single node. A graph G is *k-connected* if $\kappa(G) \geq k$. A *k-component* is a maximal k-connected subgraph of G. A graph G is *k-cohesive* if $\kappa(G) = k$. Hence, for each value of k, the k-components of a social network represented by a graph G define empirical social groups with a corresponding level k of cohesion. Subgroups with higher levels of cohesion are embedded in ones with lower cohesion since the k-components of a graph form hierarchies by inclusion. By Menger's Theorem: A graph is k-connected if and only if every pair of nodes are connected by at least k independent paths. The redundancy of multiple independent paths connecting actors is fundamental to measuring group cohesion as distinct from the proximities of actors in a network. The *social cohesion* of a group is thus the minimum number k of its actors whose removal would not allow the group to remain connected or would reduce the group to but a single member. Hence a k-cohesive block is not only k-connected but every pair of actors is connected by at least k node-independent paths.

(2) *Conditional density* measures cohesion on a finer scale, that of surplus density beyond that implied by connectivity. For each of the k-components of a graph, these two measures may be combined into an aggregate measure of social cohesion, suitable for both small- and large-scale network studies.

Connectivity is a distributive phenomenon with emergent properties – such as might define the boundaries of social groups – that are of fundamental importance to the study of social networks. Large scale cohesive sets may appear in terms of network connectivities that would not appear by the use of other measures. Correlations between hierarchical embeddedness in cohesive blocks and potential effects of cohesion (such as school attachment, or similarities in business behaviors, as in the studies of Moody and White n.d.) underscore the conceptual and substantive importance of connectivity as the primary measure of cohesiveness. In the study of social networks, both large and small, node connectivities and conditional densities are fundamental measurement concepts for social cohesion.

One of the bases of preconceptions about cohesion that is most resistant to change is the idea that in social networks, social interaction has only proximal effects, and that indirect effects quickly decay as we move from direct effects (distance 1) to effects along paths of distance 2 or 3, beyond which indirect effects are regarded as minimal. It is worth stressing once more that what this bias in preconceptions of social cohesion omits are the two fundamental properties of the redundancies created by multiple independent pathways and large node cut sets.

First, independent pathways are convergent in their indirect effects, even at a distance. Independent paths between every pair of nodes in a cohesive block defined by connectivity (equal to the minimum number of such paths) may more than compensate for the decay of effects of cohesive interaction along long paths. Studies of large-scale social diffusion, for example, typically rest upon and demonstrate the fact that long paths matter. What connectivity provides to the internal networks of cohesive groups in terms of transmission effects is the possibility for repetition along multiple independent pathways in the transmission of rumor, information, influences and cycles of dispersion/concentration in the transmission of material items.

Second, multiple independent pathways (equinumerous to minimum cuts) necessarily imply stronger bonding between pairs of nodes, regardless of distance decay. It is k times as hard to break apart a network tying nodes together by k node-independent pathways that it is to break apart a single chain that connects them.

The effects of multiple bonding and redundancy or repetition along convergent independent pathways are crucial in the formation of social coherence, social norms, sanctions and solidarities, and the emergence of socially or cultural homogeneous groups, and thus should be of focal interest to the study of social cohesion, including cohesion on a very large scale.

Acknowledgements

We thank James Moody and Phil Bonacich, as well as the anonymous AJS reviewers of the Moody and White (n.d.) and White et al. (2000) papers, for a great many useful comments and suggestions made prior to the final revisions of this paper. In making final revisions, we are deeply endebted to the commentary of the following members of the Santa Fe Institute Working Group on Co-Evolution of Markets and the State: John Padgett, Walter Powell, David Stark, Sander van der Leeuw, and Sanjay Jain.

References

Alba, Richard D. 1972. COMPLT – a program for analyzing sociometric data and clustering similarity matrices. **Behavioral Science** 17:566.

Alba, Richard D. 1973. A graph theoretic definition of a sociometric clique. **Journal of Mathematical Sociology** 3:113-26.

Alba, Richard D. 1982. Taking stock of network analysis: A decade's results. **Research in the Sociology of Organizations** 1. ADD PAGE #S

Alba, Richard. D., and Charles Kadushin. 1976. The intersection of social circles: A new measure of proximity in networks. **Sociological Methods and Research** 5:77-102.

Alba, Richard D., and Gwen Moore. 1978. Elite social circles. **Sociological Methods and Research** 7. ADD PAGE #S

Batagelj, Vladimir and Andrej Mrvar. 1997. Networks / Pajek: Program for Large Networks Analysis. University of Ljubljana, Slovenia. http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

Batagelj, Vladimir and Andrej Mrvar. 1998. Pajek: A Program for Large Networks Analysis. **Connections** 21:47-57

Blau, Peter. 1964. **Crosscutting Social Circles**. Orlando, FL: Academic Press.

Brudner, Lilyan A. and Douglas R. White. 1997. Class, Poverty, and Structural Endogamy: Visualizing Networked Histories. **Theory and Society** 26:161-208.

Carron, A. V. 1982. "Cohesiveness in Sport Groups: Implications and Considerations." **Journal of Sport Psychology** 4:123-38.

Dirac, Gabriel A. 1960. Généralisations du théorème de Menger. **Comptes Rendus de Académie des Sciences. Paris** 250:4252-4253/405-409.

Doreian, Patrick, and Katherine L. Woodward. 1994. Defining and locating cores and boundaries of social networks. **Social Networks** 16:267-293.

Even, Shimon. 1979. **Graph Algorithms**. Potomac, MD.: Computer Science Press.

Everett, Martin L. 1982a. Graph theoretic blocking k-plexes and k-cutpoints. **Journal of Mathematical Sociology** 9:75-84.

Everett, Martin L. 1982b. A graph theoretic blocking algorithm for social networks. **Social Networks** 5:323-346.

Everett, Martin L. 1983. An extension of EBLOC to valued graphs. **Social Networks** 5:395-402.

Everett, Martin L. 1984. An analysis of cyclically dense data using EBLOC. **Social Networks** 6:97-102.

Everett, Martin L. n.d. Cohesive subgroups. http://www.analytic-tech.com/networks/chapter_4_cohesive_subgroups.htm

Fararo, Thomas, and Patrick Doreian. 1998. The theory of solidarity: An agenda of problems. Pp. 1-31, *in*, eds., Patrick Doreian and Thomas Fararo, **The Problem of Solidarity: Theories and Models**. Amsterdam: Gordon and Breach.

Festinger, Leon, S. Schachter, and K. Back. 1950. **Social Pressures in Informal Groups**. New York: Harper.

Ford, L. R., and David R. Fulkerson. 1956. Maximal flow through a network. **Canadian Journal of Mathematics** 8: 399-404.

Freeman, Linton C. 1992. The Sociological Concept of 'Group': An Empirical Test of Two Models. **American Journal of Sociology**, 98:55-79.

Freeman, Linton C. 1996. Cliques, Galois lattices, and the structure of human social groups. **Social Networks** 18:173-187.

French, J. R. P., Jr. 1941. The disruption and cohesion of groups. **Journal of Abnormal and Social Psychology** 36:361-377.

Gibbons, Alan. 1985. **Algorithmic Graph Theory**. Cambridge: Cambridge University Press.

Grannis, Richard. 1998. The Importance of Trivial Streets: Residential Streets and Residential Segregation. **American Journal of Sociology** 103: 6: 1530-1564.

Gross, N., and W. E. Martin. 1952. On group cohesiveness. **American Journal of Sociology** 52:546-554.

Harary, Frank. 1969. **Graph Theory**. Reading, MA: Addison-Wesley.

Harary, Frank. 1983. Conditional Connectivity. **Networks** 13:437-357.

Harary, Frank, and Dorwin Cartwright. 1961. The number of lines in a digraph of each connectedness category. **SIAM Review** 3:309-314.

Harary, Frank, Robert Norman, and Dorwin Cartwright. 1965. **Structural Models: An Introduction to the Theory of Directed Graphs.** New York: Wiley.

Harary, Frank, and Y. Kodama. 1964. On the genus of an n-connected graph. **Fundamenta Mathematicae** 54: 7-13.

Kadushin, Charles. 1966. The friends and supporters of psychotherapy: On social circles in urban life. **American Sociological Review** 31:786-802.

Kolchin, V. F. 1999. **Random Graphs**. Cambridge: Cambridge University Press.

Laumann, Edward O., Peter V. Marsdan, and David Prensky. 1989. The boundary specification problem in network analysis. Pp. 61-87, *in*, eds., L.C. Freeman, D.R. White and A.K. Romney, **Research Methods in Social Network Analysis**. Fairfax, VA: George Mason University Press. Originally published 1982, *in*, eds., Ronald S. Burt and Michael J. Minor, **Applied Network Analysis**. Beverly Hills: Sage.

Libo, Lester M. 1953. **Measuring group cohesiveness**. Ann Arbor: Research Center for Group Dynamics, Institute for Social Research, University of Michigan.

Markovsky, Barry. 1998. Social network conceptions of solidarity. Pp. 343-372, *in*, eds., Patrick Doreian and Thomas Fararo, **The Problem of Solidarity: Theories and Models**. Amsterdam: Gordon and Breach.

Markovsky, Barry, and Edward J. Lawler. 1994. A new theory of social solidarity. Pp. 113-137 *in*, eds., Barry Markovsky, Jodi O'Brien and Karen Heimer, **Advances in Group Processes**, vol. 11. Greenwich, CN: JAI Press.

Markovsky, Barry, and Mark V. Chafee. 1995. Social identification and solidarity: A reformulation. Pp. 249-270 *in*, eds., Barry Markovsky, Jodi O'Brien and Karen Heimer, **Advances in Group Processes**, vol. 12. Greenwich, CN: JAI Press.

Menger, Karl. 1927. Zur allgemeinen Kurventheorie. **Fundamenta Mathematicae** 10:96-115.

Mizruchi, Mark S. 1992. **The American Corporate Network, 1904-1974**. Beverly Hills: Sage.

Moody, James, and Douglas R. White. n.d.. Social cohesion and connectivity: A hierarchical conception of social groups. Dept. of Sociology: Ohio State University. Submitted to AJS.

Moreno, Jacob L., and H. H. Jennings. 1937. Statistics of social configurations. **Sociometry** 1:342-374.

Morowitz, Harold. 2000. Emergence and the Law. **Complexity** 5(4): 11-12.

Mudrack, P. E. 1989. Defining Group Cohesiveness: A Legacy of Confusion? **Small Group Behavior** 20:37-49.

Palmer, Edgar N. 1985. **Graphical Evolution: An Introduction to the Theory of Random Graphs**. New York: Wiley.

Scott, John. 1991. **Social Network Analysis: A Handbook.** London: Sage Publications.

Seidman, Stephen B. 1983. Internal cohesion and LS sets in graphs. **Social Networks** 5:97-107.

Seidman, Stephen B., and Brian L. Foster. 1978. A graph-theoretic generalization of the clique concept. **Journal of Mathematical Sociology** 6:139-154.

Simmel, Georg. 1908. **Soziologie: Untersuchungen uber die Formen der Vergesellschaftung**. 2. Aufl. München: Duncker & Humbolt. Partly translated in [1950]. **The Sociology of Georg Simmel**. Editor K. H. Wolf. New York: The Free Press.

Skvoretz, John. 1998. Solidarity, social structure, and social control. Pp. 373-402, *in*, eds., Patrick Doreian and Thomas Fararo, **The Problem of Solidarity: Theories and Models**. Amsterdam: Gordon and Breach.

Tutte, William T. 1966. **Connectivity in Graphs**. Toronto: University of Toronto Press.

Wasserman, Stanley, and Katherine Faust. 1994. **Social Network Analysis: Methods and Applications**. Cambridge: Cambridge University Press.

Watts, Duncan J. 1999a. **Small Worlds: The dynamics of Networks Between Order and Randomness**. Princeton Studies in Complexity. Princeton: Princeton University Press.

Watts, Duncan J. 1999b. Networks, Dynamics, and the Small-World Phenomenon. **American Journal of Sociology** 105:493-527.

White, Douglas R. 1998. "Concepts of cohesion, old and new: Which are valid, which are not?" Dept. of Anthropology: University of California, Irvine.

White, Douglas R. 1999b. "Longitudinal Network Studies and Predictive Cohesion Theory." NSF grant #BCS-9978282, http://eclectic.ss.uci.edu/~drwhite/nsf/nsf.htm

White, Douglas R., and Ulla C. Johansen. 2000. Long-term ethnography and the collaborative study of social dynamics of a nomadic clan in Southeastern Turkey. In press, Robert V. Kemper, ed., **Long-Term Field Research in Anthropology**.

White, Douglas R., M. Schnegg, L. A. Brudner, and H. Nutini. In press. Conectividad Múltiple y sus Fronteras de Integración: Parentesco y Compadrazgo en Tlaxcala Rural ("Multiple Connectivity and its Boundaries of Integration: A Community Study."

University of California - Irvine. Spanish translation”). *In*, Jorge Gil and Samuel Schmidt, eds., **Redes Sociales: Teoría y Aplicaciones**. México, DF: UNAM Press.

Whitney, Hassler. 1932. Congruent graphs and the connectivity of graphs. **American Journal of Mathematics** 54:150-168.

Zachary, Wayne W. 1975. **The Cybernetics of Conflict in a Small Group: An Information Flow Model**. Unpublished M.A. Thesis, Department of Anthropology, Temple University.

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

ENDNOTES [34]

[1] University of California, Irvine, and New Mexico State University, Las Cruces.

Douglas R. White	Frank Harary
School of Social Sciences	Computer Science Department
University of California, Irvine	New Mexico State University
Irvine, CA 92697, USA	Las Cruces, NM 88003, USA
fax:949-824-4717	fax:505-646-6218
ph:949-452-9957	ph:505-646-6238

This paper was completed under NSF grant #BCS-9978282, “Longitudinal Network Studies and Predictive Cohesion Theory” (White 1999a). This is the foundational theoretical paper on which the hypotheses of the large-scale longitudinal network studies of the grant are based.

[2] French (1941:370), for example, discussed how a group exists as a balance between “cohesive” and “disruptive” forces. Moreno and Jennings (1937: 371) defined cohesiveness as "the forces holding the individuals within the groupings in which they are." Carron (1982:124) viewed cohesiveness as "a dynamic process that is reflected in the tendency for a group to stick together and remain united in pursuit of its goals and objectives." Gross and Martin (1952: 553) define cohesiveness by "the resistance of a group to disruptive forces" as opposed to the rather vague definition of Festinger et al.

(1950:164) of cohesiveness as "the total field of forces that act on members to remain in the group."

[3] A group with nonsymmetric relations is representable by a digraph D=(V,A) consisting of a set V of *nodes* and a set A of *arcs* (directed edges) consisting of ordered pairs of nodes in V. A more complex but also more general derivation of our results regarding measures of cohesion, applicable to digraphs, was done by Harary, Norman and Cartwright (1965:add page reference).

[4] Cohesiveness for a complete graph K_n is n-1.

[5] This two-part definition is needed because no matter how many nodes are removed from a complete graph, the remaining subgraph remains complete and hence connected until the trivial graph with one node is obtained, and we do not remove it since its removal leaves emptiness. Thus connectivity is defined as n-1 for the complete graph K_n.

[6] Scott (1991:111,189 f.n.9), owing to the fact that block has another meaning in network analysis, uses the unnecessary and unfortunate term *knot*, easily confounded with the established term with another meaning in topology. Everett (1982a, 1982b) deals separately with both types of block.

[7] To clarify the difference between graph theoretic and our sociological vocabulary, among the *blocks* of a graph, only the 2-components of order n $\geq$ 3 are *cohesive blocks.* The weakest value of *cohesiveness* has connectivity 2, in which a *cohesive block* is a 2-component. The weakest value of cohesion has connectivity 1. The *blocks* of a 1-component may consist of dyads not contained in cycles, or of larger subsets of the 1-component that have connectivity 2 or more.

[8] Another characterization of a tree is that it is connected and each edge is a bridge. A connected graph has κ'=1 if and only if it has a bridge. Connected graphs 6, 7 and 8 in Figure 5 have bridges.

[9] Node and edge-connectivity are equal for any graph in which the minimum degree $\delta(G)$ is sufficiently large so that $\delta(G) \geq n/2$ (Harary 1969:44).

[10] Alba and Kadushin (1976) define the cohesion of two nodes as the number of cycles in which they are contained. Since two cycles may differ, but have edges in common, k

cycles containing two nodes do not imply k-1 disjoint paths between them, so this measure of cohesion does not identify clear boundaries of cohesive subsets.

[11] He accomplished this as an abstract result in the study of point-set topology.

[12] A graph of connectivity k is k-connected, but a k-connected graph may have connectivity k or greater (likewise for edge-connectivity k and k-edge-connected).

[13] The local and global theorems of Menger are examples of minimax theorems in mathematics.

[14] Two or more edges are *parallel* in a multigraph M if they connect the same two nodes u,v (no parallel edges are allowed in a graph). To disconnect a multigraph, one or more sets of parallel edges must be removed.

[15] Insofar as we know, the definition of node-flow is a new concept, and the restatements are new, but its proof is obvious from Menger's theorem and the definition of node-flow.

[16] Dirac (1960) showed that this result, in which each edge e has a numerical weight w(e), is a straightforward corollary of Menger's Theorem B. Several variations on Theorems A and B are presented in Harary (1969, Ch. 5).

[17] In a separate paper, we develop the consequences of the observation that the sum of (node-flow –1) for each pair of nodes, divided by the sum of (flow) between each pair of nodes, is a measure of the decentralization of a graph.

[18] For 20 the path numbers are equal, but the path distances are weaker to A (34).

[19] It would be useful to test whether maximum node flow-minimum node cut in node-independent paths predicts the split of the club members, using the weights on edges that Zachary assigned according to the number of contexts in which each edge occurred. To do so for large networks, however, requires an algorithm that is more difficult to implement and of a higher order of complexity than the original maximum flow-minimum cut algorithm.

[20] The fact that Mr. Hi is the cutnode in a bifurcated network might help to explain – in sociological terms – why he is the instigator of the dispute in the first place: He has a set of at least five potential students who were never integrated into the larger cohesive block containing the administrator (#34), and for whom it was clear from the beginning that they would follow his leadership. He was also a strong figure for many of his other adherents.

[21] The notation for m_0 and m_1 designates that density normally varies between 0 and 1; conditional P-density ρ_2 approaches but never reaches 1 unless $m_1 = K_n$. This latter characteristic will be useful when we define cohesion as an aggregate measure consisting of the sum of connectivity κ plus conditional κ-density $\rho_2(G{:}\kappa)$.

[22] Conditional P-densities ρ and ρ_2 differ in that the denominator of the former limits density to an interval [0,1] relative to the number of edges at which property P cannot be retained. Conditional P-density, $\rho(G{:}P)$, is the number of surplus edges divided by the maximum number of surplus edges at which a graph with order n *can* still retain property P.

[23] See also Harary (1983) on conditional connectivity, and Harary and Cartwright (1961) on the number of arcs in each connectedness category of a digraph.

[24] A problem opposite to that of conditional density, covered in extremal graph theory, is *conditional connectivity*: What are the minimum and maximum connectivities for a graph of order n and size m? A 4-node graph with 5 edges, for example, must be 2-connected.

[25] See Harary (1969: 17-19) for an introduction to extremal graphs. The result $(m{:}n, \kappa>1) = \lceil n\kappa/2 \rceil$ agrees with a minimum density of $\kappa/(n-1)$ for a graph of connectivity $\kappa>1$.

[26] This is not the case for the sum of connectivity and $\rho(G,\kappa)$ conditional density, which does not give a measure of cohesion because the sum $\kappa+\rho(G,\kappa)$ for a graph with connectivity κ and size $m_1(G,\kappa)$ is the same as the sum for a graph with connectivity $\kappa+1$ and size $m_0(G,\kappa)=0$.

[27] In both their analyses, they define two measures, one the highest k-connected subgraph to which each node belongs, and the other a measure of cohesive embeddedness, discussed above. The two measures will typically be highly correlated.

[28] NSF grant #BCS-9978282, "Longitudinal Network Studies and Predictive Cohesion Theory," PI Douglas R. White and consultant Frank Harary, is focused on comparative studies of this type.

[29] Ritual kinship established between parents and godparents.

[30] Future research can combine the "small world" approach that takes as key variables the average path length of the first and second shortest independent paths between pairs of nodes (Watts 1999a, 1999b) with our k-connectivity approach. When the average path

lengths of the 1st and 2nd shortest independent paths in a network are short, the logically implication is that the average cycle length between any two pairs of nodes is also relatively short (approximated by the sum of the two shortest independent path-length averages), and thus measuring the average cycle length of biconnectivity in the network. Similarly, the sum of the averages for 1st, 2nd 3rd shortest independent paths give an approximation if the average cycle length of triconnectivity, and so forth. In the case of triconnectivity, there are two independent "shortest cycles" between pairs of nodes. The relationship of average path length in k-connectivity structures needs to be investigated both in simulations and large-scale empirical network studies.

[31] A typical critique of connectivity-based measures of cohesion might run like this: "a cycle of 1000 people (connectivity 2) running from the U.S. to China does not constitute a cohesive group." Surely not, but a group of 1000 people with a boundary at connectivity 5 (or higher, each a higher embedded level of cohesion), conditional density of 10%, and first and second shortest average path lengths of 3.5 and 4.0 (a "small world" as defined by Watts 1999a,b) is a large-scale group with considerable cohesiveness.

[32] In Simmel's (1922) conception, zones around each ego or ego-memberships in groups simply overlap or intersect to form extensive connected networks (cf. Blau 1964, Kadushin 1966), but without necessarily forming higher-order cycles or connectivity sets (but see footnote 10 regarding Alba and Kadushin's attempt to operationalize the higher-order cycles concept of cohesion).

[33] The widely used "social circles" approach to large-scale cohesion as webs of overlapping cliques (Alba 1972, 1973, 1982, Alba and Moore 1978) has the same defect as Freeman's (1996) intersecting cliques: pairs of nodes connected at some distance by multiple independent paths are not necessarily detected as part of the same cohesive subset.

Sets	Members [1,34 leaders]	n=	κ'=	Nested in set
1	1-34	34	1	
2	1-11,13-34	33	2	1
3	1-9,11-12,14,20,24-26,28-34	23	3	2
4	1-4,8-9,14,31,33-34	10	4	3

Table 1: Edge-connectivity sets for Karate Club

Sets	Members [1,34 leaders]	n=	κ=	Nested in set	ρ_2 =
1	1,5-7,11,17	6	2		.2
2	1,5-7,11	6	3	2	.54
3	1-4,8-10,13-16,18-34	28	2		.12
4	1-4,8-9,14,20,24-26,28-34	18	3	3	.12
5	1-4,8-9,14,31,33,34	10	4	4	.24

Table 2: Connectivity sets characteristics for Karate Club

Cohesion Faction	Mr. Hi (T)	Equal for T and A	John (A)	Members by id number
Mr. Hi's (T)	15			1-8,11-14,18,20,22
None	1	1	1	17, 10, 19
John's (A)			16	9,15,16,21,23-34

Table 3a: Predictions of Faction Membership from Cohesion (r=.969)

Edge Cut Faction	Mr. Hi (T)	John (A)	Members by id number
Mr. Hi's (T)	15		1-8,11-14,18,20,22
None	1	2	17, 10, 19
John's (A)		16	9,15,16,21,23-34

Table 3b: Predictions of Faction Membership from Minimum Weighted Edge Cut (r=.955)

k-Connectivity Faction	4 Mr. Hi (T)	3 Mr. Hi (T)	2 Mr. Hi (T)	2 Mr. Hi & John (T&A)	2 John (A)	3 John (A)	Members by id number
Strong - Mr. Hi's (T)	5	4	1				1-8,11,12
Weak - Mr. Hi's (T)	1	1	3				13,14,18, 20,22
None			1	1	1		17, 10, 19
Weak - John's (A)					1	3	9,16,24, 25
Strong - John's (A)					4	8	15,21,23, 26-34
Members by id number	1-4, 8,14	5-7, 11, 20	12, 13, 17, 18, 22	10	15, 16, 19, 21, 23, 27	9, 24-26, 28-34	

Table 3c: Predictions of Faction Membership from k-Connectivity (r=.929, Spearman's rho=.878)

Graph G	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
κ	0	0	0	0	0	1	1	1	2	2	3
$\rho_2(G:\kappa)$	0.0	.25	0.5	0.5	.75	0.0	0.0	.5	0.0	0.5	0.0
$\kappa+\rho_2(G:\kappa)$	0.0	.25	0.5	0.5	.75	1.0	1.0	1.5	2.0	2.5	3.0
Component S		**a,c**	**a,c**	**a,c,d**		**G**	**G**	**G**	**G**	**G**	**G**
κ		1	1	1		1	1	1			
$\rho_2(S;\kappa)$		0.0	0.0	0.0		0.0	0.0	0.5			
$\kappa+\rho_2(S;\kappa)$		1.00	1.00	1.00		1.0	1.0	1.5			
Bicomponent S					**a,c,d**			**a,c,d**	**G**	**G**	**G**
κ					2			2	2	2	
$\rho_2(S;\kappa)$					0.0			0.0	0.0	0.5	
$\kappa+\rho_2(S;\kappa)$					2.0			2.0	2.0	2.5	
Tricomponent S											**G**
κ											3
$\rho_2(S;\kappa)$											0.0
$\kappa+\rho_2(S;\kappa)$											3.0

Table 4: Connectivity k, conditional density $r_2(G:k)$, and aggregate cohesion $k+r_2(G:k)$, for the 11 graphs in Figure 3

graph sizes:	at κ	forcing κ+1
Given: order n, connectivity κ	min m_0 = $[n\kappa/2]^+$	$m_2 = 1 + \kappa + m^*$ for κ<n-1
n > 1, κ =0	0	1 + κ + m*
n > 1, κ =1 Fig. 2	n-1	1 + κ + m*
n=4, κ =2 Fig.1.9	4	6
κ =3 Fig.1.11	6	6
n=5, κ =2	5	9
κ =3	8	10
κ =4	10	10
n=6, κ =2	6	13
κ =3	9	14
κ =4	12	15
κ =5	15	15
n=7, κ =2	7	18
κ =3	11	19
κ =4	14	20
κ =5	18	21
κ =6	21	21
n=40, κ =2	40	744
κ =3	60	745
κ =4	80	756
κ =5	100	757
κ =6	120	758
κ =39	780	780

Table 5: Ranges of m_0 and m_2 for computing Conditional Connectivity k Density

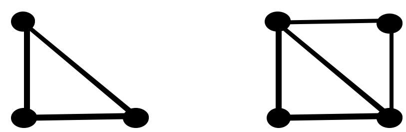


Figure 1: A disconnected graph with two components and three cliques, each a K_3.

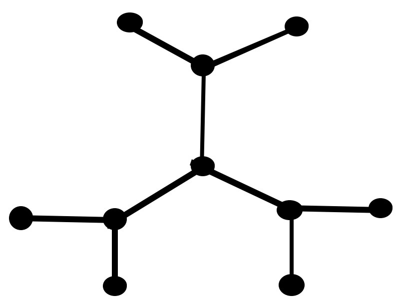


Figure 2: A graph G that is a star

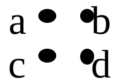
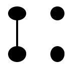
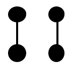
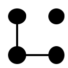
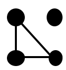
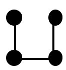
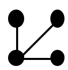
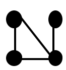
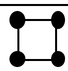
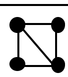
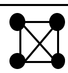
	Graph	Size	Number of Components	Connectivity	Type
1		0	4	0	
2		1	3	0	
3		2	2	0	
4		2	2	0	
5		3	2	0	
6	P_4 	3	1	1	Path (Tree)
7	$K_{1,3}$ 	3	1	1	Star (Tree)
8		4	1	1	
9	C_4 	4	1	2	Cycle
10		5	1	2	
11	K_4 	6	1	3	Complete graph

Figure 3: The eleven graphs of order 4

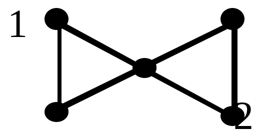


Figure 4: The bowtie graph. Its edge and node connectivities differ

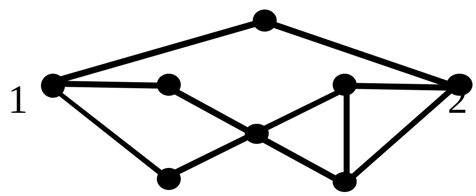


Figure 5: Graph illustrating the difference between the numbers of node-independent (2) and edge-independent (3) paths between two nodes ($\kappa = \kappa' = 2$)

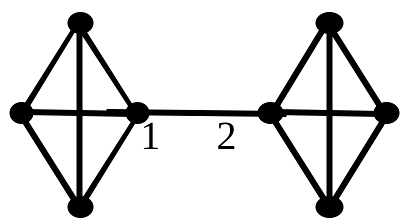


Figure 6: Graph with $\delta(G)=3$ (a 3-core and 5-plex) that lacks both 3-connectivity and 3-edge connectivity. It is not even 2-connected or 2-edge connected ($\kappa = \kappa' = 1$).

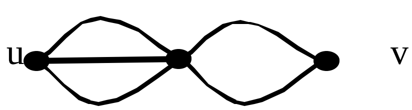


Figure 7: A Multigraph M illustrating node-flow connectivity $\kappa''(u,v) = \kappa''(M) = 2$, where the connectivity is $\kappa(u,v) = \kappa(M) = 1$.

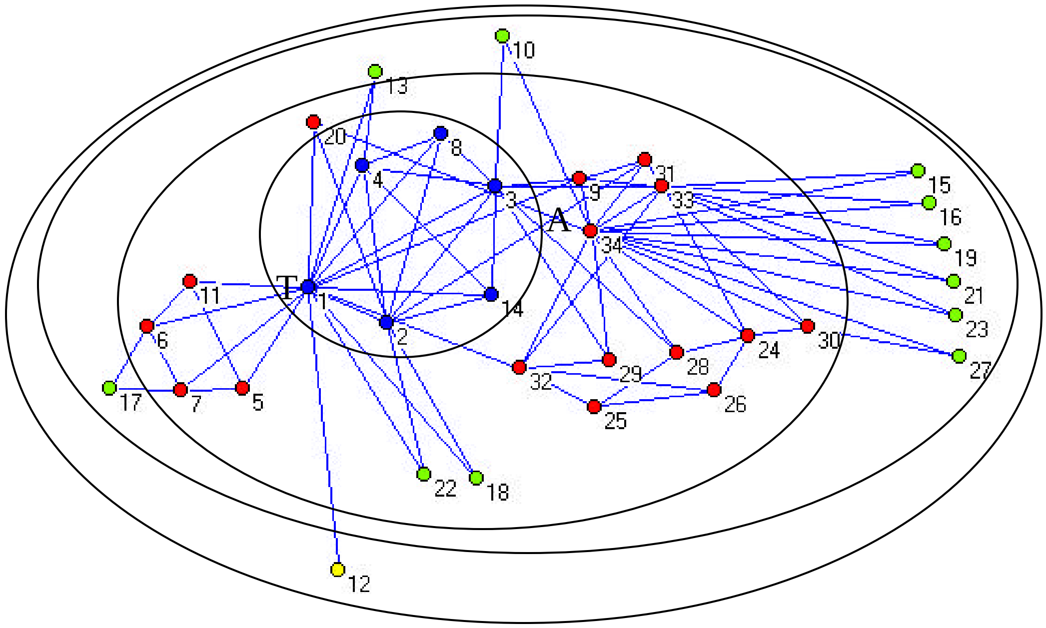


Figure 8: Nested Adhesive sets by k-edge-connectivities of 1,2,3,4 [drawn with Pajek, Batagelj and Mrvar 1997, 1998]

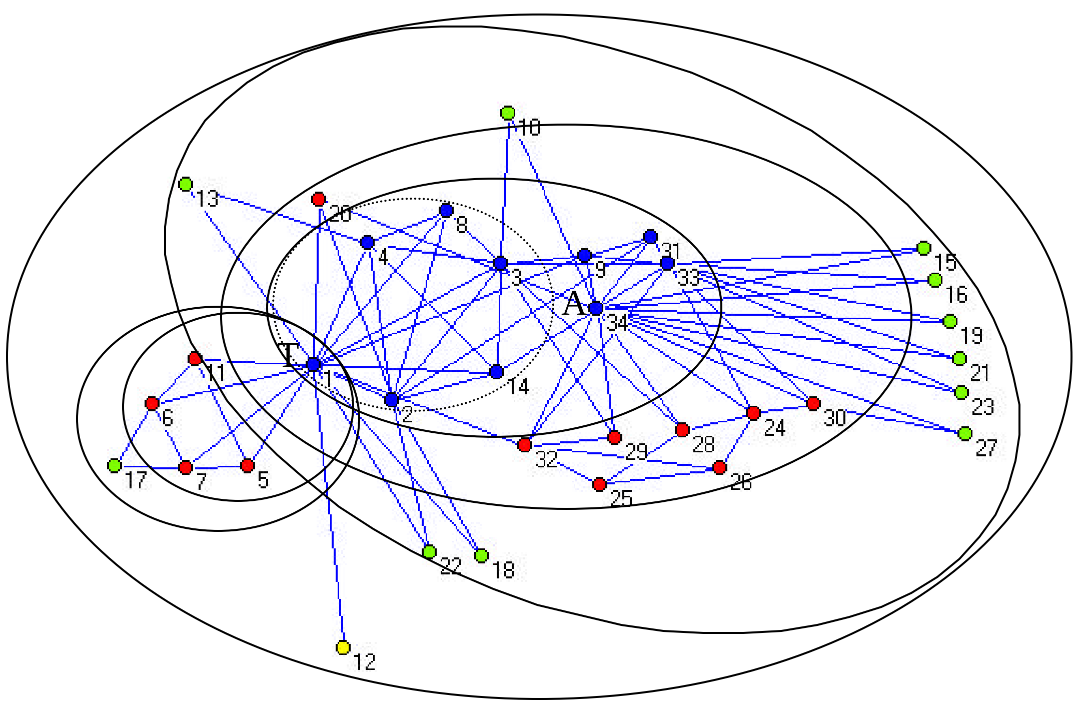


Figure 9: Cohesive blocks hierarchically ordered by connectivity into two nests (the outer dotted circle nests them all in a connected graph with connectivity 1) [drawn with Pajek, Batagelj and Mrvar 1997, 1998]

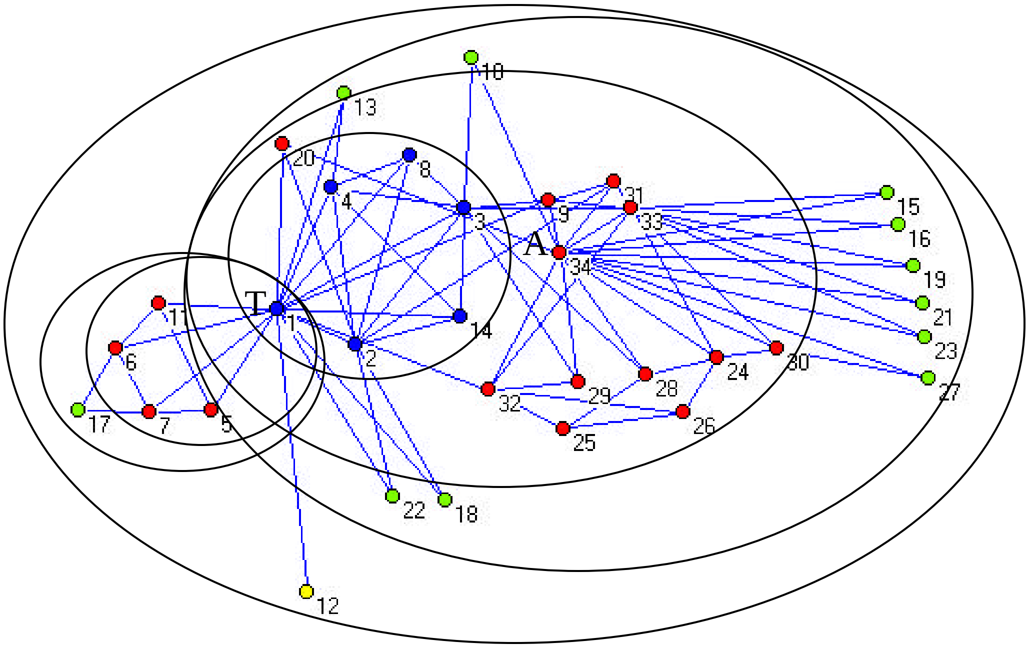


Figure 10: Nested cohesive sets by k-connectivity [drawn with Pajek, Batagelj and Mrvar 1997, 1998]

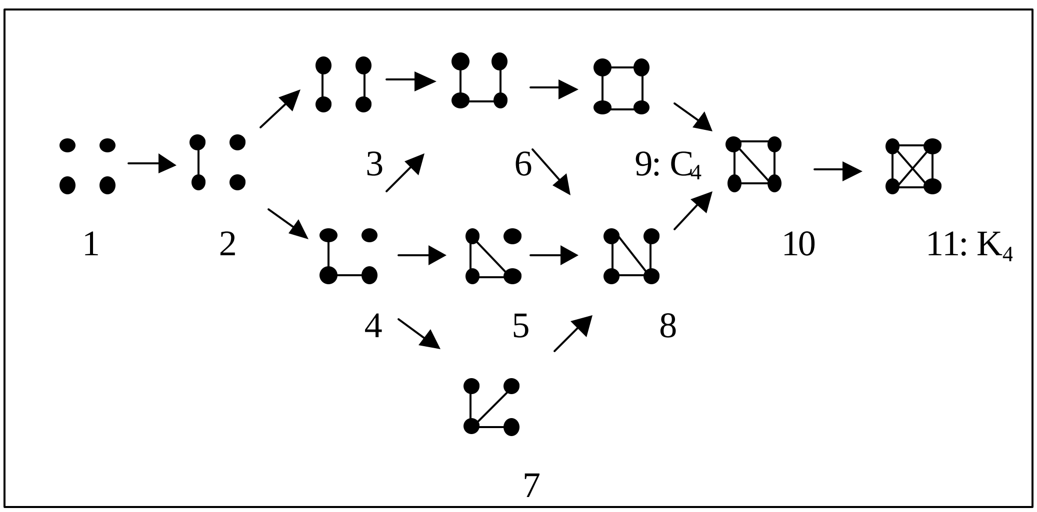


Figure 11: The 11 graphs of order 4 showing transitions by graph evolution (addition of edges)

[34] **AJS-HW6**.doc-- June 27, 2000

Note: SocialCo13.doc was the basis for this AJS-HW.doc revision, with the following use of the old Outline:

I. The Method and its Precursors ***renamed***

II. Comparison to Non-Optimal Measures of Cohesion ***dropped***

III. Conditional Density and the Aggregate Measure of Cohesion

IV. Hypotheses and Results

V. Technical Discussion – Why has connectivity been ignored? ***dropped***

VI. Summary and Conclusion

New Outline:

I. Connectivity and resistance to being pulled apart by removal of nodes.

II. Edge Connectivity and resistance to being pulled apart by removal of edges.

III. Degree, Volume and Density: Egocentric, Dyadic and Group Criteria in relation to Cohesion

IV. Connectivity and Multiple Independent Paths and node-flows as Cohesion : Menger's Isomorphism A

V. Edge-Connectivity and Multiple Edge-Independent Paths and Flows as Adhesion : Menger's Isomorphism B

VI. Hypotheses

VII. An Empirical Test Case: the Karate Club

VIII. Conditional Density

IX. Testing predictiveness of cohesion measures on a larger scale

X. Summary and Conclusion

New Finding:

Centralized node $0 < (\text{node-flow} - 1) / \text{Flow} < 1$ Decentralized Node

$\Sigma\Sigma (R_{uv} - 1) / \Sigma\Sigma F_{uv}$ is a measure of decentralization, where

R_{uv} = the minimum node cut between u and v (node-flow)

F_{uv} = the minimum edge cut between u and v (Flow)

Problem is to compute R_{uv} (See Gibbons p. 111)