SDSC 2013 Summer Institute
Structural Cohesion & Worldly Nodes
Solving for large-network connectivities

Doug White

IMBS at UC Irvine & Santa Fe Institute

and Bob Sinkovits at SDSC

structural cohesion as a measurement concept

Goal: implement Menger's 1927 theorem - **a pair of isomorphic measurements**:

Maximal k-cohesive subnetwork with **a minimum number k of node-independent paths between any x, y pairs of nodes → (hard proof: isomorphic to below)**

max above = **min** below

Maximal subnetwork with no less than a (**min**) k-node separator → (**easy proof that this entails the above**, very hard to prove the converse)

Hard part is to develop an algorithm: thought to be **NP hard** (very very slow)

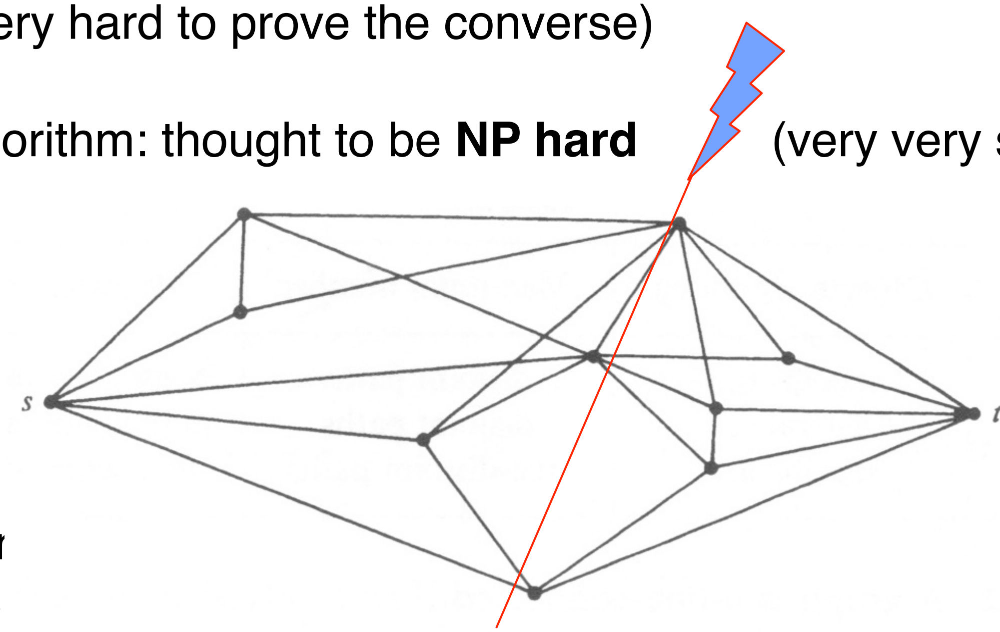


Fig. 5.5. A graph illustrating Menger s Theorem.

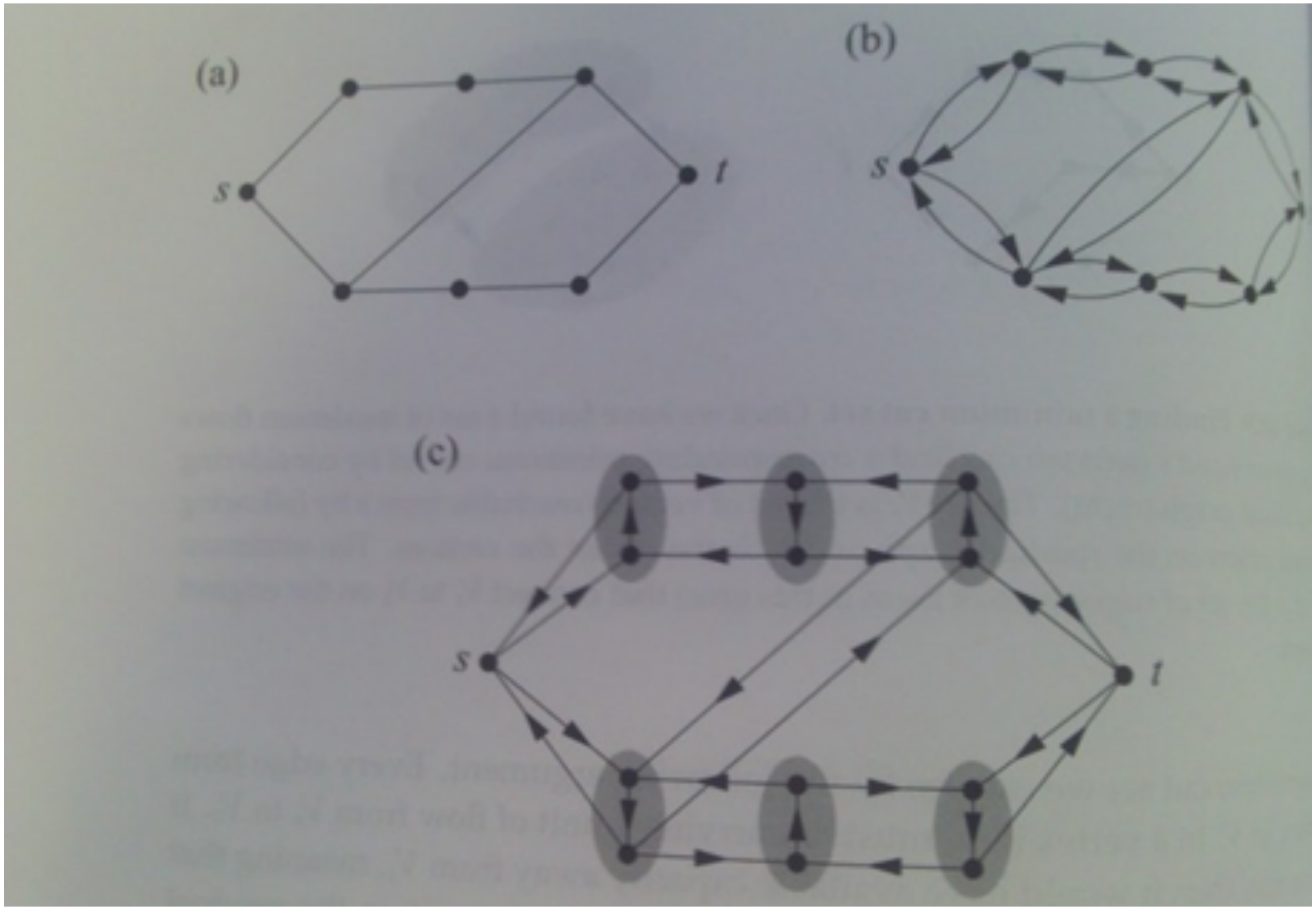
http://www.math.unm.edu/~lorir

Rediscovery of a fundamental variable

- **In graph theory and measurement, Menger's k–components are important because they uniquely define the natural units of structural cohesion that are identifiable within empirical networks: social, biological, physiological, etc.**

- **Until 1997 k–components were ignored in network research and the sciences because calculation of these cohesive units was "hard" for large graphs.** Sociologists and mathematical (Harary-trained) anthropologists led the research to identify many of the causal effects of structural cohesion.

the identification problem for vertex pair k-connectivity is hard without iGraph tricks for pairwise connectivity



First approximation by Doug White and Mark Newman 2001

Exact solution by iGraph: symmetric edges have to be transformed to reciprocal directed edges in running the iGraph algorithm

UCI's Tolga Oztan programmed this transformation in R

Definitions and Menger Theorem (1927)

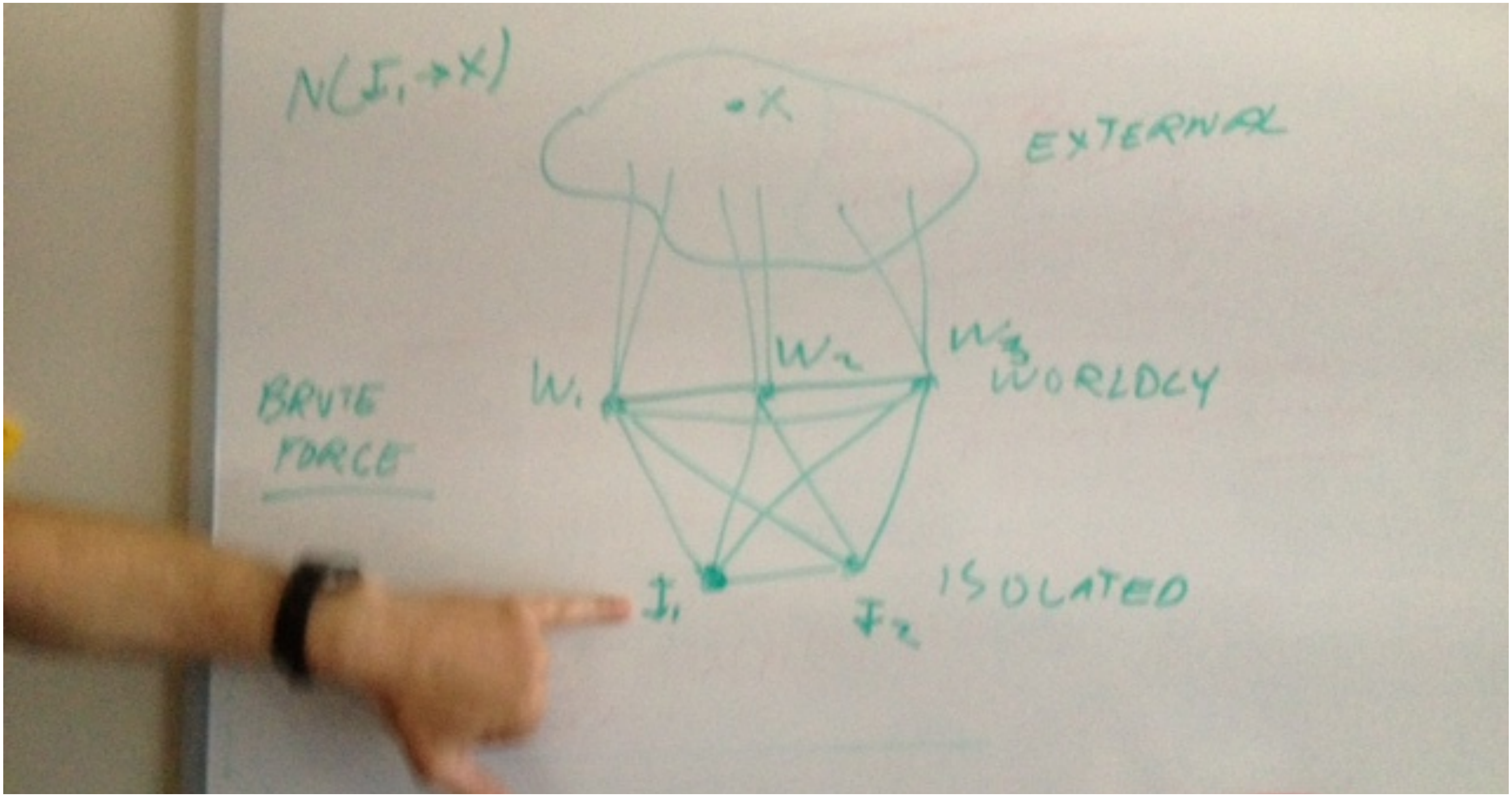
- **Structurally cohesive subgroups are a powerful and mathematically rigorous way to characterize network robustness. Their strength lies in the ability to detect strong connections among vertices that not only might have no neighbors in common, but that may be distant in the graph.**
- **1 A *k-component* of a graph G is a maximal subgraph S with the following equivalent properties:**
  - *a. multiconnectivity* k, the minimal number of node-independent paths in S connecting pairs of nodes in S.
  - *b. connectivity* k, the smallest node cutset of S is of size k.
- **Solving the identification problem for k-connectivity in large graphs is thought to be hard. Restated:**
- **2** The following White-Sinkovits GKT theorem derives from the Menger Theorem that proved equivalence between: (a) a maximal interconnected set of vertex pairs in which every pair is k-connected, and (b) a maximal set of vertices that cannot be disconnected without removal of fewer than k nodes.
- (Pairs of concepts in 1 + 2 are equivalent)

GKT Theorem: White & Sinkovits

- **The <u>Gordion Knot Theorem</u> (GKT) shows that identifying k-connectivity is not hard: The set (c) of all vertex pairs that are <u>exactly</u> 2-connected are complementary to a pairwise subset (d) of (a) in the Menger Theorem - those vertex pairs that are (k+1)-connected (a=b as before).**

- **Every vertex in the <u>union</u> of vertex-pair sets (c) and (a) adds to a growing subset of vertex pairs in some unknown final set (d). Construction depends on adding vertex pairs to (d) that have a vertex connection to (a), reaching a recursive limit in a maximal set (d) of <u>vertex pairs</u> equivalent to (a).** Is there a QED?

The Aha! Moment at Bob's whiteboard



GKT Proof

- **<u>Proof</u> (Doug): Given (a=b) between sets of vertices and sets of k-cohesive vertex pairs, the argument establishes the equivalence of (d=a)=(b) by iterative construction within a connected sets of pairs and vertices. Thus the <u>exactly</u> k-connected <u>vertex pairs</u> in a network is identified by (iGraph) vertex-pair construction of the k-connected subsets of the network.**
- **Computational efficiency identifies k-connectivities applied to the first few steps of k=2,3,4; brute force solves higher order k-connectivities by parallel computing.** We think this is QED

Gordon computation: Bob

Bob: Using existing fast algorithms, we can rapidly find all pairs of vertices that are exactly 2-connected, and then, within this set, we often find vertex pairs that are exactly 3-connected using vertex degrees and a search over just those triplets that are likely 3-vertex separator candidates.

(d)

The remaining elements of the pairwise cohesion matrix can be calculated using the power of parallel computing.

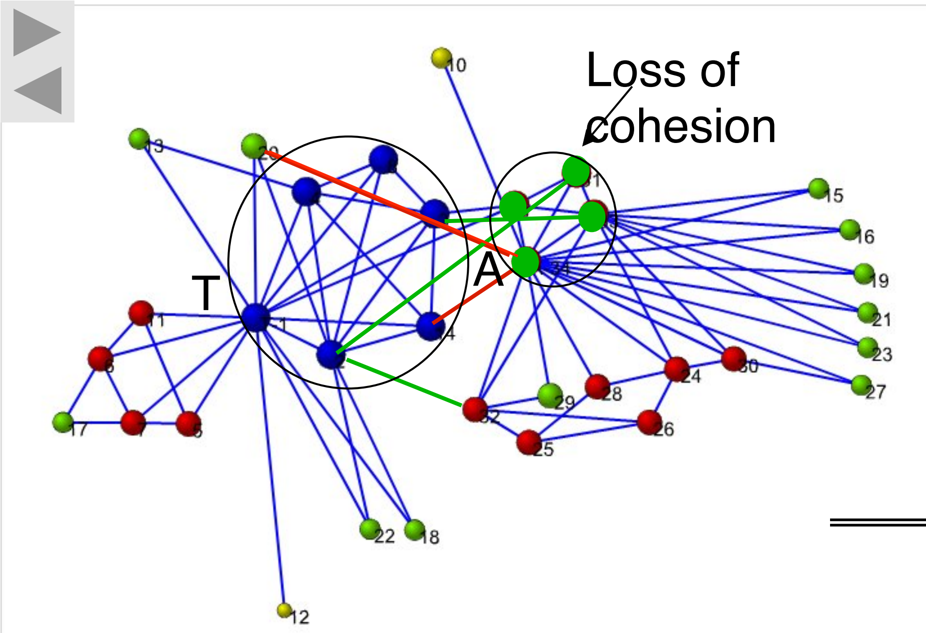
As a proof of principle, our study (Bob) presents results of cohesion analysis for a biconnected component of nearly 30,000 vertices extracted from a co-authorship data set.

New and potential applications

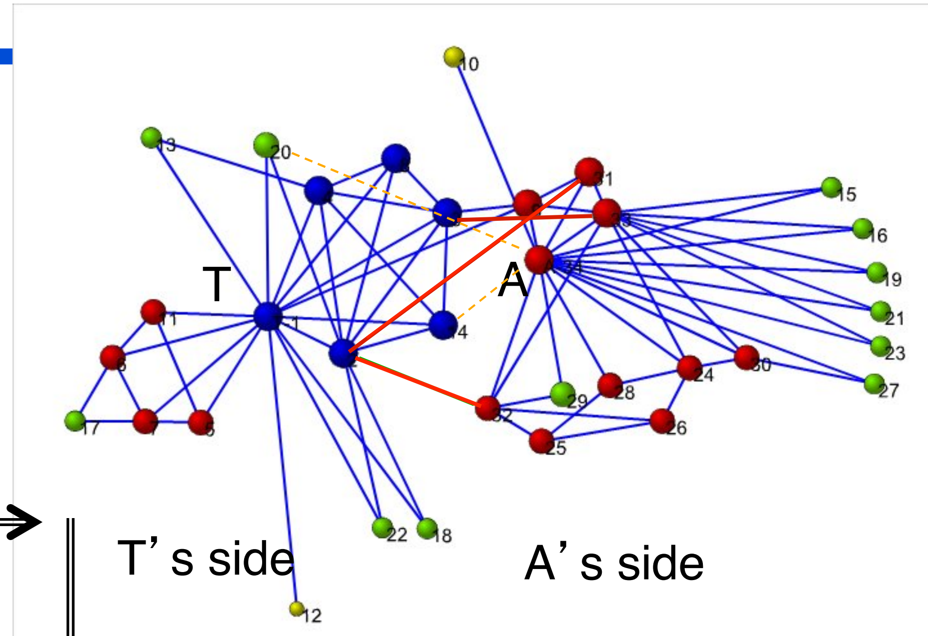
- **<u>Biologists</u> have begun to use these methods to study consequences of cohesive behavior in animal studies.**
- **<u>Historians</u> now use these methods to study consequences of cohesive structure in urban communities.**
- **<u>Biomedical</u> study of protein and interactive neural and physiological networks are good candidates for study of consequences of cohesive dynamics.**
- **<u>Evolutionary cohesive dynamics</u> now studies how groups form to create successful adaptive strategies.**

Six examples that follow

- Splitting of a conflict network along cohesive fractures (Karate club splits between owner and instructor): White & Harary 2001 <u>Sociological Methodology</u> chapter, proof of k-cohesion as a measurement concept; successful empirical prediction of the order of a Karate club split-up.
- Moody & White 2003 <u>American Sociological Review</u> (first algorithm, adopted in igraph): two sets of empirical predictions
- Ditto 1: results for 10 American schools study
- Ditto 2 Mention: Corporate attachments
- Educational studies: explanations of bullying
- Cohesion in two-mode netwks beats "cmnty detection"



T & A start to fight: some must choose sides



Opposing cohesive sides emerge

Group members with k-cohesion automatically have at least k different ways of connecting through k node-independent paths, and vice versa

T = karate teacher

A = club administrator

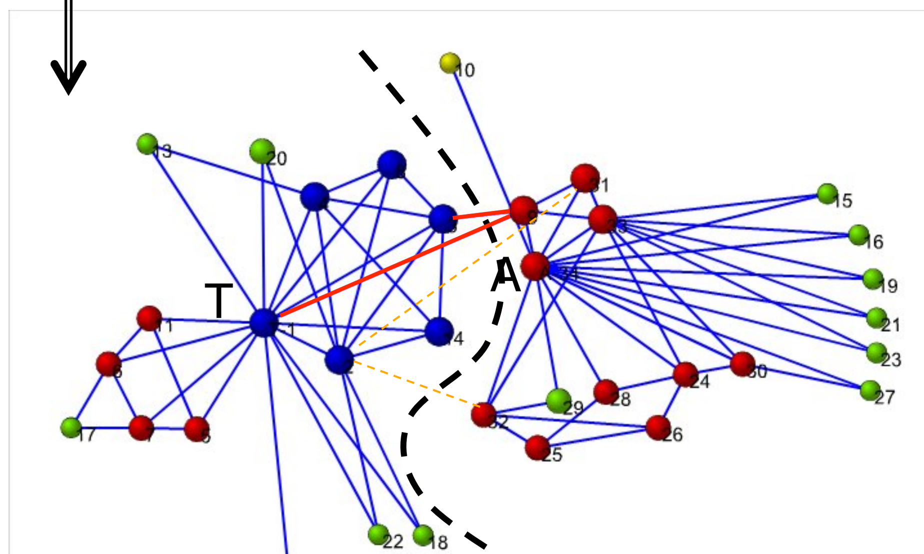
Block Connectivity:

Blue k=4 (quadricomponent)

Red k=3 (tricomponent)

Green k=2 (bicomponent)

Yellow k=1 (component)

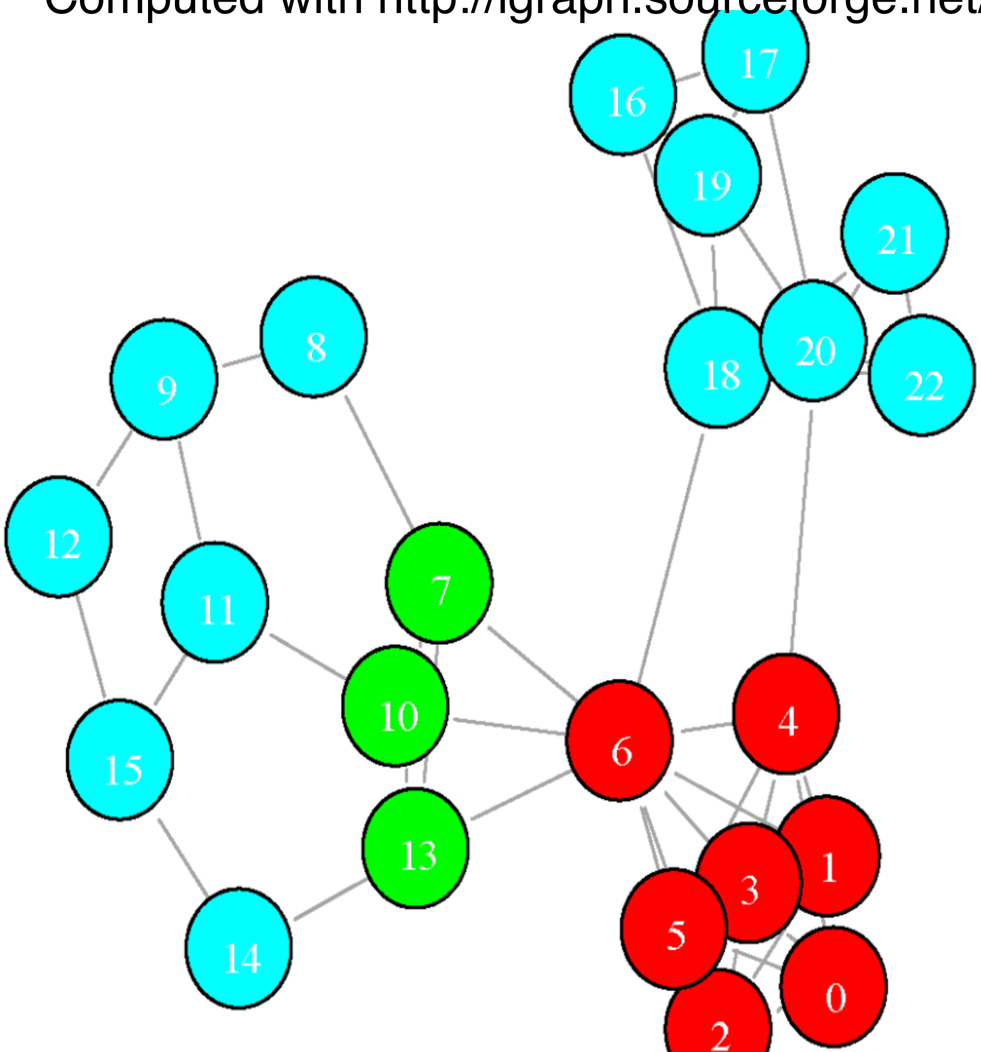


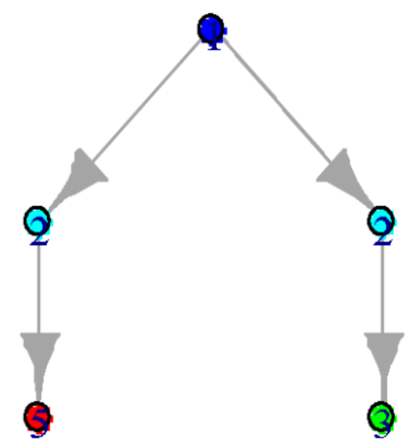
Sides separate along cohesive fractures

WHITE & HARARY 2001 SOC. METHODOLOGY

Moody & White (ASR 2003) iGraph example

Computed with http://igraph.sourceforge.net/doc/R/vertex.connectivity.html





MOODY & WHITE 2003 Cohesion of School Friends predicts student's questionnaire responses on attachment to school – in 10 U.S. Schools

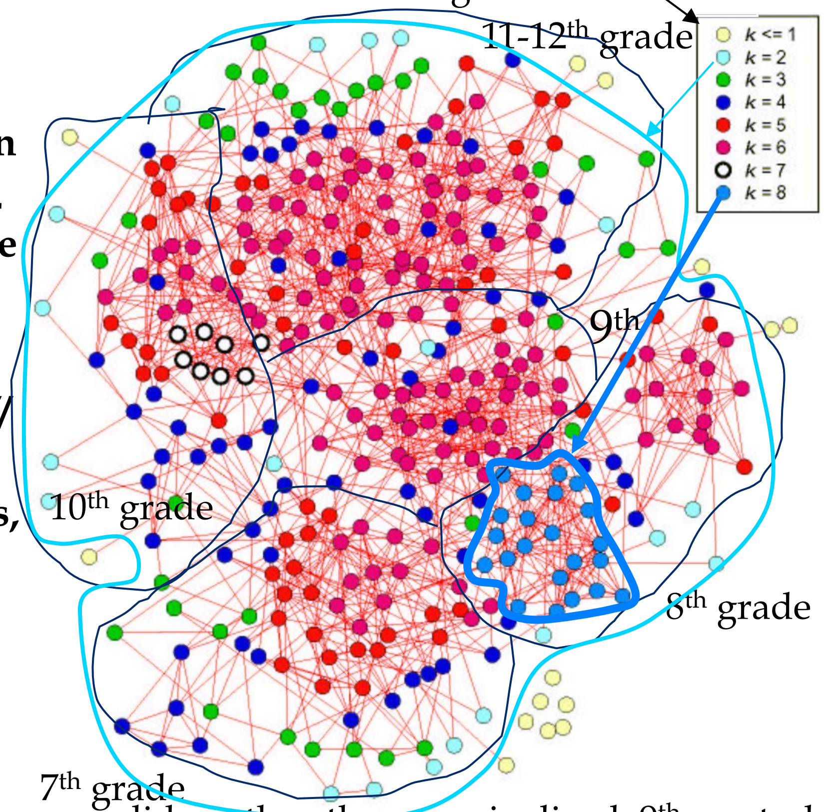
Topology: Overlapping hierarchies (Empirical Results)

The algorithm for finding social embeddedness in nested cohesive subgroups is applied to high school friendship networks (e.g., Fig 2; boundaries of grades are approximate) and to interlocking corporate directorates. The usefulness of the measures of cohesion and embeddedness are *tested against outcome variables of school attachment in the friendship study* and similarity in corporate donations to political parties in the corporate interlock study. The cohesion variables outperform other network and attribute variables in predicting the outcome variables using multiple regression.

Nearly identical findings are replicated for school attachment measures and friendship networks in 12 American high schools from the AddHealth Study (http://www.cpc.unc.edu/addhealth/), *Adolescent Risk and Vulnerability: Concepts and Measurement.* Baruch Fischhoff, Elena O. Nightingale, Joah G. Iannotta, Editors, 2002, The National Academy Press.

2003 James Moody and Douglas R. White, Social Cohesion and Embeddedness: A Hierarchical Conception of Social Groups. *American Sociological Review* 8(1)

Fig 2. Structural Cohesion of Friendships in an American high school



Interpretation: 7th-graders- core/periphery; 8th- two cliques, one hyper-solidary, the other marginalized; 9th- central transitional; 10th- hang out on margins of seniors; 11th-12th- integrated, but more freedom to marginalize

Cohesion and behavior: 3rd-4th grade surveys

classes with **egalitarian** (low-k) cohesion have **little bullying**, aggressors unpopular

<--(no class size effect)-->

classes with **hierarchical** (hi-k) cohesion **have bullying**, bullies liked, victims unpopular

All: aggression corr. with popularity but neg. corr. with social preference.

All: victimization neg. corr. with both pop.&pref but not corr. with aggression

86 Journal of Early Adolescence 30(1) Ahn et al. 85

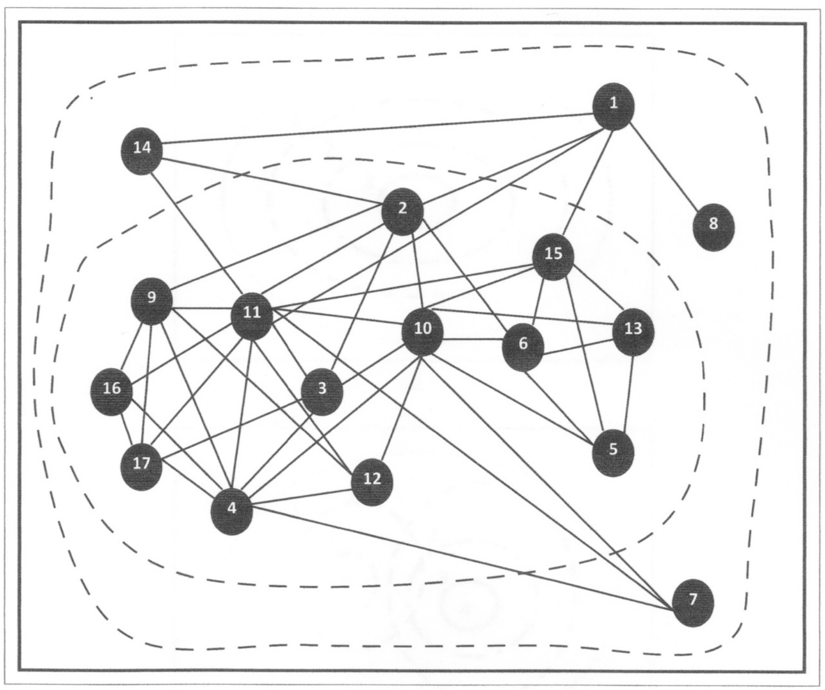


Figure 3. Affiliative network structure and ties of the classroom with lowest embeddedness
Note: Solid lines are affiliation ties, and dotted lines are the boundaries of embeddedness levels. $n = 17$ (participation rate = 71%), Classroom embeddedness (E) = 2. $Z_E = -1.66$. Classroom density (D) = 0.17. $Z_D = -0.79$.

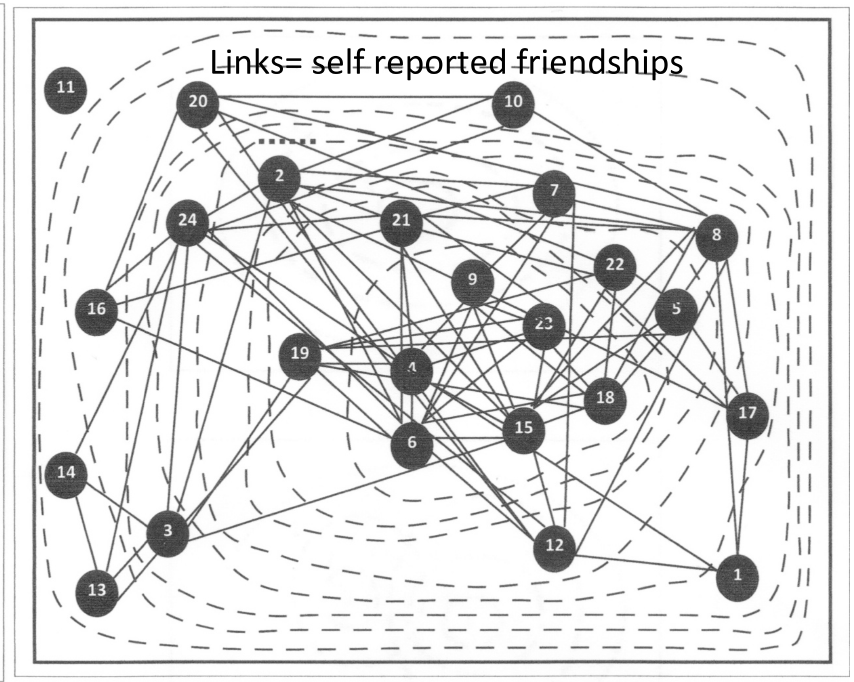


Figure 2. Affiliative network structure and ties of the classroom with highest embeddedness
Note: Solid lines are affiliation ties, and dotted lines are the boundaries of embeddedness levels. $n = 24$ (participation rate = 75%). Classroom embeddedness (E) = 8. $Z_E = 2.99$. Classroom density (D) = 0.21. $Z_D = -0.17$.

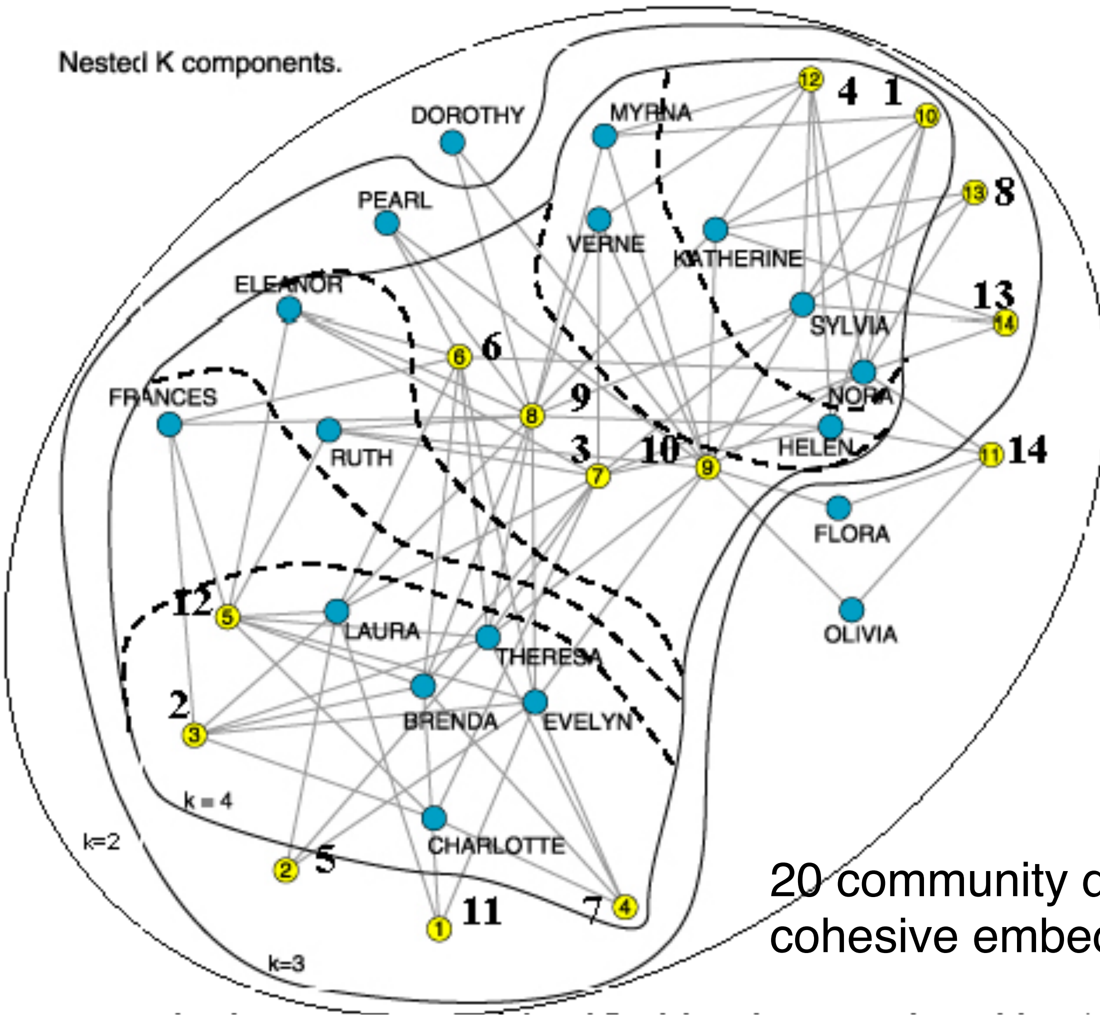
Community detection for 18 Southern Women Study: 20 of 21 say: two. Structural cohesion (& OSB00) says one.

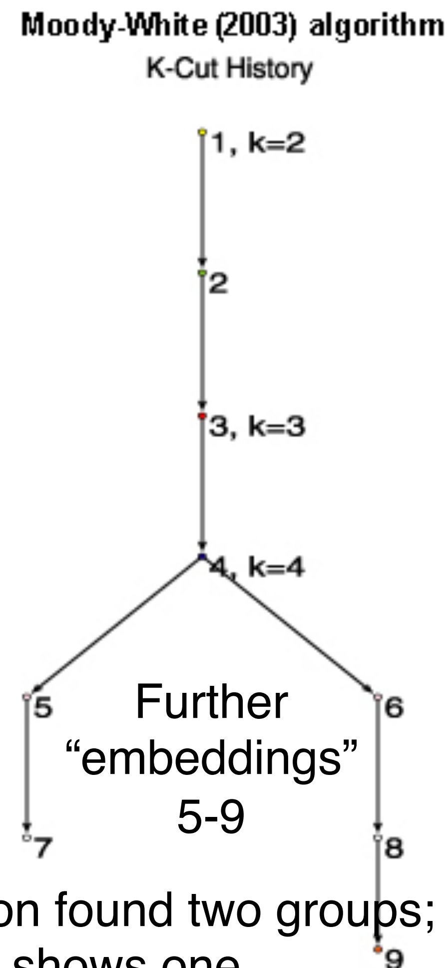
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	DGG41	W	W	W	W	W	W	W	W	WW	W	W	W	W	W	W	W	W	W
2	HOM50	W	W	W	W	W	W	W	WW			W	W	W	W	W		W	W
3	P&C72	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
4	BGR74	W	W	W	W	W	W	W		W	W	W	W	W	WW	WW		W	W
5	BBA75	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
6	BCH78	W	W	W	W	W	W				W	W	W	W	W	W			
7	DOR79	W	W	W	W	W	W	W		W	W	W	W	W	W	W			
8	BCH91	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
9	FRE92	W	W	W	W	W	W	W		W	W	W	W	W	W	W	W		
10	E&B93	W	W	W	W	W	W	W		W	W	W	W	W	W	W			
11	FR193	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
12	FR293	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
13	FW193	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	WW	W	W
14	FW293	W	W	W	W	W	W	W		W	W	W	W	W	W	W		W	W
15	BE197	W	W	W	W	W	W	W		W	W	W	W	W	W	W			
16	BE297	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
17	BE397	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
18	S&F99	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W		W	W
19	ROB00	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
20	OSB00	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W
21	NEW01	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W	W

Structural cohesion & Cohesive embedding

Applying a cohesive blocking to the Davis network

18 women at 14 social events – **2-mode network**





Further "embeddings" 5-9

20 community detection found two groups; cohesive embeddings shows one.

Longitudinal Network Studies and Predictive Social Cohesion Theory

D.R. WHITE, University of California Irvine, NSF BCS-9978282

TOPOLOGY: STACKED HIERARCHIES AND DYNAMICS (EMPIRICAL RESULTS)

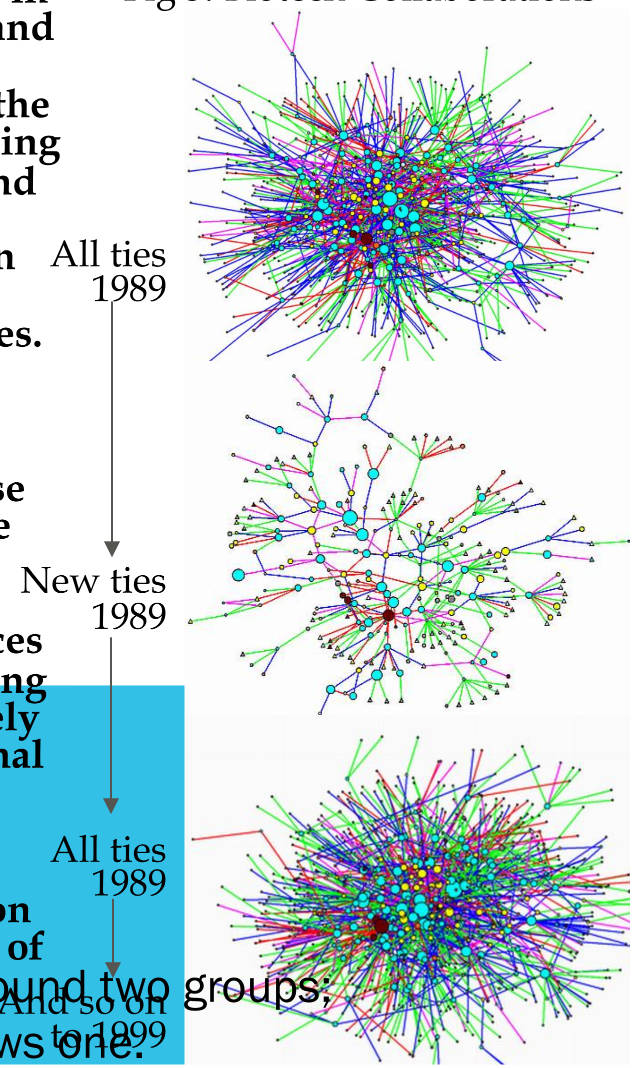
LONGITUDINAL VALIDATION OF STRUCTURAL COHESION DYNAMICS IN HUMAN BIOTECHNOLOGY WORLDWIDE STUDY (2003 POWELL, WHITE, KOPUT, OWEN-SMITH)

To account for the development of collaboration among organizations in the field of biotechnology, four logics of attachment are identified and tested: accumulative advantage, homophily, follow-the-trend, and multiconnectivity. We map the network dynamics of the field over the period 1988-99 (Fig 3 →1999). Using multiple novel methods, including analysis of network degree distributions, network visualizations, and multi-probability models to estimate dyadic attachments, we demonstrate how a preference for diversity and multiconnectivity in choice of collaborative partnerships shapes network evolution. Cohesion variables outperform scores of other independent variables.

Collaborative strategies pursued by early commercial entrants are supplanted by strategies influenced more by universities, research institutes, venture capital, and small firms. As organizations increase both the number of activities around which they collaborate and the diversity of organizations with which they are linked, cohesive subnetworks form that are characterized by multiple, independent pathways. These structural components, in turn, condition the choices and opportunities available to members of a field, thereby reinforcing an attachment logic based on connection to partners that are diversely and differently linked. The dual analysis of network and institutional evolution offers a compelling explanation for the decentralized structure of this science-based field.

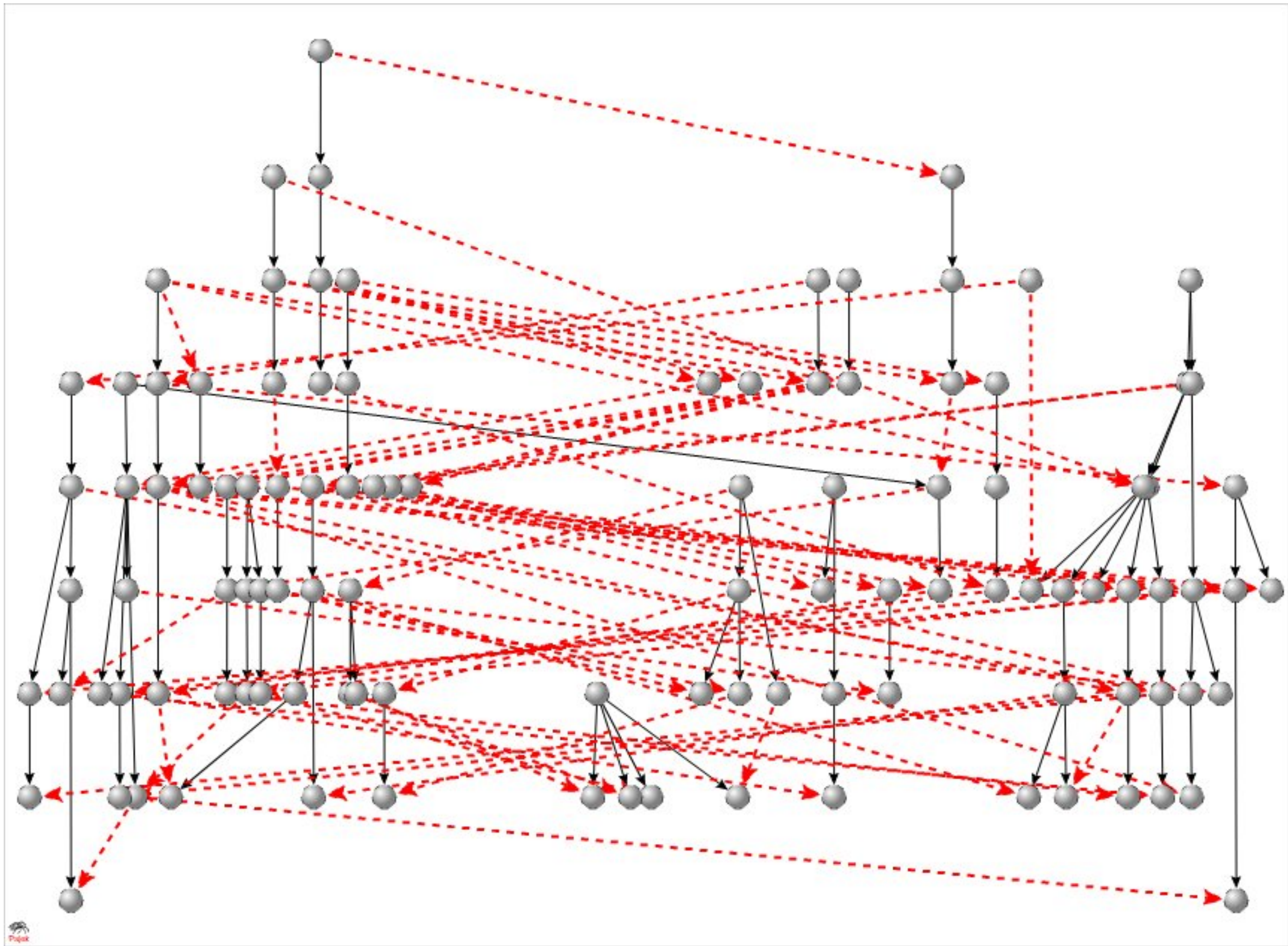
2003 Walter W. Powell, Douglas R. White, Kenneth W. Koput and Jason Owen-Smith. Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences, 1988-99. Submitted to: *American Journal of Sociology*.

Fig 3. Biotech Collaborations



20 community detection found two groups; cohesive embeddings shows one.

The origin of these studies is from earlier network-anthropological studies of kinship



1998, 1999, 2010 White; Houseman.

This is Dravidian kinship in Sri Lanka

Measures of cohesion AND measures of bipartite intra community structure

Applications of Cohesion to Structural Endogamy

Social Class: Carinthian Farmers Theory and Society (1997)

red circles 2-connected i.e., *bicomponents* with 2-family relinkings, the simplest affinal relinking. They predict inheritances and community stayers versus leavers

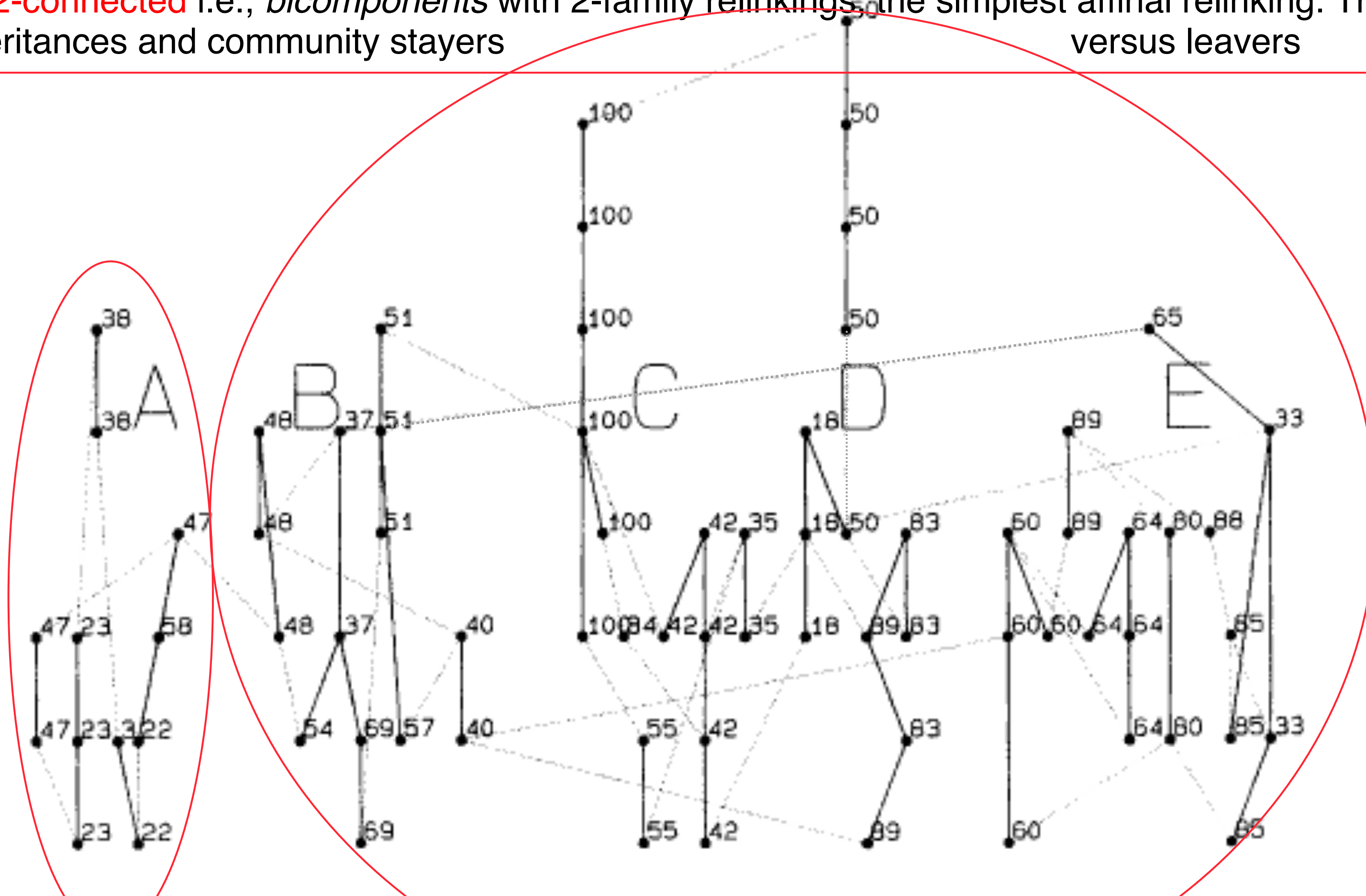


Figure 6. Two-family relinkings in Feistritz.

Structural Endogamy and Social Class: Carinthian Farmers of

Feistritz: Comparison of Relinking Frequencies for Actual and Simulated Data (*=actual freqs greater)

Number of Structurally Endogamous Marriages						
Generation	1	2	3	4	5	6
Present:	by Ancestral Levels					
Actual	8*	16*	70*	179	257	318
Simulated	0	0	32	183	273	335
Back 1 gen:	by Ancestral Levels					
Actual	8*	58*	168	246	308	339
Simulated	0	18	168	255	320	347
Back 2 gen:	by Ancestral Levels					
Actual	26*	115*	178	243	278	292
Simulated	0	98	194	262	291	310

Statistical conclusion comparing actual to random behavior: conscious relinking among families creates structural endogamy

Source: 1997 "Class, Property and Structural Endogamy: Visualizing Networked Histories," Theory and Society 25:161-208. Lilyan Brudner and Douglas White

New and potential applications (repeat)

- **Biologists** have begun to use these methods to study consequences of cohesive behavior in animal studies.
- **Historians** now use these methods to study consequences of cohesive structure in urban communities.
- **Biomedical** study of protein and interactive neural and physiological networks are good candidates for study of consequences of cohesive dynamics.
- **Evolutionary cohesive dynamics** now studies how groups form to create successful adaptive strategies.