Testing causal effects of structural k-cohesion

SFI Behavior Seminar  Abstract  Doug White  Friday Sept 12 12:15

Redundancy of independent paths of connection within groups of all sorts not only can create cohesion, depending on the type of communication, but can offset effects of distance in various ways. The talk invites discussion in covering

1. Definitions and measurement algorithms, Menger (1927) theorem
2. Guide to SFI working paper 08-09-042 – section on resistive cohesion; role of cohesion in evolutionary cooperation (models)
3. Problems of topological measurement – k-core percolation, robustness, and use knowledge of structure (e.g., hierarchy, degree distribution) to assess missing links, missing outside vertices; false positives for links and irrelevant vertices
4. Measuring causal effects – examples such as Karate club – high school attachment – partnership formation of biotech firms – LR & bootstrap
5. Meta-analysis for cohesion v. – density based community detection. Convergence if hierarchical?
6. Cohesion (Group) v. Role - World Trade (paper with J. Reichardt)
7. Effects in multiple-tie networks – which pairs, triples, etc. have greater causal effects – e.g., Karate club, World Trade. Other math for valued ties?
8. Lab experiments – like the Bavelas small group experiments, current network economic experiments (some coverage in the SFI WP) – questions of best and simplest treatment designs and simplest and best outcome variables – fully open to collaborations
9. Community discussions about lab experiment, software at InterSciWiki

K-components as cohesion measures are relevant for every sort of network with communicative or positive interaction links. All networks have a distribution of identifiable subgraphs characterized by different levels of cohesive intensity. This is measured by the integer series 0,...k, k_max where up to k_max there are sets of k-cohesive groups for each k > 0 (0 for a disconnected graph). Each ith group at level k has a maximal size expansion or extent e_i(k) that preserves the connectivity traversal property that I will define below. Holding k ties per member constant such a group can add members and expand its e_i(k) but without either randomization of ties or strategic hubs its density will decrease and its average distance between members increase, which may reduce the average effectiveness of redundancy of internal transmissions.

A k-scalability property S of k-cohesive groups specifies a function S=f(N,v,t) for the extent to which vertex and transmissibility properties (v,t – e.g., for people, cells, proteins) enable traversal redundancy to overcome distance in expanding group size. There might be only a constant cost of k new ties for every new member. Using language, symbols and technological media of communication, humans can benefit from highly scalable k-cohesive groups. This family of functions is as yet unspecified and k-scalability not well studied.
1. **Definitions and measurement algorithms.** We draw on standard graph theory (e.g., Harary 1969) to define **structural k-cohesion**: A graph $G = \{\text{vertices } V \text{ and edges } e \}$ (nearest neighbors) in $V \times V$ or subgraph $S \subseteq G$ ($S = \{\text{vertices } V' \subseteq V \text{ and edges } e' \text{ for } i, j \text{ in } V' \times V' \text{ iff } i, j \in e \text{ in } G \}$) **has the (k-) connectivity traversal property** if every pair of its vertices or “group members” is connected by (k or more) vertex-independent paths. Here, $k$ is the level of multiconnectivity and a disconnected graph has $k=0$. A subgraph of a graph $G$ is a $k$-cohesive group ($k$-component of $G$, $k>1$; for $k=1$, called a component but not cohesive) iff it is maximal with respect to the $k$-traversal property. (blackboard graphs)

Possible relations among distinct $k$-components of $G$ include **containment, overlap and mutual exclusion.** For $k>1$ every $k$-connected subgraph is contained in one with connectivity $k-1$, while overlapping $k$-components cannot have members in common whose number matches or exceeds the minimum of their two $k$-connectivities without being related by containment. Multiple overlapping $k$-components each with $k>1$ will also be contained in some lower-connectivity $k$-component.

**Computation.** For low density graphs of empirical networks, even of large $N$, the number of containment structures of $k$-components with a low order of $k_{max}$, say $< 8$, will often be sparse. So although computational complexity of $k$-components is NPC, practical computation problems might be severe only for large $N$, high mean number of ties per vertex, or tie distributions that approach those of Erdős-Rényi random graphs. Conceivably there may be computational shortcuts that offer advantages for graphs with power-law or other well-known distributions. Existing algorithms are programmed in SAS (Moody and White 2003) and [R] igraph (McMahan).
Fig. 1. Results of Applying Cohesive Blocking to the Southern Women Network. Extra-edges within the appropriate cohesive group are shown by vertical yardsticks and extra edges outside the cohesive group by diagonal yardsticks.
Figure 1: A hierarchy of nested k-components (4-cone), with cohesive contours

Applying a cohesive blocking to the Davis network

Key: The k-components identified by the Moody-White (2003) algorithm within the bipartite graph of events and actors in events are enclosed in solid-line cohesive contours and labeled k=2, 3 and 4. The larger numbers 1-14 give the temporal order of events. The branching tree at right shows the order in which the algorithm finds first the k-components (again labeled k=2, 3 and 4) and then the remaining embedded subsets that remain following removal of four central events within the 4-component. The dotted lines in the graph separate the 1- and 2-components identified at steps 5 and 7 by the algorithm, and the 1-, 2-, and 3-components identified at steps 6, 8 and 9. These sets are separated out within the 4-component by the dashed lines for subsets on opposite sides of the 4-component split. The only other four node 4-component split (also found by the algorithm but not shown) places temporal event 6 in the lower-left embedded set and removes Nora from the upper-right embedded set.
For large graphs an approximation technique has been described by White, Owen-Smith, Moody, and Powell (2004).
**Menger theorem.** The external $k$-cohesion of a group is its structural invulnerability to disconnection by removal of $k$ of its members. The Menger (1927) theorem is that the external and the internal $k$-cohesion of every subgraph are identical, which holds from their maximal or most extensive exemplars, that is, every vertex that has $k$ links to a maximal $k$-cohesive group is a member of that group.

2. *Guide to forthcoming SFI working paper* – section on resistive cohesion; role of cohesion in evolutionary cooperation (models)

**Scalability working paper.** These definitions help to follow the arguments of an SFI working paper soon to appear, submitted for the Ency of Complexity. It begins with the observation that in human groups, redundant paths in $k$-connected groups have great advantage for larger groups, even indefinitely large groups that follows from the use of language or symbols to transmit messages. Given the scalability (factor) of structural $k$-cohesion, the greater the ease of transmissibility, multiplied by the redundancy of $k$ paths between every pair of vertices, the greater the internal cohesion of the group and by the Menger theorem its external cohesive resistance as well.

Dynamics of resistive cohesion (Turchin) - European region map, backlash table and century maps

![Figure 1: Turchin's 2003 50 cultural regions used as geographical units in the statistical analysis of the relationship between metaethnic frontiers and polity size](image)

**Table 1:** Cross tabulations for polities that start on frontiers and end as empires a millennium later *(Turchin's 2003 50 cultural regions)*

<table>
<thead>
<tr>
<th>0-1000CE Starts as Frontier</th>
<th>No Frontier 50 regions</th>
<th>1000-1900CE Starts as Frontier</th>
<th>No Frontier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Becomes Empire</td>
<td>11</td>
<td>1</td>
<td>Becomes Empire</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>
3. **Problems of topological measurement** – k-core percolation, robustness, and use knowledge of structure (e.g., hierarchy, degree distribution) to assess missing links, missing outside vertices; false positives for links and irrelevant vertices.

**k-core.** The k-components of a network G are subgraphs of k-cores: largest subgraphs of G whose V' vertices have degree at least k. Because k-cores are not necessarily k-cohesive, and are not even necessarily connected, there is only one k-core in G for each k. In an Erdös-Rényi random graph, however, the k-cores are likely to approach maximal cohesion, which for k=1 corresponds to the giant component. In this context the k-core is a natural generalization for the giant k-component, whether as k-core or as k-cohesive component.

**k-core percolation.** (Thank Rudy Hanel) What percolation means in this context is not that k-cores overlap (they necessarily form a single containment hierarchy) but the question of "how does random damage change and destroy these substructures" (Dorogovtsev, Goltsev and Mendes DGM 2006), including the k-components? This percolation implies “the breakdown of the giant k-core at the threshold concentration of vertices or edges removed at random” (bootstrap percolation).

Not only are k-cores of fundamental importance but so are the “coronas” of the k-cores, subgraphs in a k-core with exact degree k.

For ER graphs, the recent results (paraphrasing DGM) are (1) "~in complex network models (infinite ER, SF, etc.) with a finite mean number Z_2 of the second nearest-neighbors, the emergence of a k-core is a unique hybrid phase transition (PhT; with a jump- emergence of the k-core as at a first order PhT but also with a critical singularity as at a continuous PhT). In contrast, if Z_2 diverges, the networks contain an infinite sequence of k-cores which are ultra-robust against random damage. (2) The k-core percolation threshold is simultaneously that of its finite corona clusters (3) the mean separation of vertices in corona clusters plays the role of correlation length and diverges at the critical point (4) deletion of a vertex in a k-corona cluster may cascade to eliminate linked k-corona clusters in the same k-core, resulting in collapse of a vast region of the k-core around the removed vertex, and the mean size of this region diverges at the critical point (5) k-core (vertex removal) percolation as an evolutionary process has an exact mapping as an evolutionary process that corresponds to a cooperative relaxation model with critical relaxation with a divergent rate at some critical moment. ~"

I might note in passing that the definition of k-cores by mathematician Steve Seidman (1983) originated in collaborative research with network anthropologist Brian Foster in an attempt to define cohesion in social organization. My network ethnography with anthropologist Ulla Johansen has a chapter that deals with the empirical process of k-core and k-component loss of cohesion as members of the Turkish nomadic nomadic group that we studied migrated to towns and cities, and their historical culture begins to break up.
4. *Measuring causal effects* – examples such as Karate club – high school attachment – partnership formation of biotech firms – LR & bootstrap

Karate club – edges valued by number of contexts of interaction
high school $k$-component groups predictive of school attachment (Moody and White 2003)
Causal effects – LR and bootstrap, McFadden's time-lagged Discrete Logistic Regression - Halbert White and Judea Perl

In Figure 7 the 18 columns represent the 18 women. And the 21 rows refer to the 21 analytic procedures. A woman in each cell is designated by a “W.” Groups are designated by colors. All the red “Ws” in a given row were assigned to the same group by the procedure designated in that row. All the blue ones were assigned to a second group. And in the fourth row there are green “W’s” that were assigned to a third group. Any woman who was assigned to two groups by the procedure in question, received a pair of color codes.

![Figure 7. Group Assignments by 21 Procedures](image-url)
All 21 are captured in *cohesive blocking*.

**Figure 1: A hierarchy of nested k-components (4-cone), with cohesive contours**

Applying a cohesive blocking to the Davis network.

Key: The k-components identified by the Moody-White (2003) algorithm within the bi-
6. **Cohesion (Group) v. Role** - World Trade (paper with J. Reichardt)

7. **Effects in multiple-tie networks** – which pairs, triples, etc. have greater causal effects – e.g., Karate club, World Trade. Other math for valued ties?

8. **Lab experiments** – like the Bavelas small group experiments, current network economic experiments (some coverage in the SFI WP) – questions of best and simplest treatment designs and simplest and best outcome variables – fully open to collaborations. - Discussion only

9. **Community discussions about lab experiment, software** at InterSciWiki
Search for:

*Testing causal effects of structural k-cohesion*

10 References


