SNA

Network data collection

Components of a social network

- Actors
  - Attributes?
    - E.g. demographic, psychological/personality traits
- Relationships
  - What are the attributes of relationship ?

Attribute data

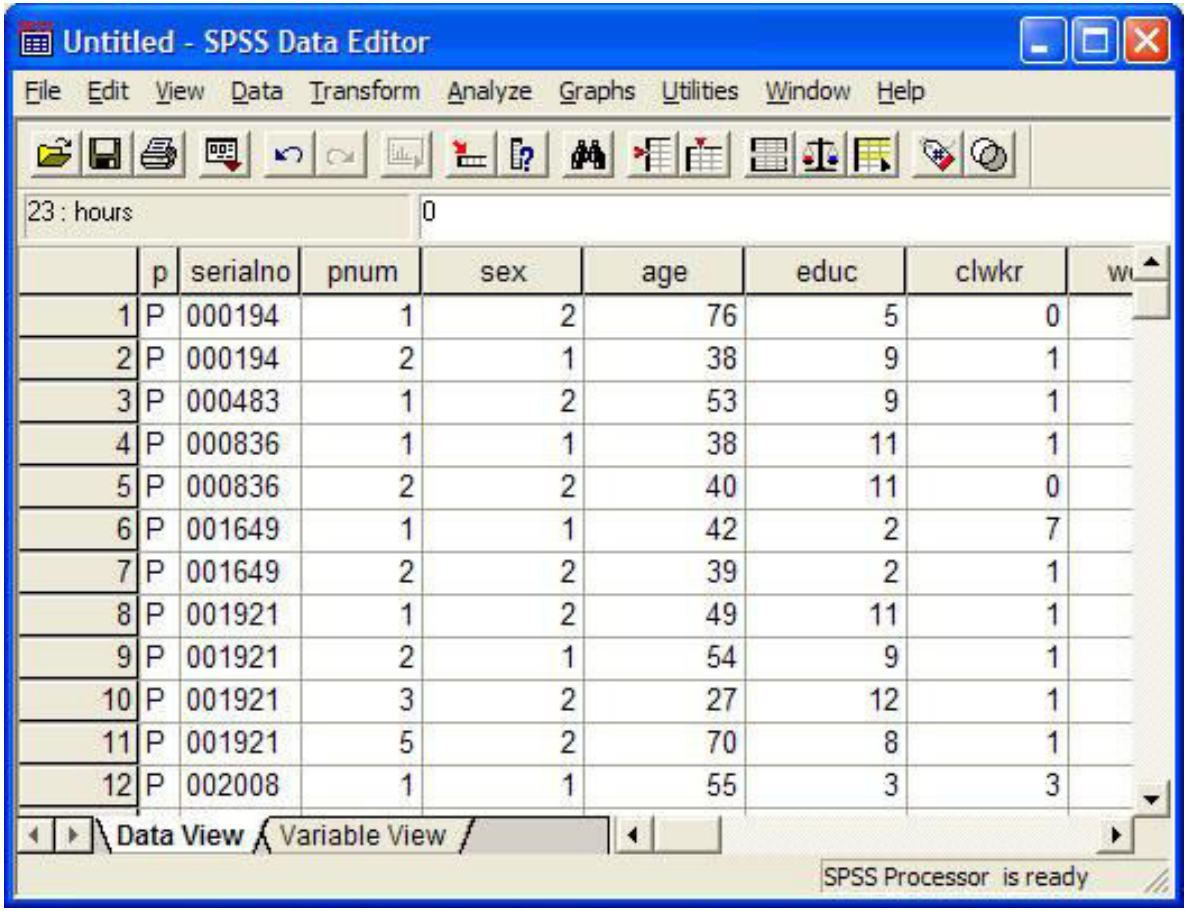
- Conventional data type-rectangular array

Attributes

Name	Sex	Age	occupation
Bob	Male	32	Teacher
Carol	Female	27	waiter
Ted	Male	29	Salesman
Alice	Female	28	airhostess

Subjects

Rectangular data array of case attributes



	p	serialno	pnum	sex	age	educ	clwkr	w
1	P	000194	1	2	76	5	0	
2	P	000194	2	1	38	9	1	
3	P	000483	1	2	53	9	1	
4	P	000836	1	1	38	11	1	
5	P	000836	2	2	40	11	0	
6	P	001649	1	1	42	2	7	
7	P	001649	2	2	39	2	1	
8	P	001921	1	2	49	11	1	
9	P	001921	2	1	54	9	1	
10	P	001921	3	2	27	12	1	
11	P	001921	5	2	70	8	1	
12	P	002008	1	1	55	3	3	

What is relational data?

- Relation
  - Data concerning relations between things (e.g. agents)
  - rather than the individual properties/attributes of those things.

Relational data

- Property of two or more units,
- Relations among units

Example:

- Interaction among individual
- Distance between locations
- Correlation between variables
- Similarity/occurrence between books/music/moves
- Diffusion of innovation between organizations

Social network data

- Network Data type-square array

	Choice			
Chooser	Bob	Carol	Ted	Alice
Bob	—	0	1	1
Carol	1	—	0	1
Ted	0	1	—	1
Alice	1	0	0	—

- Actors are described by their relation, PLUS attributes,

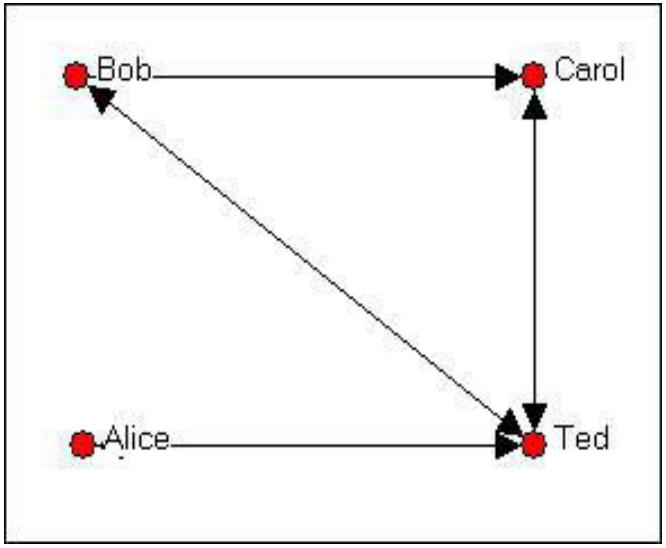
Adjacency matrix

is a matrix which describes a graph by representing which vertices are adjacent to which other vertices.

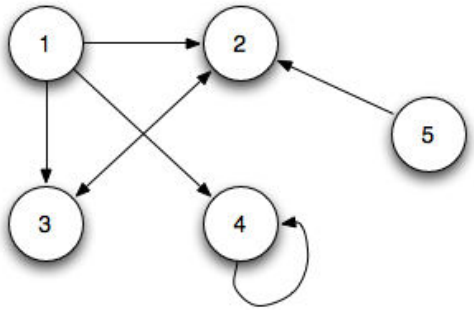
	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

Sociogram

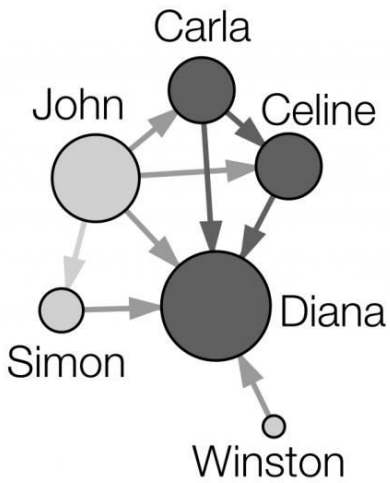
A map of relationships (formal or informal) among actors (person, organization, and others)



Edge versus Matrix data model



	1	2	3	4	5
1	0	1	1	1	0
2	0	0	1	0	0
3	0	1	0	0	0
4	0	0	0	1	0
5	0	1	0	0	0



Nodes

Id,Label,Attribute
1,John,1
2,Carla,2
3,Simon,1
4,Celine,2
5,Winston,1
6,Diana,2

Edges

Source,Target
1,2
1,3
1,4
1,6
2,4
2,6
3,6
4,6
5,6

Attributes of ties/edges

- Nature of ties
  - Content (e.g. personal, formal, collaborative…)
  - Multiple relationship between ties
- Direction of ties
  - Directed (symmetrical or asymmetrical, e.g. give or receive)
  - Undirected (co-authorship, joint membership)
- Strength of ties
  - (e.g. intimacy, frequency, duration…)

A typology of relationship studied in SNA

Similarities			Social Relations				Interactions	Flows
Location	Membership	Attribute	Kinship	Other Role	Affective	Cognitive		
e.g., Same spatial & temporal space	e.g., Same clubs, Same events etc.	e.g., Same gender, Same attitude etc.	e.g., Mother of, Sibling of	e.g., Friend of, Boss of, Student of, Competitor	e.g., Likes, Hates, etc.	e.g., Knows, knows about, sees as happy, etc.	e.g., Sex with, Talked to, Advice to, Helped, Harmed, etc.	e.g., Information, beliefs, personnel, resources, etc.

Figure 3. A typology of ties studied in social network analysis.

Table 1.3 Relations among firms.

Type	Firms as entities	Via individuals
Similarities	Joint membership in trade association; co-located in Silicon Valley	CEO of organization A sits on same board as CEO of organization B
Relations	Joint ventures; alliances; distribution agreements; own shares in; regards as competitor	Chief scientist of A is friends with chief scientist of B
Interactions	Sells product to; makes competitive move in response to	Representatives of A attend same conference as representatives of B
Flows	Technology transfers; cash infusions	Employee of A leaks information to employee of B

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. SAGE Publications Limited.

Other evidence for connection: occurrence

11.19.2001. Meeting at Brussels. Attending:
Smith, Johnson, Davis, James, Jackson

12.22.2001. Meeting at Paris. Attending:
Johnson, James, Jones, **Wilson**

1.12.2001. Meeting in New York. Attending:
Jones, **Carter**, Burns

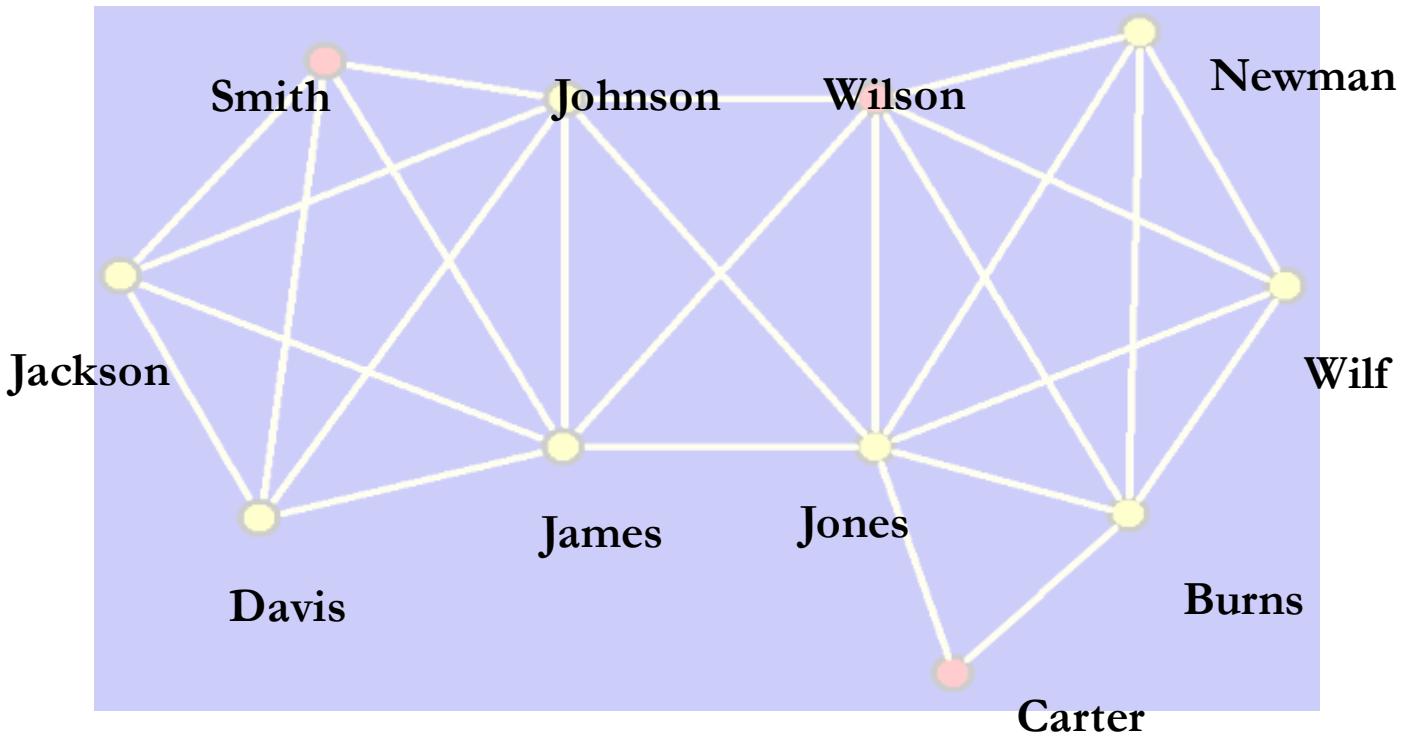
2.14.2001. Meeting in Denver. Attending:
Wilson, Burns, Wilf, Newman

(Red **bold** indicates people who are the focus of an investigation)

Adopted from Ernesto Estrada

Visualization

While perhaps not immediately apparent when looking at the list of names, a simple algorithm reveals connections among these actors.



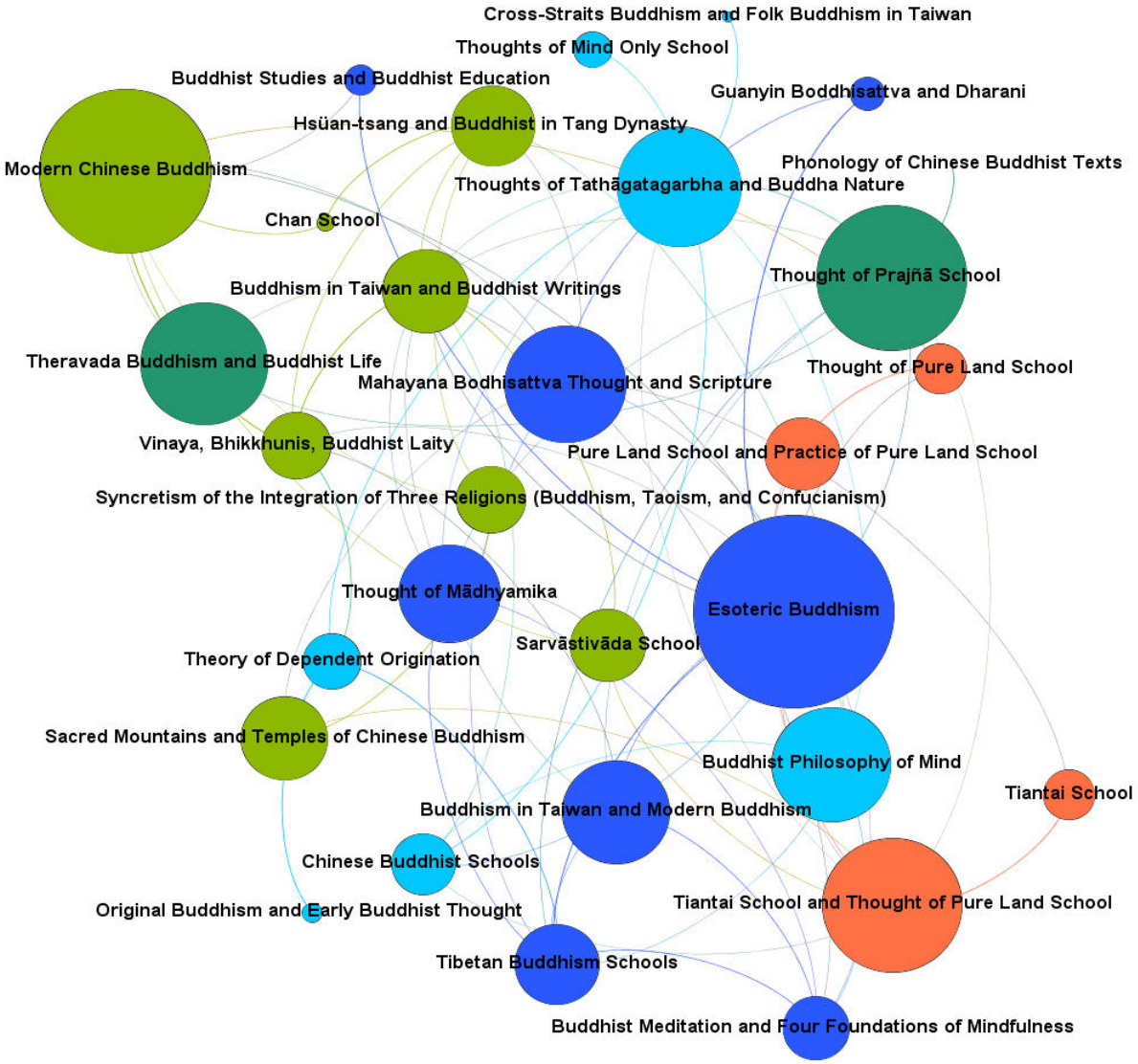
Adopted from Ernesto Estrada

Term-term similarity matrix

	Term1	Term2	Term3	Term4	Term5	Term6	Term7	Term8
Term 1		7	16	15	14	14	9	7
Term 2	7		8	12	3	18	6	17
Term 3	16	8		18	6	16	0	8
Term 4	15	12	18		6	18	6	9
Term 5	14	3	6	6		6	9	3
Term 6	14	18	16	18	6		2	16
Term 7	9	6	0	6	9	2		3
Term 8	7	17	8	9	3	16	3	

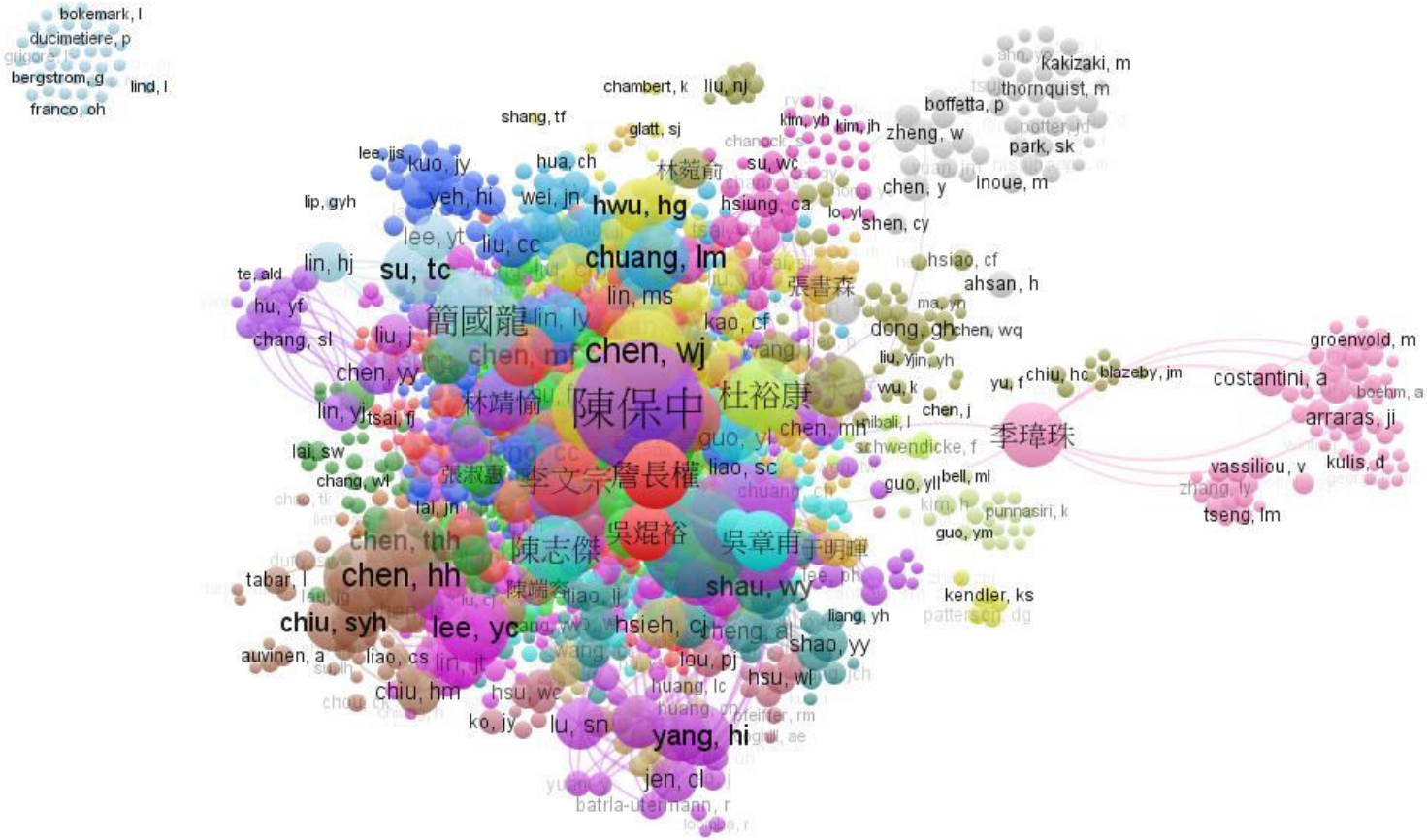
Visual thesaurus

Network of subdomains in Chinese Buddhist studies (1957–1986)



Co-authorship network of NTU public health

1,559 among 5,337 authors had published over 3 articles and had co-authorship with other authors (divided into 24 groups)



17

Co-occurrence

Two-mode to one mode transformation

	Book 1	Book 2	Book 3	Book 4	Book 5	Book 6
Customer A	X			X		
Customer B		X	X		X	
Customer C		X	X			
Customer D		X				X
Customer E	X				X	
Customer F			X		X	

		Also bought…					
		Book 1	Book 2	Book 3	Book 4	Book 5	Book 6
Customers who bought…	Book 1				1	1	
	Book 2			2		1	1
	Book 3		2			2	
	Book 4	1					
	Book 5	1	1	2			
	Book 6		1				

Geographic distances

Airline distances between US cities

	ATLANTA	CHICAGO	DENVER	HOUSTON	LA	MIAMI	NY	SF	SEATTLE	DC
ATLANTA		587	1212	701	1936	604	748	2139	2182	543
CHICAGO	587		920	940	1745	1188	713	1858	1737	597
DENVER	1212	920		879	831	1726	1631	949	1021	1494
HOUSTON	701	940	879		1374	968	1420	1645	1891	1220
LA	1936	1745	831	1374		2339	2451	347	959	2300
MIAMI	604	1188	1726	968	2339		1092	2594	2734	923
NY	748	713	1631	1420	2451	1092		2571	2408	205
SF	2139	1858	949	1645	347	2594	2571		678	2442
SEATTLE	2182	1737	1021	1891	959	2734	2408	678		2329
DC	543	597	1494	1220	2300	923	205	2442	2329	

Space distance vs. social distance

Possible directed ties between a pair

- A gives something to B, but B doesn't give A
- B gives A, but not the reverse
- Both A and B exchange one another
- Neither A nor B transacts

Directed and undirected network

- In an **undirected** network, the links are reciprocal—so if A is connected to B, B is by definition connected to A.

- In a **directed** graph or network, the links are not necessarily reciprocal—A may be connected to B, but B may not be connected to A (think of a graph with arrows indicating direction of the edges.)

Some relations are ‘undirected’ (e.g. Join the same committee, co-authorship). This is recorded in the matrix.

	Tom	Sally	Fred	Alice
Tom		1		
Sally	1			
Fred				
Alice				

Some relations are 'directed' (e.g. Liking or consult knowledge) and thus not necessarily reciprocal. This is recorded in the matrix. Tom likes Sally but she doesn't like him. She likes Alice. Or citing and cited

Target

Source

	Tom	Sally	Fred	Alice
Tom		1		
Sally	?			1
Fred				
Alice	1			

- In this case Tom's relation to Sally and hers to him are distinct and should be treated independently

Is this a directional network?

	1 Coun	2 Comm	3 Educ	4 Indu	5 Mayr	6 WRO	7 News	8 UWay	9 Welf	10 West
1 Coun	---	1	0	0	1	0	1	0	1	0
2 Comm	1	---	1	1	1	0	1	1	1	0
3 Educ	0	1	---	1	1	1	1	0	0	1
4 Indu	1	1	0	---	1	0	1	0	0	0
5 Mayr	1	1	1	1	---	0	1	1	1	1
6 WRO	0	0	1	0	0	---	1	0	1	0
7 News	0	1	0	1	1	0	---	0	0	0
8 UWay	1	1	0	1	1	0	1	---	1	0
9 Welf	0	1	0	0	1	0	1	0	---	0
10 West	1	1	1	0	1	0	1	0	0	---

Figure 13.2. Adjacency matrix for Knoke information network

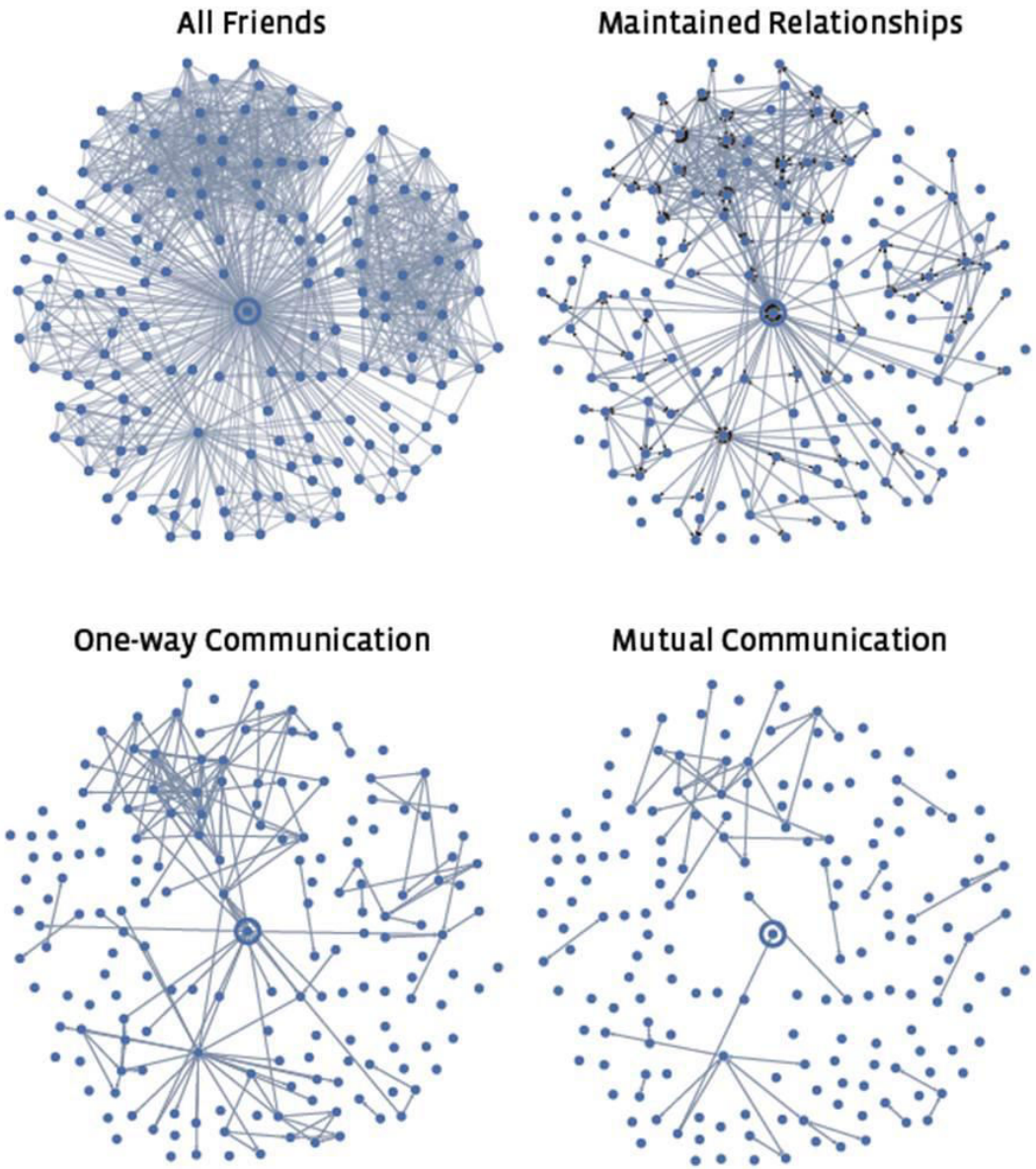
Strength of ties

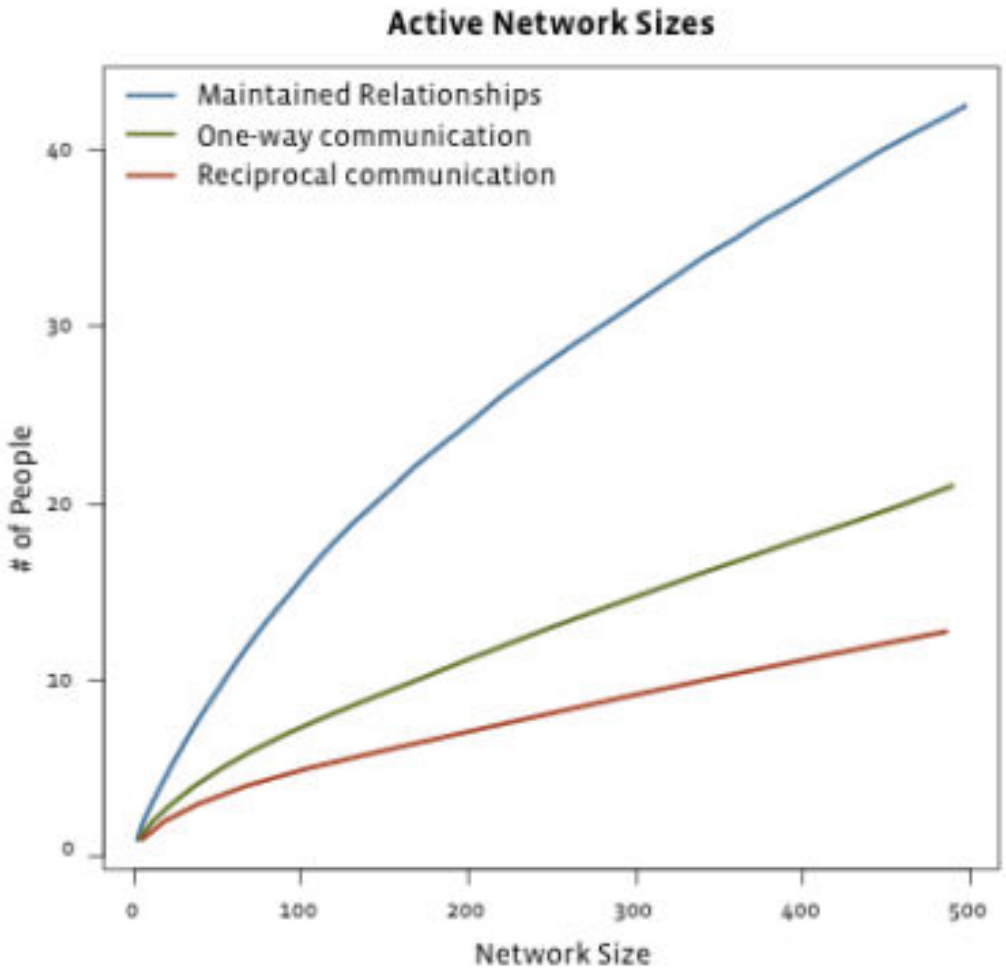
- Values can be attached to ties, signifying quantitative attributes such as
  - Strength of relationship (e.g. co-citation, co-occurrence etc.)
  - Rate of information flow
  - Probability of passing on information
  - Frequency of interaction

Tie strength: relations may be weighted in ordinal/interval manner: e.g. 0 = 'Don't like', 1='like', 2='really like'; or telephones *n* times per week.

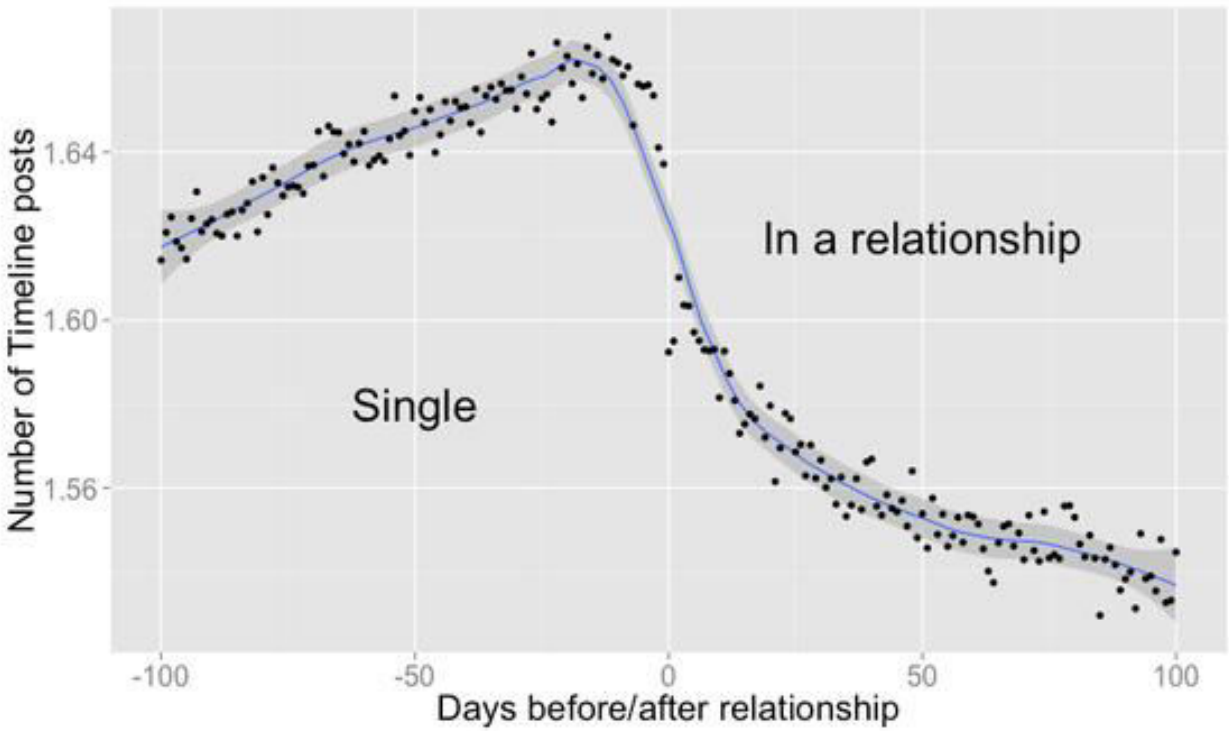
	Tom	Sally	Fred	Alice
Tom	0	1	5	4
Sally	2	0	3	5
Fred	3	2	0	8
Alice	5	3	0	0

Network at different thresholds of interaction




Active Network Sizes
Maintained Relationships
One-way communication
Reciprocal communication
# of People
0
10
20
30
40
0
100
200
300
400
500
Network Size

The evolving of relationships



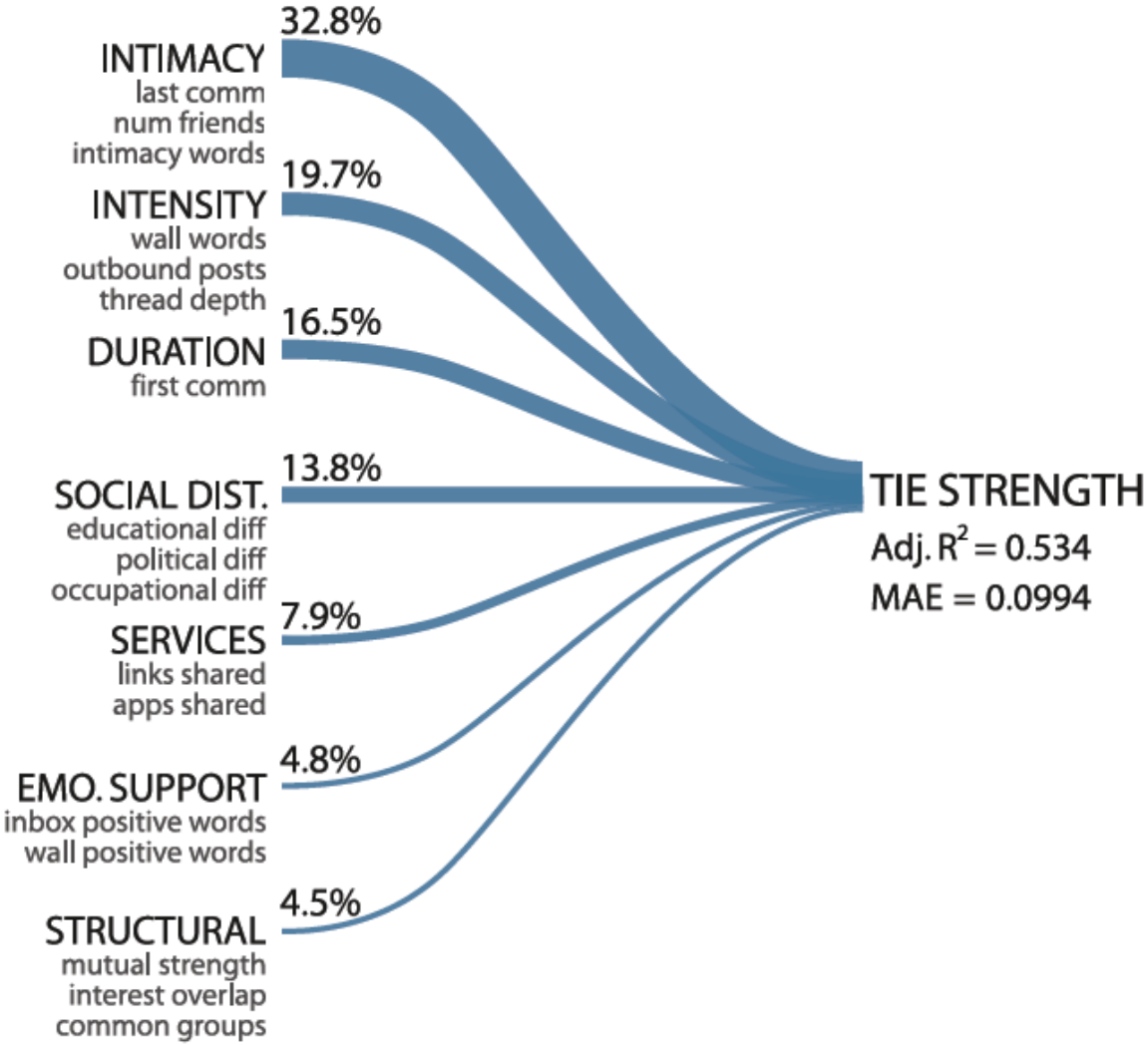
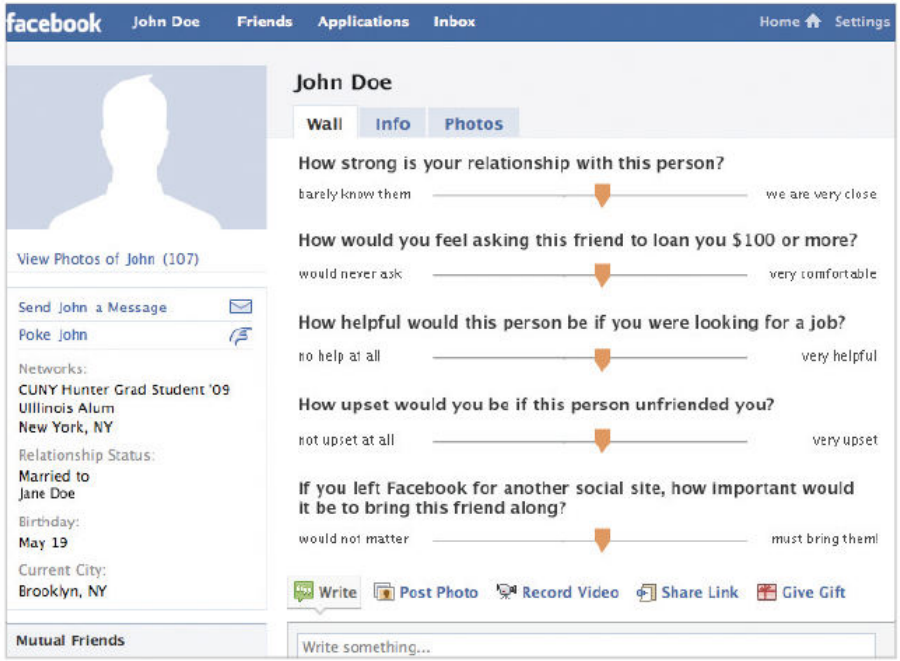


Figure 3. The predictive power of the seven tie strength dimensions, presented here as part of the *How strong?* model. A dimension's weight is computed by summing the absolute values of the coefficients belonging to it. The diagram also lists the top three predictive variables for each dimension. On average, the model predicts tie strength within one-tenth of its true value on a continuous 0–1 scale.

Tie strength



How strong is your relationship with this person?

barely know them ——— we are very close

How would you feel asking this friend to loan you $100 or more?

would never ask ——— very comfortable

How helpful would this person be if you were looking for a job?

no help at all ——— very helpful

How upset would you be if this person unfriended you?

not upset at all ——— very upset

If you left Facebook for another social site, how important would it be to bring this friend along?

would not matter ——— must bring them

Figure 1. The questions used to assess tie strength, embedded into a friend's profile as participants experienced them. An automated script guided participants through a random subset of their Facebook friends. As participants answered each question by dragging a slider, the script collected data describing the friendship. The questions reflect a diversity of views on tie strength.

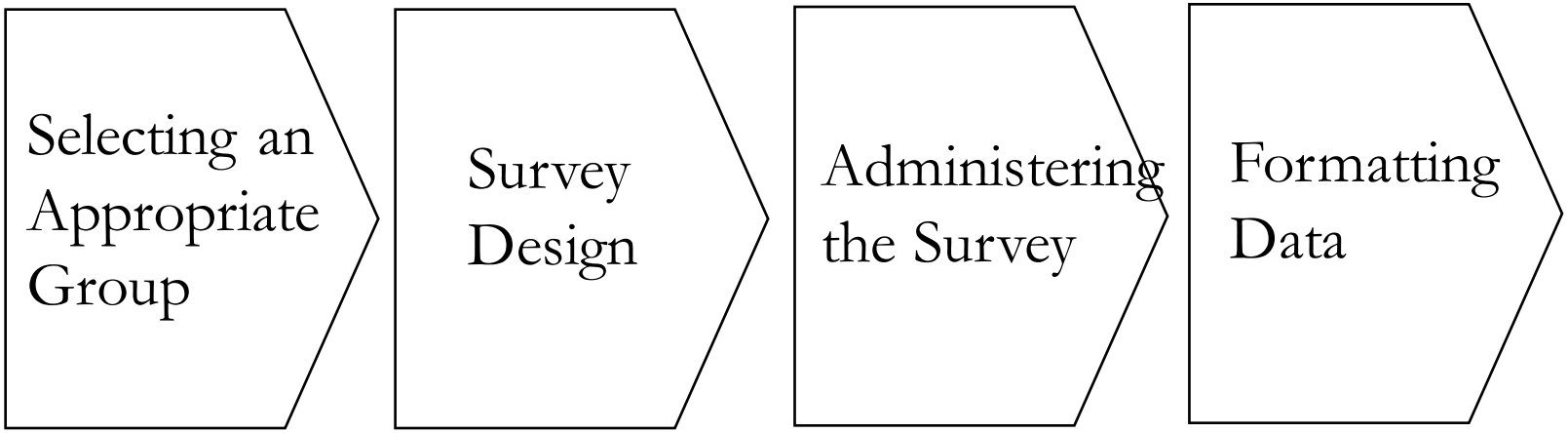
Top 15 Predictive Variables	β	F	p-value
Days since last communication	-0.76	453	< **0.001**
Days since first communication	0.755	7.55	< **0.001**
Intimacy × Structural	0.4	12.37	< **0.001**
Wall words exchanged	0.299	11.51	< **0.001**
Mean strength of mutual friends	0.257	188.2	< **0.001**
Educational difference	-0.22	29.72	< **0.001**
Structural × Structural	0.195	12.41	< **0.001**
Reciprocal Serv. × Reciprocal Serv.	-0.19	14.4	< **0.001**
Participant-initiated wall posts	0.146	119.7	< **0.001**
Inbox thread depth	-0.14	1.09	0.29
Participant's number of friends	-0.14	30.34	< **0.001**
Inbox positive emotion words	0.135	3.64	**0.05**
Social Distance × Structural	0.13	34	< **0.001**
Participant's number of apps	-0.12	2.32	0.12
Wall intimacy words	0.111	18.15	< **0.001**

Table 3. The fifteen predictive variables with highest standardized beta coefficients. The two *Days since* variables have large coefficients because of the difference between never communicating and communicating once. The utility distribution of the predictive variables forms a power-law distribution: with only these fifteen variables, the model has over half of the information it needs to predict tie strength.

Multiple relations

- There can be multiple kinds of ties between actors
- Positions in one set of relations may re-enforce or contradict positions in another

Planning and administering a network analysis: the old fashion way



Populations, samples, and boundaries

- From full network
- Survey, e.g. snowball methods
  - Each is asked to name names,
    - Need to determine which of the nodes identified in the first stage are connected with one another. (Ego-centric networks with **alter connections**)
  - Named names are tracked down till no new actors are identified.
    - The initial nodes is are important, why?

Network boundary

- Formal definition
  - Network studies are more likely to include all the actors who occur within some (usually naturally occurring) boundary. E.g. the classroom, the LIS community, virtual communities etc.
- Relevance based
  - For example, in a word of mouth study, any referral can come across group boundary

1) *Complete*

- Well defined boundary
- Data on all actors within a particular (relevant) boundary
- Never exactly complete, but boundaries are set
- Example: Co-authorship data among all writers in the social sciences

2) Partial network

- Ego networks plus some amount of tracing to reach contacts of contacts

- Something less than full account of connections among all pairs of actors in the relevant population

- Example: CDC Contact tracing data for STDs
- Criminal and terrorist network

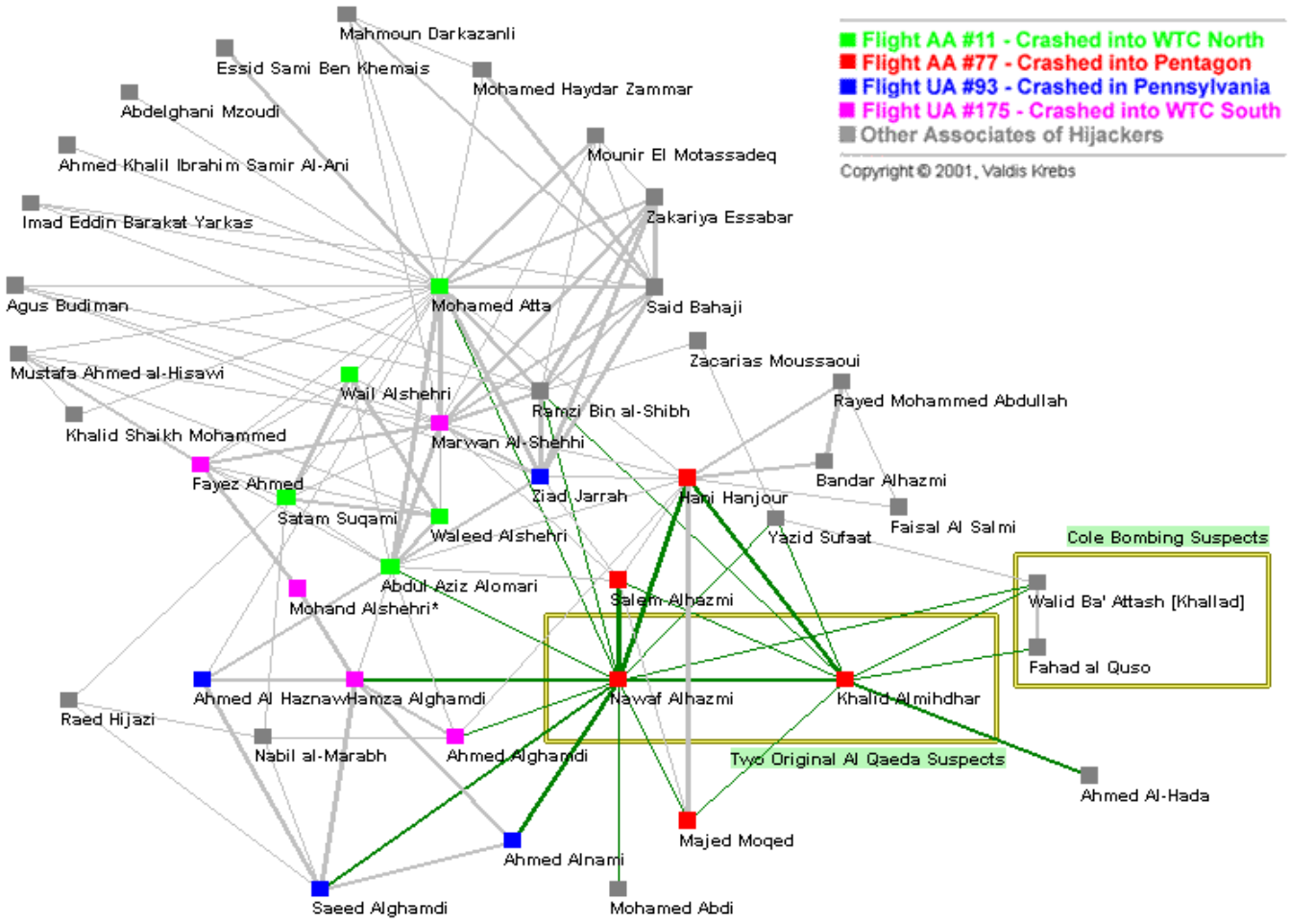


Figure 3 - All Nodes within 2 steps / degrees of original suspects

2) Ego-network

- Have data on a set of respondents (ego) and the people they are connected to (alters)
- May include estimates of connections among alters
- Data based on naming of alters associated with unconnected informants

Populations, samples, and boundaries

- From full network
- Survey, e.g. snowball methods
  - Each is asked to name names,
    - Need to determine which of the nodes identified in the first stage are connected with one another. (Ego-centric networks with alter connections)
  - Named names are tracked down till no new actors are identified.
    - The initial nodes is are important, why?

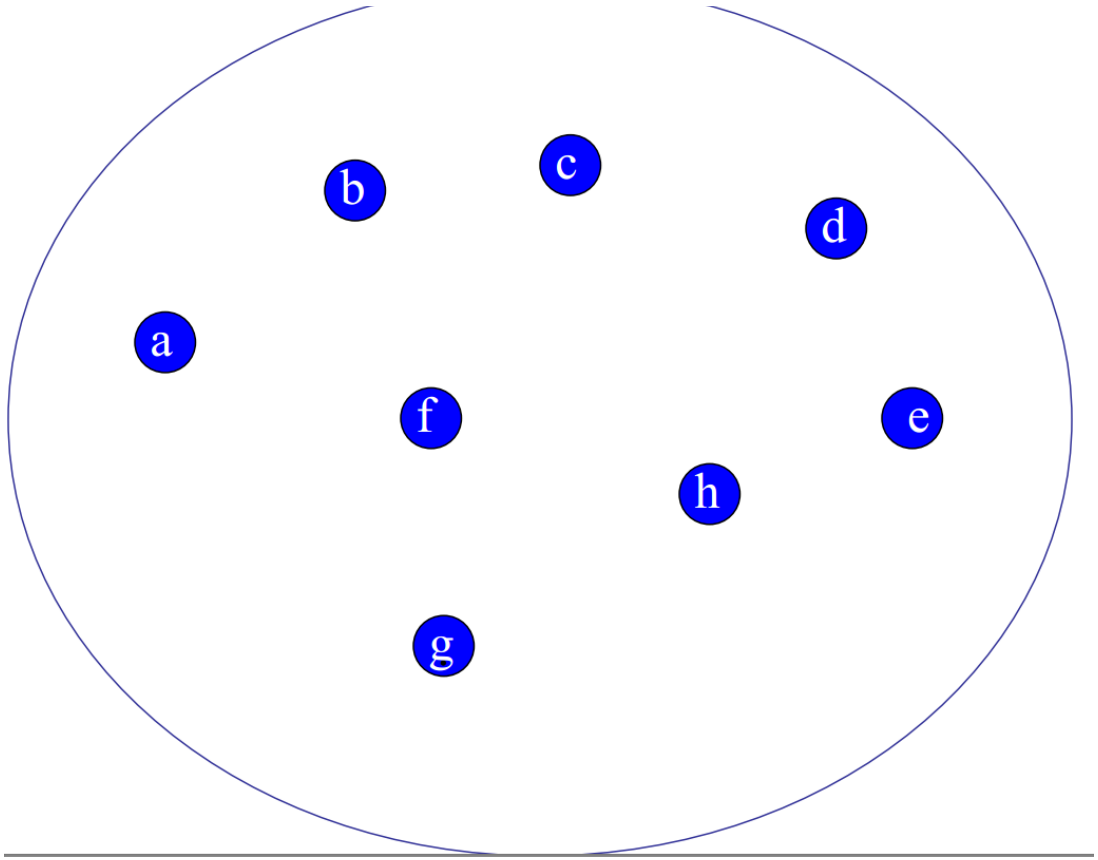
Whole network studies

- Boundary delamination approaches
  - Through formal definition
    - All the employees in the R&D department
    - All the faculty members that appear in the thesis committees of LIS thesis in the past five years
  - "Relational" strategy for boundary specification
    - Snowballing
      - A sampling of informants who offer more names
      - E.g. partners of drug users and prostitutes

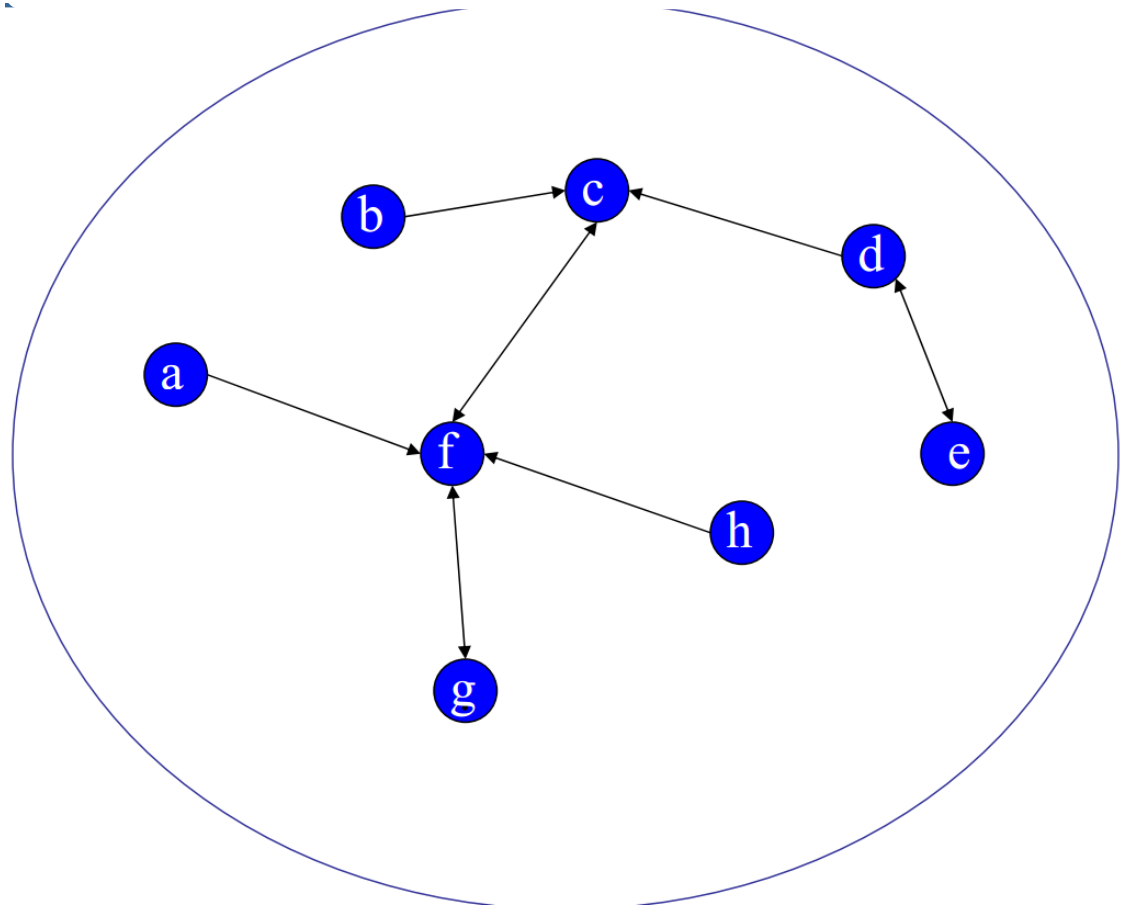
Populations, samples, and boundaries

- From full network
- Survey, e.g. snowball methods
  - Each is asked to name names,
    - Need to determine which of the nodes identified in the first stage are connected with one another. (Ego-centric networks with alter connections)
  - Named names are tracked down till no new actors are identified.
    - The initial nodes is are important, why?

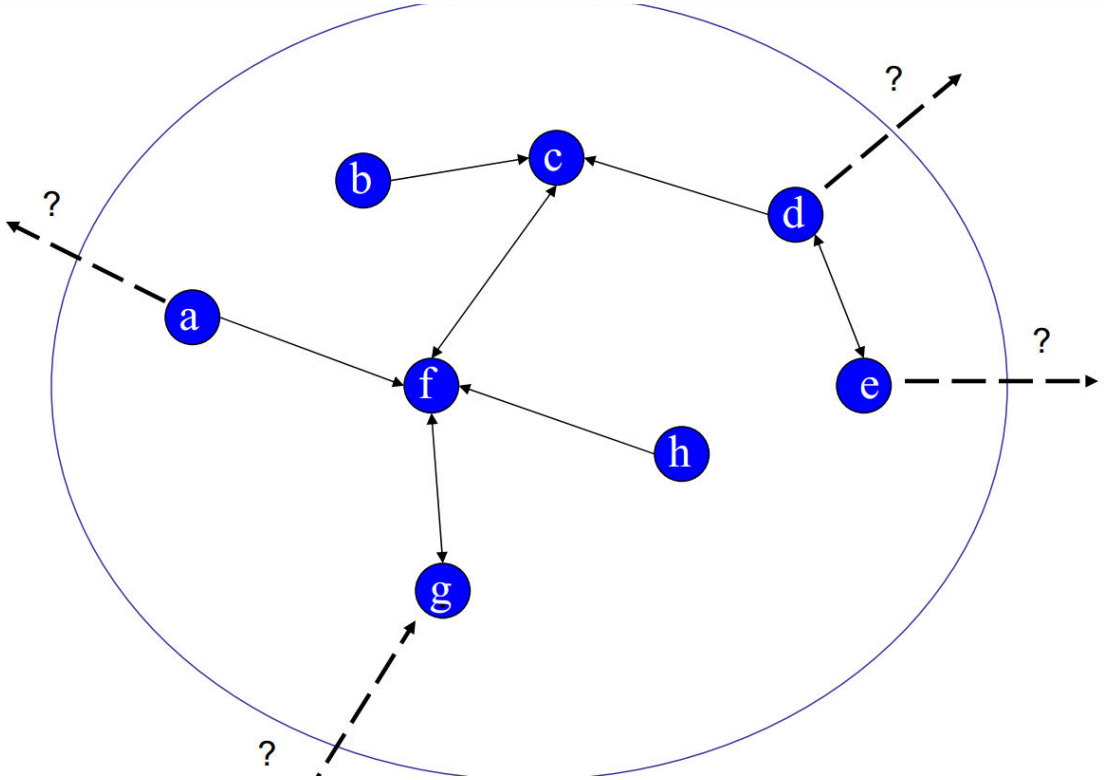
Population of actors



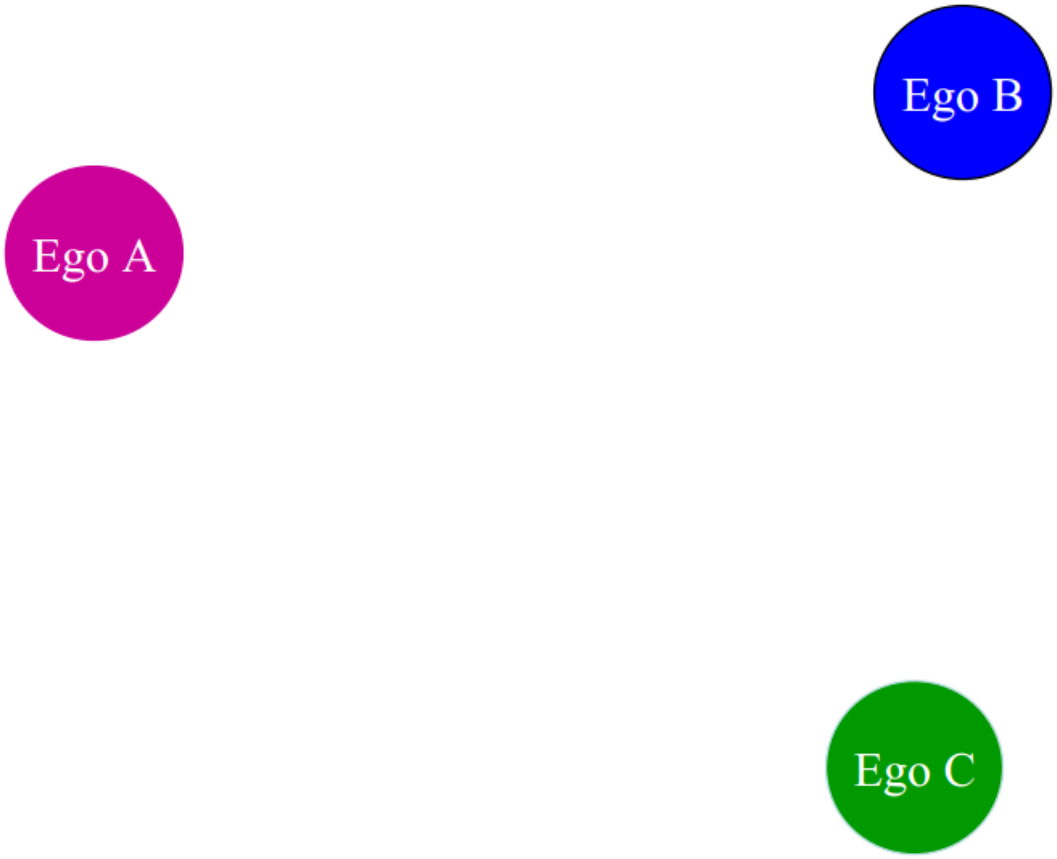
Ties between all actors



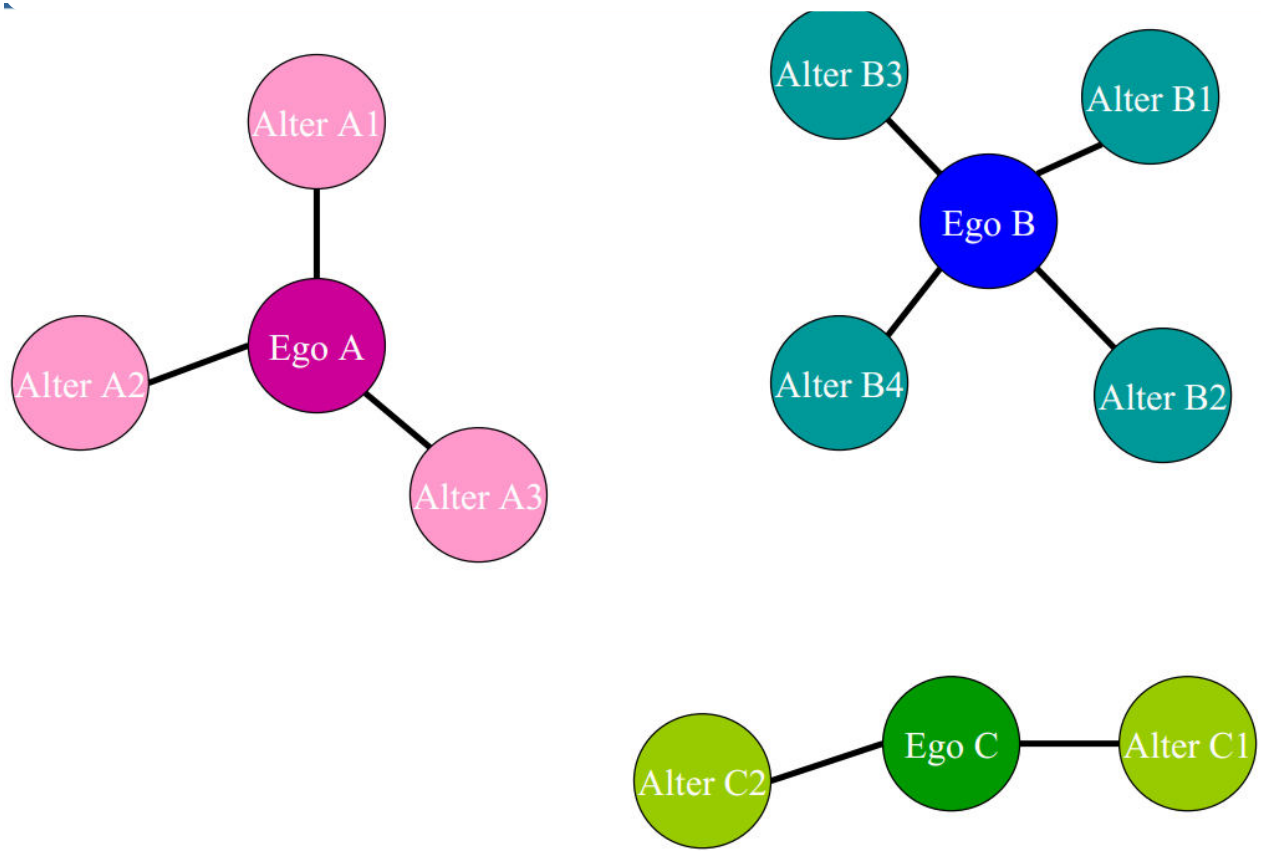
Unknown ties with outside



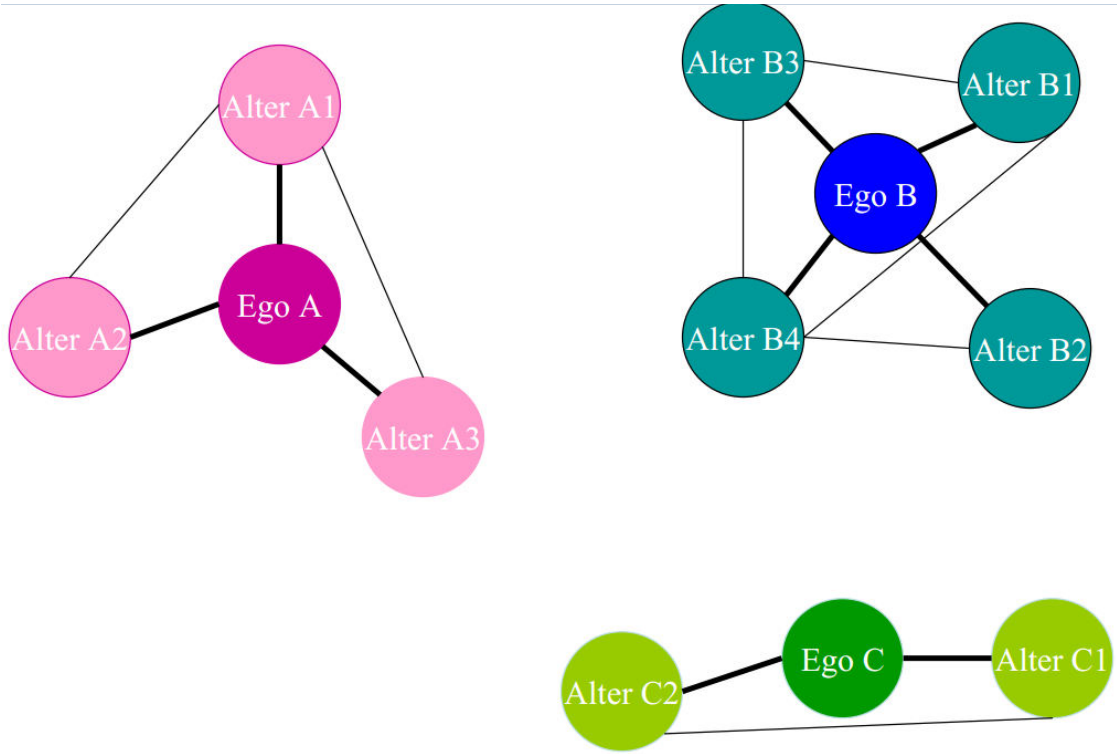
Snowballing: sample respondent "egos"



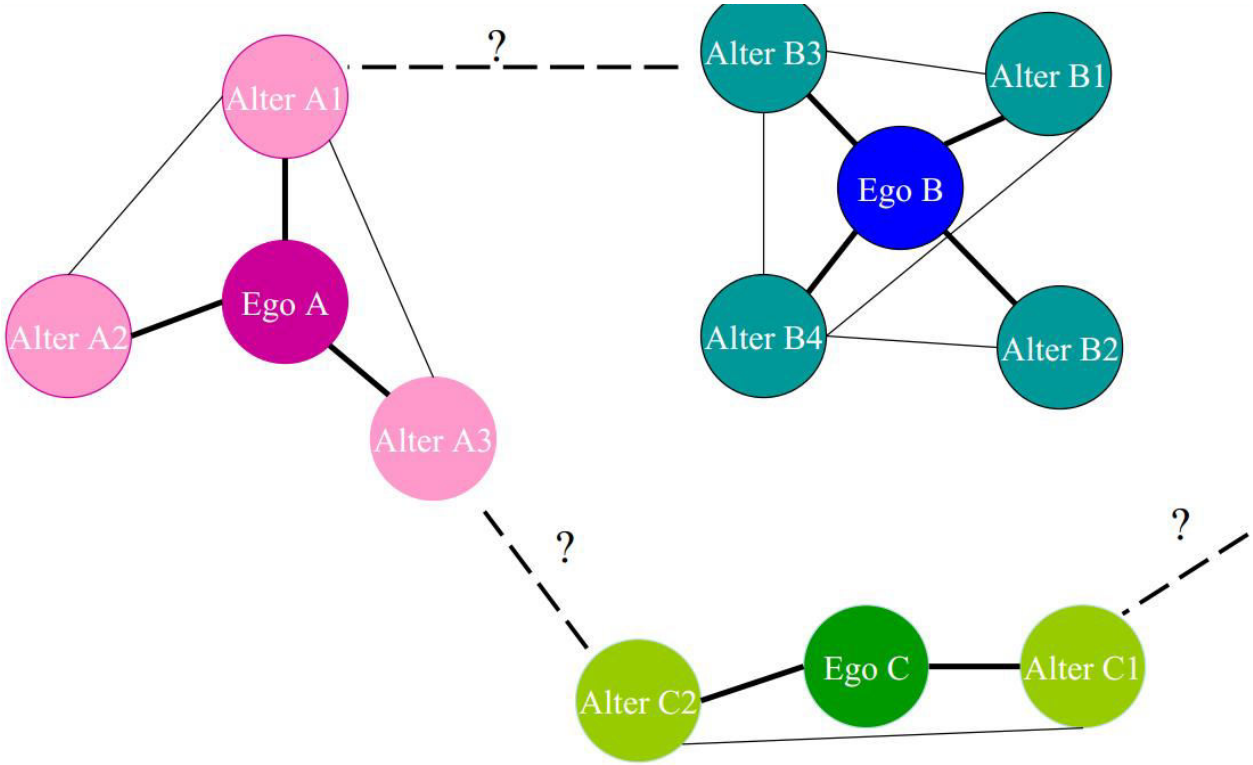
Elicit lists of “alters”



Ties between “alters”?



Unknown ties



Using questionnaire to collect relationship data

- Row-based
  - Each questionnaire forms one row in the adjacency matrix
- Row and Column based
  - Data collected in pairs
- Consensus based
  - The relationship between each pair judged by the group members/informants

Using questionnaire to collect relationship data

- Row-based
  - Each questionnaire forms one row in the adjacency matrix
- Row and Column based
  - Data collected in pairs
- Consensus based
  - The relationship between each pair judged by the group members/informants

Table 1.1 Examples of research questions by level of analysis and type of node.

	Type of node	
Level of analysis	**Individuals**	**Collectivities**
Dyad level $O(n^2)$	Are employees whose offices are near each other more likely to develop friendships than employees whose offices are further apart?	Are firms with similar organizational cultures more likely to form joint ventures with each other?
Node level $O(n^1)$	Are employees who are more central in their organization's friendship network less likely to leave for another company?	Are firms with more diverse technology partners more likely to introduce innovative products into the market?
Network level $O(n^0)$	When a network of employees is characterized by many redundant paths between all pairs of persons, is the network less disrupted by individuals leaving the firm?	When a network of firms is densely connected, does this place the network at greater risk of catastrophic failure (because of cascade effects)?

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. SAGE Publications Limited.

Table 1.4 Types of network studies classified by direction of causality and level of analysis.

	Network variables as independent/ explanatory	**Network variables as dependent/ outcomes**
Dyad level	Friendship between pairs of farmers to predict which pairs of farmers make the same decision about going organic	Similarity of interests (e.g., sky diving) to predict who becomes friends with each other
Node level	Centrality in organizational trust network to predict who is chosen for promotion	Extraversion to predict who becomes central in friendship network
Network level	Shortness of paths in a group's communication network to predict group's ability to solve problems	Type of organizational culture (emphasizing either cooperation or competition) to predict structure of the trust network

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. SAGE Publications Limited.

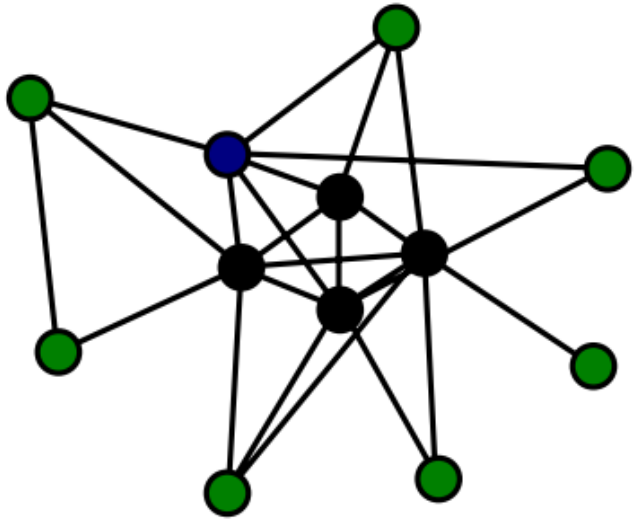
SNA

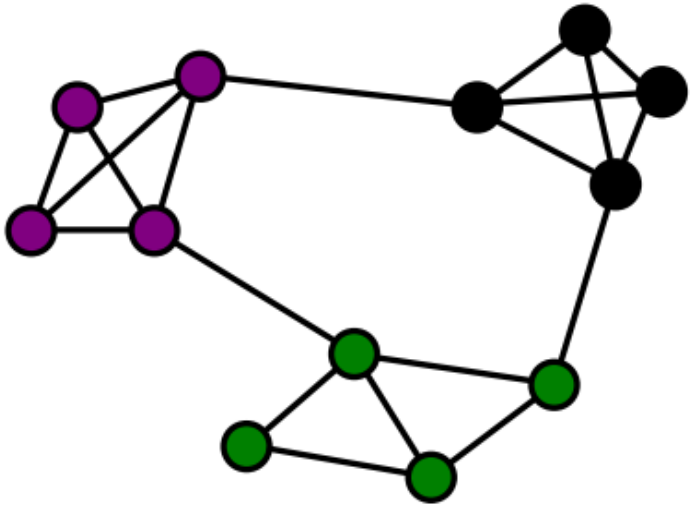
Graph, sociogram, cohesive and connectivity

Graph and metrics

- From intuition to apply algorithms to calculate mathematical measures of graph properties

Network characteristics





Network metrics

- Structural metrics provide a measurement of a structural property of a graph
  - Global metrics refer to a whole graph
  - Local metrics refer to a single node in a graph

e.g. social capital

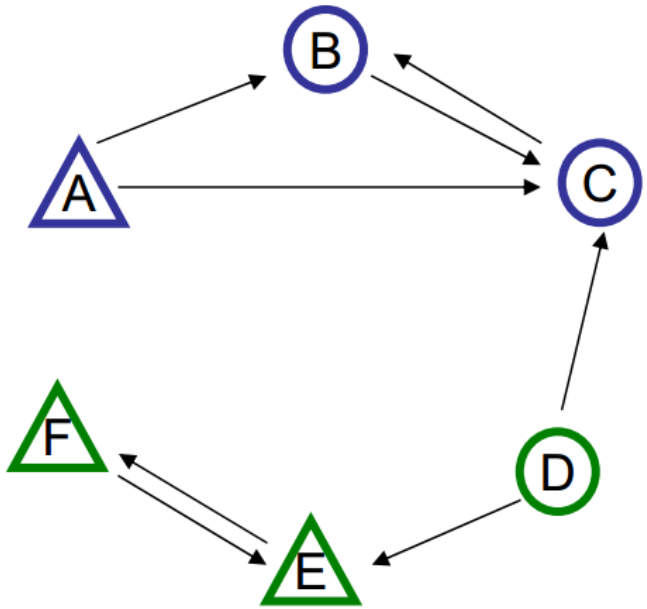
Basic components in a network

Graph arcs

(AB)
(AC)
(BC)
(CB)
(DC)
(DE)
(EF)
(FE)

a.k.a. edge list

Sociogram



Sociomatrix

	A	B	C	D	E	F
A	-	1	1	0	0	0
B	0	-	1	0	0	0
C	0	1	-	0	0	0
D	0	0	1	-	1	0
E	0	0	0	0	-	1
F	0	0	0	0	1	-

adjacency matrix

Which is more efficient representation of networks that are large and sparse?

Graph: definitions

- Vertex (nodes)
  - Basic element
  - Drawn as a *node* or a *dot*.
  - **Vertex set** of G is usually denoted by $V(G)$, or V
- Edge (lines)
  - A set of two elements
  - Drawn as a line connecting two vertices, called end vertices, or endpoints.
  - The edge set of G is usually denoted by E(G), or E.

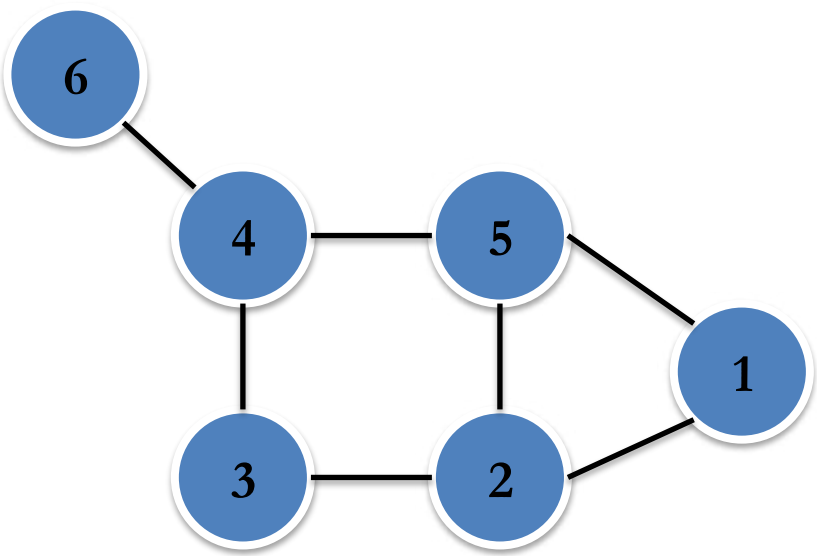
Graph

- A graph G = (V,E) is a data structure composed of:

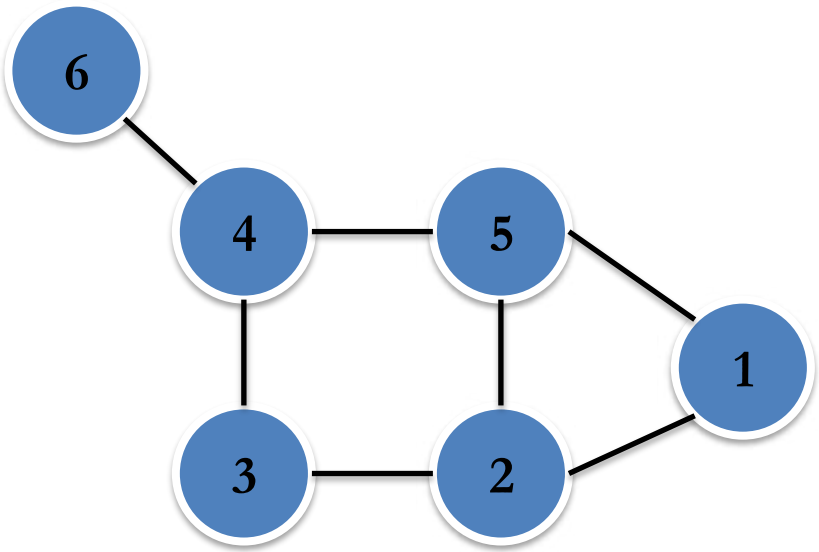
  V: set of vertices

  E: set of edges connecting the vertices in V

  An edge e = (u,v) is a pair of vertices



Example



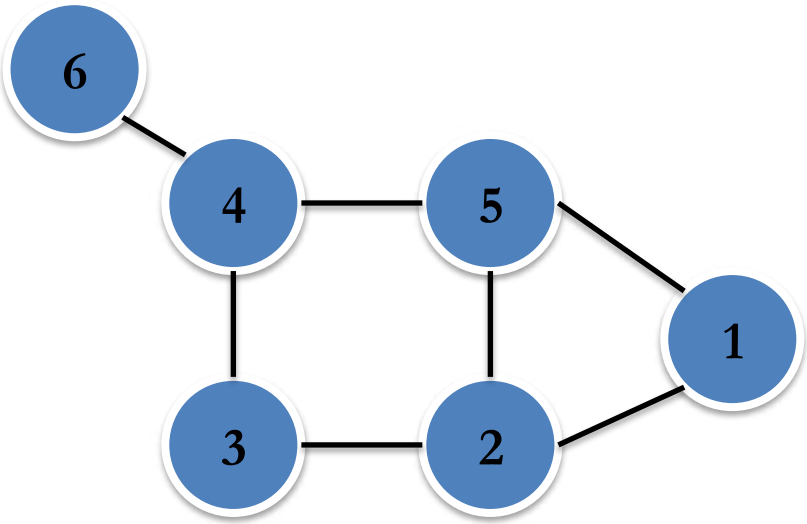
- A graph G = (V,E)
- V:={1,2,3,4,5,6}
- E:={{1,2},{1,5},{2,3},{2,5},{3,4},{4,5},{4,6}}

Nodal Degree (undirected graph)

- Number of edges associated with a node
- Sociologically, what does high nodal degree represent?

The degree of 5 is ?

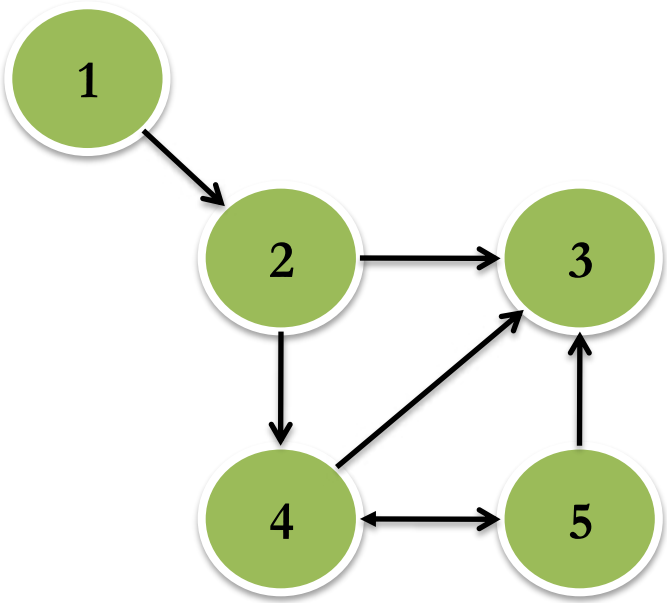
The maximum number of degrees any node in a N nodes network can have is?



If directional: in degree and out degree

Directed graph

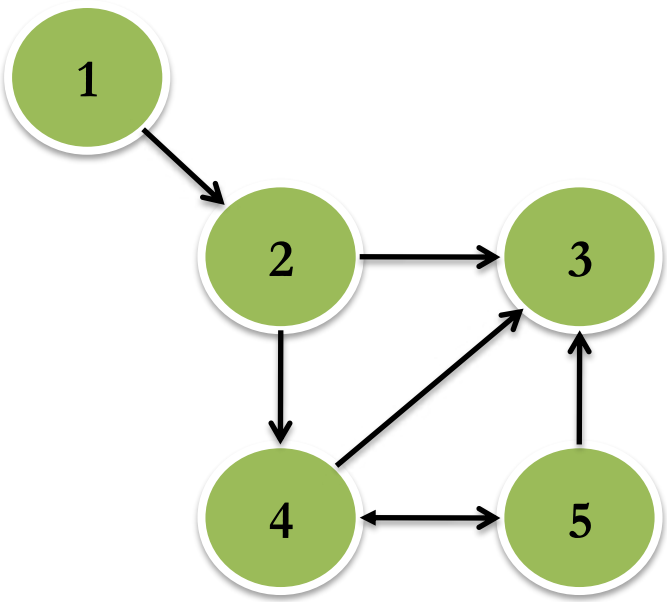
- When edges have direction



"In an undirected path, hyperlinks can be followed forward or backward, a technique available to search engine spiders but not to a person using a Web browser."

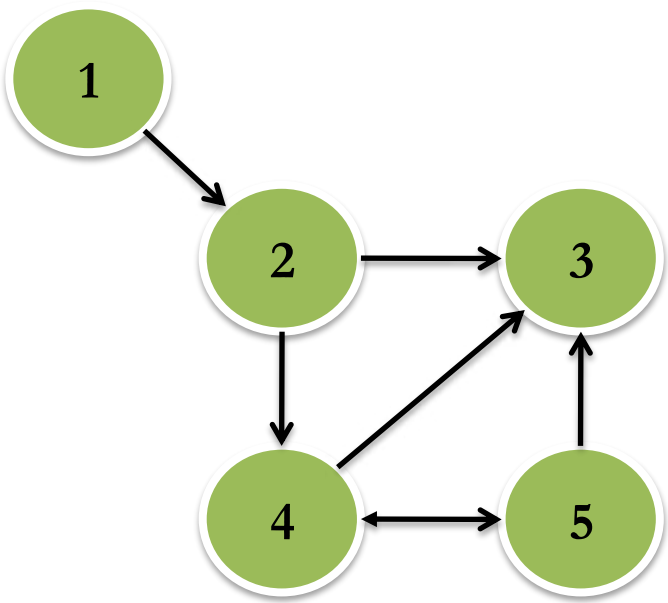
Formally...

- An undirected graph is one in which the pair of vertices in a edge is unordered, $(v_0, v_1) = (v_1, v_0)$
- A directed graph is one in which each edge is a directed pair of vertices, $\langle v_0, v_1 \rangle \mathrel{!=} \langle v_1, v_0 \rangle$



Degree (Directed Graphs)

- In-degree: Number of edges entering
- Out-degree: Number of edges leaving
- Total degree = indeg + outdeg



Outdeg (1)=1
indeg (1)=0

Outdeg (2)=2
indeg (2)= 1

Outdeg (3)=0
indeg (3)= ?

The maximum number of connections/degrees in a directional graph is?

Table 1. Nine newsgroups, with some basic statistics.

Name *January 2001 except as noted*	authors	replies	% threads with		% authors with	
			2 messages	5+ messages	degree > 10	1 message
Technical Newsgroups						
comp.soft-sys.matlab	437	712	28%	20%	3%	41%
microsoft.public.windows.server.general, *Jan. 2004*	855	1489	33%	10%	3%	51%
Discussion Newsgroups						
rec.kites	276	924	19%	32%	12%	32%
rec.music.phish	1263	6105	23%	22%	20%	29%
Political Newsgroup						
alt.politics	2187	16059	17%	40%	20%	26%
Flame Newsgroups						
alt.flame	618	3802	16%	34%	13%	32%
alt.alien-vampire.flonk.flonk.flonk	755	6494	17%	50%	16%	33%
Social Support Newsgroups						
alt.support.divorce	339	2937	14%	53%	21%	23%
alt.support.autism	188	3095	17%	60%	32%	20%

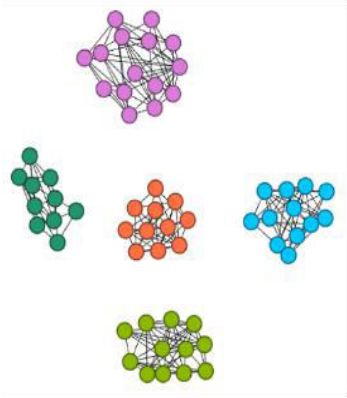
Fisher, D.; Smith, M.; Welser, H.T., "You Are Who You Talk To: Detecting Roles in Usenet Newsgroups," *System Sciences, 2006. HICSS '06. Proceedings of the 39th Annual Hawaii International Conference on*, vol.3, no., pp.59b,59b, 04-07 Jan. 2006
doi: 10.1109/HICSS.2006.536

Basic SNA constructs

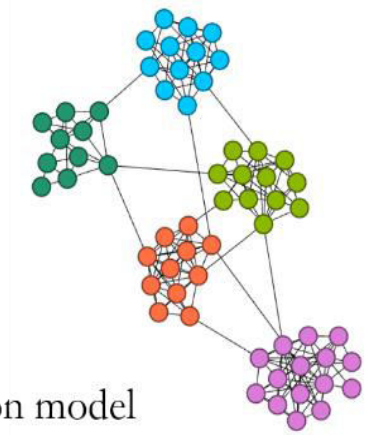
- Cohesion
  - Density, connectivity, clustering coefficient, Reachability, distance, , E-I etc.
- Centrality
  - Centralization
- Cluster/sub-network
  - Grouping cohesive sets of actors
  - Grouping according to structural equivalence

Models of network cohesion

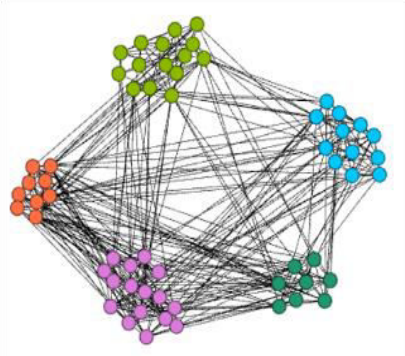
Plural-worlds



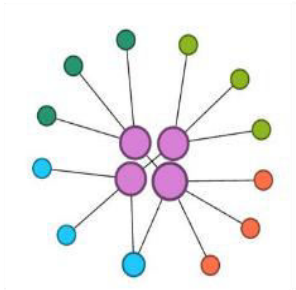
Small-world



Structural cohesion



Star production model



Cohesion

- Dictionary definition: the state of cohering, uniting, or sticking together
- Sociologically
  - Solidarity, shared norms, collective behaviors
  - Identify, as people who interact intensively are likely to consider themselves a social group.

Translate network properties into sociologically meaningful constructs

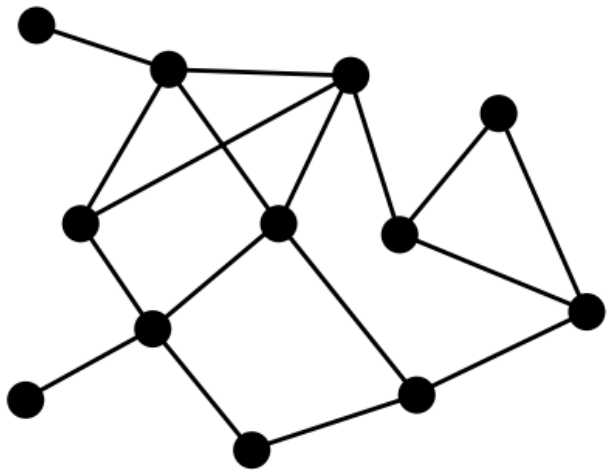
Fun intro to network density, broadly defined

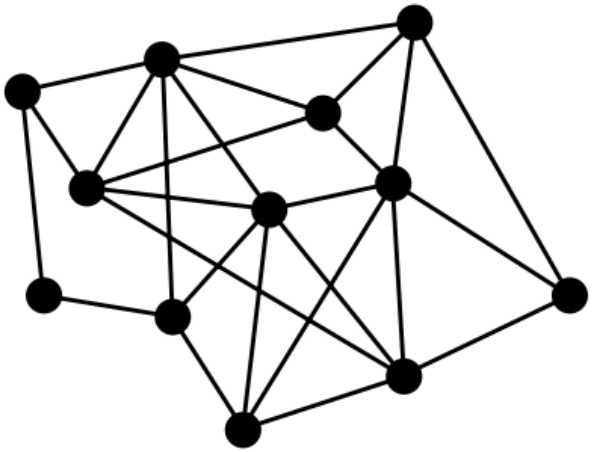
Measurement of network cohesion

- Density
  - Level of interconnectedness
- Distance
  - How fast/easily information spread
- Connectivity
  - How robust/cohesive a network is
  - Bridging, in-betweenness

Network density

- Percentage of number of ties to the total possible ties





Consider binary or valued; direct and non-directed

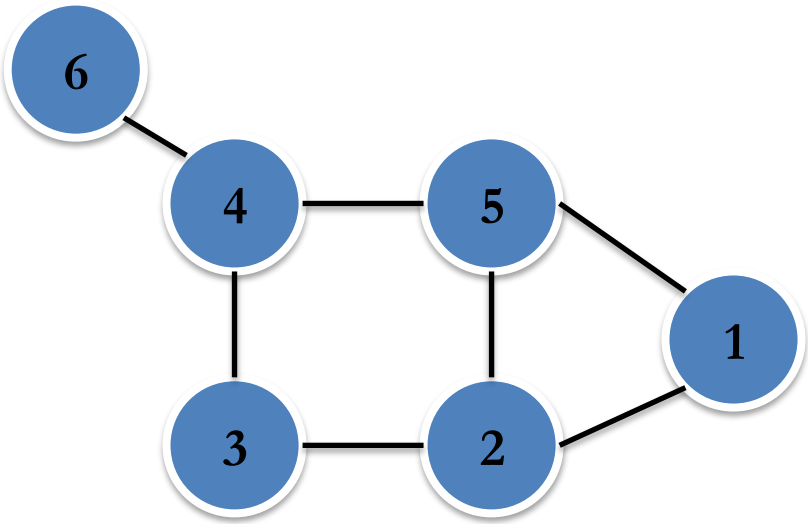
Density

- The actual number of connections in a network expressed as a proportion of the total possible number of connections.
- Sociologically, high density should generate greater: trust, cultural homogeneity and diffusion speed.
- Not easy to make meaningful comparisons of density across networks of different sizes (or involving different types of relations).

Density

- The proportion of all possible ties that are actually present.
- The maximum number of connections/degrees is?
- $2 * |E| / (|V| * (|V| - 1))$
- Network>Cohesion>Overall Density

Compare KNOKI AND KNOKM



Density in a valued network

- For a valued network it is the total of all values divided by the number of possible ties.
- Defined as the sum of all tie strength that is actually present to the number of possible ties.

Try Network>Cohesion>Overall Density

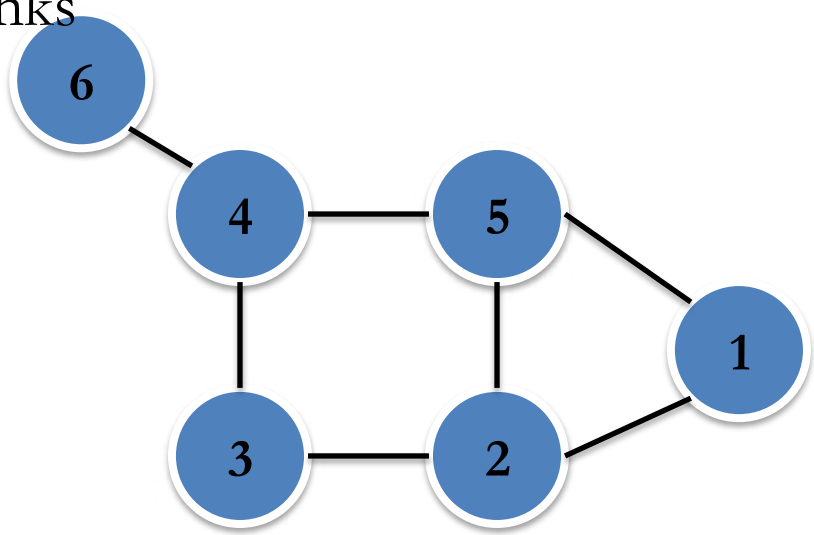
Simplified ntu/ntnu

Save into excel file then open with Gephi

Average degree

The average nodal degrees is **the sum of all the degrees** divided by the number of actors in the graph:
2*|E|/|V|.
Could be misleading because the wide variation in link distribution, :a few prominent nodes might possess disproportionally large amount of links

E: link
V: node



Notice we are talking about degree, not |E|, or number of "matches"

Global network attributes

- Density
  - Level of interconnectedness
- Reachability/Distance/Diameter
  - How fast information spread
- Connectivity
  - How robust/cohesive a network is

Walks, Trails, Paths and distance

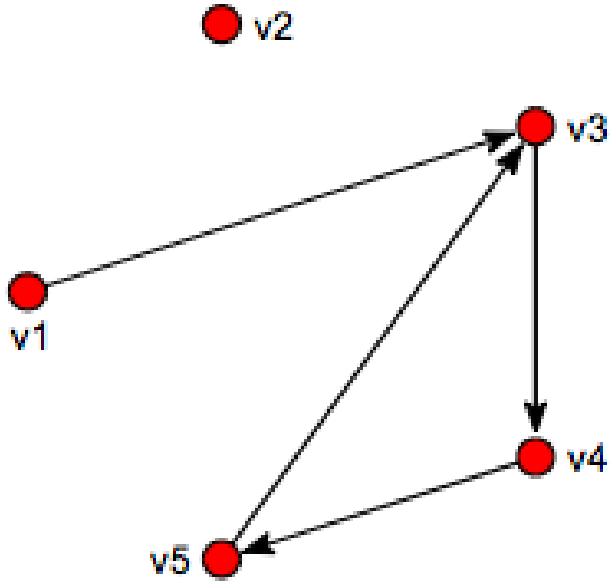


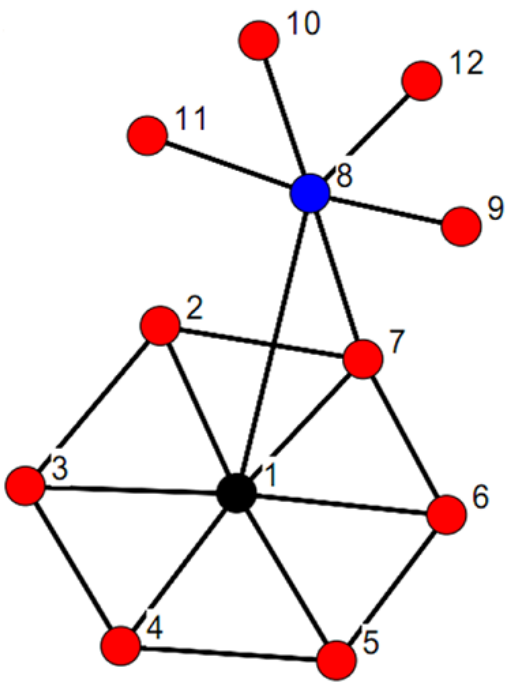
Figure 2 - A small simple directed network.

Consider lines as roads

How do we define "distance?" hint: the length of the lines between two nodes

Walk and its length

- A walk is an alternating sequence of nodes and lines in which each node is adjacent to its preceding and following lines
- All the nodes and lines can be traversed repeatedly
- The length of a walk is the number of lines it contains

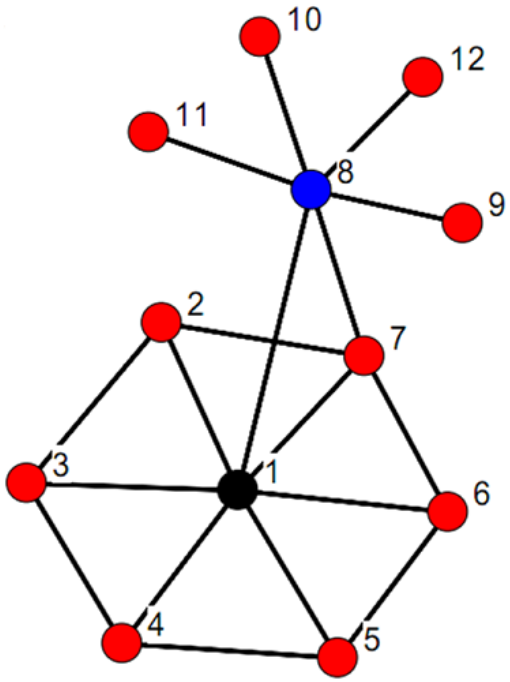


(Knoke and Yong, 2002, p. 47)

Path

- A walk with entirely distinct nodes and lines, i.e. no node or line can be included more than once

For example:
10-8-7-1-3
Is 10-8-7-1-8 a path?

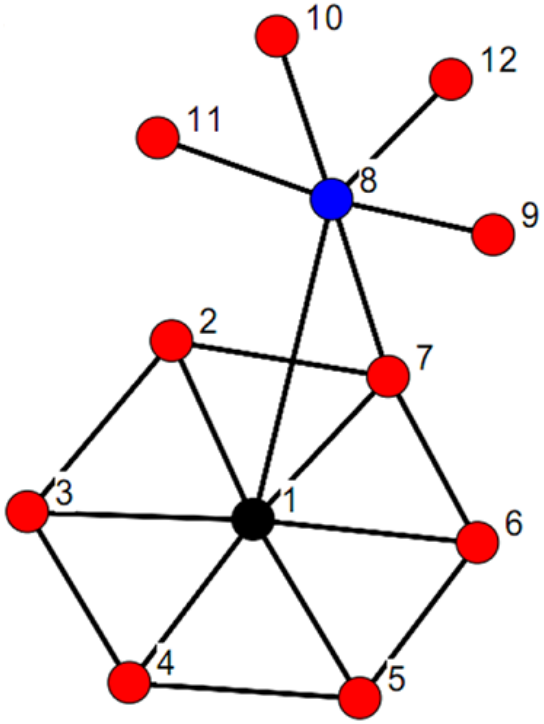


Adopted from Borgatti, 2010

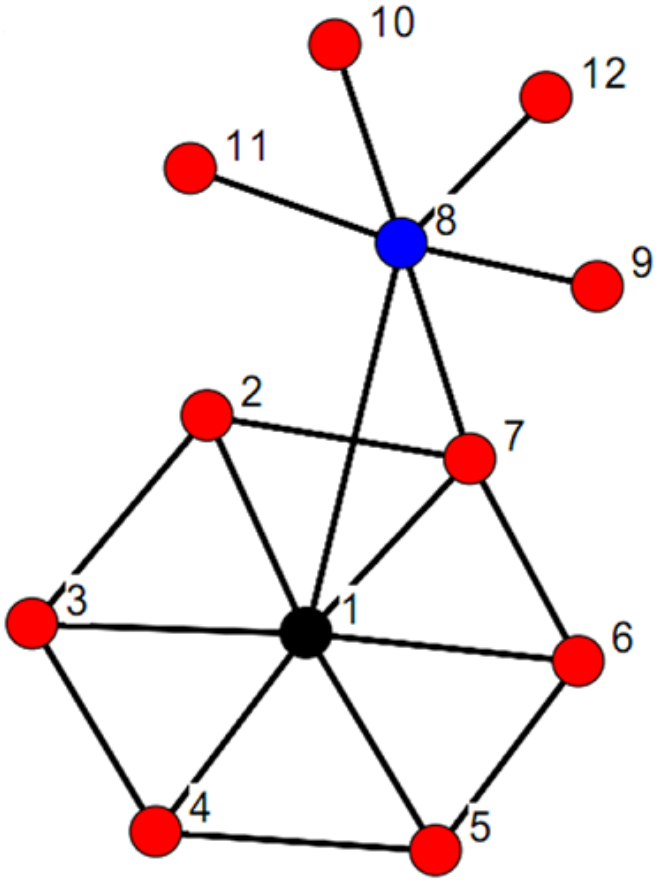
Trial

- A trial is a walk consisting of distinct lines, although some nodes may be visited more than once

Is 10-8-7-1-8 a trail?
1-2-3-1-2-7?



Walks, Trails, Paths

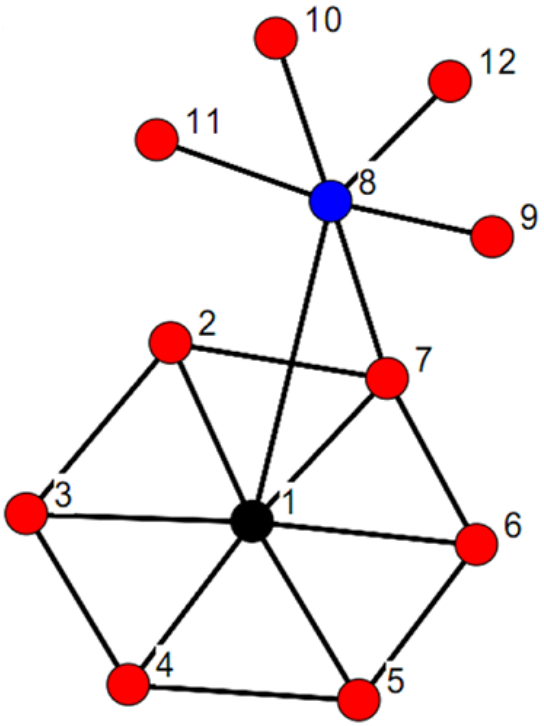


- Walk : unrestricted
  - 1-2-3-1-2-7-1-7-1
- Path : all nodes are distinct
  - 1-2-3-4-5-6-7-8-
  - But not 7-1-2-7-6
- Trail : all edges are distinct
  - 7-1-2-7-6
  - 1-2-3-1-7-8
  - But not 7-1-2-7-1-4

Geodesic path

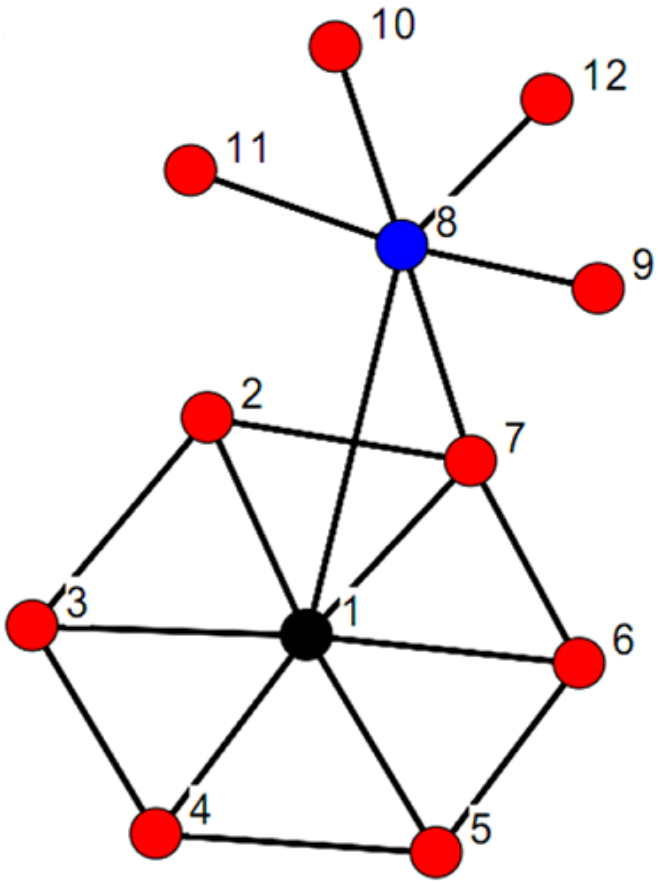
- Several paths may exist between two nodes, but the shortest path between them is the geodesic

The geodesic distance between "10" and "4"?



Types of transmission routes

- Walks: unrestricted – can repeat edges (4-1-8-1-7-1...) MONEY BILLS
- Paths: can't visit a node more than once (4-3-2-1-8) VIRUS
- Trails: can't use any line more than once (4-1-7-8-1-6) GOSSIP
- Geodesics: shortest path between two nodes (4-1-8) PACKAGE



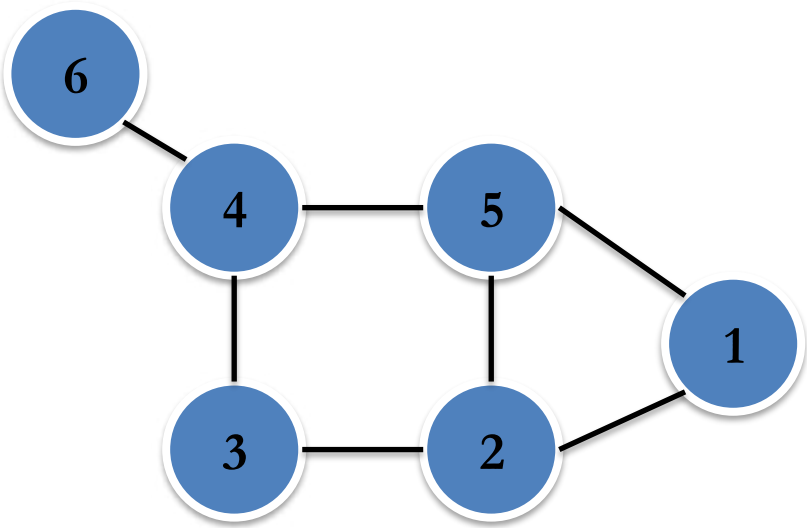
Adopted from Borgatti, 2010

Geodesic distance

- Given a graph G, The **distance** d(x,y) between two vertices x and y is **the length of the shortest path** from x to y, considering all possible paths in G from x to y.
- The distance between any node and itself is 0. If there is no path from x to y then d(x,y) is infinity.

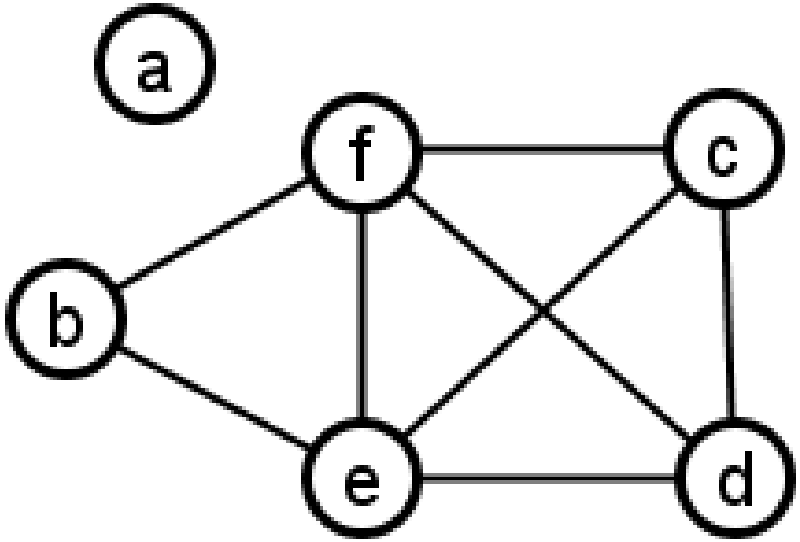
Geodesic distance

- The geodesic distance from 6 to 1 is ?

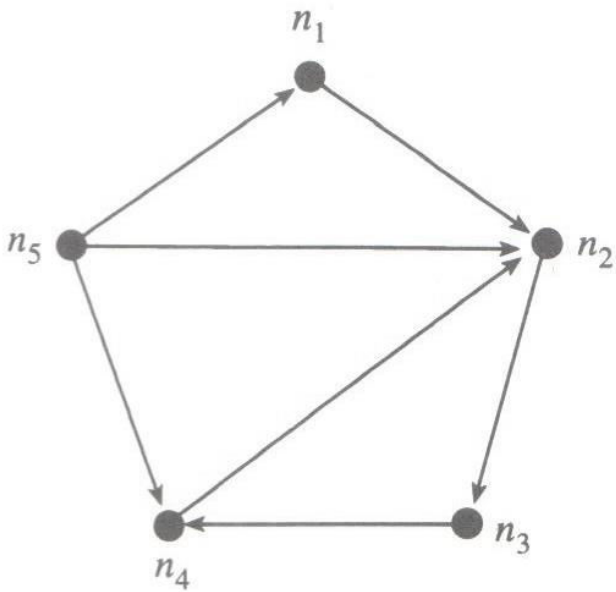


Geodesic distance

The distance from b to c is?



Directed graph



Directed walk	$n_5\ n_1\ n_2\ n_3\ n_4\ n_2\ n_3$
Directed path	$n_5\ n_4\ n_2\ n_3$
Semipath	$n_1\ n_2\ n_5\ n_4\ n_3$
Cycle	$n_2\ n_3\ n_4\ n_2$
Semicycle	$n_1\ n_2\ n_5\ n_1$

Fig. 4.18. Directed walks, paths, semipaths, and semicycles

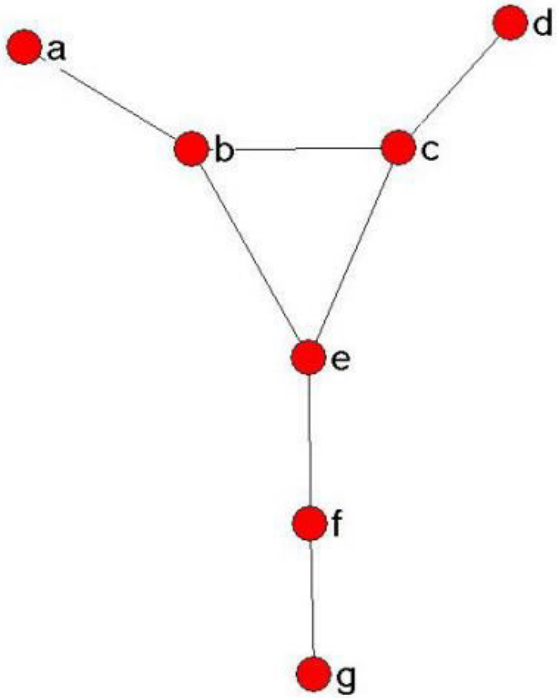
Geodesic in a directional network

- For a directional network, the geodesic is considered as the shortest directed path between two nodes.
- Considering that in a directed path all the arrows have to be in the same direction, the geodesic from n_i to n_j will not always be the same geodesic from n_j to n_i.

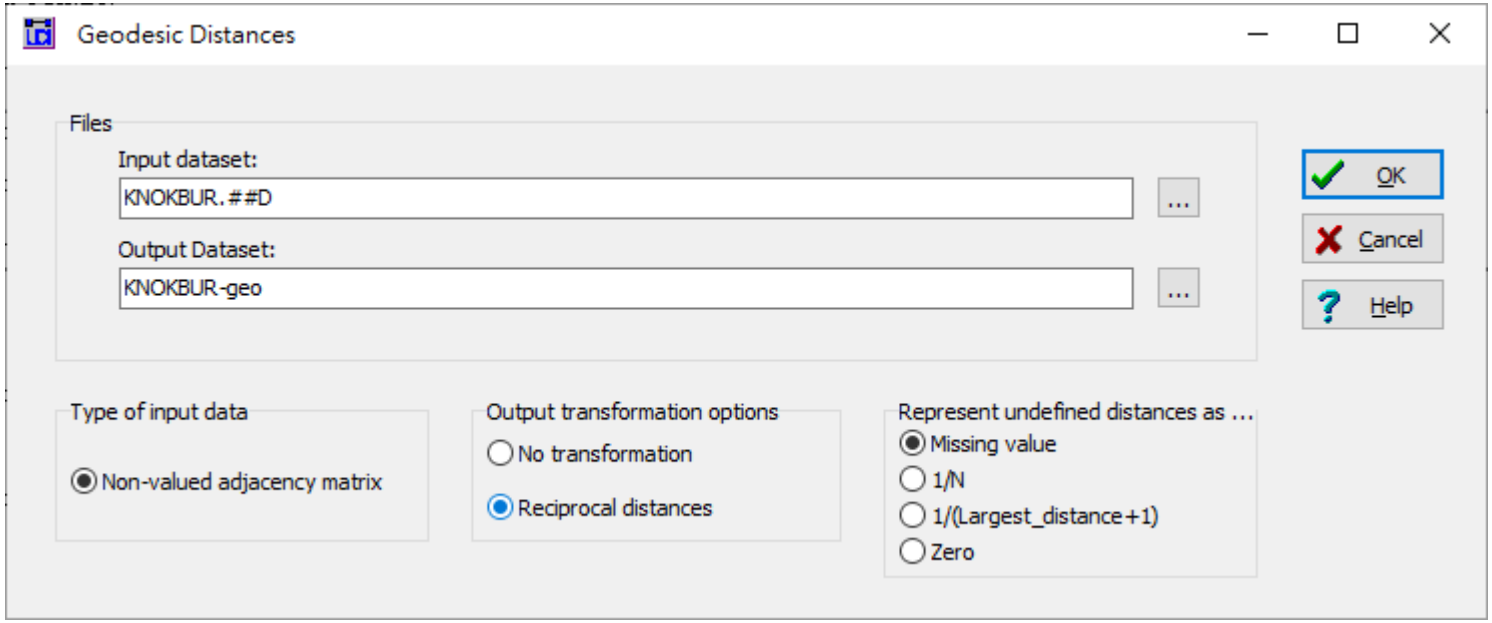
Geodesic distance matrix as opposed to adjacency matrix

- Network>Cohesion> Geodesic distance

	a	b	c	d	e	f	g
a	0	1	2	3	2	3	4
b	1	0	1	2	1	2	3
c	2	1	0	1	1	2	3
d	3	2	1	0	2	3	4
e	2	1	1	2	0	1	2
f	3	2	2	3	1	0	1
g	4	3	3	4	2	1	0



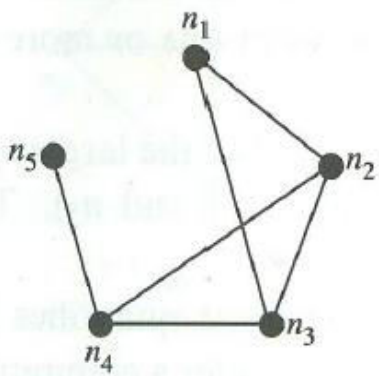
Geodesic distance



Eccentricity of a node

- Consider the geodesic distances between a given node and the other g -1 nodes in a connected graph,
- The eccentricity of a node is the largest geodesic distance between that node and any other node.

Diameter: the largest geodesic distance between any pair of nodes



Geodesic distances

$d(1,2) = 1$
$d(1,3) = 1$
$d(1,4) = 2$
$d(1,5) = 3$
$d(2,3) = 1$
$d(2,4) = 1$
$d(2,5) = 2$
$d(3,4) = 1$
$d(3,5) = 2$
$d(4,5) = 1$

Diameter of graph $= \max d(i,j) = d(1,5) = 3$

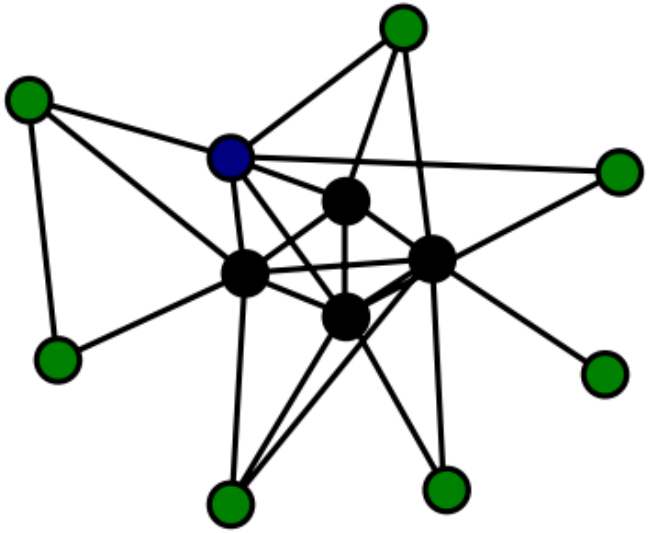
Fig. 4.9 Graph showing geodesics and diameter

Or equivalently, the largest nodal eccentricity

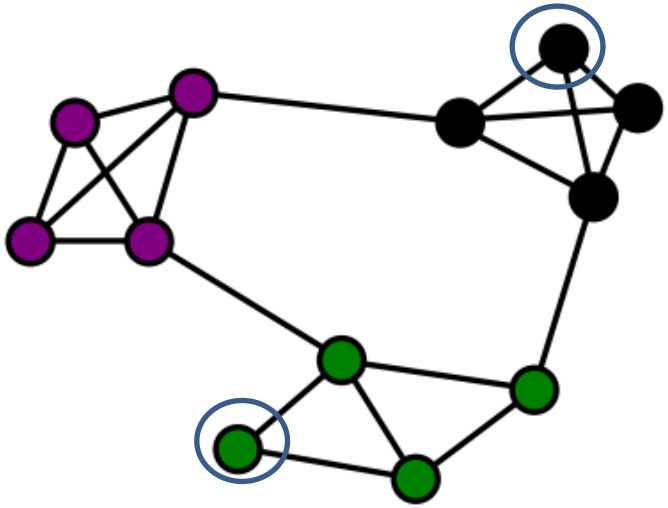
Diameter

- The maximum internode distance
  - the "longest shortest path (geodesic path)" between any two nodes

Diameter = 3



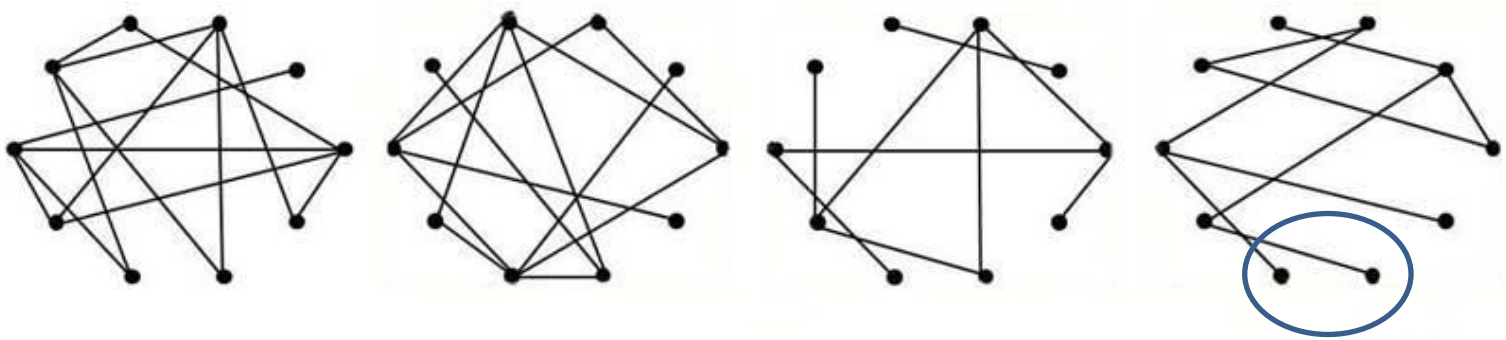
Diameter = 4



The lower the diameter of a network the shorter the time to send a message from one node to the node farthest away from it

Adopted from Borgatti, 2010

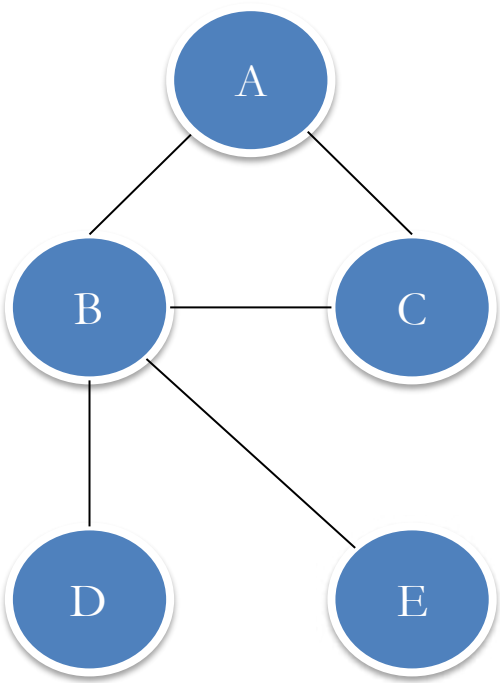
Graph Diameter



- The graphs above have diameters of 3, 4, 5, and 7, respectively.
- What does it mean? Or how does it compare with density as a indication of group cohesion?

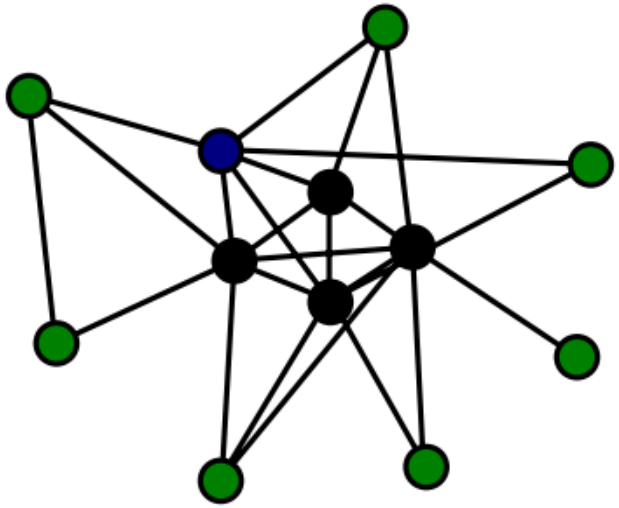
Average path length

the average number of steps along the shortest paths for all possible pairs of nodes. It is a measure of the efficiency of information or mass transport on a network

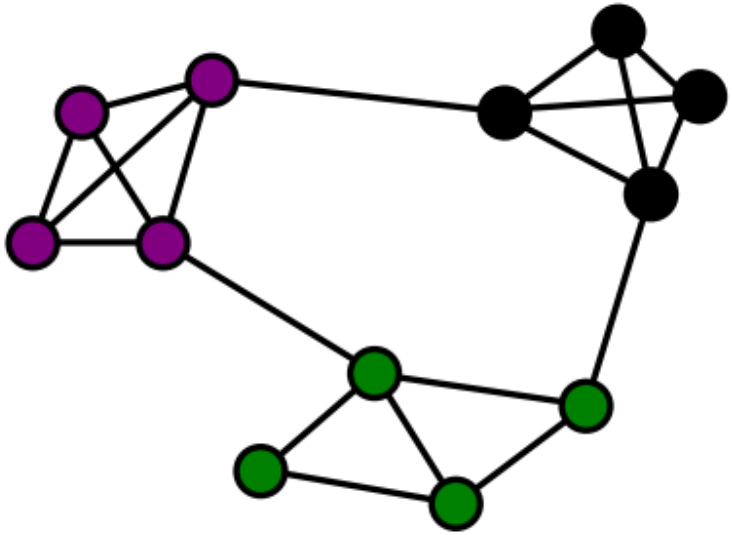


- Shortest paths are
  - AB, AC, ABD, ABE, BC, BD, BE, CBD, CBE, DBE
- Lengths
  - 1, 1, 2, 2, 1, 1, 1, 2, 2, 2
- Average path length
  - 1.5

Average path length



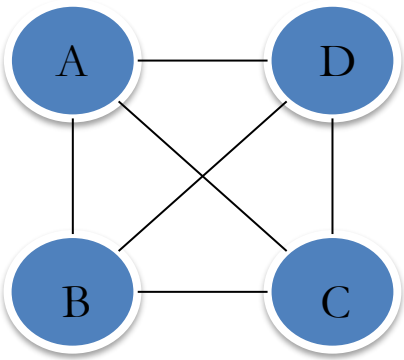
Average distance = 1.9



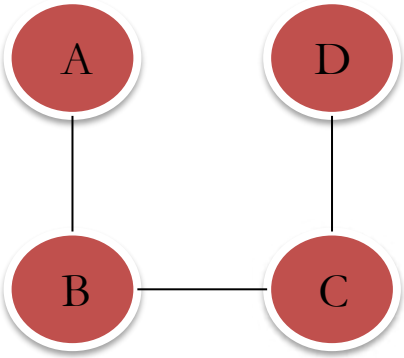
Average distance = 2.4

Adopted from Borgatti, 2010

Examples



Average path length = ?



Average path length = ?

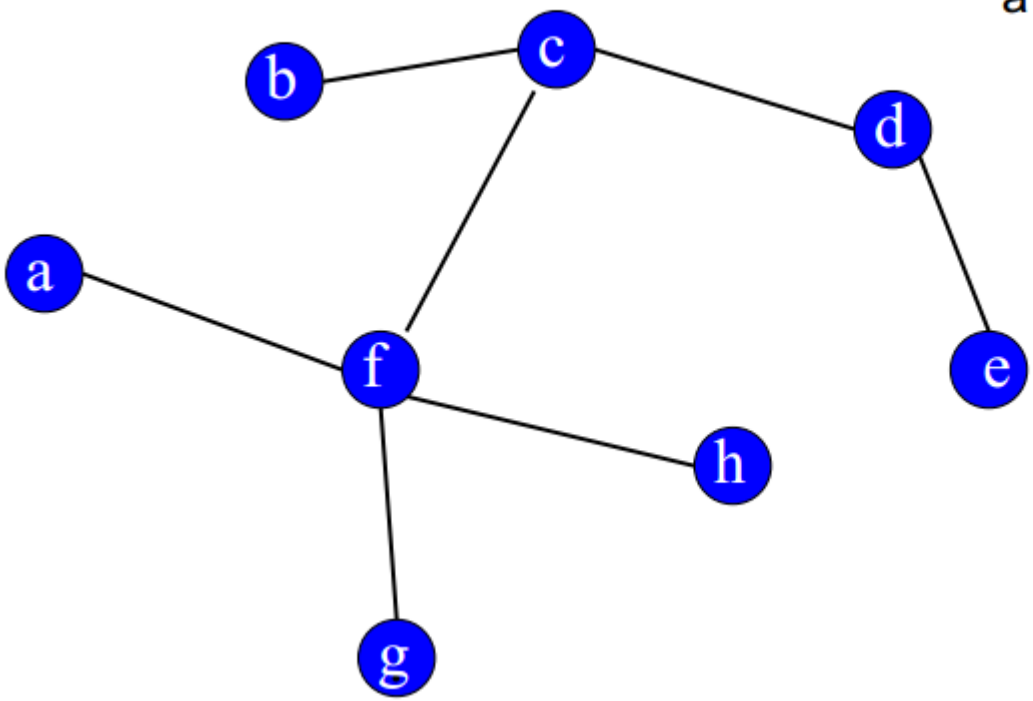
Reachability

- Adjacent
  - Two nodes are adjacent if there is an edge between them, or incident
- Reachable
  - Two nodes are reachable if there is a "path" between them

Adjacent

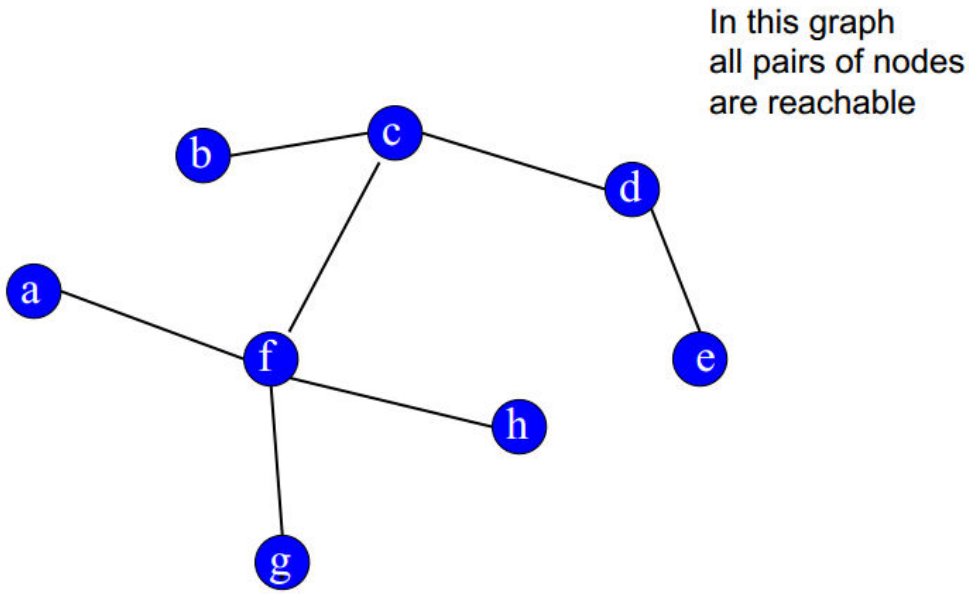
- Two nodes are adjacent if there is an edge between them

b and c are adjacent
a and b are not



Reachability

- Two nodes are reachable if there exists any set of connections by which we can trace from the source to the target actor,
- if there is a path between them



Network>Cohesion>Reachability 0,1

Not very useful metric as the network grows large, why?

Global network attributes

- Density
  - Level of interconnectedness
- Distance
  - How fast information spread
    - Geodesic path
    - Diameter
- Connectivity
  - How robust a network is
  - In-betweenness, bridging, brokage

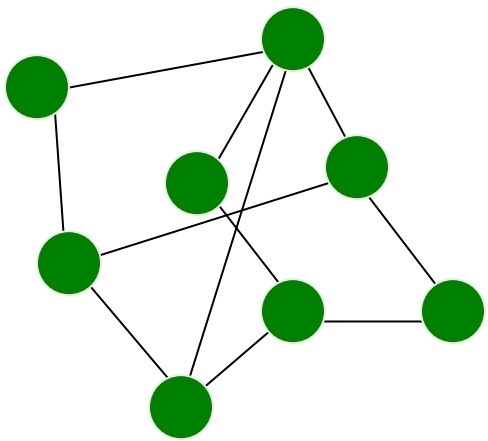
Network connectivity

- Network connectivity measures the resiliency of a network and its ability to continue operation despite disabled components

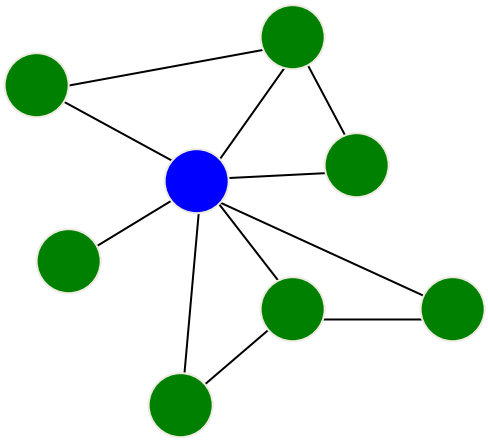
The connectivity view of cohesion
Complementary to density

Cohesion increases as the number of independent paths in the network increases.

Ties through a single person are minimally cohesive.

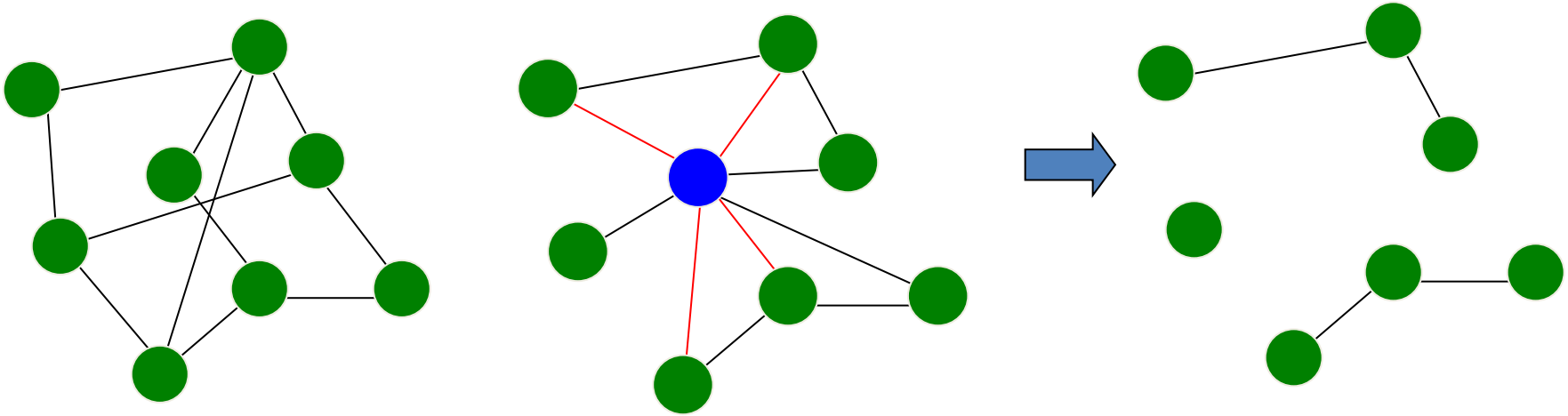


D = . 39



D = . 39

Consider this graph, is density a good indicator of cohesion?
What if density increases, but through a single person?



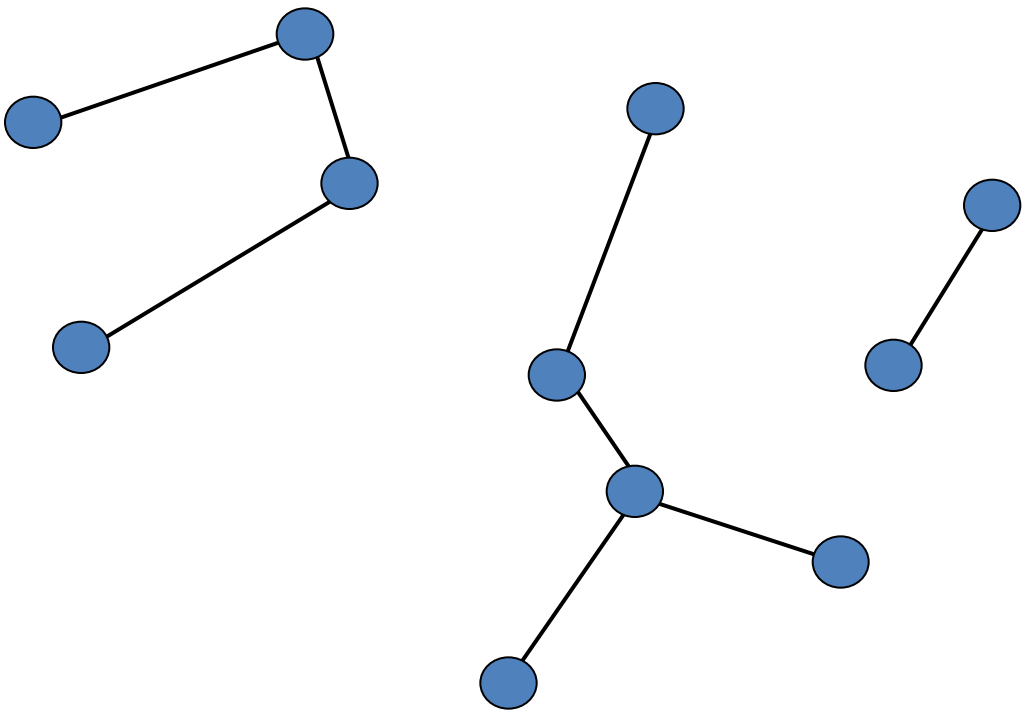
Removal of 1 person destroys the group.

Sociologically,

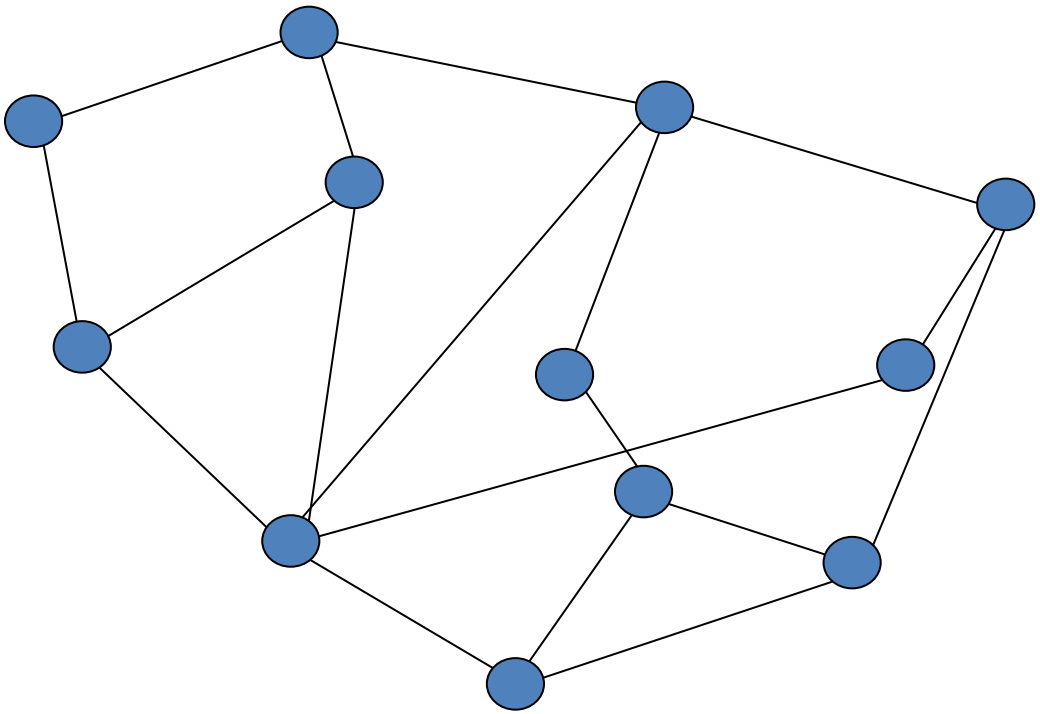
An intuitive definition of structural cohesion:

A collectivity is structurally cohesive to the extent that the social relations of its members hold it together.

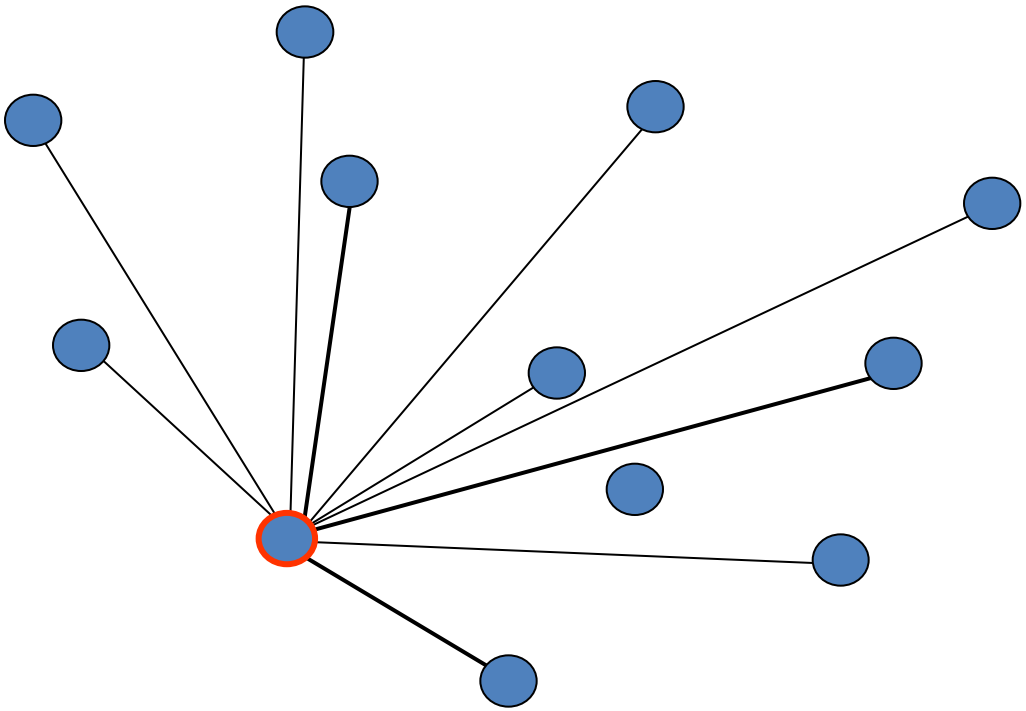
A scattered network



Spreading relations around the structure makes it robust to node removal.



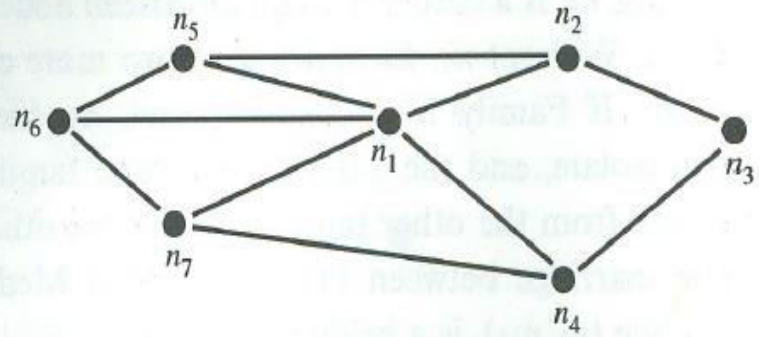
When focused on a single person, the network is fragile.



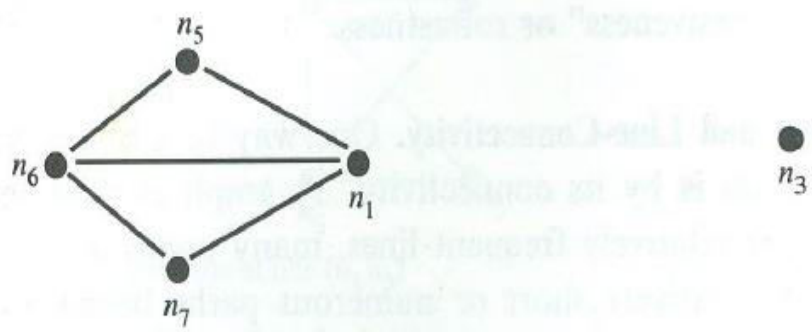
Global connectivity

Redundant routes enhance resistance to disruption

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



n_2 and n_4 comprise a 2-node cut



The graph without n_2 and n_4

Fig. 4.12. Connectivity in a graph

Local connectivity

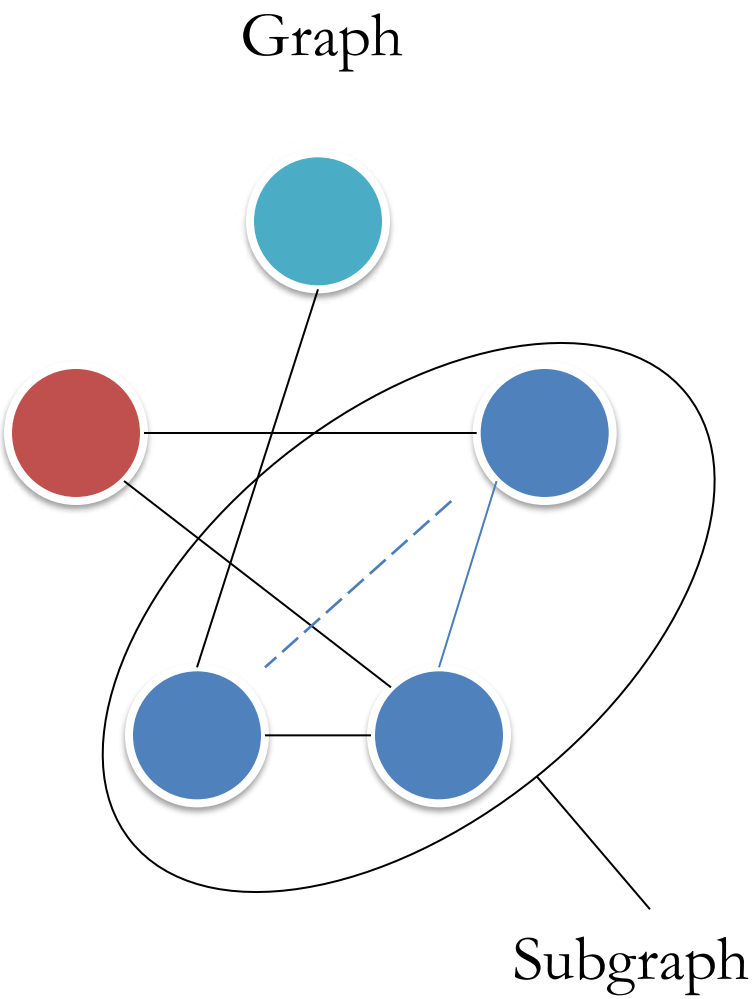
- *Connectivity* calculates the number of nodes that would have to be removed in order for one actor to no longer be able to reach another.
- If there are many different pathways that connect two actors, they have high "connectivity" in the sense that there are multiple ways for a signal to reach from one to the other.

Point/local connectivity

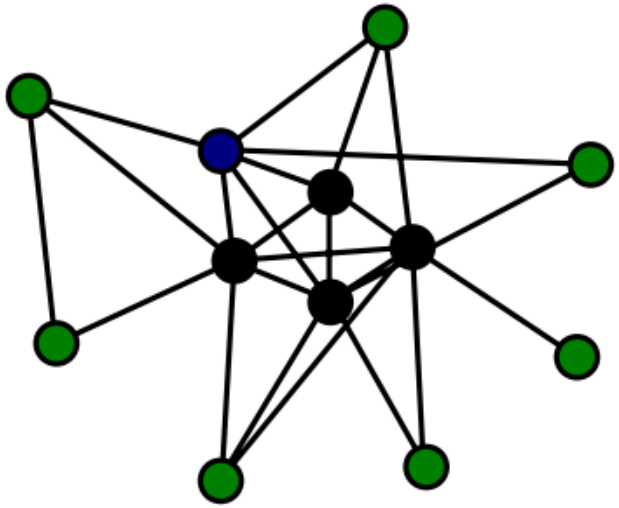
- **NETWORK>COHESION>POINT CONNECTIVITY**
  PURPOSE Compute the local point connectivity between all pairs of nodes in a network.
  DESCRIPTION The local (point) connectivity of two non-adjacent vertices is the number of vertices that need to be deleted so that no path connects them, this is equal to the maximum number of vertex disjoint paths connecting them.
- NOTE: This procedure only operates on the first matrix in a dataset.
- Hint: data > unpack

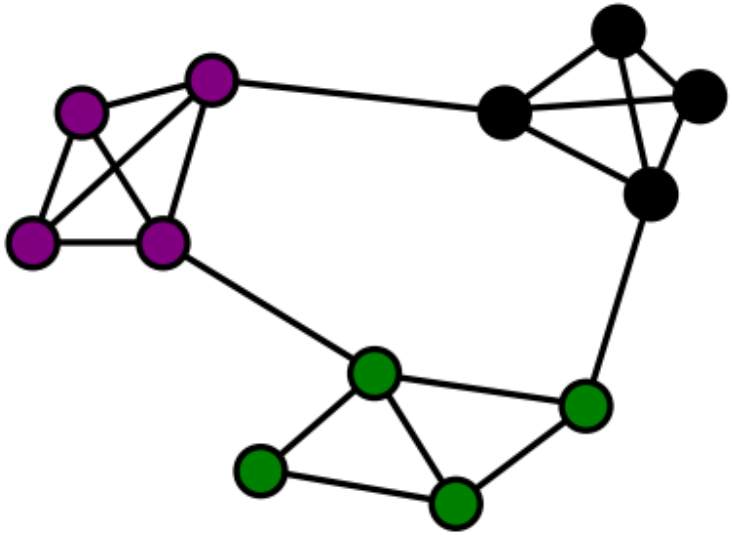
Subgraphs

- A subgraph consists of a subset of the nodes and edges of a graph
- Subgraphs are also graphs



Are there apparent subgroups?





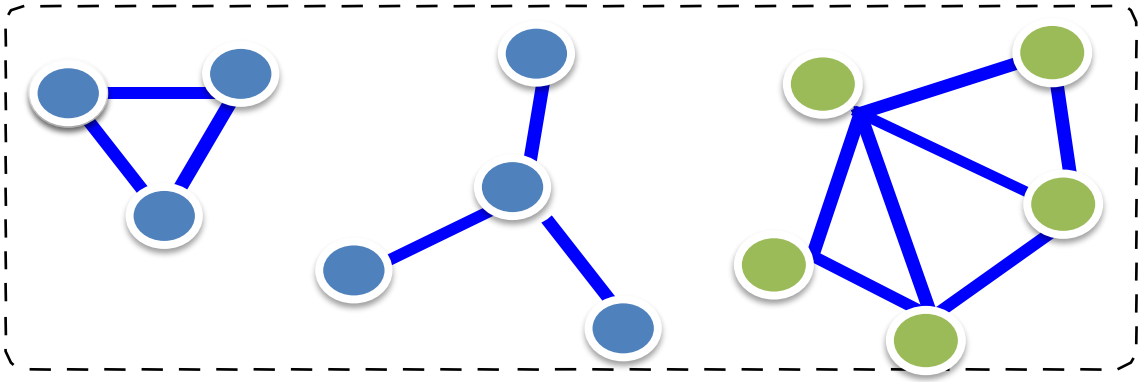
Adopted from Borgatti, 2010

Basic form of subgraphs

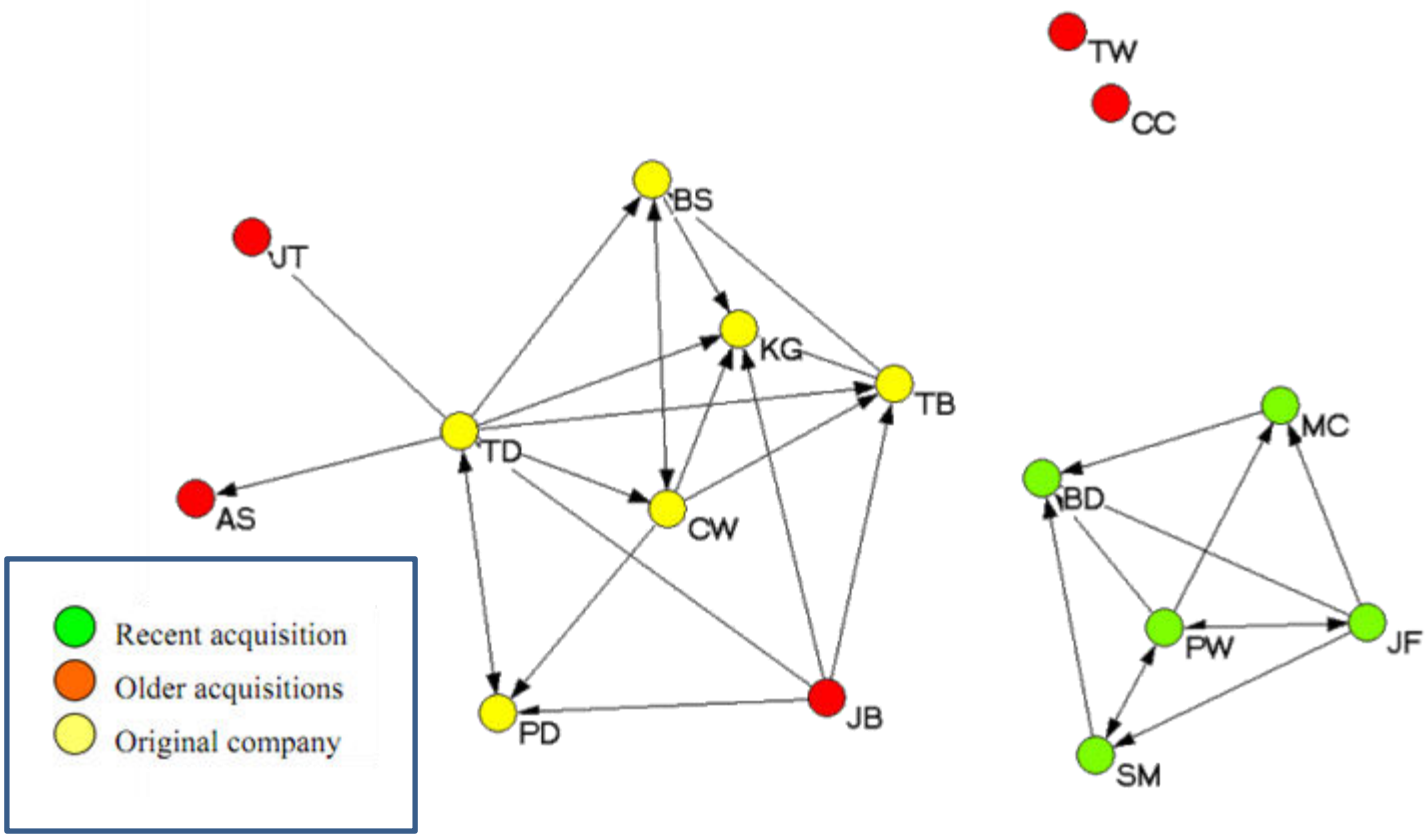
- Components
  - a maximal subgraph in which all nodes are reachable from every other nodes
- Cliques AND N-clique
  - fully interconnected subgroups of actors
- K-core..
  - Subgroups of highly interconnected actor

Connected components

- Maximal sets of nodes in which every node can **reach** every other by some path (no matter how far) i.e. the largest possible subgraph



A network with how many components?



Strong and weak components

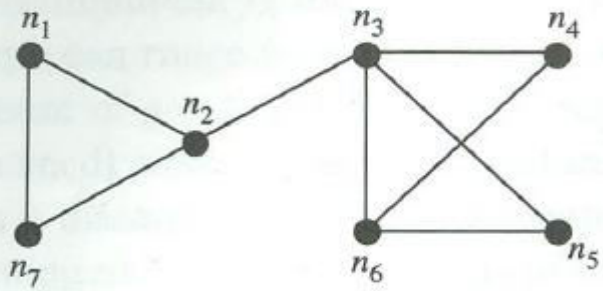
- For directed graphs, there strong components and weak components.
- A strong component is a maximal subgraph in which there is a path from every point to every point following all the edges in the direction they are pointing.
- A weak component is a maximal subgraph which would be connected if we ignored the direction of the edges.

Cutpoint and bridge

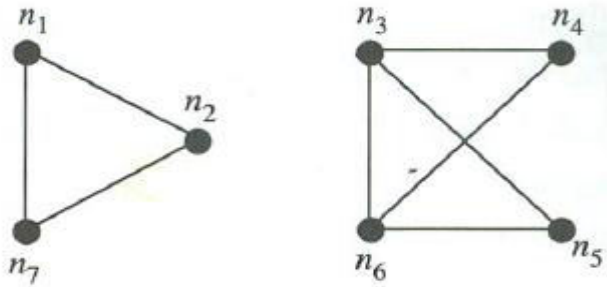
- An edge is a bridge if its removal would disconnect a graph into two or more components
- A node is a cutpoint if its removal would disconnect the graph
- A cutpoint set is a collection of nodes whose removal disconnects a graph.

Bridge

A bridge is a line such that the graph containing the line has fewer components than the subgraph that is obtained after the line is removed



Line $(n_2\ n_3)$ is a bridge



Graph without line $(n_2\ n_3)$

Fig. 4.11. Example of a bridge in a graph

Bridge and cutpoint

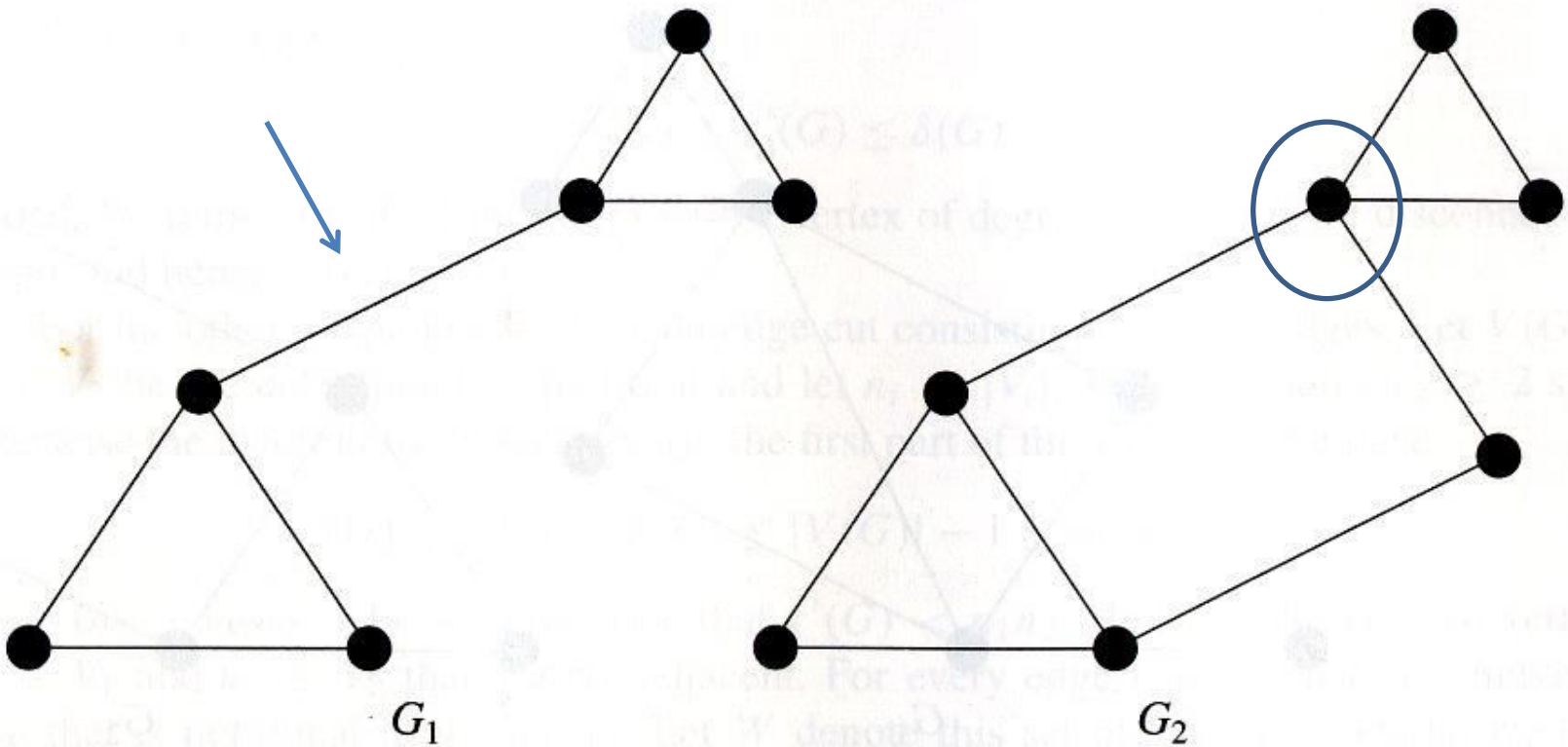


Figure 6.3: G_1 has a bridge and a cut-vertex; G_2 has no bridge but has a cut-vertex.

Edge cut

Sometimes we have to remove more than one line to disconnect the graph

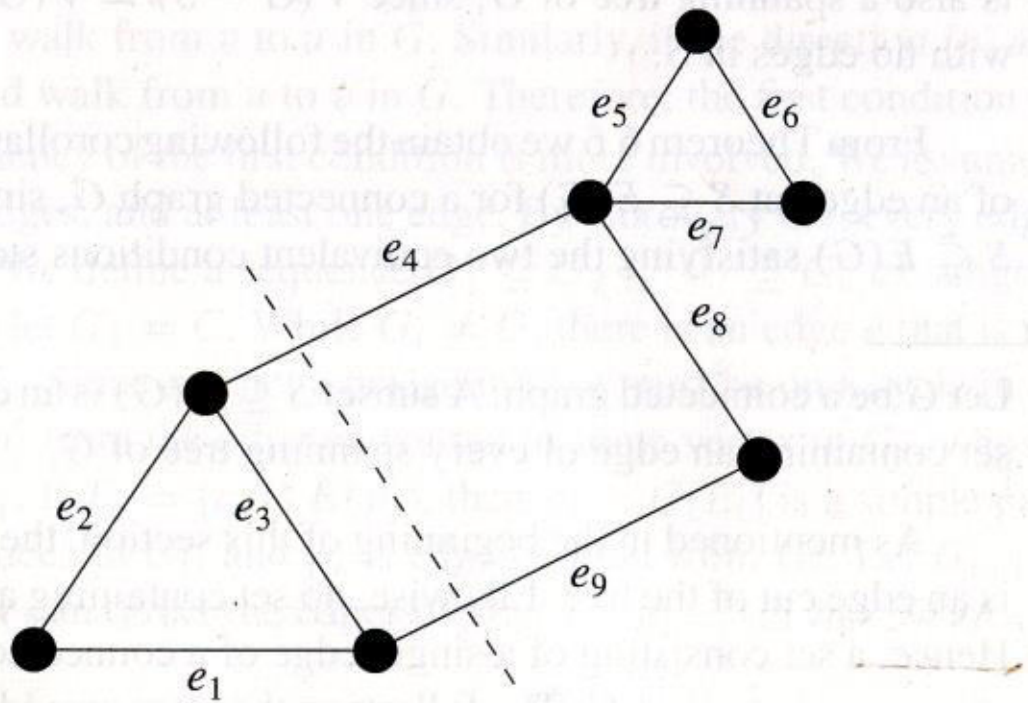
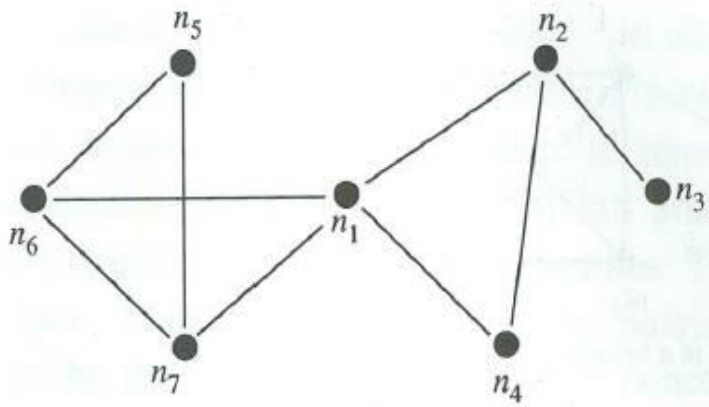
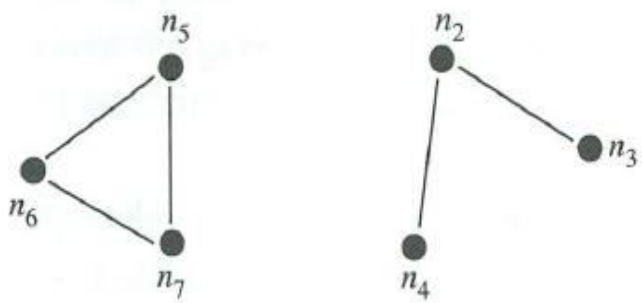


Figure 6.1: The removal of edges e_4 and e_9 disconnects the graph, but the removal of either one of these edges separately does not.

Cutpoint & connectivity



Node n_1 is a node cut, or cutpoint



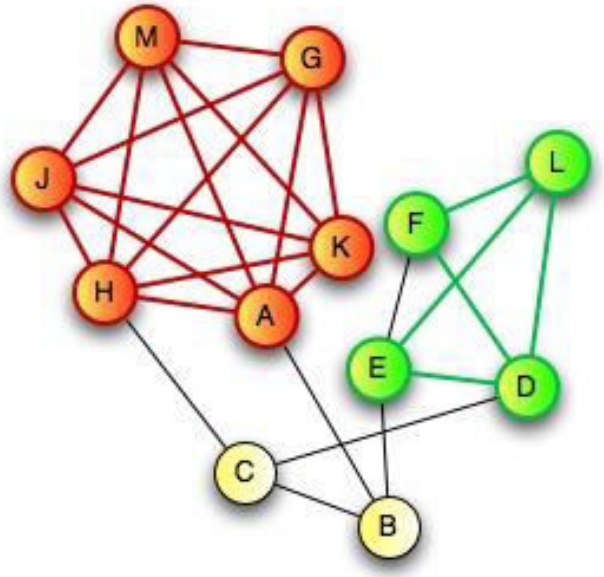
The graph without node n_1

Fig. 4.10. Example of a cutpoint in a graph

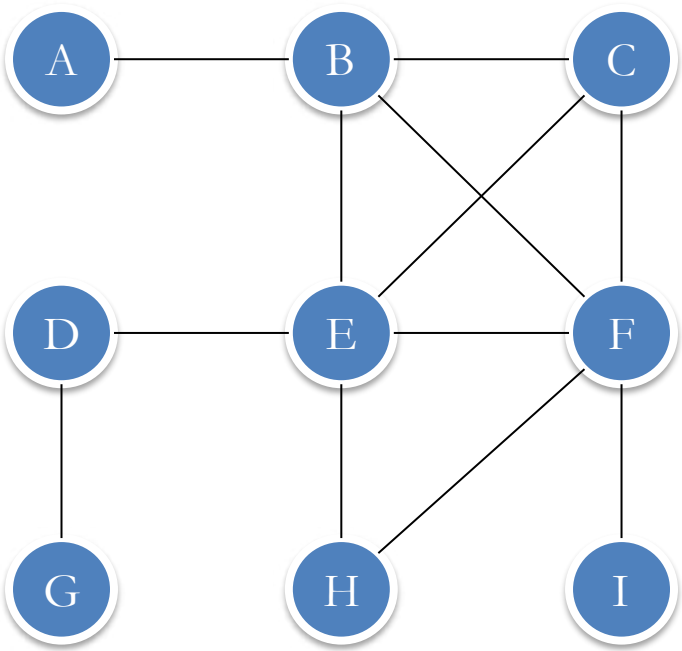
Clique: fully connected component

- The maximal number of actors who have all possible ties (i.e. fully connected) present among them themselves

The geodesic distance for every pair of nodes
in a clique is ?



Clique



- B, C, E and F form a clique of size 4
- E, F and H form a clique of size 3
- A, D, G and I are not part of any clique

How many components/cliques?

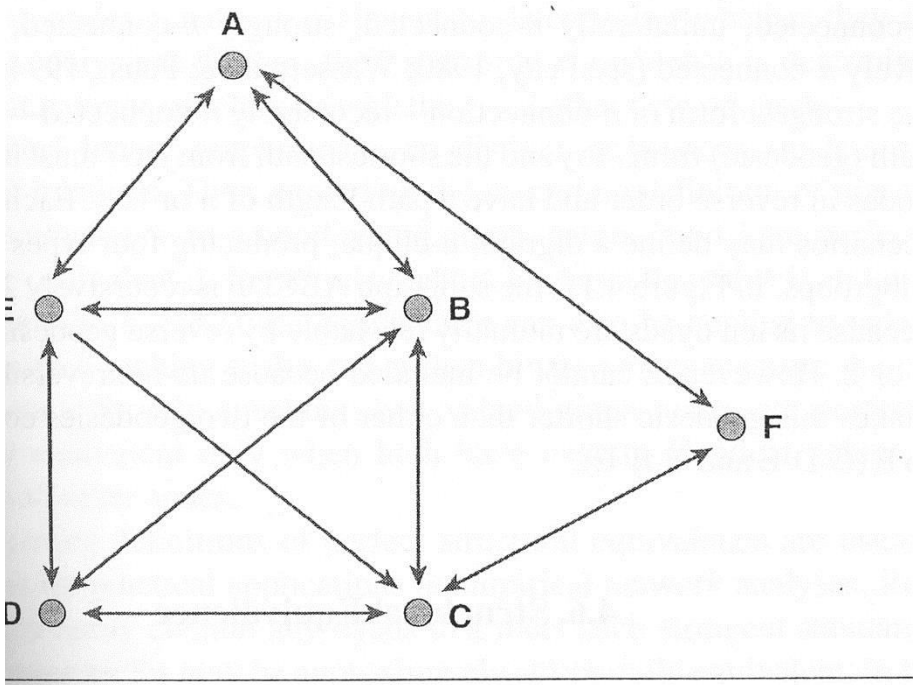
Figure 1: A disconnected graph with two components and three cliques, each a K_3.

Cliques in a directed graphs

A subgraph with three or more nodes, all directly connected to one another,

Where each pair is **reciprocally connected**

Can you identify the three triadic cliques?



Cliques in a Directed Graph

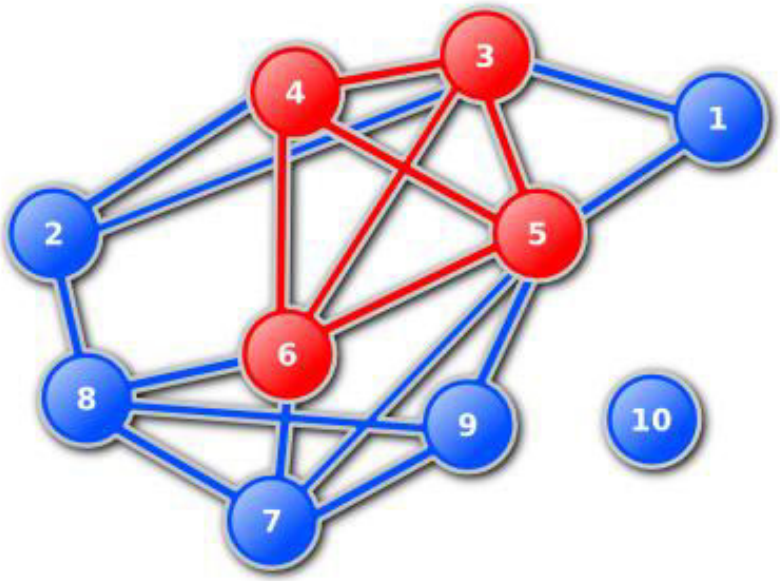
Identify cohesive subgroups

- Searching for subgroups
- As a method for identifying cohesive subgroups, the clique criteria impose very strict conditions
- To relax the stringent requirements for clique detection using nodal connectivity
  - N-Clique
  - K-Core

Clique and N-Clique

Clique: not very useful in the network get large or sparse

N-cliques, which define an actor as a member of a clique if they are connected to distances *greater than one* (more inclusive, i.e. relaxing the complete connection requirement)

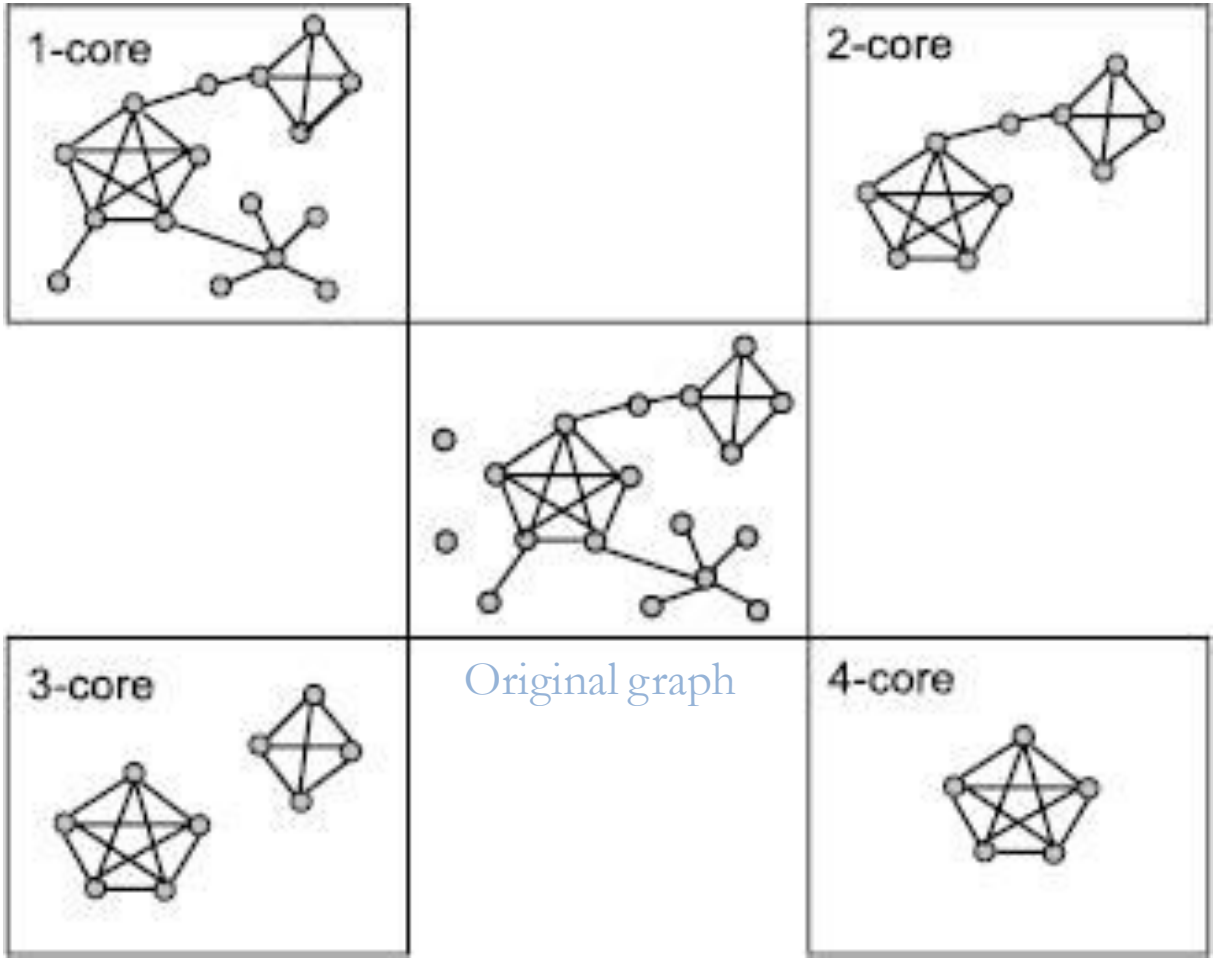


All nodes except 10 constitute a 2 Clique

K-core

- A subgroup is identified as *a k-core* if every member has direct ties to at **least *k* other nodes.**
- What happens if we raise the number of K?
- Useful when dealing with large network
  - Filter the network by raising "K"

Raise “k”, raise the threshold



Zhang et al. (2008)

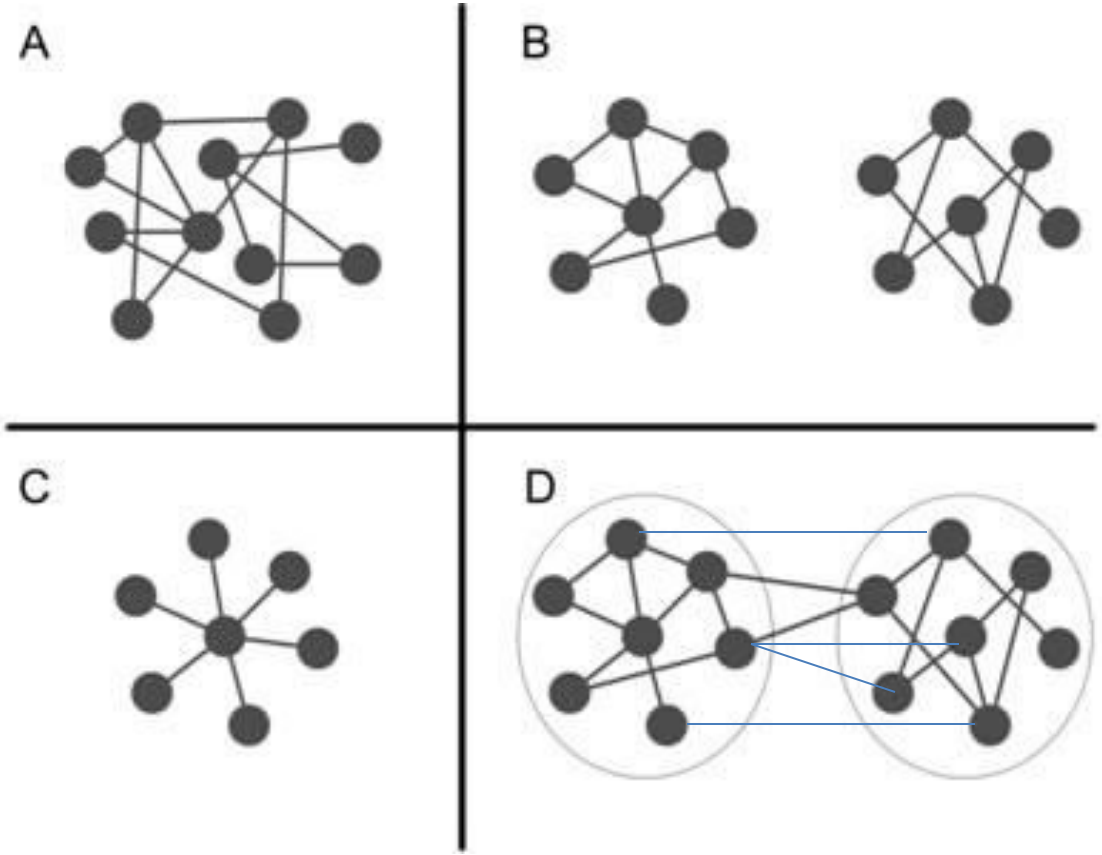
SNA

Cohesion, geodesic distance

Cohesion

- **Don't confuse the measure with the construct**
- Many ways to define or theoretically conceive of cohesion
- What is the mechanism that would relate cohesion to the outcome of interest?
- Define cohesion consistent with this mechanism
- For each way, we can then devise an operational measurement

Cohesion



Cohesions

- Network>Cohesion>Density
- Network>Cohesion>Reachability
- Network>Cohesion>No. of Geodesics
  - Geodesics?
- Network>Cohesion>Point Connectivity

More functions…

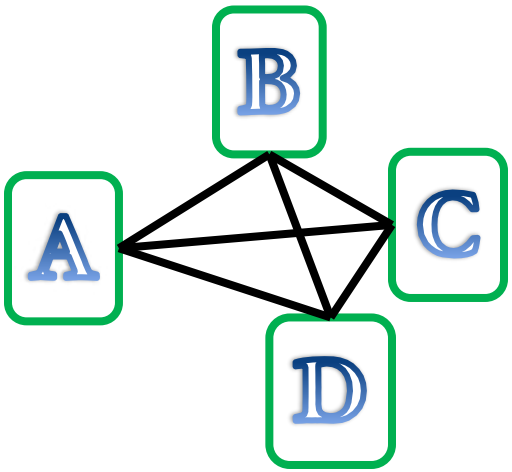
- Network>Cohesion>Density>Density of subgroups
- Network>Network properties>Reciprocity
  - For directional network
- Network>Cohesion>E-I index

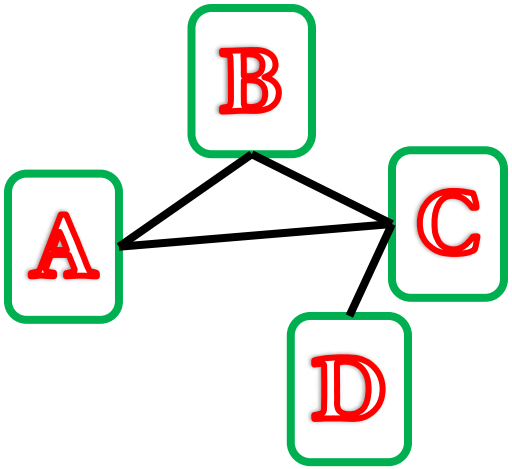
Density

- An index of the degree of dyadic connection in a group
  - Binary data
    - The ration of the number of adjacencies that are present divided by the number of all possible pairs

Density

- The density of a <u>binary</u> network is the total number of ties divided by the total number of possible ties





Density

- Valued data
  - **The sum of the values of all ties divided by the number of possible ties**
  - The average strength of ties of across all possible (not all actual) ties

Density by group

- Identify subgroups in a network
- Calculate the density, sum or average value **within** and **between** groups in a network or matrix.

Density by groups

Open KNOKBUR

Create an attribute file (##d ##h)

- Group 1: governmental agencies: coun, educ, mayr 2 out of 3 links present
- Group 2: non-governmental agencies : comm, indu, news
- Group 3: welfare specialists: wro, uway, welf, west

Density in UCINET and Gephi

Try Network>Cohesion>Overall Density

Try Network>Cohesion>Density by group

KNOKBUR

You gonna need attribute data

##D#, ##H

Also show the attributes using NetDraw

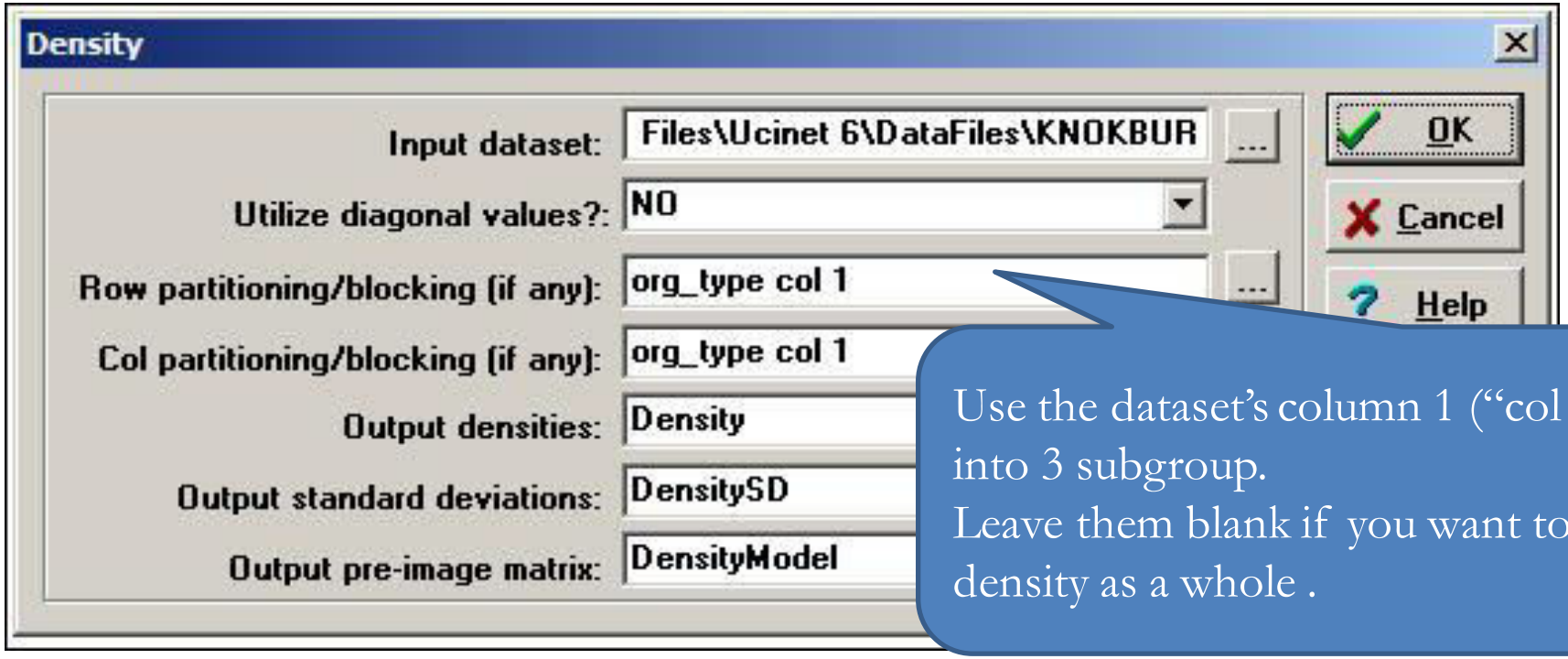
Do the same exercise with friendship data

Density

- Exercise/unpack KRACK-HIGH-TECH
  - Open KRACKAD and KRACKFR from NetDraw
  - Use UCINET to compare the density between friendship and advice seeking density
  - Network>cohesion>density (overall)
    - * by group ? (Please use the old procedure, I have run into all kinds of problems with the "New")

Within group density

- *Network>Cohesion>Density*
  - To calculate the density of whole populations, or of partitions
  - We can use an attribute or partition to divide the cases



Column Block	Old Code	Members:
1	1	COUN EDUC MAYR
2	2	COMM INDU NEWS
3	3	WRO UWAY WELF WEST

Relation: KNOKI

	1 C	3 E	5 M	2 C	4 I	7 N	6 W	8 U	9 W	10 W
1			1	1		1			1	
3			1	1	1	1	1			1
5	1	1		1	1	1		1	1	1
2	1	1	1		1	1		1	1	
4	1		1	1		1				
7			1	1	1					
6		1				1			1	
8	1		1	1	1	1			1	
9			1	1		1				
10	1	1	1	1		1				

This shows the way you partitioned the cases. "Old Code" is from column 1.

The three sub-groups and the connections WITHIN them.

```
Density / average value within blocks

            1       2       3
            1       2       3
        ------  ------  ------
  1 1   0.6667  0.8889  0.5000
  2 2   0.6667  1.0000  0.1667
  3 3   0.5833  0.6667  0.1667


Standard Deviations within blocks

            1       2       3
            1       2       3
        ------  ------  ------
  1 1   0.4714  0.3143  0.5000
  2 2   0.4714  0.0000  0.3727
  3 3   0.4930  0.4714  0.3727
```

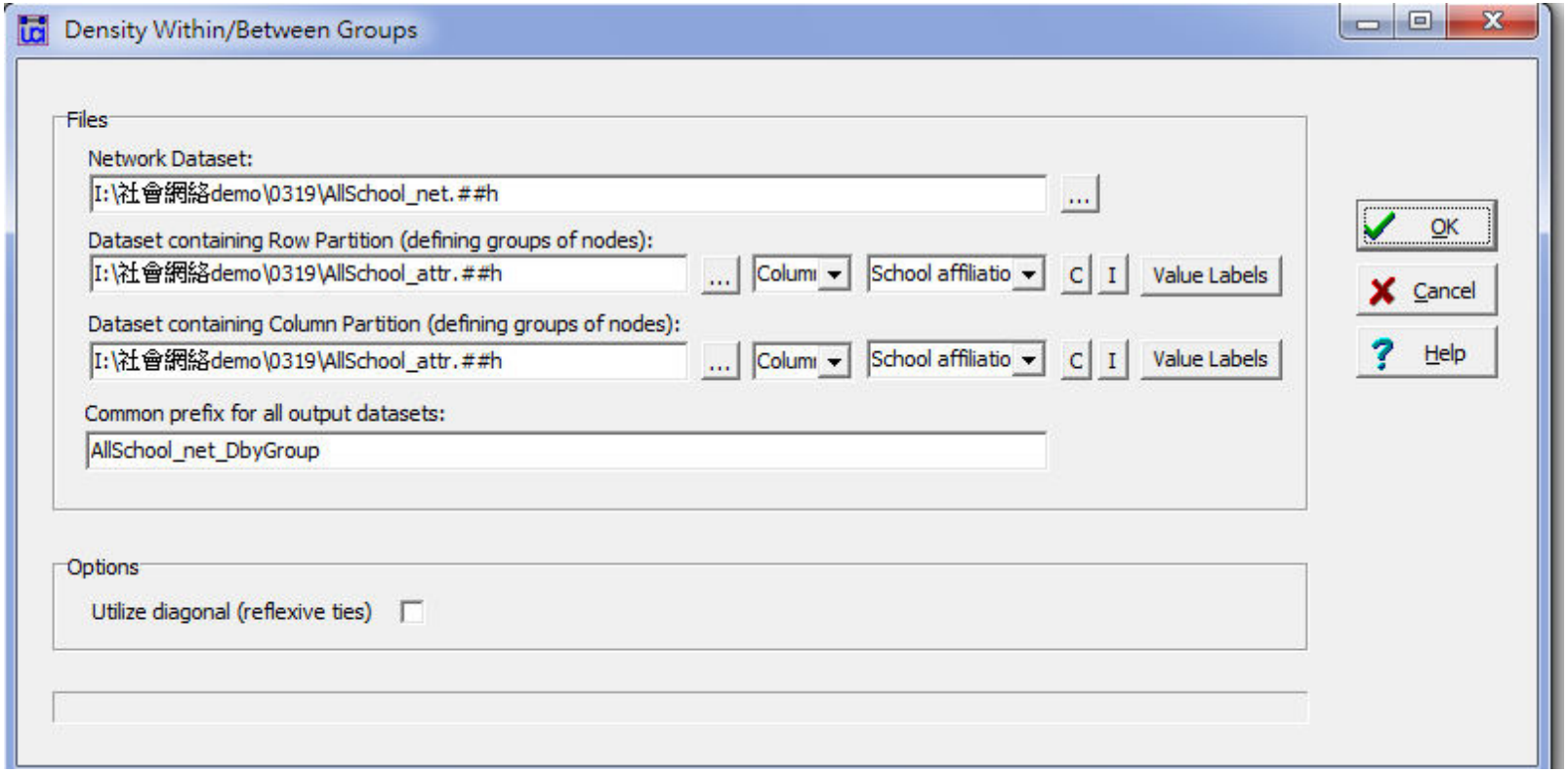
The density within- and between-blocks.

The SD within- and between-blocks.

Exercise, apply density analysis on NTNU data

Simplified NTNU

Save into excel file then open with Gephi



Density by groups

Number of ties/Sum of tie-strengths

	1 0	2 1	3 2	4 3	5 4	6 5	7 6	8 7	9 8
1 0	0.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
2 1	1.000	138.000	35.000	19.000	25.000	14.000	19.000	0.000	16.000
3 2	0.000	36.000	79.000	21.000	16.000	6.000	30.000	4.000	47.000
4 3	0.000	10.000	25.000	118.000	17.000	9.000	9.000	6.000	85.000
5 4	0.000	4.000	1.000	12.000	59.000	12.000	15.000	2.000	14.000
6 5	1.000	20.000	14.000	12.000	36.000	30.000	24.000	5.000	12.000
7 6	0.000	7.000	11.000	11.000	11.000	16.000	61.000	2.000	15.000
8 7	0.000	11.000	6.000	2.000	5.000	4.000	1.000	19.000	10.000
9 8	0.000	3.000	9.000	16.000	1.000	4.000	1.000	7.000	11.000

Density (prop of ties) / Average tie strength

	1 0	2 1	3 2	4 3	5 4	6 5	7 6	8 7	9 8
1 0	0.000	0.009	0.010	0.000	0.000	0.000	0.000	0.000	0.000
2 1	0.009	0.197	0.052	0.032	0.042	0.027	0.050	0.000	0.008
3 2	0.000	0.053	0.132	0.038	0.029	0.013	0.086	0.020	0.026
4 3	0.000	0.017	0.045	0.255	0.035	0.022	0.029	0.034	0.054
5 4	0.000	0.007	0.002	0.025	0.128	0.029	0.049	0.011	0.009
6 5	0.013	0.039	0.029	0.029	0.086	0.088	0.090	0.033	0.009
7 6	0.000	0.019	0.031	0.036	0.036	0.060	0.335	0.018	0.015
8 7	0.000	0.051	0.030	0.011	0.028	0.026	0.009	0.339	0.018
9 8	0.000	0.002	0.005	0.010	0.001	0.003	0.001	0.012	0.002

Reachability

- An actor is reachable by another if there exists any set of connections by which we can trace from the source to the target actor, **regardless of how many others fall between them.**

Reachability of Knoke "I" and "M" relations

Matrix #1

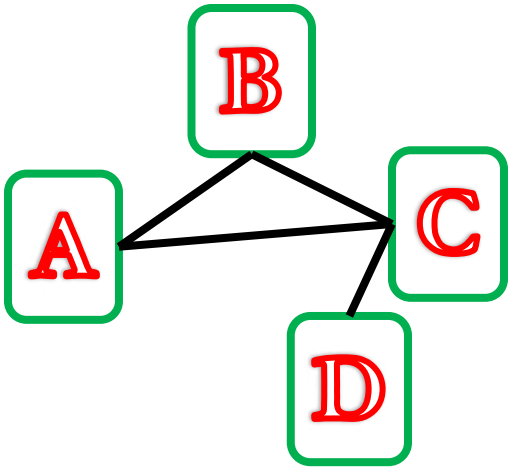
	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	1	1	1	1	1	1	1	1	1
2	1	0	1	1	1	1	1	1	1	1
3	1	1	0	1	1	1	1	1	1	1
4	1	1	1	0	1	1	1	1	1	1
5	1	1	1	1	0	1	1	1	1	1
6	1	1	1	1	1	0	1	1	1	1
7	1	1	1	1	1	1	0	1	1	1
8	1	1	1	1	1	1	1	0	1	1
9	1	1	1	1	1	1	1	1	0	1
10	1	1	1	1	1	1	1	1	1	0

Matrix #2

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	1	1	0	1	0	0	1	1	1
2	0	0	1	0	0	0	0	1	1	1
3	0	0	0	0	0	0	0	1	1	1
4	0	1	1	0	0	0	1	1	1	1
5	0	1	1	0	0	0	0	1	1	1
6	0	0	0	0	0	0	0	0	0	0
7	0	1	1	0	0	0	0	1	1	1
8	0	0	1	0	0	0	0	0	1	1
9	0	0	1	0	0	0	0	1	0	1
10	0	0	0	0	0	0	0	0	0	0

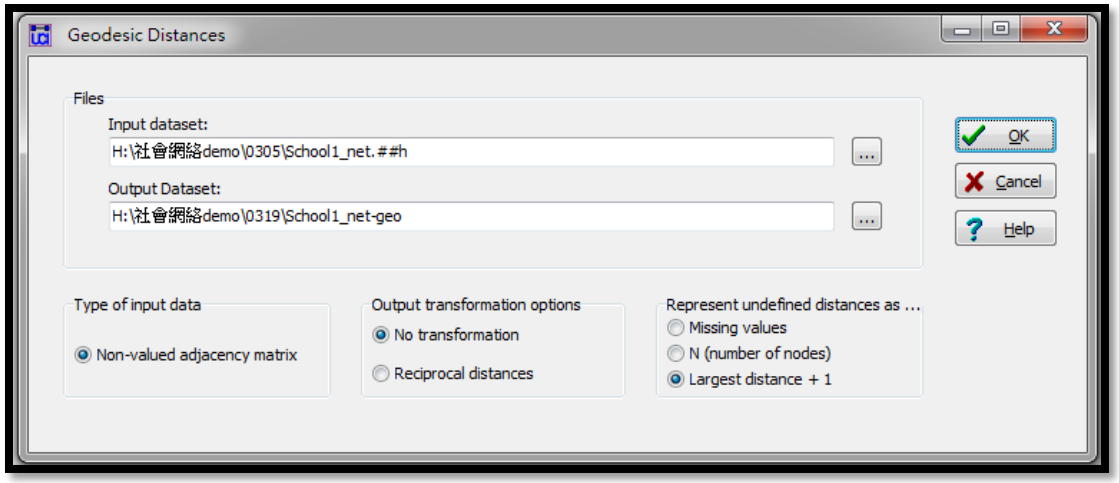
(Geodesic) Distance

- The distance between two nodes is the length of the shortest path.
- Isolates?



Distance

- Choose input, choose output.
- Use default settings.
- Suggestion: **save the text file** for average and frequencies.



	Actor1	Actor2	Actor3	Actor4	Actor5	Actor6	Actor7
Actor1	0	2	2	1	1	2	3
Actor2	2	0	1	2	1	2	3
Actor3	2	3	0	2	1	2	3
Actor4	1	1	1	0	1	1	2
Actor5	1	2	1	1	0	1	2
Actor6	1	1	1	1	1	0	1
Actor7	2	2	2	1	1	1	0

Geodesic distances for Knoke information exchange

Geodesic Distances

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	1	2	2	1	3	1	2	1	2
2	1	0	1	1	1	2	1	1	1	2
3	2	1	0	1	1	1	1	2	2	1
4	1	1	2	0	1	3	1	2	2	2
5	1	1	1	1	0	2	1	1	1	1
6	3	2	1	2	2	0	1	3	1	2
7	2	1	2	1	1	3	0	2	2	2
8	1	1	2	1	1	3	1	0	1	2
9	2	1	2	2	1	3	1	2	0	2
10	1	1	1	2	1	2	1	2	2	0

Compare 6 to 4 and 4 to 6, are they symmetrical

Save the output file and open it with netdraw
Reverse the value?

Who is close to the CEO ? Krack-high-tech

Distance in a valued graph: tie strength

- How to measure distance with valued data?
- The distance between two actors is defined as the strength of the weakest link in the shortest path.
  - If A sends 6 units to B, B sends 4 units to C, the strength of the path from A to C (**assuming A to B to C is the shortest path**) is 4.
  - Wolf/fn (first save the data into separate file, then use the OLD procedure)

No. of Geodesics

- Counts the number of different geodesic paths connecting each pair of nodes.
- Binary dataset only, or UCINET will warn you.
- What does the mean?

	Actor1	Actor2	Actor3	Actor4	Actor5	Actor6	Actor7	Actor8
Actor1	1	1	2	1	1	2	2	1
Actor2	1	1	1	1	1	1	1	2
Actor3	1	2	1	1	1	1	1	2
Actor4	1	1	1	1	1	1	1	2
Actor5	1	2	1	1	1	1	1	2
Actor6	1	1	1	1	1	1	1	1
Actor7	3	2	3	1	1	1	1	1
Actor8	0	0	0	0	0	0	0	1

Who can influence the CEO ? Krack-high-tech

Number of geodesic paths for Knoke information exchange

# of Geodesic Paths

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	1	1	2	3	1	2	1	2	1	1
2	1	1	1	1	1	1	1	1	1	2
3	4	1	1	1	1	1	1	2	3	1
4	1	1	2	1	1	2	1	2	3	1
5	1	1	1	1	1	1	1	1	1	1
6	9	3	1	2	3	1	1	6	1	1
7	3	1	2	1	1	2	1	2	2	1
8	1	1	2	1	1	2	1	1	1	1
9	2	1	2	3	1	2	1	2	1	1
10	1	1	1	4	1	1	1	2	3	1

Next: Run on WOLFN data

Flow

- Complementing geodesic paths
- Taking into account ALL connections between pairs of actors, **not just the most efficient ones**.
- Network>Cohesion>Maximum Flow
  - How many different actors in the NEIGHBORHOOD of a source leads to pathways to a target.

Flow

- "If I need to send a message to you, and there is only one other person to whom I can send this for retransmission, my connection is weak– **even if the person I send it to may have many ways of reaching you**.
- If, on the other hand, there are four people to whom I can send my message, **each of whom has one or more ways of retransmitting my message to you**, then my connection is stronger."

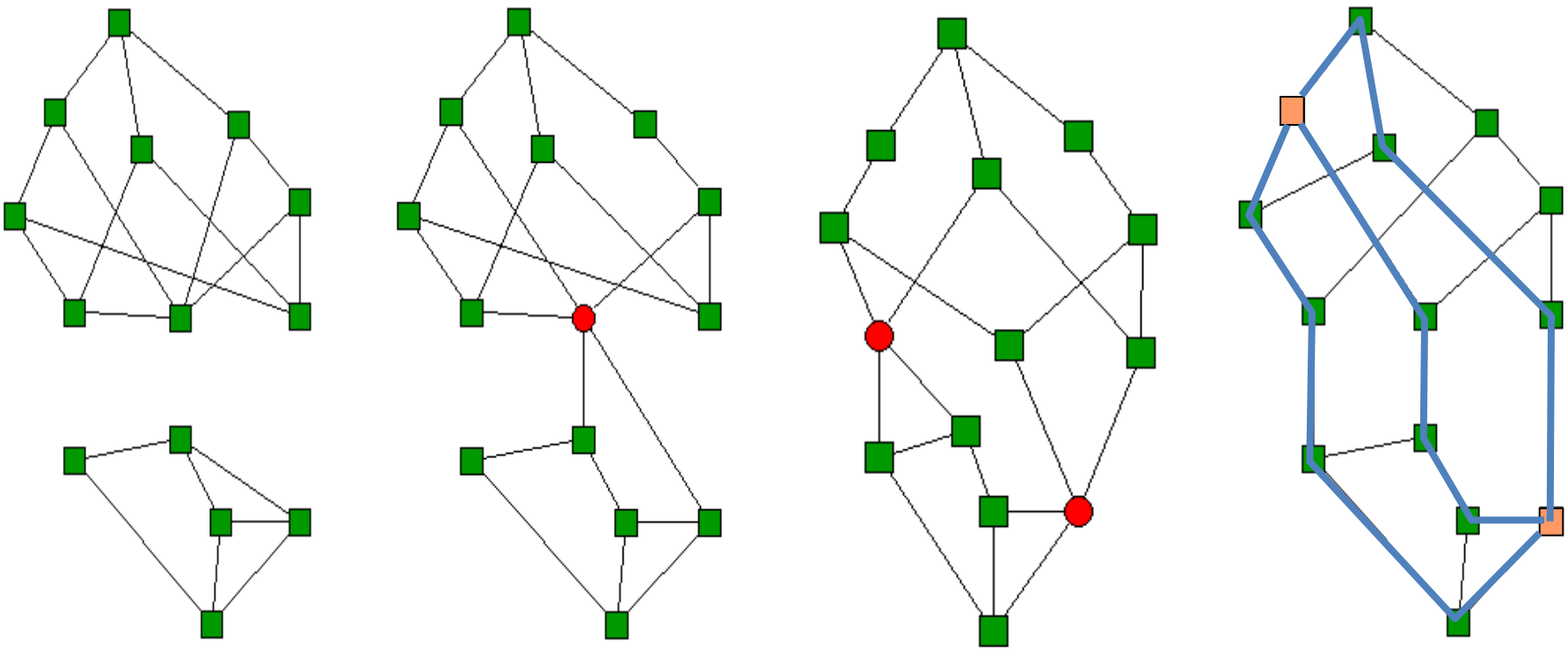
Maximum flow for Knoke information network

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	4	3	4	4	1	4	2	4	2
2	5	0	3	5	7	1	7	2	5	2
3	5	6	0	5	6	1	6	2	5	2
4	4	4	3	0	4	1	4	2	4	2
5	5	8	3	5	0	1	8	2	5	2
6	3	3	3	3	3	0	3	2	3	2
7	3	3	3	3	3	1	0	2	3	2
8	5	6	3	5	6	1	6	0	5	2
9	3	3	3	3	3	1	3	2	0	2
10	5	5	3	5	5	1	5	2	5	0

Who can influence the CEO ? Krack-high-tech

Global Connectivity

- Networks are structurally cohesive if they remain connected even when nodes are removed



0 1 2 3

Node/point connectivity

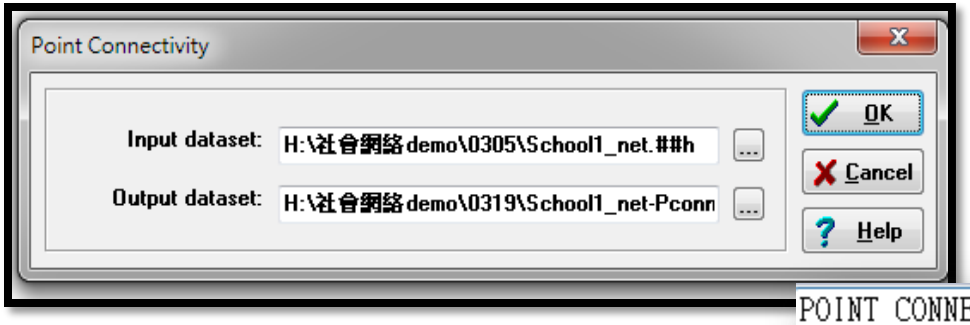
- The minimum number of nodes that need to be removed to disrupt the path **between any two pair of nodes.**
- Equal to the maximum number of **node** independent path between any two pair of nodes.

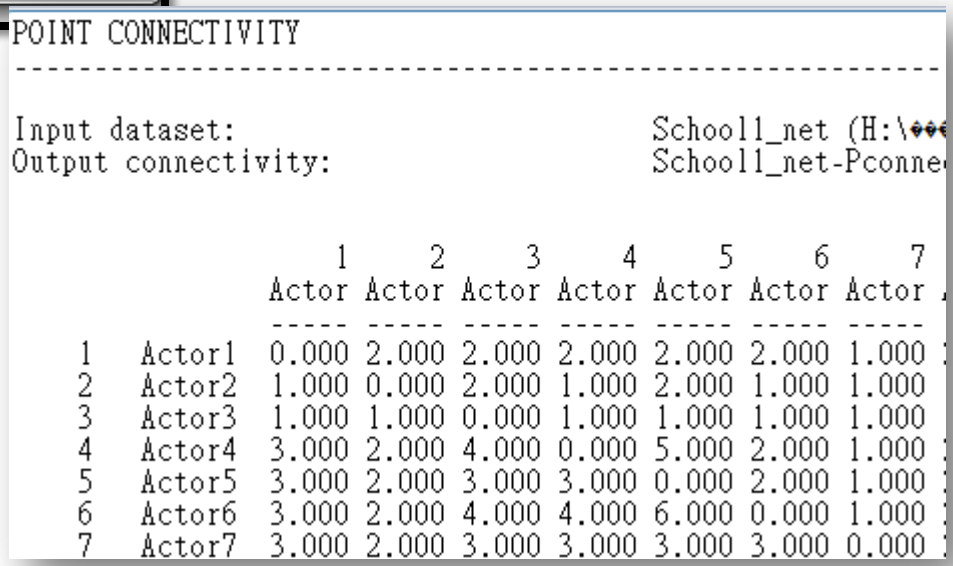
Point/local connectivity

- **NETWORK>COHESION>POINT CONNECTIVITY**
  PURPOSE Compute the local point connectivity between all pairs of nodes in a network.
  DESCRIPTION The local (point) connectivity of two non-adjacent vertices is the number of vertices that need to be deleted so that no path connects them, this is equal to the maximum number of vertex disjoint paths connecting them.
- NOTE: This procedure only operates on the first matrix in a dataset.
- Hint: data > unpack

Point Connectivity

- Compute the local point connectivity between all pairs of nodes in a network





POINT CONNECTIVITY

Input dataset: School1_net (H:\
Output connectivity: School1_net-Pconne

		1 Actor	2 Actor	3 Actor	4 Actor	5 Actor	6 Actor	7 Actor
1	Actor1	0.000	2.000	2.000	2.000	2.000	2.000	1.000
2	Actor2	1.000	0.000	2.000	1.000	2.000	1.000	1.000
3	Actor3	1.000	1.000	0.000	1.000	1.000	1.000	1.000
4	Actor4	3.000	2.000	4.000	0.000	5.000	2.000	1.000
5	Actor5	3.000	2.000	3.000	3.000	0.000	2.000	1.000
6	Actor6	3.000	2.000	4.000	4.000	6.000	0.000	1.000
7	Actor7	3.000	2.000	3.000	3.000	3.000	3.000	0.000

Local connectivity: the minimum # of nodes need to be removed to disconnect a pair of nodes

	1	2	3	4	5	6	7	8	9	10
	C	C	E	I	M	W	N	U	W	W
1	5	5	3	4	5	1	6	4	4	3
2	5	8	3	5	8	1	6	5	3	4
3	3	3	4	4	3	1	4	3	3	3
4	5	5	3	5	5	1	5	4	3	4
5	5	8	3	5	8	1	6	5	3	5
6	1	1	1	1	1	1	2	1	2	1
7	5	6	3	5	6	1	6	4	2	3
8	5	5	3	5	5	1	5	5	4	4
9	3	3	3	3	3	1	3	3	3	3
10	4	5	3	4	5	1	4	4	3	5

Flow

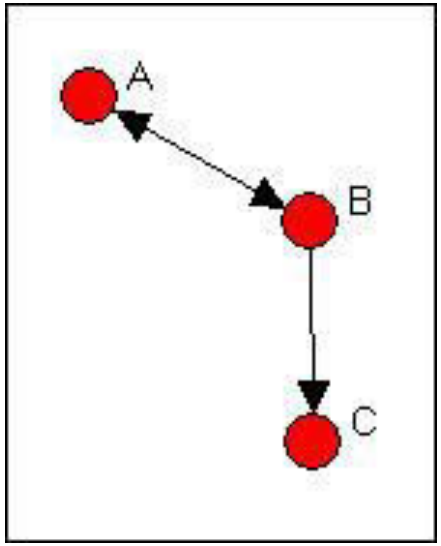
- Complementing geodesic paths
- Taking into account ALL connections between pairs of actors, **not just the most efficient ones**.
- Network>Cohesion>Maximum Flow
  - How many different actors in the NEIGHBORHOOD of a source leads to pathways to a target.

Flow

- "If I need to send a message to you, and there is only one other person to whom I can send this for retransmission, my connection is weak– **even if the person I send it to may have many ways of reaching you**.
- If, on the other hand, there are four people to whom I can send my message, **each of whom has one or more ways of retransmitting my message to you**, then my connection is stronger."

Reciprocity in directional network

- Four possible dyadic relationships with directed data
- A network that has a predominance of **null or reciprocated ties** over asymmetric connections may be a more "equal" or "stable" network than **one with a predominance of asymmetric connections**



Reciprocity

- *Network>Cohesion>Reciprocity*
- Method
  - Arc-based:
    - "Hybrid"

(A, B), (B, C) (C, A) > 1/3

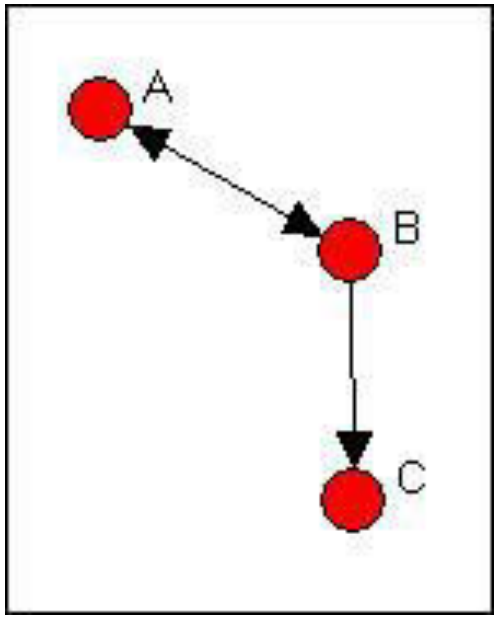
  - Dyad-based

All actual pairs
(A, B), (B, C) > 1/2

AB, BA, BC, CB, CA, AC > 1/3

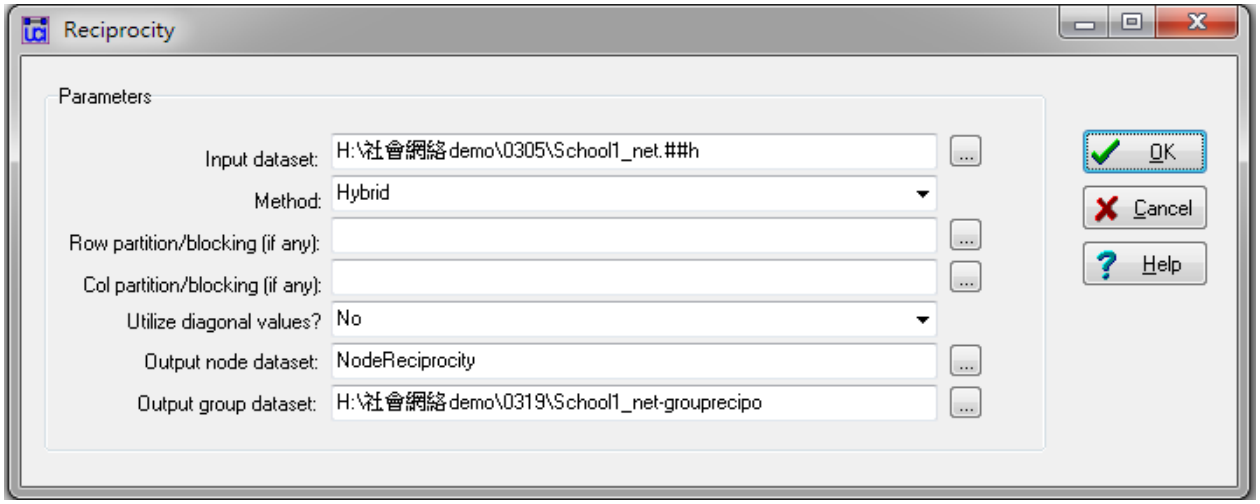
All actual ties
AB, BA, BC, CB, CA, AC > 2/3

Consider all possible pairs
Is less sensible in a large network, why?

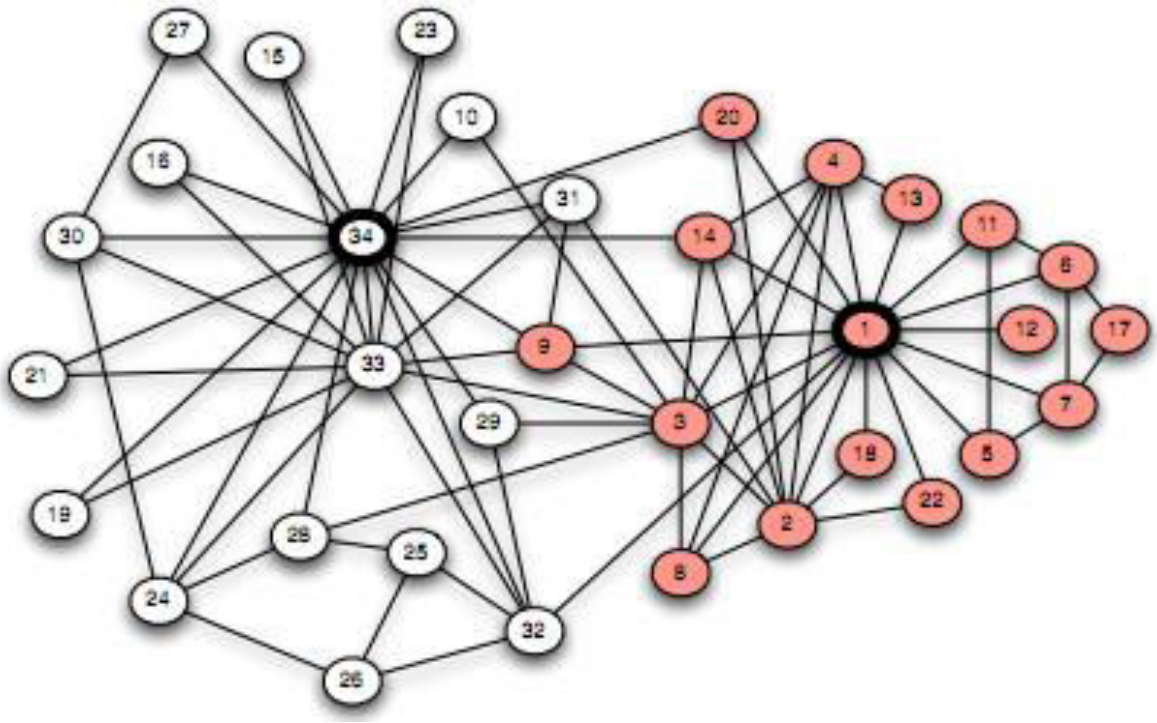


Reciprocity

- Calculate the amount of ties that are reciprocated in a network.
- Use default settings.
- Suggestion:
  save the text file for detail info.



E-I: Group External vs. Internal ties



E-I: Group External vs. Internal ties

- Takes the number of external group ties, subtracts the number of internal ties, and divides by the total number of ties
- The lower the E-I, the higher the degree of **closure**
- Range from -1 (all ties are internal; homophile or closure) to +1 (all ties are external)
- Directions of ties are ignored

E-I index

- Units of analysis
  - Entire population
  - Each sub-group
  - Individual

External and internal ties

- Network>Cohesion>E-I index
  - Given **a partition** of a network into a number of **mutually exclusive <u>groups</u>**, then
  - the E-I index is the number of ties external to the groups minus the number of ties that are internal to the group divided by the total number of ties.

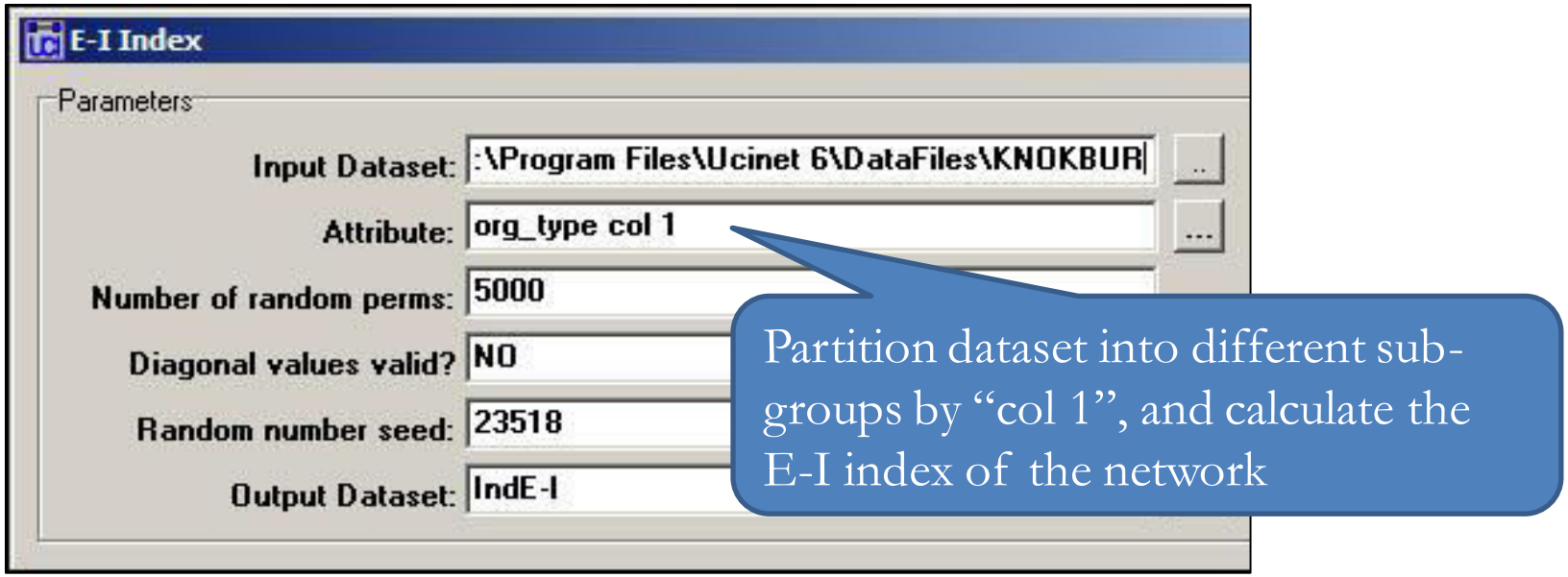
You can also try

Network>Cohesion>Homophily

more about this later

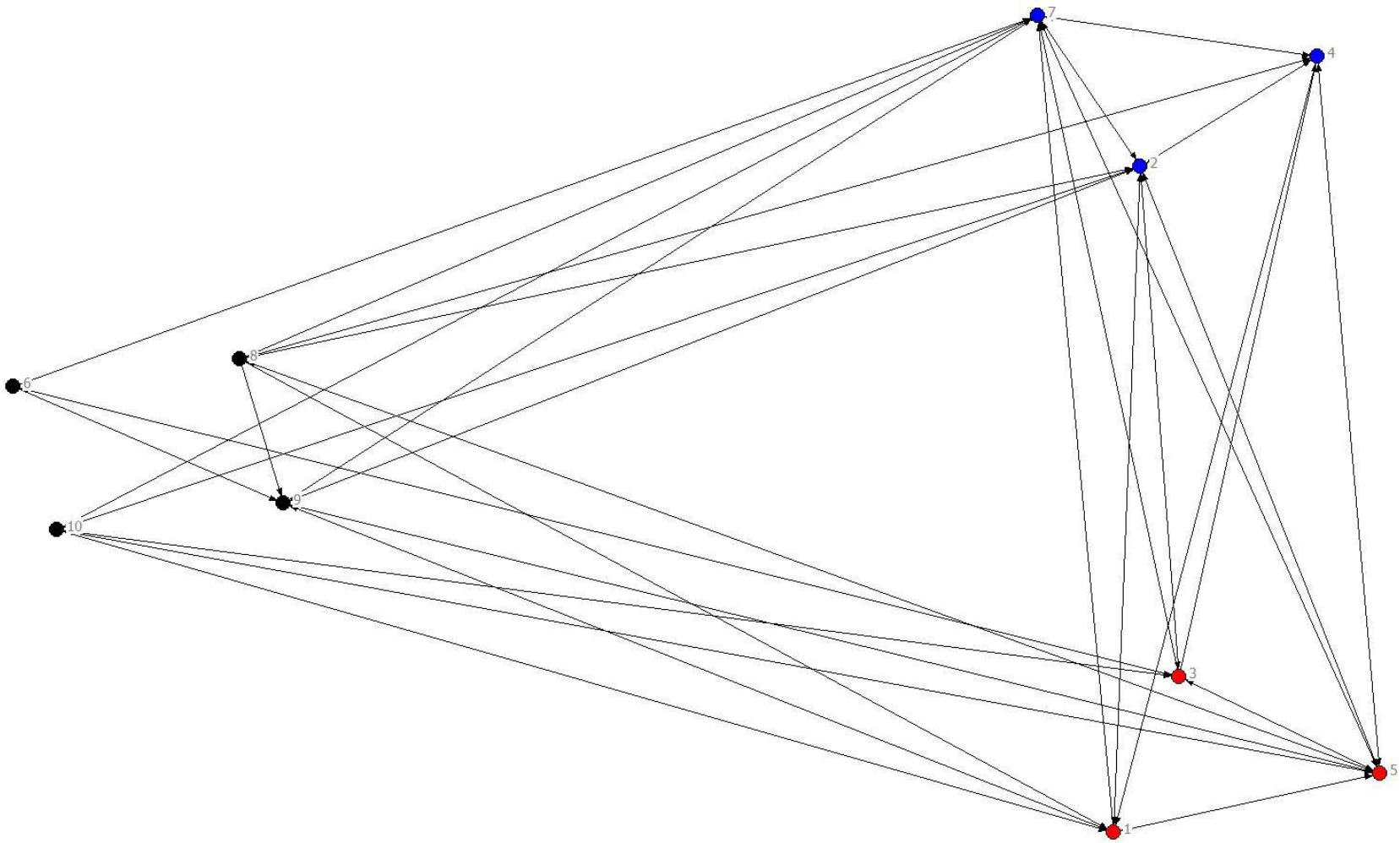
E-I Index

- *Network>Cohesion>E-I Index*
- Compare the numbers of ties within groups and between groups.
  - The directions of ties are ignored



nologies\datafiles\High-Tec-Attributes.##h" col 1

E-I



	1 Freq	2 Pct	3 Possib	4 Densit
1 Internal	14.000	0.219	24.000	0.583
2 External	50.000	0.781	66.000	0.758
3 E-I	36.000	0.563	42.000	0.467

```
Max possible external ties: 66.000
Max possible internal ties: 24.000

E-I Index: 0.563
Expected value for E-I index is: 0.467

Max possible E-I given density & group sizes: 1.000
Min possible E-I given density & group sizes: 0.250

Re-scaled E-I index: -0.167
```

the expected value under a random distribution (given the size of the group and the overall density)

Re-scaled E-I Index by the demographic constraints and overall density.

```
Permutation Test
Number of iterations = 5000
```

P=.203, not significant

	1 Obs	2 Min	3 Avg	4 Max	5 SD	6 P >= Ob	7 P <= Ob
1 Internal	0.219	0.625	0.733	0.844	0.039	1.000	0.000
2 External	0.781	0.156	0.267	0.375	0.039	0.000	1.000
3 E-I	0.563	0.250	0.467	0.688	0.078	0.203	0.953

Group level E-I Index

	1 Intern	2 Extern	3 Total	4 E-I
1 1	4.000	17.000	21.000	0.619
2 2	6.000	17.000	23.000	0.478
3 3	4.000	16.000	20.000	0.600

Individual Level E-I Index

	1 Inter	2 Exter	3 Total	4 E-I
1	1.000	6.000	7.000	0.714
2	2.000	6.000	8.000	0.500
3	1.000	5.000	6.000	0.667
4	2.000	4.000	6.000	0.333
5	2.000	6.000	8.000	0.500
6	1.000	2.000	3.000	0.333
7	2.000	7.000	9.000	0.556
8	1.000	5.000	6.000	0.667
9	2.000	4.000	6.000	0.333
10	0.000	5.000	5.000	1.000

Raw counts of ties within and between each group.

E-I Index for each group.

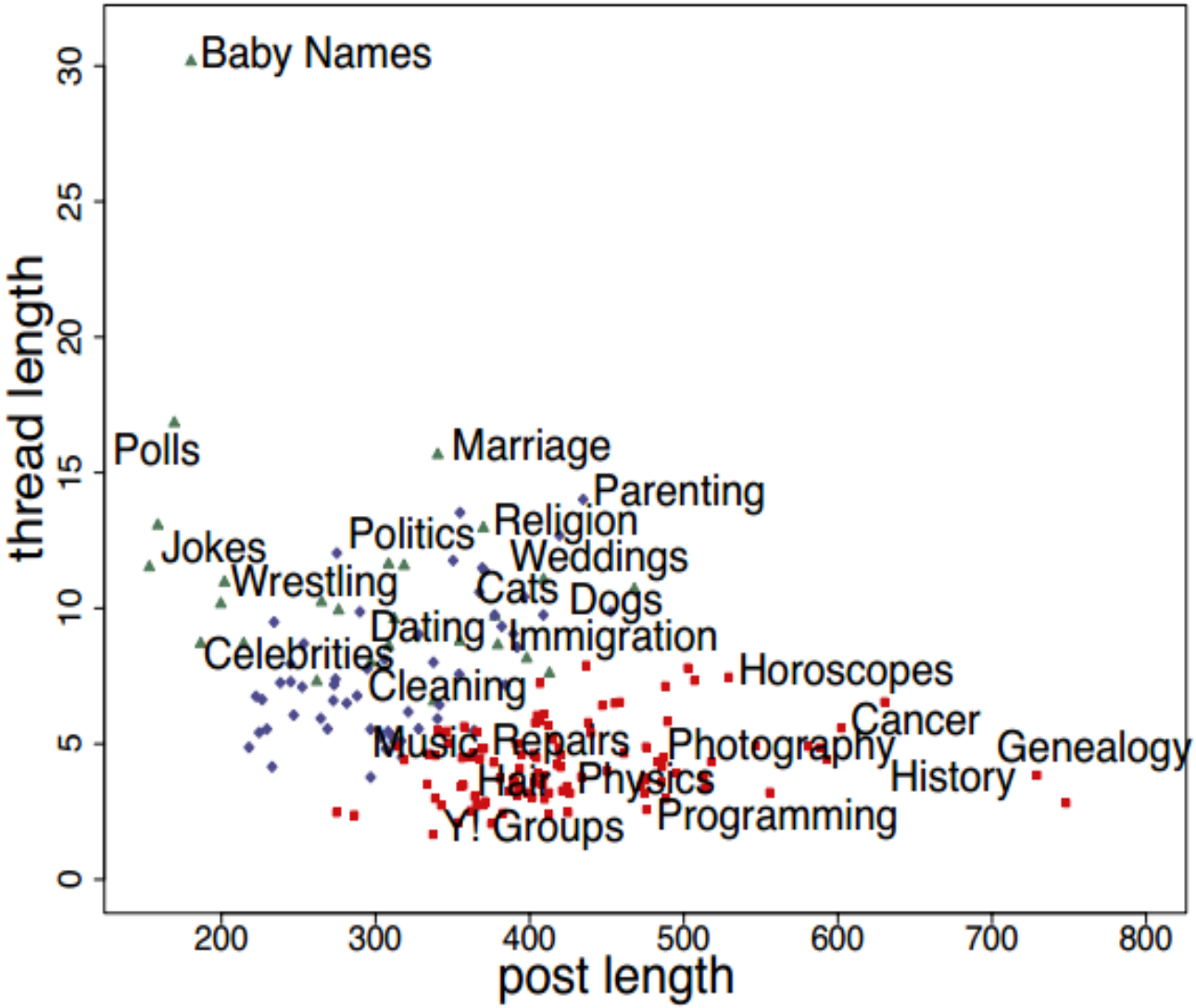
Exercise I

- Use Friendship data and do the following analyses
- 1. Run density within Gender and Club
- 2. Run E-I Index on both attributes and see if there exists homophily by either attribute
- 3. Identify the most outward/Inward - looking group and individuals
- 4. Which gender is more outward looking?

Exercise II

- Apply E-I index analysis on Krack High Tech, and Try both level (column 3) and department (column 4)
  - The procedure only runs on the first spreadsheet
    - Run on report-to network and check the E-I value

Yahoo Answers (YA)



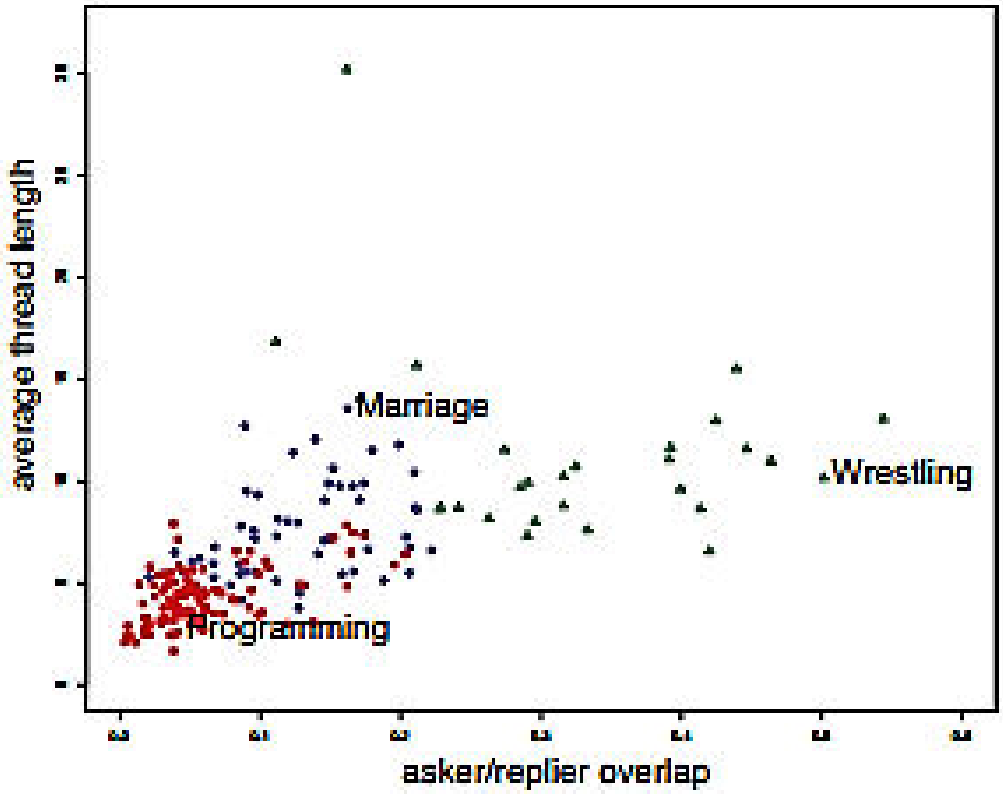


Figure 2: Clustering of categories by thread length and overlap between askers and repliers

Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008, April). Knowledge sharing and yahoo answers: everyone knows something. In*Proceedings of the 17th international conference on World Wide Web* (pp. 665-674). ACM.

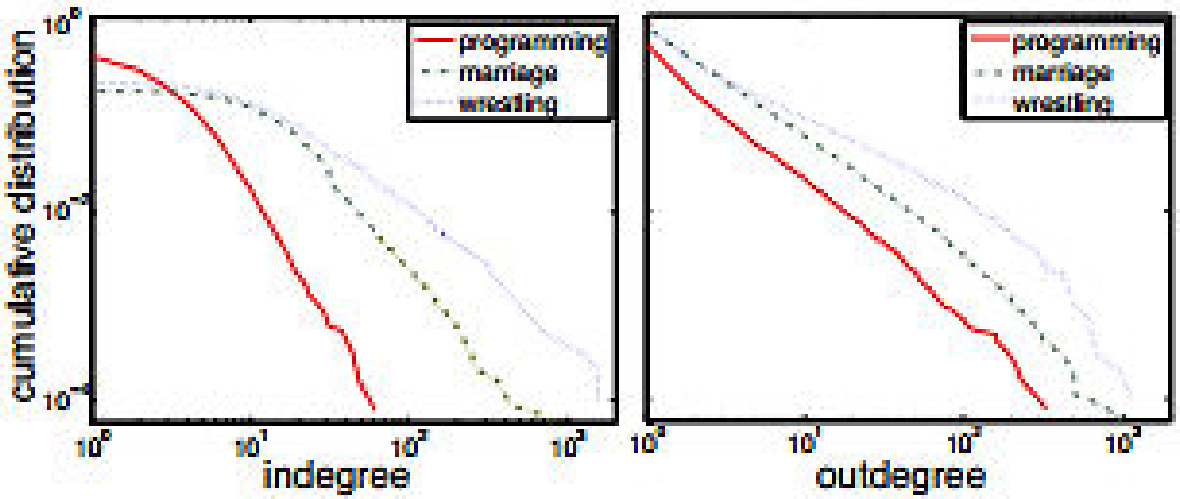


Figure 3: Distributions of indegree (number of users one has received answers from) and outdegree (number of users one has answered)

Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008, April). Knowledge sharing and yahoo answers: everyone knows something. In*Proceedings of the 17th international conference on World Wide Web* (pp. 665-674). ACM.

Table 1: Summary statistics for selected QA networks

Category	Nodes	Edges	Avg. deg.	Mutual edges	SCC
Wrestling	9,959	56,859	7.02	1,898	13.5%
Program.	12,538	18,311	1.48	0	0.01%
Marriage	45,090	164,887	3.37	179	4.73%

Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008, April). Knowledge sharing and yahoo answers: everyone knows something. In *Proceedings of the 17th international conference on World Wide Web* (pp. 665-674). ACM.

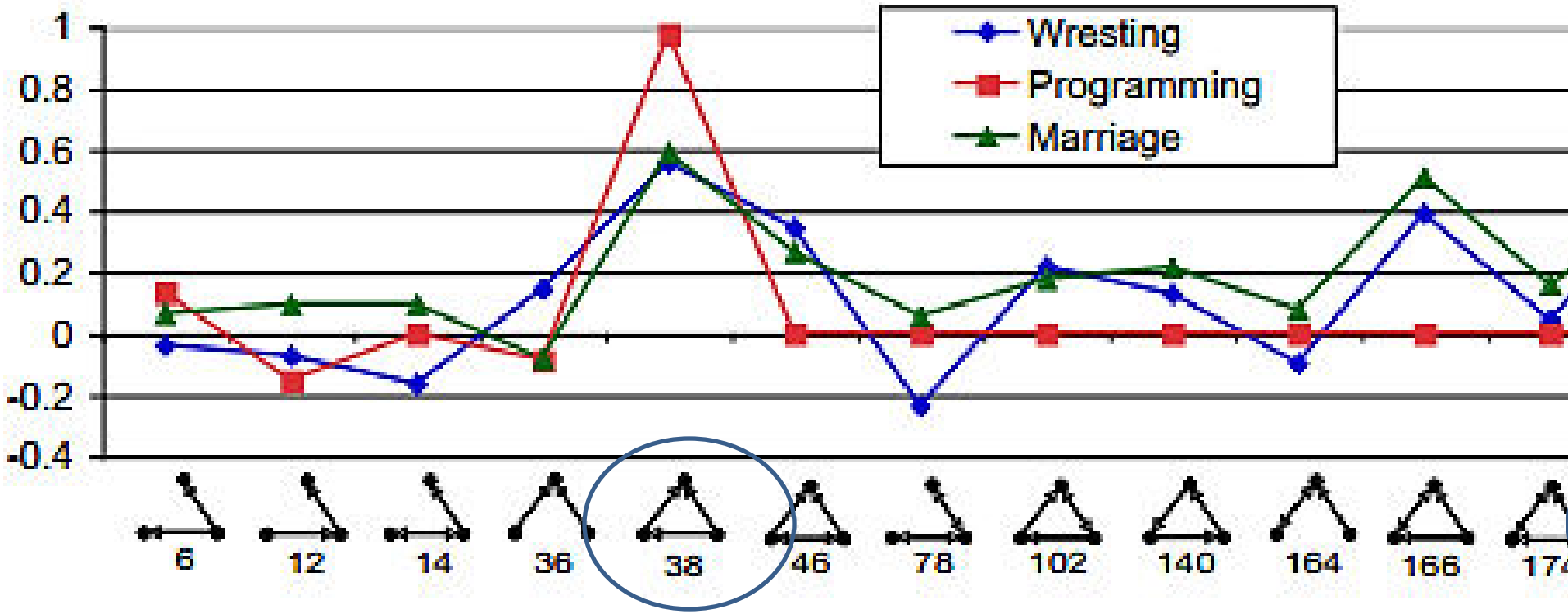


Figure 5: Motif profiles of selected categories

Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008, April). Knowledge sharing and yahoo answers: everyone knows something. In*Proceedings of the 17th international conference on World Wide Web* (pp. 665-674). ACM.

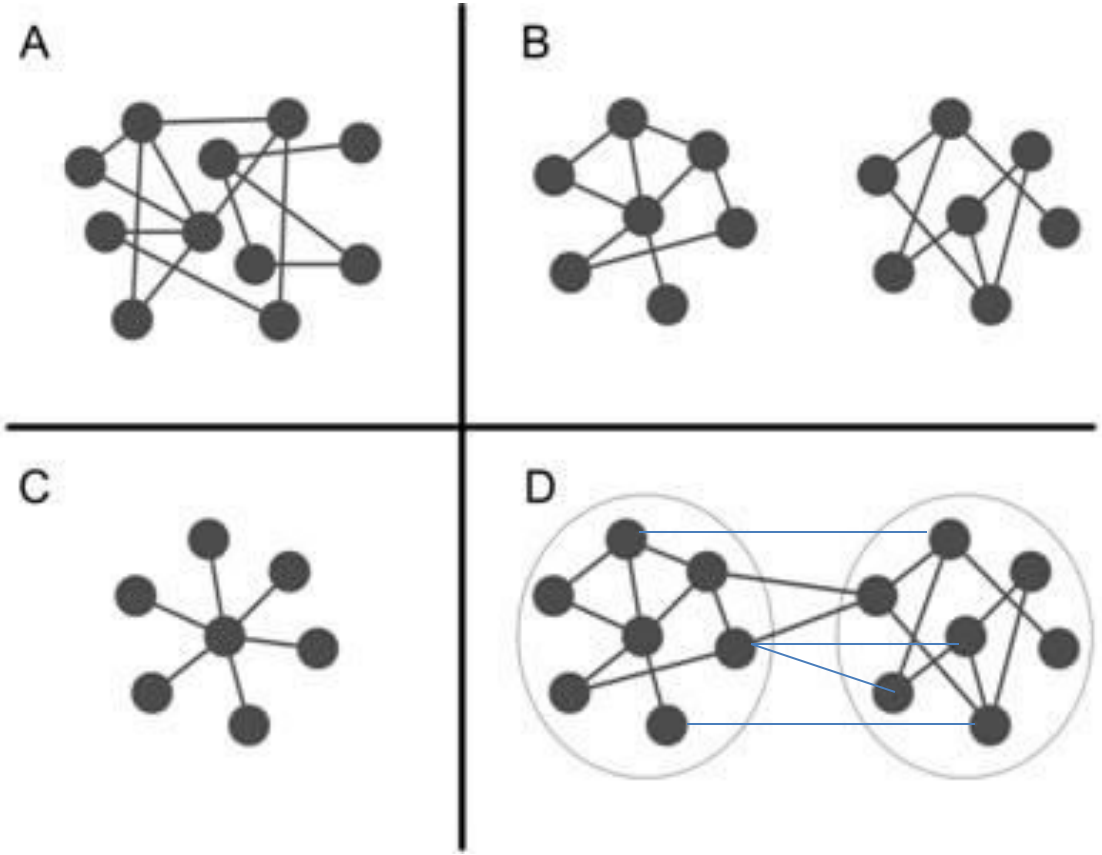
SNA

Cohesion, geodesic distance

Cohesion

- **Don't confuse the measure with the construct**
- Many ways to define or theoretically conceive of cohesion
- What is the mechanism that would relate cohesion to the outcome of interest?
- Define cohesion consistent with this mechanism
- For each way, we can then devise an operational measurement

Cohesion



Cohesions

- Network>Cohesion>Density
- Network>Cohesion>Reachability
- Network>Cohesion>No. of Geodesics
  - Geodesics?
- Network>Cohesion>Point Connectivity

More functions…

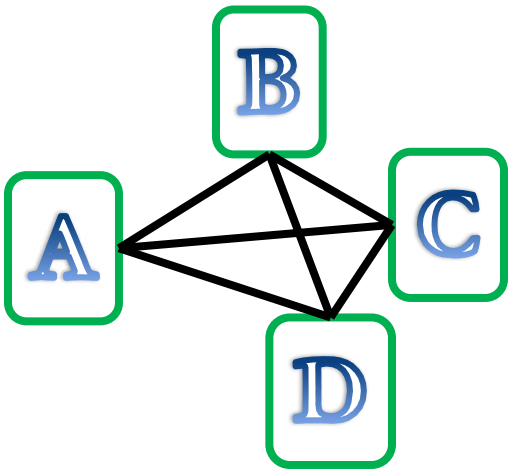
- Network>Cohesion>Density>Density of subgroups
- Network>Network properties>Reciprocity
  - For directional network
- Network>Cohesion>E-I index

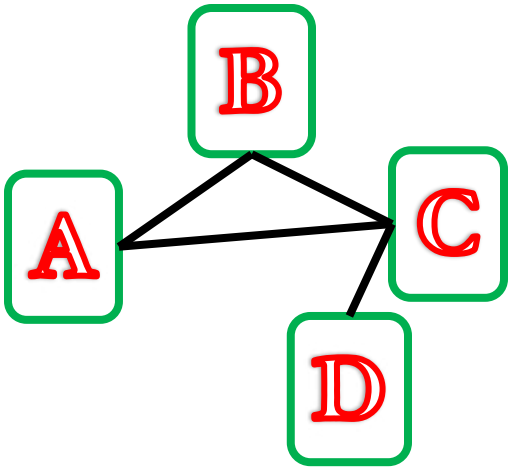
Density

- An index of the degree of dyadic connection in a group
  - Binary data
    - The ration of the number of adjacencies that are present divided by the number of all possible pairs

Density

- The density of a <u>binary</u> network is the total number of ties divided by the total number of possible ties





Density

- Valued data
  - **The sum of the values of all ties divided by the number of possible ties**
  - The average strength of ties of across all possible (not all actual) ties

Density by group

- Identify subgroups in a network
- Calculate the density, sum or average value **within** and **between** groups in a network or matrix.

Density by groups

Open KNOKBUR

Create an attribute file (##d ##h)

- Group 1: governmental agencies: coun, educ, mayr 2 out of 3 links present
- Group 2: non-governmental agencies : comm, indu, news
- Group 3: welfare specialists: wro, uway, welf, west

Density in UCINET and Gephi

Try Network>Cohesion>Overall Density

Try Network>Cohesion>Density by group

KNOKBUR

You gonna need attribute data

##D#, ##H

Also show the attributes using NetDraw

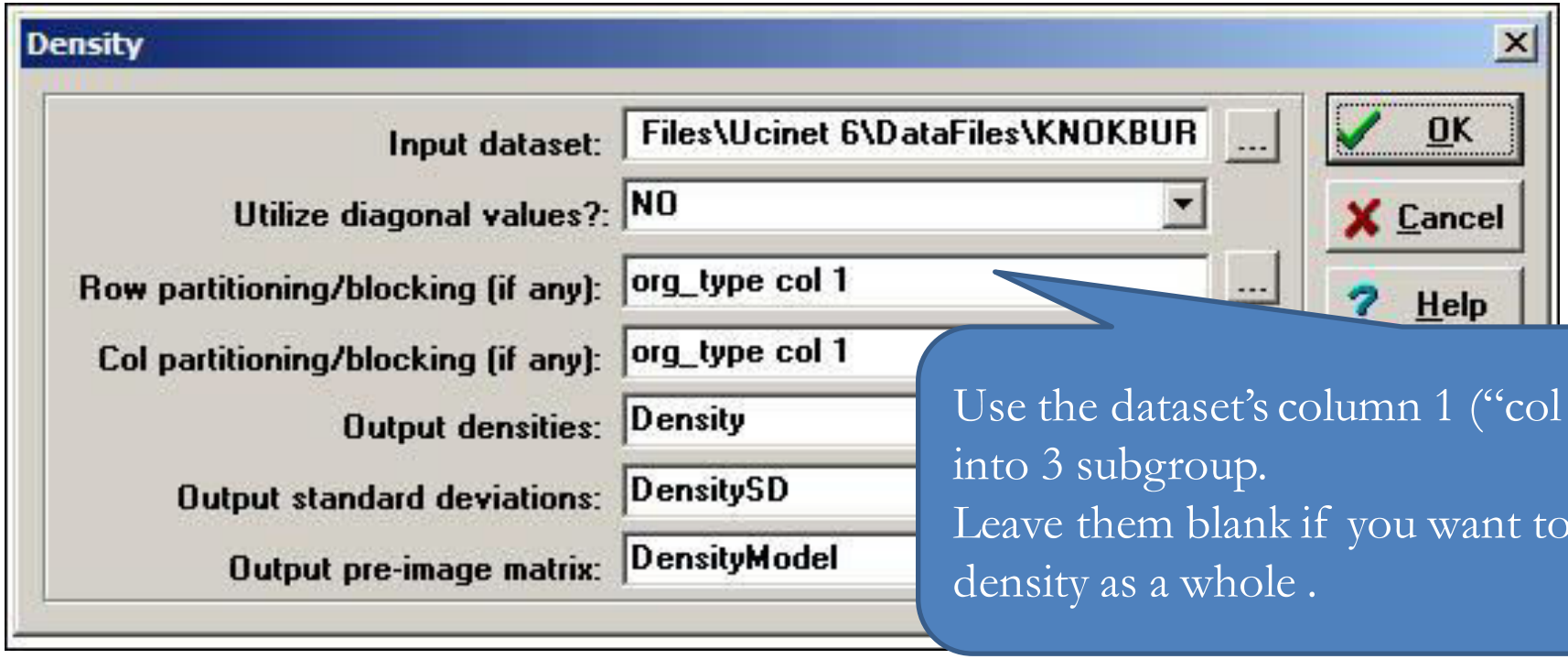
Do the same exercise with friendship data

Density

- Exercise/unpack KRACK-HIGH-TECH
  - Open KRACKAD and KRACKFR from NetDraw
  - Use UCINET to compare the density between friendship and advice seeking density
  - Network>cohesion>density (overall)
    - * by group ? (Please use the old procedure, I have run into all kinds of problems with the "New")

Within group density

- *Network>Cohesion>Density*
  - To calculate the density of whole populations, or of partitions
  - We can use an attribute or partition to divide the cases



Column Block	Old Code	Members:
1	1	COUN EDUC MAYR
2	2	COMM INDU NEWS
3	3	WRO UWAY WELF WEST

Relation: KNOKI

	1 C	3 E	5 M	2 C	4 I	7 N	6 W	8 U	9 W	10 W
1			1	1		1			1	
3			1	1	1	1	1			1
5	1	1		1	1	1		1	1	1
2	1	1	1		1	1		1	1	
4	1		1	1		1				
7			1	1	1					
6		1				1			1	
8	1		1	1	1	1			1	
9			1	1		1				
10	1	1	1	1		1				

This shows the way you partitioned the cases. "Old Code" is from column 1.

The three sub-groups and the connections WITHIN them.

```
Density / average value within blocks

            1       2       3
            1       2       3
        ------  ------  ------
  1 1   0.6667  0.8889  0.5000
  2 2   0.6667  1.0000  0.1667
  3 3   0.5833  0.6667  0.1667


Standard Deviations within blocks

            1       2       3
            1       2       3
        ------  ------  ------
  1 1   0.4714  0.3143  0.5000
  2 2   0.4714  0.0000  0.3727
  3 3   0.4930  0.4714  0.3727
```

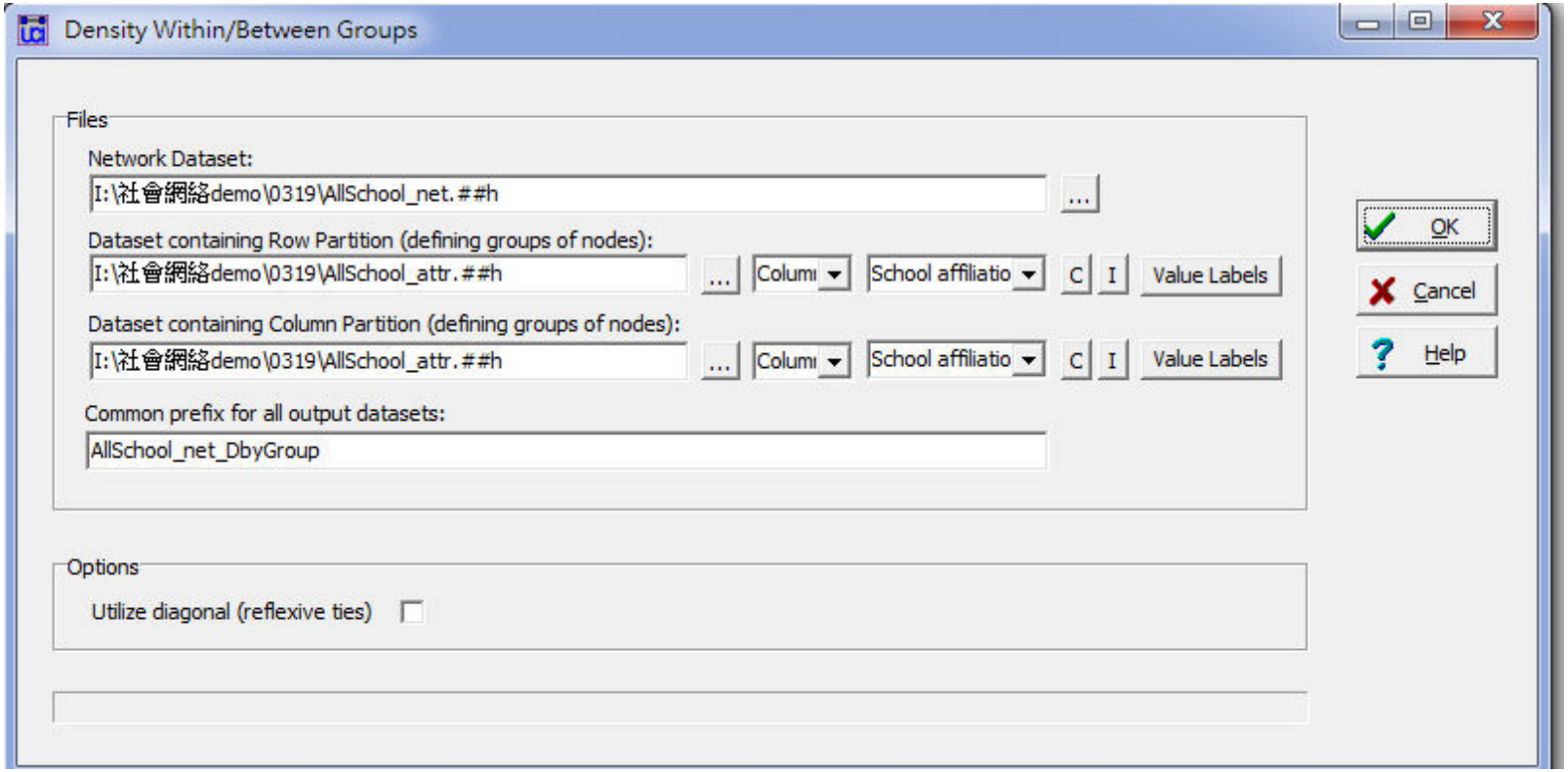
The density within- and between-blocks.

The SD within- and between-blocks.

Exercise, apply density analysis on NTNU data

Simplified NTNU

Save into excel file then open with Gephi



Density by groups

Number of ties/Sum of tie-strengths

	1 0	2 1	3 2	4 3	5 4	6 5	7 6	8 7	9 8
1 0	0.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
2 1	1.000	138.000	35.000	19.000	25.000	14.000	19.000	0.000	16.000
3 2	0.000	36.000	79.000	21.000	16.000	6.000	30.000	4.000	47.000
4 3	0.000	10.000	25.000	118.000	17.000	9.000	9.000	6.000	85.000
5 4	0.000	4.000	1.000	12.000	59.000	12.000	15.000	2.000	14.000
6 5	1.000	20.000	14.000	12.000	36.000	30.000	24.000	5.000	12.000
7 6	0.000	7.000	11.000	11.000	11.000	16.000	61.000	2.000	15.000
8 7	0.000	11.000	6.000	2.000	5.000	4.000	1.000	19.000	10.000
9 8	0.000	3.000	9.000	16.000	1.000	4.000	1.000	7.000	11.000

Density (prop of ties) / Average tie strength

	1 0	2 1	3 2	4 3	5 4	6 5	7 6	8 7	9 8
1 0	0.000	0.009	0.010	0.000	0.000	0.000	0.000	0.000	0.000
2 1	0.009	0.197	0.052	0.032	0.042	0.027	0.050	0.000	0.008
3 2	0.000	0.053	0.132	0.038	0.029	0.013	0.086	0.020	0.026
4 3	0.000	0.017	0.045	0.255	0.035	0.022	0.029	0.034	0.054
5 4	0.000	0.007	0.002	0.025	0.128	0.029	0.049	0.011	0.009
6 5	0.013	0.039	0.029	0.029	0.086	0.088	0.090	0.033	0.009
7 6	0.000	0.019	0.031	0.036	0.036	0.060	0.335	0.018	0.015
8 7	0.000	0.051	0.030	0.011	0.028	0.026	0.009	0.339	0.018
9 8	0.000	0.002	0.005	0.010	0.001	0.003	0.001	0.012	0.002

Reachability

- An actor is reachable by another if there exists any set of connections by which we can trace from the source to the target actor, **regardless of how many others fall between them.**

Reachability of Knoke "I" and "M" relations

Matrix #1

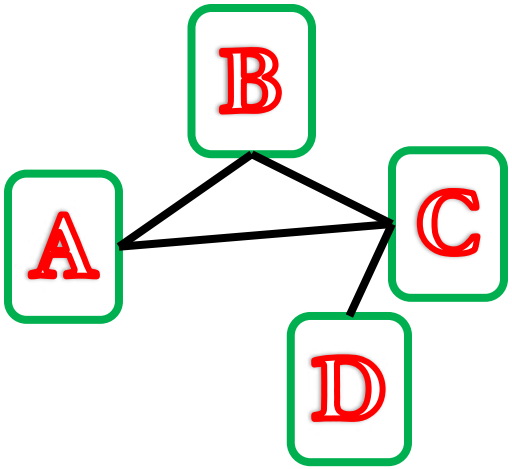
	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	1	1	1	1	1	1	1	1	1
2	1	0	1	1	1	1	1	1	1	1
3	1	1	0	1	1	1	1	1	1	1
4	1	1	1	0	1	1	1	1	1	1
5	1	1	1	1	0	1	1	1	1	1
6	1	1	1	1	1	0	1	1	1	1
7	1	1	1	1	1	1	0	1	1	1
8	1	1	1	1	1	1	1	0	1	1
9	1	1	1	1	1	1	1	1	0	1
10	1	1	1	1	1	1	1	1	1	0

Matrix #2

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	1	1	0	1	0	0	1	1	1
2	0	0	1	0	0	0	0	1	1	1
3	0	0	0	0	0	0	0	1	1	1
4	0	1	1	0	0	0	1	1	1	1
5	0	1	1	0	0	0	0	1	1	1
6	0	0	0	0	0	0	0	0	0	0
7	0	1	1	0	0	0	0	1	1	1
8	0	0	1	0	0	0	0	0	1	1
9	0	0	1	0	0	0	0	1	0	1
10	0	0	0	0	0	0	0	0	0	0

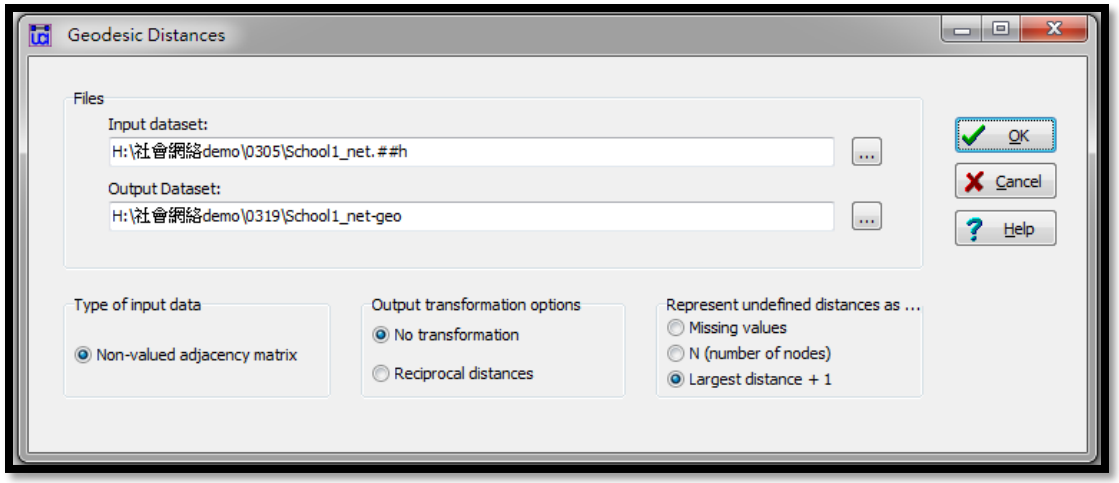
(Geodesic) Distance

- The distance between two nodes is the length of the shortest path.
- Isolates?



Distance

- Choose input, choose output.
- Use default settings.
- Suggestion: **save the text file** for average and frequencies.



	Actor1	Actor2	Actor3	Actor4	Actor5	Actor6	Actor7
Actor1	0	2	2	1	1	2	3
Actor2	2	0	1	2	1	2	3
Actor3	2	3	0	2	1	2	3
Actor4	1	1	1	0	1	1	2
Actor5	1	2	1	1	0	1	2
Actor6	1	1	1	1	1	0	1
Actor7	2	2	2	1	1	1	0

Geodesic distances for Knoke information exchange

Geodesic Distances

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	1	2	2	1	3	1	2	1	2
2	1	0	1	1	1	2	1	1	1	2
3	2	1	0	1	1	1	1	2	2	1
4	1	1	2	0	1	3	1	2	2	2
5	1	1	1	1	0	2	1	1	1	1
6	3	2	1	2	2	0	1	3	1	2
7	2	1	2	1	1	3	0	2	2	2
8	1	1	2	1	1	3	1	0	1	2
9	2	1	2	2	1	3	1	2	0	2
10	1	1	1	2	1	2	1	2	2	0

Compare 6 to 4 and 4 to 6, are they symmetrical

Save the output file and open it with netdraw
Reverse the value?

Who is close to the CEO ? Krack-high-tech

Distance in a valued graph: tie strength

- How to measure distance with valued data?
- The distance between two actors is defined as the strength of the weakest link in the shortest path.
  - If A sends 6 units to B, B sends 4 units to C, the strength of the path from A to C (**assuming A to B to C is the shortest path**) is 4.
  - Wolf/fn (first save the data into separate file, then use the OLD procedure)

No. of Geodesics

- Counts the number of different geodesic paths connecting each pair of nodes.
- Binary dataset only, or UCINET will warn you.
- What does the mean?

	Actor1	Actor2	Actor3	Actor4	Actor5	Actor6	Actor7	Actor8
Actor1	1	1	2	1	1	2	2	1
Actor2	1	1	1	1	1	1	1	2
Actor3	1	2	1	1	1	1	1	2
Actor4	1	1	1	1	1	1	1	2
Actor5	1	2	1	1	1	1	1	2
Actor6	1	1	1	1	1	1	1	1
Actor7	3	2	3	1	1	1	1	1
Actor8	0	0	0	0	0	0	0	1

Who can influence the CEO ? Krack-high-tech

Number of geodesic paths for Knoke information exchange

# of Geodesic Paths

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	1	1	2	3	1	2	1	2	1	1
2	1	1	1	1	1	1	1	1	1	2
3	4	1	1	1	1	1	1	2	3	1
4	1	1	2	1	1	2	1	2	3	1
5	1	1	1	1	1	1	1	1	1	1
6	9	3	1	2	3	1	1	6	1	1
7	3	1	2	1	1	2	1	2	2	1
8	1	1	2	1	1	2	1	1	1	1
9	2	1	2	3	1	2	1	2	1	1
10	1	1	1	4	1	1	1	2	3	1

Next: Run on WOLFN data

Flow

- Complementing geodesic paths
- Taking into account ALL connections between pairs of actors, **not just the most efficient ones**.
- Network>Cohesion>Maximum Flow
  - How many different actors in the NEIGHBORHOOD of a source leads to pathways to a target.

Flow

- "If I need to send a message to you, and there is only one other person to whom I can send this for retransmission, my connection is weak– **even if the person I send it to may have many ways of reaching you**.
- If, on the other hand, there are four people to whom I can send my message, **each of whom has one or more ways of retransmitting my message to you**, then my connection is stronger."

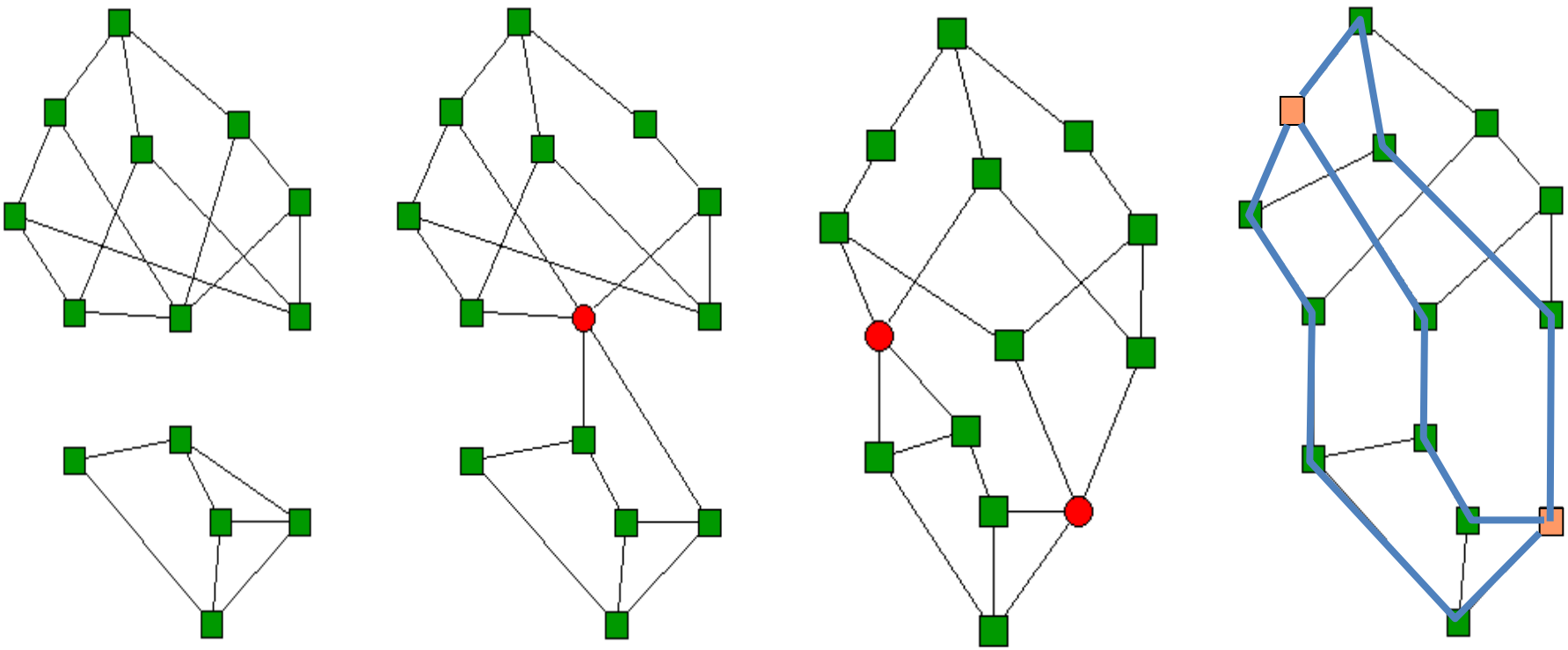
Maximum flow for Knoke information network

	1 C	2 C	3 E	4 I	5 M	6 W	7 N	8 U	9 W	10 W
1	0	4	3	4	4	1	4	2	4	2
2	5	0	3	5	7	1	7	2	5	2
3	5	6	0	5	6	1	6	2	5	2
4	4	4	3	0	4	1	4	2	4	2
5	5	8	3	5	0	1	8	2	5	2
6	3	3	3	3	3	0	3	2	3	2
7	3	3	3	3	3	1	0	2	3	2
8	5	6	3	5	6	1	6	0	5	2
9	3	3	3	3	3	1	3	2	0	2
10	5	5	3	5	5	1	5	2	5	0

Who can influence the CEO ? Krack-high-tech

Global Connectivity

- Networks are structurally cohesive if they remain connected even when nodes are removed



0 1 2 3

Node/point connectivity

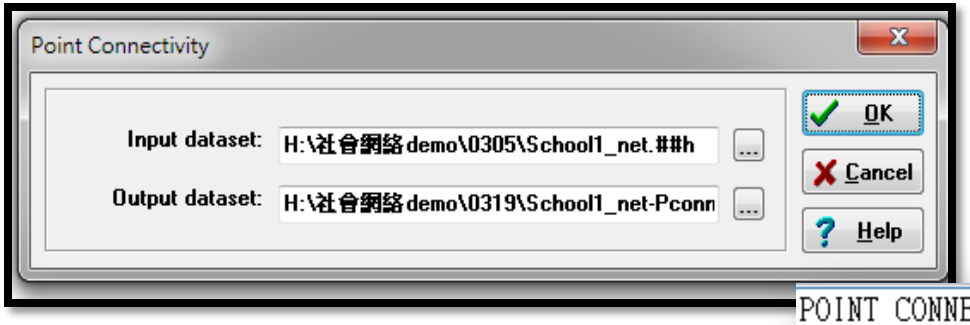
- The minimum number of nodes that need to be removed to disrupt the path **between any two pair of nodes.**
- Equal to the maximum number of **node** independent path between any two pair of nodes.

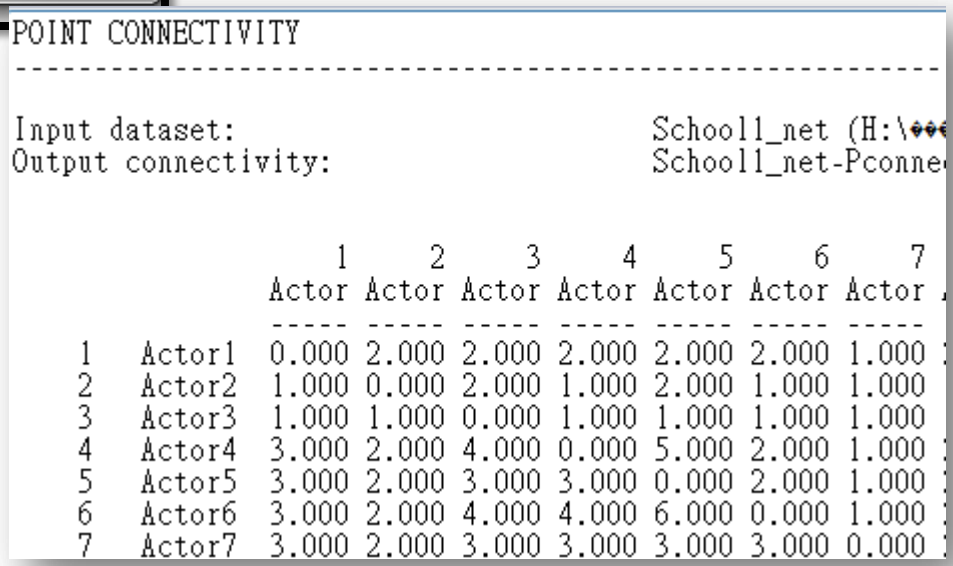
Point/local connectivity

- **NETWORK>COHESION>POINT CONNECTIVITY**
  PURPOSE Compute the local point connectivity between all pairs of nodes in a network.
  DESCRIPTION The local (point) connectivity of two non-adjacent vertices is the number of vertices that need to be deleted so that no path connects them, this is equal to the maximum number of vertex disjoint paths connecting them.
- NOTE: This procedure only operates on the first matrix in a dataset.
- Hint: data > unpack

Point Connectivity

- Compute the local point connectivity between all pairs of nodes in a network





POINT CONNECTIVITY

Input dataset: School1_net (H:\
Output connectivity: School1_net-Pconne

		1 Actor	2 Actor	3 Actor	4 Actor	5 Actor	6 Actor	7 Actor
1	Actor1	0.000	2.000	2.000	2.000	2.000	2.000	1.000
2	Actor2	1.000	0.000	2.000	1.000	2.000	1.000	1.000
3	Actor3	1.000	1.000	0.000	1.000	1.000	1.000	1.000
4	Actor4	3.000	2.000	4.000	0.000	5.000	2.000	1.000
5	Actor5	3.000	2.000	3.000	3.000	0.000	2.000	1.000
6	Actor6	3.000	2.000	4.000	4.000	6.000	0.000	1.000
7	Actor7	3.000	2.000	3.000	3.000	3.000	3.000	0.000

Local connectivity: the minimum # of nodes need to be removed to disconnect a pair of nodes

	1	2	3	4	5	6	7	8	9	10
	C	C	E	I	M	W	N	U	W	W
1	5	5	3	4	5	1	6	4	4	3
2	5	8	3	5	8	1	6	5	3	4
3	3	3	4	4	3	1	4	3	3	3
4	5	5	3	5	5	1	5	4	3	4
5	5	8	3	5	8	1	6	5	3	5
6	1	1	1	1	1	1	2	1	2	1
7	5	6	3	5	6	1	6	4	2	3
8	5	5	3	5	5	1	5	5	4	4
9	3	3	3	3	3	1	3	3	3	3
10	4	5	3	4	5	1	4	4	3	5

Flow

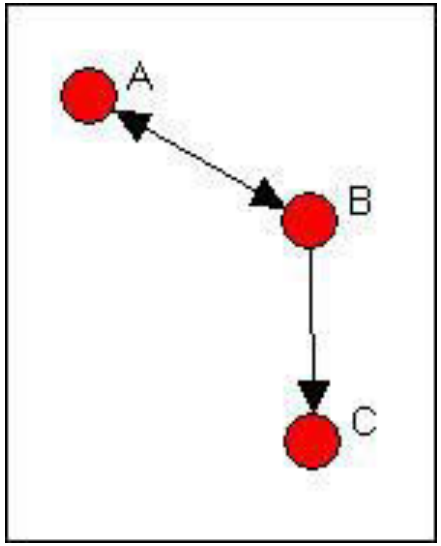
- Complementing geodesic paths
- Taking into account ALL connections between pairs of actors, **not just the most efficient ones**.
- Network>Cohesion>Maximum Flow
  - How many different actors in the NEIGHBORHOOD of a source leads to pathways to a target.

Flow

- "If I need to send a message to you, and there is only one other person to whom I can send this for retransmission, my connection is weak– **even if the person I send it to may have many ways of reaching you**.
- If, on the other hand, there are four people to whom I can send my message, **each of whom has one or more ways of retransmitting my message to you**, then my connection is stronger."

Reciprocity in directional network

- Four possible dyadic relationships with directed data
- A network that has a predominance of **null or reciprocated ties** over asymmetric connections may be a more "equal" or "stable" network than **one with a predominance of asymmetric connections**



Reciprocity

- *Network>Cohesion>Reciprocity*
- Method
  - Arc-based:
    - "Hybrid"

(A, B), (B, C) (C, A) > 1/3

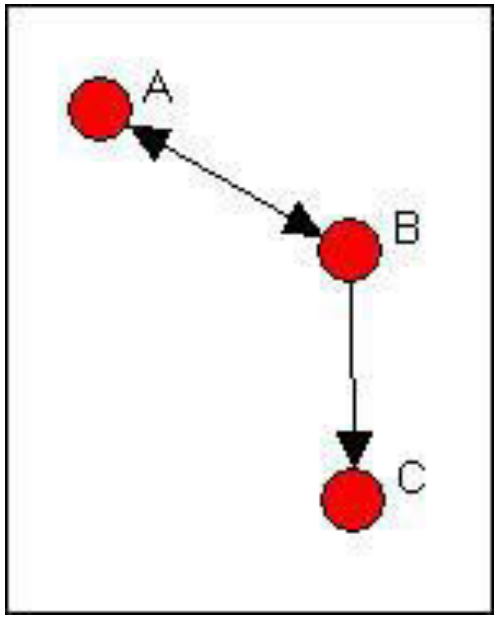
  - Dyad-based

All actual pairs
(A, B), (B, C) > 1/2

AB, BA, BC, CB, CA, AC > 1/3

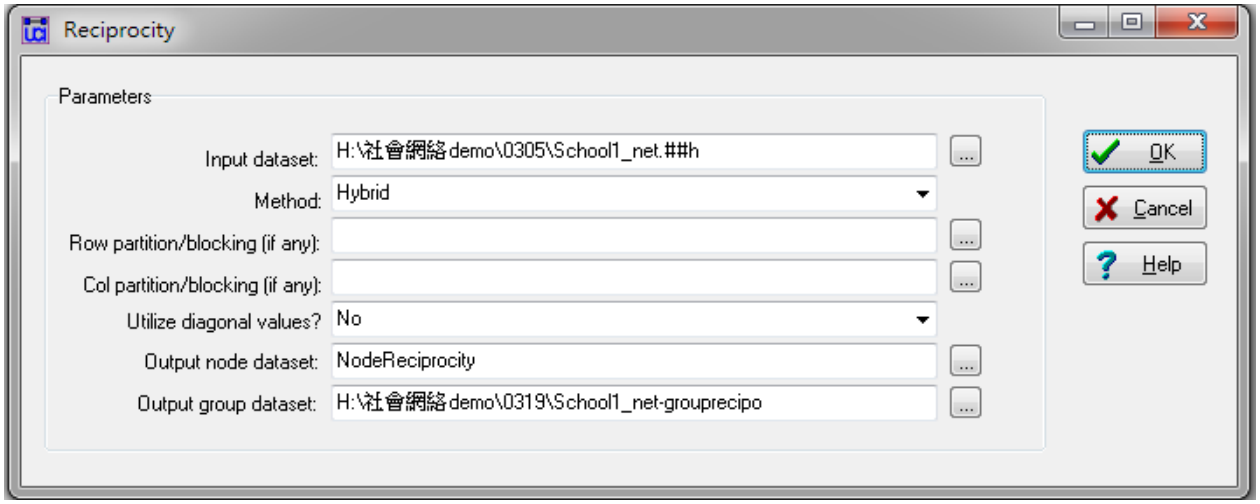
All actual ties
AB, BA, BC, CB, CA, AC > 2/3

Consider all possible pairs
Is less sensible in a large network, why?

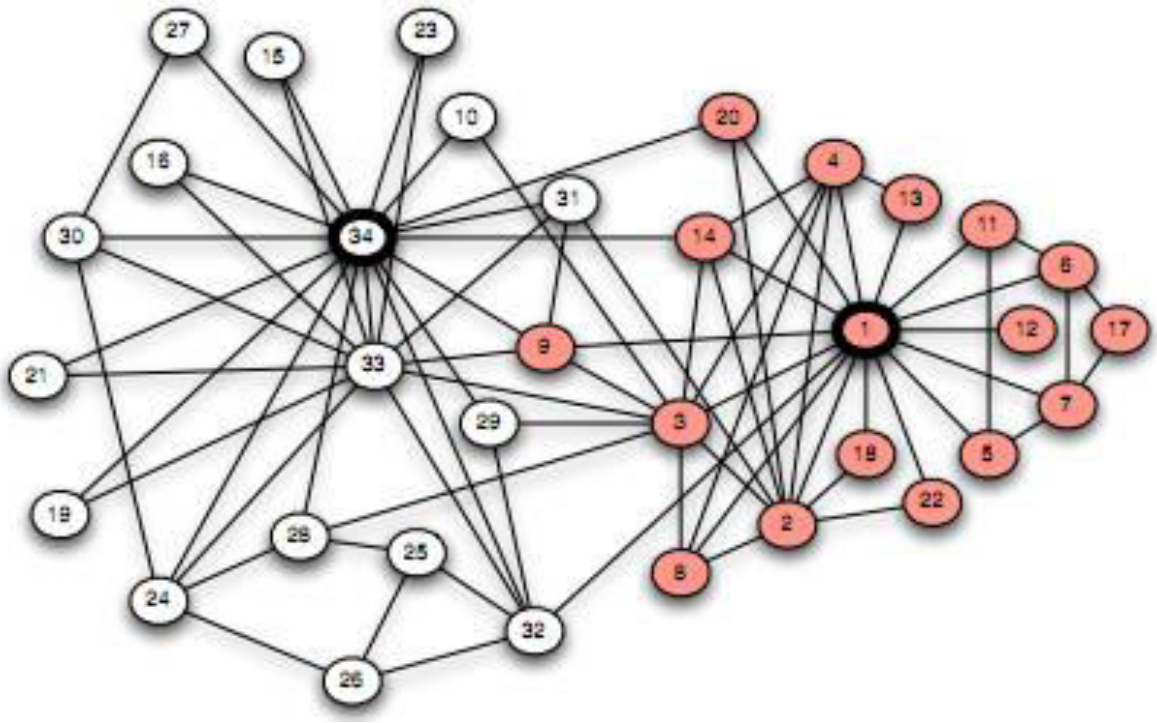


Reciprocity

- Calculate the amount of ties that are reciprocated in a network.
- Use default settings.
- Suggestion:
  save the text file for detail info.



E-I: Group External vs. Internal ties



E-I: Group External vs. Internal ties

- Takes the number of external group ties, subtracts the number of internal ties, and divides by the total number of ties
- The lower the E-I, the higher the degree of **closure**
- Range from -1 (all ties are internal; homophile or closure) to +1 (all ties are external)
- Directions of ties are ignored

E-I index

- Units of analysis
  - Entire population
  - Each sub-group
  - Individual

External and internal ties

- Network>Cohesion>E-I index
  - Given **a partition** of a network into a number of **mutually exclusive <u>groups</u>**, then
  - the E-I index is the number of ties external to the groups minus the number of ties that are internal to the group divided by the total number of ties.

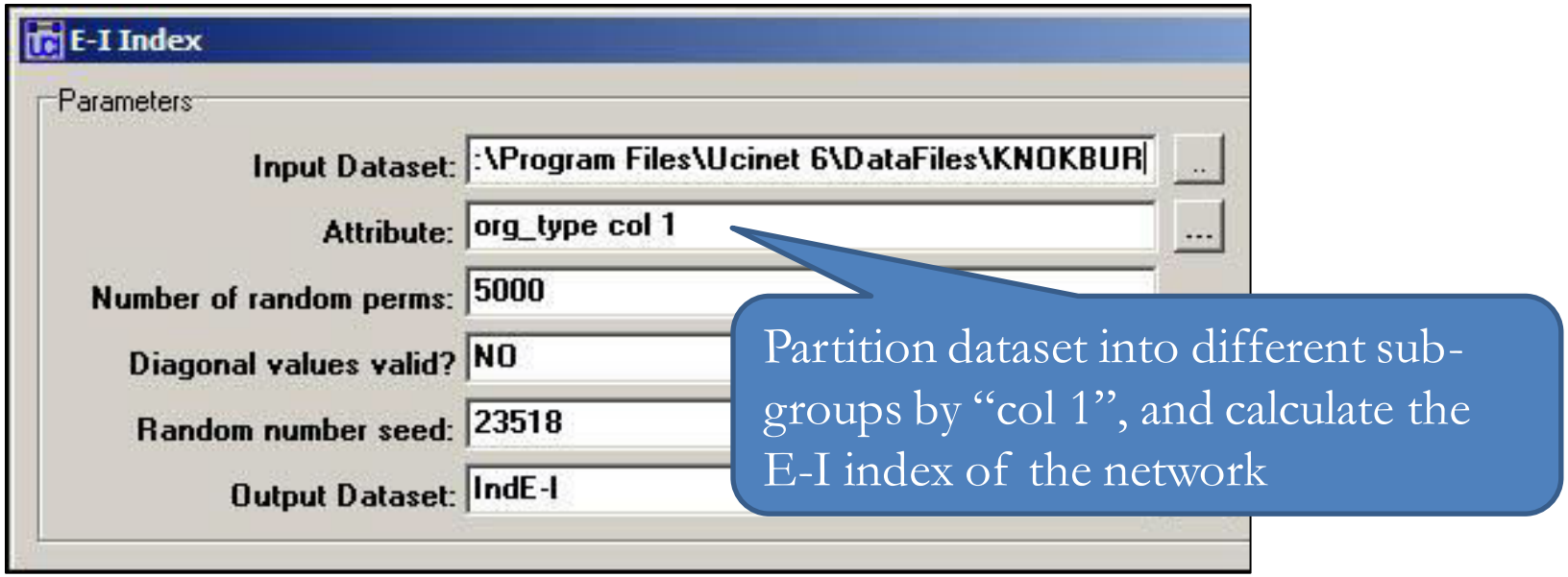
You can also try

Network>Cohesion>Homophily

more about this later

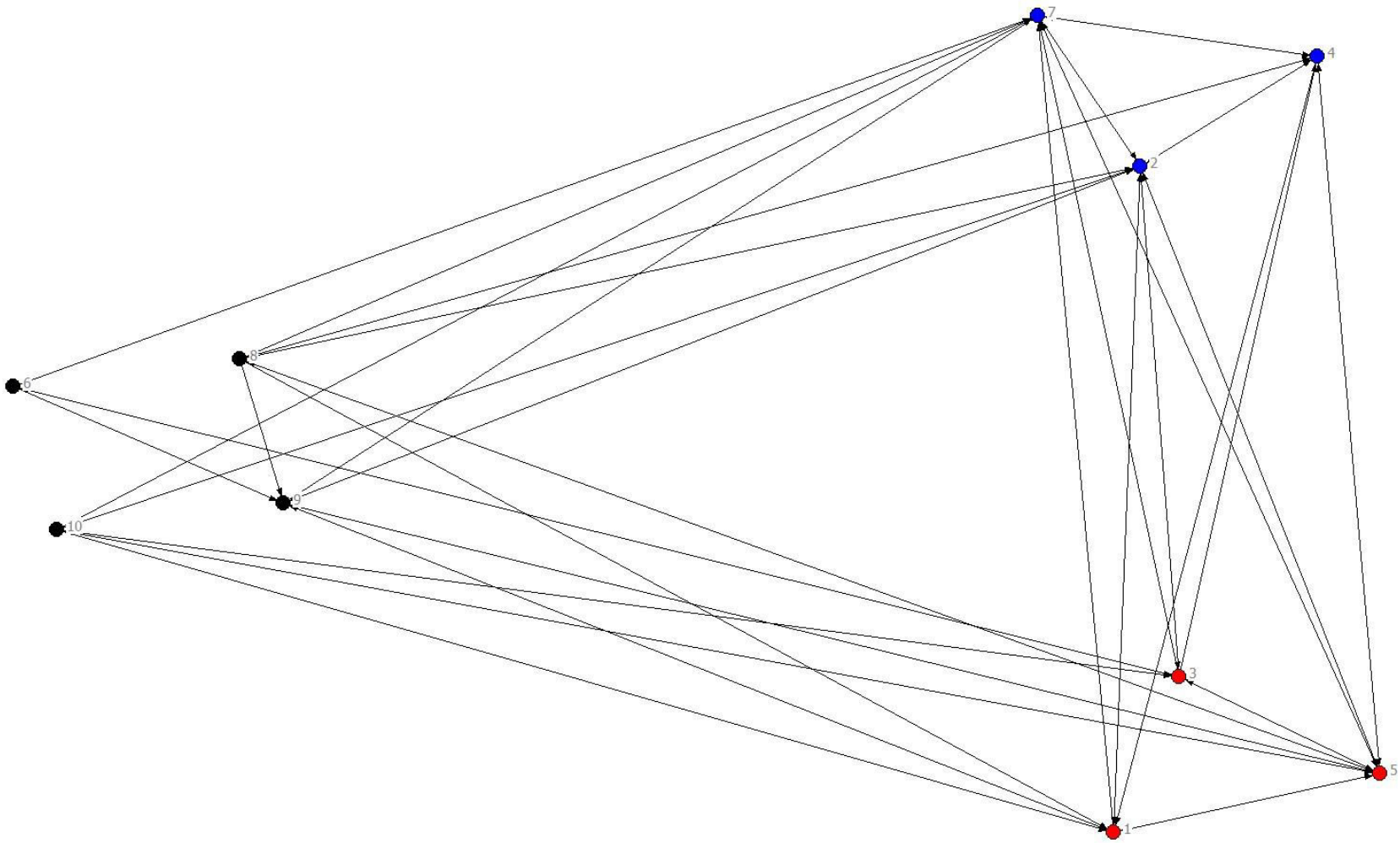
E-I Index

- *Network>Cohesion>E-I Index*
- Compare the numbers of ties within groups and between groups.
  - The directions of ties are ignored



nologies\datafiles\High-Tec-Attributes.##h" col 1

E-I



	1 Freq	2 Pct	3 Possib	4 Densit
1 Internal	14.000	0.219	24.000	0.583
2 External	50.000	0.781	66.000	0.758
3 E-I	36.000	0.563	42.000	0.467

```
Max possible external ties: 66.000
Max possible internal ties: 24.000

E-I Index: 0.563
Expected value for E-I index is: 0.467

Max possible E-I given density & group sizes: 1.000
Min possible E-I given density & group sizes: 0.250

Re-scaled E-I index: -0.167
```

the expected value under a random distribution (given the size of the group and the overall density)

Re-scaled E-I Index by the demographic constraints and overall density.

```
Permutation Test
Number of iterations = 5000
```

P=.203, not significant

	1 Obs	2 Min	3 Avg	4 Max	5 SD	6 P >= Ob	7 P <= Ob
1 Internal	0.219	0.625	0.733	0.844	0.039	1.000	0.000
2 External	0.781	0.156	0.267	0.375	0.039	0.000	1.000
3 E-I	0.563	0.250	0.467	0.688	0.078	0.203	0.953

Group level E-I Index

	1 Intern	2 Extern	3 Total	4 E-I
1 1	4.000	17.000	21.000	0.619
2 2	6.000	17.000	23.000	0.478
3 3	4.000	16.000	20.000	0.600

Individual Level E-I Index

	1 Inter	2 Exter	3 Total	4 E-I
1	1.000	6.000	7.000	0.714
2	2.000	6.000	8.000	0.500
3	1.000	5.000	6.000	0.667
4	2.000	4.000	6.000	0.333
5	2.000	6.000	8.000	0.500
6	1.000	2.000	3.000	0.333
7	2.000	7.000	9.000	0.556
8	1.000	5.000	6.000	0.667
9	2.000	4.000	6.000	0.333
10	0.000	5.000	5.000	1.000

Raw counts of ties within and between each group.

E-I Index for each group.

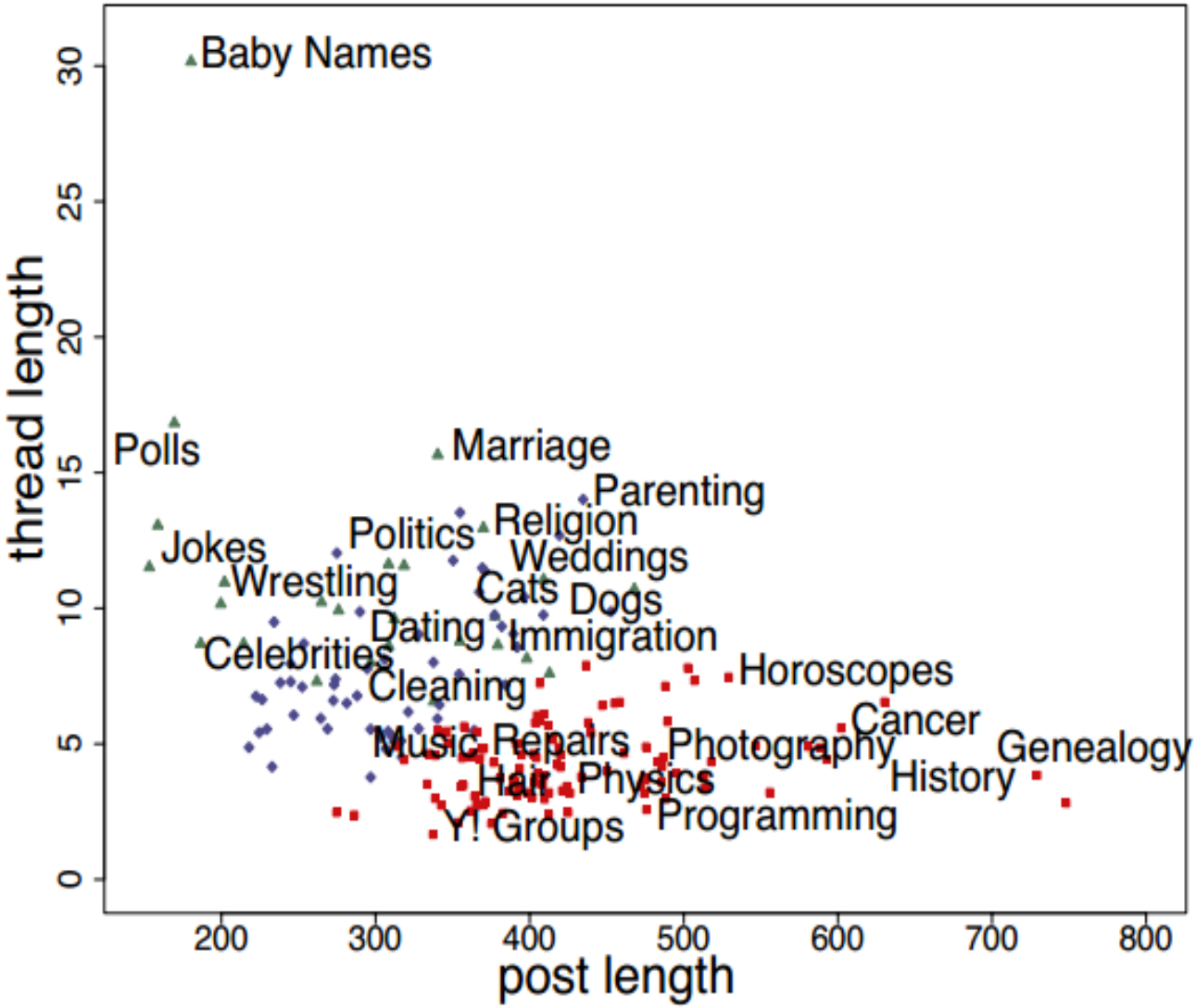
Exercise I

- Use Friendship data and do the following analyses
- 1. Run density within Gender and Club
- 2. Run E-I Index on both attributes and see if there exists homophily by either attribute
- 3. Identify the most outward/Inward - looking group and individuals
- 4. Which gender is more outward looking?

Exercise II

- Apply E-I index analysis on Krack High Tech, and Try both level (column 3) and department (column 4)
  - The procedure only runs on the first spreadsheet
    - Run on report-to network and check the E-I value

Yahoo Answers (YA)



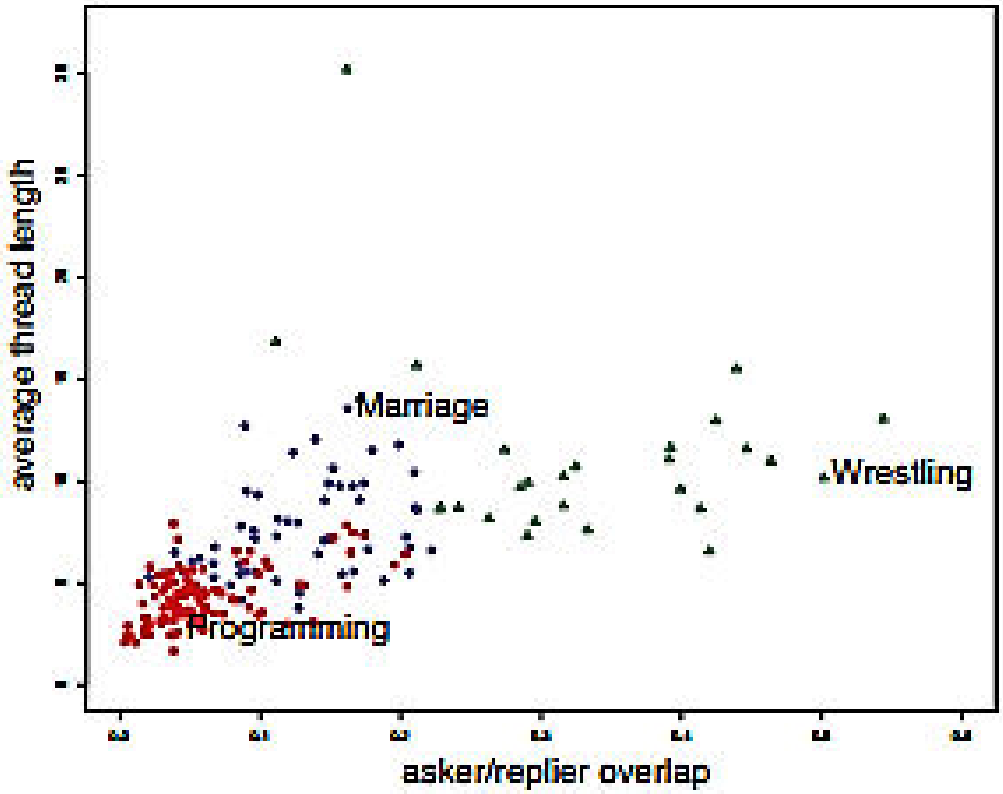


Figure 2: Clustering of categories by thread length and overlap between askers and repliers

Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008, April). Knowledge sharing and yahoo answers: everyone knows something. In*Proceedings of the 17th international conference on World Wide Web* (pp. 665-674). ACM.

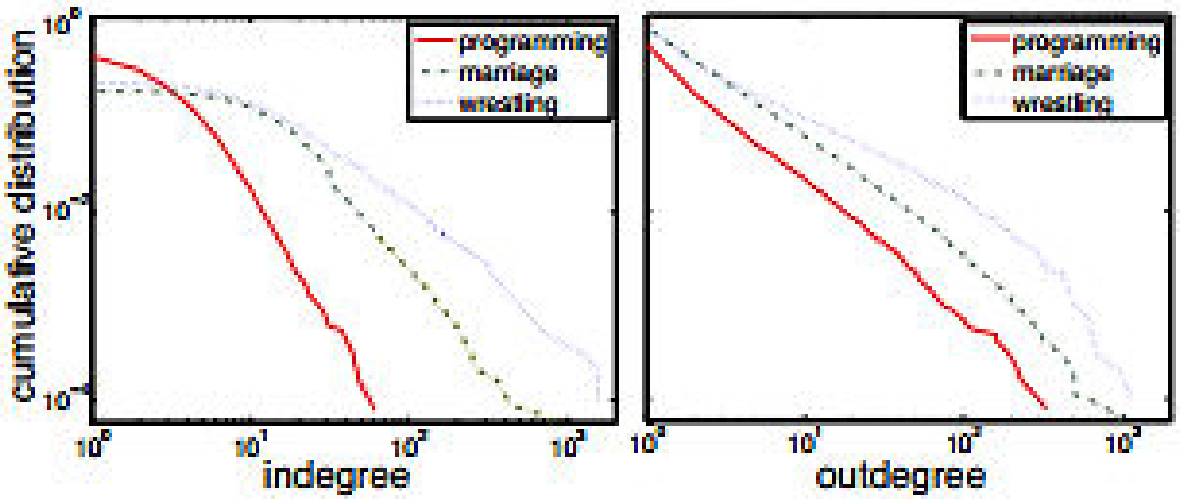


Figure 3: Distributions of indegree (number of users one has received answers from) and outdegree (number of users one has answered)

Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008, April). Knowledge sharing and yahoo answers: everyone knows something. In*Proceedings of the 17th international conference on World Wide Web* (pp. 665-674). ACM.

Table 1: Summary statistics for selected QA networks

Category	Nodes	Edges	Avg. deg.	Mutual edges	SCC
Wrestling	9,959	56,859	7.02	1,898	13.5%
Program.	12,538	18,311	1.48	0	0.01%
Marriage	45,090	164,887	3.37	179	4.73%

Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008, April). Knowledge sharing and yahoo answers: everyone knows something. In *Proceedings of the 17th international conference on World Wide Web* (pp. 665-674). ACM.

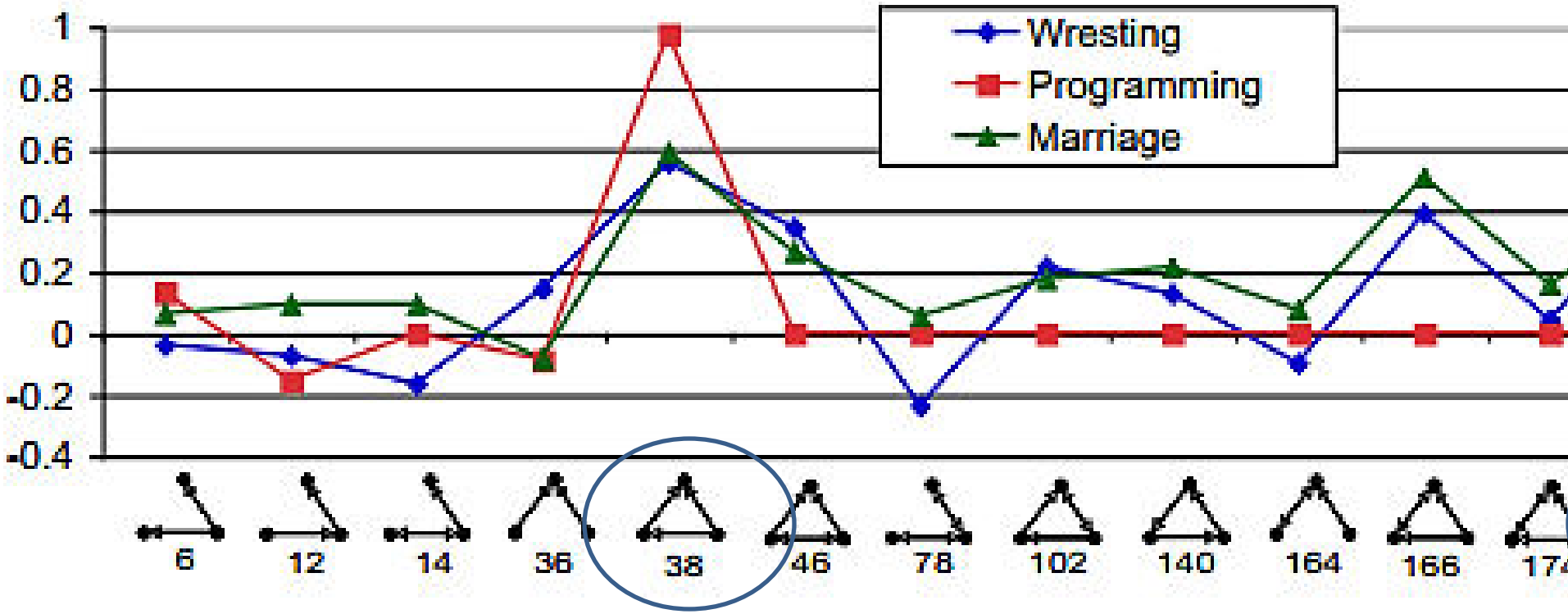


Figure 5: Motif profiles of selected categories

Adamic, L. A., Zhang, J., Bakshy, E., & Ackerman, M. S. (2008, April). Knowledge sharing and yahoo answers: everyone knows something. In*Proceedings of the 17th international conference on World Wide Web* (pp. 665-674). ACM.

SNA

Homophily, faction, and coreness

Grouping

- **Grouping/partition as owning an attribute**
  - **Pre-defined**
  - **Exogenous** (from the Greek words "exo" and "gignomi", meaning "outside" and "to come to be") refers to an action or object coming from outside a system.
- **Grouping/partition as interconnectedness**
  - **Emergent**
  - **Endogenous** substances are those that originate from within

Homophily

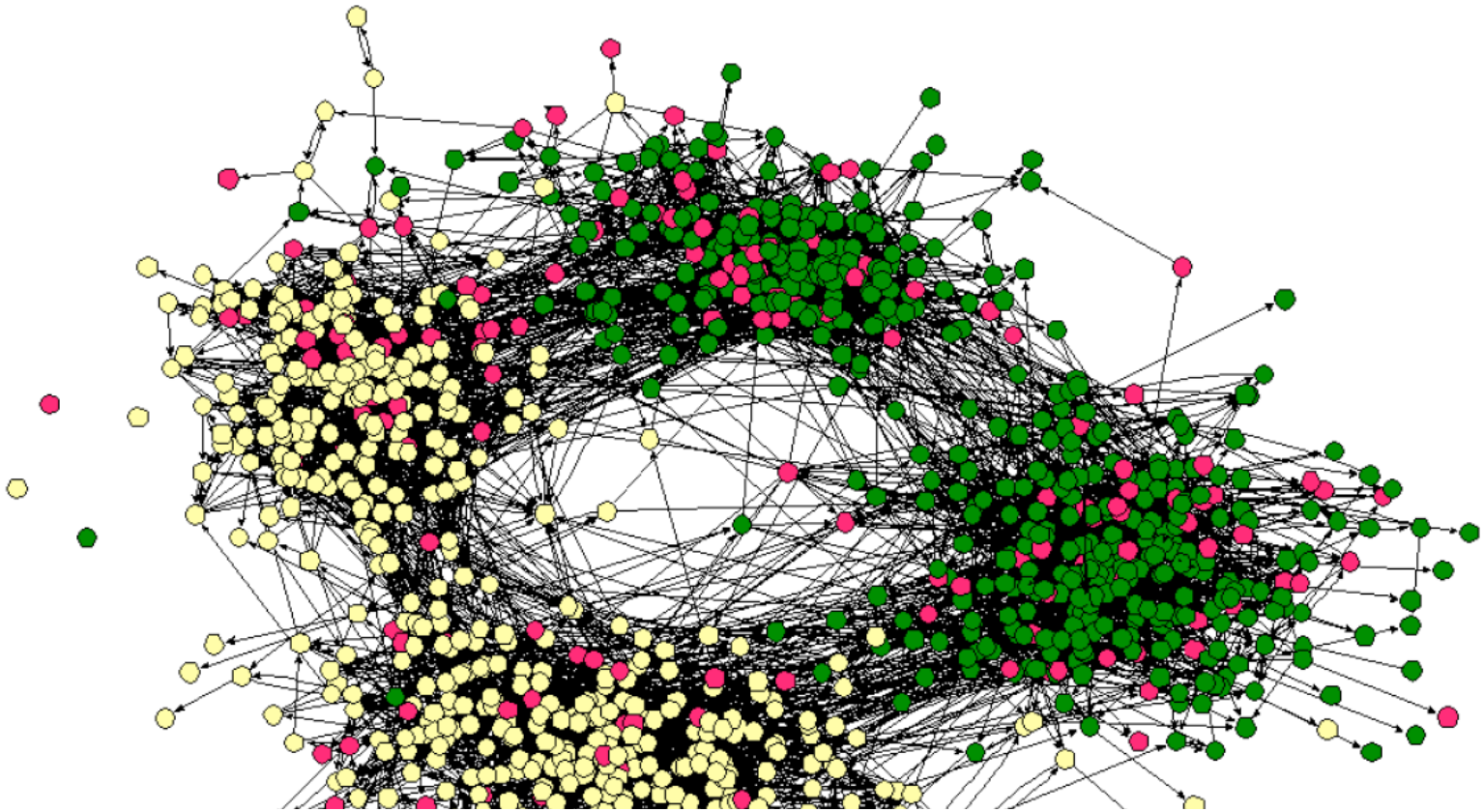
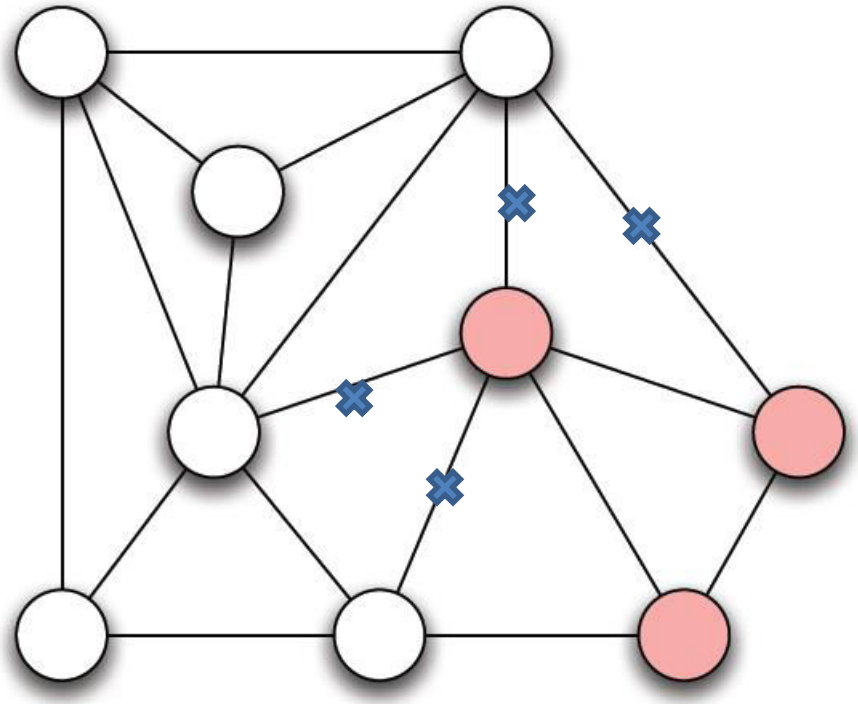


Figure 4.1: Homophily can produce a division of a social network into densely-connected, homogeneous parts that are weakly connected to each other. In this social network from a town's middle school and high school, two such divisions in the network are apparent: one based on race (with students of different races drawn as differently colored circles), and the other based on friendships in the middle and high schools respectively [304].

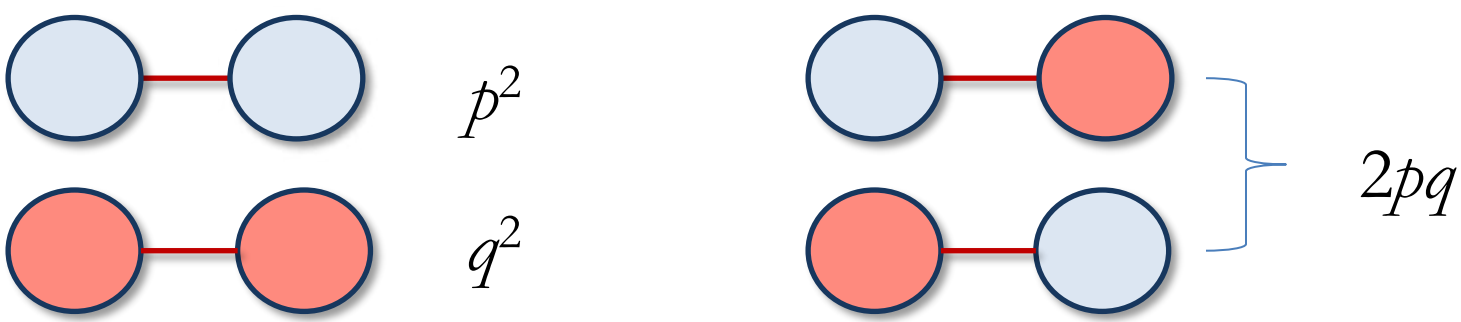
Homophily test

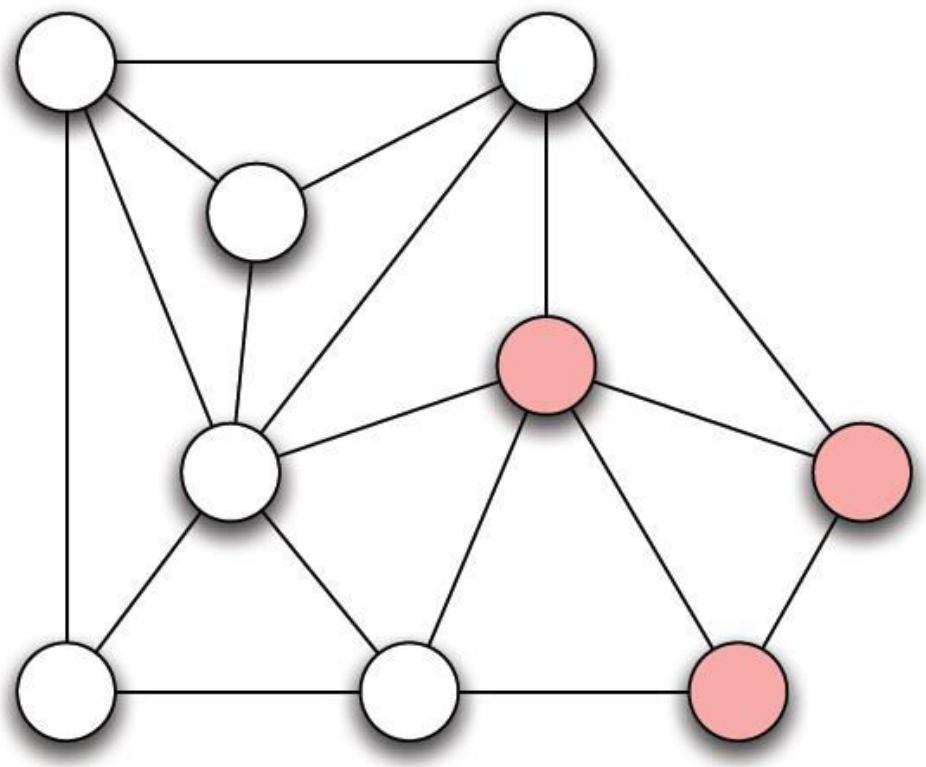


p fraction are male (white node),
q fraction are female (pink node)

Randomly Generated Network

- Probability of both ends of edge being male: p^2
  female: q^2
- male on one end and female on other: 2pq





Homophily Test: If the fraction of heterogeneous (cross-gender) edges is significantly less than $2pq$ then there is evidence for homophily

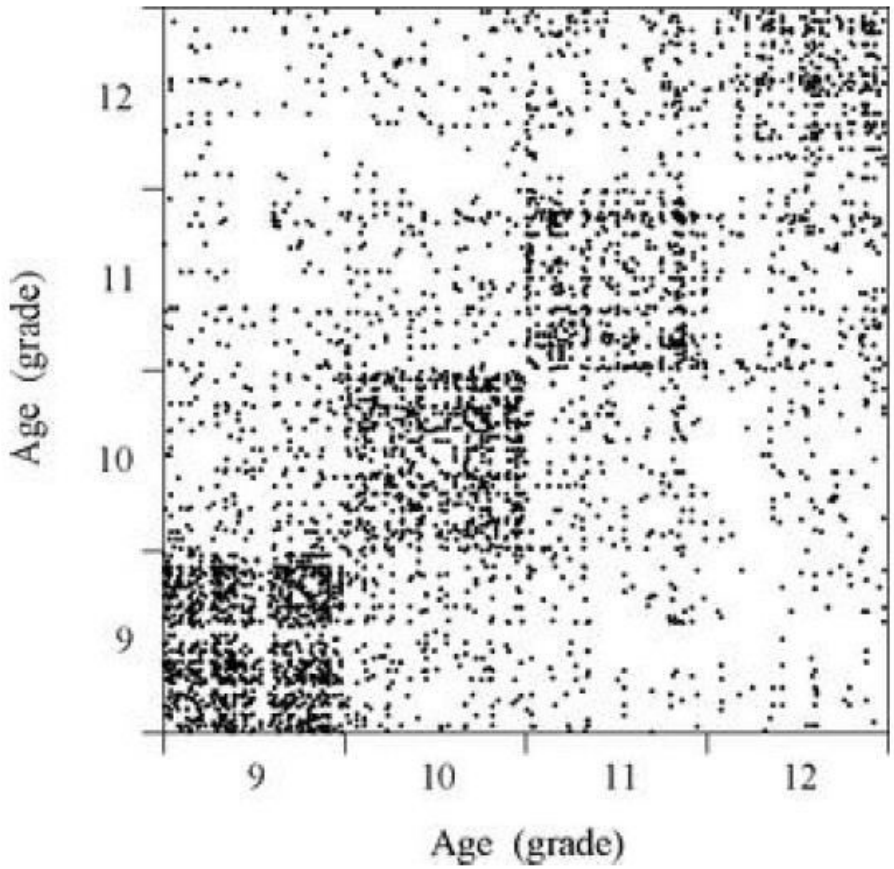
Cross-gender edges:
5 of 18

$p = 6/9 = 2/3$
$q = 3/9 = 1/3$

If no homophily, # of cross-gender edges should be $2pq = 4/9 = 8$ out of 18

$\therefore$ Evidence of homophily

Homophily: correlation between network and attribute variables



Homophily test with UCINET

- **NETWORK>COHESION>HOMOPHILY**

  1. Make a partition of a network into a number of mutually exclusive groups
- 2. The E-I index , which ranges from 1 to -1 and can be seen as a measure of the extent a group chooses themselves a value of -1 showing homophily and a value of +1 showing heterophily.
- 3. For valued data it is the sum of the tie strengths instead of the number of ties.
- 4. Find the correlation between a network and an attribute variable by correlating the dichotomous variable "is in the same group as" with the corresponding entry in the data matrix.

Exercises

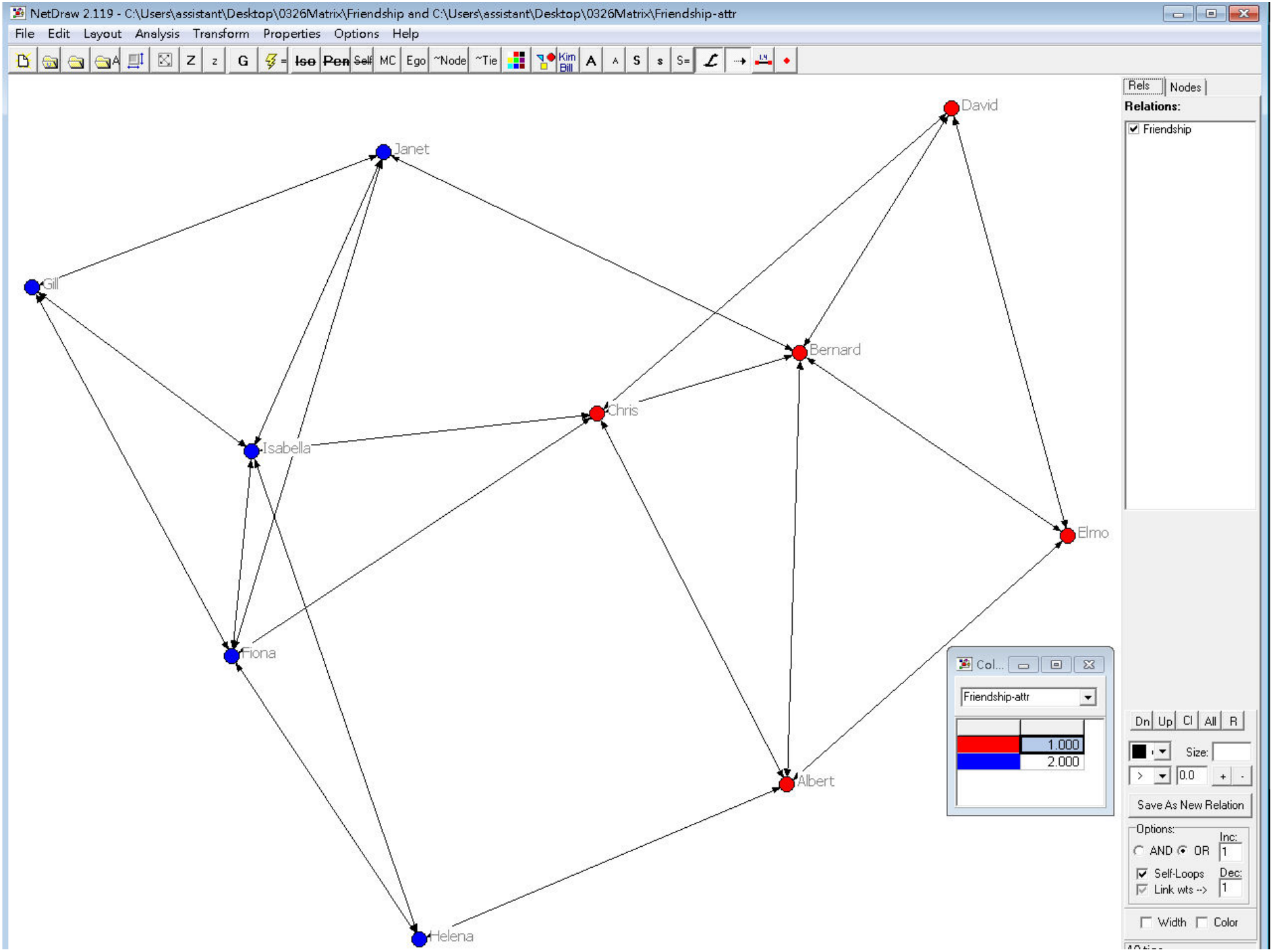
- Use the excel file friendship and friend-attr files to run both

  NETWORK>COHESION>HOMOPHILY

  AND

  Network>Cohesion>E-I Index

  Then compare their results


NetDraw 2.119 - C:\Users\assistant\Desktop\0326Matrix\Friendship and C:\Users\assistant\Desktop\0326Matrix\Friendship-attr
File Edit Layout Analysis Transform Properties Options Help
Iso Pen Self MC Ego ~Node ~Tie Kim Bill
Rels Nodes
Relations:
Friendship
David
Janet
Gill
Bernard
Chris
Isabella
Elmo
Fiona
Albert
Helena
Col...
Friendship-attr
1.000
2.000
Dn Up Cl All R
Size:
0.0
Save As New Relation
Options:
AND OR
Inc:
1
Self-Loops
Link wts -->
Dec:
1
Width Color

NODE PARTITION

Block	Value	Freq	Members:
1	1	5	Albert Bernard Chris David Elmo
2	2	5	Fiona Gill Helena Isabella Janet

No. of ties within and between groups

	1	2
1	16	4
2	4	16

Whole Network Homophily Measures

	1 E-I Ind	2 PBSC
Matrix 1	-0.6000	0.6400

PBSC = point biserial correlation.

```
Density matrix

              1     2
              1     2
           ----- -----
    1  1  0.800 0.160
    2  2  0.160 0.800

40 ties.

Whole Network Results

                      1       2       3       4
                   Freq     Pct Possibl Density
                 ------- ------- ------- -------
    1  Internal   32.000   0.800  40.000   0.800
    2  External    8.000   0.200  50.000   0.160
    3       E-I  -24.000  -0.600  10.000   0.111


Max possible external ties: 50.000
Max possible internal ties: 40.000

E-I Index: -0.600
Expected value for E-I index is: 0.111

Max possible E-I given density & group sizes: 1.000
Min possible E-I given density & group sizes: -1.000

Re-scaled E-I index: -0.600


Permutation Test
Number of iterations = 5000

                      1       2       3       4       5       6       7
                    Obs     Min     Avg     Max      SD P >= Ob P <= Ob
                 ------- ------- ------- ------- ------- ------- -------
    1  Internal    0.800   0.200   0.446   0.800   0.091   0.009   1.000
    2  External    0.200   0.200   0.554   0.800   0.091   1.000   0.009
    3       E-I   -0.600  -0.600   0.108   0.600   0.181   1.000   0.009


E-I Index is significant (p < 0.05)
```

Group level E-I Index

		1	2	3	4
		Intern	Extern	Total	E-I
1	1	16.000	4.000	20.000	-0.600
2	2	16.000	4.000	20.000	-0.600

Individual Level E-I Index

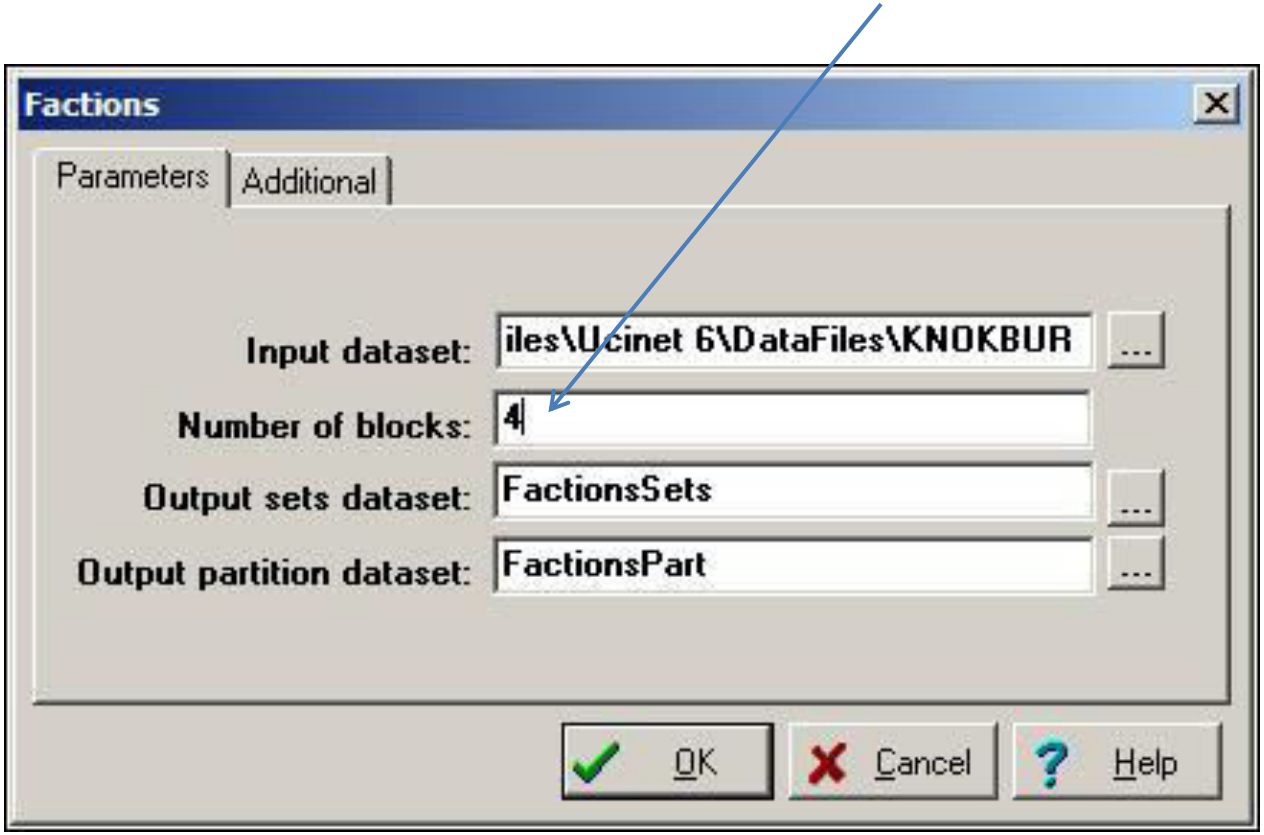
		1	2	3	4
		Intern	Extern	Total	E-I
1	Albert	3.000	1.000	4.000	-0.500
2	Bernard	4.000	1.000	5.000	-0.600
3	Chris	3.000	2.000	5.000	-0.200
4	David	3.000	0.000	3.000	-1.000
5	Elmo	3.000	0.000	3.000	-1.000
6	Fiona	4.000	1.000	5.000	-0.600
7	Gill	3.000	0.000	3.000	-1.000
8	Helena	2.000	1.000	3.000	-0.333
9	Isabella	4.000	1.000	5.000	-0.600
10	Janet	3.000	1.000	4.000	-0.500

Individual E-I values saved as dataset C:\Users\assistant\Desktop\0326Matrix\IndE-I

Factions (Endogenous grouping)

- Factions: a subgroup in which each person is closely tied to all others within the group and no or few connections at all to other sub-groups
- Network>Subgroups>Factions
  - Maximize within-group connections/density, minimize between-group connections
  - Errors (sum of zeros within, and ones between, i.e., noon-diagonal blocks)
  - **Use in an exploratory manner** (i.e. trail and errors)

Need to specify the # of factions



The assignment of actors into 4 groups

Show within and between density

How to count classification "errors"/badness of fit?

(sum of zeros within, and ones between, i.e., noon-diagonal blocks)

```
Number of factions:                          4
Measure of fit:                              Hamming
Input dataset:                               KNOKBUR

Initial proportion correct: 0.567

...Badness of fit: 27.000
...Badness of fit: 27.000
...Badness of fit: 29.000

Final proportion correct: 0.700

Group Assignments:

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

Grouped Adjacency Matrix
```

Density Table

	1	2	3	4
1	0.83	0.67	0.17	0.17
2	0.50		0.00	0.00
3	0.67	0.00		0.50
4	0.42	0.50	0.50	1.00

	1 C	2 C	8 U	4 I	5 M	7 N	9 W	10 W	6 W	3 E
1		1			1	1	1			
2	1		1	1	1	1	1			1
8	1	1		1	1	1	1			
4	1	1			1	1				
5	1	1	1	1		1	1	1		1
7		1		1	1					
9		1			1	1				
10	1	1			1	1				1
6						1	1			1
3		1		1	1	1		1	1	

```
Number of factions:                    3
Measure of fit:                        Hamming
Input dataset:                         KNOKBUR

Initial proportion correct: 0.656

...Badness of fit: 25.000
...Badness of fit: 25.000
...Badness of fit: 25.000

Final proportion correct: 0.722

Group Assignments:

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

Grouped Adjacency Matrix
```

Density Table

	1	2	3
1	1.00	0.50	0.57
2	0.50		0.29
3	0.21	0.00	0.76

	3 E	10 W	6 W	2 C	4 I	1 C	7 N	8 U	9 W	5 M
3		1	1	1	1		1			1
10	1			1		1	1			1
6	1						1		1	
2	1				1	1	1	1	1	1
4				1		1	1			1
1				1			1		1	1
7				1	1					1
8				1	1	1	1		1	1
9				1			1			1
5	1	1		1	1	1	1	1	1	

```
Number of factions:                    4
Measure of fit:                        Hamming
Input dataset:                         KNOKBUR

Initial proportion correct: 0.567

...Badness of fit: 27.000
...Badness of fit: 27.000
...Badness of fit: 29.000

Final proportion correct: 0.700

Group Assignments:

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

Grouped Adjacency Matrix
```

Density Table

	1	2	3	4
1	0.83	0.67	0.17	0.17
2	0.50		0.00	0.00
3	0.67	0.00		0.50
4	0.42	0.50	0.50	1.00

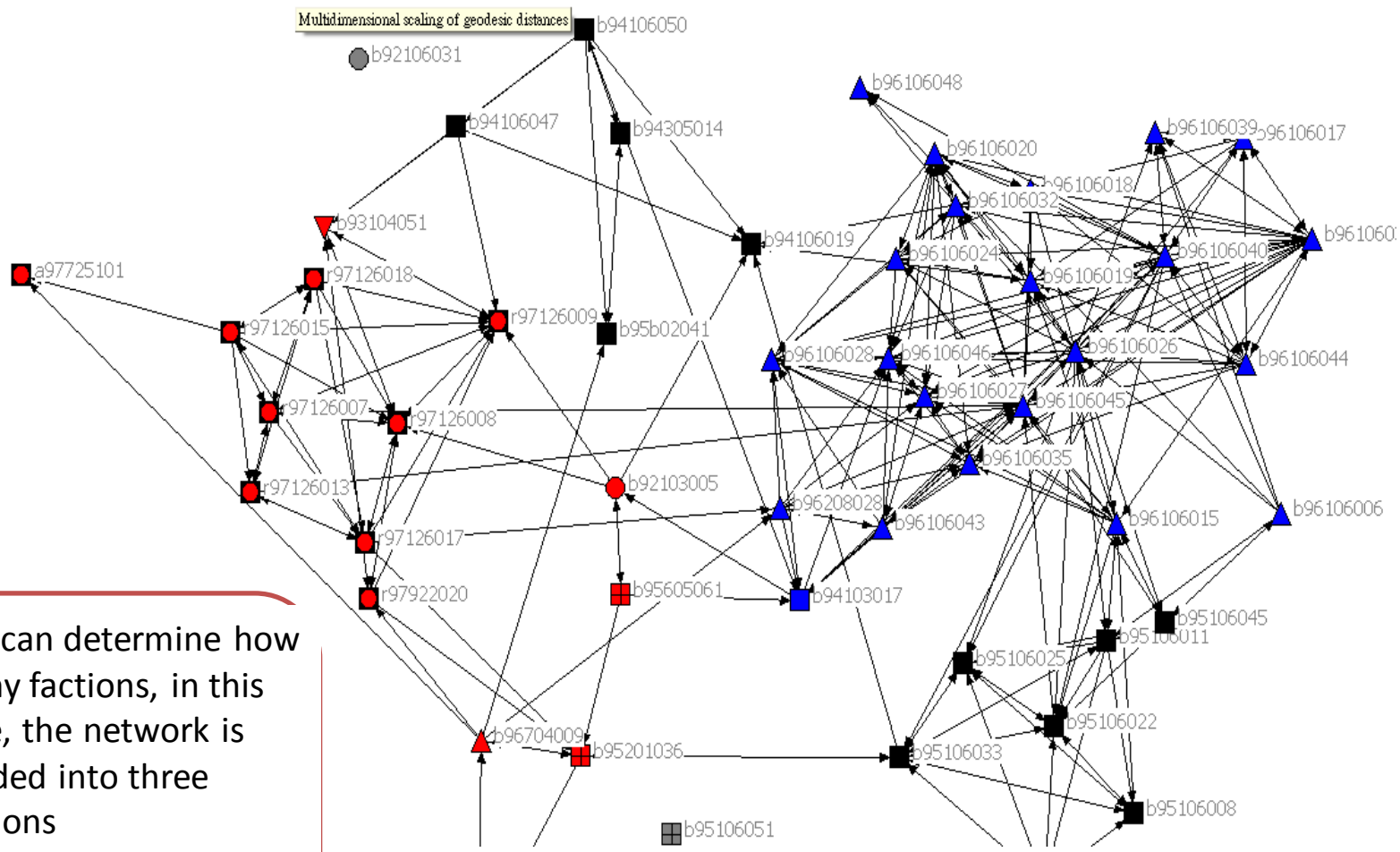
	1 C	2 C	8 U	4 I	5 M	7 N	9 W	10 W	6 W	3 E
1		1			1	1	1			
2	1		1	1	1	1	1			1
8	1	1		1	1	1	1			
4	1	1			1	1				
5	1	1	1	1		1	1	1		1
7		1		1	1					
9		1			1	1				
10	1	1			1	1				1
6						1	1			1
3		1		1	1	1		1	1	

Compare "goodness" of fit

Factions with Netdraw

- *Analysis>Subgroups>Factions* (select number).
- A "faction" is a part of a graph in which the nodes are more tightly connected to one another than they are to members of other "factions."

Subgroups: *Analysis > Subgroups > Factions*



You can determine how many factions, in this case, the network is divided into three factions

Exercises

- Try setting faction number at 2 and see the correction rate
- Try assign different faction number to the NTNU faculty data
- Try visualization of the NTNU data with Netdraw
  - *Analysis>Subgroups>Factions* (select number).

Core/periphery structures

- Classification of community members by connectivity
  - A dense, cohesive core, and
  - A parse, unconnected periphery, only connect to the core, but not to each other

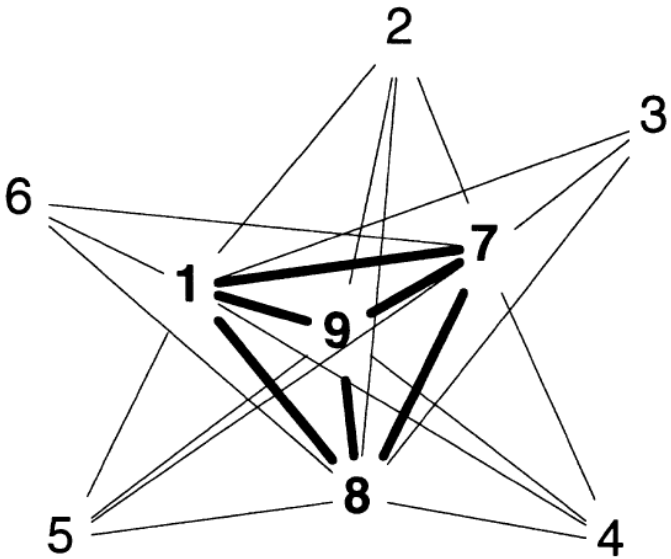
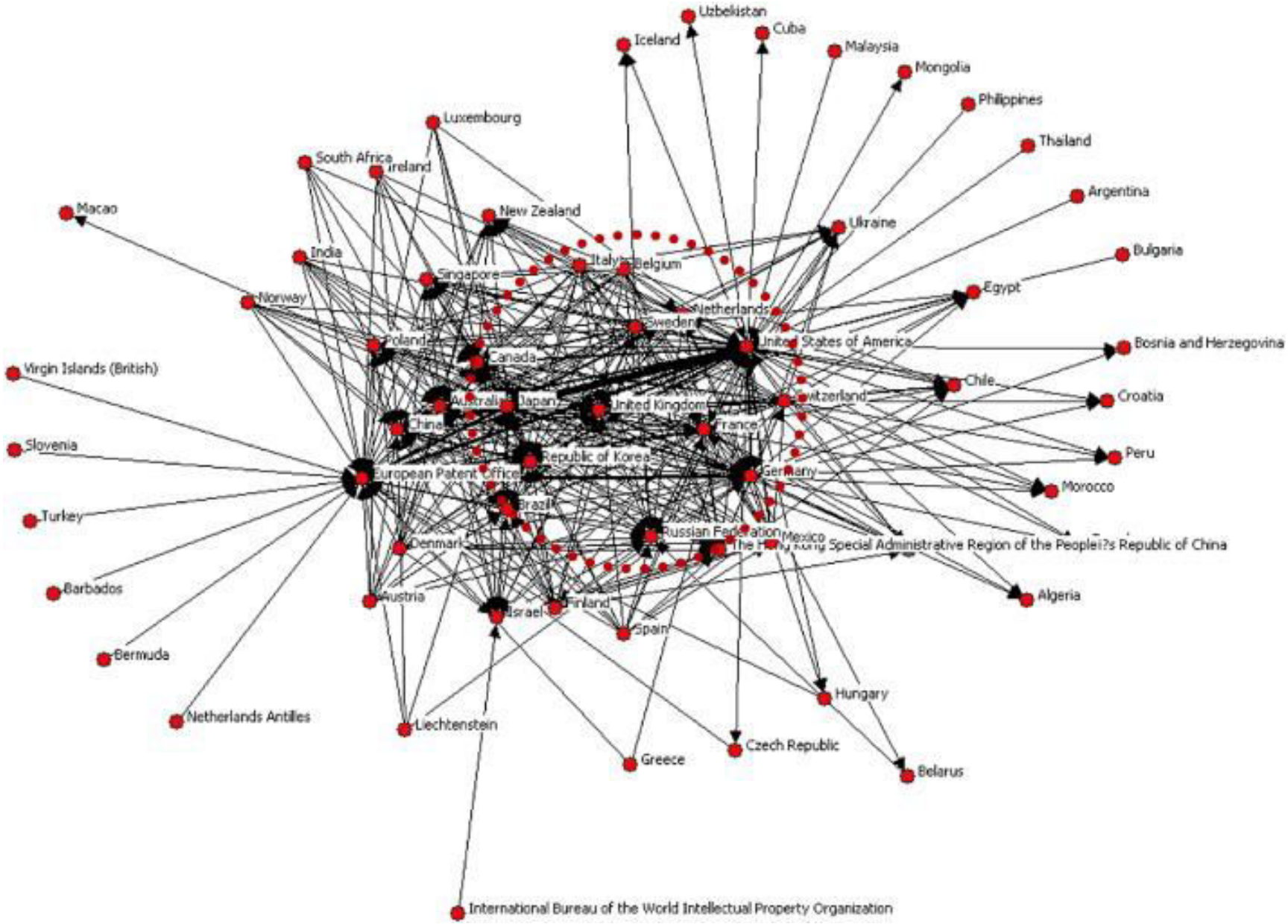


Fig. 3. Core/periphery structure.


Uzbekistan
Cuba
Iceland
Malaysia
Mongolia
Philippines
Thailand
Argentina
Luxembourg
South Africa
Ireland
Macao
New Zealand
Ukraine
India
Italy
Belgium
Bulgaria
Singapore
Egypt
Norway
Netherlands
Sweden
Poland
United States of America
Bosnia and Herzegovina
Canada
Virgin Islands (British)
Chile
Australia
Japan
United Kingdom
Switzerland
Croatia
China
France
Slovenia
Republic of Korea
Peru
European Patent Office
Germany
Morocco
Turkey
Brazil
Russian Federation
Mexico
Denmark
The Hong Kong Special Administrative Region of the People's Republic of China
Barbados
Austria
Algeria
Israel
Finland
Spain
Bermuda
Hungary
Netherlands Antilles
Liechtenstein
Czech Republic
Greece
Belarus
International Bureau of the World Intellectual Property Organization

Core/periphery

- To identify a core-periphery structure, we compare an **observed** block structure to an **ideal** block structure

Idealized core/periphery structure

	1	2	3	4	5	6	7	8	9	10
1		1	1	1	1	1	1	1	1	1
2	1		1	1	1	1	1	1	1	1
3	1	1		1	1	1	1	1	1	1
4	1	1	1		1	1	1	1	1	1
5	1	1	1	1		0	0	0	0	0
6	1	1	1	1	0		0	0	0	0
7	1	1	1	1	0	0		0	0	0
8	1	1	1	1	0	0	0		0	0
9	1	1	1	1	0	0	0	0		0
10	1	1	1	1	0	0	0	0	0	

Compare an observed block structure to an ideal block structure in order to identify the core/periphery structure

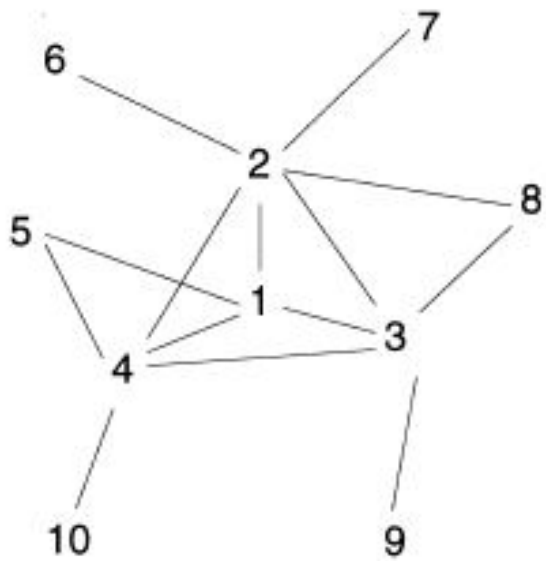


Fig. 1. A network with a core/periphery structure.

	1	2	3	4	5	6	7	8	9	10
1		1	1	1	1	0	0	0	0	0
2	1		1	1	0	1	1	1	0	0
3	1	1		1	0	0	0	1	1	0
4	1	1	1		1	0	0	0	0	1
5	1	0	0	1		0	0	0	0	0
6	0	1	0	0	0		0	0	0	0
7	0	1	0	0	0	0		0	0	0
8	0	1	1	0	0	0	0		0	0
9	0	0	1	0	0	0	0	0		0
10	0	0	0	1	0	0	0	0	0	

(observed blocked network)

(Ideal CP blocked network)

	1	2	3	4	5	6	7	8	9	10
1		1	1	1	1	1	1	1	1	1
2	1		1	1	1	1	1	1	1	1
3	1	1		1	1	1	1	1	1	1
4	1	1	1		1	1	1	1	1	1
5	1	1	1	1		0	0	0	0	0
6	1	1	1	1	0		0	0	0	0
7	1	1	1	1	0	0		0	0	0
8	1	1	1	1	0	0	0		0	0
9	1	1	1	1	0	0	0	0		0
10	1	1	1	1	0	0	0	0	0	

(observed blocked network)

	1	2	3	4	5	6	7	8	9	10
1		1	1	1	1	0	0	0	0	0
2	1		1	1	0	1	1	1	0	0
3	1	1		1	0	0	0	1	1	0
4	1	1	1		1	0	0	0	0	1
5	1	0	0	1		0	0	0	0	0
6	0	1	0	0	0		0	0	0	0
7	0	1	0	0	0	0		0	0	0
8	0	1	1	0	0	0	0		0	0
9	0	0	1	0	0	0	0	0		0
10	0	0	0	1	0	0	0	0	0	

A core periphery structure exists to the extent that the correlation between the ideal structure and the observed structure is high. We can search for cores by simply proposing a partition (many times) and then selecting the best fitting partition.

Core/periphery structures

- When a network cannot be subdivide into exclusive cohesive subgroups (such as components, cliques etc.)
- The assumption:
  - The network consists of just one group to which all actors belong to a greater or lesser extent

Core/periphery with UCINET

- **NETWORK > CORE/PERIPHERY > CATEGORICAL**

  PURPOSE Uses a genetic algorithm to fit a core/periphery model to the data.

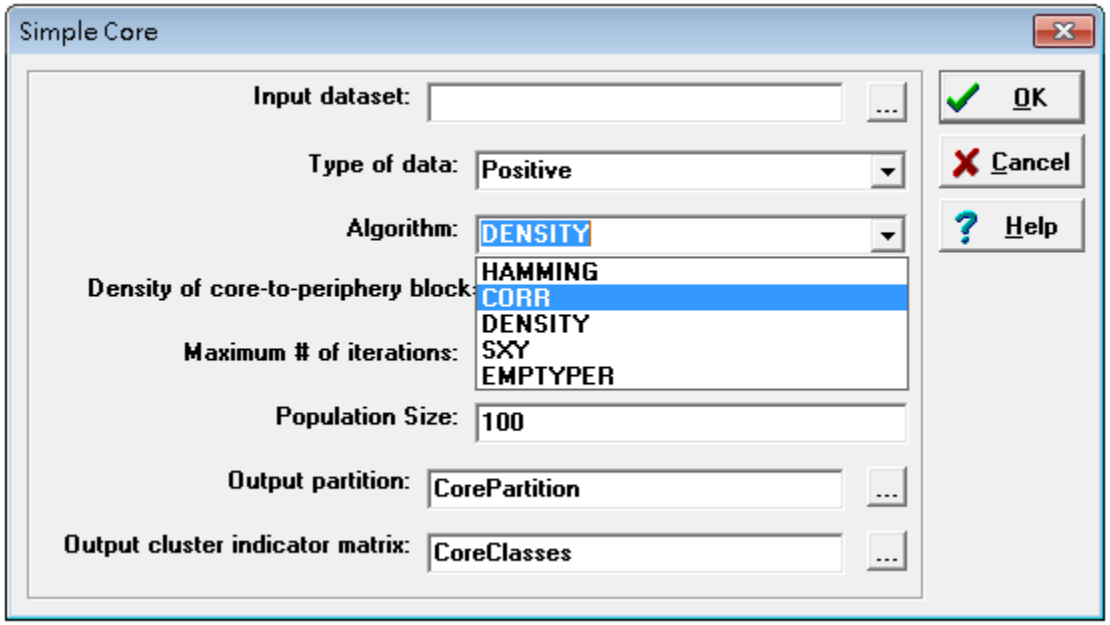
  DESCRIPTION Simultaneously fits a core/periphery model to the data network, and identifies which actors belong in the core and which belong in the periphery.

Alternative core/periphery model

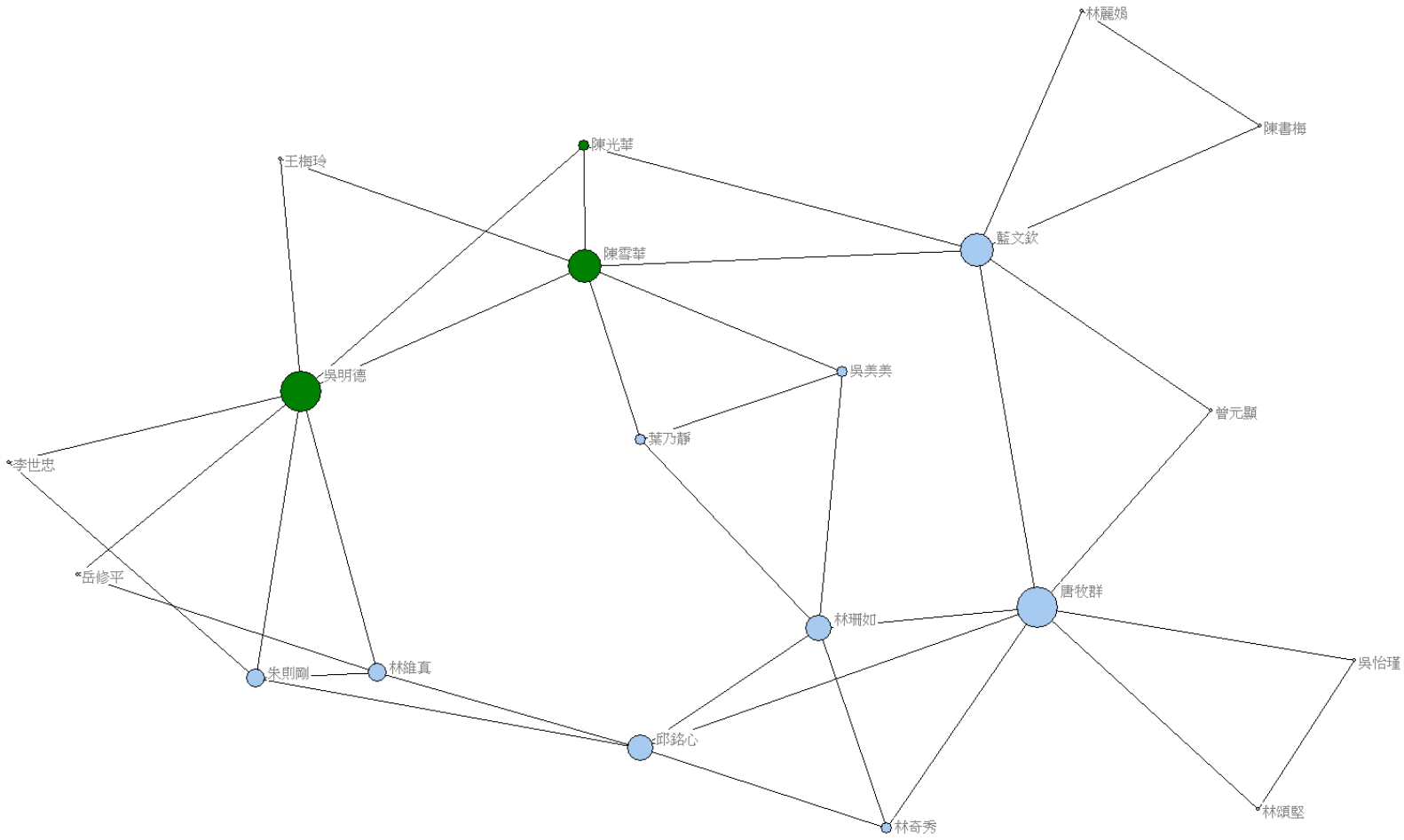
Apply core/periphery in bibliometrics

		6 C	7 C	13 J	16 S	15 S	20 S	17 S	4 C	9 F	8 C	1 A	12 A	3 B	14 P	5 C	11 J	12 J	18 S	19 S	10 I
6	CW	1	1	1	1	1	1	1	1						1	1					
7	CYSR	1	1		1	1	1	1	1	1					1	1					
13	JSWE	1		1	1	1	1	1					1	1					1	1	
16	SSR	1	1	1	1	1	1	1			1		1					1	1		
15	SCW	1	1	1	1	1	1	1	1	1	1		1			1	1		1	1	
20	SWRA	1	1	1	1	1	1	1					1								
17	SW	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1		1	1	1
4	CAN	1	1			1		1	1						1						
9	FR		1			1		1		1											
8	CSWJ				1	1		1			1										
1	AMH							1				1									
2	ASW			1	1	1	1	1					1		1						
3	BJSW			1				1						1					1		
14	PW	1	1					1	1				1		1						
5	CCQ	1	1			1										1					
11	JGSW					1		1									1				
12	JSP				1													1			
18	SWG			1	1	1		1						1					1	1	
19	SWHC			1		1		1											1	1	
10	IJSW							1													1

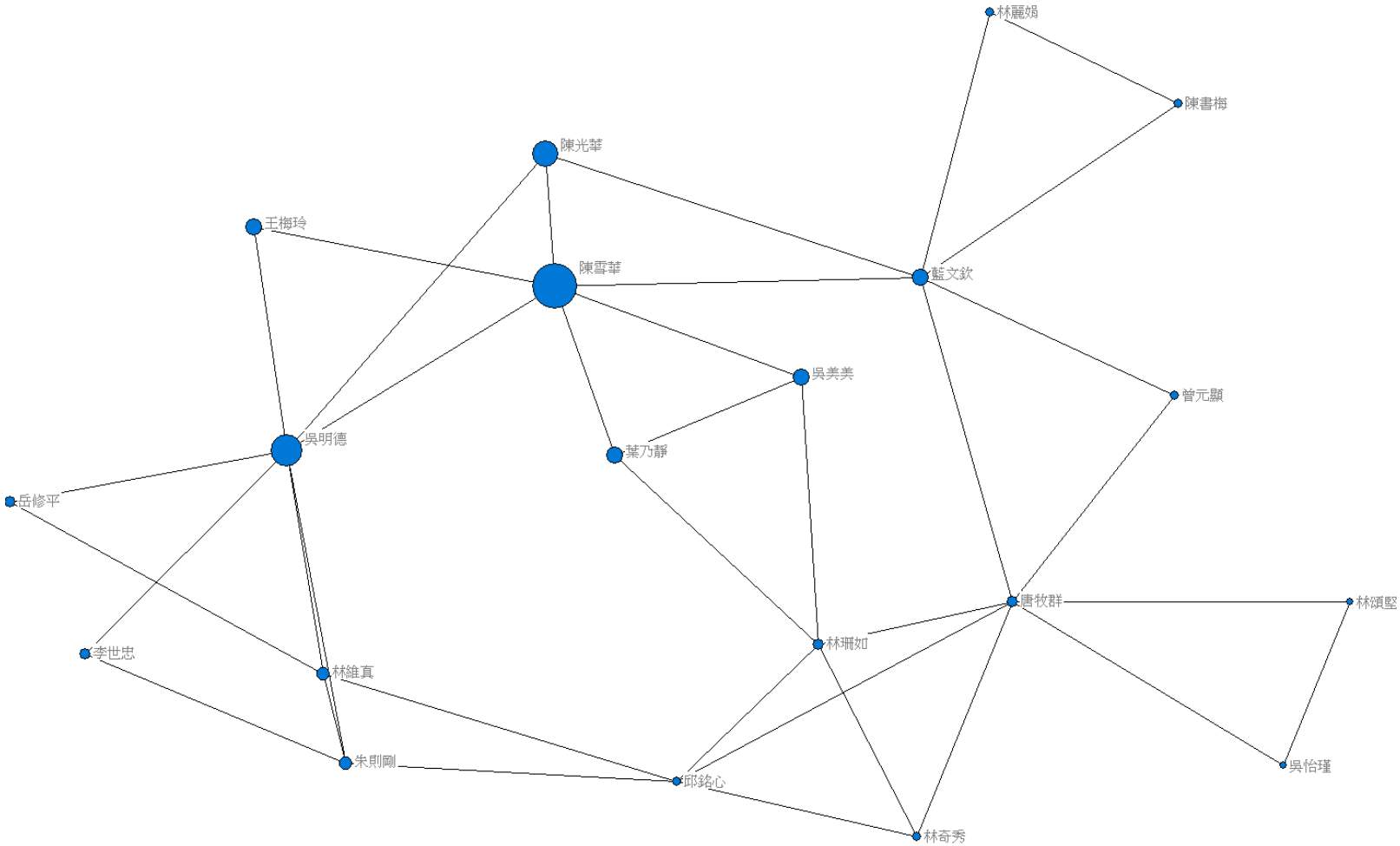
Core/periphery with categorical data



Try ntu_TwomodRows ##d



Performed with categorical core-periphery procedure
Size = degree
Color = core



Performed with continuous core-periphery procedure
Size = Coreness
More credits to the library director

Make sure you know the location of these two output files

Coreness Partition

Coreness Classes

UCINET Spreadsheet - C:\Users\assistant\Documents\UCINET da

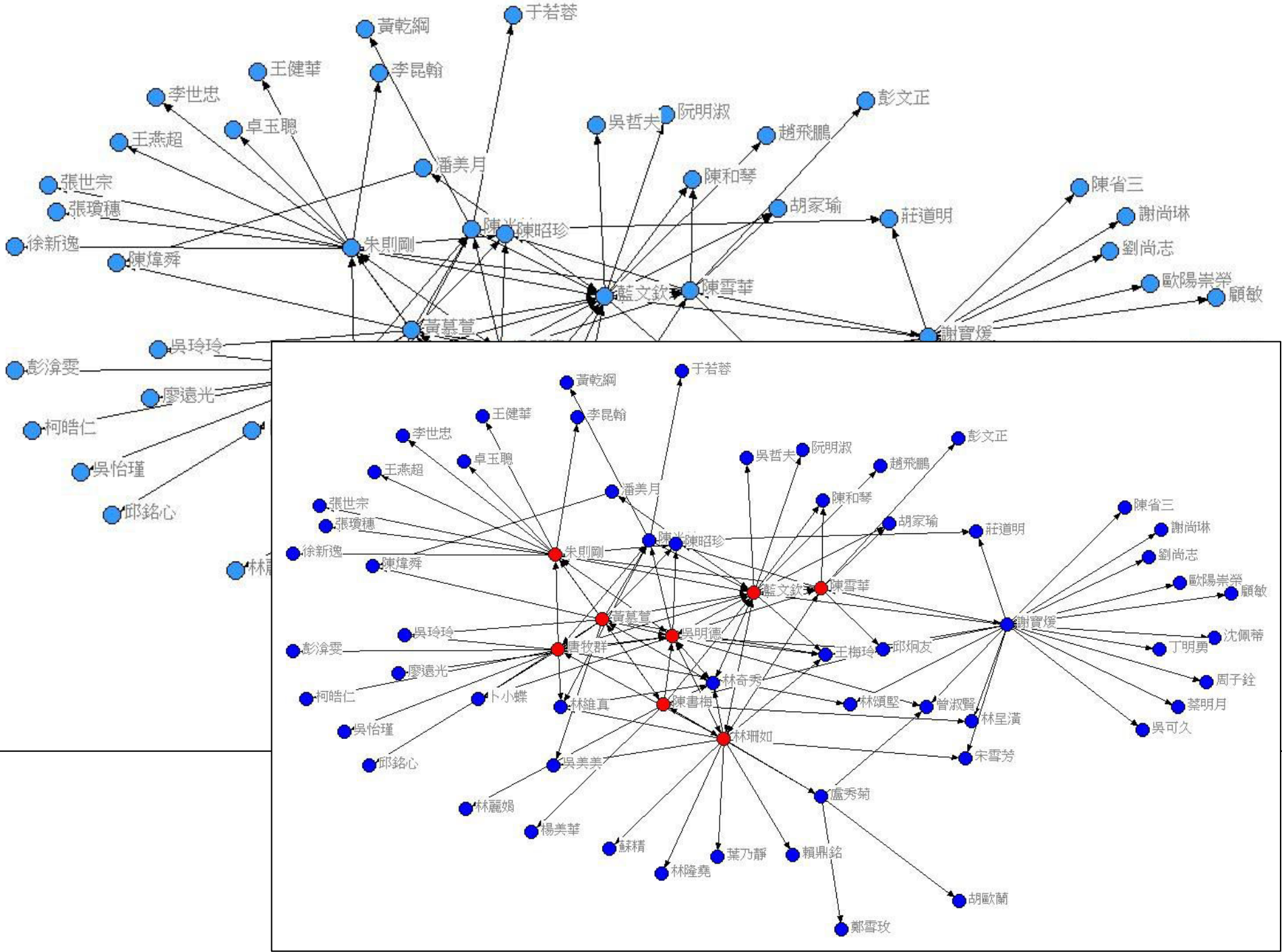
File Edit Transform Fill Labels Options Help

	PARTITION
黃慕萱	1
朱則剛	1
吳明德	1
林奇秀	2
藍文欽	1
林維真	2
陳雪華	1
謝寶煖	2
唐牧群	1
林珊如	1
陳光華	2
陳書梅	1
潘美月	2
盧秀菊	2
吳可久	2
鄭雪玫	2
卜小蝶	2
吳美美	2
陳昭珍	2
王梅玲	2
宋雪芳	2
林頌堅	2
莊道明	2
林呈潢	2
黃乾綱	2
吳玲玲	2
邱銘心	2

UCINET Spreadsheet - C:\Users\assistant\Documents\UCINET data\CoreClasses.##h

File Edit Transform Fill Labels Options Help

	黃慕萱	朱則剛	吳明德	林奇秀	藍文欽	林維真	陳雪華	謝寶煖	唐牧群	林珊如	陳光華
1	1	1	1	0	1	0	1	0	1	1	0
2	0	0	0	1	0	1	0	1	0	0	1


于若蓉
黃乾綱
王健華
李昆翰
李世忠
卓玉聰
王燕超
張世宗
張瓊穗
徐新逸
陳煒舜
朱則剛
潘美月
陳昭珍
吳哲夫
阮明淑
趙飛鵬
彭文正
陳和琴
胡家瑜
莊道明
陳省三
謝尚琳
劉尚志
歐陽崇榮
顧敏
藍文欽
陳雪華
黃慕萱
謝寶煖
吳玲玲
彭渰雯
廖遠光
柯皓仁
吳怡瑾
邱銘心
林
唐牧群
吳明德
王梅玲
邱炯友
沈佩蒂
丁明勇
周子銓
蔡明月
吳可久
林奇秀
林雄真
陳書梅
林頌堅
曾淑賢
林呈潢
林珊如
宋雪芳
卜小蝶
吳美美
盧秀菊
林麗娟
楊美華
蘇精
葉乃靜
賴鼎銘
林隆堯
胡歐蘭
鄭雪玫

Categorical

ucinetlog4.txt - 記事本

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)

Blocked Adjacency Matrix

	1 黃慕萱	2 朱則剛	3 吳明德	12 陳書梅	5 藍文欽	10 林珊如	7 陳雪華	9 唐牧群	8 謝寶煖	6 林維真	11 陳光華	4 林奇秀	13 潘美月	14 盧秀菊	15 吳可久	16 鄭雪玫	17	18	19	20	21	22	23	24
1 黃慕萱		1.000	8.000	4.000	2.000		4.000		1.000	2.000			3.000	1.000	1.000	2.000								
2 朱則剛			9.000	1.000								1.000												
3 吳明德	7.000	4.000		1.000	2.000	4.000	2.000		1.000	2.000			1.000		4.000	3.000								
12 陳書梅	7.000	1.000	3.000		7.000	2.000			2.000		2.000													1.000
5 藍文欽			2.000		1.000	2.000	1.000		1.000	1.000						1.000								
10 林珊如			3.000	1.000	5.000	2.000	5.000	3.000		1.000		1.000		2.000		9.000						2.000	4.000	
7 陳雪華		1.000		2.000	3.000			1.000							1.000									
9 唐牧群		1.000	1.000		3.000	1.000			1.000	3.000				1.000								2.000		
8 謝寶煖				1.000		1.000							1.000					1.000	3.000		1.000	10.000	8.000	
6 林維真																								
11 陳光華				2.000			4.000			1.000											1.000		1.000	
4 林奇秀	1.000		1.000		2.000						1.000													
13 潘美月				1.000																				
14 盧秀菊													1.000											
15 吳可久																								
16 鄭雪玫																								

Coreness

- A continuous version of "coreness" can be had by generalizing the ideal image seen above. Instead of just 0/1, pairs of "high core" nodes have a very strong tie connecting them, and core-periphery nodes have a very low score.
- **Nodes with high coreness are more likely to be at the center of a core-periphery structure**.

Coreness

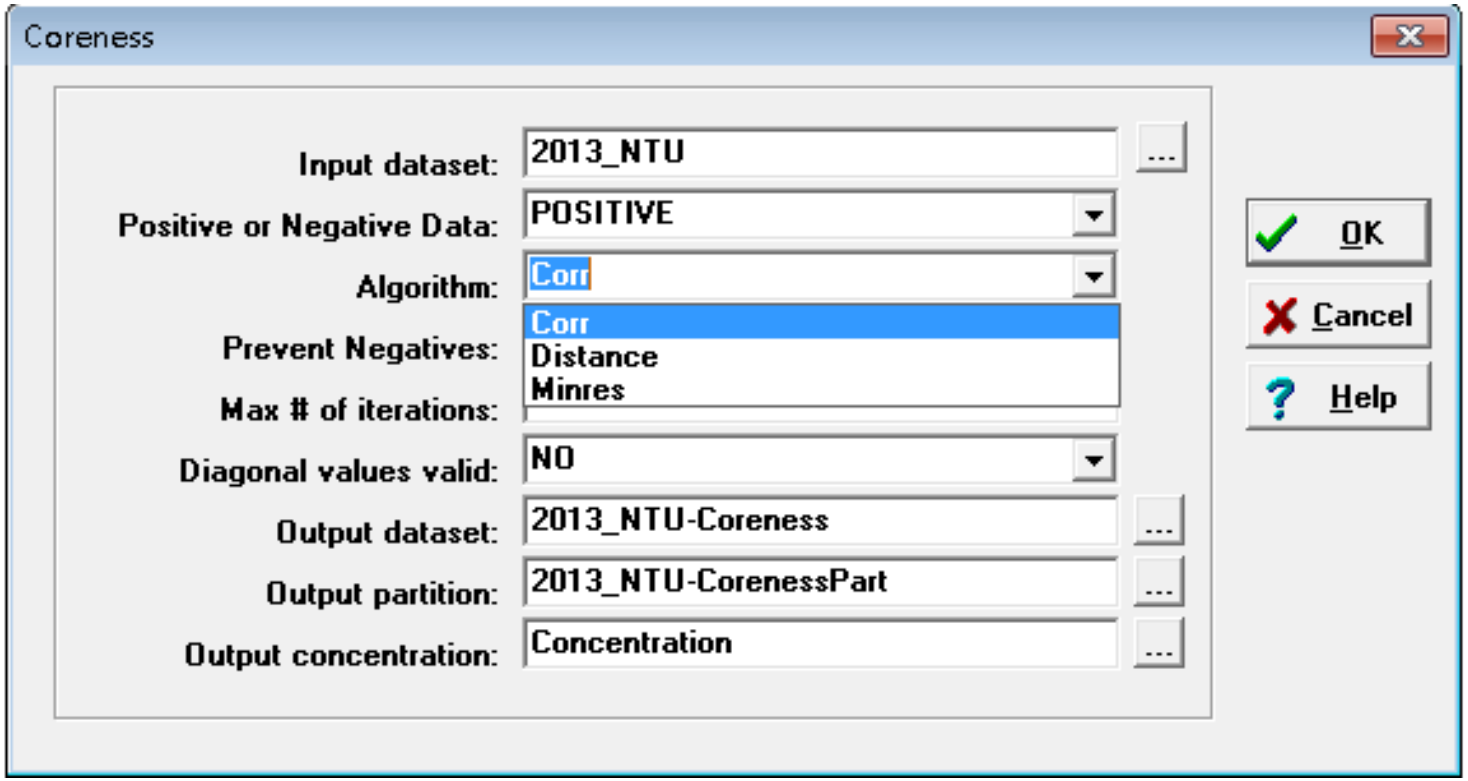
- Coreness can thus be defined as a type of centrality, **but one that assumes a particular underlying structure to the network**.
- Coreness is essentially Eigenvector centrality, and **UCINET sorts nodes by eigenvector centrality and build the "core" until the correlation between ideal/observed drops.**

Again, correlation between network and attribute variables

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
		M	M	M	M	M	F	F	F	F	F	F	F	F	F	F	F	F	F	F	F
1	M		2	10	4	5	5	9	7	4	3	3	7	3	2	5	1	4	1	0	1
2	M	2		5	1	3	1	4	2	6	2	5	4	3	2	2	6	3	1	1	1
3	M	10	5		8	9	5	11	7	8	8	14	17	9	11	11	5	9	4	6	5
4	M	4	1	8		4	0	3	4	2	3	5	3	11	4	7	0	4	3	3	0
5	M	5	3	9	4		3	5	7	4	3	5	6	3	4	4	1	2	1	3	3
6	F	5	1	5	0	3		5	2	3	2	2	4	4	3	1	1	2	0	1	2
7	F	9	4	11	3	5	5		5	4	6	3	9	5	5	4	2	6	3	2	2
8	F	7	2	7	4	7	2	5		3	0	3	4	2	1	3	0	1	1	1	0
9	F	4	6	8	2	4	3	4	3		1	3	2	4	5	4	3	4	1	3	2
10	F	3	2	8	3	3	2	6	0	1		4	7	5	5	7	2	2	3	3	2
11	F	3	5	14	5	5	2	3	3	3	4		9	3	4	4	2	4	2	3	1
12	F	7	4	17	3	6	4	9	4	2	7	9		7	7	8	3	7	2	4	3
13	F	3	3	9	11	3	4	5	2	4	5	3	7		8	11	3	8	2	5	3
14	F	2	2	11	4	4	3	5	1	5	5	4	7	8		8	1	5	4	4	1
15	F	5	2	11	7	4	1	4	3	4	7	4	8	11	8		2	5	2	2	1
16	F	1	6	5	0	1	1	2	0	3	2	2	3	3	1	2		6	1	0	1
17	F	4	3	9	4	2	2	6	1	4	2	4	7	8	5	5	6		4	3	3
18	F	1	1	4	3	1	0	3	1	1	3	2	2	2	4	2	1	4		2	1
19	F	0	1	6	3	3	1	2	1	3	3	3	4	5	4	2	0	3	2		6
20	F	1	1	5	0	3	2	2	0	2	2	1	3	3	1	1	1	3	1	6	

Interactions among a troop of monkeys

Network> core/periphery > continuous

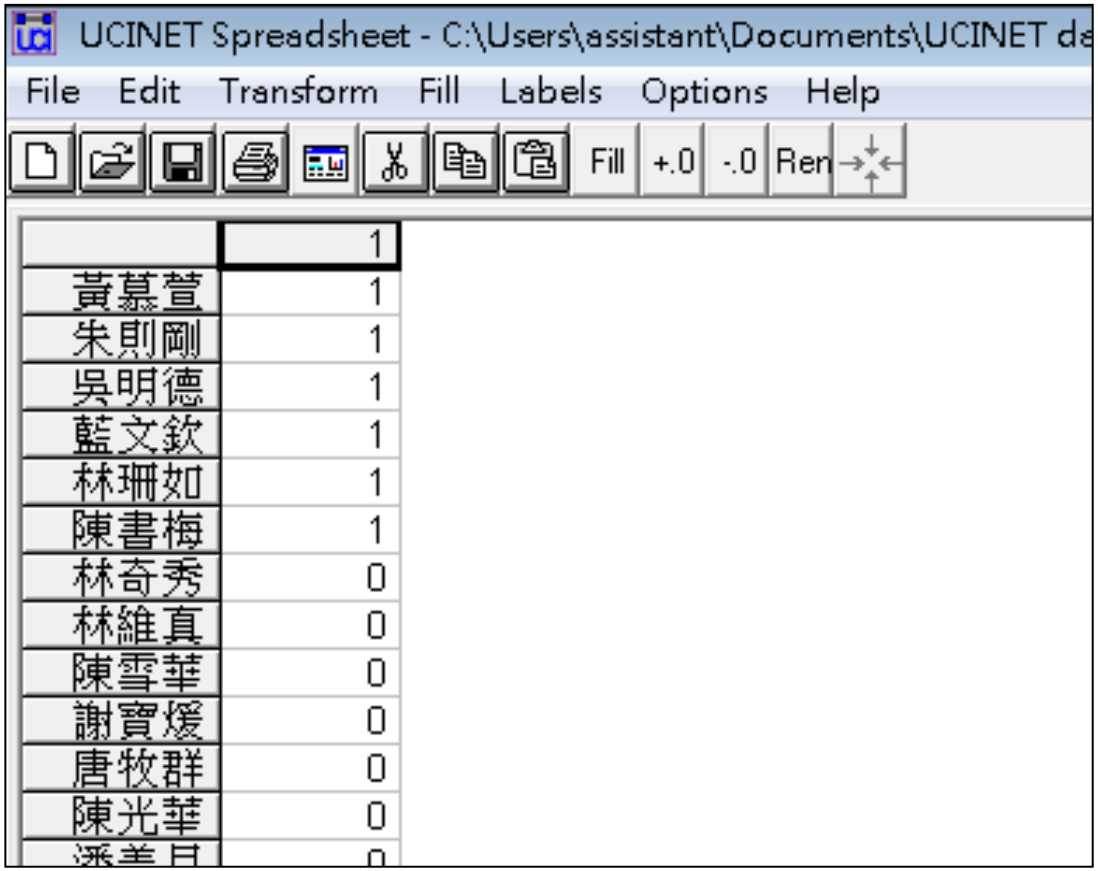


Coreness exercise

Output：

檔名-Coreness、檔名-CorenessPart

檔名-CorenessPart



	1
黃慕萱	1
朱則剛	1
吳明德	1
藍文欽	1
林珊如	1
陳書梅	1
林奇秀	0
林維真	0
陳雪華	0
謝寶煖	0
唐牧群	0
陳光華	0
潘美月	0

檔名-Coreness

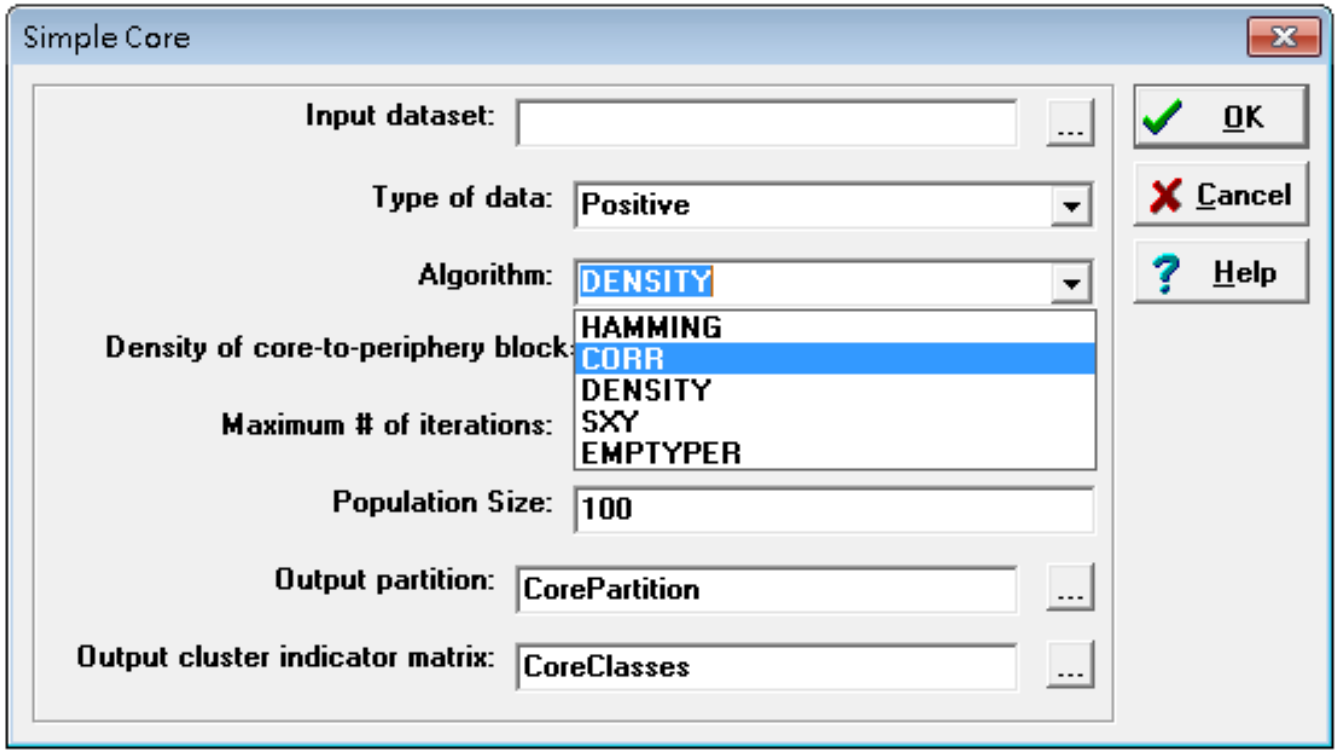
UCINET Spreadsheet - C:\Users\assistant\Documents\UCINET data\2

File Edit Transform Fill Labels Options Help

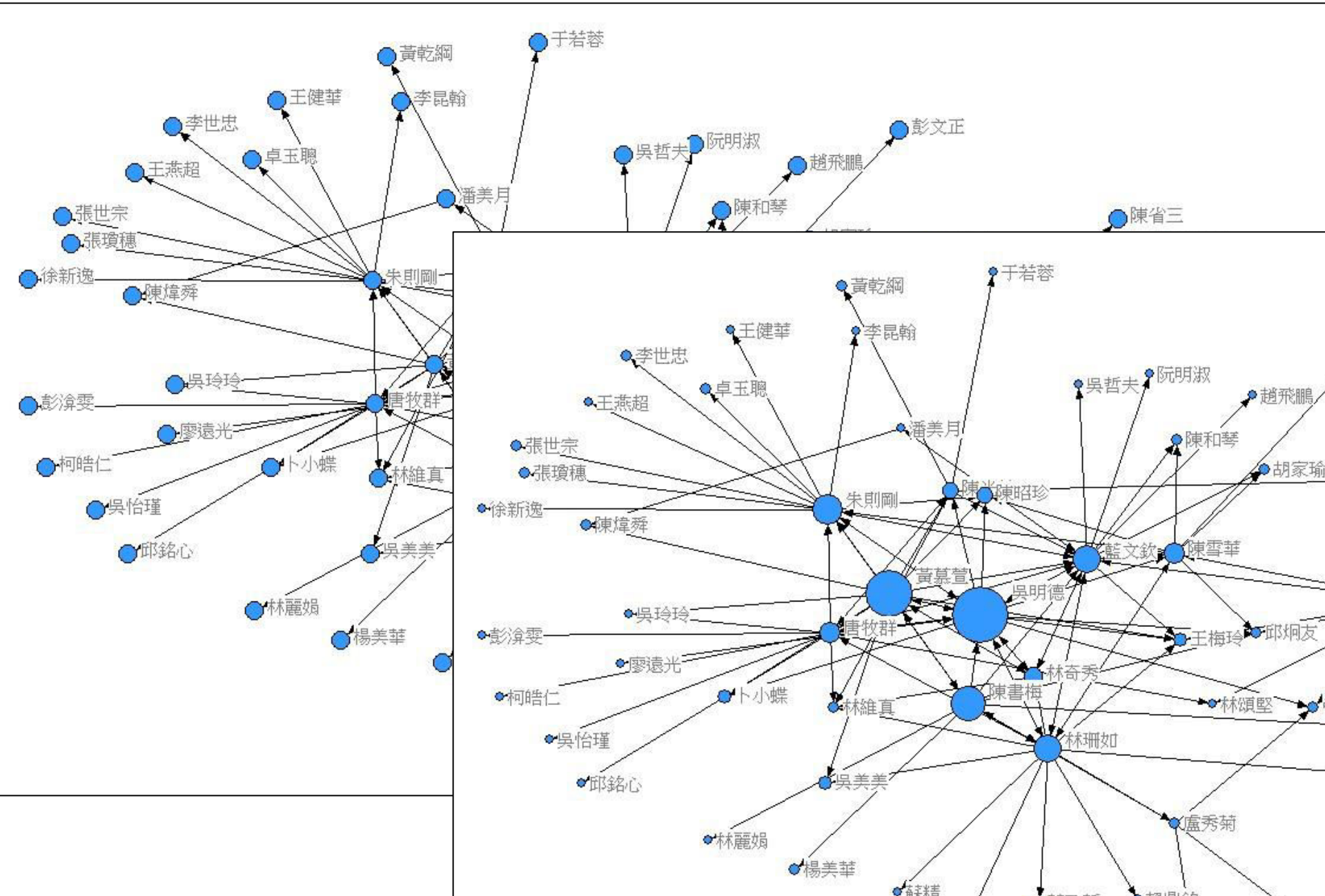
	Coreness
吳明德	0.60771095752716 1
黃慕萱	0.48911952972412 1
陳書梅	0.32168805599212 6
朱則剛	0.26822048425674 4
林珊如	0.25574007630348 2
藍文欽	0.23022437095642 1
唐牧群	0.16308997571468 4
陳雪華	0.14237113296985 6
林奇秀	0.11541780829429 6
陳光華	0.09423986077308 65

Use this file as attribute

Network> core/periphery > categorical

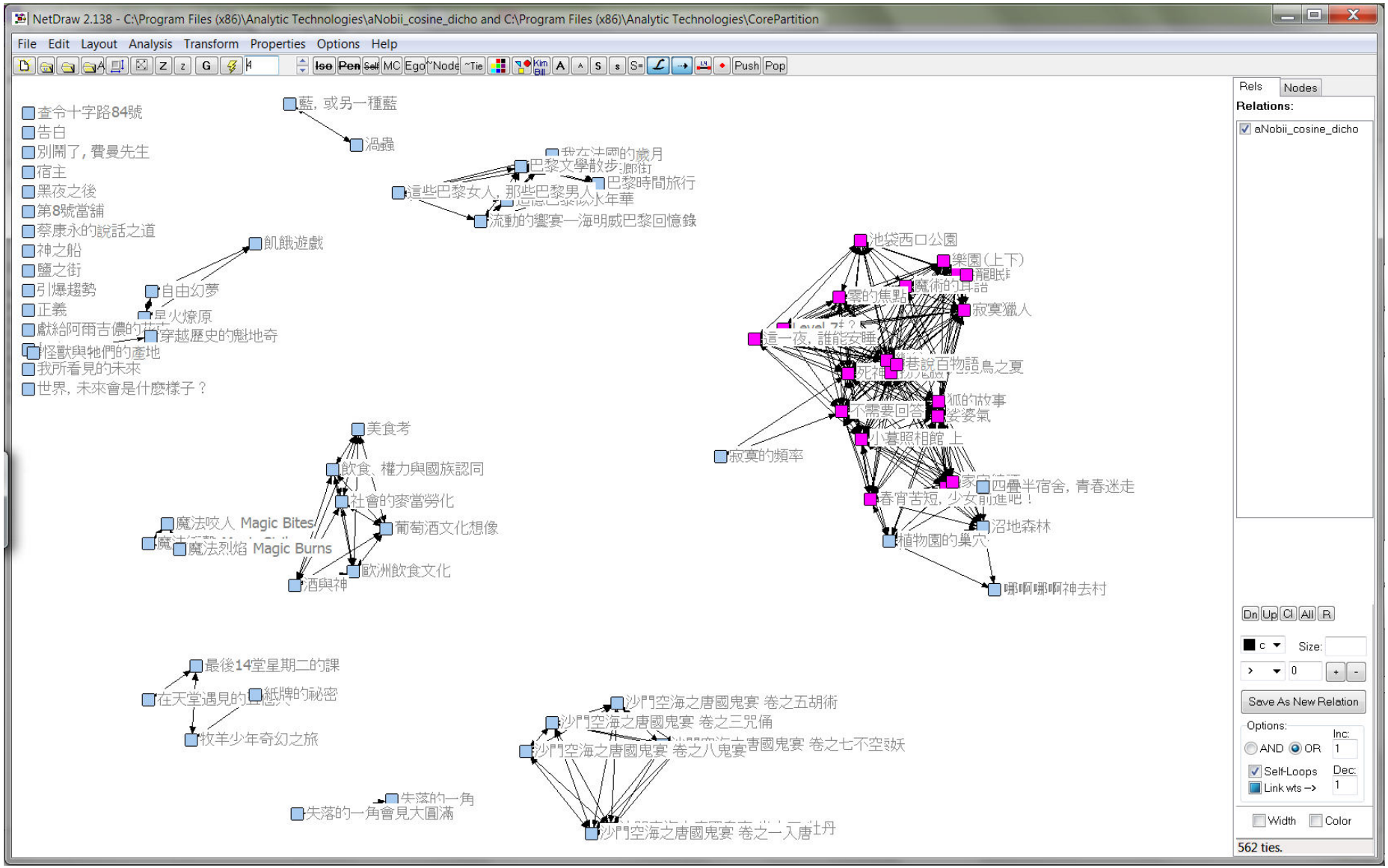


Output：CorePartition、CoreClasses


于若蓉
黃乾綱
王健華
李昆翰
李世忠
卓玉聰
王燕超
潘美月
張世宗
張瓊穗
徐新逸
陳煒舜
朱則剛
吳玲玲
彭渰雯
唐牧群
廖遠光
柯皓仁
卜小蝶
林維真
吳怡瑾
邱銘心
吳美美
林麗娟
楊美華
吳哲夫
阮明淑
趙飛鵬
彭文正
陳和琴
陳省三
胡家瑜
陳昭珍
藍文欽
陳雪華
黃慕萱
吳明德
王梅玲
邱炯友
林奇秀
陳書梅
林頌堅
林珊如
盧秀菊

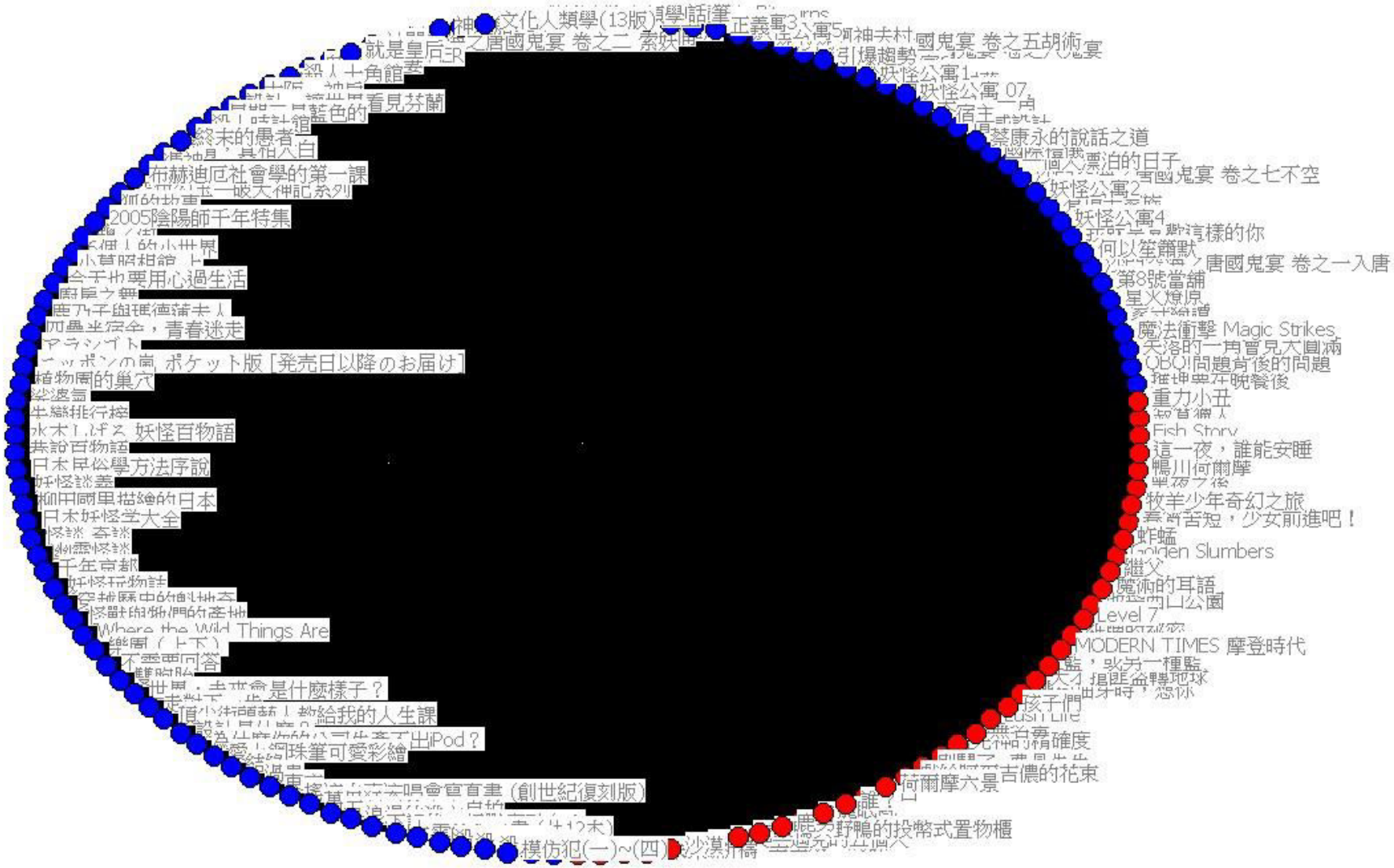
Exercise

- 1. Use aNobii cosine (aNobii cosin ##d ; aNobii cosin ## h) to
  - A. Open the file in UCINET and NetDraw to examine the cells
  - B Dichotomize by the threshold of 0.7, then run categorical core/periphery procedure, note where CorePartition file is saved
  - B. Import CorePartition into NetDraw
  - C. Visualize the raw data with NetDraw then set edge >0.7
  - E. Run continuous core/periphery procedures and save the coreness output
  - F. Input partition and coreness as attributes and redraw the book-book network, again set edge > 0.7

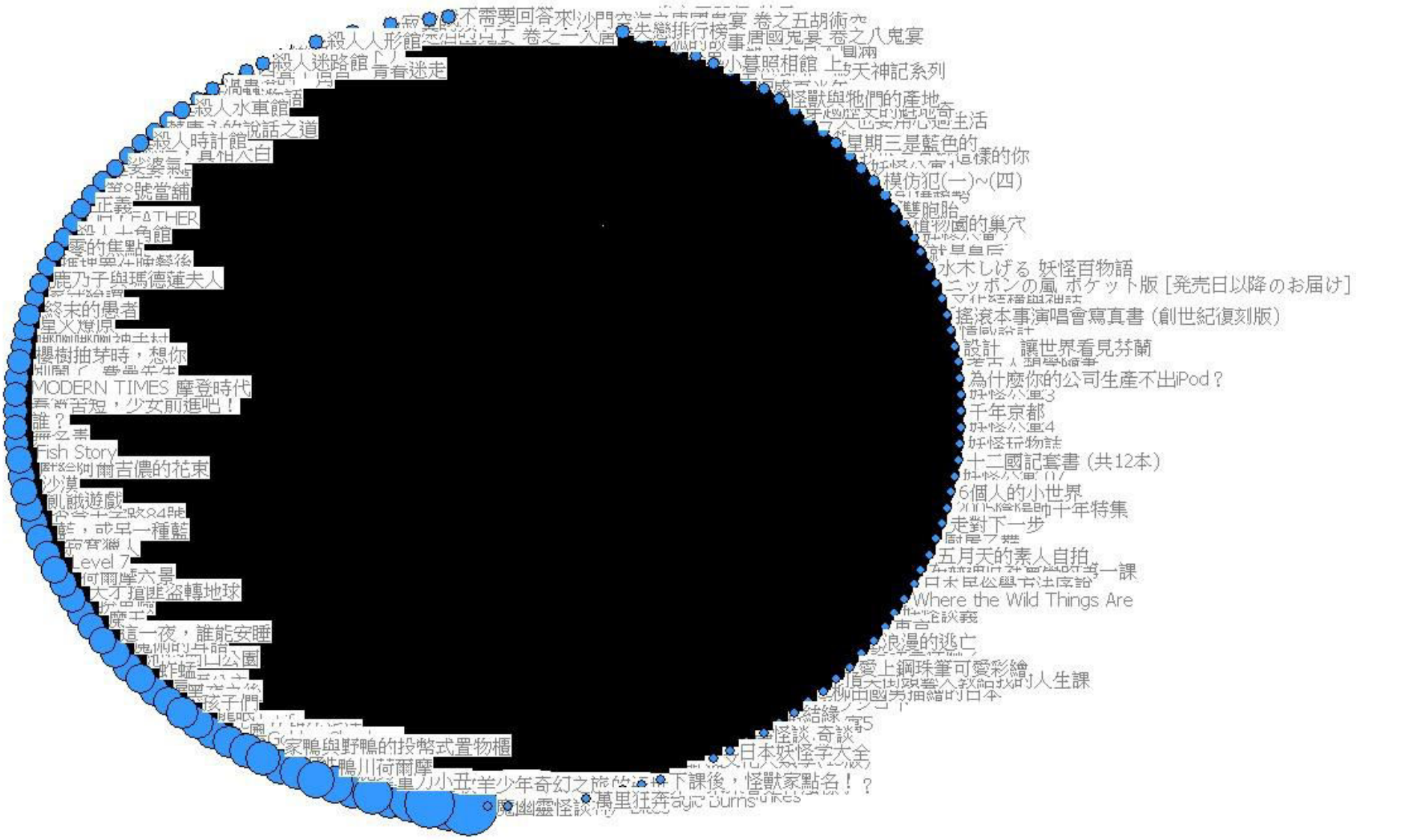

NetDraw 2.138 - C:\Program Files (x86)\Analytic Technologies\aNobii_cosine_dicho and C:\Program Files (x86)\Analytic Technologies\CorePartition
File Edit Layout Analysis Transform Properties Options Help
查令十字路84號
告白
別鬧了，費曼先生
宿主
黑夜之後
第8號當舖
蔡康永的說話之道
神之船
鹽之街
引爆趨勢
正義
獻給阿爾吉儂的花束
怪獸與牠們的產地
我所看見的未來
世界，未來會是什麼樣子？
藍，或另一種藍
渦蟲
巴黎文學散步
巴黎時間旅行
這些巴黎女人，那些巴黎男人
流動的饗宴—海明威巴黎回憶錄
飢餓遊戲
自由幻夢
星火燎原
穿越歷史的魁地奇
池袋西口公園
樂園（上下）
魔術的耳語
寂寞獵人
這一夜，誰能安睡
巷說百物語鳥之夏
狐的故事
婆婆氣
小暮照相館 上
寂寞的頻率
四疊半宿舍，青春迷走
春宵苦短，少女前進吧！
沼地森林
植物園的巢穴
哪啊哪啊神去村
美食考
飲食、權力與國族認同
社會的麥當勞化
葡萄酒文化想像
歐洲飲食文化
酒與神
魔法咬人 Magic Bites
魔法烈焰 Magic Burns
最後14堂星期二的課
在天堂遇見的紙牌的祕密
牧羊少年奇幻之旅
失落的一角會見大圓滿
沙門空海之唐國鬼宴 卷之五胡術
沙門空海之唐國鬼宴 卷之三咒俑
沙門空海之唐國鬼宴 卷之八鬼宴
沙門空海之唐國鬼宴 卷之一入唐
Rels
Nodes
Relations:
aNobii_cosine_dicho
Save As New Relation
Options:
AND
OR
Self-Loops
Link wts
Inc:
Dec:
Width
Color
562 ties.

Golden Slumbers
愛上鋼珠筆可愛彩繪
家守綺譚
龍眠
在天堂遇見的五個人
魔法衝擊 Magic Strikes
終末的愚者
鹿乃子
獻給阿爾吉儂的花束
門空海之唐國鬼宴 卷之
重力小丑
星空
黑夜之後
巷說百物語
沙漠
我所看見的未來
有頂天家族
沙妖怪談義唐國鬼宴 卷之三咒俑
妖怪公寓 07
殺人時計館
寂寞獵人
天才搶匪盜轉地球
今天也要用心過生活
設計是什麼？
星火燎原
怪談.奇談
布赫迪厄社會學的第一課
姑獲鳥之夏
柳田國男描繪的日本
世界，未來會是什麼
Level 7
就是皇后
植物園的巢穴
告白
鹿男
Fish Story
失落的一角
水木しげる 妖怪百物語
怪獸與牠們的產地
四疊半宿舍，青春迷走
哲學的慰藉
蔡康永的說話之道
星期三是藍色的前進吧
沙門空海之唐國鬼宴 卷之八鬼宴
妖怪公寓6
不需要回答
我就是喜歡這樣的你
魔法烈焰 Magic Burns
豐臣公主
沙門空海之唐國鬼宴 卷之五胡術
下課後，怪獸家點名！
雙胞胎
飢餓遊戲
妖怪公寓3
幽靈怪談
Where the Wild Things Are
妖怪公寓5
日本妖怪学大全
設計 讓世界看見芬蘭
搖滾本事演唱會寫真書（創世紀復刻版）
國際禮儀
十二國記套書（共12本）
結緣
QBQ!問題背後的問題
情感設計
繼父
櫻樹抽芽時，想你
蚱蜢
索妖
穿越歷史的
為什麼你的公司生產不出iPod？
無名毒
妖怪公寓1
沼地森林
廚房之舞
孩子們
頂尖街頭藝人教給我的人生課
推理要在晚餐後
鬼太郎之妻
6個人的小世界
模仿犯(一)~(四)
五月天的素人自拍
藍，或另一種藍
妖怪公寓4
樂園（上下）
殺人人形館
歡迎光臨人類學
正義
日本民俗學方法序說
誰？
引爆趨勢
魔法咬人 Magic Bites
失戀排行榜
浪漫的逃亡
池袋西口公園
寂寞的頻率
這一夜，誰能安睡
走對下一步，曼先生
OH！FATHER
殺人迷路館
殺人水車館
牧羊少年奇幻之旅
見大圓滿
東京
千年京都
查令十字路84號
盛夏光年

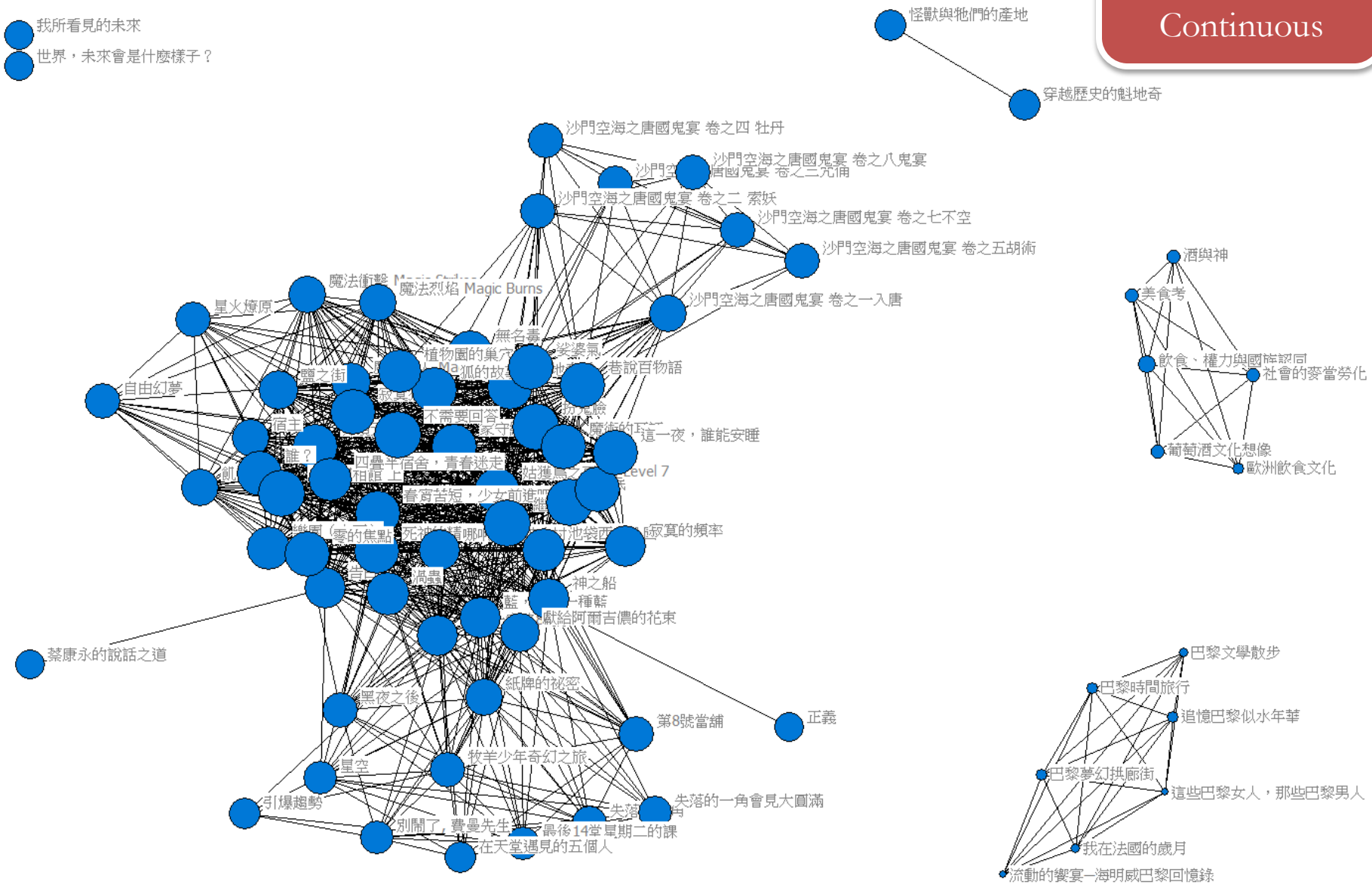
aNobii-Raw data – Categorical



aNobii-Raw data
Continuous



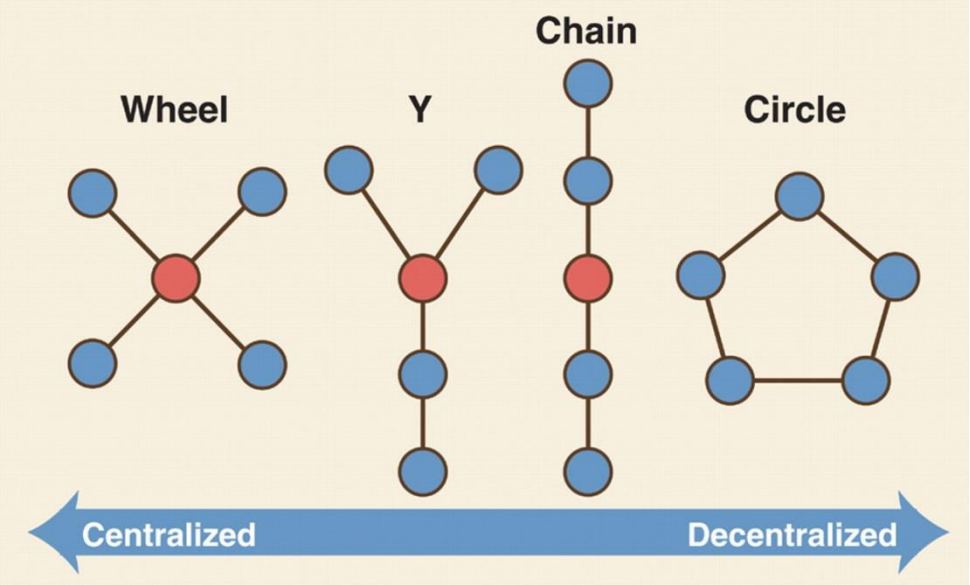
aNobii-dichotomized–Continuous



SNA

Centrality

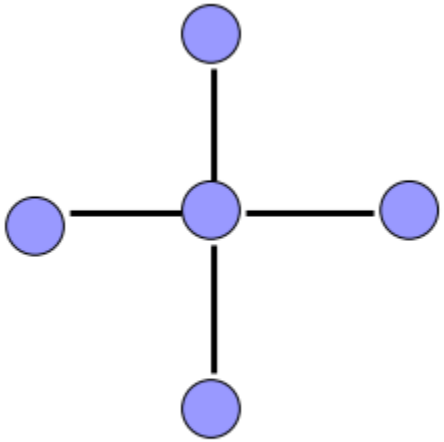
The Bavelas-Leavitt Experiment



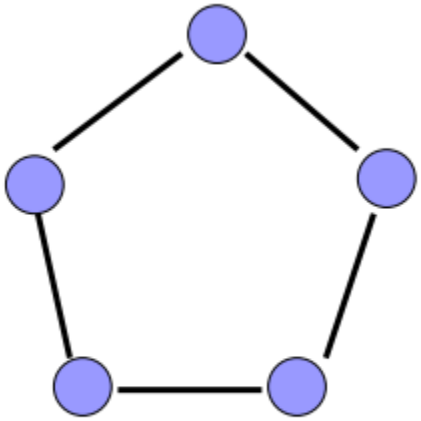
Variables	Simple Task	Complex task
Fewest messages	Centralized	Centralized
Least errors	Centralized	Decentralized
Least time	Centralized	Decentralized
Satisfaction	Decentralized	Decentralized

Network centralization

- One or few actors are quite central, other actors are not central
- Intuition



Centralized



Not centralized

Centrality

- Actor centrality (Micro)
  - Identify important or prominent actors
- Network centralization (Macro)
  - Characterize the structure of the network
    - Equality or evenness
  - If a system is very loosely coupled (sparse linkage) not much power can be exerted; in high density systems there is the potential for greater power.

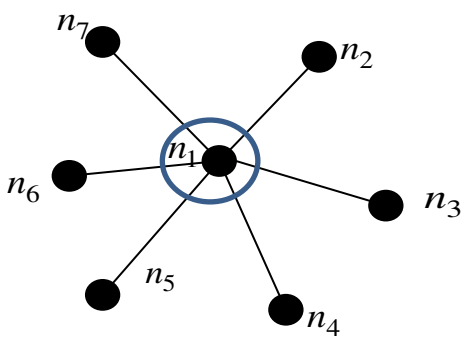
Prominent individuals in a network

- Not all nodes are created equal
  - Centrality
    - Prominent actors have higher involvement in many relations
  - Prestige (in a directional network)
    - Where prominent actors initiates few relations but receives many directed ties.
    - E.g. PageRank

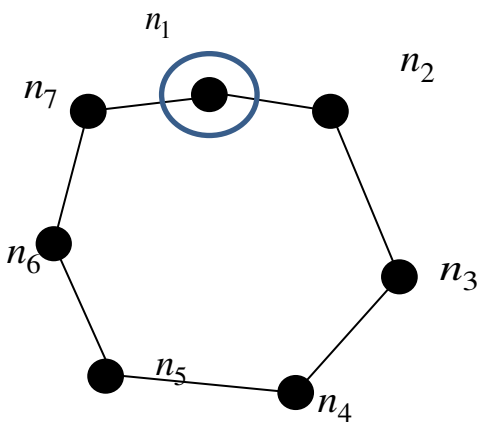
Centrality

- Network view of "power"
  - Manifest itself in social relationship
  - An ego's power is alter's dependence
- Centrality analysis identifies individuals having a favored position, having more opportunities and fewer constraints, and presumably possess more power

Micro and Macro view of centrality

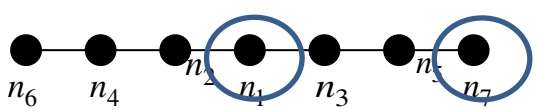


(a) Star Graph



(b) Circle Graph

(c) Line Graph



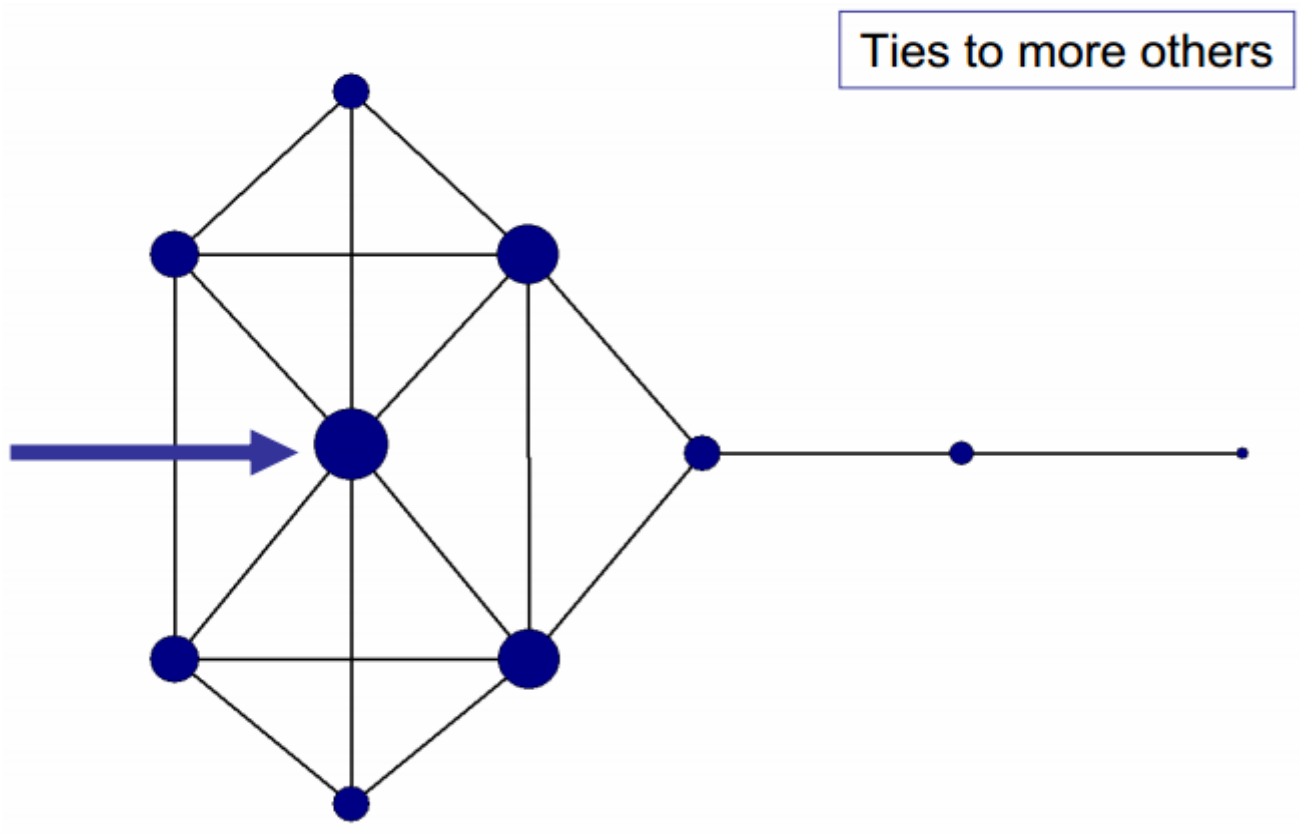
Micro: which node is the most central in each graph?

Macro: which graph has the most unevenly distributed degree ??

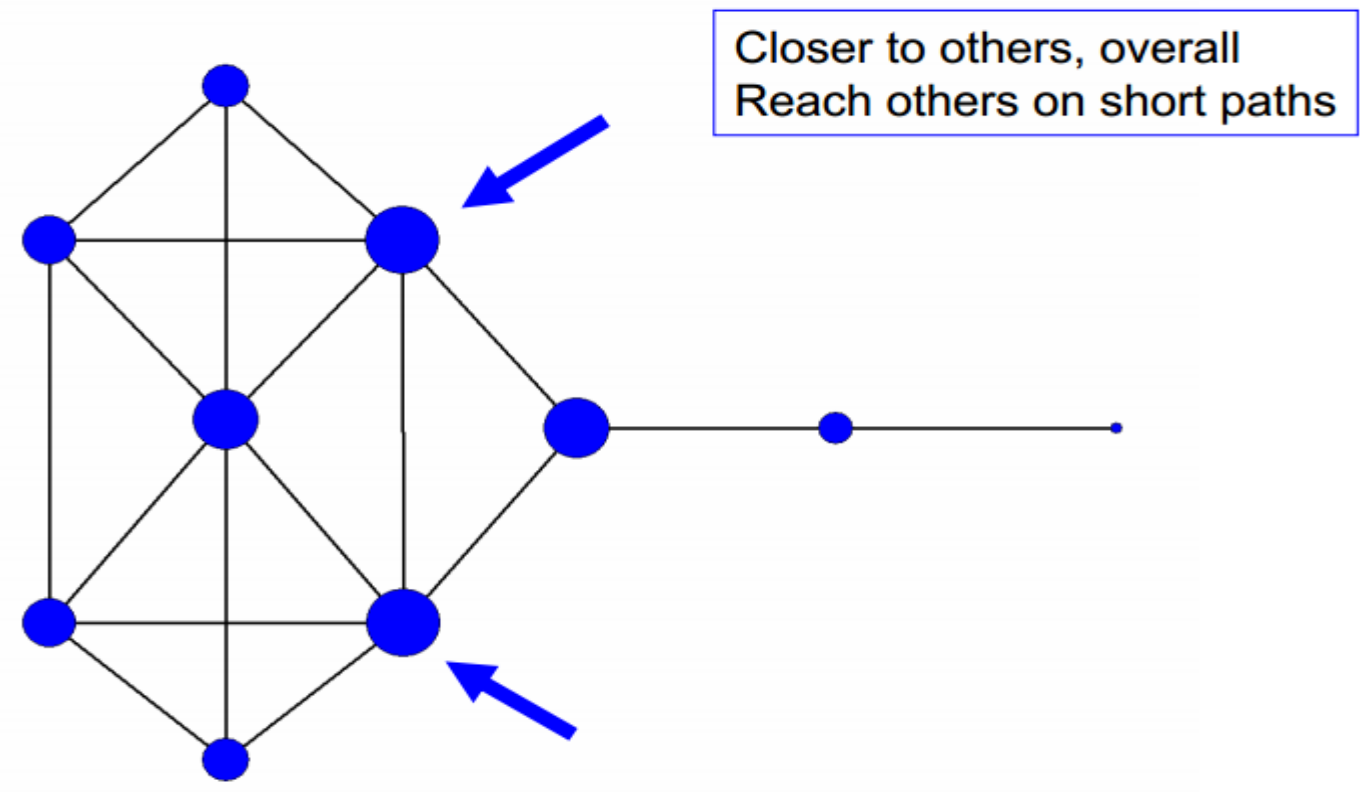
Centrality

- Degree centrality
  - Counting the number of paths of length 1 emanating from a node
- Closeness centrality
  - The total geodesic distance from a given node to all other nodes
- Betweenness centrality
  - The extent that that node falls on the geodesic paths between other pairs of nodes
- Eigenvector centrality
  - assigns relative scores to all nodes in the network based on the principle that **connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes**.

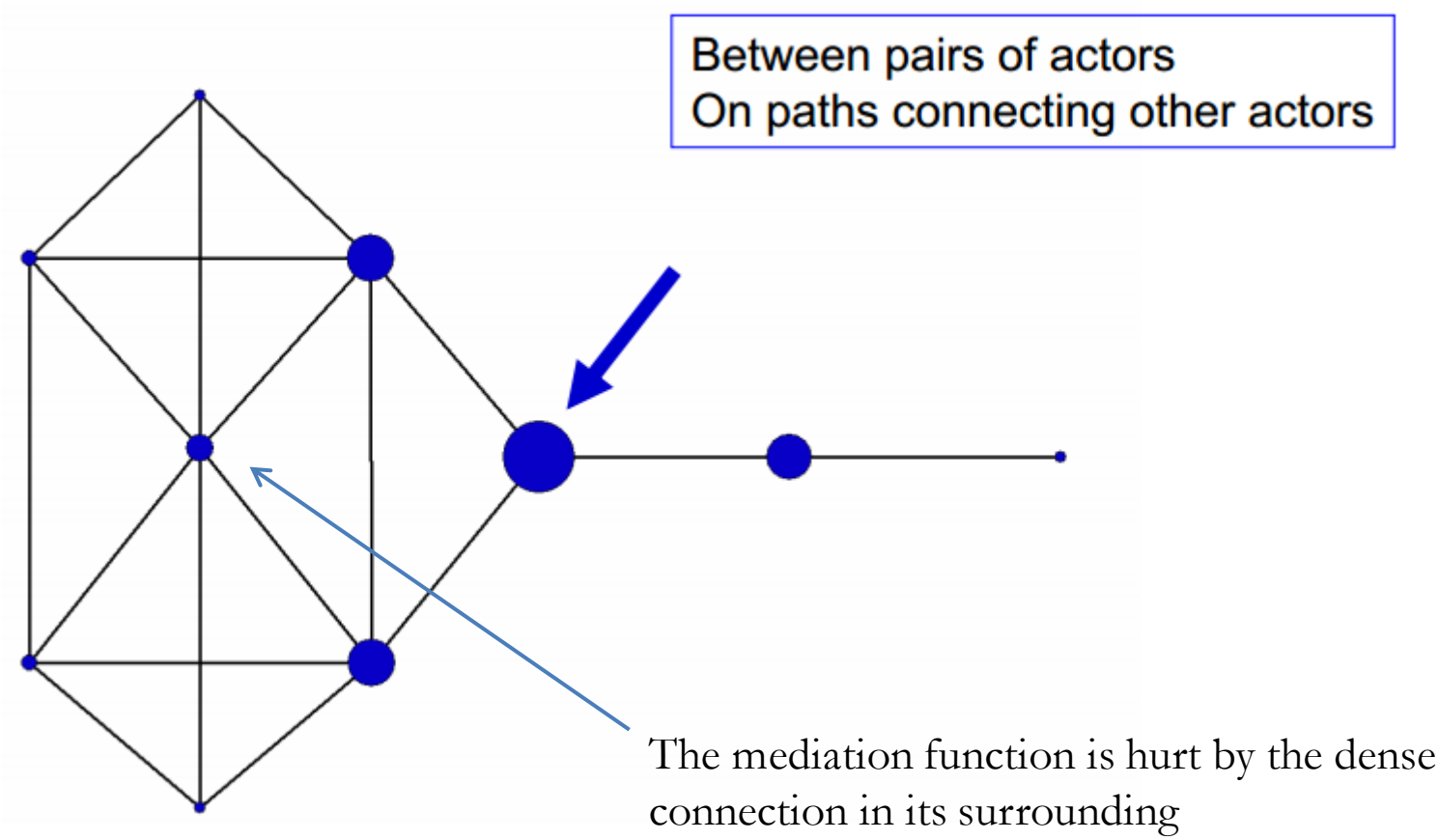
Actor degree centrality



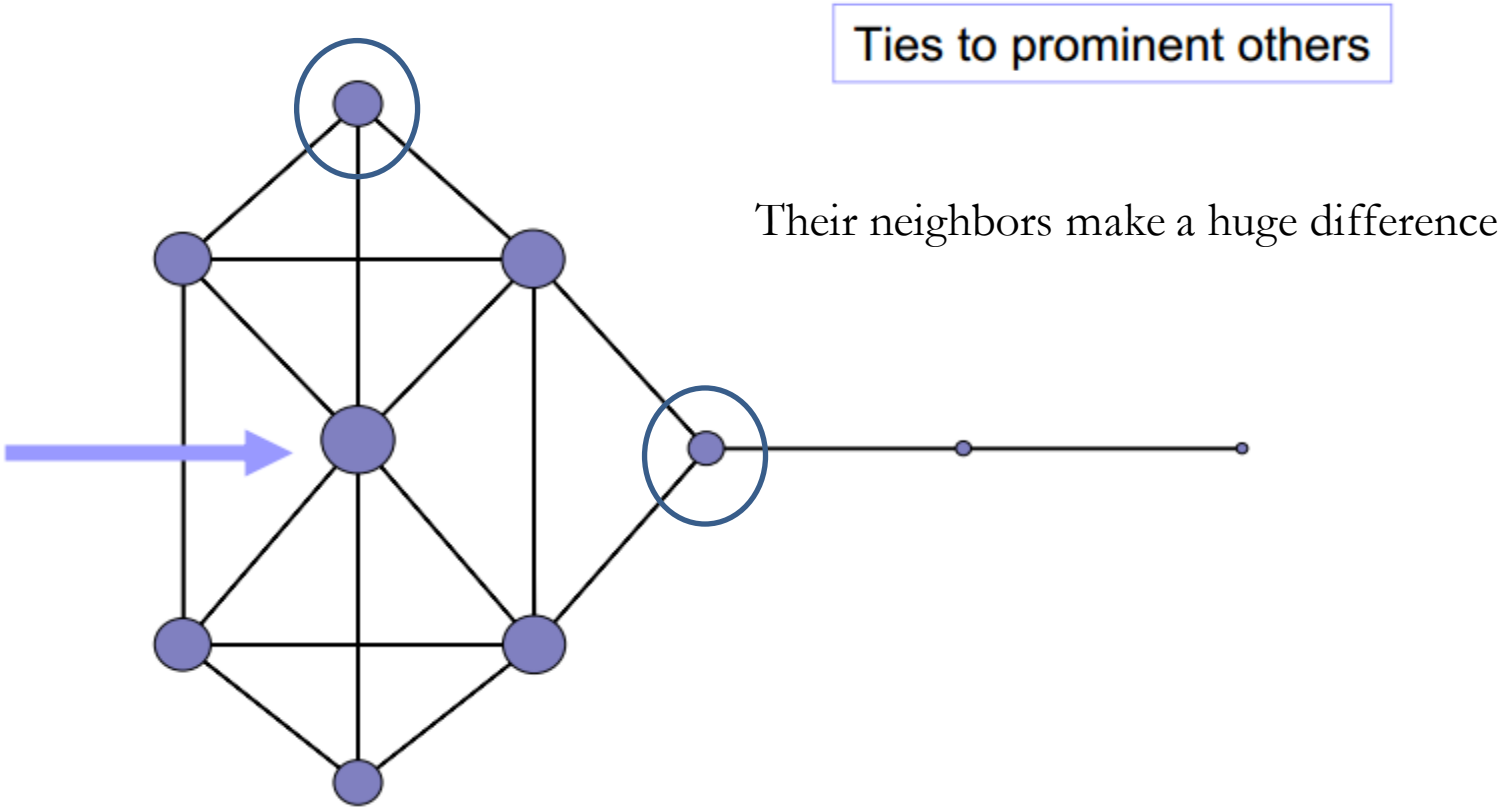
Actor closeness centrality



Actor betweenness centrality



Actor eigenvector centrality



Structural intuitions for actor centrality

- Intuition
  - Activity, visibility
  - Brokerage, control
  - Efficiency
  - Status, rank
- Measure
  - Degree
  - Betweenness
  - Closeness
  - Eigenvector

Actor Degree Centrality

- Measures the extent to which a node connect to all other nodes

$$C_D(Ni) = \sum_{j=1}^{g} X_{ij} \ (i \neq j)$$

For a network of size g,
Counts the number of direct ties that node i has to the g-1 other J nodes

General form of network attributes

1. Raw Measure—raw measures of network characteristics reflect a given characteristic that is not sensitive to the size or boundaries of a network or any of its websites
2. Normalized Measures—take on the general form of:

$$S_x = \frac{raw\ measure_x}{maximum\ value\ of\ measure\ given\ network\ characteristics}$$

$\rightarrow \mathbf{0 \leq S_x \leq 1}$

Normalized/standardized Actor Degree Centrality

- To eliminate the effect of variation in network size
  - Allow meaningful comparison of actors across different networks
- Vary from 0, indicating no connections with any actors (i.e. an isolate) to 1, reflecting direct ties to everyone

$$C'_D(Ni) = \frac{C_D(Ni)}{g - 1}$$

Freeman normalize group centralization measure

- Macro-view: measure the extent to which the actors in a social network differ from one another in their individual degree of centrality
- Express the degree of variability in the degrees of all actors in the network
- Normalized by the theoretical maximums (i.e. of a perfect star network of the same size)

Normalized Centrality and Centralization

- **Actor** centrality measure
  - Raw
  - Normalized
- Normalize **group** centralization
  - **Variation measure**

The largest actor degree centrality observed

$$C_A = \frac{\sum_{i=1}^{g}[C_A(N^*) - C_A(Ni)]}{\max \sum_{i=1}^{g} C_A(N^*) - C_A(Ni)]}$$

Theoretical maximum hint: star network

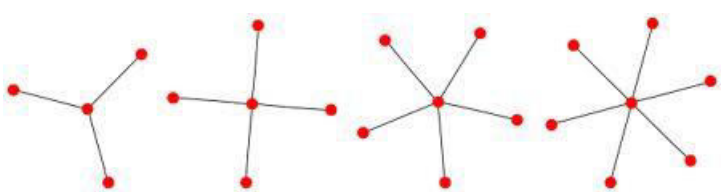
Group Degree Centralization (Macro)

- Measure the extent to which the actors in a social network differ from one another in their individual degree centralities
- A variance measure
- How does it translate into social world?

The largest actor degree centrality observed

$$C_A = \frac{\sum_{i=1}^{g}[C_A(N^*) - C_A(Ni)]}{\max \sum_{i=1}^{g} C_A(N^*) - C_A(Ni)]}$$

a perfect star network of the same size.



Network>Centrality>Degree

KNOKBUR

	1 OutDegree	2 InDegree	3 NrmOutDeg	4 NrmInDeg
1	4.000	5.000	44.444	55.556
2	7.000	8.000	77.778	88.889
3	6.000	4.000	66.667	44.444
4	4.000	5.000	44.444	55.556
5	8.000	8.000	88.889	88.889
6	3.000	1.000	33.333	11.111
7	3.000	9.000	33.333	100.000
8	6.000	2.000	66.667	22.222
9	3.000	5.000	33.333	55.556
10	5.000	2.000	55.556	22.222

DESCRIPTIVE STATISTICS

		1 OutDegree	2 InDegree	3 NrmOutDeg	4 NrmInDeg
1	Mean	4.900	4.900	54.444	54.444
2	Std Dev	1.700	2.625	18.889	29.165
3	Sum	49.000	49.000	544.444	544.444
4	Variance	2.890	6.890	356.790	850.617
5	SSQ	269.000	309.000	33209.875	38148.148
6	MCSSQ	28.900	68.900	3567.901	8506.173
7	Euc Norm	16.401	17.578	182.236	195.316
8	Minimum	3.000	1.000	33.333	11.111
9	Maximum	8.000	9.000	88.889	100.000

Network Centralization (Outdegree) = 38.272%
Network Centralization (Indegree) = 50.617%

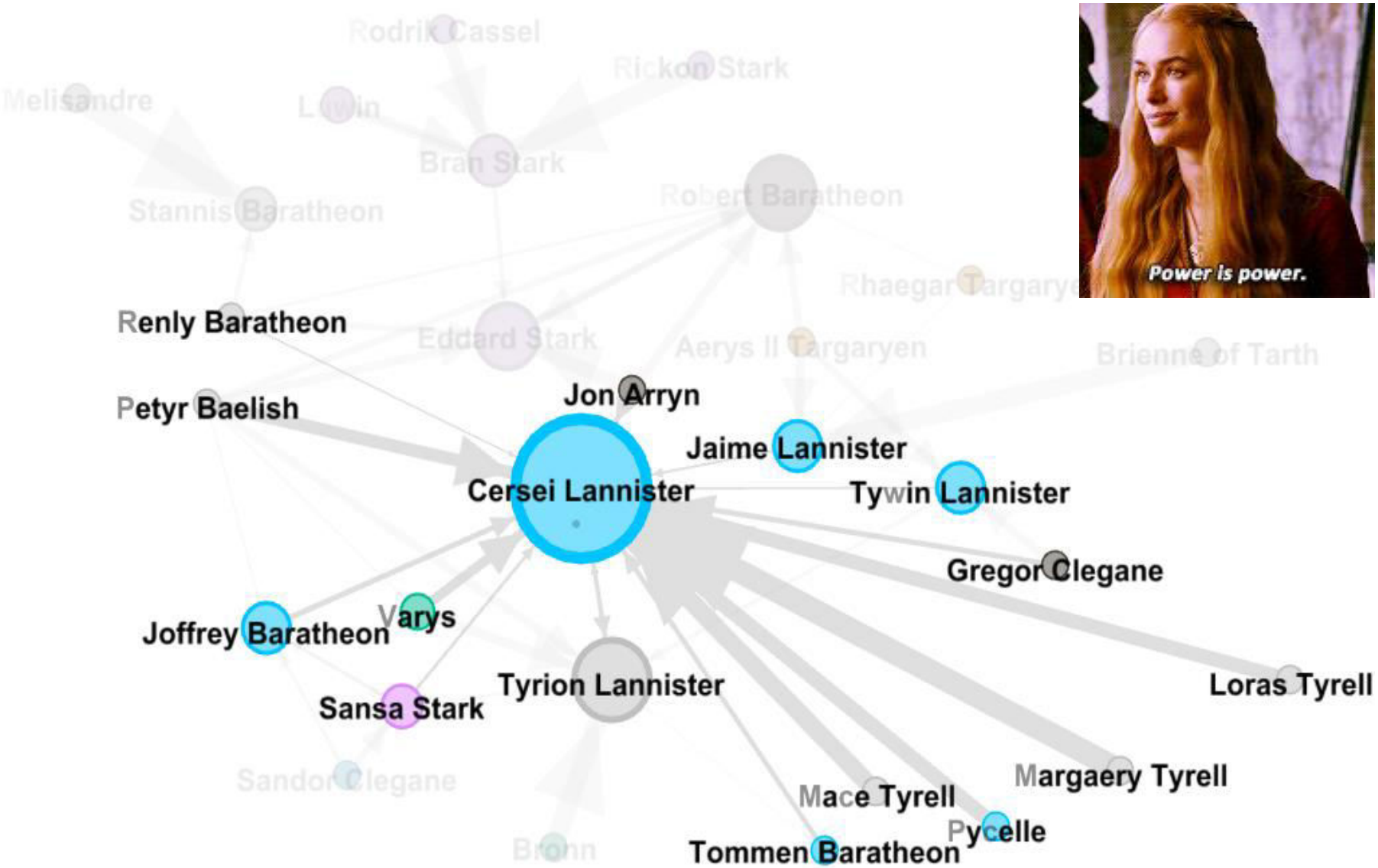
Drawbacks of degree measure

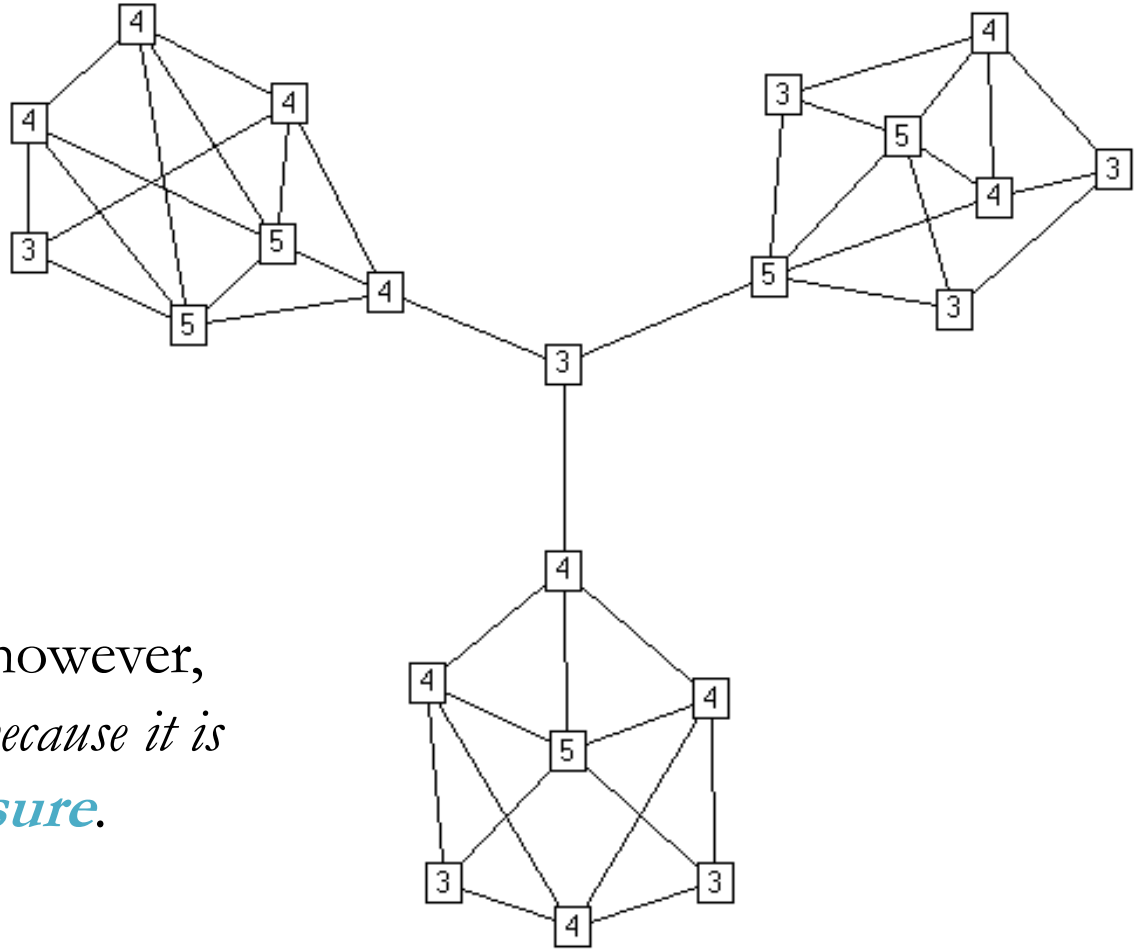
- An actor could be not quite central even with lots of connection
  - Overlook indirect links
  - One actor might be tied to a large number of others, but those others might be rather disconnected from the network as a whole. (occur more likely in a large network)
  - A "local" measure

- Or, a person with few ties yet exert great influence

Cersei is the main character?!

We use the size of node represents the In-degree





Degree centrality, however, can be deceiving, *because it is a purely* ***local measure***.

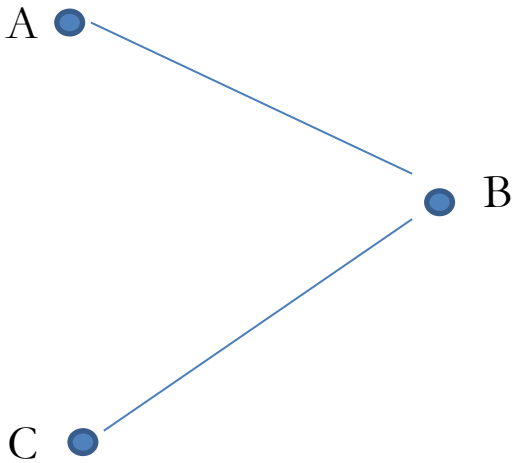
Actor Closeness Centrality

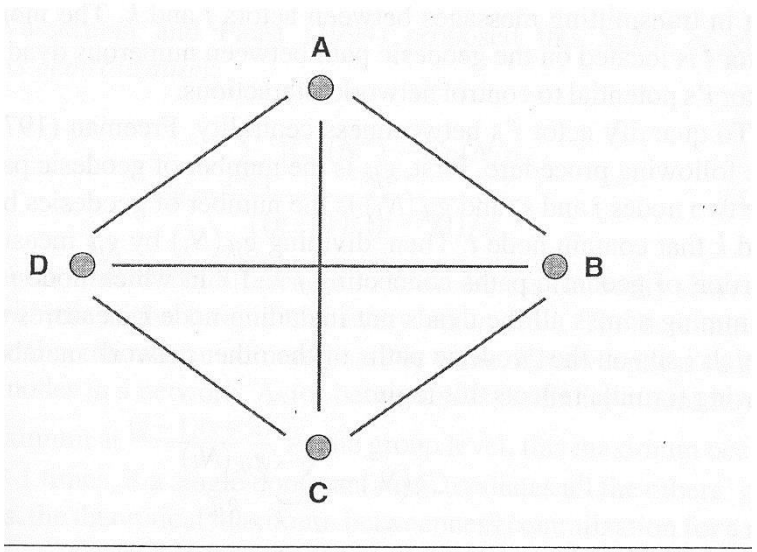
- Emphasizes the distance of an actor to all (not just the neighbors) others
- Reflect how far/near a node is to the other nodes in a social network
- How quickly an actor can interact with others

$$C_C(Ni) = \frac{1}{[\sum_{j=1}^{g} d(N_i, N_j)]} (i \neq j)$$

Can't be computed for isolated nodes, whose distances from i are undefined

Closeness centrality normalization





A Four-Node Completely Connected Network

The closeness centrality of A is
1/(1+2) = 1/3 (i.e. ab, ac)

The normalized closeness centrality is?
2/3

The closeness centrality of A is ?
AB, AC, AD **1/(1+1+1)**

The normalized closeness centrality is? 1

Need for normalization (n-1)

Normalized Actor Closeness Centrality

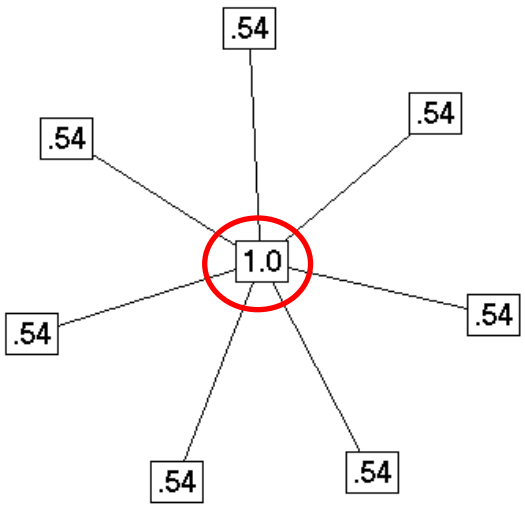
Computed as the inverse of the sum of the geodesic distances between actor i and the g – 1 other actors

$$C_C(Ni) = \frac{1}{[\sum_{j=1}^{g} d(N_i, N_j)]} (i \neq j)$$

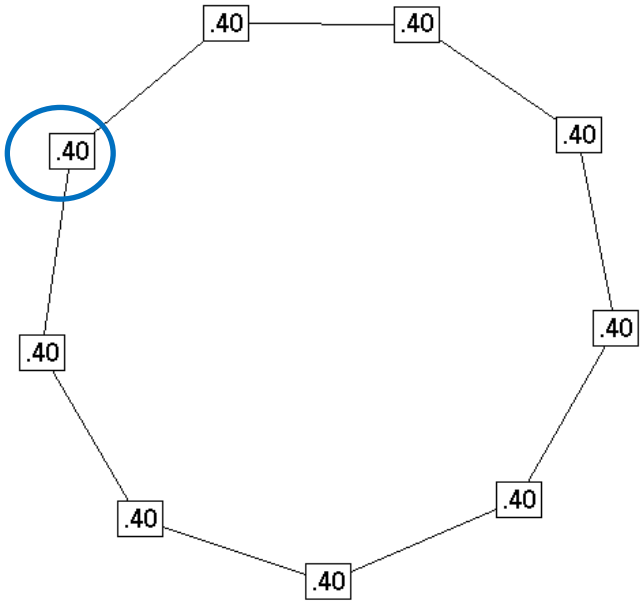
which varies with network size,
to control for the size of the network

$$C'_C(N_i) = (g-1)(C_C(N_i))$$

Closeness Centrality of two graphs



Distance								closeness	normalized
0	1	1	1	1	1	1	1	.143	1.00
1	0	2	2	2	2	2	2	.077	.538
1	2	0	2	2	2	2	2	.077	.538
1	2	2	0	2	2	2	2	.077	.538
1	2	2	2	0	2	2	2	.077	.538
1	2	2	2	2	0	2	2	.077	.538
1	2	2	2	2	2	0	2	.077	.538
1	2	2	2	2	2	2	0	.077	.538



Distance									closeness	normalized
0	1	2	3	4	4	3	2	1	.050	.400
1	0	1	2	3	4	4	3	2	.050	.400
2	1	0	1	2	3	4	4	3	.050	.400
3	2	1	0	1	2	3	4	4	.050	.400
4	3	2	1	0	1	2	3	4	.050	.400
4	4	3	2	1	0	1	2	3	.050	.400
3	4	4	3	2	1	0	1	2	.050	.400
2	3	4	4	3	2	1	0	1	.050	.400
1	2	3	4	4	3	2	1	0	.050	.400

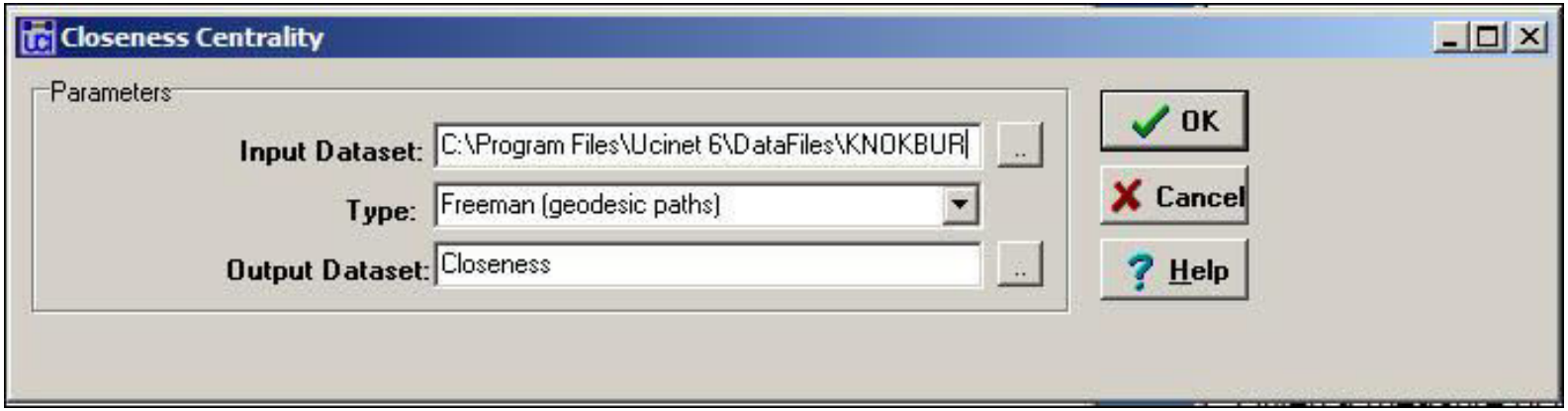
Index of Group Closeness Centralization (Macro)

- The extent to which actors in a given network differ in their closeness centrality
- A dispersion measure indicating the hierarchy of closeness centralities within a network (p. 67)
- It's sociological implications?

The largest actor closeness centrality observed

$$C_C = \frac{\sum_{i=1}^{g} [C'_C(N^*) - C'_C(N_i)]}{[(g-2)(g-1)]/(2g-3)}$$

Path distances

- Network>Centrality>Closeness
- 

Closeness : the reciprocal of "farness"
(i.e. the sum of the lengths of the shortest paths from an ego to all other nodes

Closeness Centrality Measures

	1 inFarness	2 outFarness	3 inCloseness	4 outCloseness
7	9.000	16.000	100.000	56.250
5	10.000	10.000	90.000	90.000
2	10.000	11.000	90.000	81.818
4	13.000	15.000	69.231	60.000
9	13.000	16.000	69.231	56.250
1	14.000	15.000	64.286	60.000
3	14.000	12.000	64.286	75.000
10	16.000	13.000	56.250	69.231
8	17.000	13.000	52.941	69.231
6	22.000	17.000	40.909	52.941

Statistics

		1 inFarness	2 outFarness	3 inCloseness	4 outCloseness
1	Mean	13.800	13.800	69.713	67.072
2	Std Dev	3.682	2.227	17.584	11.616
3	Sum	138.000	138.000	697.133	670.721
4	Variance	13.560	4.960	309.201	134.925
5	SSQ	2040.000	1954.000	51691.488	46335.906
6	MCSSQ	135.600	49.600	3092.015	1349.255
7	Euc Norm	45.166	44.204	227.358	215.258
8	Minimum	9.000	10.000	40.909	52.941
9	Maximum	22.000	17.000	100.000	90.000

Network in-Centralization = 71.51%
Network out-Centralization = 54.14%

Measure of inequality in the distribution of distances across the actors

Actor 6 has the largest sum of geodesic distances from other actors (inFarness of 22) and to other actors (outFarness of 17).

The farness figure can then be re-expressed as nearness and normed relative to the greatest nearness observed in the graph (the inCloseness of actor 6)

Normalized Freeman closeness using new closeness procedure

Matrix: KNOKI

		1 OutFreeClo	2 InFreeClo	3 OutValClo	4 InValClo	5 OutRecipClo	6 InRecipClo
1	COUN	0.600	0.643	0.778	0.815	0.704	0.759
2	COMM	0.818	0.900	0.926	0.963	0.889	0.944
3	EDUC	0.750	0.643	0.889	0.815	0.833	0.722
4	INDU	0.600	0.692	0.778	0.852	0.704	0.778
5	MAYR	0.900	0.900	0.963	0.963	0.944	0.944
6	WRO	0.529	0.409	0.704	0.519	0.630	0.463
7	NEWS	0.563	1.000	0.741	1.000	0.648	1.000
8	UWAY	0.692	0.529	0.852	0.704	0.815	0.593
9	WELF	0.563	0.692	0.741	0.852	0.648	0.778
10	WEST	0.692	0.563	0.852	0.741	0.778	0.611

Different operationalization of closeness centrality

Sum of geodesic distances(Freeman) distances are lengths of shortest paths (geodesic), the standard Freeman measure.

Sum of reciprocal distances As an alternative to taking the reciprocal after the summation, the reciprocals can be taken before. In this case the closeness is the sum of the reciprocated distances (i.e. undefined) so that infinite distances contribute a value of zero. (Harmonic closeness centrality), **used when graph is NOT strongly connected THIS IS CALLED "HARMONIC CLOSENESS CENTRALITY" IN GEPHI**

Avg of reversed distances (Valente-Foreman) the reversed distance is the diameter minus the geodesic distance.

*When a graph is not strongly connected, a widespread idea is that of using the sum of reciprocal of distances, instead of the reciprocal of the sum of distances, with the convention :why?

Betweenness Centrality

- Reflects how other actors control or mediate the relations between dyads that are not directly connected
- Views an actor as being in a favored position to the extent that the actor falls on the geodesic paths between other pairs of actors in the network. That is, the more people depend on me to make connections with other people, the more power I have.

Betweenness centrality

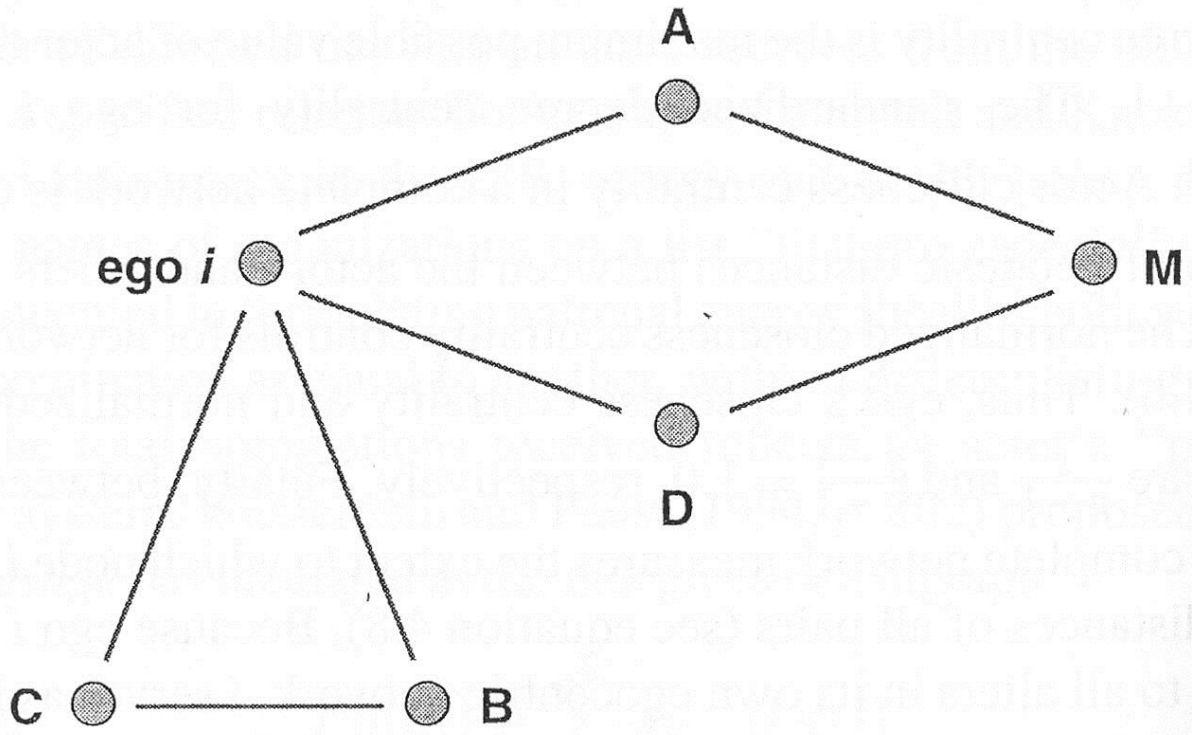


Figure 4.9 Betweenness Centrality in an Egocentric Network

Actor Betweenness Centrality

- Measures the extent to which other actors lie on the geodesic path (shortest distance) between pairs of actors in the network.

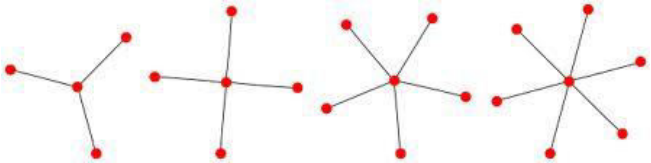
$$C_B = (N_i) = \sum_{j<k} \frac{g_{jk}(N_i)}{g_{jk}}$$

Sum the number of geodesic path between J and K that Contain node i

The number of geodesic path between J and K

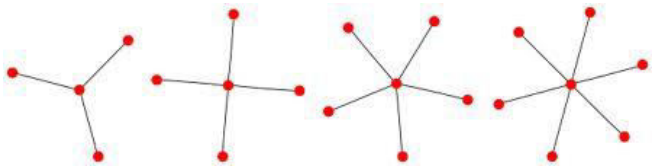
Normalized Actor Betweenness Centrality

- Normed by the maximum theoretical value of actor betweenness
- When node i falls on every geodesic path for all dyads, assuming each pair has only one geodesic path



Normalized Actor Betweenness Centrality

The normalized value varies from 0 to 1



$$C'_B(N_i) = \frac{C_B(N_i) X 2}{(g-1)(g-2)}$$

Excluding node i, the total number of geodesic paths among the (g-1) nodes will be?
(g-1)(g-2)/2

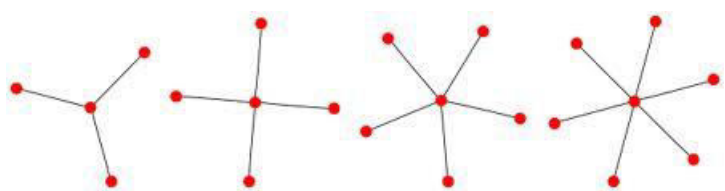
Index of Group Betweenness Centralization

- Reach 1 when a single dominant actor sits on all geodesic paths, reaches 0 when every node has the same betweenness centrality

Sum the differences in betweenness centrality for the actor with the highest value and every other actor

$$C_B = \frac{\sum_{i=1}^{g}[C_B(N^*) - C_B(N_i)]}{((g-1)^2(g-2))/2}$$

Maximum possible value of betweenness centralities for all nods in a network



Network>Centrality>Betweenness>Nodes

	1 Betweenness	2 nBetweenness
5	17.833	24.769
2	12.333	17.130
3	11.694	16.242
7	2.750	3.819
9	1.222	1.698
4	0.806	1.119
1	0.667	0.926
10	0.361	0.502
6	0.333	0.463
8	0.000	0.000

DESCRIPTIVE STATISTICS FOR EACH MEASURE

		1 Betweenness	2 nBetweenness
1	Mean	4.800	6.667
2	Std Dev	6.220	8.639
3	Sum	48.000	66.667
4	Variance	38.689	74.632
5	SSQ	617.290	1190.760
6	MCSSQ	386.890	746.316
7	Euc Norm	24.845	34.507
8	Minimum	0.000	0.000
9	Maximum	17.833	24.769

Network Centralization Index = 20.11%

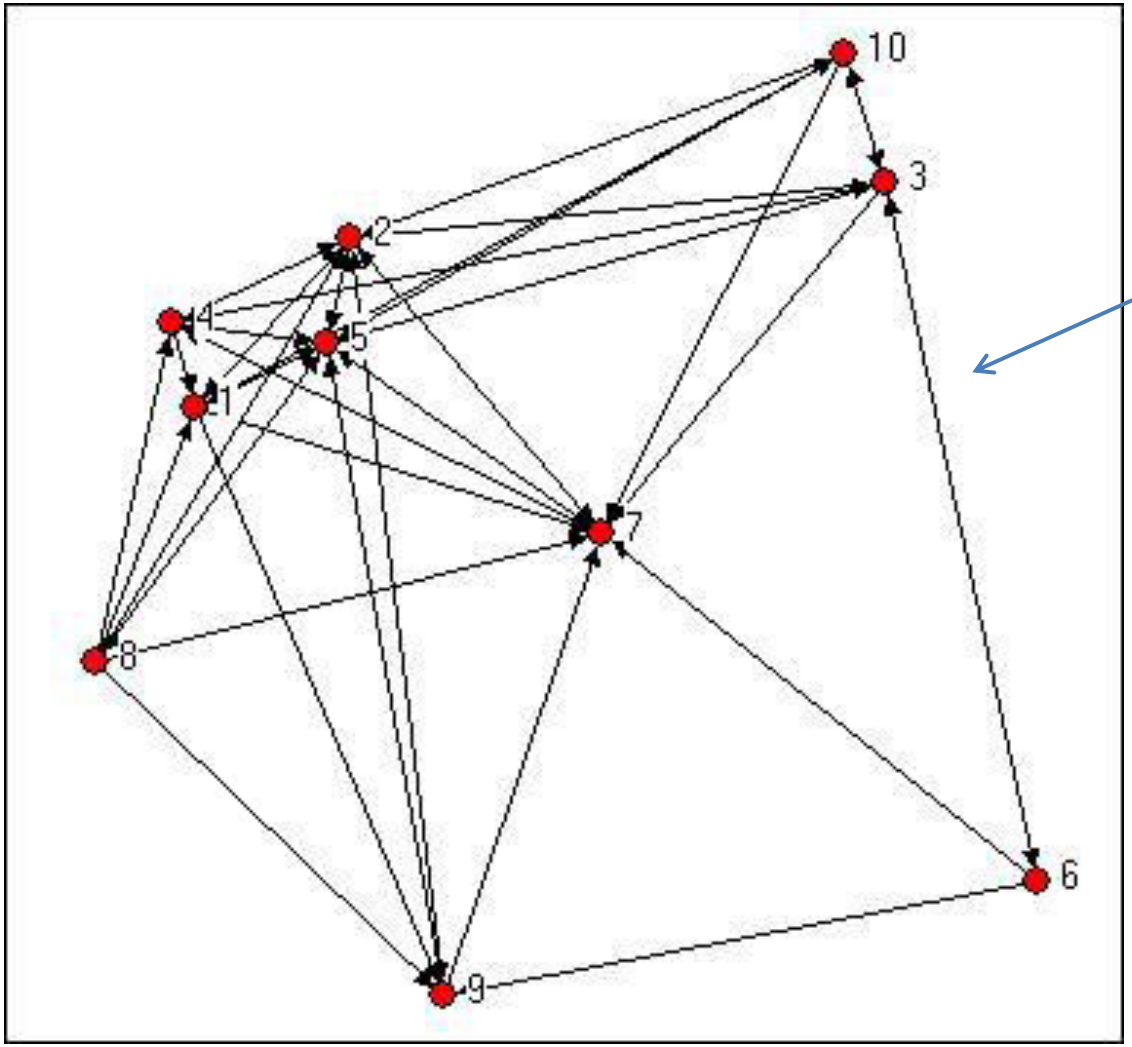
Kind of low, despite wide
Variation in betweenness
From 17.833 to 0
why?

Network>Centrality>Betweenness>Lines (edges)

Another way to think about betweenness is to ask which **relations** are most central, rather than which **actors**. Freeman's definition can be easily applied: a relation is between to the extent that it is part of the geodesic between pairs of actors.

Edge Betweenness

	1 COUN	2 COMM	3 EDUC	4 INDU	5 MAYR	6 WRO	7 NEWS	8 UWAY	9 WELF	10 WEST
1	0.000	2.833	0.000	0.000	3.833	0.000	1.333	0.000	1.667	0.000
2	2.417	0.000	7.500	1.917	1.500	0.000	1.000	4.500	2.500	0.000
3	0.000	2.694	0.000	2.111	2.694	9.333	1.000	0.000	0.000	2.861
4	2.139	2.833	0.000	0.000	3.833	0.000	1.000	0.000	0.000	0.000
5	2.417	1.000	7.000	1.917	0.000	0.000	1.000	4.500	2.500	6.500
6	0.000	0.000	3.944	0.000	0.000	0.000	2.833	0.000	2.556	0.000
7	0.000	3.944	0.000	2.861	4.944	0.000	0.000	0.000	0.000	0.000
8	1.000	2.000	0.000	1.000	3.000	0.000	1.000	0.000	1.000	0.000
9	0.000	3.944	0.000	0.000	4.944	0.000	1.333	0.000	0.000	0.000
10	1.694	2.083	2.250	0.000	2.083	0.000	1.250	0.000	0.000	0.000



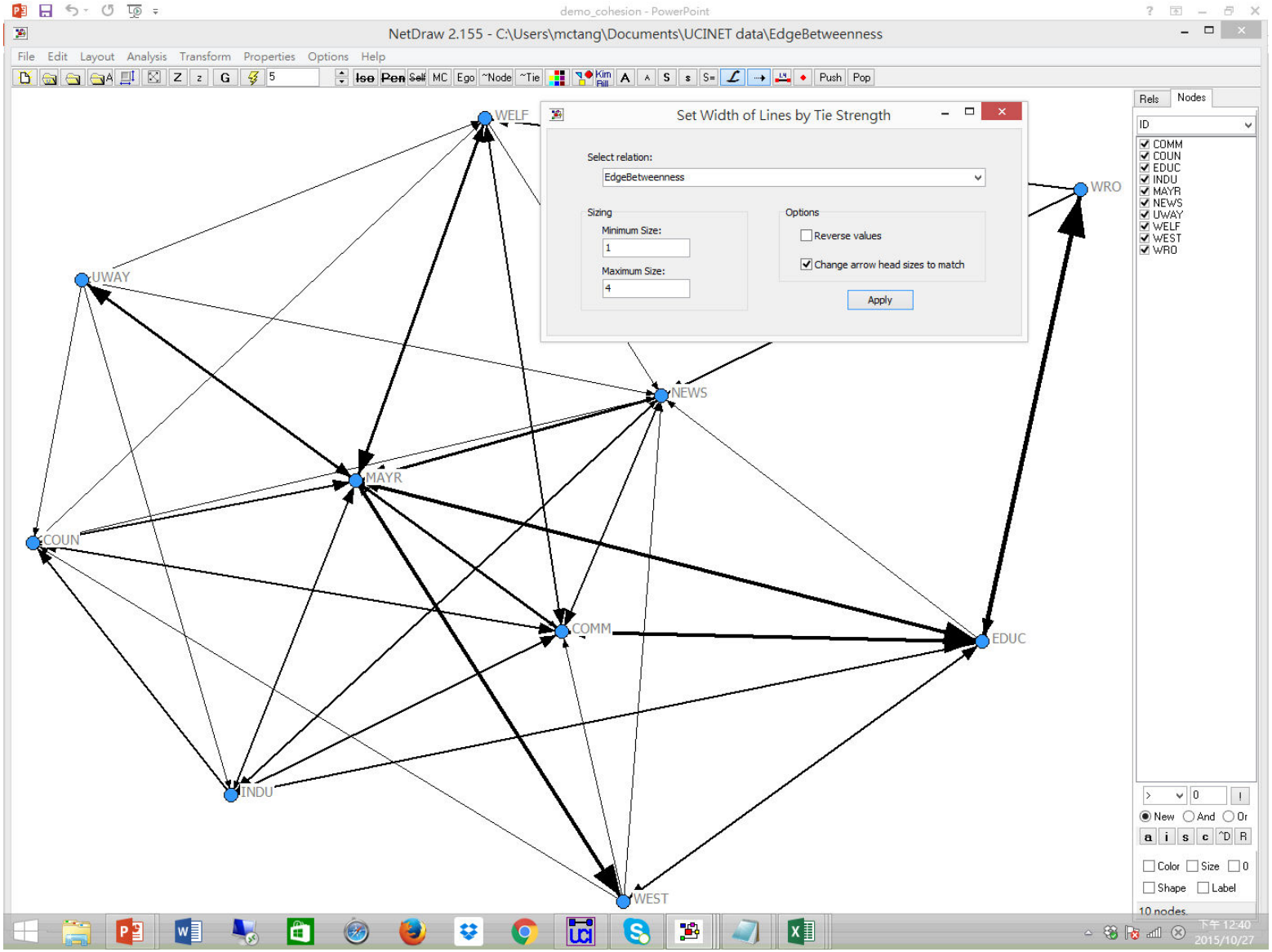
Betweenness between 3 and 6, 3.944
Relative high

This particular high value arises because without the tie to actor 3, actor 6 would be largely isolated.

Visualizing centrality

- Netdraw
  - Analysis>Centrality Measures
    - High tech (advise, friend, report-to)

Visualize line-betweenness



Betweenness: flow centrality

- Assume that actors will use all pathways (instead of the most efficient one only; blocked by a reluctant broker)
- Adds up how involved that actor is in all of the flows between all other pairs of actors (i.e. independent paths)
- Normalized by network size and density

Network>Centrality>Betweenness>Flow Betweenness

	1 FlowBet	2 nFlowBet
1	3.854	5.352
2	20.783	28.866
3	16.954	23.547
4	4.220	5.861
5	25.876	35.939
6	1.500	2.083
7	8.401	11.668
8	2.954	4.102
9	4.054	5.630
10	4.092	5.683

Network Centralization Index = 25.629%

DESCRIPTIVE STATISTICS FOR EACH MEASURE

		1 FlowBet	2 nFlowBet
1	Mean	9.269	12.873
2	Std Dev	8.230	11.430
3	Sum	92.687	128.732
4	Variance	67.725	130.642
5	SSQ	1536.335	2963.609
6	MCSSQ	677.249	1306.421
7	Euc Norm	39.196	54.439
8	Minimum	1.500	2.083
9	Maximum	25.876	35.939

	1 Betweenness	2 nBetweenness
5	17.833	24.769
2	12.333	17.130
3	11.694	16.242
7	2.750	3.819
9	1.222	1.698
4	0.806	1.119
1	0.667	0.926
10	0.361	0.502
6	0.333	0.463
8	0.000	0.000

DESCRIPTIVE STATISTICS FOR EACH MEASURE

		1 Betweenness	2 nBetweenness
1	Mean	4.800	6.667
2	Std Dev	6.220	8.639
3	Sum	48.000	66.667
4	Variance	38.689	74.632
5	SSQ	617.290	1190.760
6	MCSSQ	386.890	746.316
7	Euc Norm	24.845	34.507
8	Minimum	0.000	0.000
9	Maximum	17.833	24.769

Network Centralization Index = 20.11%

Actor eigenvector centrality

- Ties to prominent others gives you more power
- *Network>Centrality>Eigenvector* routine calculates individual actor centrality, and graph centralization using weights on the first eigenvector.
- The assumption is that each node's centrality is the sum of the centrality values of the nodes that it is connected to.

Review of centrality in undirected networks

Comparison

Comparing across these 3 centrality values

- Generally, the 3 centrality types will be positively correlated
- When they are not (low) correlated, it probably tells you something interesting about the network.

	Low Degree	Low Closeness	Low Betweenness
High Degree		Embedded in cluster that is far from the rest of the network	Ego's connections are redundant - communication bypasses him/her
High Closeness	Key player tied to important important/active alters		Probably multiple paths in the network, ego is near many people, but so are many others
High Betweenness	Ego's few ties are crucial for network flow	Very rare cell. Would mean that ego monopolizes the ties from a small number of people to many others. As in a hierarchy? (unidirectional)	

slide: Jim Moody

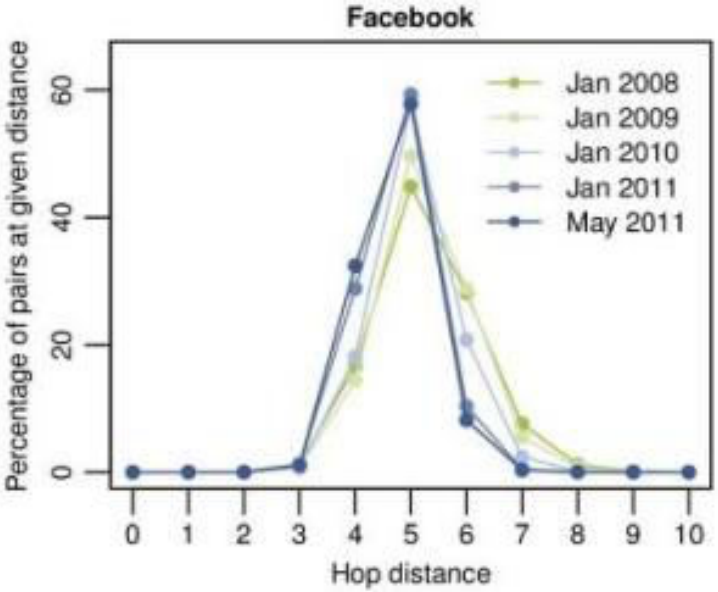
SNA

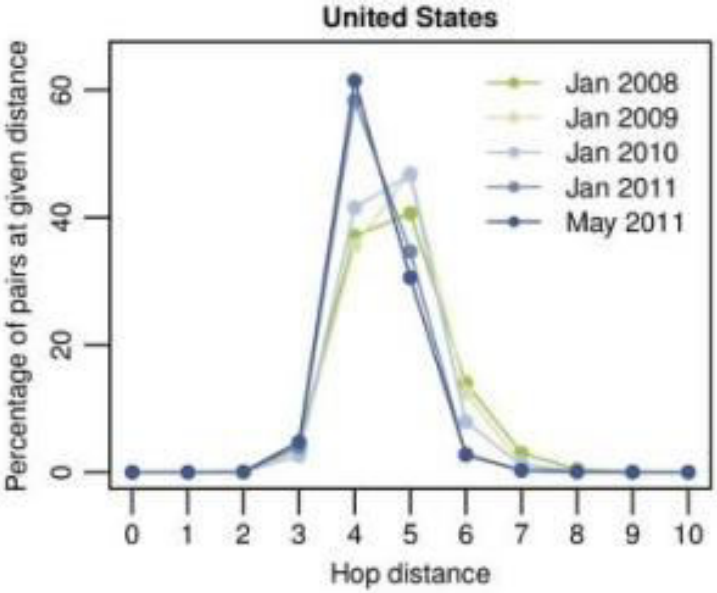
Small world, clustering coefficient

Small world

- that even when most of our connections are local, any pair of people can be connected by a fairly small number of relational steps.
  - Six degree of Kelvin Bacon

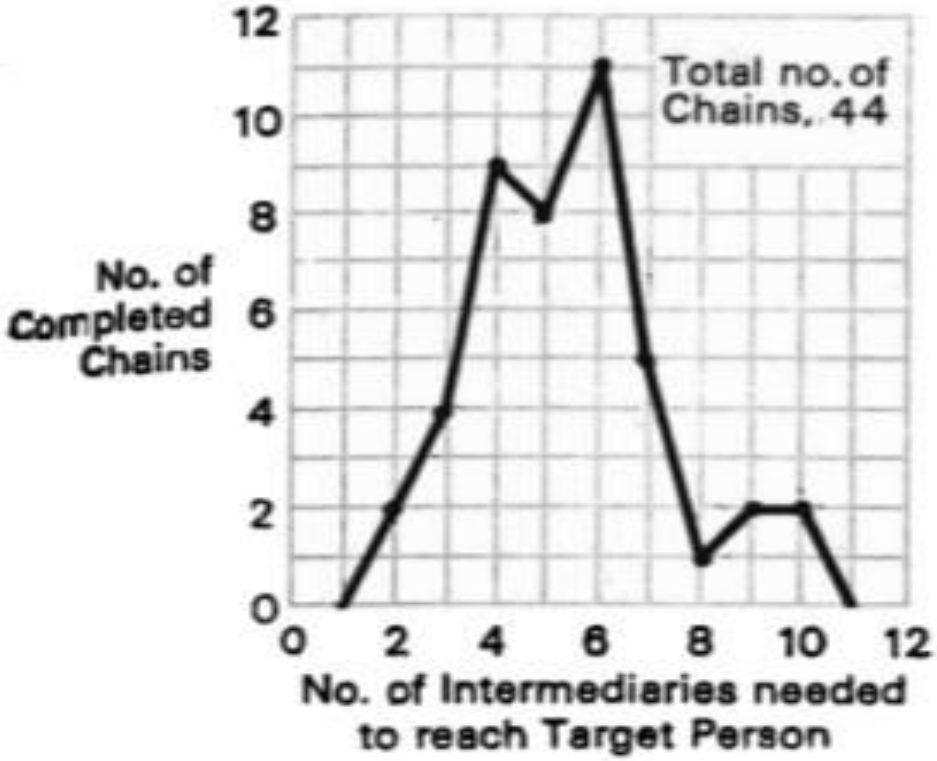
6 degree on the Facebook





Today it's 3 and half degrees of separation on FB
https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/

Milgram's 1967 "small world experiment



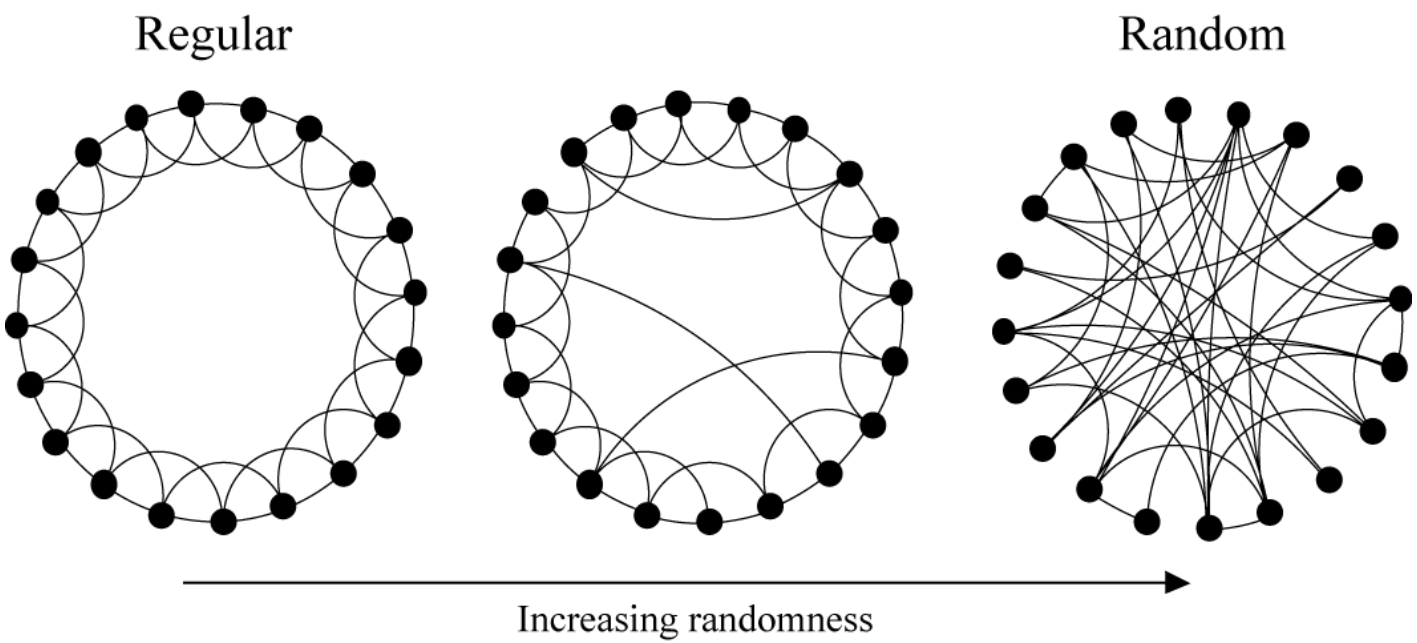
It's a Small World, After All

- This is essentially the "six degrees of separation" idea—that the number of "steps" or "links" needed to connect any one arbitrarily chosen individual to any other is low
- In Milgram's 1967 "small world experiment", individuals were asked to reach a particular target individual by passing a message along a chain of acquaintances.
- For successful chains, the an average (medium?) # of intermediaries needed was 5 (that is, 6 steps)—although note that most chains were not completed. (Easley and Kleinberg, 2010; p. 36, 37)

How to model the small world phenomenon ?

- Two possible graphs (almost at opposite ends of a spectrum) are “random graphs” and “regular graphs”.
- A “small world” can be thought of in-between a random and a regular graph.

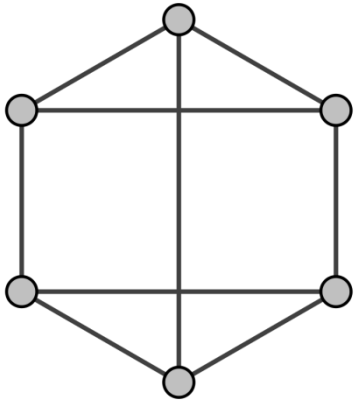
Modeling of the small world phenomenon



Regular Graphs

- A **regular graph** is a network where each node has the same number (k) of neighbors (that is, each node or vertex has the same degree k).

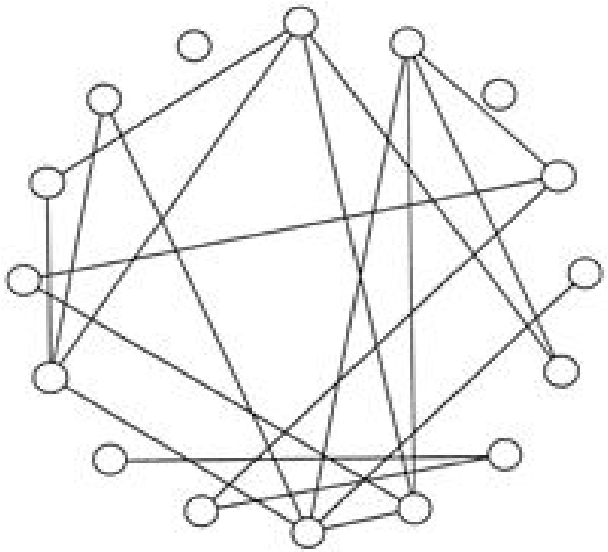
- A k-degree graph is seen at the left. k = 3 (each node is connected to three other nodes)



Is it a good model for real social network?

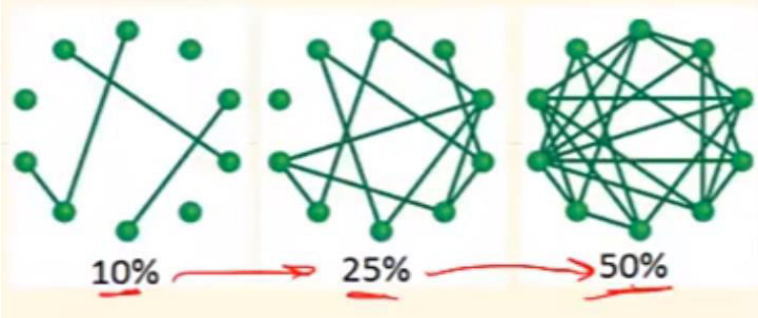
Random Graphs

- This graph has 16 nodes,
  - 120 possible connections, and
  - 19 actual connections
  - about a 1/7 probability that any two nodes will be connected to each other (i.e. density).
- Any two vertices have some *uniform probability* of being connected (i.e. independent event)



Random graphs

- Random graph formation
  - Consider each pair of nodes
  - Establish link with certain probability
  - What will happen if we have a fixed number of nodes and gradually raise the probability



Random Graphs

- Studied by P. Erdös A. Rényi in 1960s
- How to build a random graph
  - **Take *n nodes***
  - **Connect each pair of nodes with an edge with some probability *p***
- For a network of n nodes, this condition implies that the expected degree (K) of an actor is

Average degree $k = p\,(n-1)$

Let n = number of nodes
P = a fix probability that any pair of nodes are connected
E = total number of edges in the graph

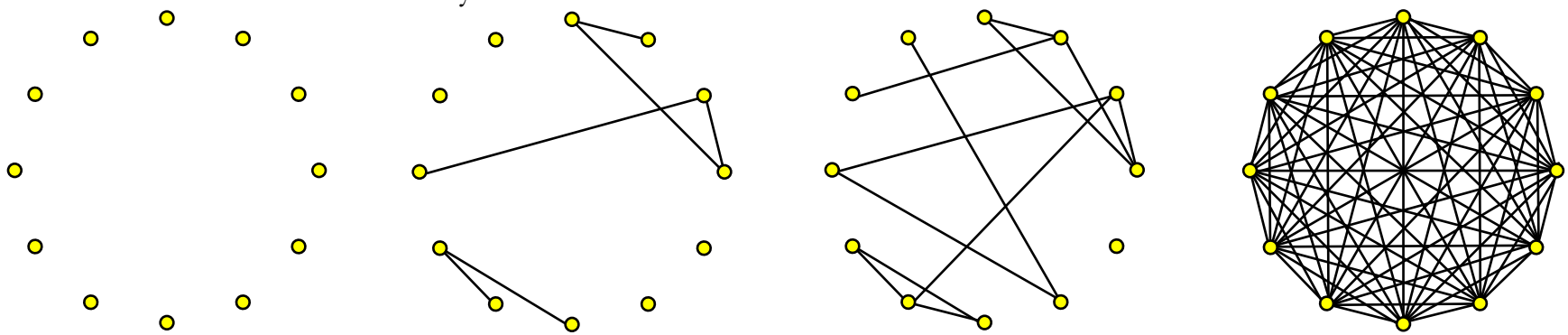
Average degree K = 2E/n (why 2E?; E = P * all possible pairs, which is?)
= P * n (n-1)/n
= P * (n-1)

$$K = \frac{n(n-1)p}{n} = (n-1)p \approx np$$

Random Graphs

Preexistence of a fixed number of nodes (N : number of nodes : 12)
uniform probability of being connected (P : the probability of two nodes being connected).
Average degree, $k = p\ (N-1)$

*Slides modified from Kentaro Toyama



$p = 0.0$; $k = 0$ | $p = 0.09$; $k = 1$ | $p = 0.181$; $k = 2$ | $p = 1.0$; $k \approx N$

Size of Giant component

1 | 5 | **11** | 12

Let's try generate random graph using Gephi (directed/)

Phase transition and the critical threshold

- "Random network theory tells us that **as the average number of links per node increases beyond the critical one, the number of nodes left out of the giant cluster decreases exponentially... The more link we add, the harder it is to find a node that remains isolated. "**
- You start with a large number of isolated nodes. Then you randomly add links between the nodes, mimicking the random encounters between the quests.
- When you add enough links such **that each node has an average of one line**, a miracle happens: A unique giant cluster emerges. That is, most nodes will be part of a single cluster such that, starting from any node, we can get to any other by navigating along the links between the nodes. ...The network, after placing a critical number of links, drastically changes.

Simulation of a phase transition

Phase transition

*Slide adopted from Kentaro Toyama

Erdős and Renyi (1959)

If $k < 1$:

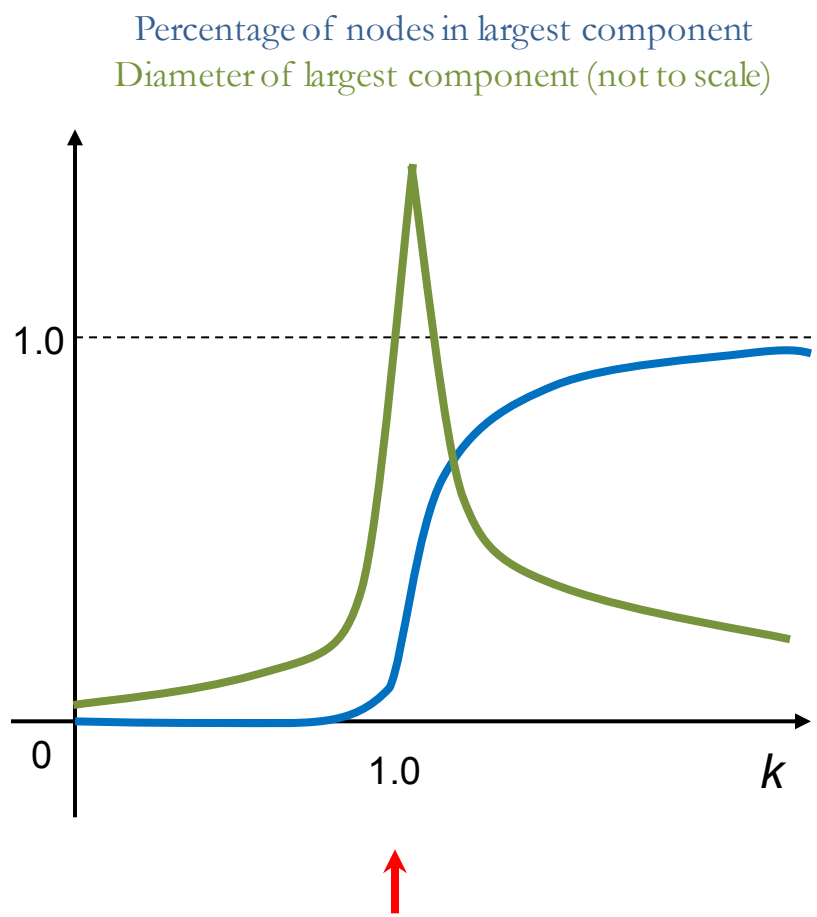
- small, isolated clusters
- small diameters
- short path lengths

At k = 1:

- a *giant component* appears
- diameter peaks
- path lengths are high

For k > 1:

- almost all nodes connected
- diameter shrinks
- path lengths shorten



As it is easy for each person to have at least one friend, therefore Small world

Random vs. Real Social Networks

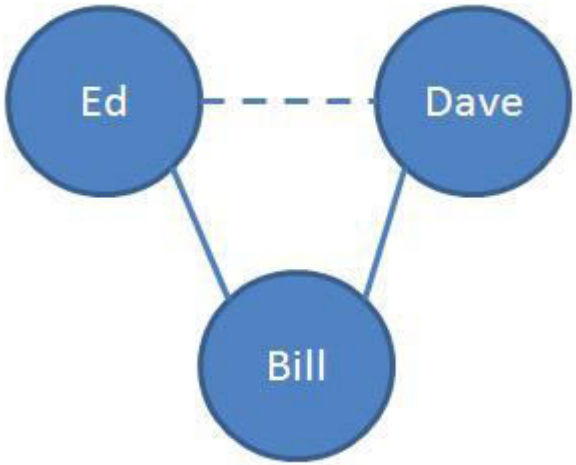
- Two problematic assumption of random network
  - Link formations are independent from each other
    - The presence of a connection between A and B as well as a connection between B and C will not influence the probability of a connection between A and C. (independence)
  - A link between any pair of nodes has a uniform probability
    - The degree distribution following bell curve
    - But, everyone is equally popular?

Clustering

- Most actors live in local neighborhoods where most others are also connected to one another
  - *Preferential attachment*
    - *Easier to make friends with people nearby*

Triadic closure

- If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in future.



Real Networks

- Local clustering
  - If A know B, and C, there is higher probability that B also know C
- Supper-connector
  - It was found in Milgram's experiment that approximately 60% of the transmissions passed the same four people
  - Network dynamics

In summary, random graphs

- Do not generate local clustering and triadic closures. ER graphs have a low clustering coefficient.
- Do not account for the formation of hubs. Formally, the degree distribution of ER graphs converges to a Poisson distribution, rather than a power law observed in many real-world, scale-free networks.

Modeling Real Network

- Duncan J. Watts and Steven Strogatz , 1998
  - Clustering
- Barabási, A.-L.; R. Albert (1999).
  - "Scale-free" network and preferential attachment
  - Look into the dynamic of link formation

Small-World Model

- Watts-Strogatz (1998) first introduced small world mode
- In between the regular and random networks
  - Regular Graphs have a high clustering coefficient, but also a high diameter
    - The length of the shortest path connecting two nodes grow very slowly, i.e., in general logarithmically, with the size of the network
  - Random Graphs have a low diameter, but a low clustering coefficient,

From Caveman graph to small world

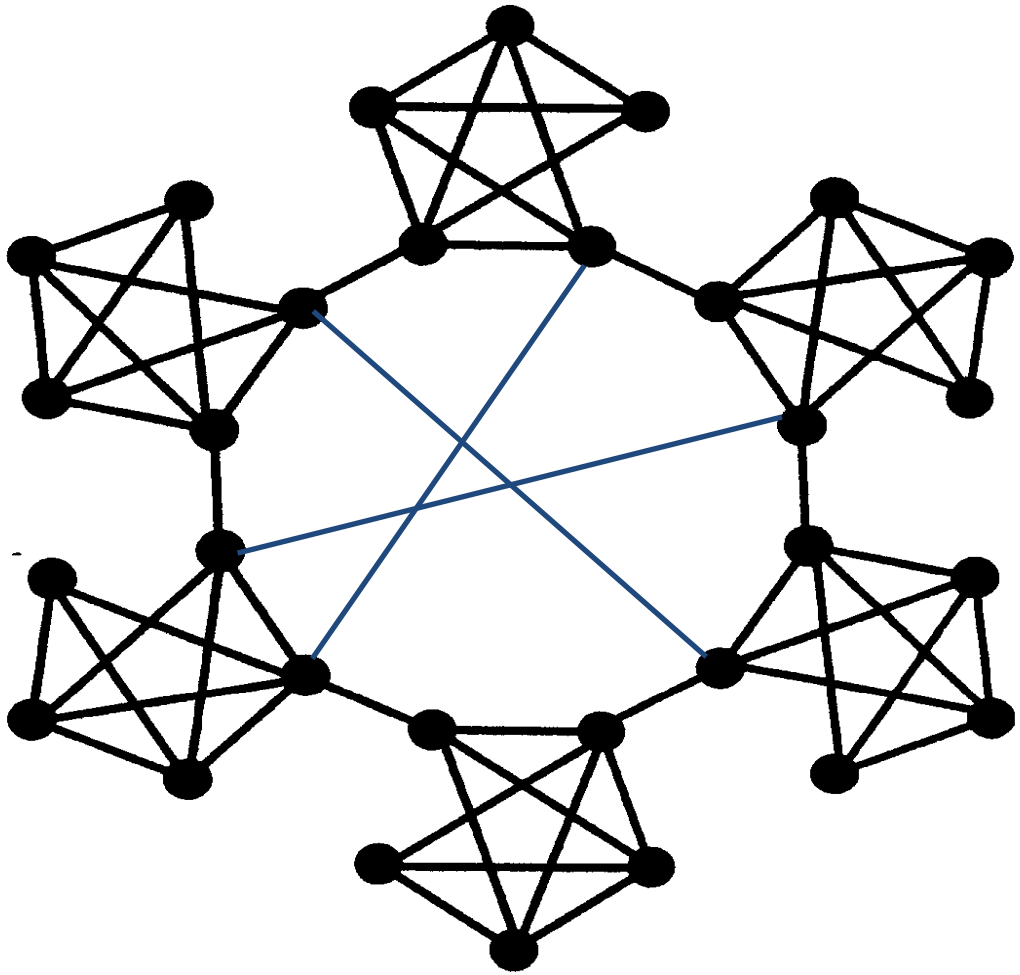


FIG. 1.—Schematic of the connected caveman construction

In a highly clustered, ordered network, a single random connection will create a shortcut that lowers L *(average path length)* dramatically

Small Worlds

- Note that it is highly clustered—a higher proportion (than one would expect randomly) of each node's neighbors are actually connected to *each other*
  - *Preferential attachment*
    - *Easier to make friends with people nearby*
- It also has a small diameter/short average path length, relative to the number of nodes.

Small Worlds

- A graph is small-world if it has significantly higher average clustering coefficient than a random graph constructed with the same number of edges, and if the graph has a short mean-shortest path length.
- These two characteristics are often mutually exclusive in random graphs—but do describe a wide variety of real-life situations.

Local cluster + random shortcuts

Watts-Strogatz's small world model

- Add edges to nodes, as in random graphs, but makes links more likely when two nodes have a common friend
- Add random links reduce diameter greatly

To quantify a small world

- Average path length
  - Measures the average number of intermediaries, that is, the degrees of separation, between any two actors in the network along their shortest path of intermediaries.
- Clustering coefficient
  - How many of an actor's contacts are connected to each other

Clustering coefficient

- Measure how close a node's neighbors are to being a clique
- If A is connected to B, and B is connected to C, then it is likely that A is connected to C
  - "A friend of your friend is MORE LIKELY TO BE your friend"

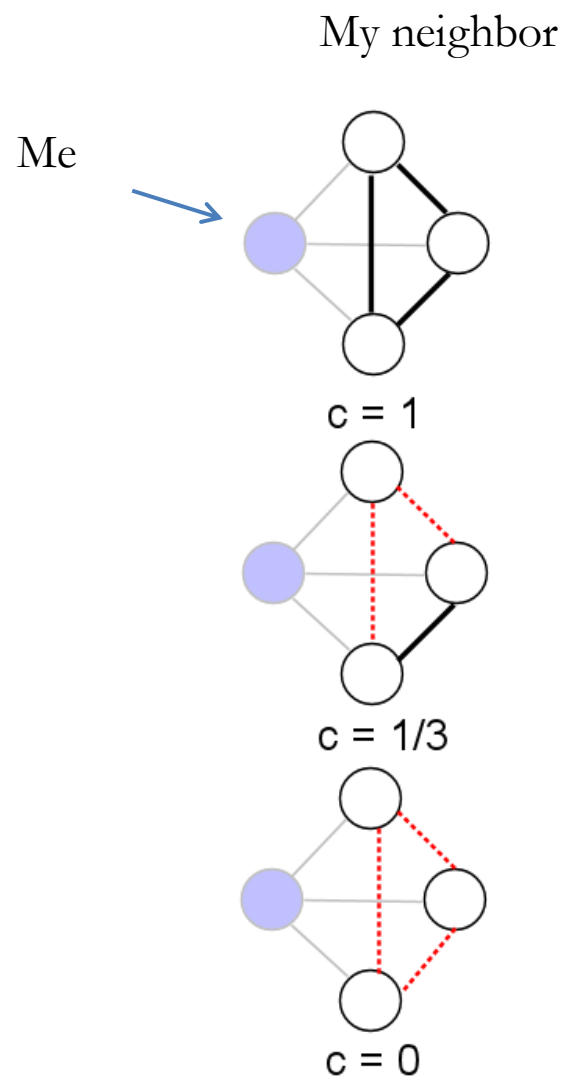
Clustering Coefficients

- As a way to measure how close a node (or vertex) and its neighbors are from being a clique, or a complete graph within a larger graph or network.

- The clustering coefficient of a node is the number of actual connections across the neighbors of a particular node, as a percentage of possible connections.

Actor clustering coefficient

the number of edges connecting the neighbors of the vertex divided by the maximum number of such edges (red dots).



CC of actor i

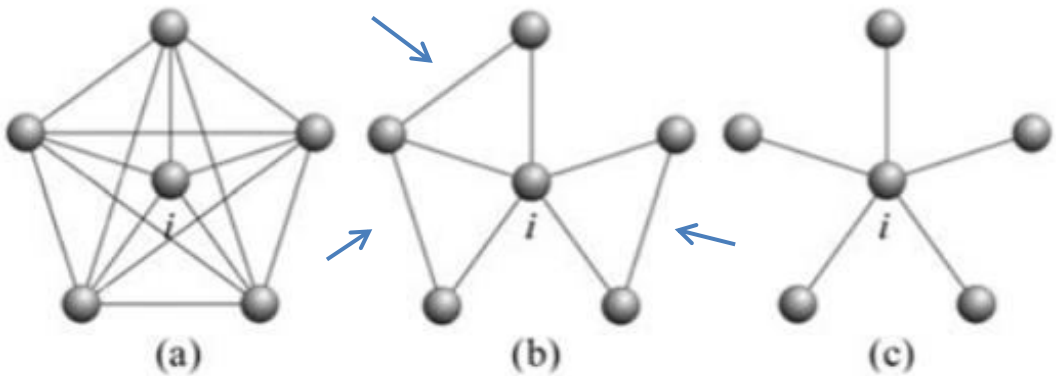


Figure 4 - Example of three networks and respective clustering coefficients (see Eq. (1)). In (a), $cc_i = \frac{10(2)}{5(4)} = 1$ (the vertices around i are fully connected), (b) $cc_i = \frac{3(2)}{5(4)} = 0.3$ and (c) $cc_i = \frac{0(2)}{5(4)} = 0$. The maximum number of edges among the neighbors of i is given by $k_i(k_i - 1)/2$.

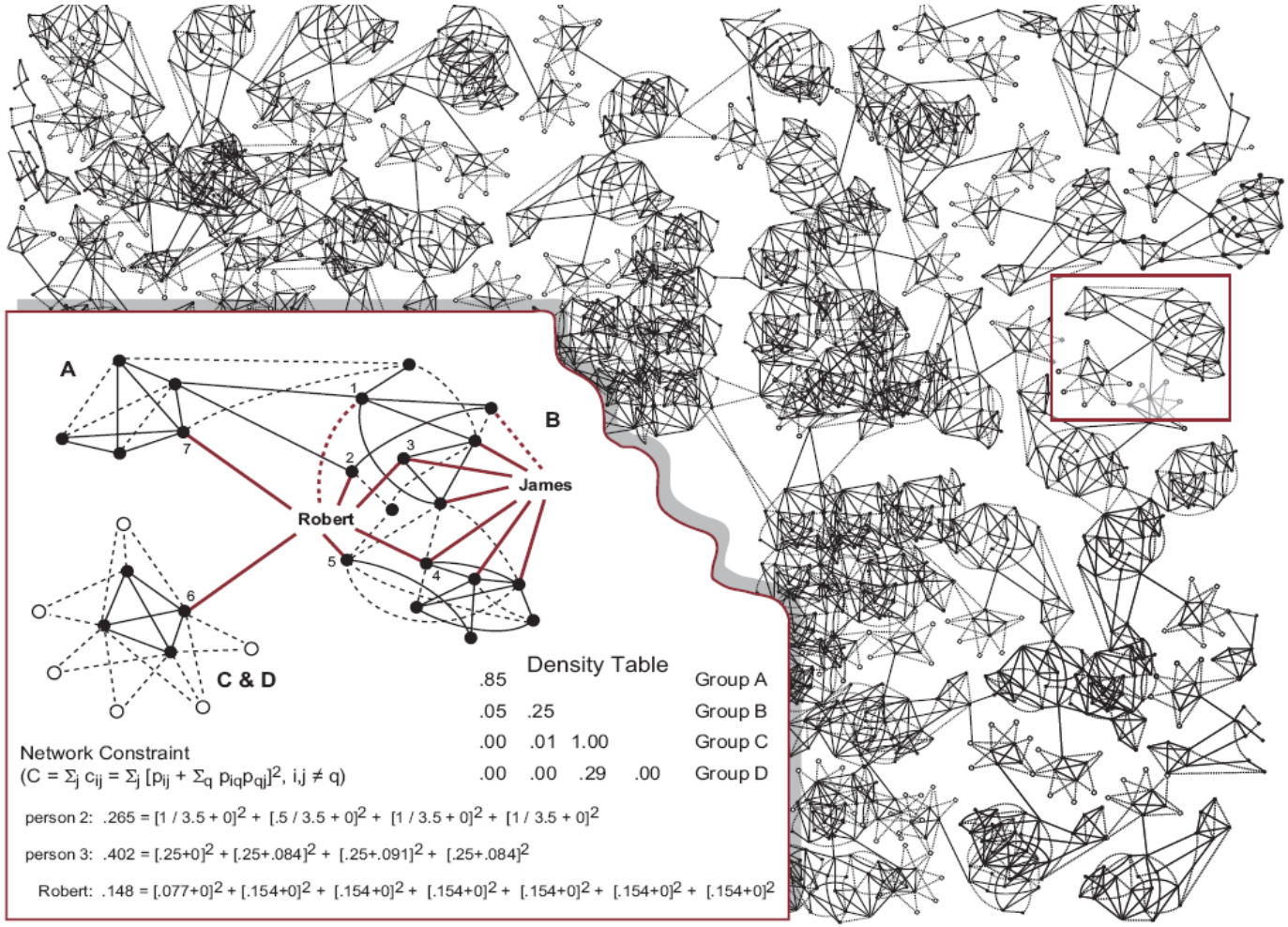


Figure 1. The Small World of Markets and Organizations

Who has a higher clustering coefficient? Robert or James?

Local CF: Average Clustering Coefficients

- This is the formula the clustering coefficient for a group. N=number of nodes. C=clustering coefficient for each node *i*.

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i .$$

Clustering coefficient

- Network>Cohesion>Clustering Coefficient (群聚係數, CC)
- The clustering coefficient of an actor is the density of its open neighborhood.

Clustering coefficient

- To examine the local neighborhood of an actor (that is, all the actors directly connected to ego), and to calculate the density in this neighborhood (leaving out the ego)
- Take the average of the density of the local neighborhoods of all actors (overall approach)
- Weighted approach: actors with larger neighborhoods get more weight in computing the average density

Clustering

- *Network>Cohesion>Clustering Coefficient*

```
Input dataset:                    C:\Program Files\Ucinet 6

Relation: KNOKI
---------------
Overall graph clustering coefficient: 0.607
Weighted Overall graph clustering coefficient: 0.599
```

- Overall: the average of the densities of the neighborhoods of all of the actors.
- Weighted: gives weight to the neighborhood densities proportional to their size.

Weighted clustering coefficient

- Varies from 0 to 1
- A value of 0.65 means that 65% of the triads are closed
- Node level

Node Clustering Coefficients

	1 Clus C	2 nPairs
1	0.667	21.000
2	0.536	28.000
3	0.567	15.000
4	0.733	15.000
5	0.518	28.000
6	0.333	3.000
7	0.514	36.000
8	0.800	15.000
9	0.600	15.000
10	0.800	10.000

- In assessing the degree of clustering, it is usually wise to compare the cluster coefficient to the overall density.
- Why? (strawman)

Expected clustering of a random graph

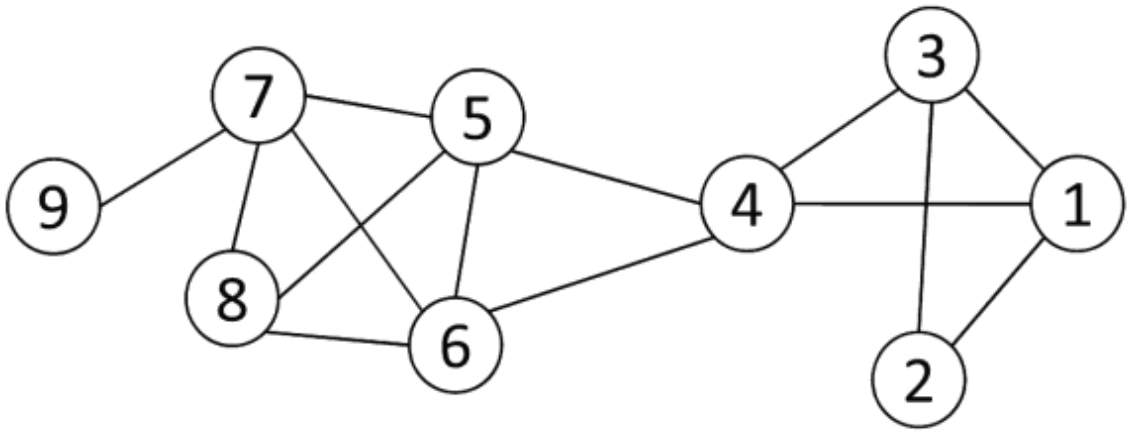
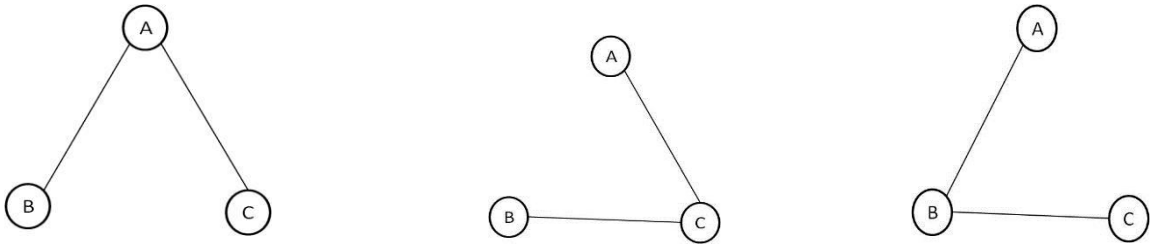


Figure 1.1: A social network of 9 actors and 14 connections. The diameter of the network is 5. The clustering coefficients of nodes 1-9 are: $C_1 = 2/3$, $C_2 = 1$, $C_3 = 2/3$, $C_4 = 1/3$, $C_5 = 2/3$, $C_6 = 2/3$, $C_7 = 1/2$, $C_8 = 1$, $C_9 = 0$. The average clustering coefficient is 0.61 while the expected clustering coefficient of a random graph with 9 nodes and 14 edges is $14/(9 \times 8/2) = 0.19$.

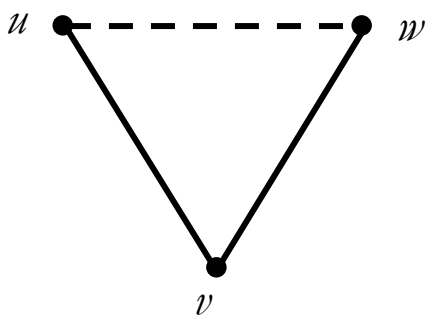
DENSITY

Global clustering coefficient

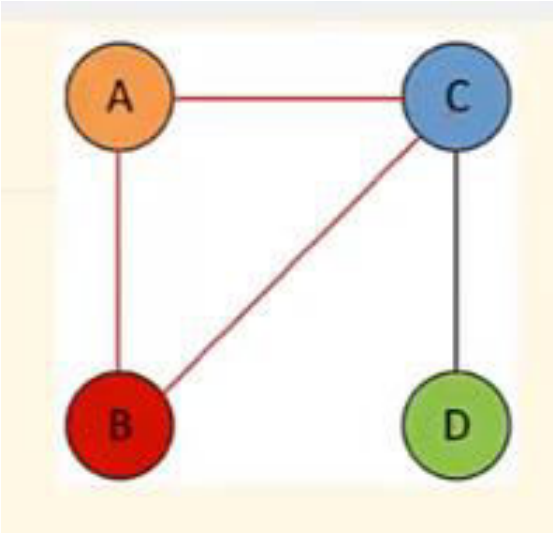
$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}.$$



Triad closure

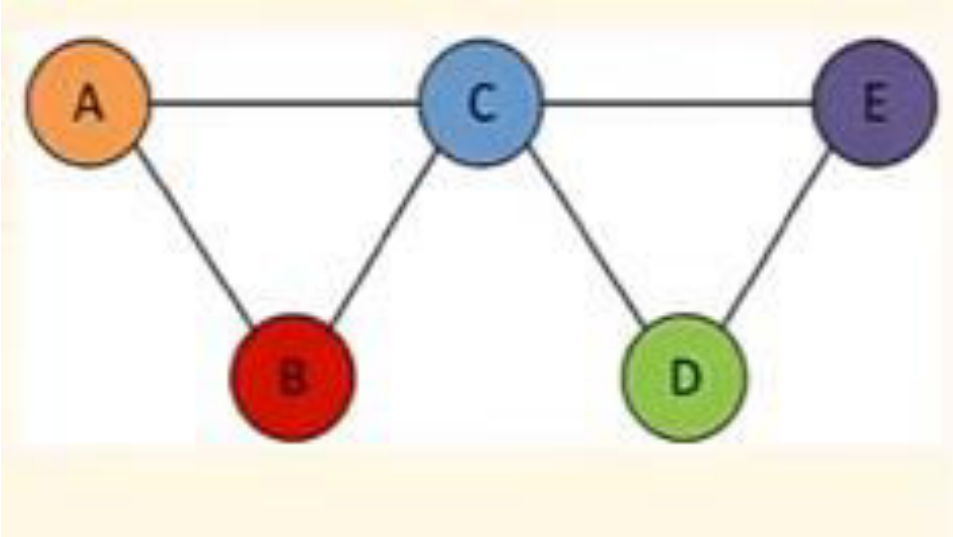


The path *uvw* (solid edges) is said to be closed if the third edge directly from *u* to *w* is present (dashed edge)



How many closed triads?
How many unclosed triads?

Clustering coefficient?



6/10

Ace
Acd
Bcd
Bce

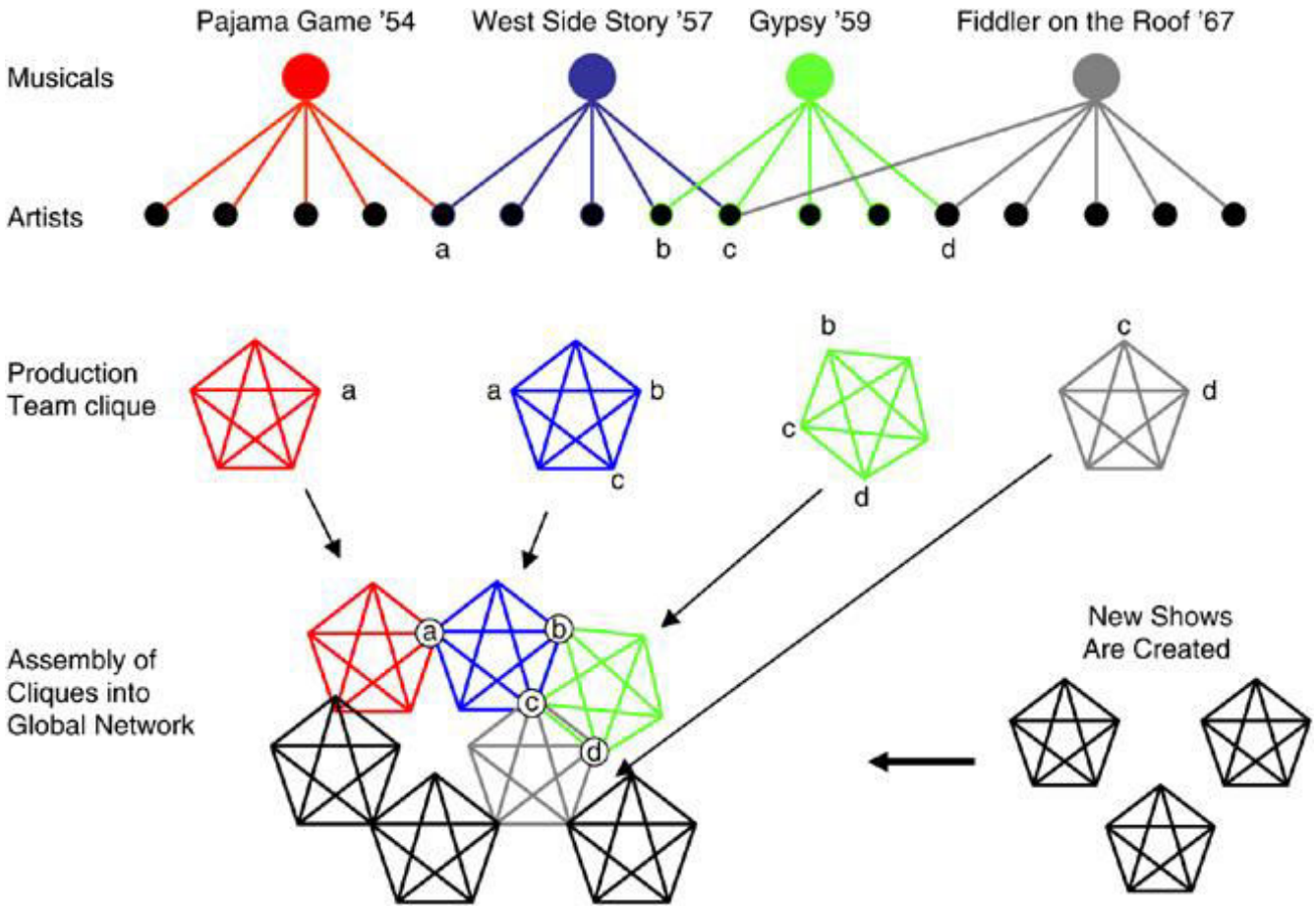
Small world and group creativity

- Good for idea's development and diffusion
  - By engendering trust, cohesive clustering encourages sharing, widespread and lateral communication
- Not so good for the seminal generation of an idea
  - Downside of clustering

To escape this fundamental conundrum:

Cohesive clusters with bridging connections

Real Social Network



Broadway from 1945 to 1989, including shows that were axed before opening night. The final database cataloged 474 musicals and listed 2,092 artists,

Small world quotient

- Short global separation and high local clustering
- Average path length (PL) and high clustering coefficient (CC)
- PL ratio (PL of the actual network/PL of a random graph)
- CC ration (CC of the actual network/CC of the random graph)
- The small world quotient then is

  CC ratio/Pl ratio

Broadway performance: financial success

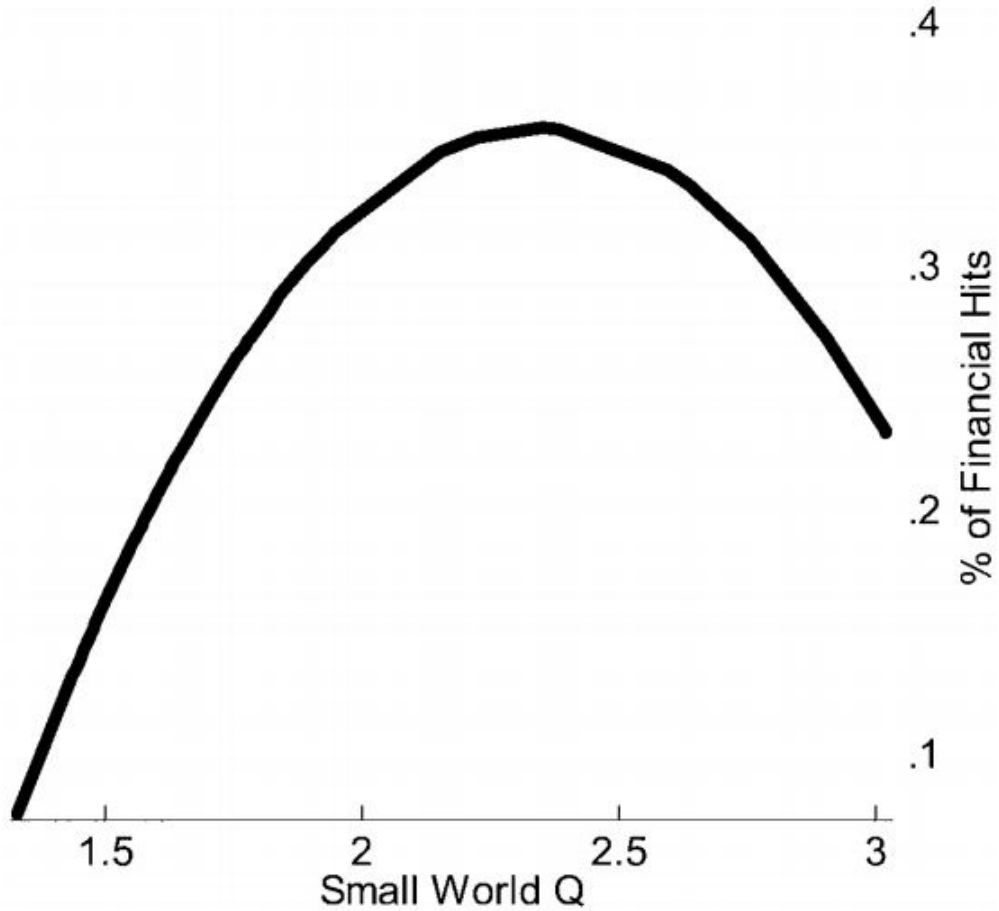


FIG. 6.—Financial success of a season

Broadway performance: artistic success

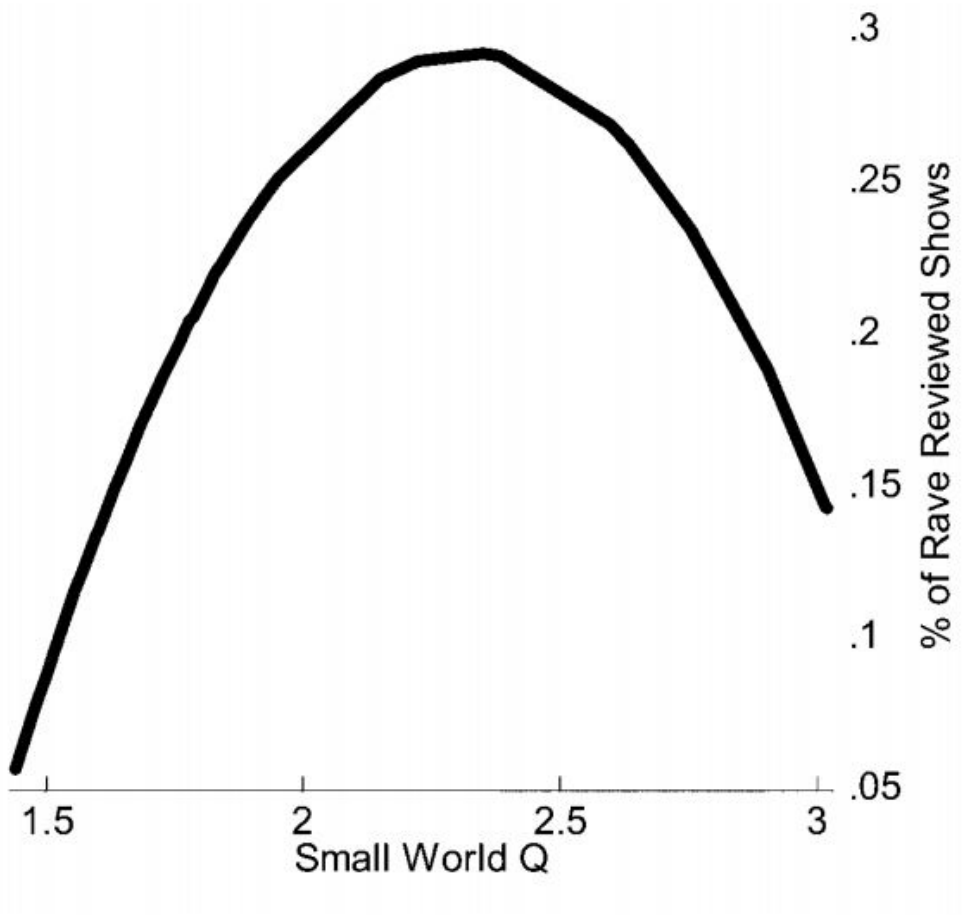


FIG. 7.—Artistic success of a season

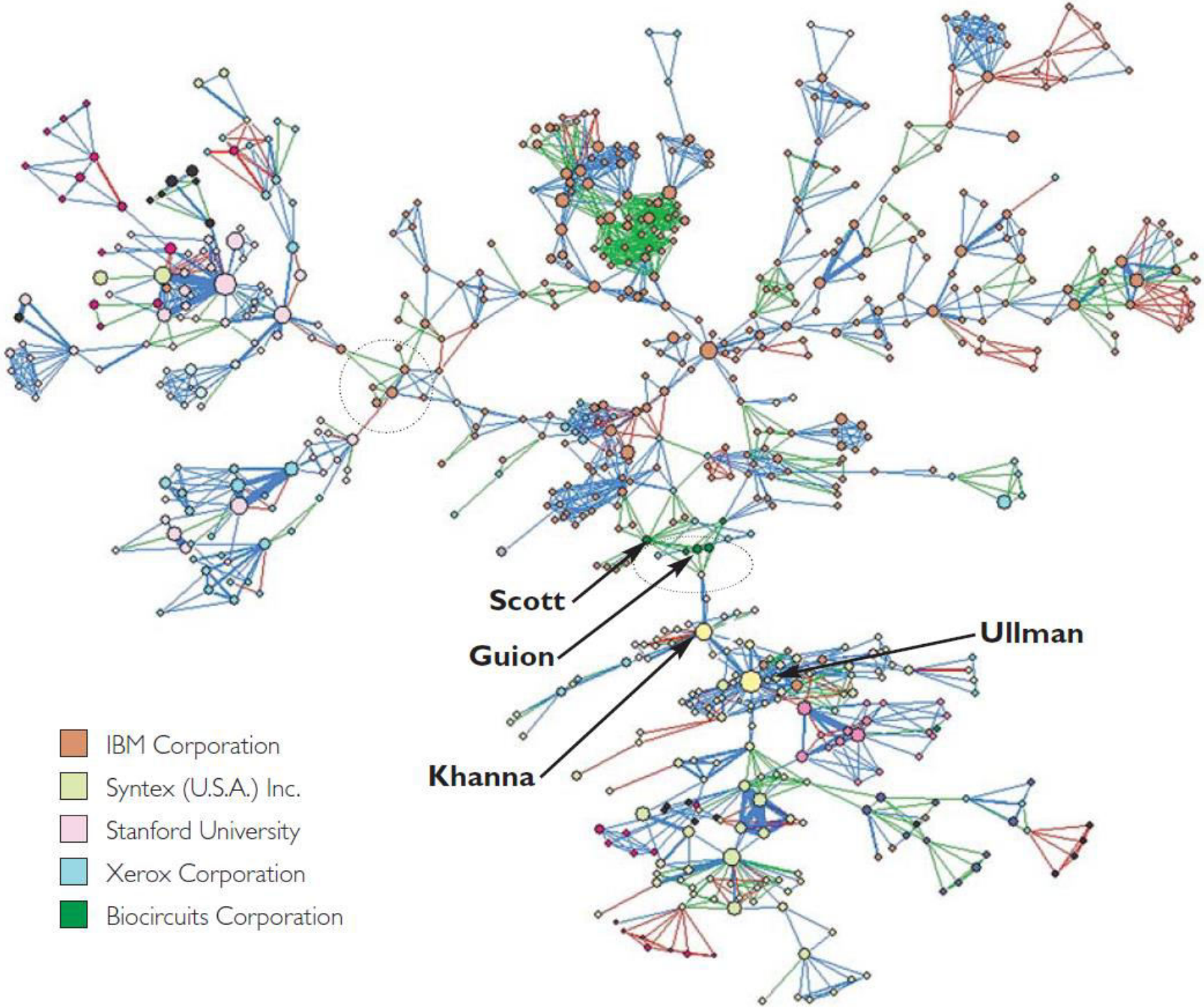


FIGURE 3. Inventors in Silicon Valley's Largest Connected Cluster circa 1986-1990

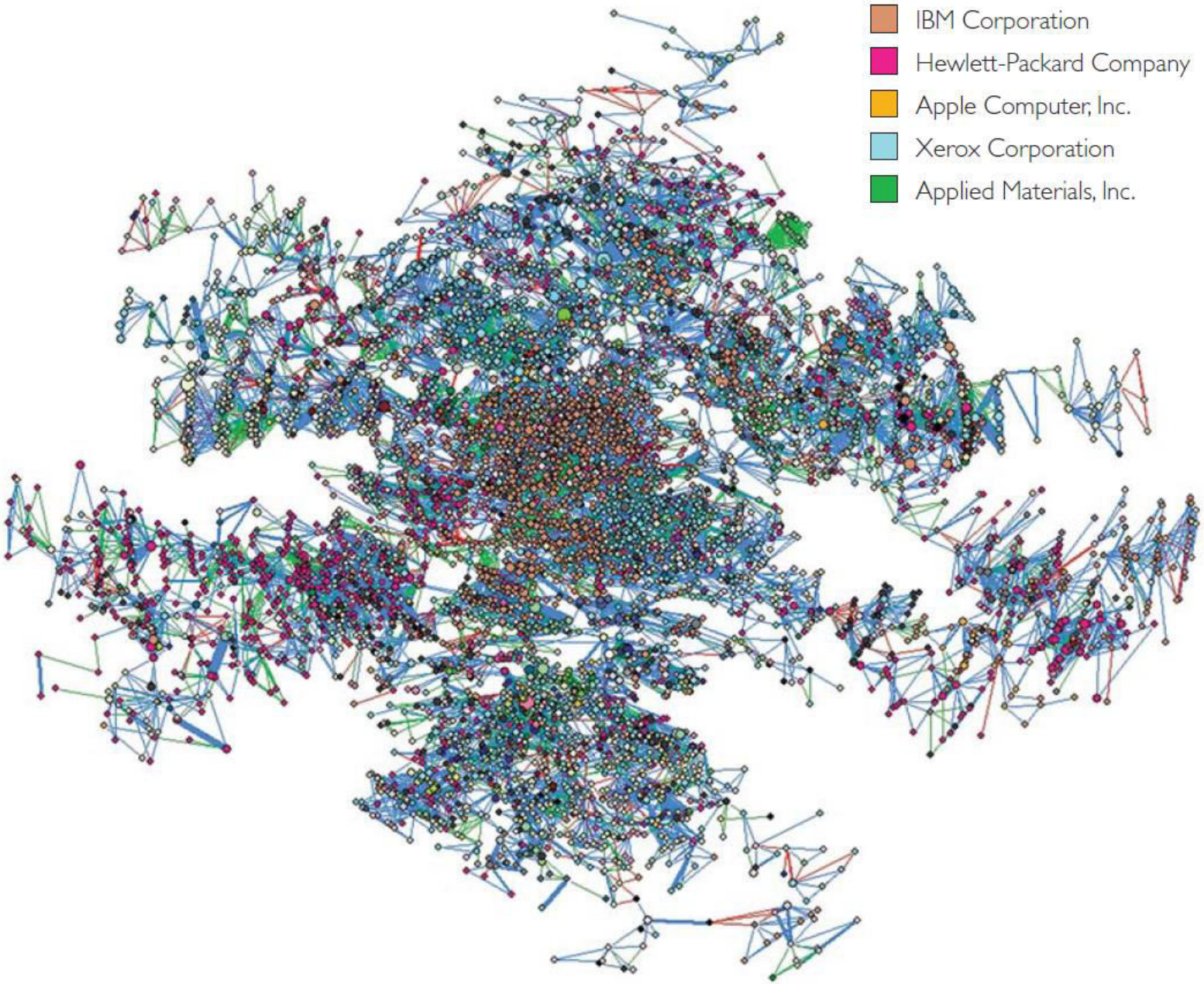


FIGURE 1. The 7,244 Inventors in Silicon Valley's Largest Collaborative Cluster in the Mid-1990s

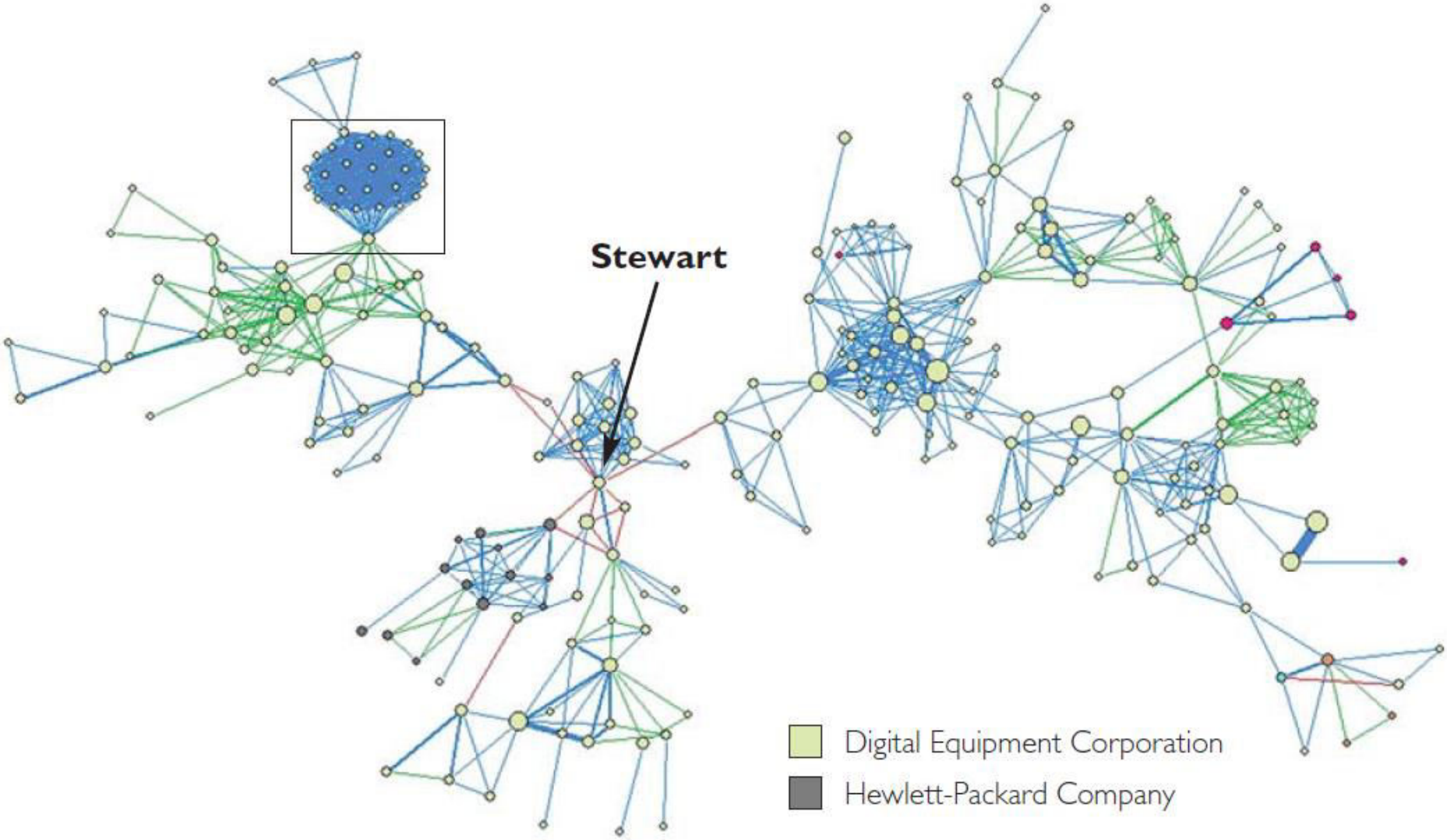


FIGURE 2. Inventors in Boston's Largest Connected Cluster circa 1986-1990

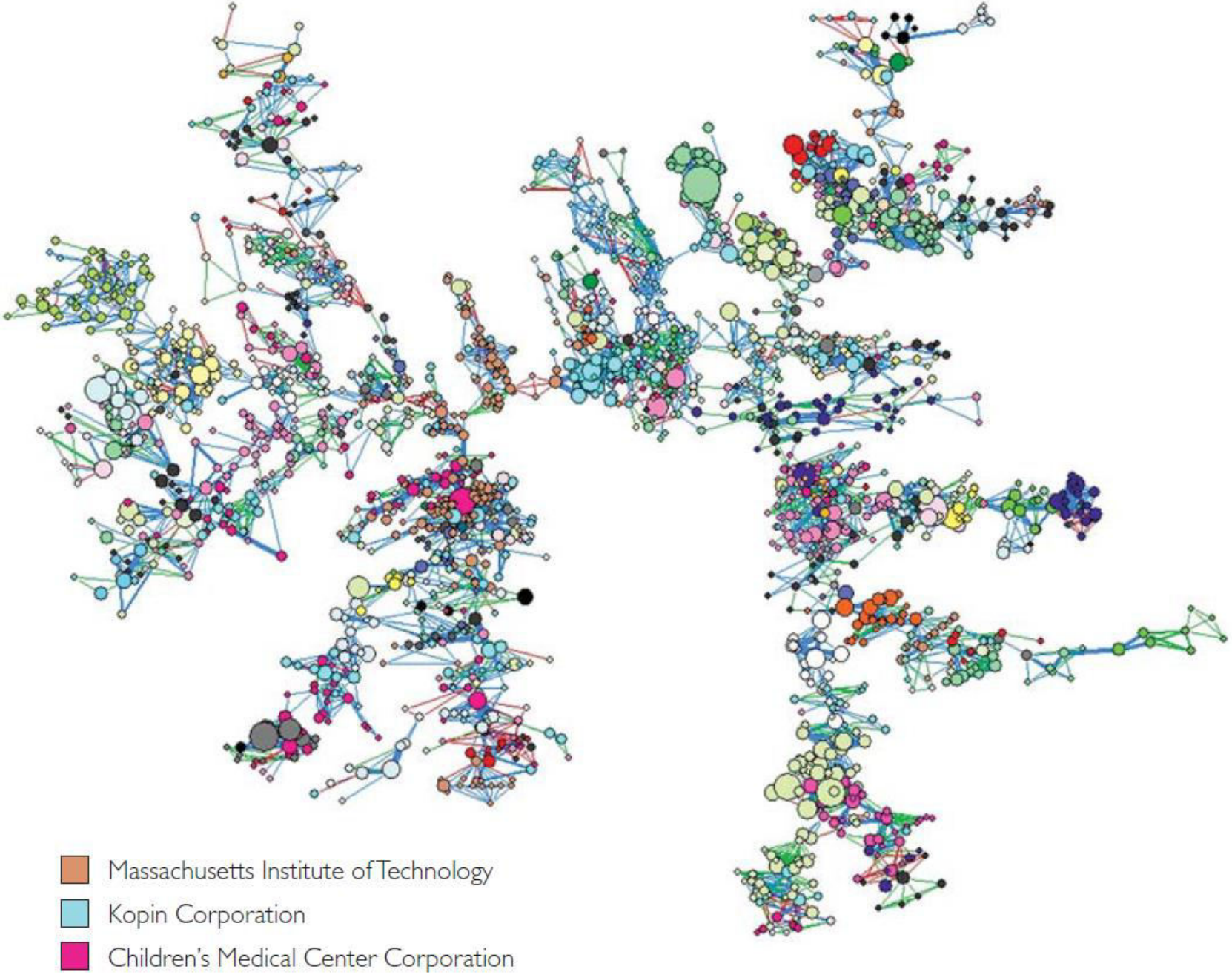


FIGURE 4. Boston's Largest Connected Cluster circa 1989-1993

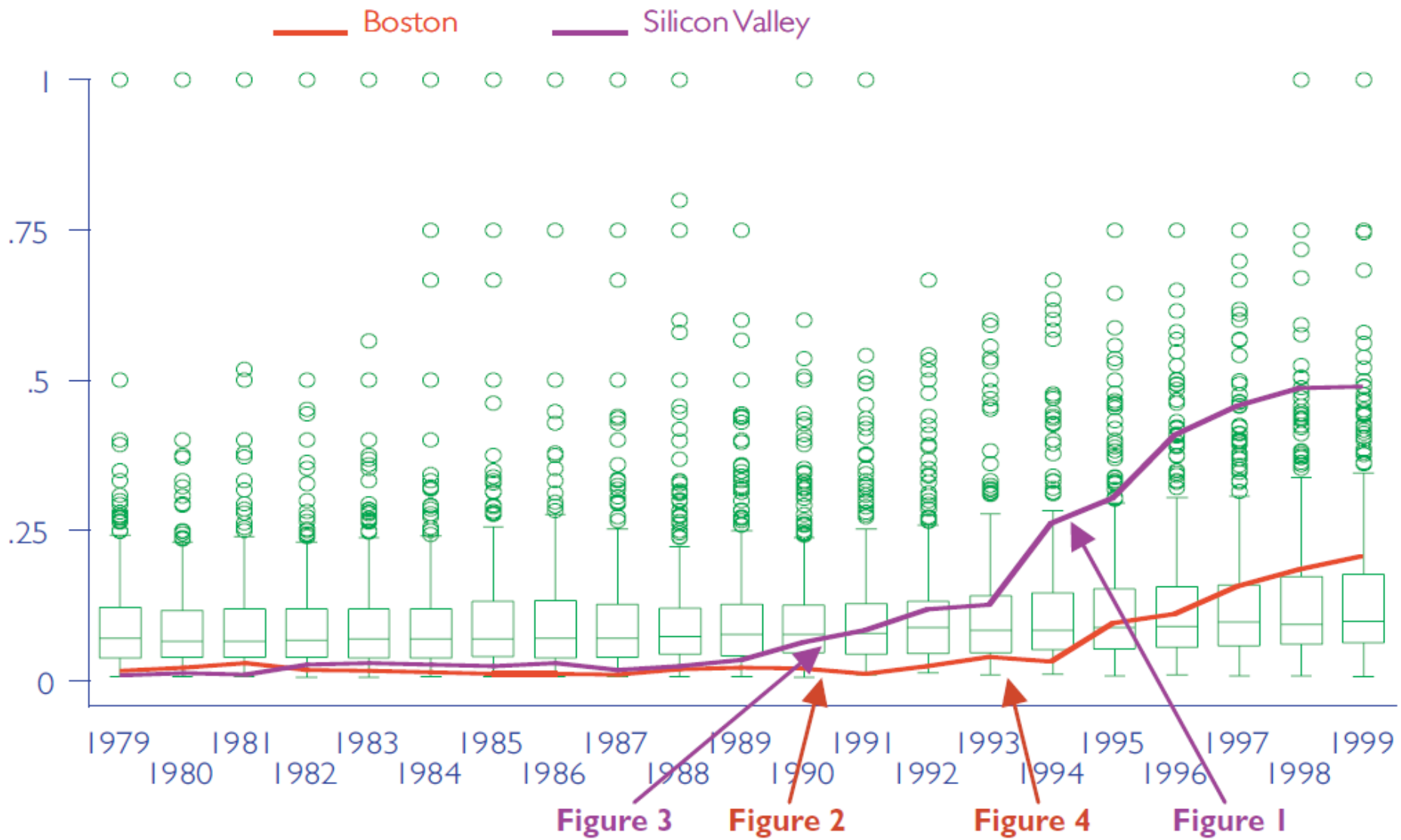
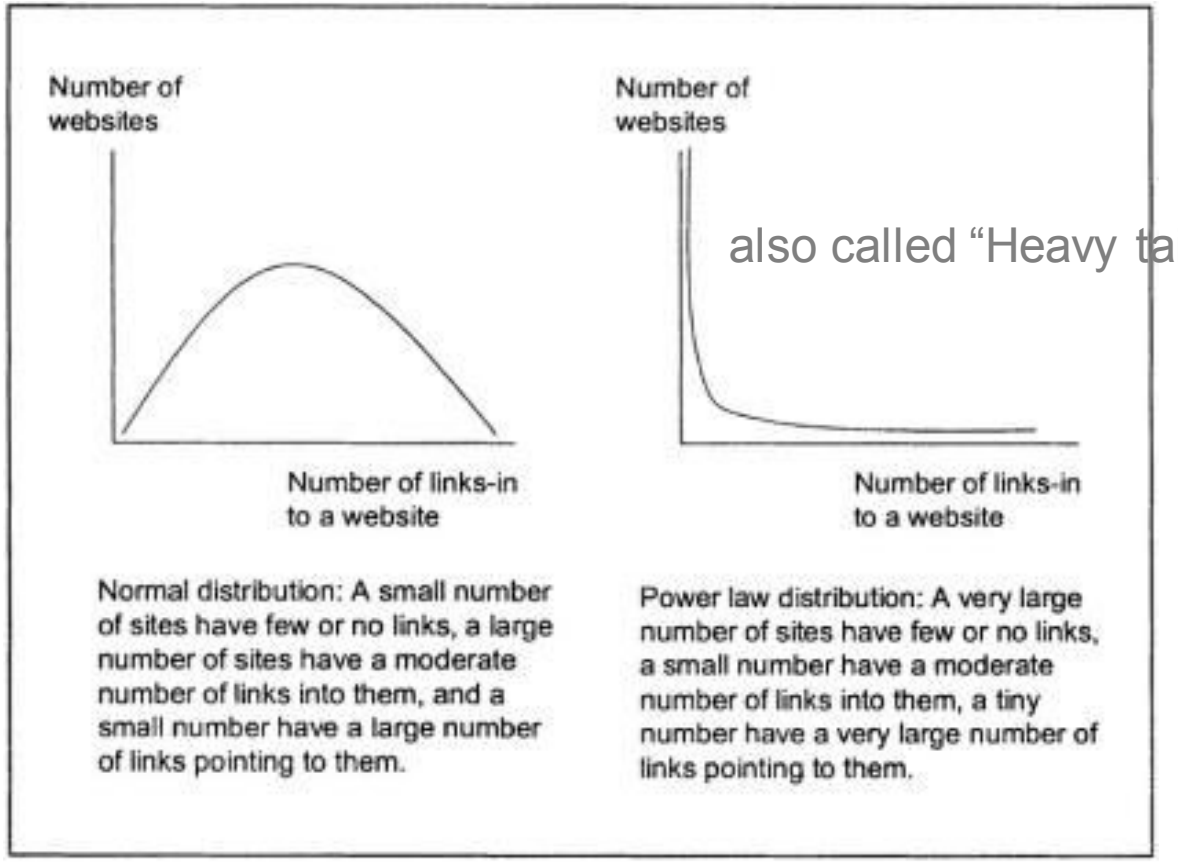


FIGURE 5. Box Plot of Proportion of Inventors within the Largest Collaborative Cluster in each U.S. Metropolitan Statistical Area (MSA)

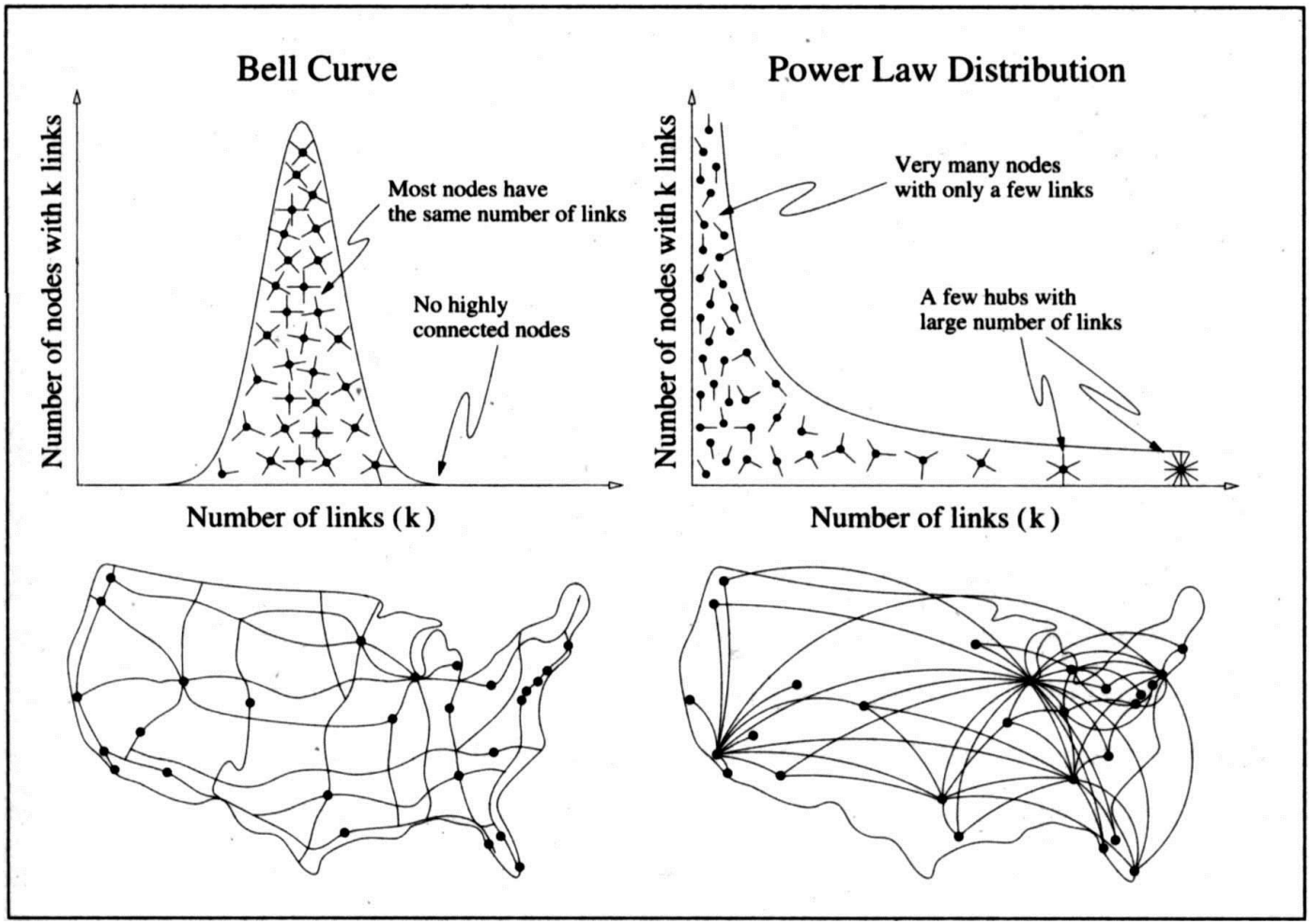
Power law distribution



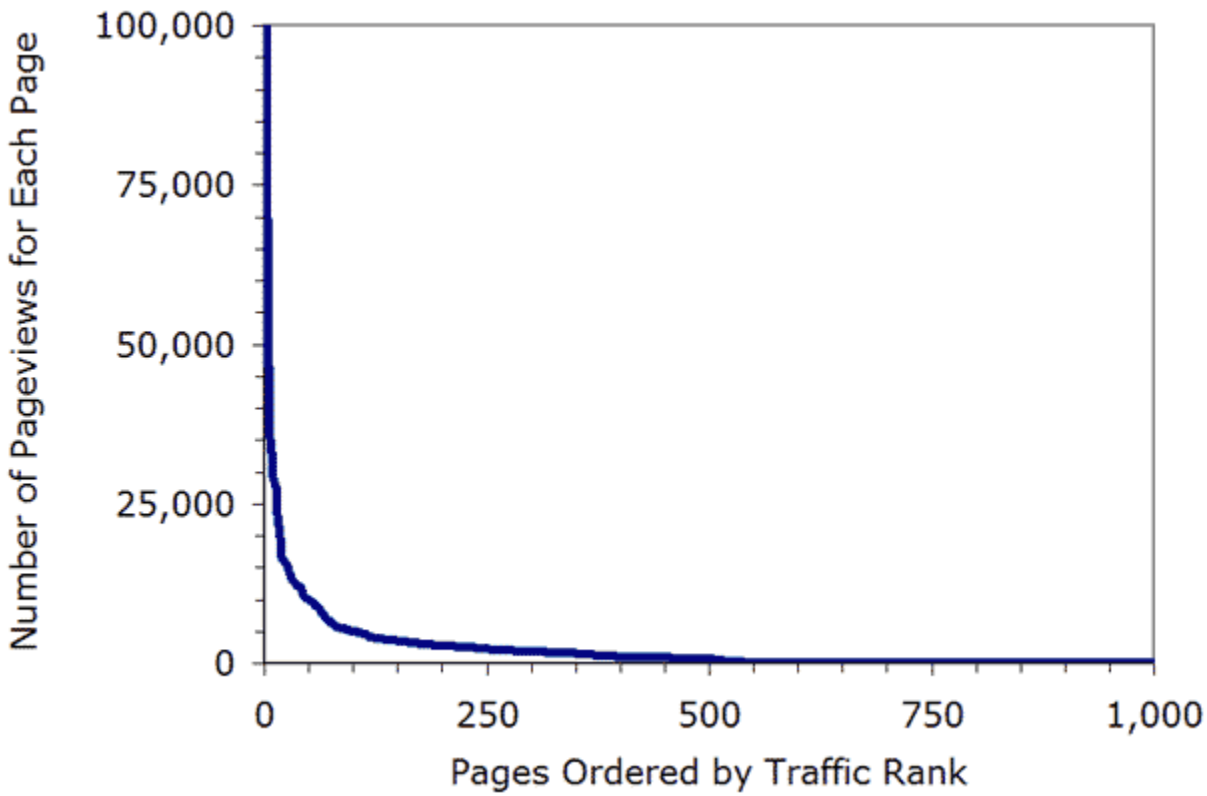
also called “Heavy tailed” distribution

Figure 7.4: Illustration of How Normal Distribution and Power Law Distribution Would Differ in Describing How Many Web Sites Have Few or Many Links Pointing at Them

“Scale-free” distribution

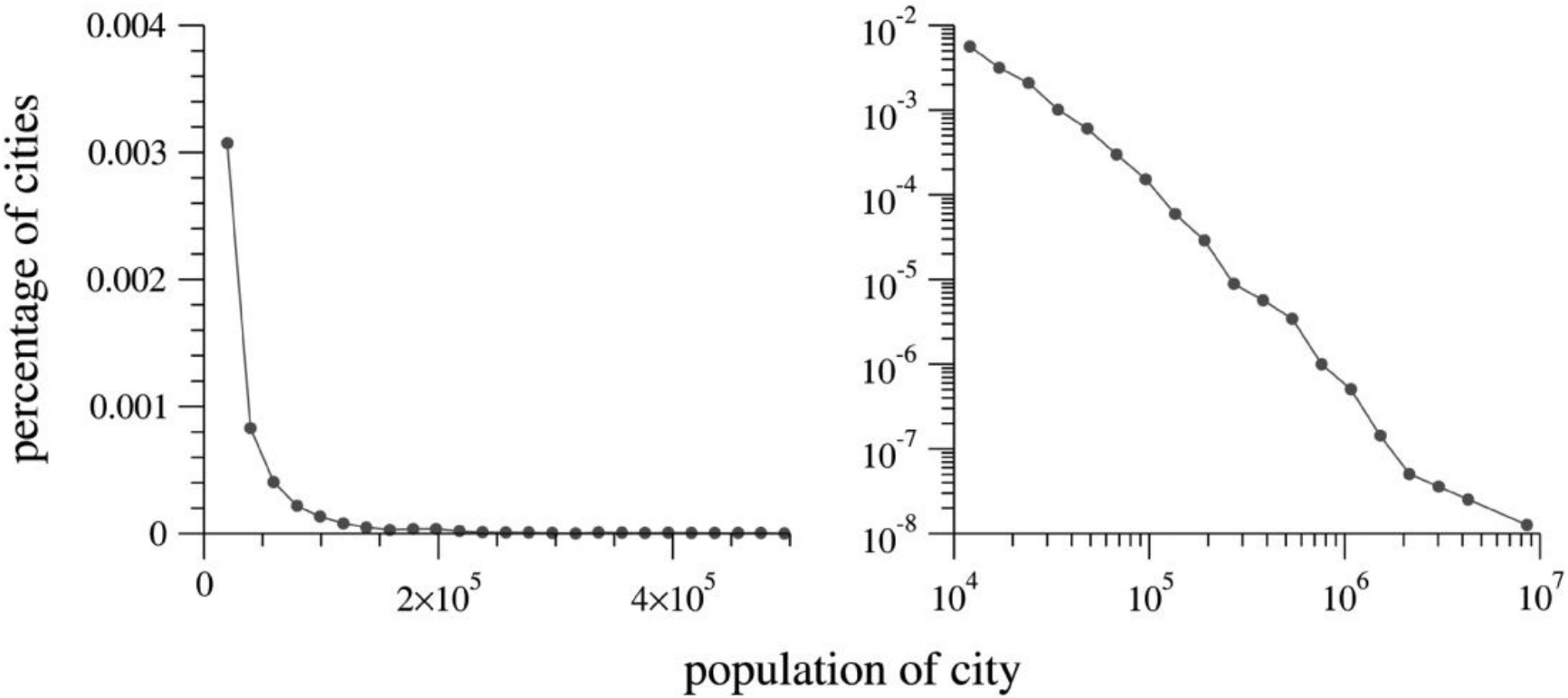


Zipf's law in webpage popularity



Power law distribution

on a log-log scale



The probability of observing X varies inversely as a power of X

$P(x) = ax^{-r}$

an earthquake twice as large is four times as rare.

Power – law distribution

for evolving self-organized networks

proposed by Barabasi and collaborators

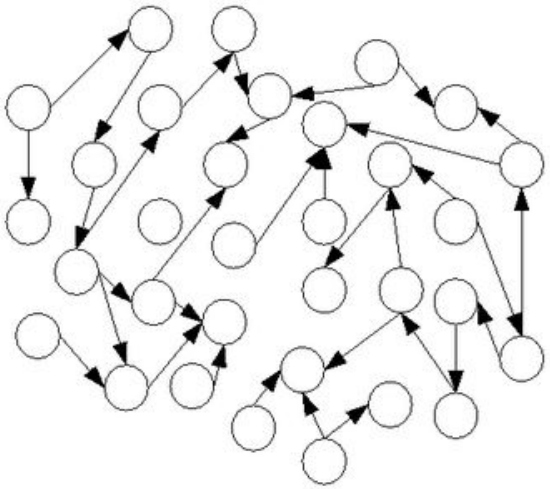
$$P(k) \propto k^{-\gamma}$$

Typical range $2 < \gamma < 3$

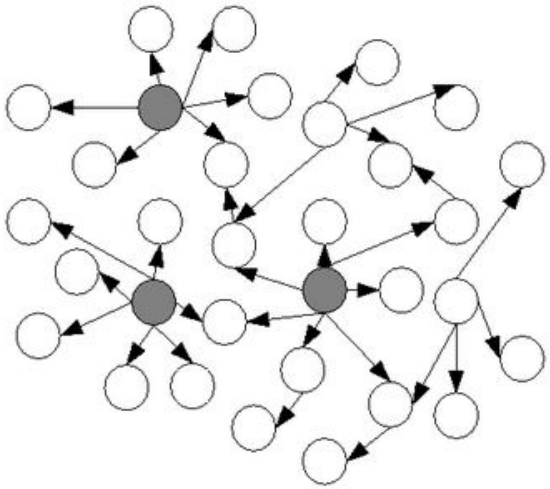
Probability of a node having K link is the power of K
Nodes twice this size is four times as rare

Random vs. Real Social networks

- Random network models introduce
  - A fixed number of nodes
  - an edge between any pair of vertices with a uniform probability *p*
- Real networks are not exactly like these
  - Nodes added to the network with time, it's evolving, instead of static
  - Tend to have a relatively few nodes of high connectivity (the "Hub" nodes)



(a) Random network



(b) Scale-free network

Scale-free networks

- A small number of nodes act has "highly connected hubs" (high degree), but most nodes have a low degree
  - No single node can be seen as "typical", no intrinsic "scale"
- The scale-free model has a shorter average path length than a random
- Thanks to the hub nodes

The rich get richer

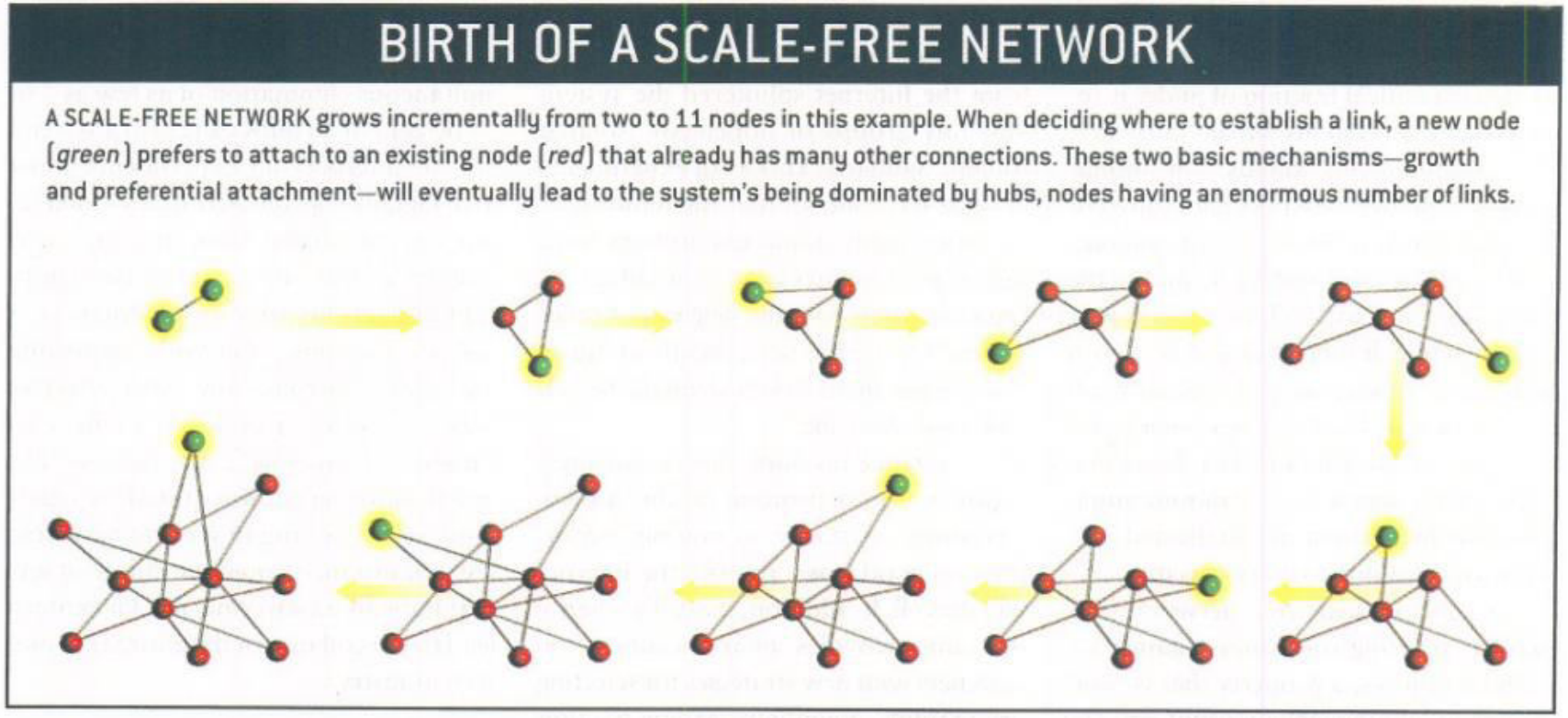
Power-law distribution of node-degree arises if

(but *not* "only if")

- As Number of nodes grow edges are added in proportion to the number of edges a node already has.
- Alternative: Copy model—where the new node copies a random subset of the links of an existing node
  - Sort of close to the WEB reality
- *Preferential attachment*
  - *More incentive to connect with those who own more links*

Preferential attachment (with network effect)

- Network dynamics for evolving self-organized networks



Rich gets richer

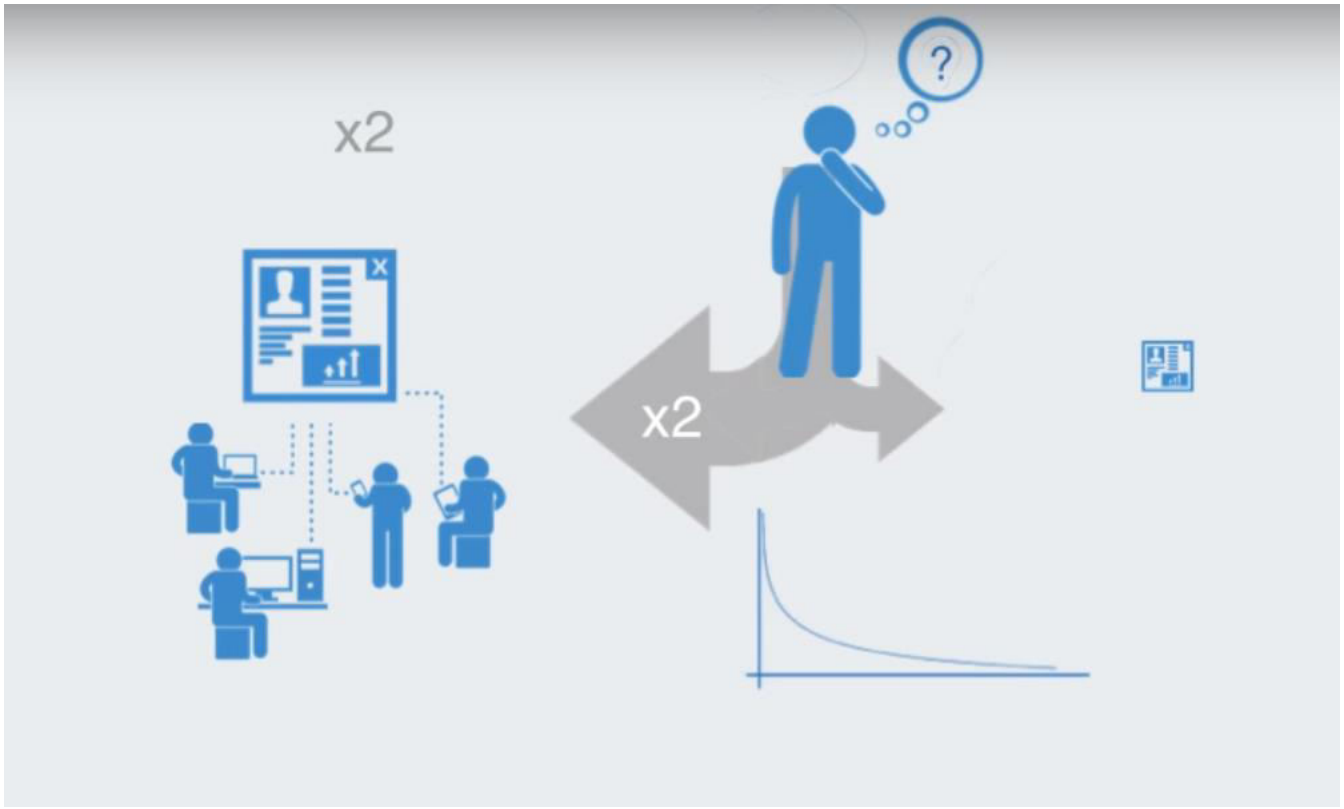
Preferential attachment/rich get richer

- that the more connected a node is, the more likely it is to receive new links. Nodes with higher degree have stronger ability to grab links added to the network
- New nodes are added to the network one at a time. Each new node is connected to *m* existing nodes with a probability that is proportional to the number of links that the existing nodes already have.
- Formally, the probability p_i that the new node is connected to node *i* is

$$p_i = \frac{k_i}{\sum_j k_j},$$

grow edges are added in proportion to the number of edges a node already has.

Preferential attachment (incentive)



Informative introduction to network dynamics

Matthew effect

- "the rich get richer and the poor get poorer"
  - For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath.

    Matthew 25:29

SNA

Weak tie, Bridge, and Embededness

The Strength of Weak Ties

- Granovetter's "The Strength of Weak Ties" (considered one of the most important sociology papers written in recent decades) argued that "weak ties" could actually be more advantageous in politics or in seeking employment than "strong ties", because weak ties allowed an individual to reach a higher number of other individuals.

Granovetter 1973

- Interviewed MBA graduates and asked: "How did you find your job?"
  - Kept getting the same answer:
    "through an acquaintance or "friend's friend", not a friend"
- Implications
  - Weak ties are surprisingly valuable because they are more likely to lean novel information and ideas

Strength of a tie and influence

- M. S. Granovetter: *The Strength of Weak Ties*, AJS, 1973:

- finding a job through a contact that one saw
  - frequently (2+ times/week) 16.7%
  - occasionally (more than once a year but < 2x week) 55.6%
  - rarely 27.8%
- but… length of path is short
  - contact directly works for/is the employer
  - or is connected directly to employer

The Strength of Weak Ties

- Granovetter observed that the presence of weak ties often reduced path lengths (distance) between any two individuals—which led to quicker diffusion of information.

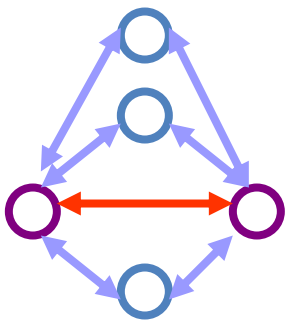
Granovetter's theory of weak tie

- Strong ties are embedded in high homophilous clusters
- Weak ties connect to diversity
- Weak ties a source of novel information

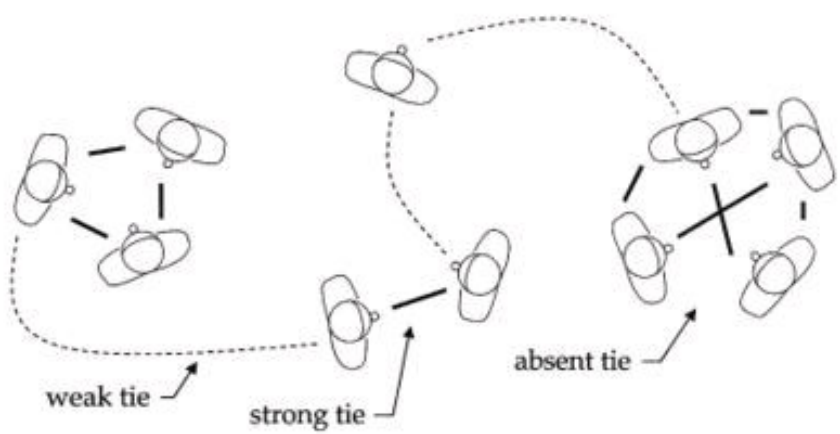
Adopted from Borgatti, 2004

Strong tie defined

- A strong tie
  - frequent contact
  - affinity
  - many mutual contacts
    - embeddness, similar to node clustering coefficient, but now we consider edge (i.e. pair of nodes))
- Less likely to be a bridge (or a local bridge)



Source: Granovetter, M. (1973). "



- Is it always true?
- Can you think of exceptions?
- The strength of strong tie?

Granovetter's theory of

- Strong ties results in triad closure
- Ties that are part of closed triples cannot be bridges or local bridges
- Therefore, only weak ties can be bridges

Triads

- A basic unit of social network analysis is the dyad (a pair of parties who may or may not share a social relation)
- A triad is a set of three parties, which also consists of three dyads.
- Sociologists pointed out that three actors in a triad may allow qualitatively different social dynamics that cannot be reduced to individuals or dyads.

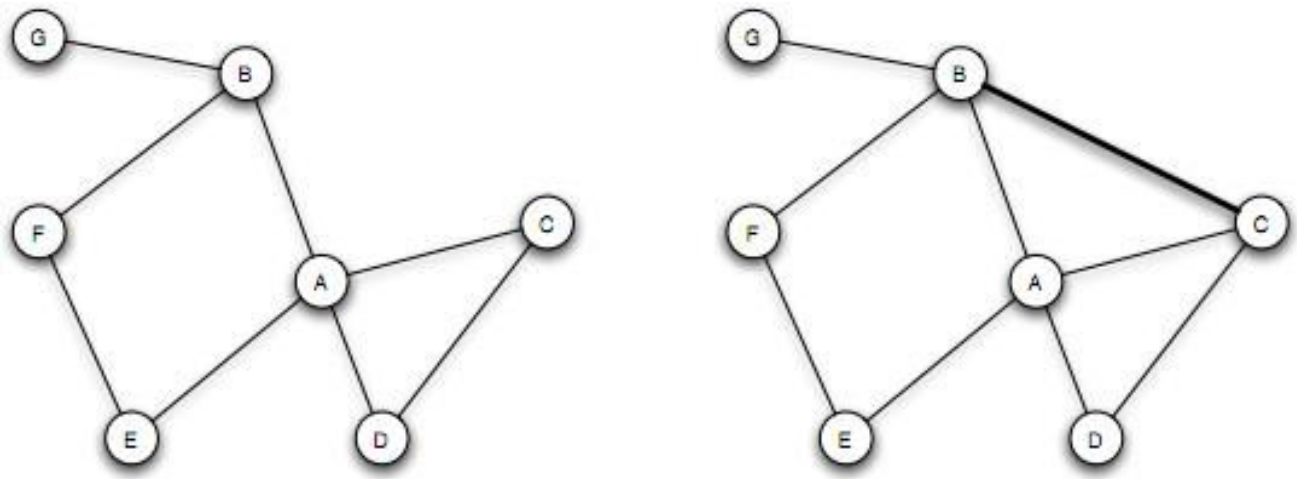
Transitivity

- A triad is transitive if when there is a tie from A to B and from B to C, THEN there is also a tie from A to C.

Triadic closure

- If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future.
- Clustering/transitivity

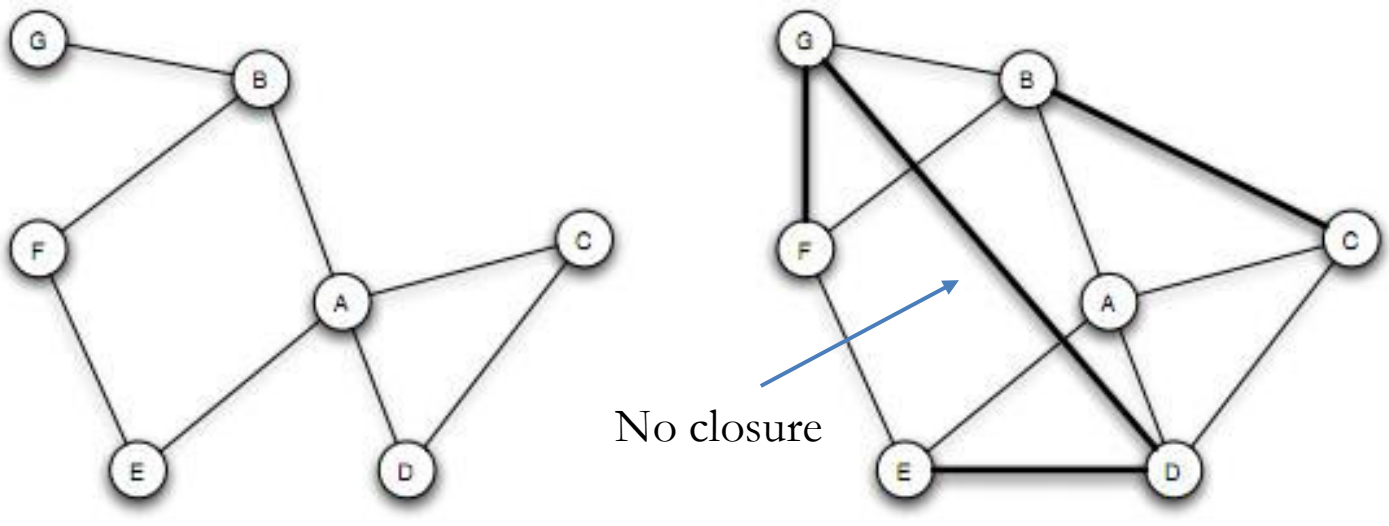
Triadic closure



(a) *Before B-C edge forms.* (b) *After B-C edge forms.*

The formation of the edge between B and C illustrate the effects of triadic closure, since they have a common neighbor, A.

Strong and weak ties



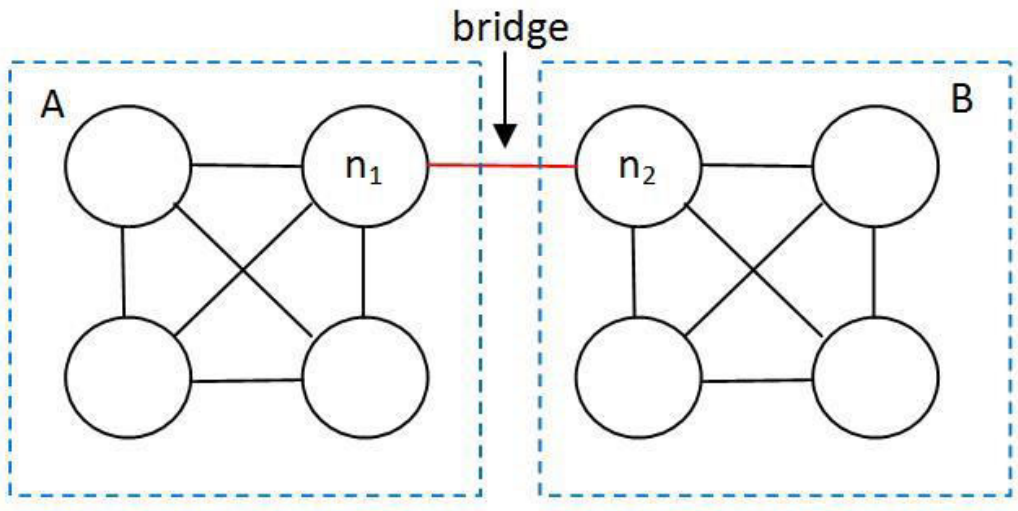
(a) *Before new edges form.* (b) *After new edges form.*

If we watch a network for a longer span of time, we can see multiple edges forming; Some form through triadic closure while others (such as ?) form even though the two nodes have no neighbors in common

Reasons for Triadic Closure

- Opportunity
- Trusting
- Incentive

Bridge

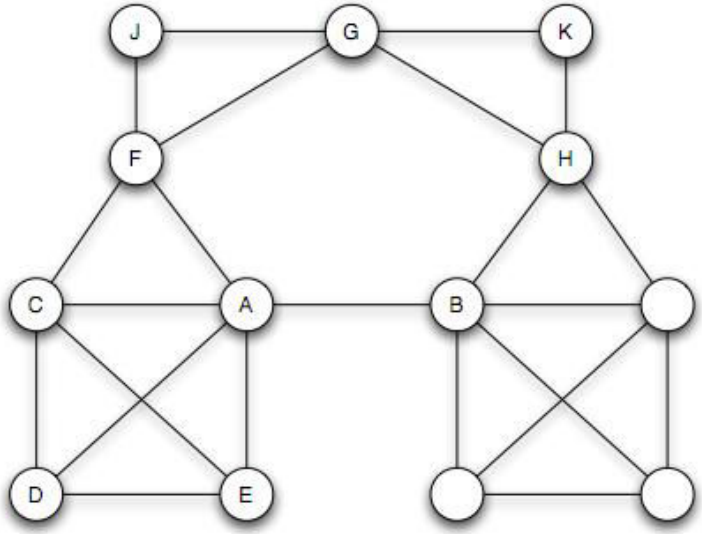


An edge that joins two nodes n1 and n2 in a graph is called a bridge if deleting The edge would cause n1 and n2 in two different components.

Bridges

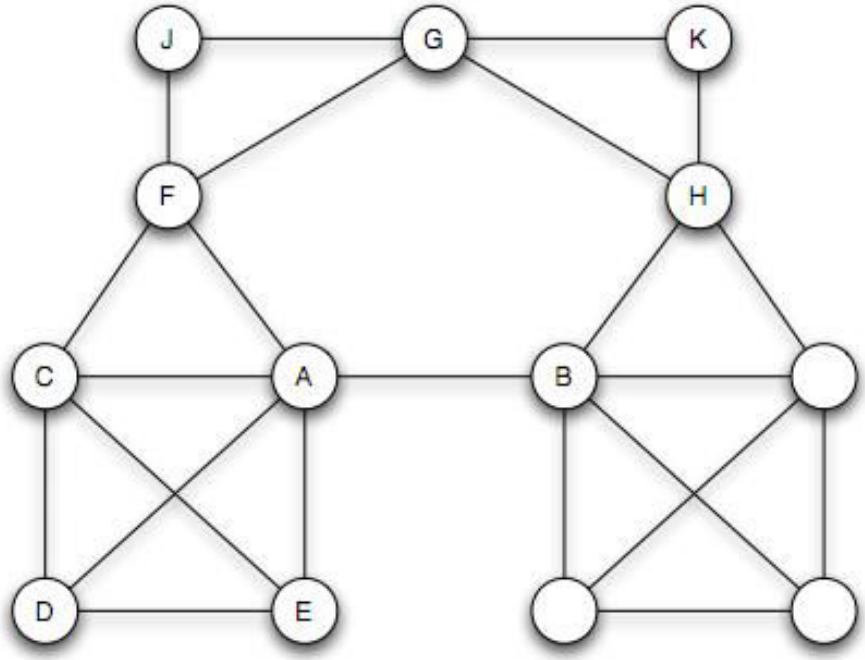
- Bridges allow diffusion of information between otherwise disconnected communities.
- Local bridges bring otherwise distant communities together
- The breadth of information spread depends heavily on the number and connectedness of the bridges available to the originators of the information.
- Author Malcolm Gladwell characterizes people that habitually act as bridges as Connectors in his book The Tipping Point.

Local bridge



1. A-B edge is not the only path that connects its two endpoints; though they may not realize it.
2. An edge joining two nodes A and B in a graph is a local bridge if its endpoint A and B have no friends in common – in other words, deleting the edge would increase the distance between A and B to a value strictly more than 2.

The "span" of a local bridge: : a matter of efficiency

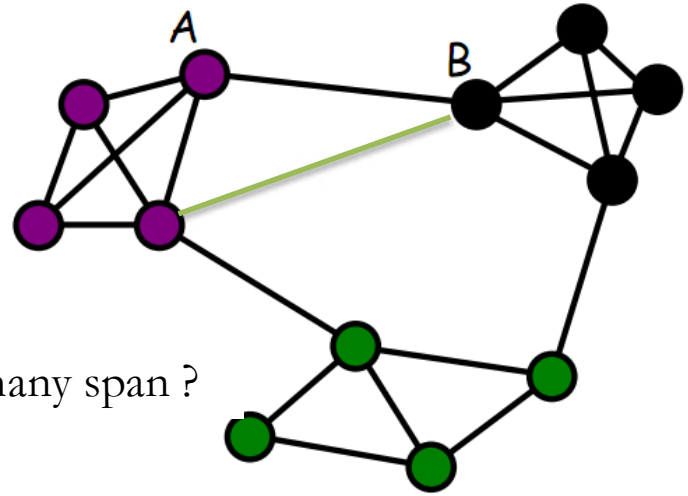


The "span" of a local bridge is the distance its endpoints would be from each other if the edge were deleted.

The A-B edge is a local bridge of span 4, since the removal of this edge would increase the distance between A and B to 4 .

Local Bridge of Span/Degree K

- A edge/tie that connect nodes that would otherwise be at least *k* steps apart



How many span ?

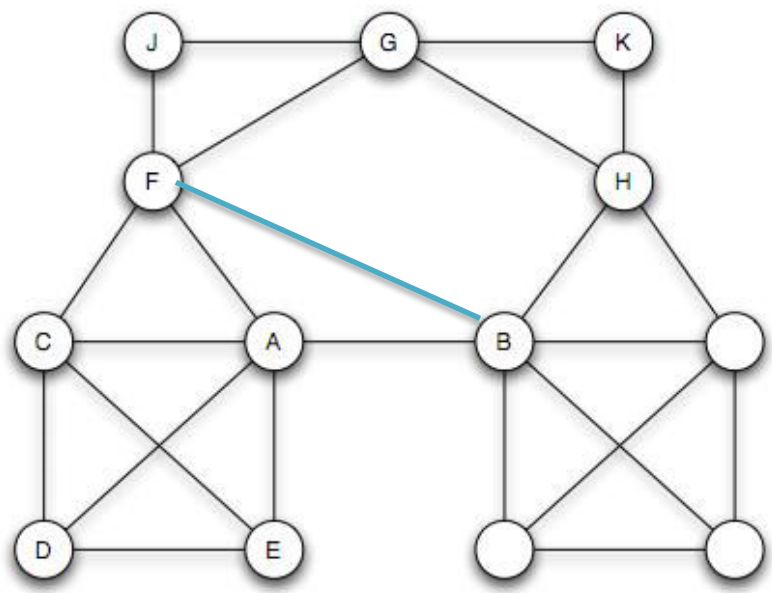
"Bridgeness": Neighborhood overlap measure

- Relax the strict definition of bridge
  - Some edges are "almost" local bridges

# of nodes who are neighbor of both A and F divided by

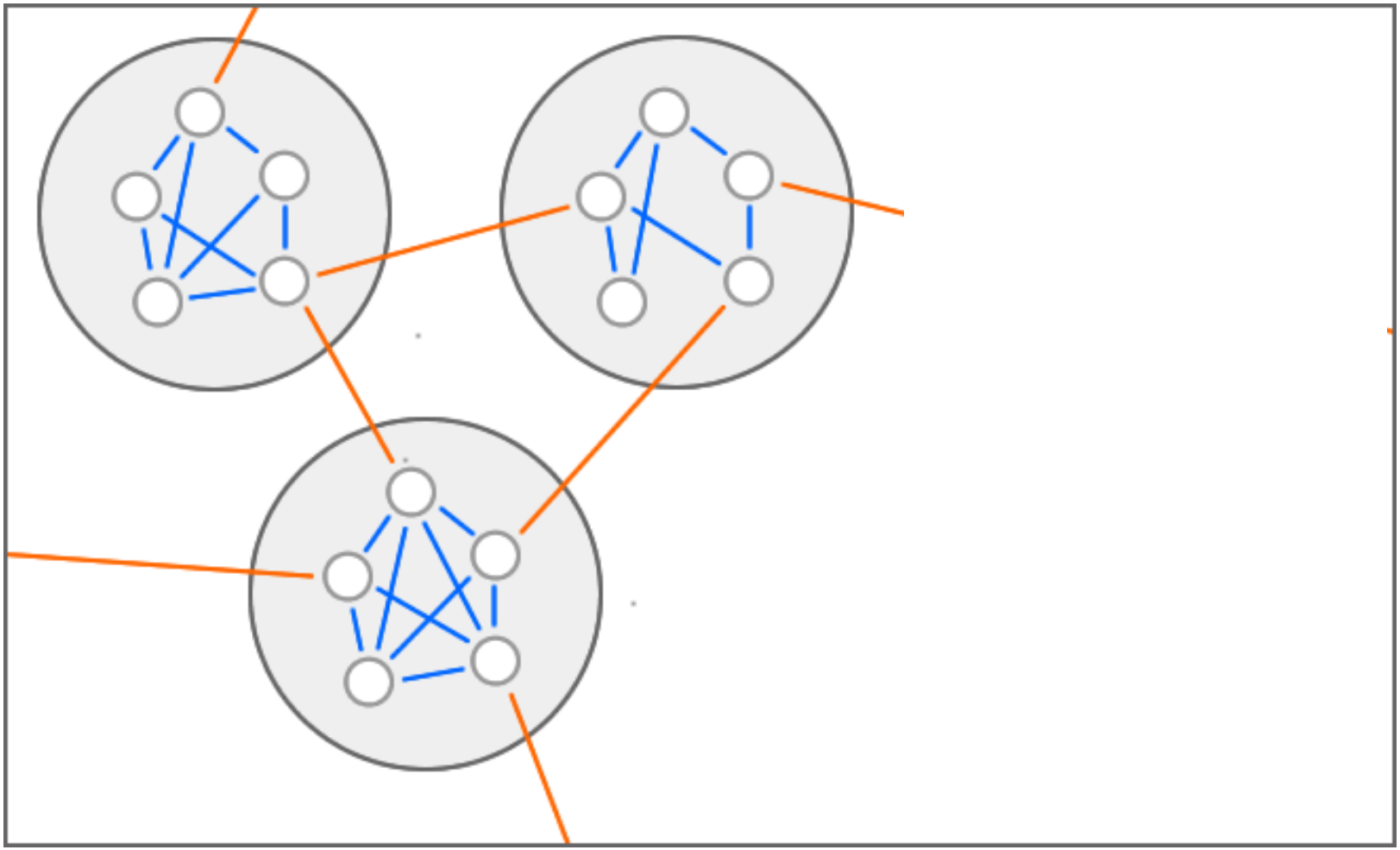
# of nodes who are neighbors of at least on of A or F

$$\frac{\text{number of nodes who are neighbors of } both\ A \text{ and } B}{\text{number of nodes who are neighbors of } at\ least\ one\ of\ A \text{ or } B'}$$



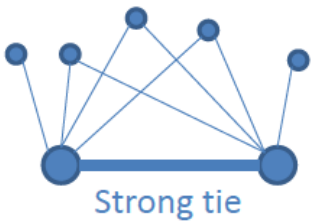
Compare Bridgeness of
Edge AB
And
Edge AF

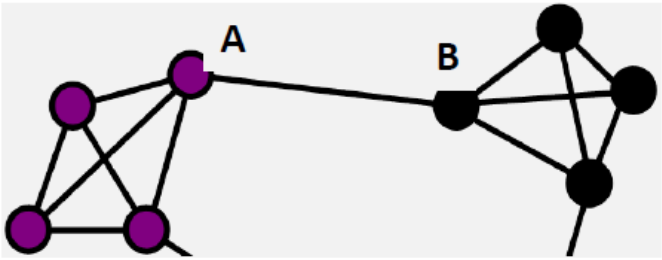
Why connection between groups tend to be weak ties?

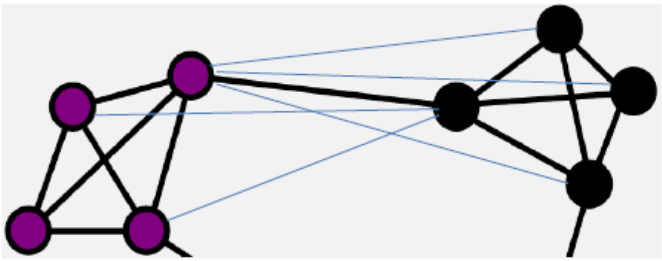


Intuitively

- Supposed A – B is a strong tie, they are likely to have many acquaintances (weak ties) in common, which prevents it to become a bridge

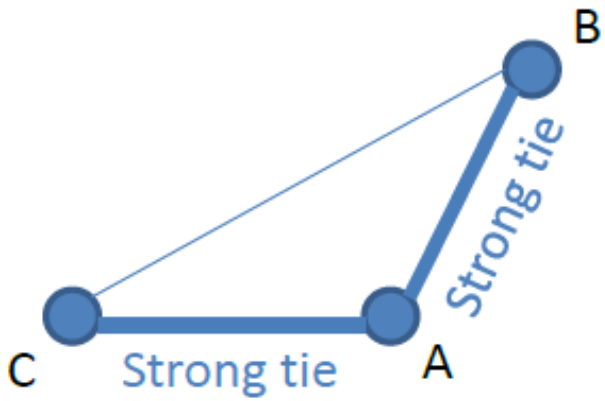




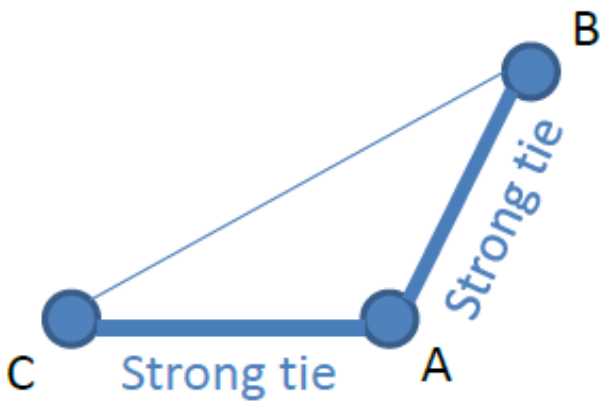


The strong triadic closure property

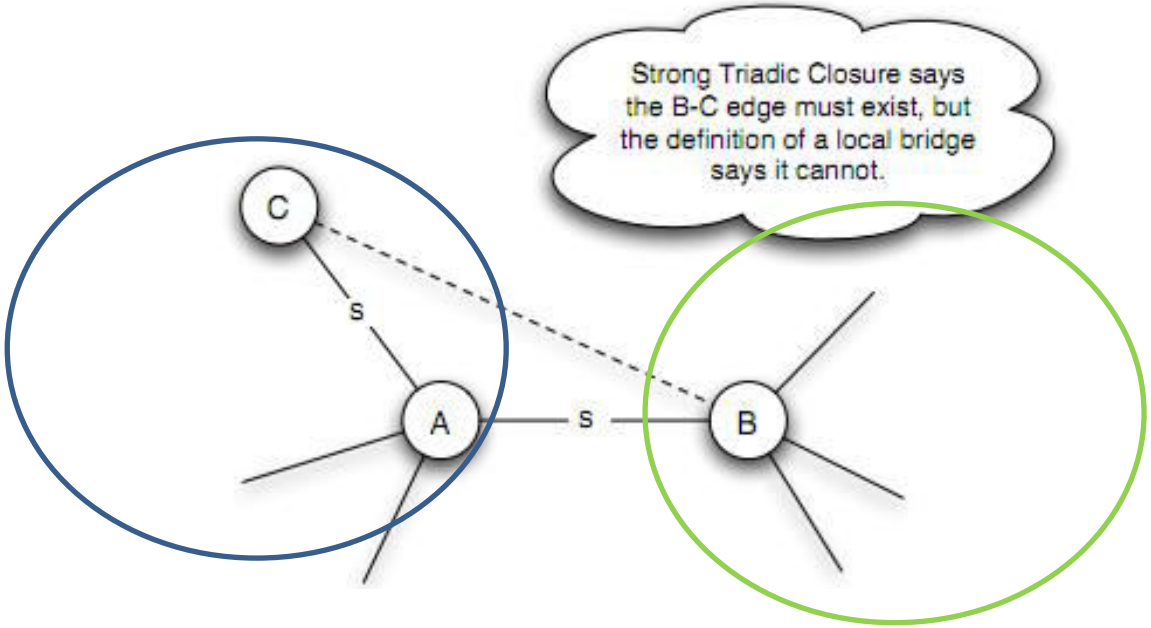
- "We say that a node A violates the Strong Triadic Closure property if it has strong ties to two other nodes B and C, and there is no edge at all (either a strong or a weak tie) between B and C.
- We say that a node A satisfies the Strong triadic Closure property if it does not violate it."



In other words, it's very likely a relationship of some sorts will eventually form between B-C



Strong and weak ties



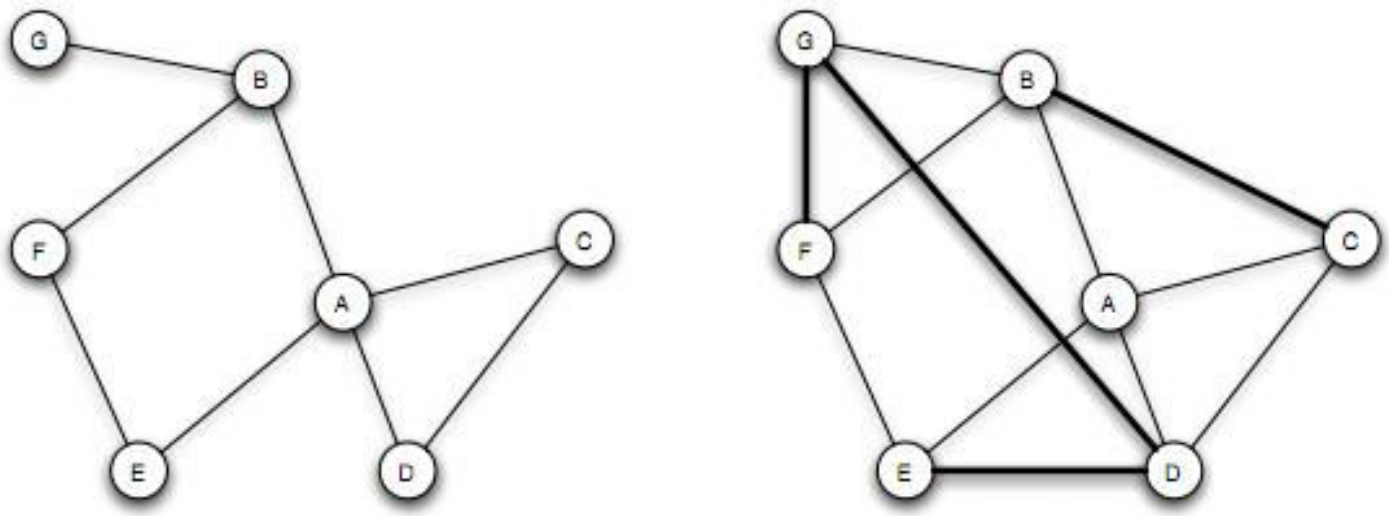
Why A – B can't be both bridge and strong tie

If a node satisfies Strong Triadic Closure and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie. Why?

If the A-B edge is a strong tie, then there must also be an edge between B and C, meaning that the A-B edge cannot be a local bridge

A related concept

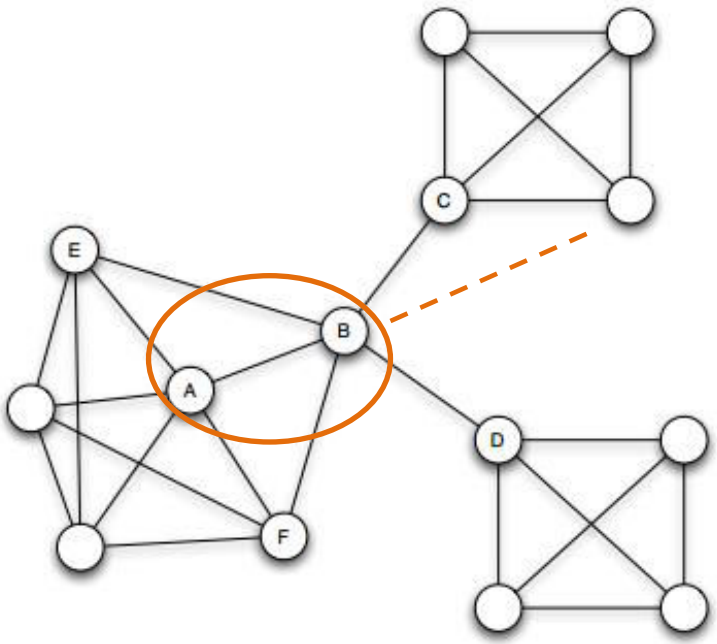
- Clustering coefficient with time
  - Measure the probability that two randomly selected friends of A are friends with each other.



(a) *Before new edges form.* (b) *After new edges form.*

Embeddedness

- The embeddedness of an edge in a network is defined as the number of common neighbors the two endpoints have.



Local bridge are precisely the edges that have embeddedness of ?

The stronger the tie, the more like the endpoints of the ties share neighbors

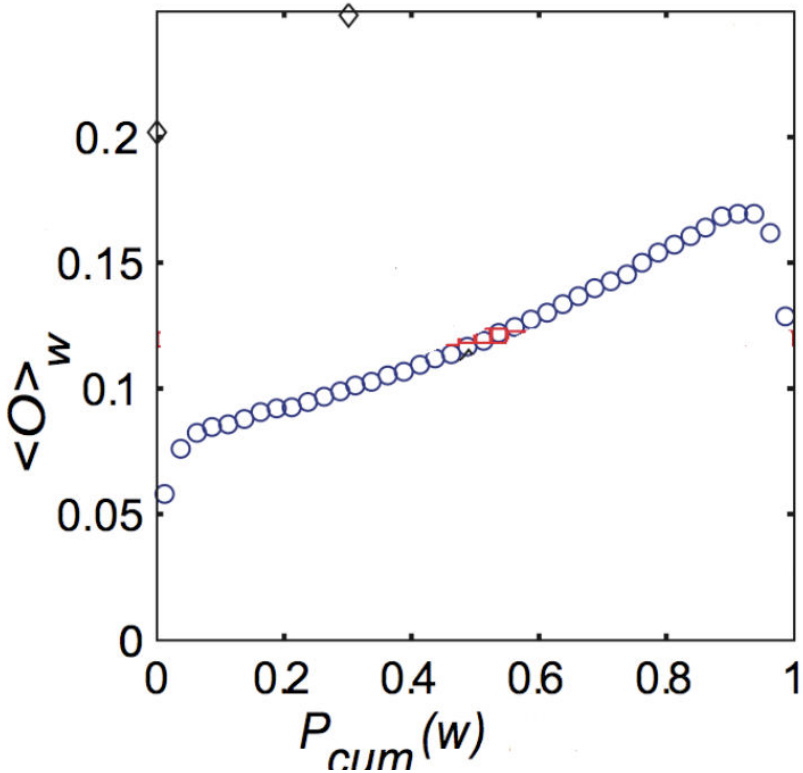
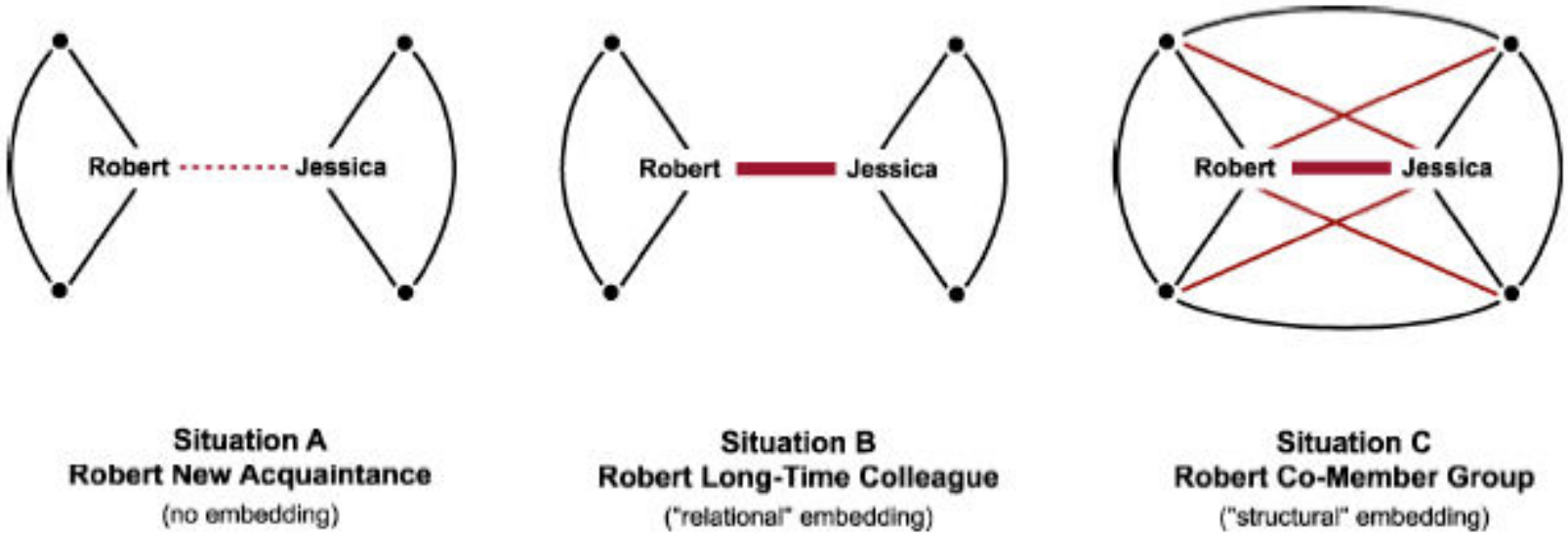


Figure 3.7: A plot of the neighborhood overlap of edges as a function of their percentile in the sorted order of all edges by tie strength. The fact that overlap increases with increasing tie strength is consistent with the theoretical predictions from Section 3.2. (Image from [334].)

Embedding and trust

Should Jessica trust Robert?



Embeddedness and trust in a network

- It has been shown in a long line of research that "If two individual are connected by an embedded edge, then this makes easier for them to trust one another, and to have confidence in the integrity of the transactions take place between them"
- Why is that?
- The presence of mutual friends puts the interactions between two people "on display" in a social sense...

Social capital

- The ability of actors to secure benefits by virtue of membership in social networks or other social structure
  - Bridging social capital: brokerage/bridging
  - Bounding social capital: Embeddedness (p.65, Easley, Kleinberg, 2010)
    - The embeddedness of an edge in a network to be the number of common neighbors the two endpoints have

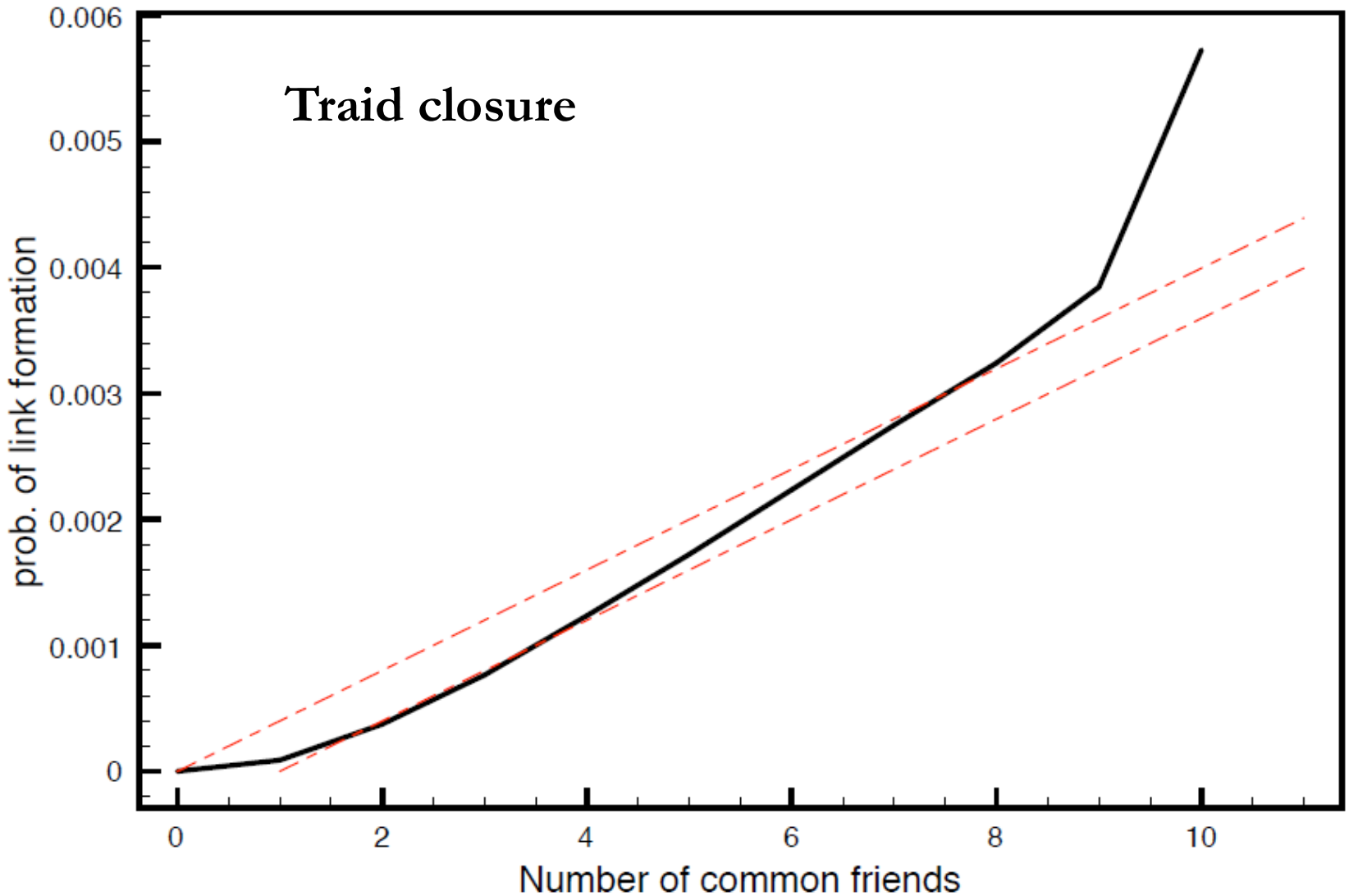


Figure 4.9: Quantifying the effects of triadic closure in an e-mail dataset [259]. The curve determined from the data is shown in the solid black line; the dotted curves show a comparison to probabilities computed according to two simple baseline models in which common friends provide independent probabilities of link formation.

Membership closure

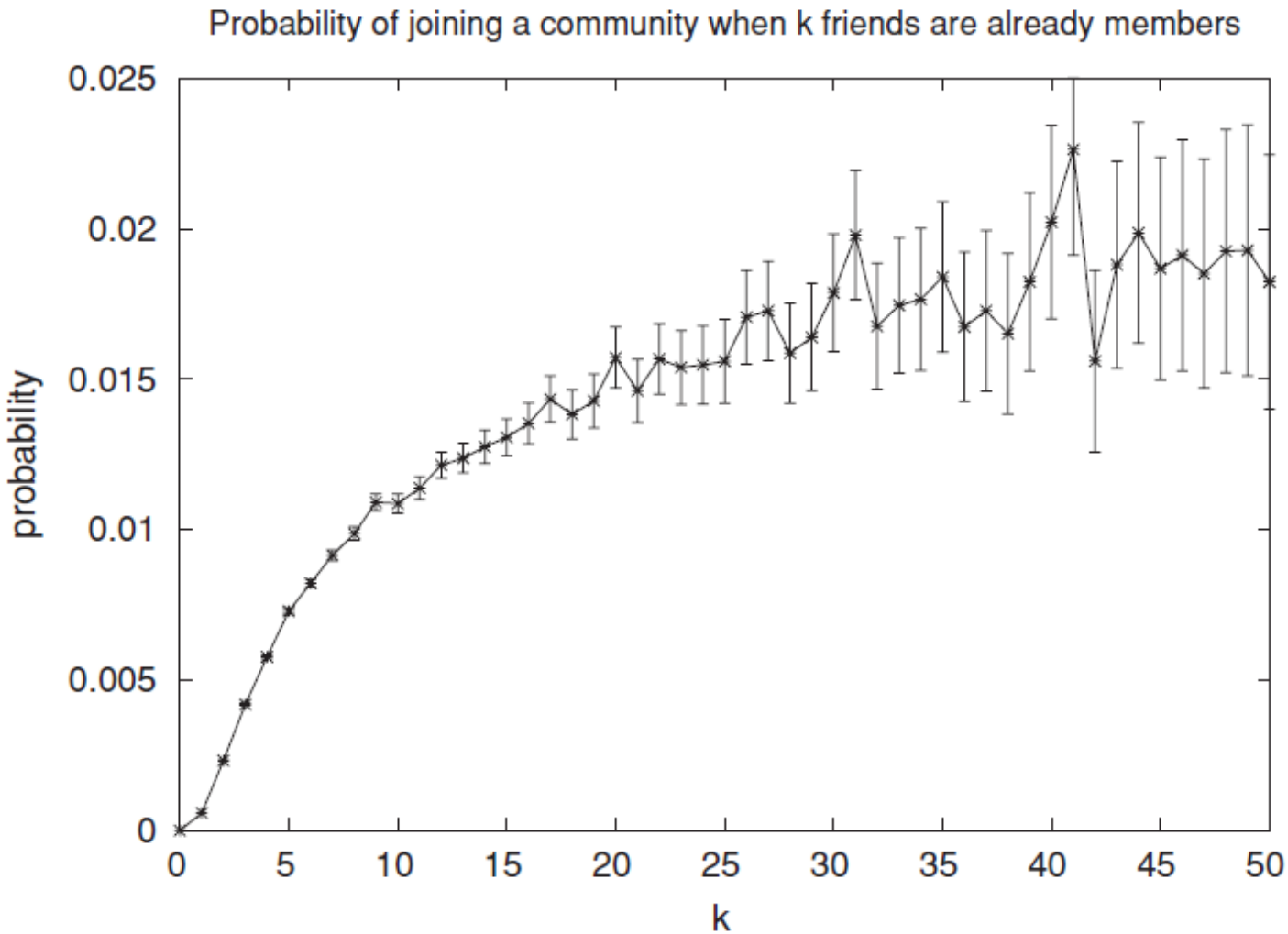


Figure 4.11: Quantifying the effects of membership closure in a large online dataset: The plot shows the probability of joining a LiveJournal community as a function of the number of friends who are already members [32].

Membership closure

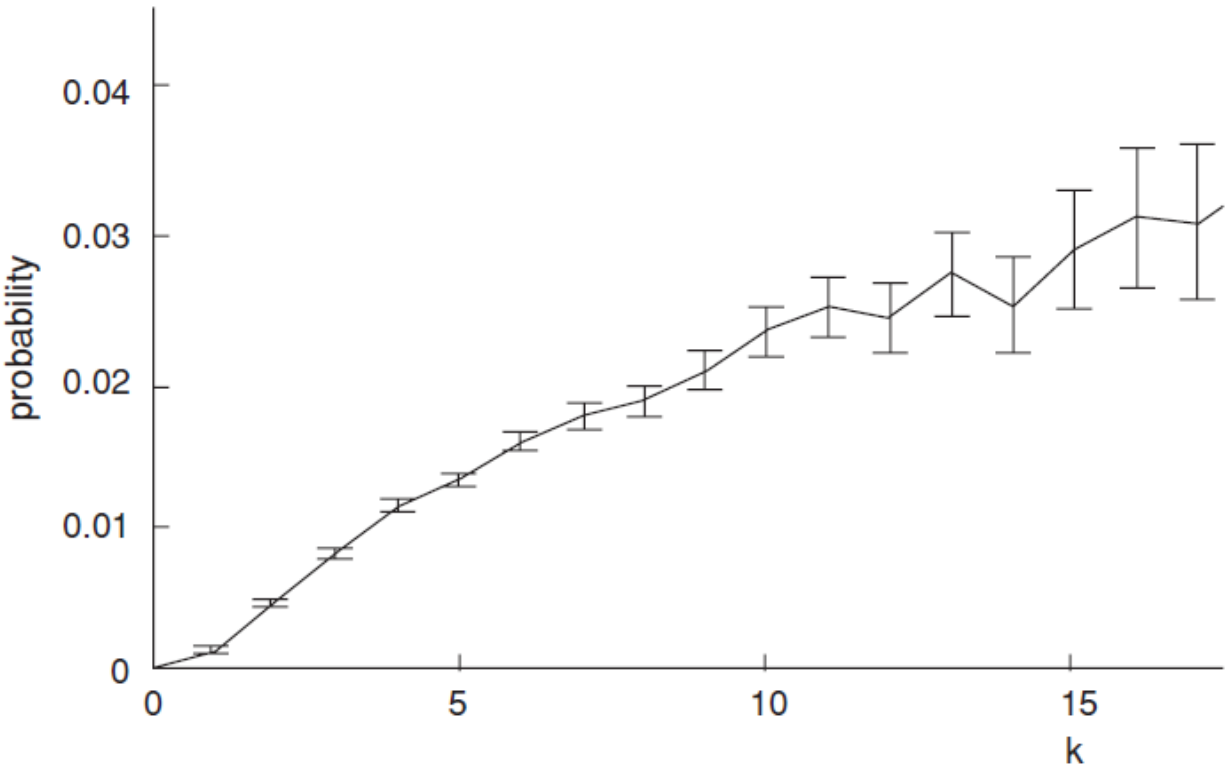


Figure 4.12: Quantifying the effects of membership closure in a large online dataset: The plot shows the probability of editing a Wikipedia articles as a function of the number of friends who have already done so [122].

Foci closure

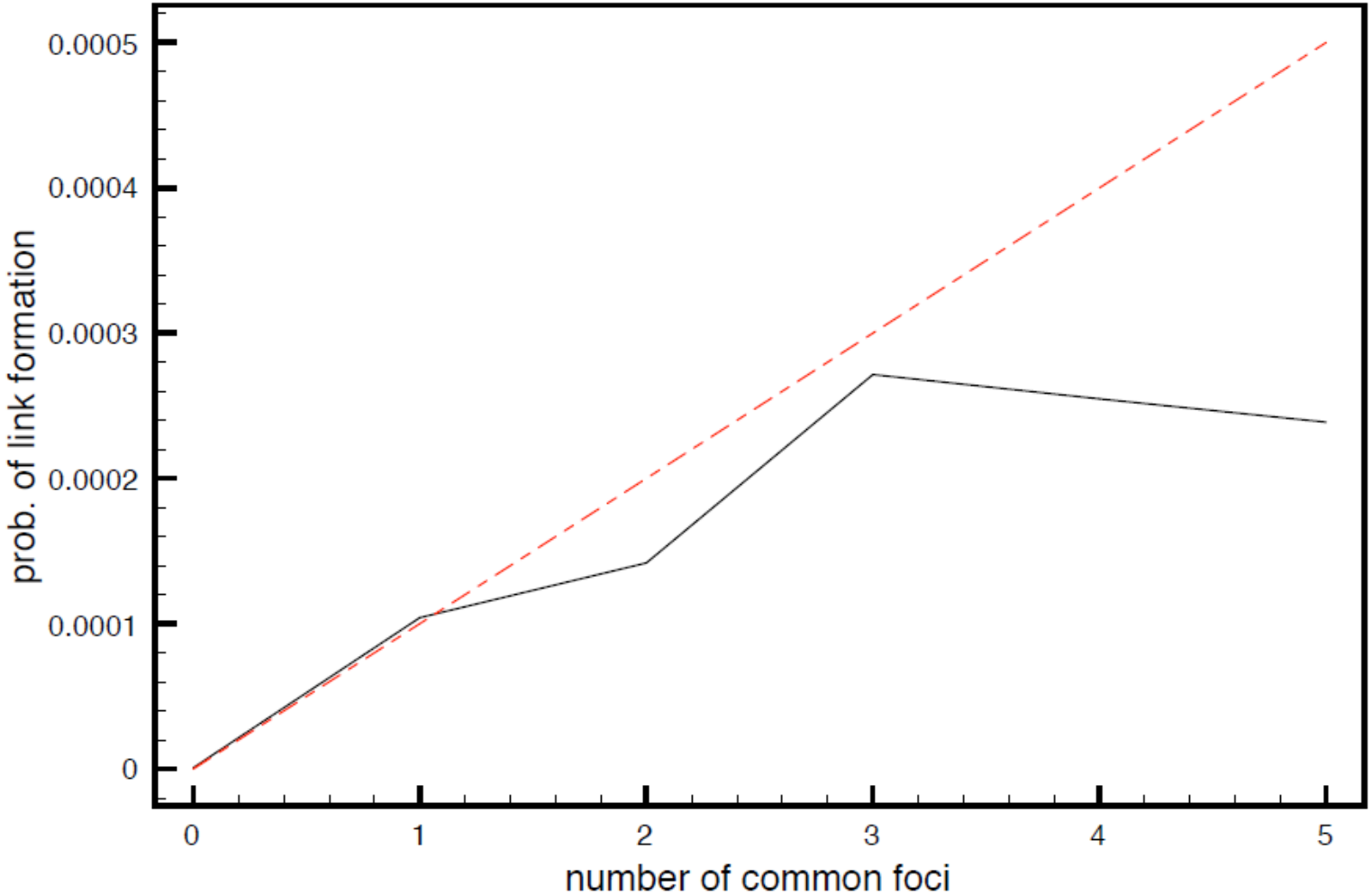


Figure 4.10: Quantifying the effects of focal closure in an e-mail dataset [259]. Again, the curve determined from the data is shown in the solid black line, while the dotted curve provides a comparison to a simple baseline.

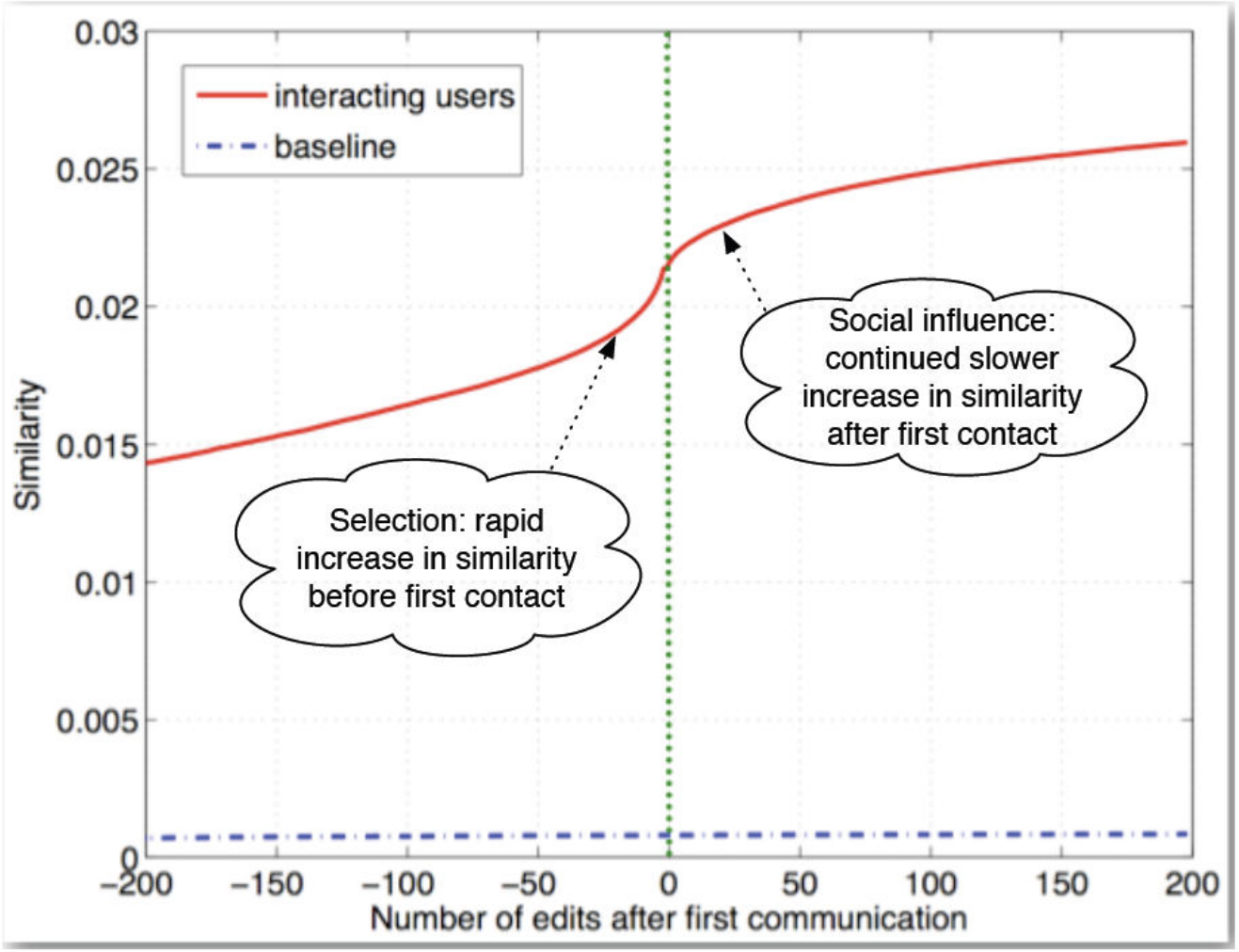


Figure 4.13: The average similarity of two editors on Wikipedia, relative to the time (0) at which they first communicated [122]. Time, on the x-axis, is measured in discrete units, where each unit corresponds to a single Wikipedia action taken by either of the two editors. The curve increases both before and after the first contact at time 0, indicating that both selection and social influence play a role; the increase in similarity is steepest just before time 0.

Mechanisms underlying homophily

- Selection
  - Tendency of people to form friendships with others who are like them
- Social influence
  - People modify their behaviors to bring them more closely into alignment with the behaviors of their friends

Causal relationships

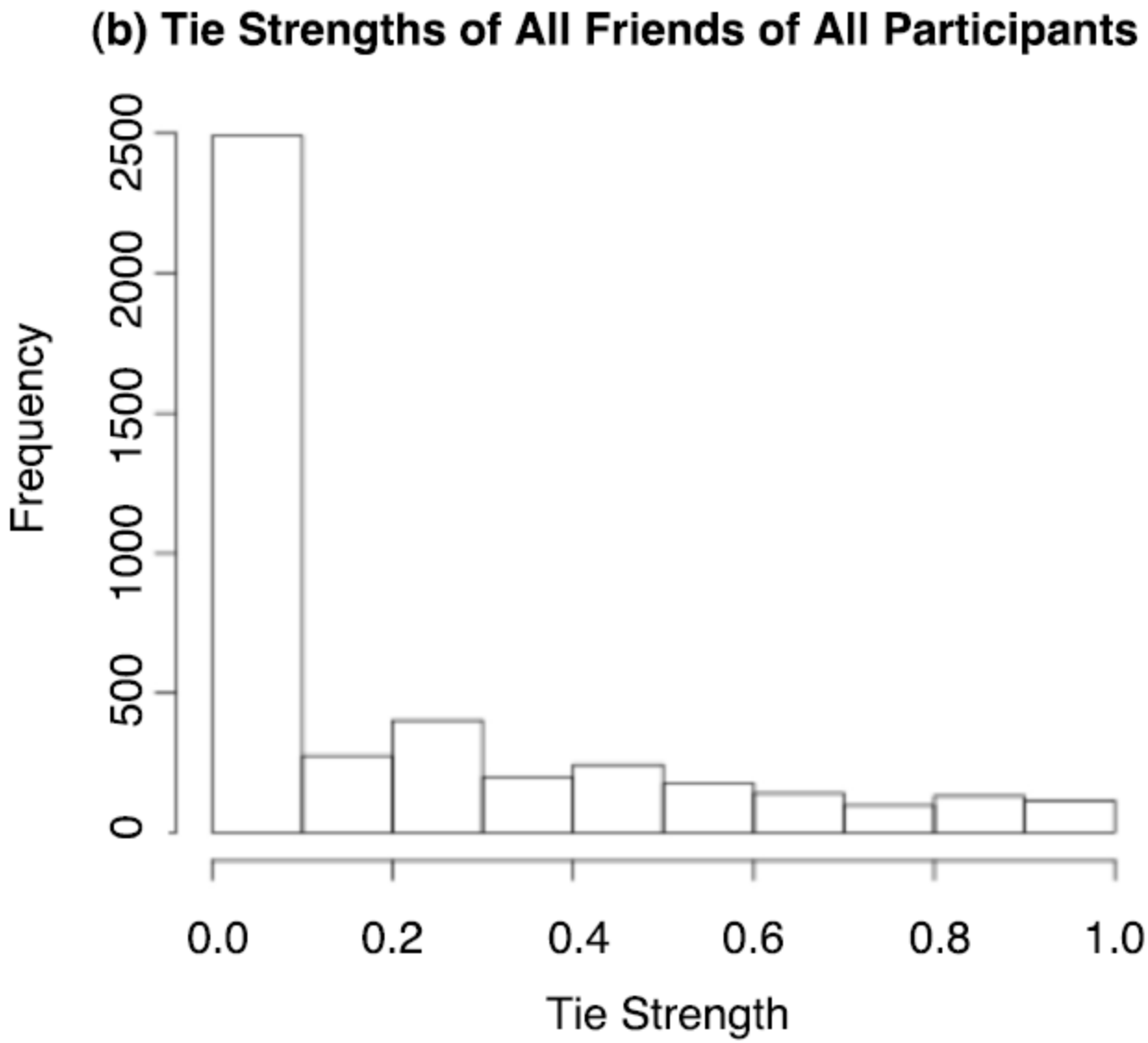
How tie-strength can be measured

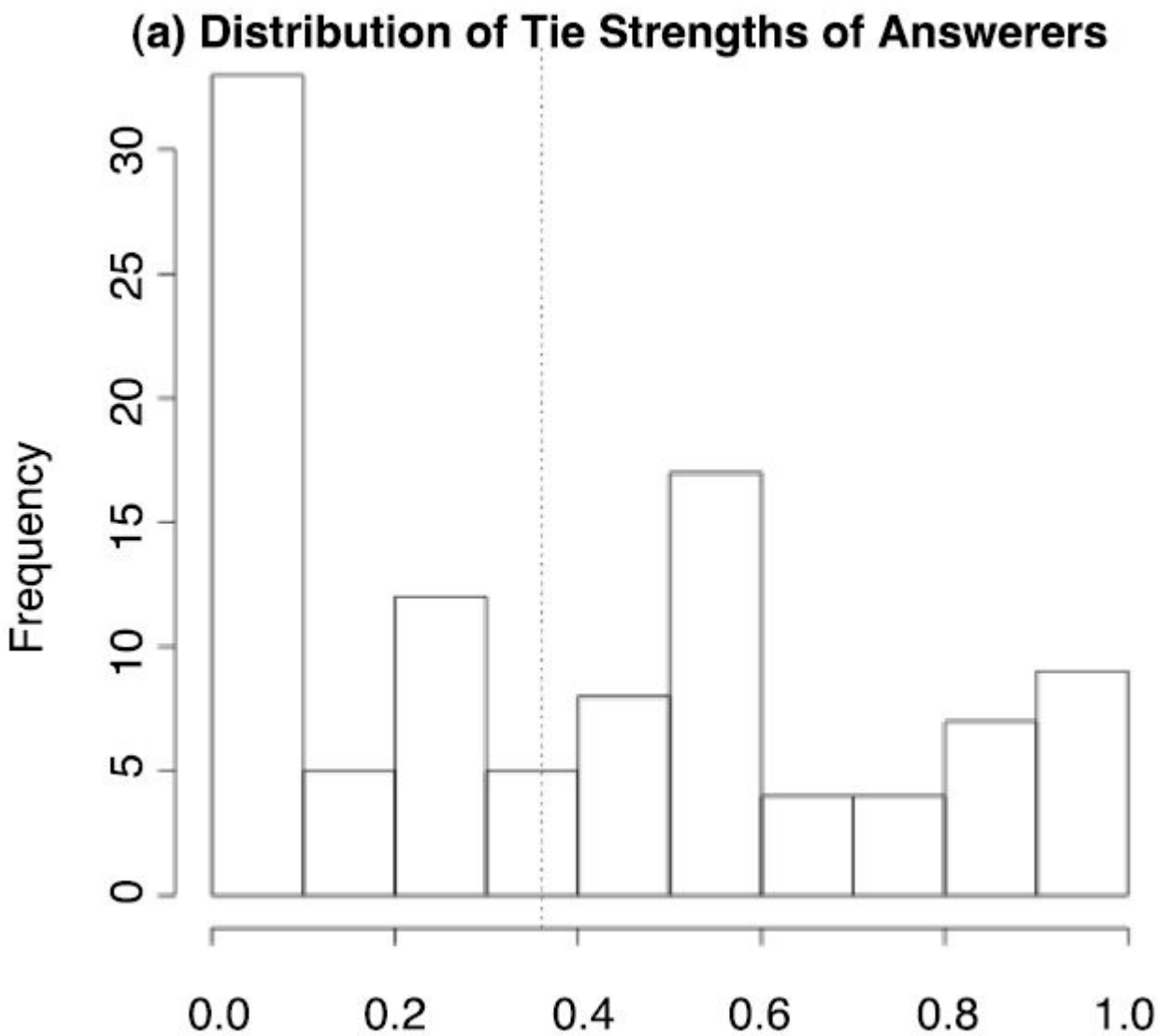
- Strong ties and weak ties on social media or, the web, or in real life
- Nominal, Ordinal, Interval, Ratio
- On social media, it's easy to form links yet strong ties are relative scarce

Panovich, K., Miller, R., & Karger, D. (2012, February). Tie strength in question & answer on social network sites. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 1057-1066). ACM.

Multiple-regression model of tie strength

Our Predictive Variables	β
Days since last communication	-0.76
Days since first communication	0.755
Words exchanged	0.299
Mean tie strength of mutual friends	0.257
Positive emotion words	0.135
Intimacy words	0.111


(b) Tie Strengths of All Friends of All Participants
Frequency
0
500
1000
1500
2000
2500
0.0
0.2
0.4
0.6
0.8
1.0
Tie Strength


(a) Distribution of Tie Strengths of Answerers
Frequency
0
5
10
15
20
25
30
0.0
0.2
0.4
0.6
0.8
1.0

Find no significant correlations between tie strength and

1. How on topic an answer was
2. How knowledgeable a friend was about a topic
3. How supportive the friend was perceived to be
4. How interesting an answer was
5. How novel the information was
6. How much the answer was trusted

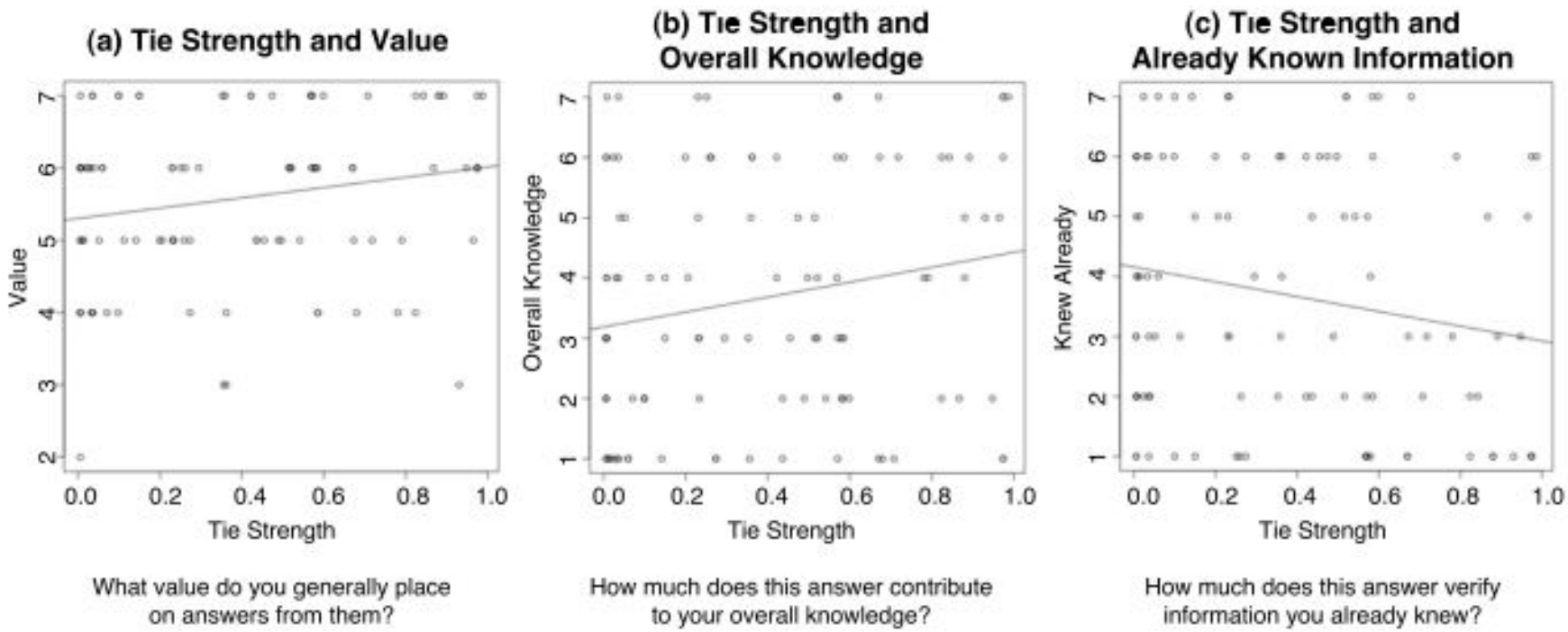


Figure 4. Tie strength and (a) value of answers in general, (b) overall knowledge gained from each answer, and (c) information already known by the participant.

SNA

Cliques and sub-groups

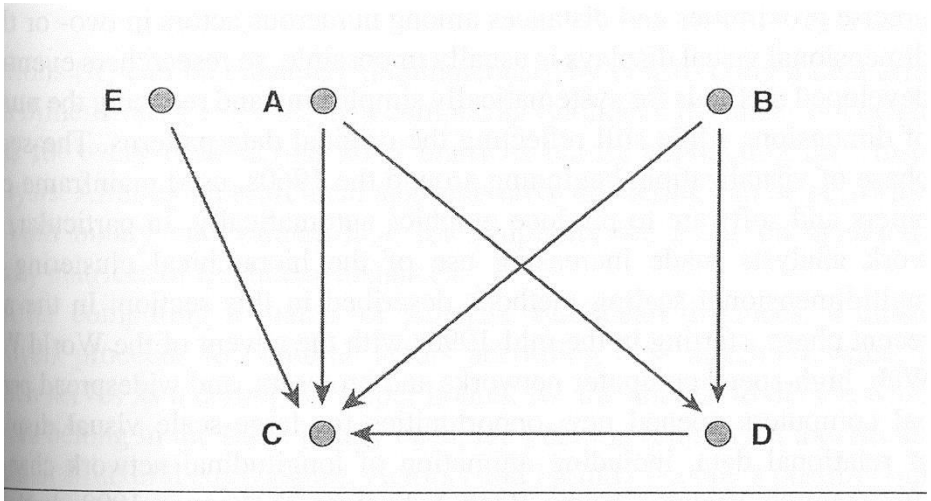
Identify subgroups

- A primary interest in Social Network Analysis is the identification of "significant social **subgroups**" – some smaller collection of nodes in the graph that can be considered, at least in some senses, as a "unit" **based on the pattern, strength, or frequency of ties**.
- Exogenous vs. endogenous groups

Grouping/clustering in SNA

- Cohesive subgroups (cohesion)
  - Areas of the network in which actors are more closely related to each other than actors outside the group
- Structural equivalent groups (Similarity)
  - Actors are classified based on how similar their interaction profiles are
  - Topic for another day

Finding sub-groups with similar network position: Structural equivalence (later)



ure 4.12 Digraph of a Five-Node Network for Structural Equivalence

	A	*B*	*C*	*D*	*E*
A	-	0	1	1	0
B	0	-	1	1	0
C	0	0	-	0	0
D	0	0	1	-	0
E	0	0	1	0	-

Cohesive subgroup/community

- Mutuality of ties
  - Everyone in the group has ties to one another
- Compactness
  - Closeness or reachability of group members in small number of steps, not necessarily adjacent.
- Separation
  - Higher frequency of ties among group members compared to non-members
    - E.g. fraction
- Density of edges
  - High frequency of ties within the group

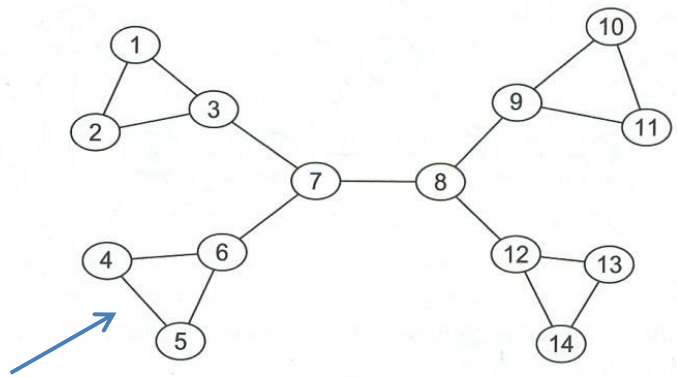
Finding cohesive sub-groups

- Bottom-up (agglomerative)
  - When two actors have a tie, they form a "group", to think about the structure of a network begins with this most base group, and see how far things can go from here.
- Top-down (divisive)
  - Start with the entire network and identify substructures that are locally denser than the field as a whole.
  - Looking for "holes", "vulnerabilities" and "weak spots"

Cohesion in directed & weighted networks

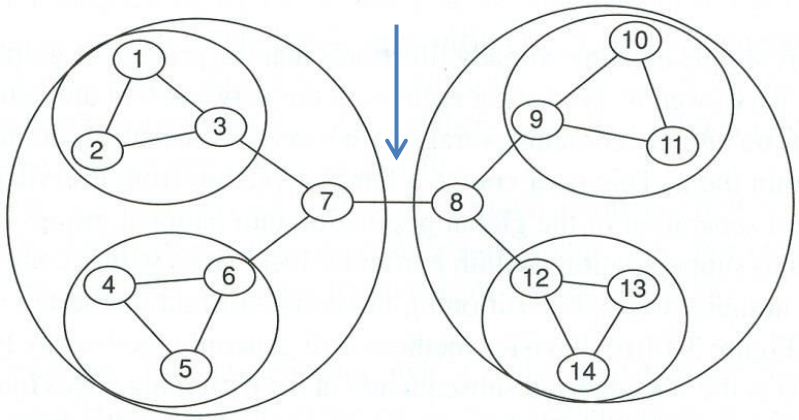
- Keep only a subset of ties
  - Reciprocal ties
    - Strongly connected components
  - Edge weight above a threshold
    - Newer algorithm can handle weighed edges

Group Partitioning: bottom up vs. top down



(a)

Plus



(b)

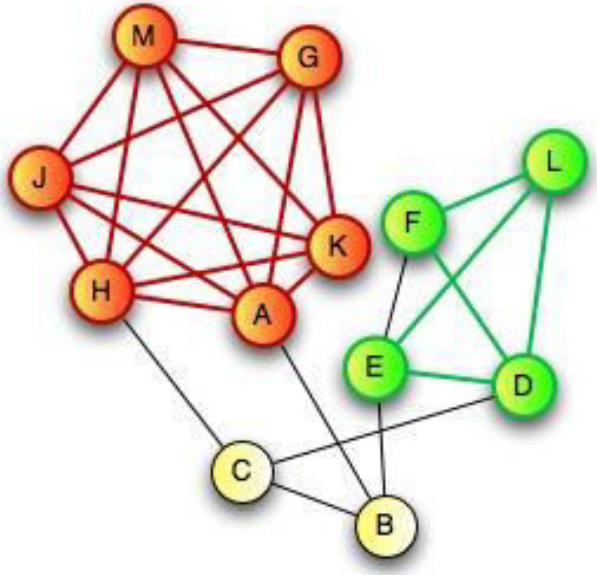
Minus

Finding sub-groups

- Bottom-up (clustering)
  - Cliques; N-Cliques
  - K-Core
  - Components
- Top-down
  - Faction
  - Betweenness
  - Blocks and cut-points
  - Lambda sets and bridges

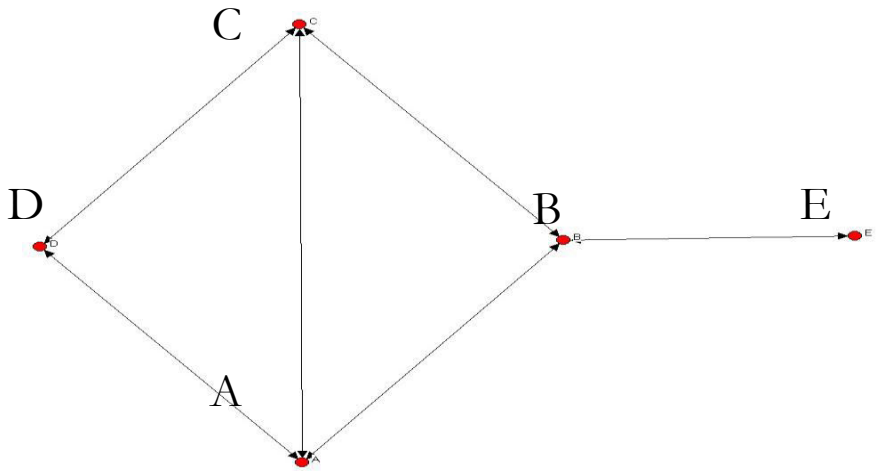
Cliques

- Cliques are groupings that are derived from a given network structure, on the basis of a definition of **complete connectedness** (every clique has a density of one).



Cliques

- The most important features of cliques are twofold:
  - (1) they are cohesive in the sense that all members within a clique are tied to (adjacent) one another, and
  - (2) they are not mutually exclusive, so that a node can belong to more than one clique and two cliques can overlap in terms of membership.

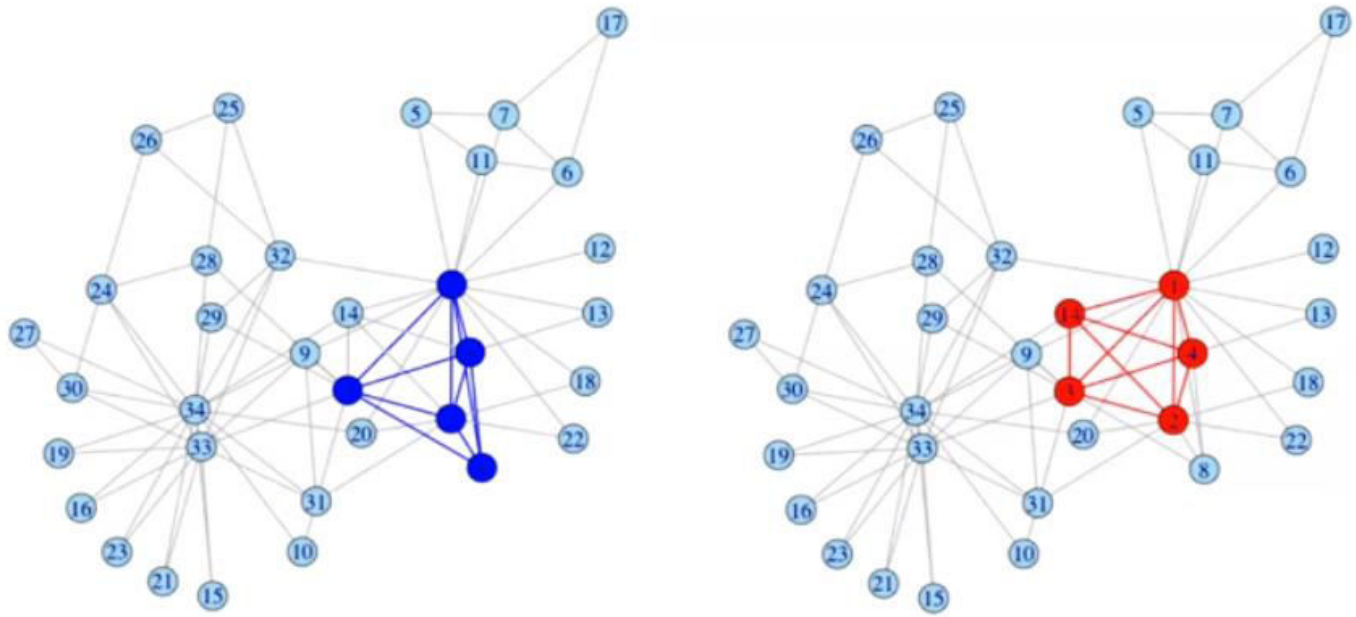


A occurs in two cliques (I, II)

Clique member matrix

Node	Clique		
	I	**II**	**III**
A	1	1	0
B	1	0	1
C	1	1	0
D	0	1	0
E	0	0	1

Maximal cliques



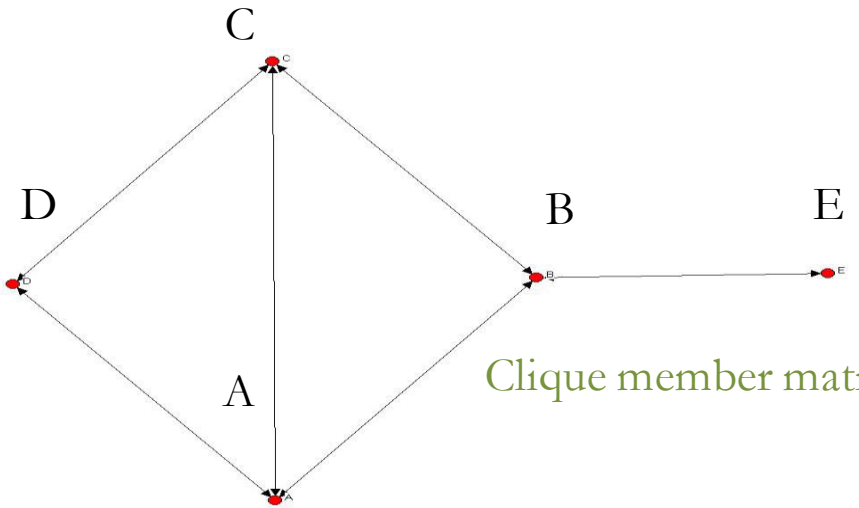
Maximal cliques:

Clique size:	2	3	4	5
Number of cliques:	11	21	2	2

ZACHE FOR GEPHI)

From Clique to similarity network

- Use clique co-membership to obtain a measure of association between each pair of actors
- Hierarchical clustering procedure can then be performed using average link method



Clique member matrix

Node	Clique		
	I	**II**	**III**
A	1	1	0
B	1	0	1
C	1	1	0
D	0	1	0
E	0	0	1

Actor-by-Actor clique co-membership matrix

A and C co-occur in two cliques (I and II)

	A	**B**	**C**	**D**	**E**
A	2	1	2	1	0
B	1	2	1	0	1
C	2	1	2	1	0
D	1	0	1	1	0
E	0	1	0	0	1

Clique-by-clique overlap matrix

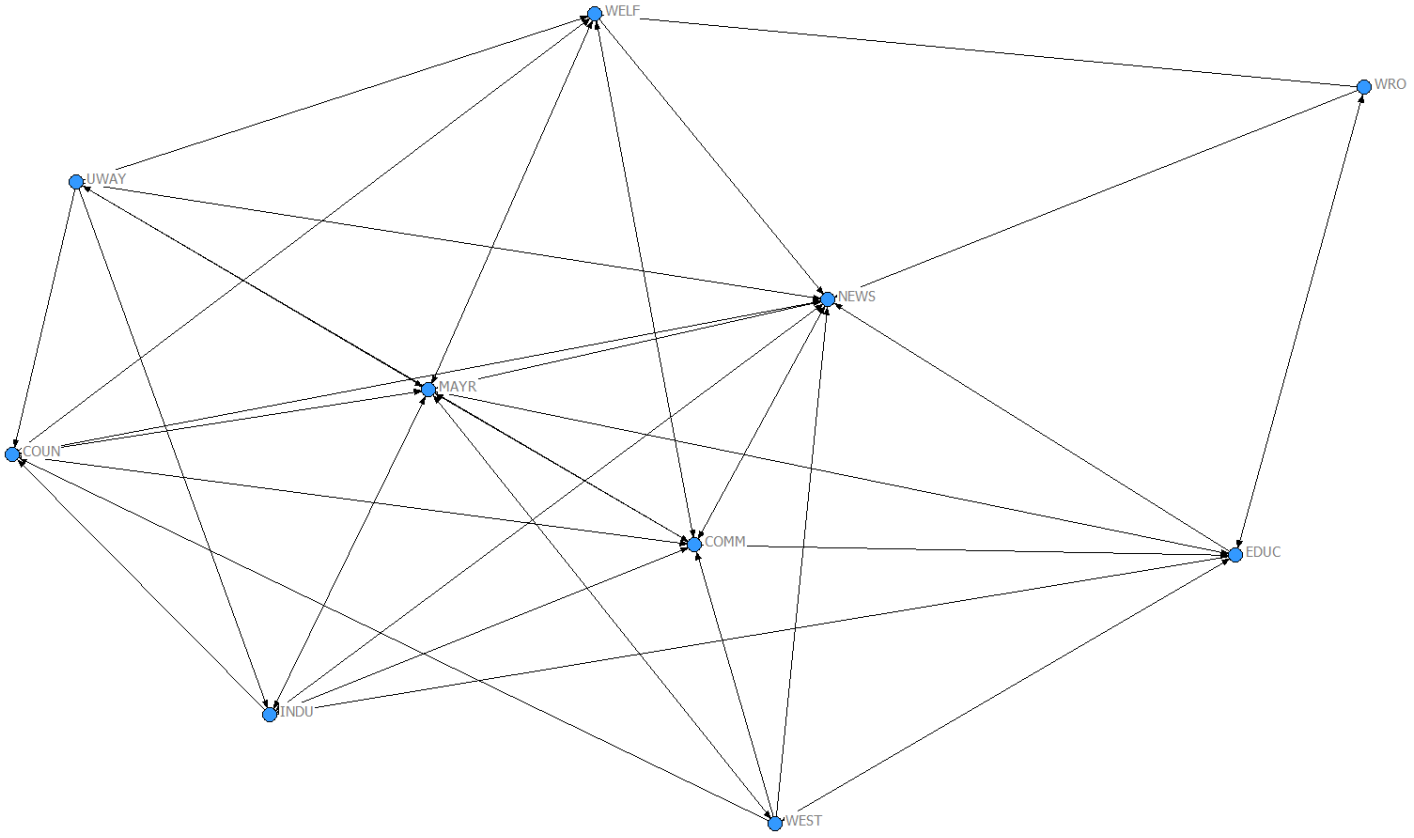
Clique I and II have two members in common (A, C)

	I	**II**	**III**
I	3	2	1
II	2	3	0
III	1	0	2

Cliques and N-Cliques

- *Network>Subgroups>Cliques / N –Cliques*
  - Number of cliques found
  - How "adjacent" each actor (row) is to each clique (column)
    - Degree of connection of actor to the members of a clique, in percentage
  - Hierarchical clustering based on co-membership
  - Co-membership matrix: the extent to which these sub-groups overlap

Original Knokbur

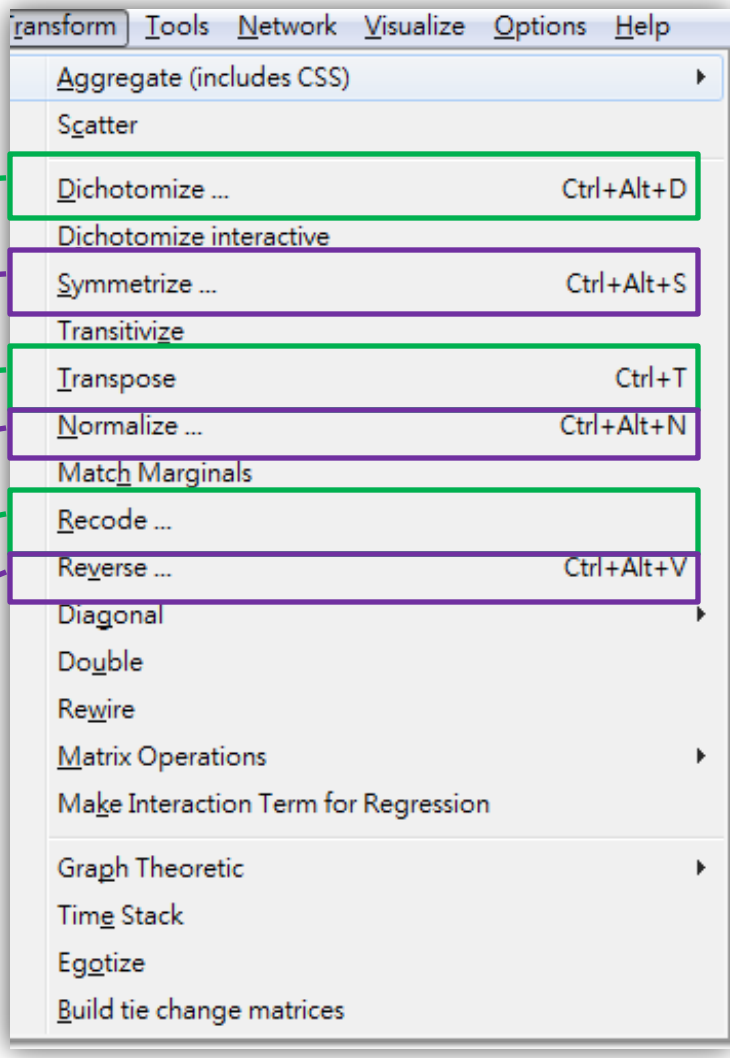


Cliques are based on symmetric (non-directional) binary relations. Non-symmetric relations cannot be partitioned into cliques without somehow converting them into symmetric relations. Also, valued relations must be converted into binary ones in order to allow derivation of cliques

Cliques are not mutually exclusive in terms of members. However, no clique can be a proper subset of another clique. Any two cliques derived from a given network must differ with respect to at least one node.

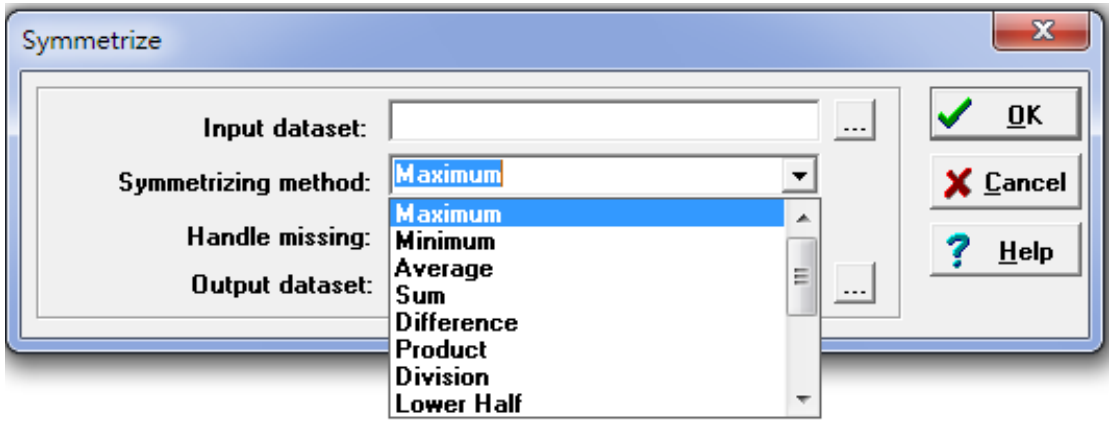
Transforming a matrix

- Transforming data values
  - 二元化
  - 對稱
  - 互換
  - 標準化
  - 代換
  - 相似-距離反轉



Symmetrize Knokbur

- Change an **unsymmetrical** matrix into a symmetric matrix by using one of a variety of criteria
  - Max.
  - Min.
  - Average



- Transform>Symmetrize
  - Choose method & output dataset

Symmetrize

Origin

	Actor1	Actor2	Actor3	Actor4	Actor5
Actor1	0	0	0	4	2
Actor2	0	0	1	0	2
Actor3	0	0	0	0	2
Actor4	1	1	1	0	3
Actor5	1	0	1	3	0

Max

	Actor1	Actor2	Actor3	Actor4	Actor5	Actor6	Actor7	A
Actor1	0	0	0	4	2	1	0	
Actor2	0	0	1	1	2	3	0	
Actor3	0	1	0	1	2	1	0	
Actor4	4	1	1	0	3	2	1	
Actor5	2	2	2	3	0	2	2	
Actor6	1	3	1	2	2	0	2	
Actor7	0	0	0	1	2	2	0	
Actor8	3	0	0	0	0	3	0	

Min

	Actor1	Actor2	Actor3	Actor4	Actor5	Actor6	Actor7
Actor1	0	0	0	1	1	0	0
Actor2	0	0	0	0	0	0	0
Actor3	0	0	0	0	1	0	0
Actor4	1	0	0	0	3	1	0
Actor5	1	0	1	3	0	1	0
Actor6	0	0	0	1	1	0	2
Actor7	0	0	0	0	0	2	0
Actor8	0	0	0	0	0	0	0

Average

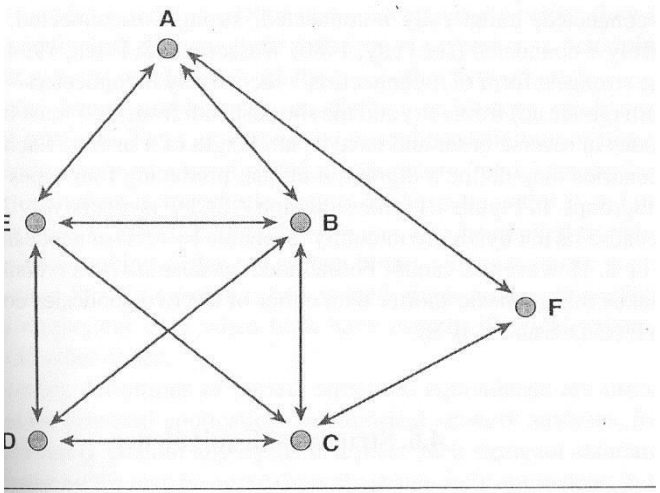
	Actor1	Actor2	Actor3	Actor4	Actor5	Actor6	Actor7
Actor1	0	0	0	2.5	1.5	0.5	0
Actor2	0	0	0.5	0.5	1	1.5	0
Actor3	0	0.5	0	0.5	1.5	0.5	0
Actor4	2.5	0.5	0.5	0	3	1.5	0.5
Actor5	1.5	1	1.5	3	0	1.5	1
Actor6	0.5	1.5	0.5	1.5	1.5	0	2
Actor7	0	0	0	0.5	1	2	0
Actor8	1.5	0	0	0	0	1.5	0

Cliques in a directed graphs

A subgraph with three or more nodes, all directly connected to one another,

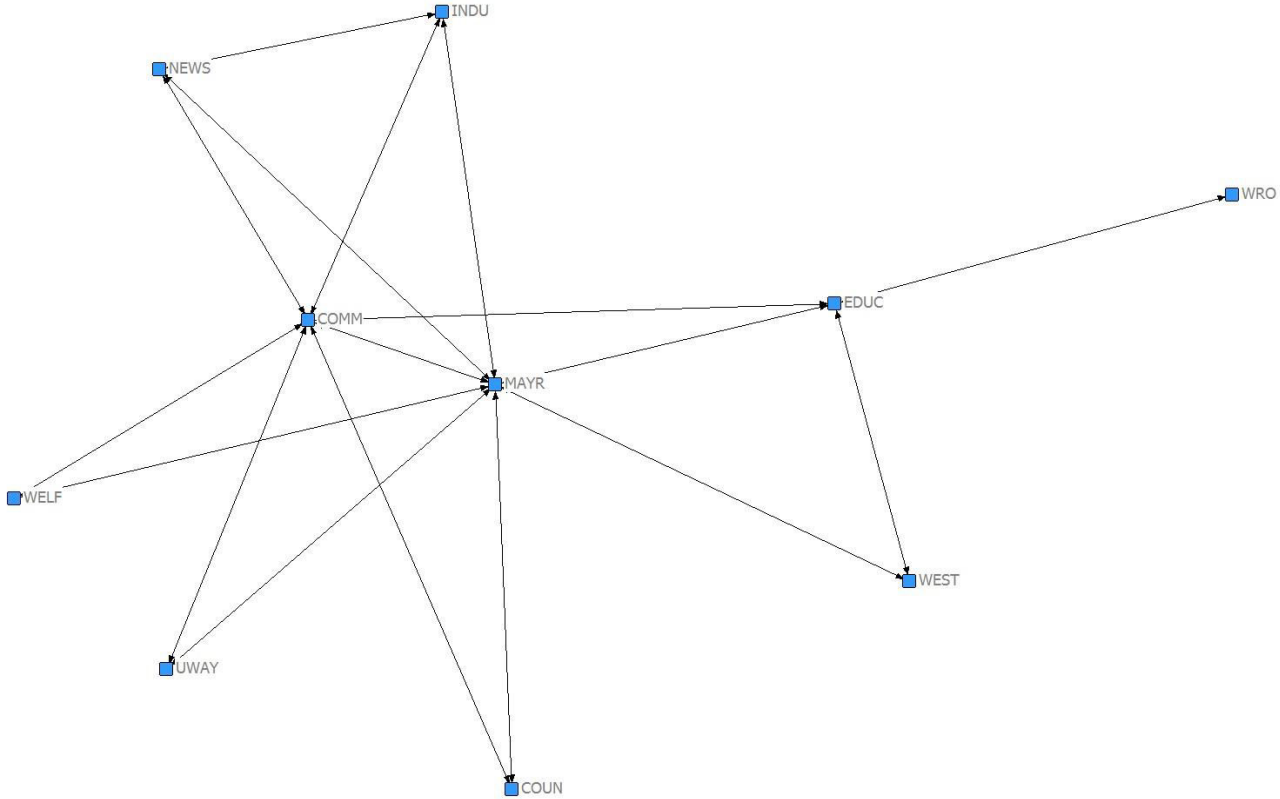
Where each pair is **reciprocally connected/ hint: product method in binary data (1, 0)**

Can you identify the three triadic cliques?



Cliques in a Directed Graph

Symmetrized knokbur, by reciprocal, i.e. product method



NOTE: Directed graph. You may prefer to symmetrize first.

6 cliques found.

1: COMM INDU MAYR NEWS
2: COMM EDUC MAYR
3: COUN COMM MAYR
4: COMM MAYR UWAY
5: COMM MAYR WELF
6: EDUC MAYR WEST

There are 6 maximal complete sub-graphs
Membership overlap among the sub-groups

ucinetlog1 - Notepad

File Edit Format View Help

```
NOTE: Directed graph. You may prefer to symmetrize first.
6 cliques found.

   1:  COMM INDU MAYR NEWS
   2:  COMM EDUC MAYR
   3:  COUN COMM MAYR
   4:  COMM MAYR UWAY
   5:  COMM MAYR WELF
   6:  EDUC MAYR WEST

Clique Participation Scores: Prop. of clique members that each node is adjacent to

            1     2     3     4     5     6
         ----- ----- ----- ----- ----- -----
  1 COUN  0.500 0.667 1.000 0.667 0.667 0.333
  2 COMM  1.000 1.000 1.000 1.000 1.000 0.667
  3 EDUC  0.500 1.000 0.667 0.667 0.667 1.000
  4 INDU  1.000 0.667 0.667 0.667 0.667 0.333
  5 MAYR  1.000 1.000 1.000 1.000 1.000 1.000
  6  WRO  0.000 0.333 0.000 0.000 0.000 0.333
  7 NEWS  1.000 0.667 0.667 0.667 0.667 0.333
  8 UWAY  0.500 0.667 0.667 1.000 0.667 0.333
  9 WELF  0.500 0.667 0.667 0.667 1.000 0.333
 10 WEST  0.250 0.667 0.333 0.333 0.333 1.000
```

Shows "connectedness" of each member to the 6 cliques"

Clique Participation Scores: Prop. of clique members that each node is adjacent to

For example, actor 1 is adjacent to 2/3 of the members of clique 5

Actor-by-Actor Clique Co-Membership Matrix

```
                1
      1 2 3 4 5 6 7 8 9 0
      C C E I M W N U W W
      ----------
 1  1 1 1 0 0 1 0 0 0 0 0
 2  2 1 5 1 1 5 0 1 1 1 0
 3  3 0 1 2 0 2 0 0 0 0 1
 4  4 0 1 0 1 1 0 1 0 0 0
 5  5 1 5 2 1 6 0 1 1 1 1
 6  6 0 0 0 0 0 0 0 0 0 0
 7  7 0 1 0 1 1 0 1 5 0 0
 8  8 0 1 0 0 1 0 0 1 0 0
 9  9 0 1 0 0 1 0 0 0 1 0
10 10 0 0 1 0 1 0 0 0 0 1
```

Show # of cliques each pair of actors are both members of

HIERARCHICAL CLUSTERING OF OVERLAP MATRIX

```
      C E C M   I N U W W
    W O D O A N E W E E
    R U U M Y D W A L S
    O N C M R U S Y F T

              1
Level  6 1 3 2 5 4 7 8 9 0
-----  ----------
5.000  . . . X X X . . . . .
1.667  . . X X X X X . . . .
1.000  . . X X X X X X X X . . .
0.750  . X X X X X X X X X . . .
0.600  . X X X X X X X X X X . . .
0.240  . X X X X X X X X X X X X . .
0.111  . X X X X X X X X X X X X X X .
0.052  . X X X X X X X X X X X X X X X X X
0.000  X X X X X X X X X X X X X X X X X X X
```

Hierarchical clustering based on co-membership

Open “CliqueParticipation” in Netdraw using “two-mode network”

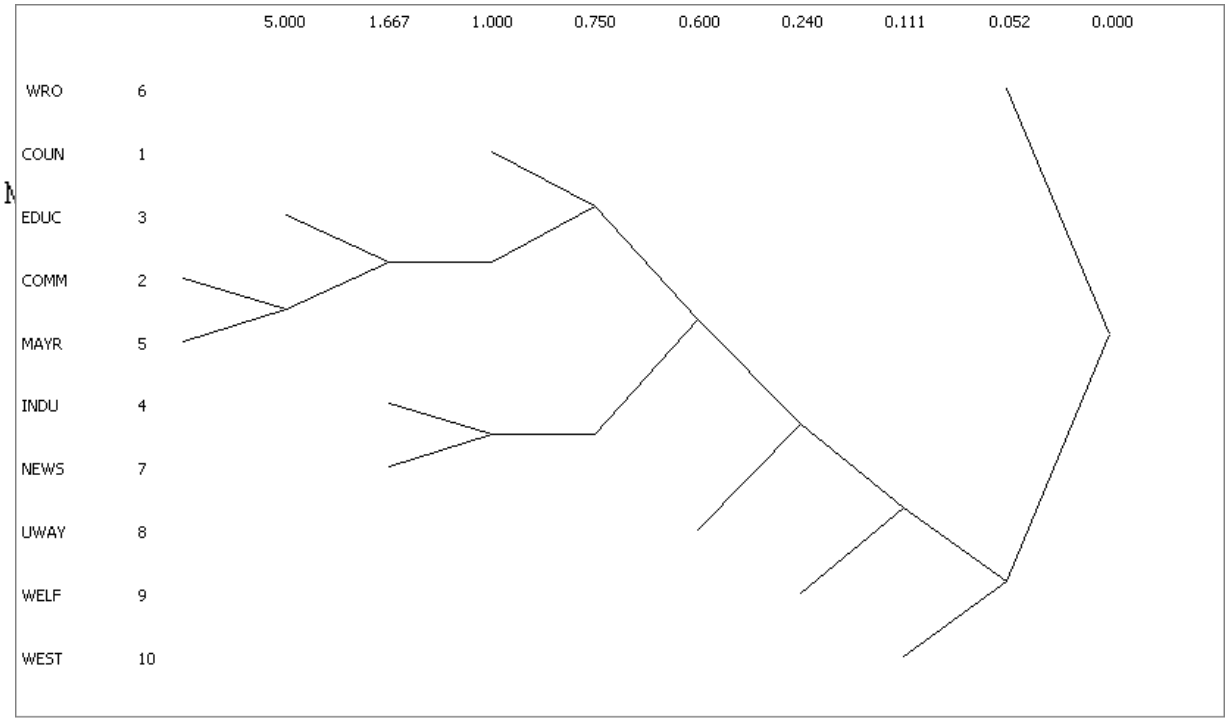
Actor-by-Actor Clique Co-Membership Matrix

```
                1
       1 2 3 4 5 6 7 8 9 0
       C C E I M W N U W W
       - - - - - - - - - -
  1  1 1 1 0 0 1 0 0 0 0 0
  2  2 1 5 1 1 5 0 1 1 1 0
  3  3 0 1 2 0 2 0 0 0 0 1
  4  4 0 1 0 1 1 0 1 0 0 0
  5  5 1 5 2 1 6 0 1 1 1 1
  6  6 0 0 0 0 0 0 0 0 0 0
  7  7 0 1 0 1 1 0 1 0 0 0
  8  8 0 1 0 0 1 0 0 1 0 0
  9  9 0 1 0 0 1 0 0 1 0 0
 10 10 0 0 1 0 1 0 0 0 0 1
```

HIERARCHICAL CLUSTERING OF OVERLAP M

```
       C E C M   I N U W W
     W O D O A N E W E E
     R U U M Y D W A L S
     O N C M R U S Y F T

                1
Level  6 1 3 2 5 4 7 8 9 0
-----  - - - - - - - - - -
5.000  . . . X X X . . . . .
1.667  . . X X X X X . . . . .
1.000  . . X X X X X X X X . . .
0.750  . X X X X X X X X X . . .
0.600  . X X X X X X X X X X X . . .
0.240  . X X X X X X X X X X X X X . .
0.111  . X X X X X X X X X X X X X X X .
0.052  . X X X X X X X X X X X X X X X X X
0.000  X X X X X X X X X X X X X X X X X X X
```

Dendrogram



Clique-by-Clique Actor Co-membership matrix

	1	2	3	4	5	6	7
1	4	2	2	2	2	1	1
2	2	3	2	2	2	1	2
3	2	2	3	2	2	2	1
4	2	2	2	3	2	1	1
5	2	2	2	2	3	1	1
6	1	1	2	1	1	3	2
7	1	2	1	1	1	2	3

HIERARCHICAL CLUSTERING OF OVERLAP MATRIX

```
Level    1 2 3 4 5 6 7
-----    - - - - - - -
2.000    XXXXXXXXX XXX
1.072    XXXXXXXXXXXXX
```

Co-membership matrix: the extent to which these sub-groups overlap

Some Limitations

- Many SNA scholars avoid clique-based analysis primarily because cliques do not make for discrete partitions of the network. In many cases the **overlap** in clique membership is extensive to the point that many cliques are almost identical in terms of the identity of their members.
- As noted, the derivation of cliques is limited to binary symmetrical (i.e., nondirectional) relations. Relations need to be dichotomized. They can be dichotomized by setting any positive value to one, or by establishing relational thresholds.
- This definition of cliques suggest that large networks can have a huge number of cliques and their analysis might be very difficult.
- The derivation of cliques is based only on first-order ties. Cliques formed by indirect ties are not tapped in this approach to network partitioning.

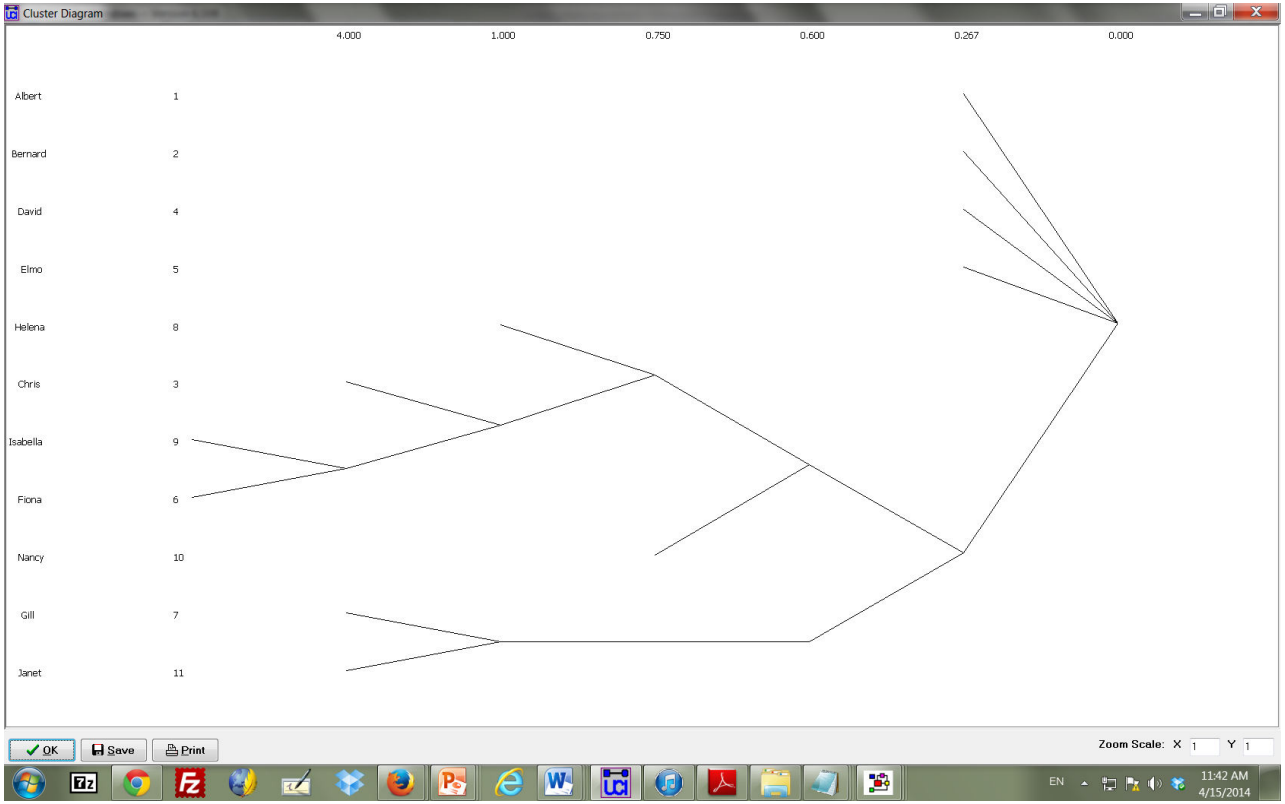
Exercise

1. Perform a clique analysis on "friendship" and zach data

Hint: zache binary unsymmetrical, symmetrized;

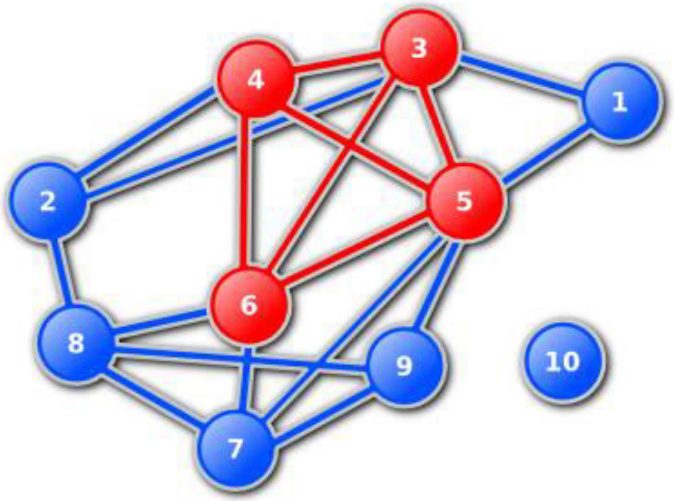
zachc numerical, dichotomized? How to determine the threshold?)

2. Visualize the networks of actor by actor clique co-membership results using NETDRAW



Cliques and N-Cliques

- *Network>Subgroups>N–Cliques*
  - *To loosen the definition to N steps*
  - *Try 2 cliques and see what happens*



Cliques are made more inclusive by relaxing definition of group membership

```
2 2-cliques found.

   1:  COUN COMM EDUC INDU MAYR NEWS UWAY WELF WEST
   2:  COMM EDUC MAYR WRO WEST


                                   1
         1 2 3 4 5 6 7 8 9 0
         C C E I M W N U W W
         - - - - - - - - - -
  1   1  1 1 1 1 1 0 1 1 1 1
  2   2  1 2 2 1 2 1 1 1 1 2
  3   3  1 2 2 1 2 1 1 1 1 2
  4   4  1 1 1 1 1 0 1 1 1 1
  5   5  1 2 2 1 2 1 1 1 1 2
  6   6  0 1 1 0 1 1 0 0 0 1
  7   7  1 1 1 1 1 0 1 1 1 1
  8   8  1 1 1 1 1 0 1 1 1 1
  9   9  1 1 1 1 1 0 1 1 1 1
 10  10  1 2 2 1 2 1 1 1 1 2

HIERARCHICAL CLUSTERING OF OVERLAP MATRIX

         C I N U W   M E C W
         O N E W E W A D O E
         U D W A L R Y U M S
         N U S Y F O R C M T

                                   1
Level    1 4 7 8 9 6 5 3 2 0
-----    - - - - - - - - - -
2.000    . . . . . . XXXXXXX
1.000    XXXXXXXXX XXXXXXXXX
0.833    XXXXXXXXXXXXXXXXXXX
```

Exercise: What happens if you run a 2-clique analysis on "friendship" data?

K-Cores

- A maximal group of actors, all of whom are connected to some number (k) of other members of the group. As K becomes smaller, group sizes will increase
- *Exercise*

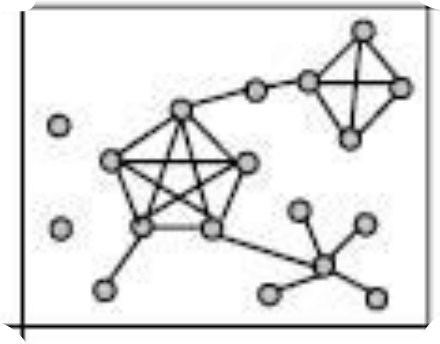
  Network>Regions>K-Core

  - *What happen if you run K-core on "friendship" data*
    - *Symmetrized > Minimum*
      - *When to use Minimum, Maximum, Average ?*
    - *Then redo K-core*
    - *Compare the differences*

  Use K-Cores filtering on Friendship Gephi hint: first visualized the node size by degree

Raise “k”, raise the threshold

Original graph



Zhang et al. (2008)

Components

- Sub-graphs that are connected within, but disconnected between sub-graphs
- *Network>Regions>Components>Simple graphs*
- *Network>Regions>Components>Valued Graphs*
  - *To examine the hierarchy of components as the cut-off value of tie strength is increasingly relaxed.*

Components cont.

- Strong component
  - A weak component is a set of nodes that are connected, regardless of the direction of ties.
- Weak component
  - A strong components requires that there be a directed path from A to B in order the two to be in the same component.

HIERARCHICAL COMPONENTS

	PAM_OMIDYAR	PIER_OMIDYAR	TERACH_OFTHEAGES	FEDERATION_OF_SCHOOLS	DEMOCRACY_PROTESTERS	SERVICE_IMPLOYEES	STATE_EMPLOYEES	REDEFINING_HASSLE	BUILDING_HASSLE	HELTER_PACKARD	CHEVRON	GRANITE_CONSTRUCTION	ROBERT_EINTEL	JOHN_TALLOW	DONALD_FISHER	ELI_BROADING	STEPHEN_LIGHT	THIMBLE_PAPER
Value	1	2	3	55	37	23	29	74	53	19	47	96	78	99	43	76	21	58	90	26	4
13.000	.	.	.	.	XXX	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
12.000	.	.	XXX	XXX	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
11.000	.	.	XXX	XXX	.	.	.	.	.	XXX	.	.	.	.	.	.	.	.	.	.	.
10.000	.	.	XXXXXXXXXXX						.	.	XXX	.	.	.	.	.	.	.	.	.	.
9.000	.	.	XXXXXXXXXXXXX							.	XXXXX	.	.	.	.	.	.	.	.	.	.
8.000	.	.	XXXXXXXXXXXXX							XXXXXXXXX			.	.	.	.	.	.	.	.	.
7.000	.	.	XXXXXXXXXXXXXXXXXXXXXXXXXX											.	.	.	.	.	.	.	.
6.000	.	.	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX															.	.	.	.
5.000	.	.	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX																		.
4.000	.	.	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX																		
3.000	.	.	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX																		
2.000	.	.	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX																		
1.000	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX																			
0.000	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX																			
-1.000	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX																			
-2.000	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX																			

Top-down
Blocks and Cutpoints (*Bi-components*)

- Cutpoints (node)
  - If a node were removed, would the structure become divided into un-connected parts (Blocks or Bi-components)
  - Limitation: allow only two groups
  - *Network>Regions>Bi-Component*

```
2 blocks found.


BLOCKS:

Block    1:  EDUC WRO
Block    2:  COUN COMM EDUC INDU MAYR NEWS UWAY WELF WEST

Articulation points

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

Faction

- 1. Determine the number of grouping
- 2. Arbitrarily assign every actor to one of the groups and then calculate the FIT to see how good the partition is
- 3. Move some actors from one group to another , calculate the fit again and see if there is any improvement
- 4. Continue until till no more improvement is possible
- Pitfall, will fine the best possible solution even though the data contain no real groups

Factions (Endogenous grouping)

- Factions: a subgroup in which each person is closely tied to all others within the group and no or few connections at all to other sub-groups
- Network>Subgroups>Factions (also in NetDraw)
  - Maximize within-group connections/density, minimize between-group connections
  - Errors (sum of zeros within, and ones between, i.e., noon-diagonal blocks)
  - **Use in an exploratory manner** (i.e. trail and errors)

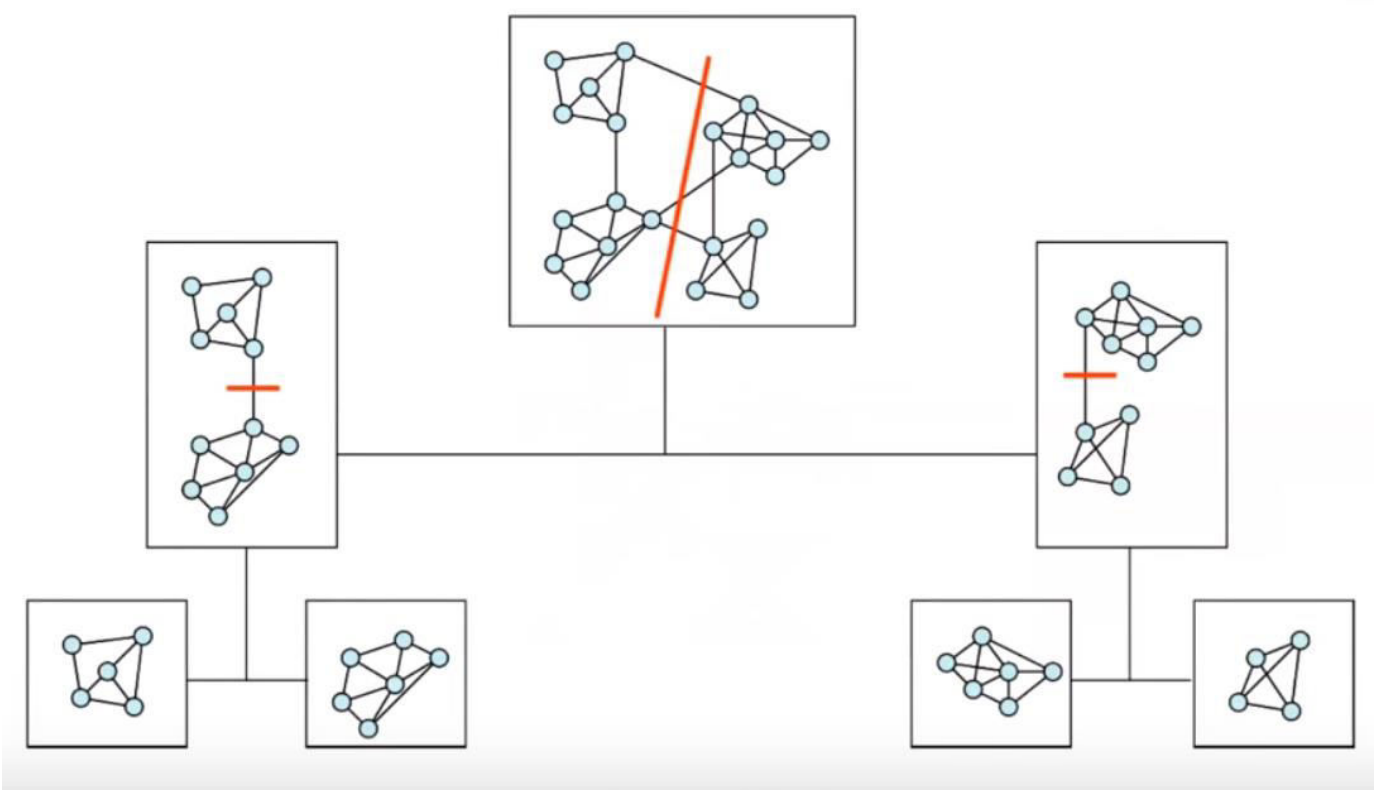
Girvan-Newman method of partitioning

- Based on "betweenness" measure
- Identifies **links** in a network that lie between communities and then removes them, leaving behind just the communities themselves.

Girvan-Newman method of partitioning

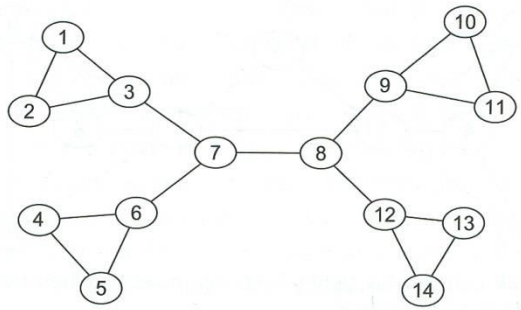
- 1. Find the edge or edges (if there is a tie) of highest betweenness, which causes the graph to separate into multiple components
  - Edge betweenness is similarity to Node betweenness
  - The number of times each edge is part of geodesic path.
- 2. Recalculate all betweennesses, and again remove the edge or edges of high
- 3. Proceed in the this way as long as edges remain in graph, in each step recalculating all betweennesses and removing the edge(s) with highest betweenness

Recursive partitioning

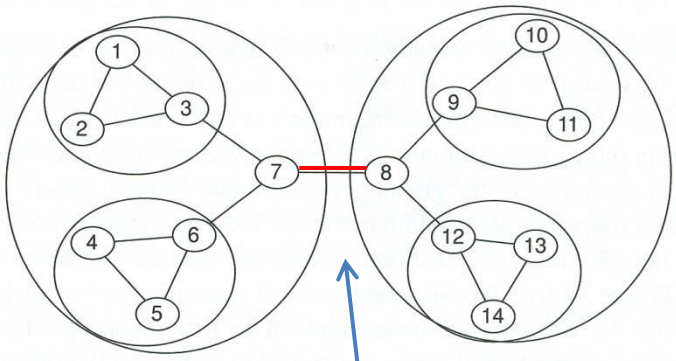


Give you a hierarchical structure

Divisive methods

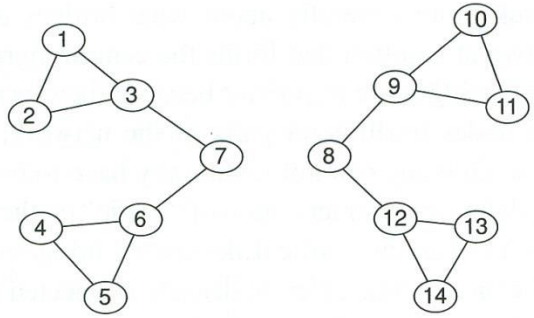


(a)

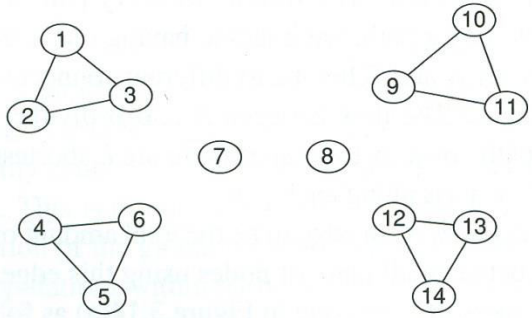


(b)

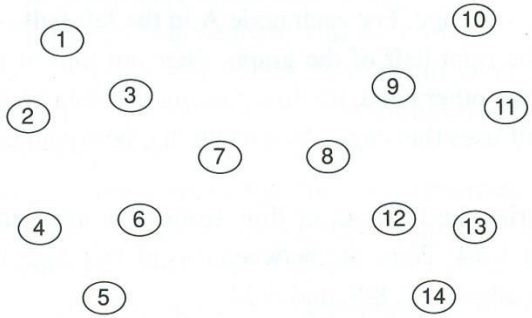
The full unit of flow passes through 7-8 edge
7x7 =49 (directed graph)


1
2
3
7
4
6
5
10
9
11
8
12
13
14

(a)

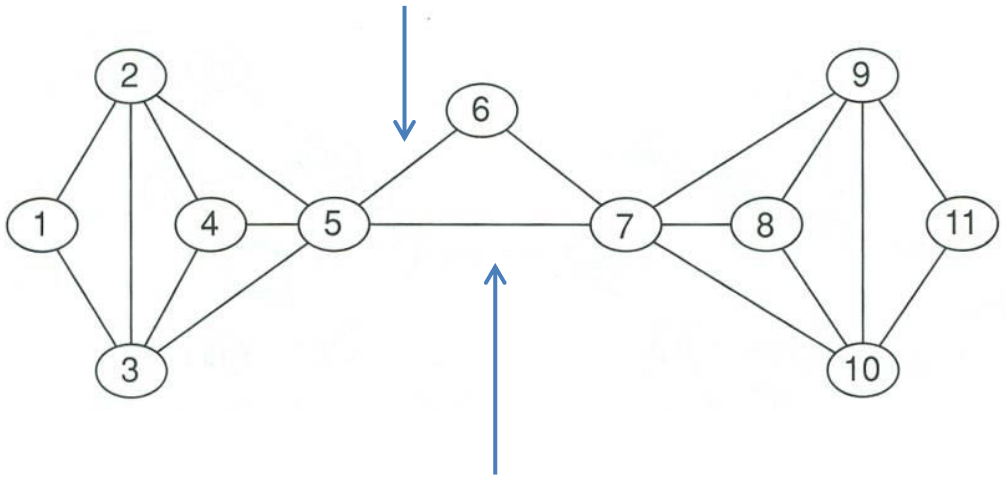

1
2
3
10
9
11
7
8
4
6
5
12
13
14

(b)

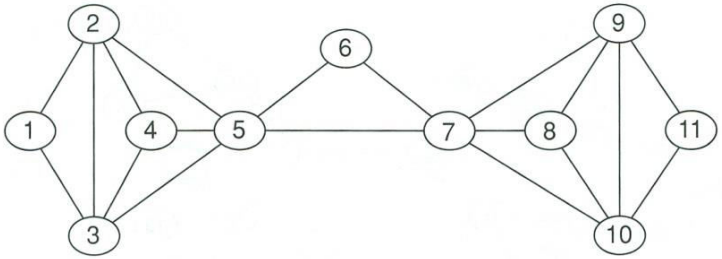

10
1
9
3
11
2
7
8
4
6
12
13
5
14

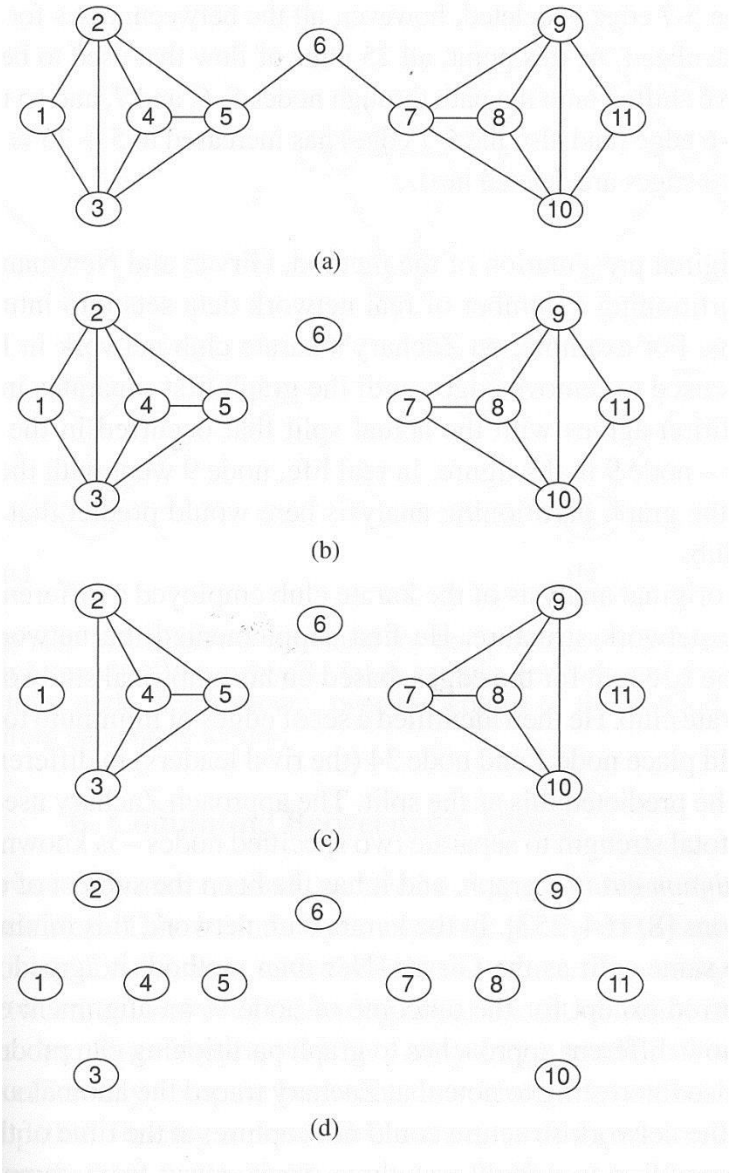
(c)

Carry a betweenness of 5
For it carries all the flow from nodes 6 to each of the 1-5 node



Carry a betweenness of 25
For it carries all the flow from nodes 1-5 to nodes 7-11


2
6
9
1
4
5
7
8
11
3
10


(a)
(b)
(c)
(d)

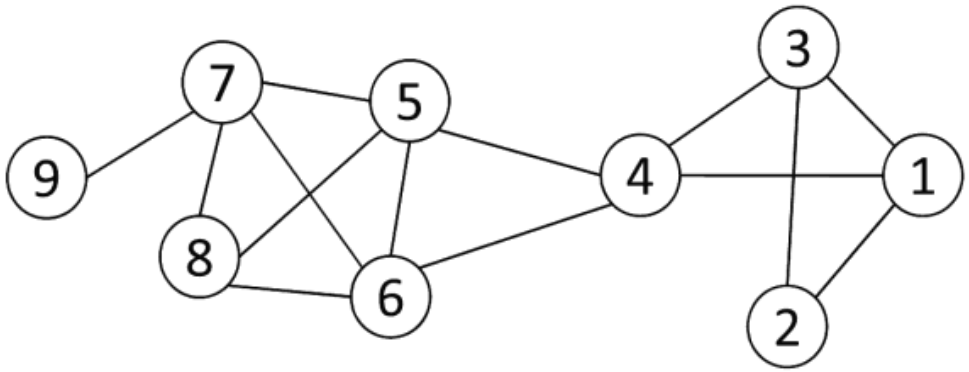
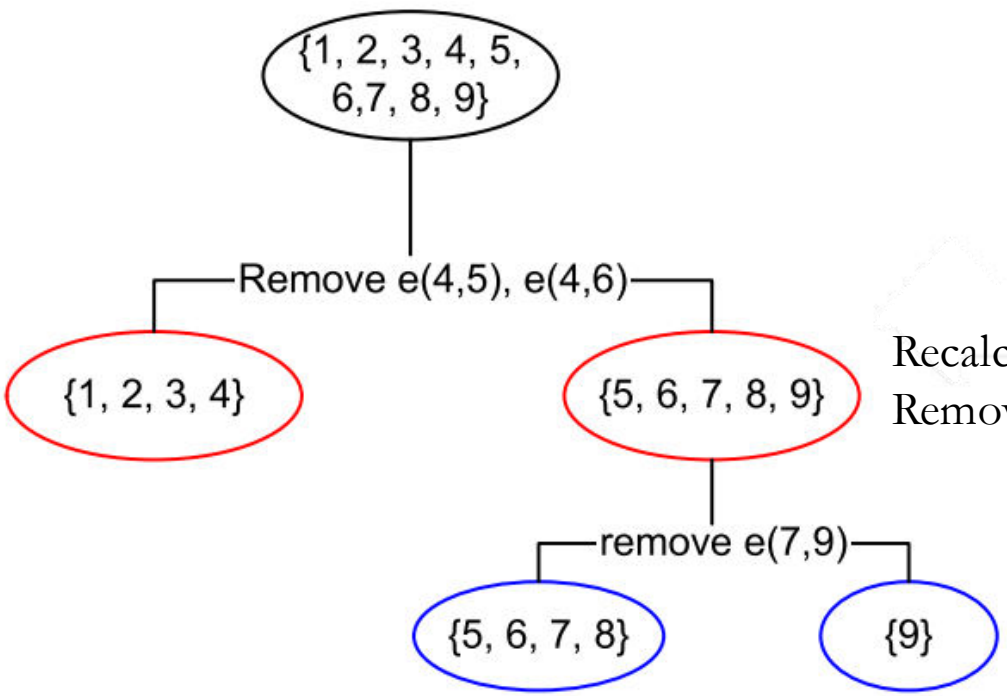


Table 3.3: Edge Betweenness

	1	2	3	4	5	6	7	8	9
1	0	4	1	9	0	0	0	0	0
2	4	0	4	0	0	0	0	0	0
3	1	4	0	9	0	0	0	0	0
4	9	0	9	0	10	10	0	0	0
5	0	0	0	10	0	1	6	3	0
6	0	0	0	10	1	0	6	3	0
7	0	0	0	0	6	6	0	2	8
8	0	0	0	0	3	3	2	0	0
9	0	0	0	0	0	0	8	0	0

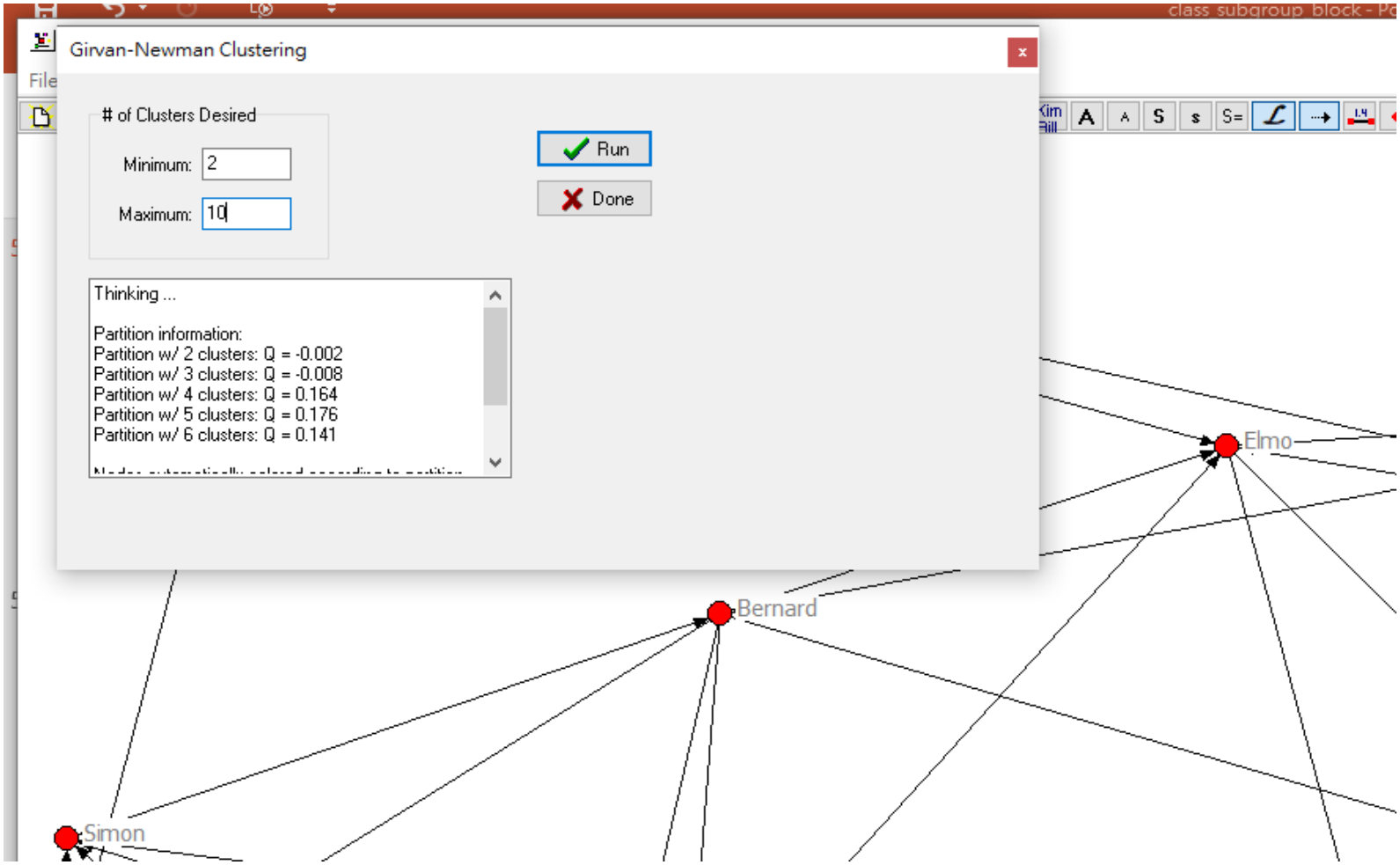


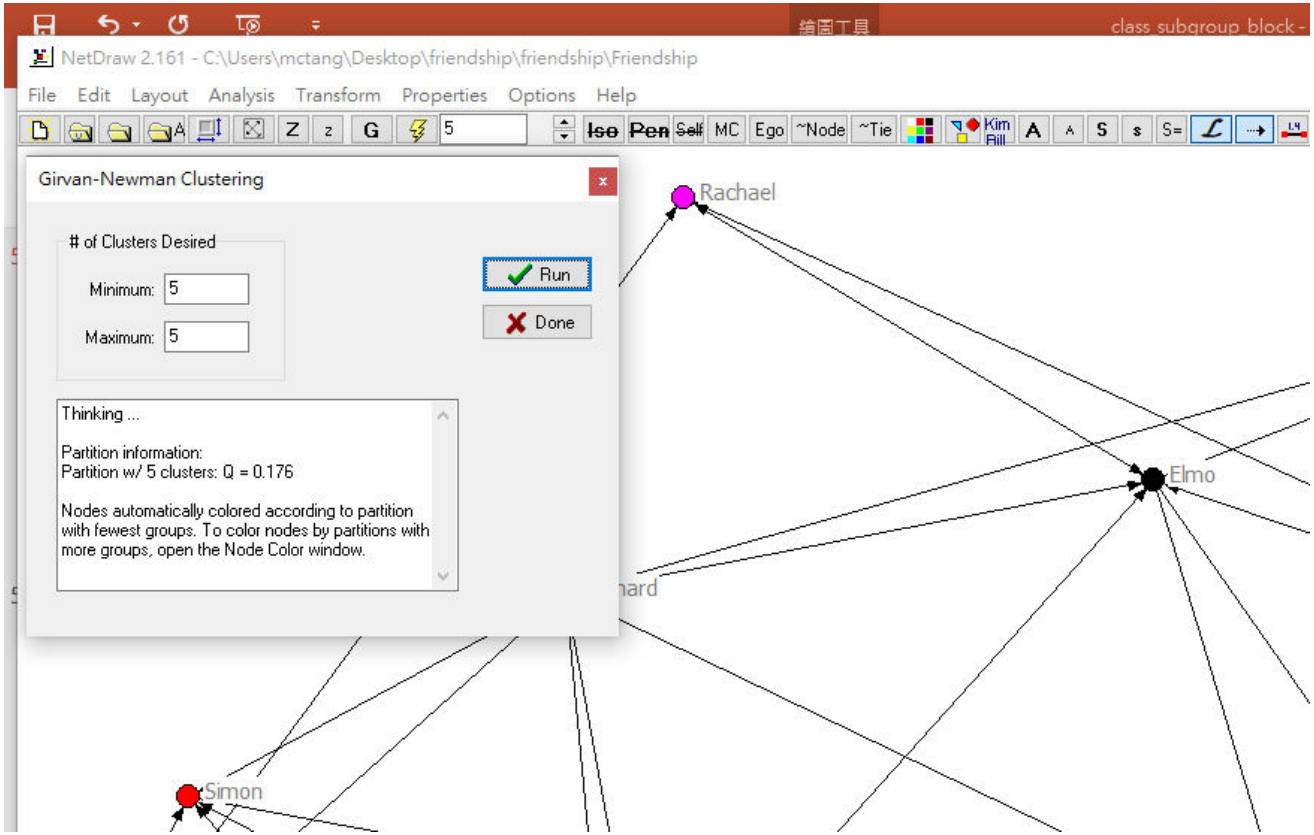
Recalculate betweenness, and again
Remove the edge(s) with highest betweenness

Exercise

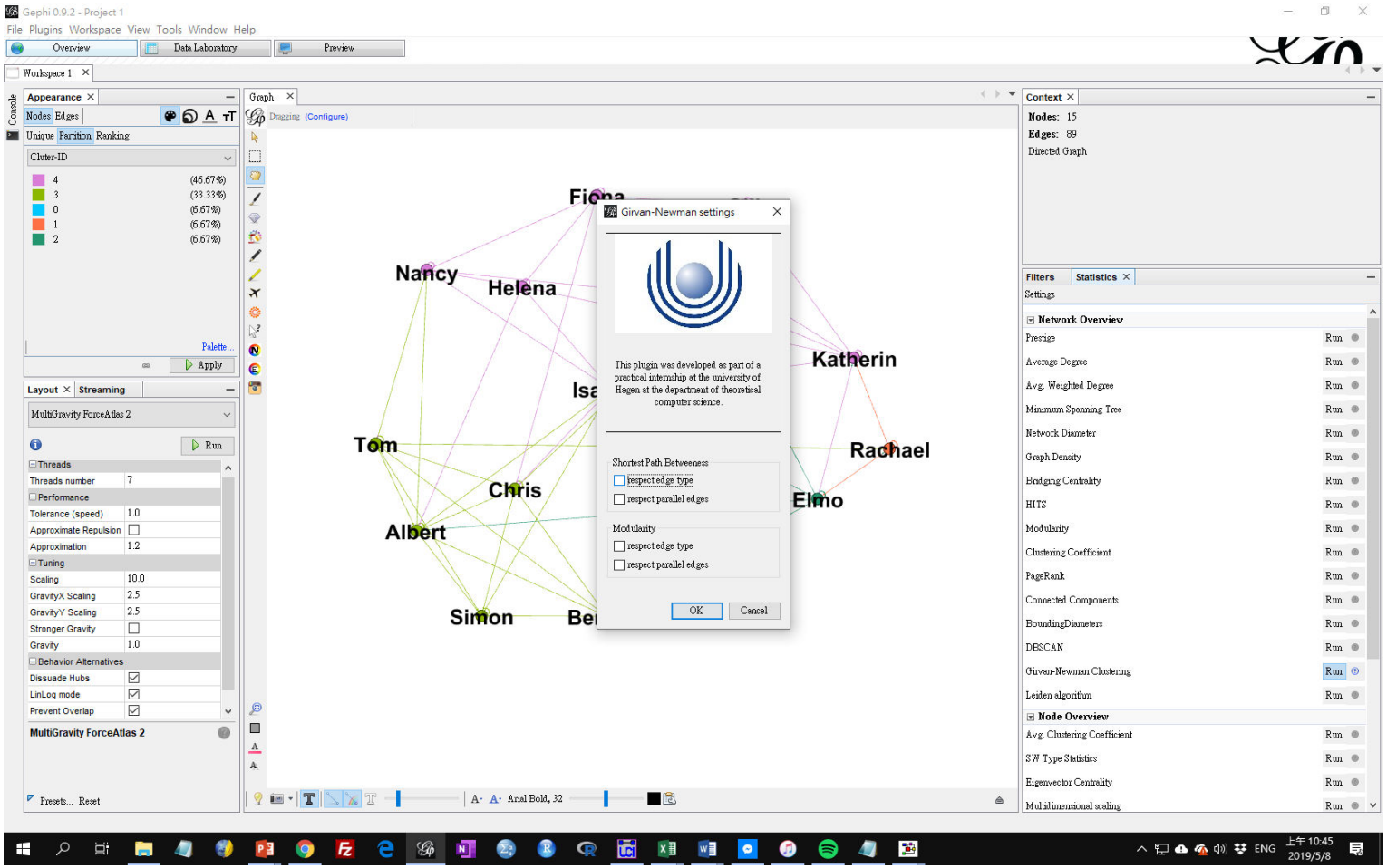
- Perform Girvan-Newman clustering and modularity using
  - UCINET AND NetDraw on "friendship"
    - The routine can be performed in UCINET using both
      - Girvan-Newman and
        - » Notice "Data were symmetrized via the maximum method"
        - » Perform your own symmertrization if you have a directional network
      - Newman Community Partition (New in 2019)
  - Gephi on friendship data
    - Compare the result of clicking and non-clicking "respect edge type"
  - Compare the results of Girvan-Newman and actor by actor clique co-membership results and Faction (in NetDraw, setting the partition at 5)

Identify the highest Q value

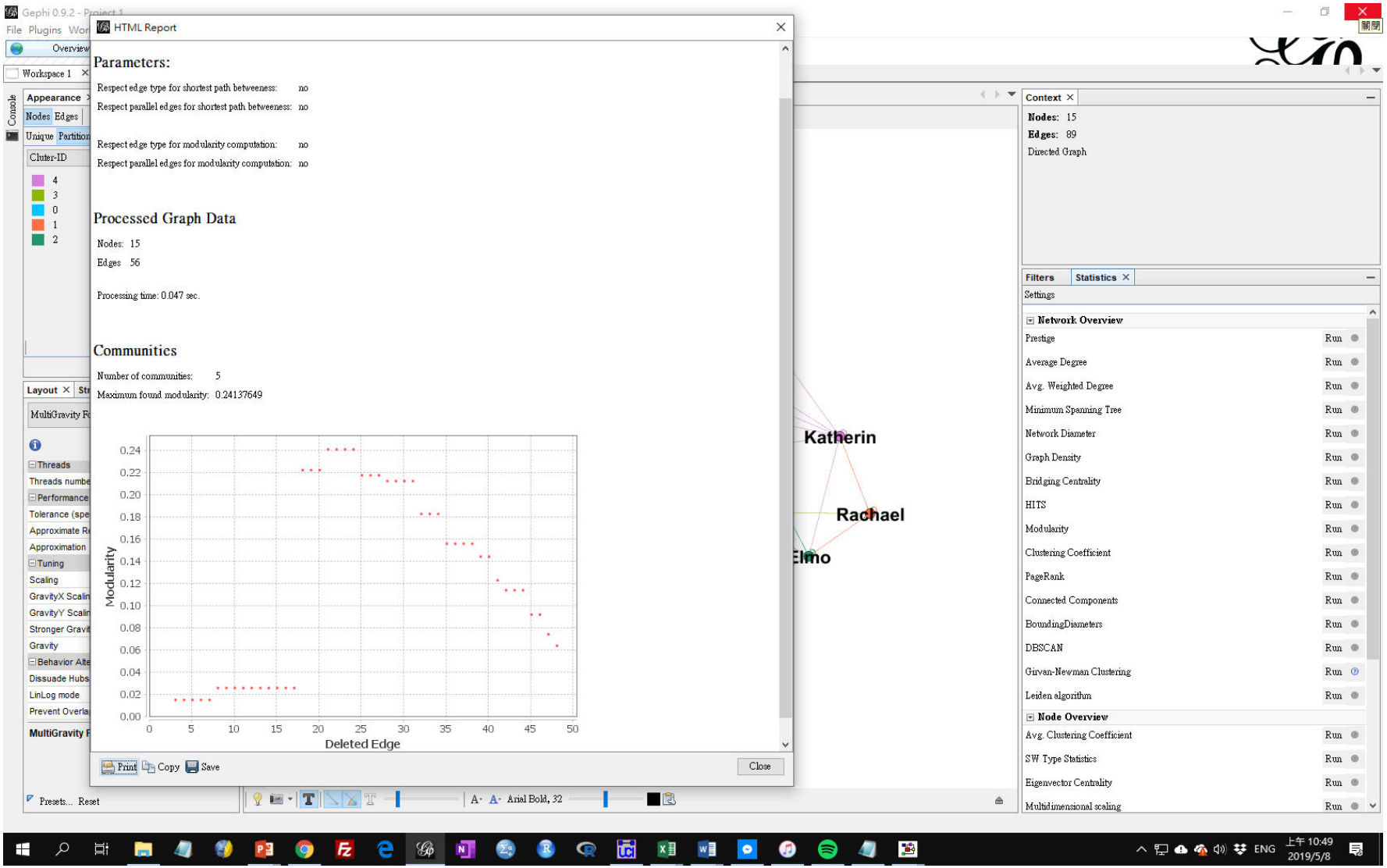



class subgroup block -
NetDraw 2.161 - C:\Users\mctang\Desktop\friendship\friendship\Friendship
File Edit Layout Analysis Transform Properties Options Help
Girvan-Newman Clustering
# of Clusters Desired
Minimum: 5
Maximum: 5
Run
Done
Thinking ...
Partition information:
Partition w/ 5 clusters: Q = 0.176
Nodes automatically colored according to partition with fewest groups. To color nodes by partitions with more groups, open the Node Color window.
Rachael
Elmo
Simon

In Gephi (edge type?)

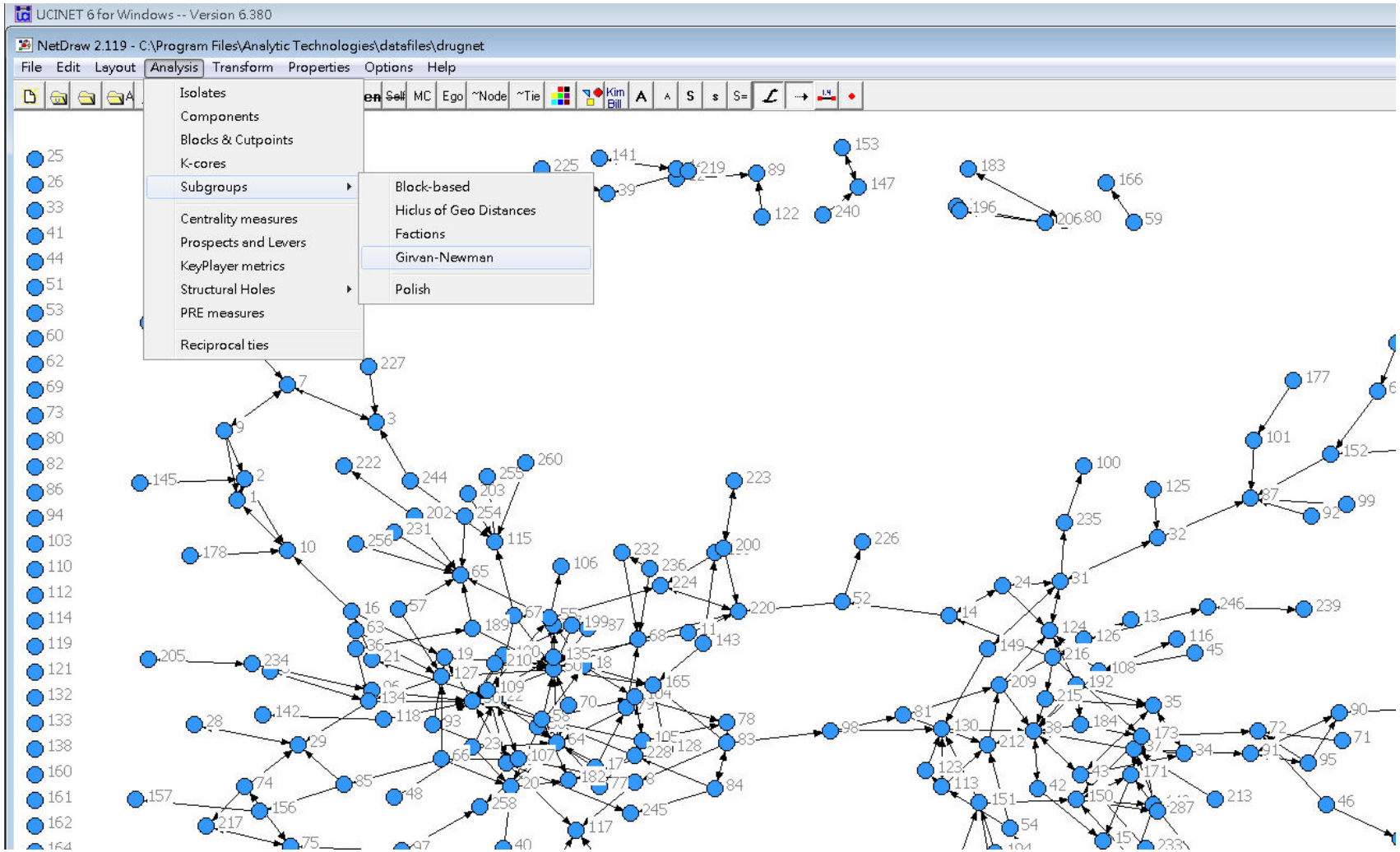


The highest Q value

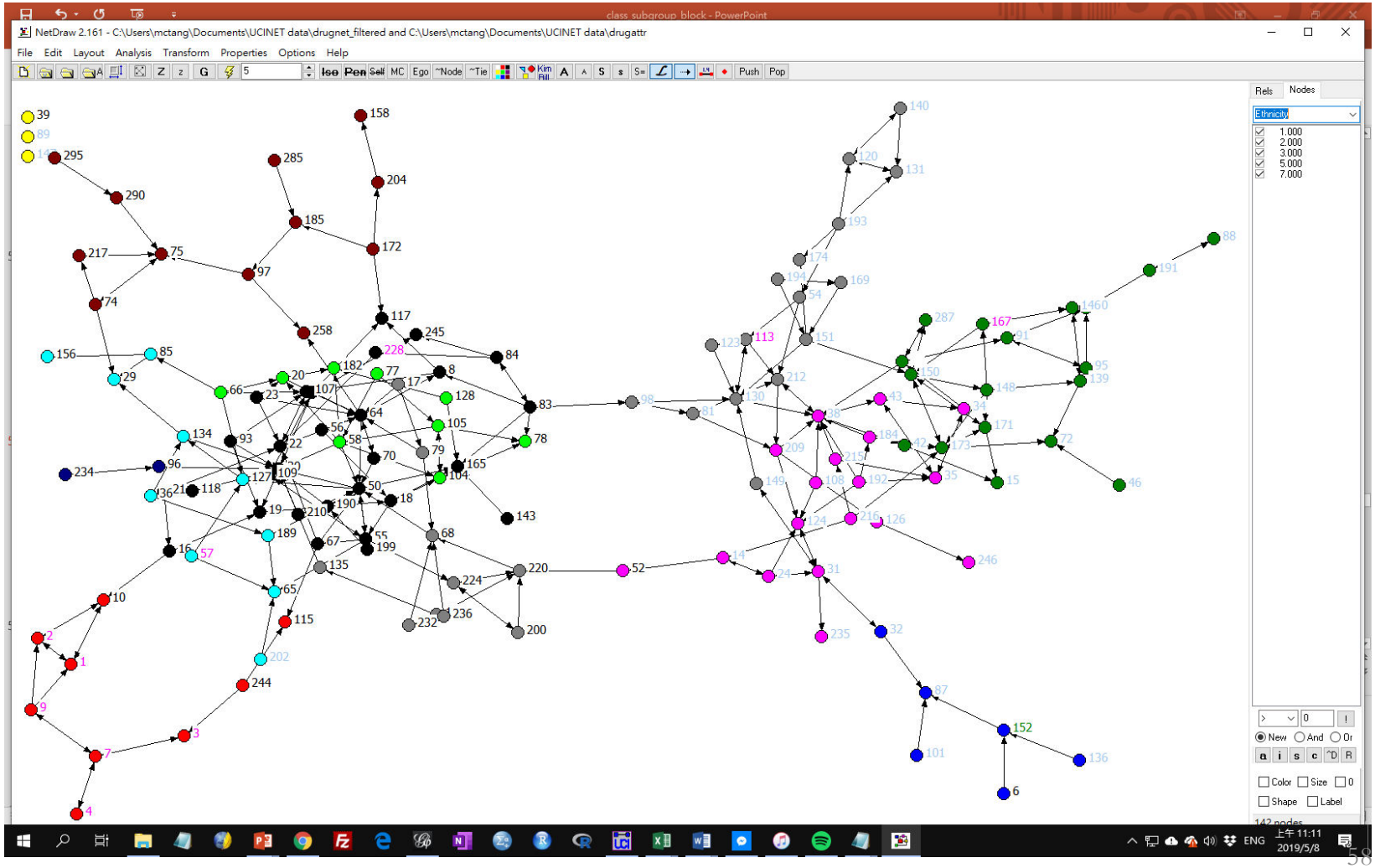


Girvan-Newman method in NetDraw

- Open "drugnet" in NetDraw
  - First remove isolated and pendants
  - And save data as drugnet_filtered
- NetDraw>Analysis>Subgroup>Girvan-Newman
- Determine the max and mim # of clusters
- Examine Q to determine the appropriate # of clusters
- Import drugatt
  - Compare visually group partitioned by GN and ethnicity
- Go to layout>group by attributes and select categorical – Ngpartition
- Can you do the same with Gephi?


UCINET 6 for Windows -- Version 6.380
NetDraw 2.119 - C:\Program Files\Analytic Technologies\datafiles\drugnet
File Edit Layout Analysis Transform Properties Options Help
Isolates
Components
Blocks & Cutpoints
K-cores
Subgroups
Centrality measures
Prospects and Levers
KeyPlayer metrics
Structural Holes
PRE measures
Reciprocal ties
Block-based
Hiclus of Geo Distances
Factions
Girvan-Newman
Polish

Is there homophile on the ground of ethnicity?



```
Number of errors: 27
Number of errors: 27
Number of errors: 27
Number of errors: 27
Number of errors: 27
Number of errors: 27

Final number of errors: 27

Group Assignments:

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

Grouped Adjacency Matrix

                        1
      1 2 8 4 5 7       0     6 3     9
      C C U I M N       W     W E     W
     ---------------------------------
  1 |   1       1 1 |     |       | 1 |
  2 | 1     1 1 1 1 |     |     1 | 1 |
  8 | 1 1     1 1 1 |     |       | 1 |
  4 | 1 1       1 1 |     |       |   |
  5 | 1 1 1 1     1 | 1   |     1 | 1 |
  7 |   1     1 1   |     |       |   |
    ---------------------------------
 10 | 1 1       1 1 |     |     1 |   |
    ---------------------------------
  6 |             1 |     |     1 | 1 |
  3 |   1     1 1 1 | 1   | 1     |   |
    ---------------------------------
  9 |   1       1 1 |     |       |   |
     ---------------------------------

Density Table

        1     2     3     4
      ---- ---- ---- ----
  1   0.83 0.17 0.17 0.67
  2   0.67      0.50 0.00
  3   0.42 0.50 1.00 0.50
  4   0.50 0.00 0.00
```

Eta is the correlation between the data matrix and an ideal structure matrix in which x(i,j)=1 if i and j are in the same cluster and 0 otherwise.

Lambda sets and bridge

- Bridges: if there are certain connections in the graph which, if removed, would result in a disconnected structure
- Lambda set approach: ranks each of the relationships in the network in terms of importance by evaluating how much of the flow among actors go through each link

Lambda set

- A maximal subset of actors who have more edge-independent paths connecting them to each other (within-group linkages) than to outsiders (between-group linkages)
- *Network>Subgroups> Lambda Set*
  - *Locate the vulnerable bridges between pairs of actors*
  - *The more "traffic" a linkage carries, the more important it is*

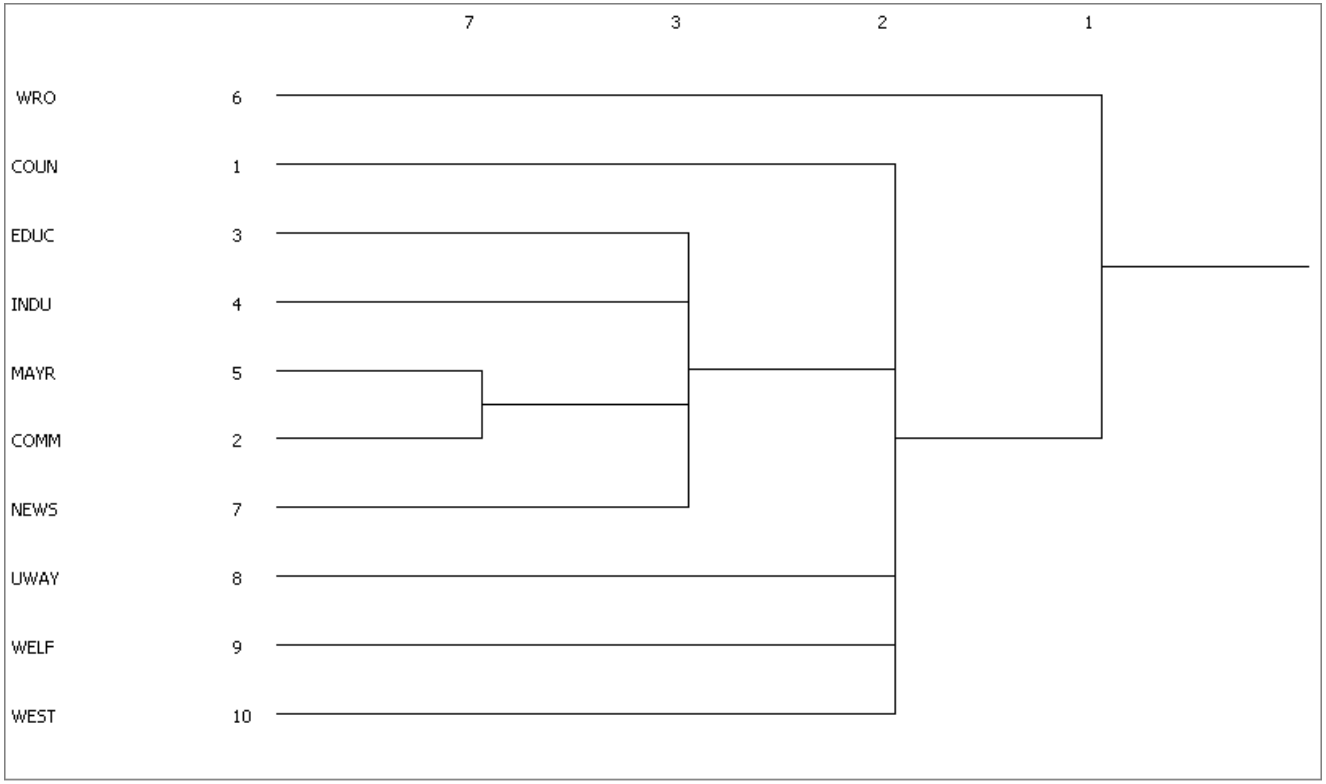
```
LAMBDA SETS

HIERARCHICAL LAMBDA SET PARTITIONS

            U W C E I M C N W
          W W E O D N A O E E
          R A L U U D Y M W S
          O Y F N C U R M S T

                            1
Lambda    6 8 9 1 3 4 5 2 7 0
------    - - - - - - - - - -
     7    . . . . . . XXX . .
     3    . . . XXXXXXXXXXXXX
     2    . XXXXXXXXXXXXXXXXX
     1    XXXXXXXXXXXXXXXXXXX
```

the connectivity within each set is given by the level of clustering

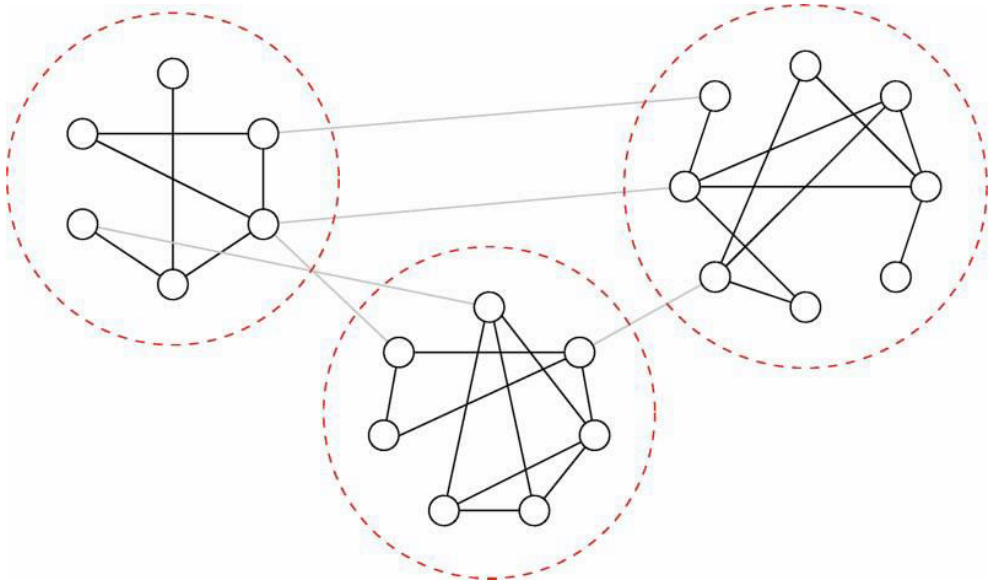
Every cluster is a lambda set

Identify #2 and #5 linkage as the most important – in the sense that it carries a great deal of traffic (i.e. independent paths travel though), and the graph would be most disrupted if it were removed...

Dendrogram


7
3
2
1
WRO 6
COUN 1
EDUC 3
INDU 4
MAYR 5
COMM 2
NEWS 7
UWAY 8
WELF 9
WEST 10

Goodness of classification: within and between linkages



How many subgroups/communities?
How good a particular division is?
E-I index; Q index

Indexes for goodness of classification

- E-I
- Q/Modularity index
  - Compare the numbers of internal links in the groups to how many you would expect if they were distributed at random
- Eta
  - Correlation between observed and ideal

```
************************************************************
Matrix KNOKI in dataset Dissim-KNOKBUR
************************************************************

HIERARCHICAL CLUSTERING

                                 1
 Level    6 3 2 5 8 4 7 1 9 0
------    - - - - - - - - - -
0.0000    . . . . . . . XXX .
1.0000    . . XXX . XXX XXX .
1.2222    . . XXX . XXXXXXX .
1.6667    . . XXXXX XXXXXXX .
2.1111    . . XXXXX XXXXXXXXX
2.8815    . . XXXXXXXXXXXXXX
3.3091    . XXXXXXXXXXXXXXXX
4.2158   XXXXXXXXXXXXXXXXXX
```

Hamming distance

Measures of cluster adequacy

		1	2	3	4	5	6	7	8
1	Eta	-0.308	-0.422	-0.578	-0.634	-0.687	-0.547	-0.487	
2	Q	-0.118	-0.143	-0.159	-0.173	-0.171	-0.074	-0.044	
3	Q-prime	-0.132	-0.167	-0.190	-0.216	-0.228	-0.111	-0.087	
4	E-I	1.000	0.970	0.894	0.848	0.727	0.015	-0.409	

Size of each cluster, expressed as a proportion of the total population clustered

		1	2	3	4	5	6	7	8
1	CL1	0.200	0.200	0.400	0.400	0.500	0.800	0.900	1.000
2	CL2	0.100	0.200	0.200	0.300	0.300	0.100	0.100	
3	CL3	0.100	0.100	0.100	0.100	0.100	0.100		
4	CL4	0.100	0.200	0.100	0.100	0.100			
5	CL5	0.100	0.100	0.100	0.100				
6	CL6	0.100	0.100	0.100					
7	CL7	0.100	0.100						
8	CL8	0.100							
9	CL9	0.100							
10	CL10								

	1 Freq	2 Pct	3 Possib	4 Densit
1 Internal	14.000	0.219	24.000	0.583
2 External	50.000	0.781	66.000	0.758
3 E-I	36.000	0.563	42.000	0.467

```
Max possible external ties: 66.000
Max possible internal ties: 24.000

E-I Index: 0.563
Expected value for E-I index is: 0.467

Max possible E-I given density & group sizes: 1.000
Min possible E-I given density & group sizes: 0.250

Re-scaled E-I index: -0.167
```

the expected value under a random distribution (given the size and density of the group)

Re-scaled E-I Index by the demographic constraints and overall density.

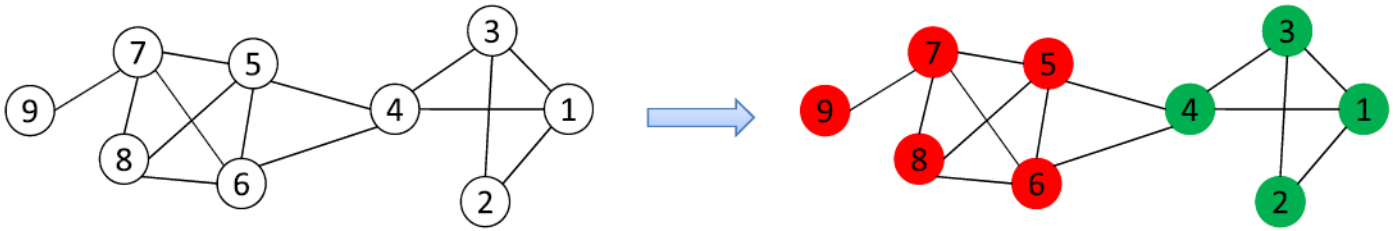
```
Permutation Test
Number of iterations = 5000
```

P=.203, not significant

	1 Obs	2 Min	3 Avg	4 Max	5 SD	6 P >= Ob	7 P <= Ob
1 Internal	0.219	0.625	0.733	0.844	0.039	1.000	0.000
2 External	0.781	0.156	0.267	0.375	0.039	0.000	1.000
3 E-I	0.563	0.250	0.467	0.688	0.078	0.203	0.953

Q index

Suppose we have a division of (1,2,3,4) and (5,6,7,8,9),



how good is the classi

M (# of edges) = 13

Hint: count the total number of “”1

Table 1.3: Adjacency Matrix

Node	1	2	3	4	5	6	7	8	9
1	-	1	1	1	0	0	0	0	0
2	1	-	1	0	0	0	0	0	0
3	1	1	-	1	0	0	0	0	0
4	1	0	1	-	1	1	0	0	0
5	0	0	0	1	-	1	1	1	0
6	0	0	0	1	1	-	1	1	0
7	0	0	0	0	1	1	-	1	1
8	0	0	0	0	1	1	1	-	0
9	0	0	0	0	0	0	1	0	-

only within-group edges are considered (“1” along the diagonal)

Q index

A good division of a network into communities is not merely one in which there are more edges within individual communities

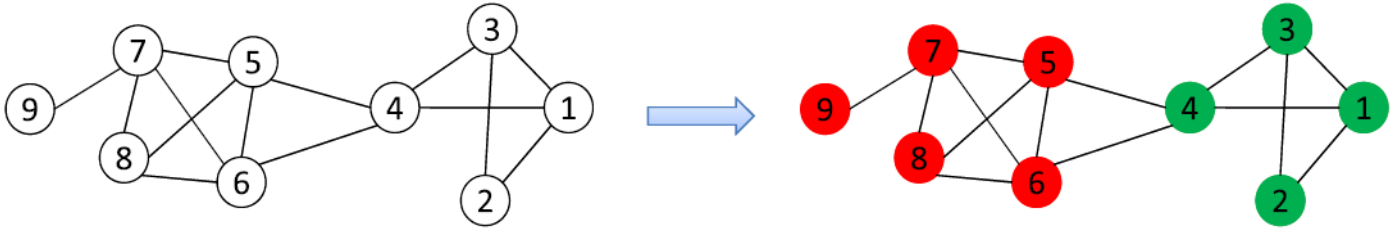
A good division is one where there are more than expected edges within the groups

Compare within group density
To that of a random graph

Table 1.3: Adjacency Matrix

Node	1	2	3	4	5	6	7	8	9
1	-	1	1	1	0	0	0	0	0
2	1	-	1	0	0	0	0	0	0
3	1	1	-	1	0	0	0	0	0
4	1	0	1	-	1	1	0	0	0
5	0	0	0	1	-	1	1	1	0
6	0	0	0	1	1	-	1	1	0
7	0	0	0	0	1	1	-	1	1
8	0	0	0	0	1	1	1	-	0
9	0	0	0	0	0	0	1	0	-

Q index

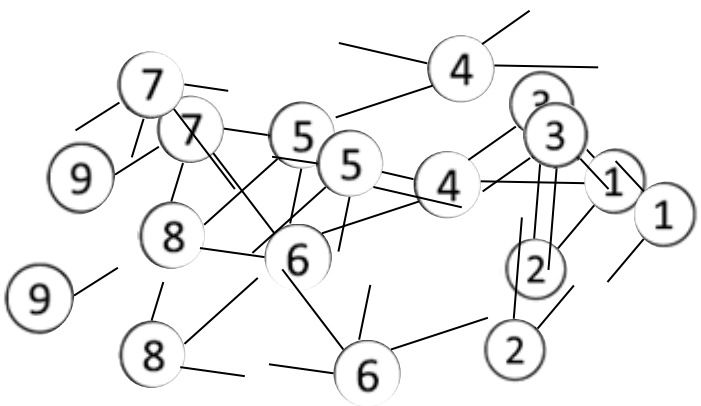


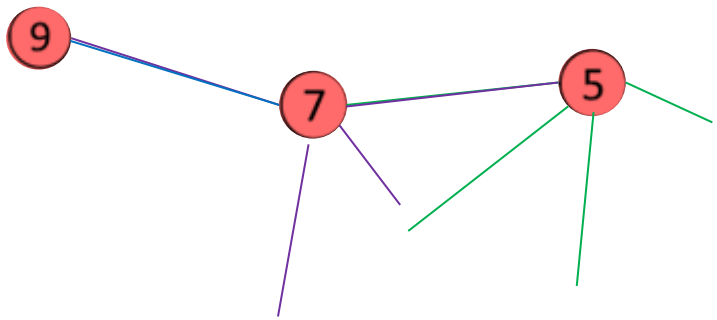
how good is the classification?
Is the link within the group greater than chance?

Table 1.3: Adjacency Matrix

Node	1	2	3	4	5	6	7	8	9
1	-	1	1	1	0	0	0	0	0
2	1	-	1	0	0	0	0	0	0
3	1	1	-	1	0	0	0	0	0
4	1	0	1	-	1	1	0	0	0
5	0	0	0	1	-	1	1	1	0
6	0	0	0	1	1	-	1	1	0
7	0	0	0	0	1	1	-	1	1
8	0	0	0	0	1	1	1	-	0
9	0	0	0	0	0	0	1	0	-

Modularity/Q index





edge (9, 7):1x 4/2x14

edge (5, 7): 4x 4/2x14

$$\text{Expectation of full edges between } v \text{ and } w = \frac{(\text{Full edges between } v \text{ and } w)}{(\text{Total number of rewiring possibilities})}$$

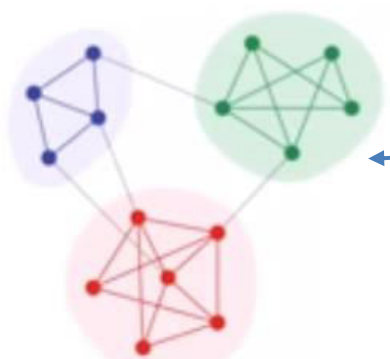
$$\sum_{i \in C, j \in C} A_{ij} - d_i d_j / 2m$$

- Let n_c - number of classes, c_i - class label per node
- Compare fraction of edges within the cluster to expected fraction if edges were distributed at random
- Modularity:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j), \quad \delta(c_i, c_j)\text{- kronecker delta}$$

"1" if i =j, otherwise "0"
In other words, only within group edges are considered ("1" along the diagonal)

Normalized, keep Q within 1



- The higher the modularity score - the better is community
- Modularity score range $Q \in [-1/2, 1)$
- Single class, $\delta(c_i, c_j) = 1$, $Q = 0$

Q index

- Measures the fraction of the links that connect nodes of the same type (within the same community) minus the expected value of the same quantity with the same community divisions but random connection between nodes
- A value of Q = 0 indicates that the community structure is no stronger than would be expected by random chance and values other than zero represent deviations from randomness
- 1 is maximum; typically range from 0.3 to 0.7

Modularity

- Modularity or Q
  - Measure the strength of a community partition by taking into account the degree distribution of nodes
  - Given a network of n nodes and m edges, the expected number of edges between node V_i and V_j is $d_i d_j / 2m$,

$$\sum_{i \in C, j \in C} A_{ij} - d_i d_j / 2m$$

Modularity maximization (Louvain) algorithm

- 1. A greedy stages followed aggregating stage
  - 1. Each node its own community
  - 2. For each node calculate Q when merged with adjacent nodes
  - 3. If this is positive, merged the node into the community with the largest gain (i.e. less likely to connect by chances)
  - 4. Repeated the process till its impossible to increase Q (greedy)
  - 5. second aggregating stage where the communities in previous stage becomes nodes, their edges determined by the sum of the weights of all the edges connecting each pair of communities
  - 6. Repeat the algorithm on communities resulted from the first greedy stage
  - 7. Repeat till no increase in Q in found

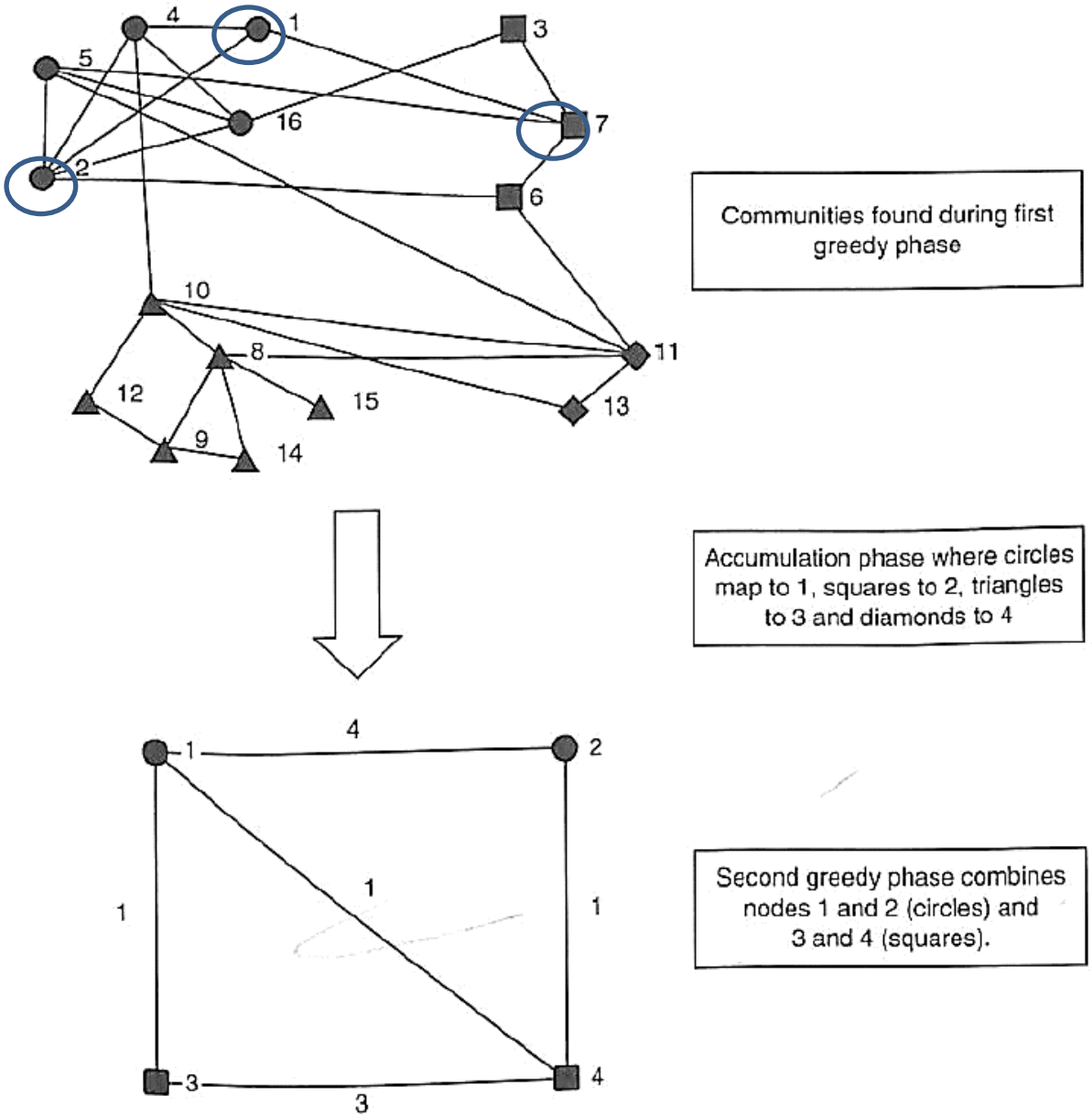
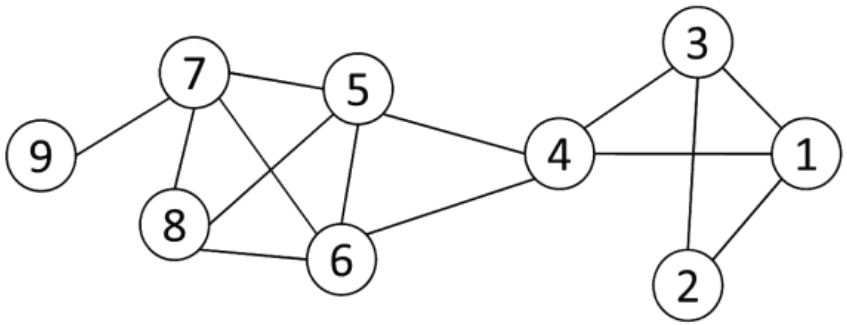


Figure 14.5 A simple visual representation of the Louvain algorithm.

Modularity maximization (Louvain) algorithm

- Start will all nodes as communities
- Follow a greedy strategy:
  - Successively join communities with the greatest increase in Q
  - Stop when the maximum possible increase in Q from joining any two
  - https://www.youtube.com/watch?v=dGa-TXpoPz8

Q index



$$\text{Expectation of full edges between } v \text{ and } w = \frac{(\text{Full edges between } v \text{ and } w)}{(\text{Total number of rewiring possibilities})}$$

$$\sum_{i \in C, j \in C} A_{ij} - d_i d_j / 2m$$

$$Q = \frac{1}{2m} \sum_{l=1}^{k} \sum_{j \in C_l, j \in C_l} \left(A_{ij} - d_i d_j / 2m\right)$$

$$\sum_{l=1}^{k} \sum_{i \in C_l, j \in C_l} \left(A_{ij} - d_i d_j / 2m\right)$$

Exercise

- Run modularity analysis using drugnet_flitered using in UCINET and Gephi
- Run modularity analysis using karate club in Gephi
  - Core-periphery??
    - What's the correlation?

Exercise:
Do a Girvan-Newman partition on aNobii popular book network after dichotomization
With UCINET and NetDraw
Use Q index to determine the best partition

Run modularity analysis in Gephi using aNobii popular book network Gephi data
Filtering?