Social Network Analysis

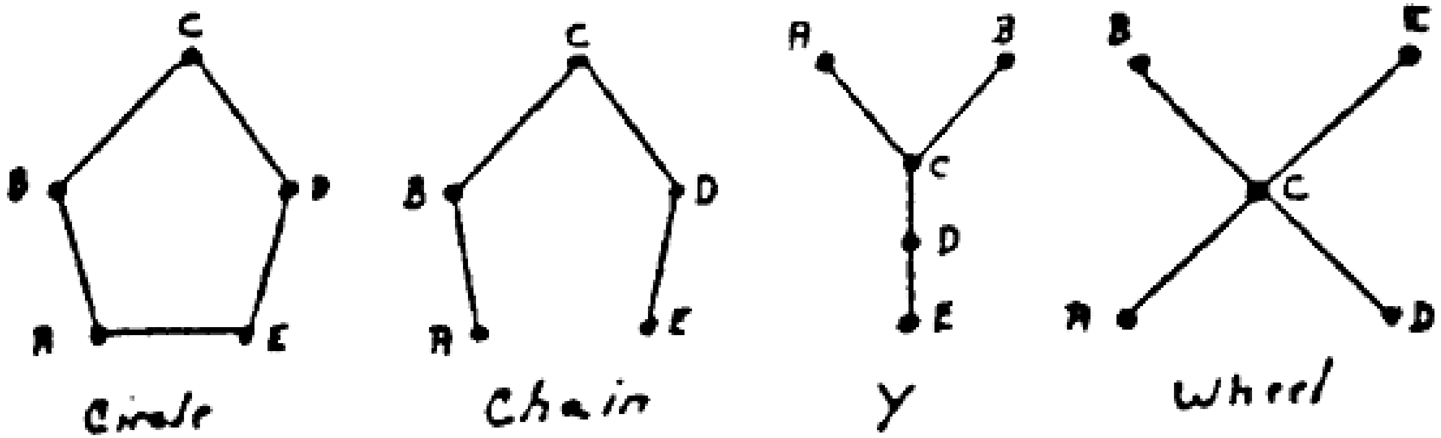
Day 3 - Global Properties

Measures of Group Cohesion

Whole Network Measures

- Density & Average degree
- Average Distance and Diameter
- Component measures (# & Ratio)
- Fragmentation (reachable & distance-weighted)
- Connectivity
- Centralization
- Core/Peripheriness

Bavelas-Leavitt experiments



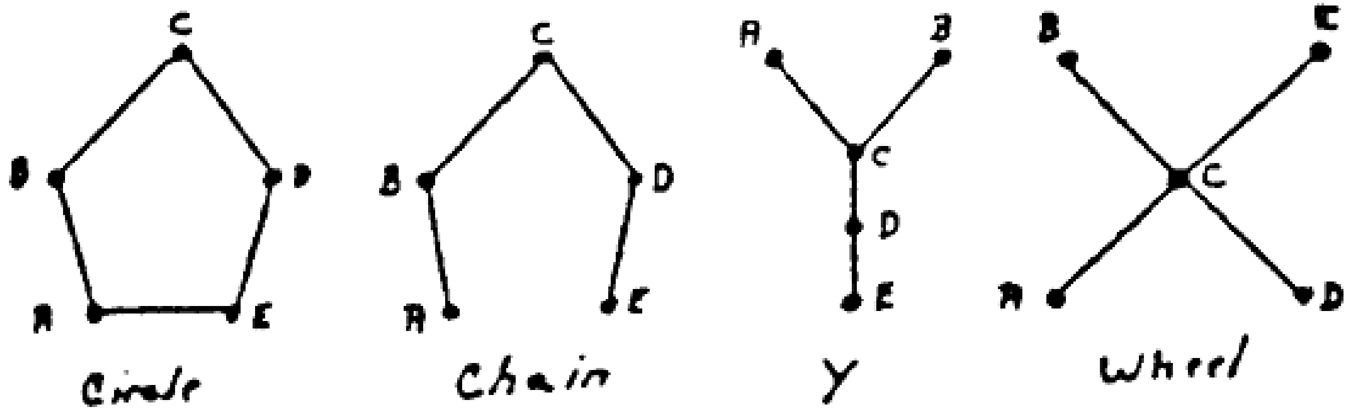
FPT*	3	5	4	5

*Fastest possible time in units of number of moves

Each person can only send one message at a time.

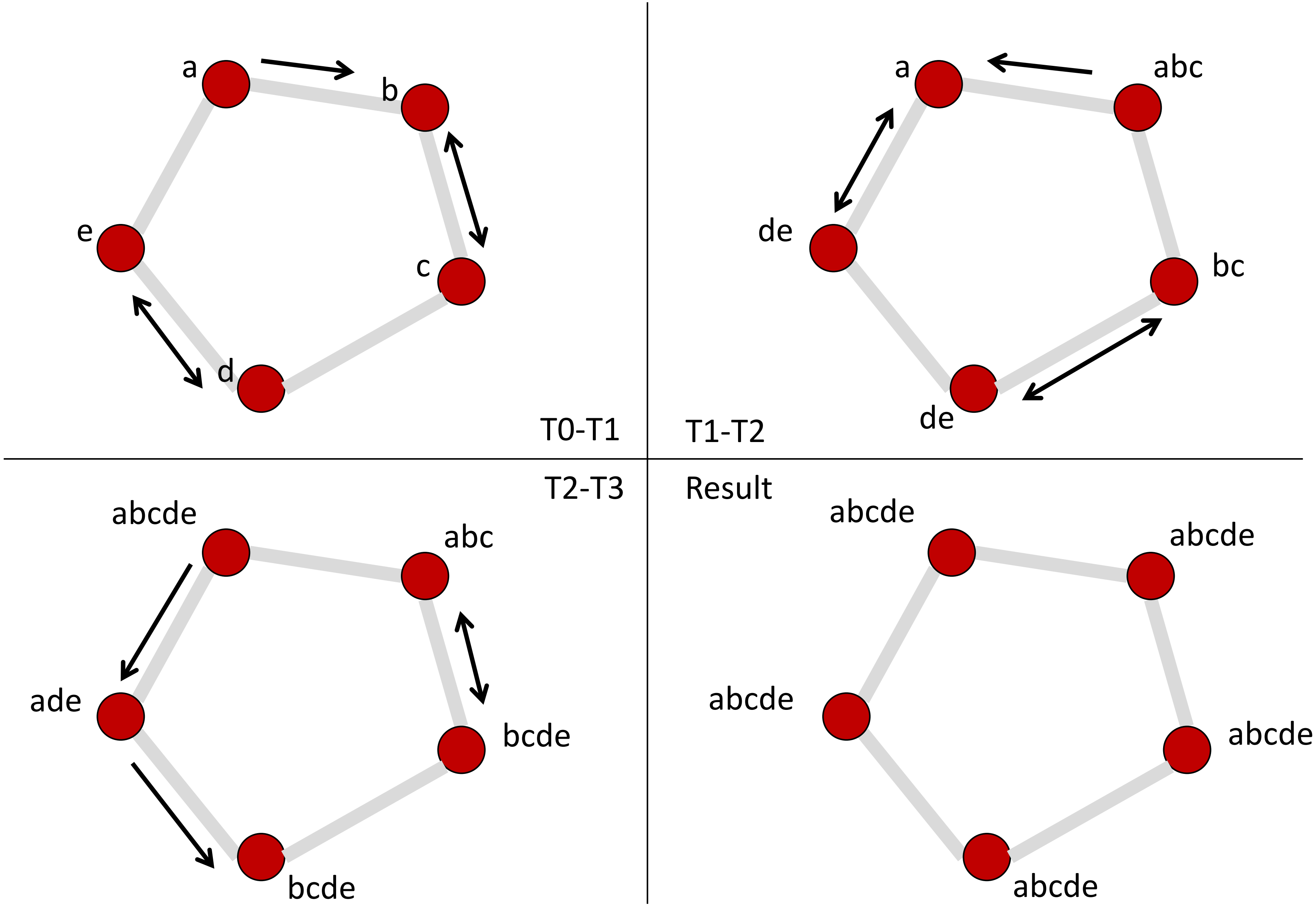
Bavelas-Leavitt experiments

Each person can only send one message at a time.



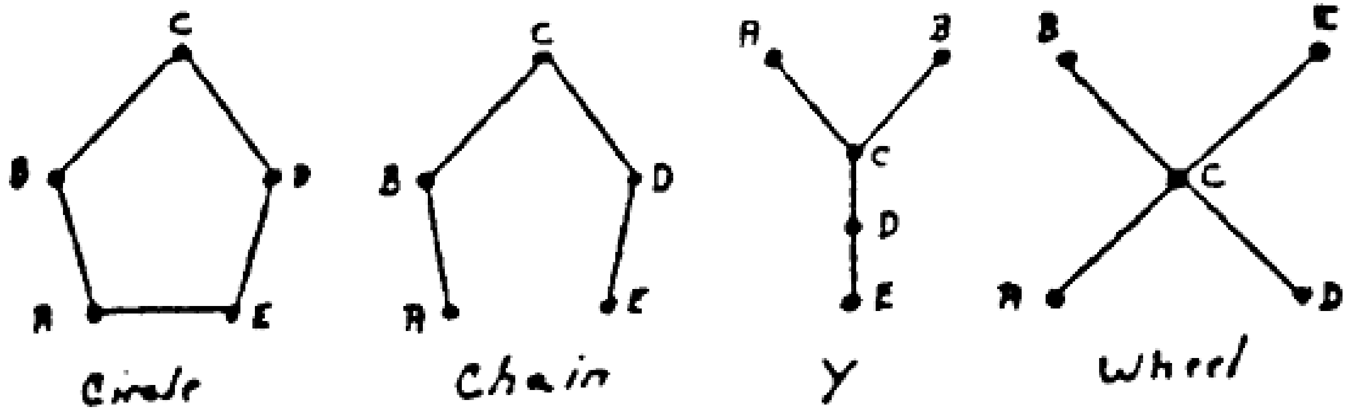
FPT*	3	5	4	5

*Fastest possible time in units of number of moves



Each person sends just one message aeceive multiple messages at one time.

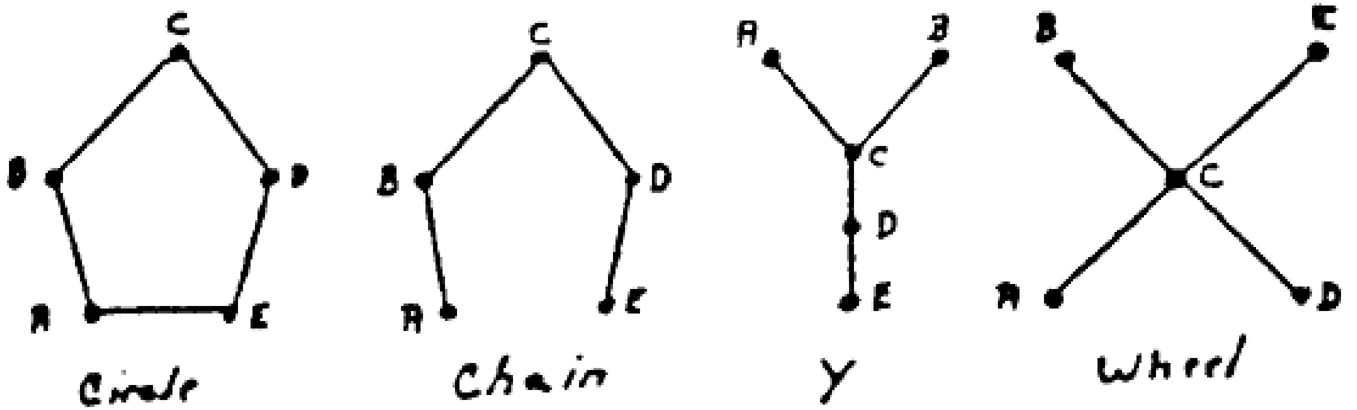
Bavelas-Leavitt experiments



FPT	3	5	4	5
Time	50.4	53.2	35.4	32
No. of errors	7.6	2.8	0	0.6
No. of msgs	high	low	low	low

Survey outcomes

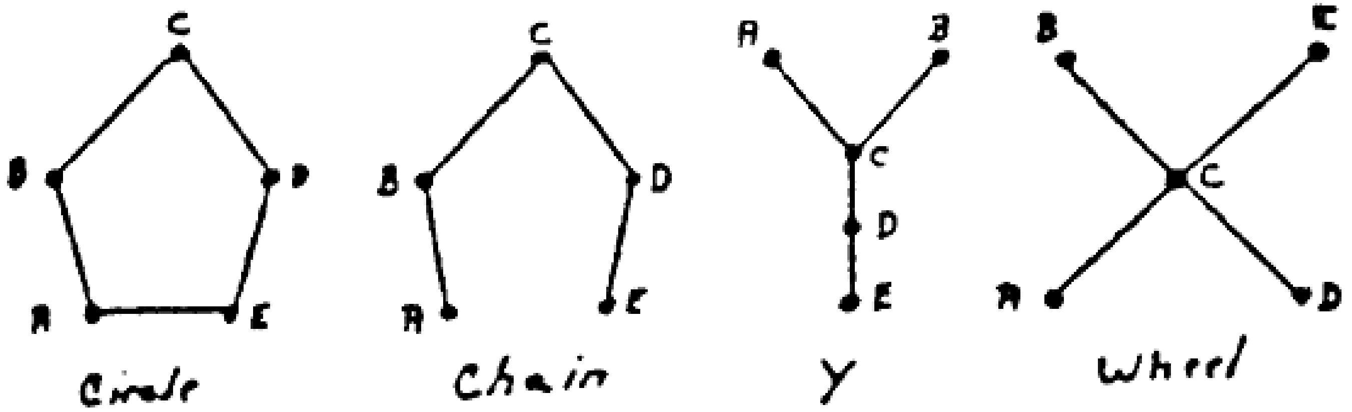
- Each person asked about “mr. green” or “mr. red”
- Satisfaction
  - Circle → wheel
- Organization
  - How to improve?
  - Did you have a system?
  - What was it?
- Leadership
  - Did you have a leader?
  - Who was it?



Bavelas-Leavitt interpretation

- In centralized networks, the distance from the "natural integrator"
- Centralization is good for simple, routine tasks

Total distance from "natural integrator"			
6	6	5	4



Measuring Bavelas centralization

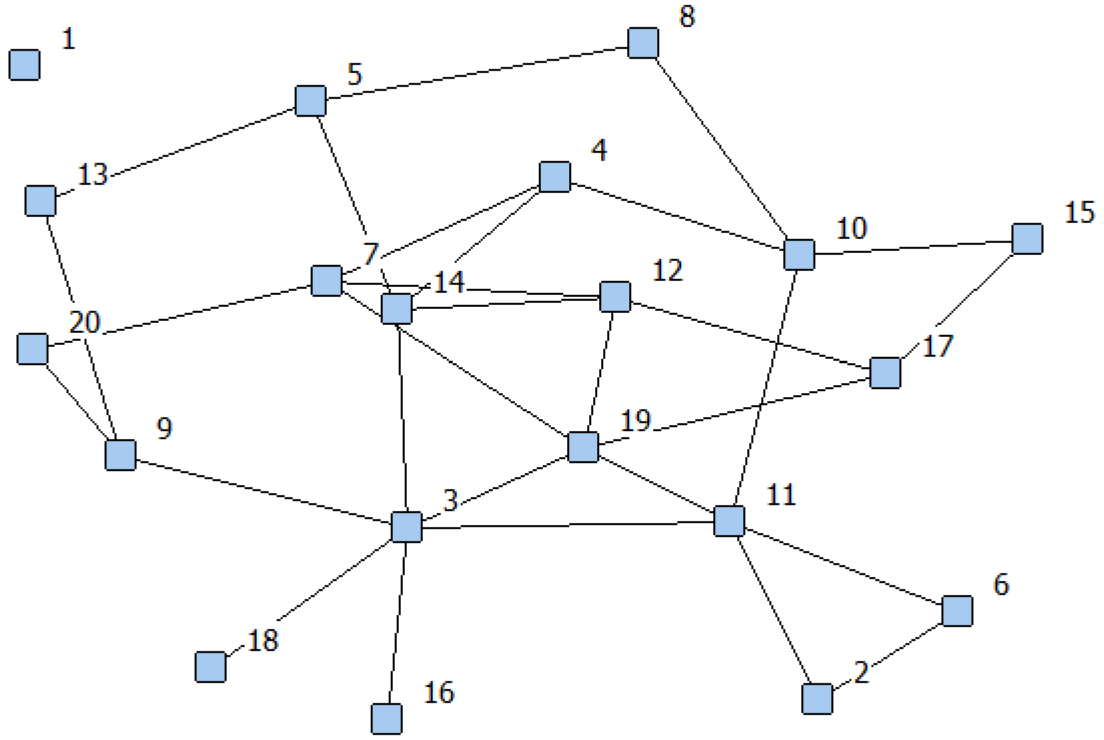
- Calculate graph-theoretic distances between every node and every other
- Find the node least far from all the others (e.g., smallest avg dist)
  - Call this the center
- Sum the the distances of every node to the center
  - This is Bavelas centralization
- See also Freeman's closeness centralization

Characterizing whole networks

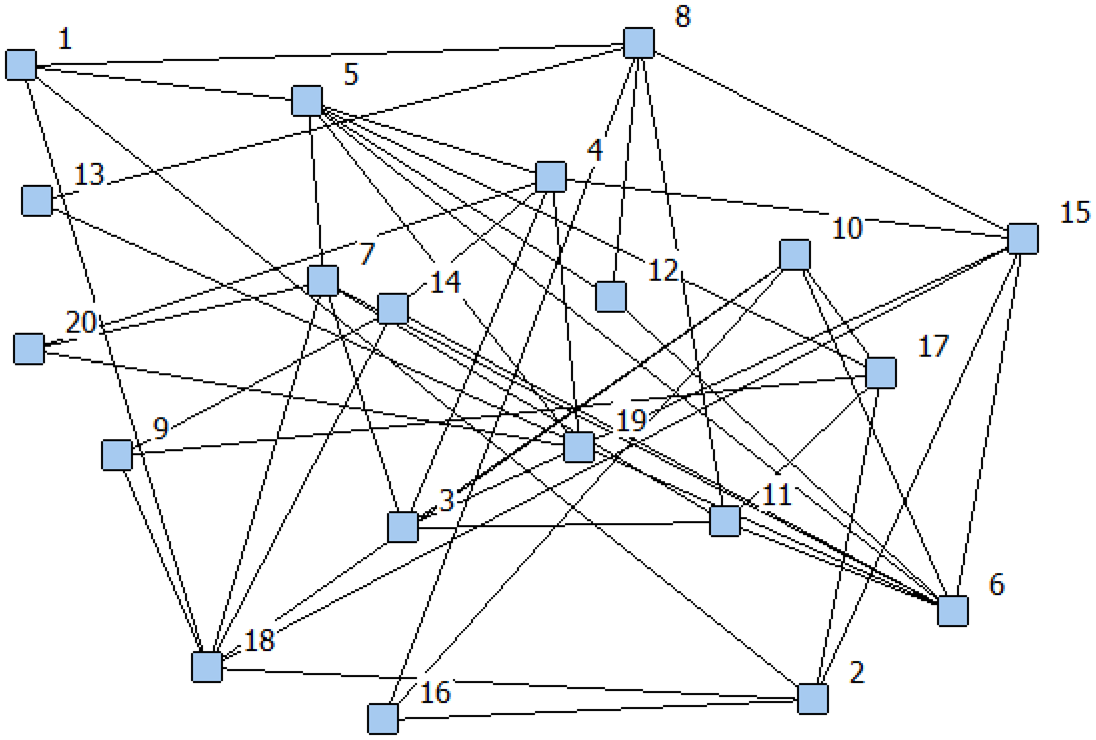
- Cohesion is biggest topic
  - Most measures of cohesion come from summarizing lower-level indices
    - E.g. average tie strength (aka density)
- There are also measures of shape
  - Many of these are “configural” in the sense that they are not simple aggregations of lower-level measures
    - E.g., core-periphery measures

Density

- Number of ties, expressed as proportion of # possible



Density = 0.15



Density = 0.25

Density

- Density is the number of ties in the network as a whole, expressed as proportion of # possible

	Reflexive	Non-Reflexive
Undirected	$=\frac{T}{n^2/2}$	$=\frac{T}{n(n-1)/2}$
Directed	$=\frac{T}{n^2}$	$=\frac{T}{n(n-1)}$

T = number of ties in network
n = number of nodes

Density as aggregated dyadic cohesion (or normalized node degree)

	HOLLY	BILL	DON	HARRY	MICHAEL	PAM	JENNIE	ANN	PAULINE	PAT	CAROL	LEE	JOHN	BRAZEY	GERY	STEVE	BERT	RUSS	Avg
HOLLY		0	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0.294
BILL	0		1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.176
DON	1	1		1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.235
HARRY	1	1	1		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.235
MICHAEL	1	1	1	1		0	0	0	0	0	0	0	0	0	1	0	0	0	0.294
PAM	1	0	0	0	0		1	1	1	0	1	0	0	0	0	0	0	0	0.294
JENNIE	0	0	0	0	0	1		1	0	1	0	0	0	0	0	0	0	0	0.176
ANN	0	0	0	0	0	1	1		1	0	0	0	0	0	0	0	0	0	0.176
PAULINE	0	0	0	0	0	1	0	1		1	1	0	1	0	0	0	0	0	0.294
PAT	1	0	0	0	0	0	1	0	1		1	0	0	0	0	0	0	0	0.235
CAROL	0	0	0	0	0	1	0	0	1	1		0	0	0	0	0	0	0	0.176
LEE	0	0	0	0	0	0	0	0	0	0	0		0	1	0	1	1	0	0.176
JOHN	0	0	0	0	0	0	0	0	1	0	0	0		0	1	0	0	1	0.176
BRAZEY	0	0	0	0	0	0	0	0	0	0	0	1	0		0	1	1	0	0.176
GERY	0	0	0	0	1	0	0	0	0	0	0	0	1	0		1	0	1	0.235
STEVE	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1		1	1	0.294
BERT	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1		1	0.235
RUSS	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1		0.235
Avg	0.29	0.18	0.24	0.24	0.29	0.29	0.18	0.18	0.29	0.24	0.18	0.18	0.18	0.18	0.24	0.29	0.24	0.24	0.229

Density tables

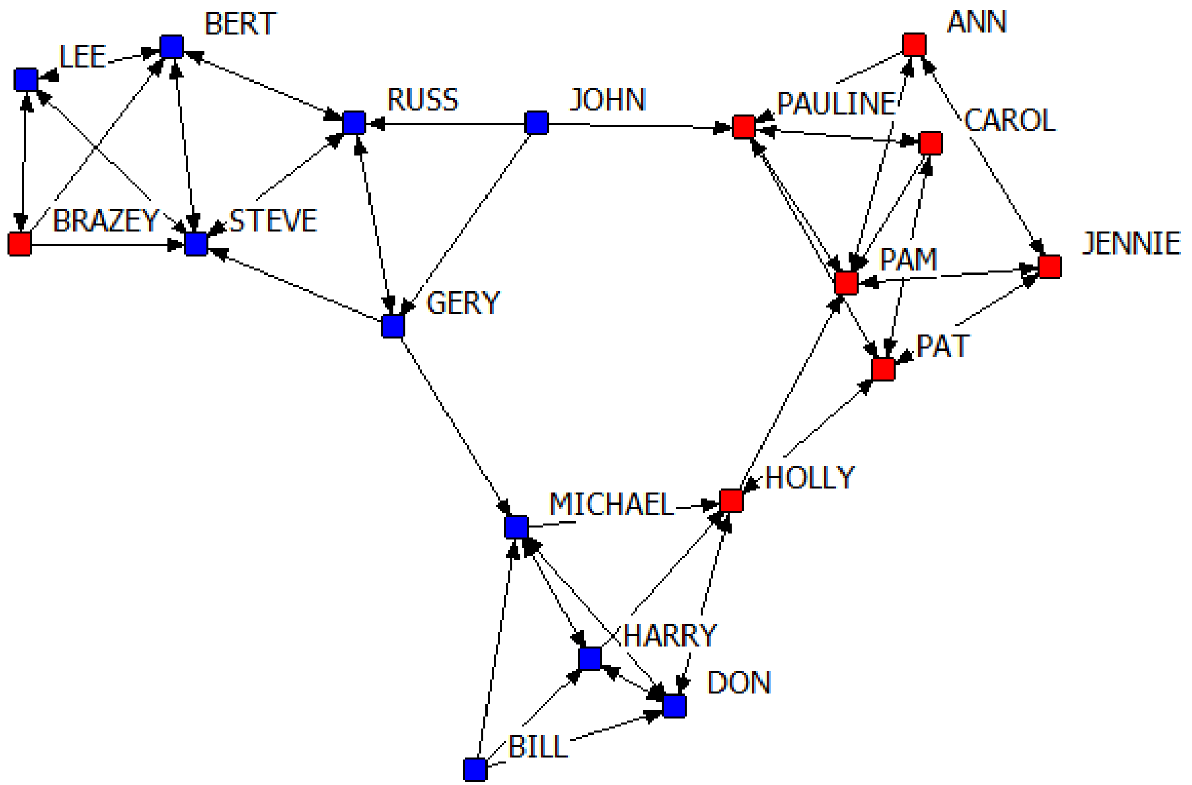
- Density of ties within and between *a priori* groups

Number of ties

	1	2
1	20	4
2	5	25

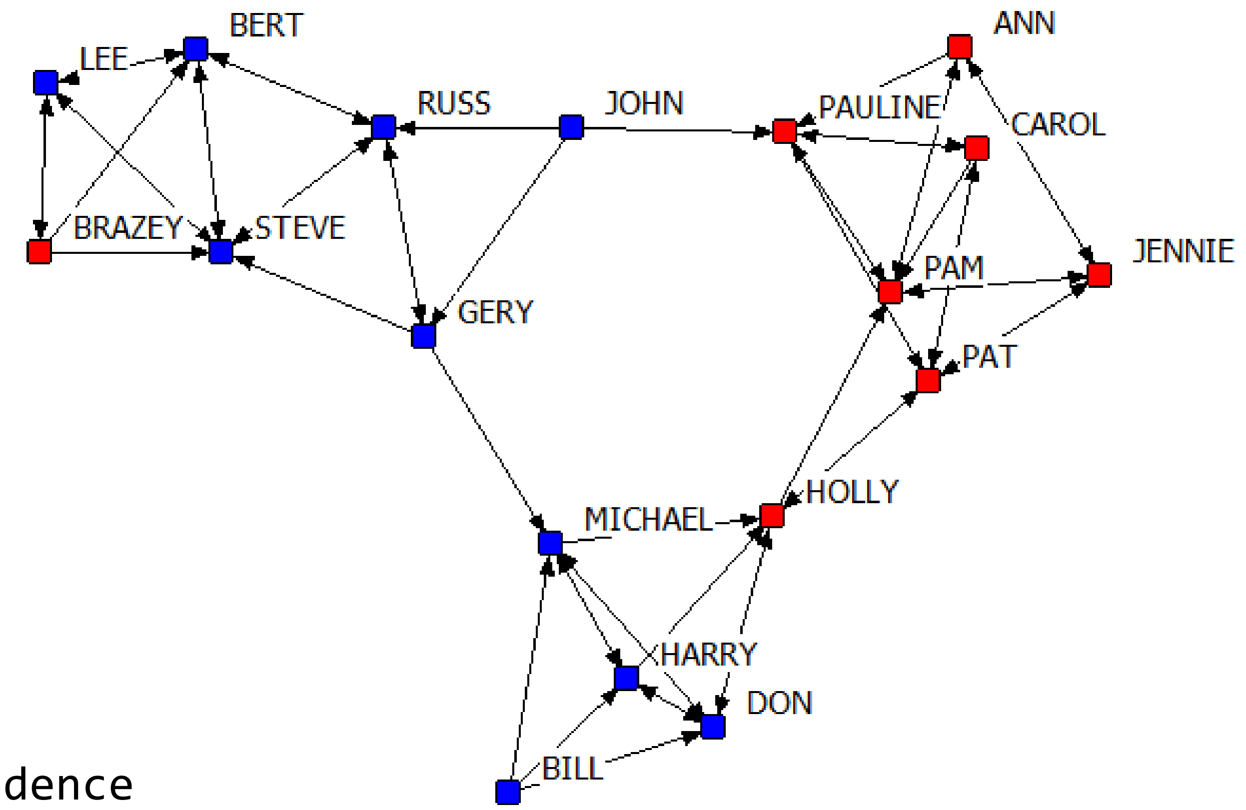
Density of ties

	1	2
1	0.357	0.050
2	0.063	0.278



Density tables

- Density of ties within and between *a priori* groups



Number of ties

	1	2
1	20	4
2	5	25

Expected Values Under Model of Independence

	1	2
1	9.88	14.12
2	14.12	15.88

Observed chi-square value = 28.732
Significance = 0.000100

“De-Energizing” Work Ties

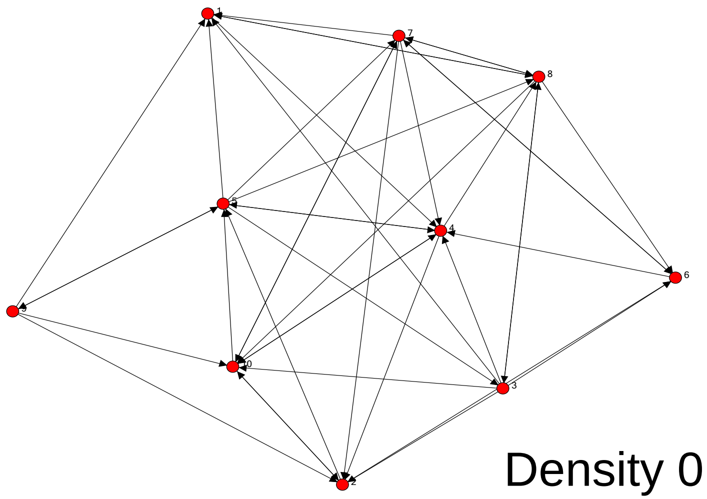
tie = "who tends to de-energize you?", run at a pizza supplier, symmetrized.

- Cross department Interactions
- 36 dept-to-dept work interaction pairs
- 7 pairs have >= 10% de-energizing work interactions
- Departments #6 and #9 have 50% de-energizing interactions between them

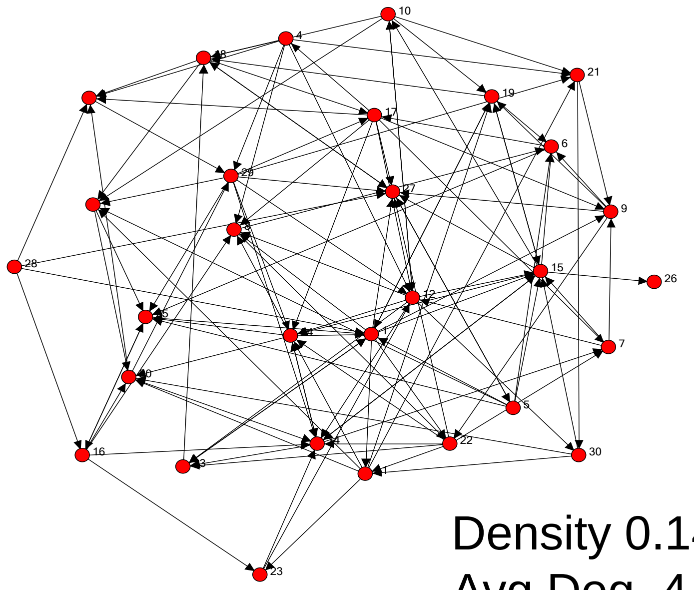
	1	2	3	4	5	6	7	8	9
1									
2	7%								
3	5%	0%							
4	5%	3%	2%						
5	0%	0%	6%	0%					
6	0%	0%	13%	0%	0%				
7	13%	2%	0%	3%	0%	11%			
8	0%	0%	7%	2%	6%	11%	14%		
9	9%	14%	0%	7%	0%	50%	0%	0%	

Average Degree

- Average number of links per person
- Is same as density*(n-1), where n is size of network
    - Density is just normalized avg degree
    - Sometimes more intuitive than density



Density 0.47
Avg Deg 4



Density 0.14
Avg Deg 4

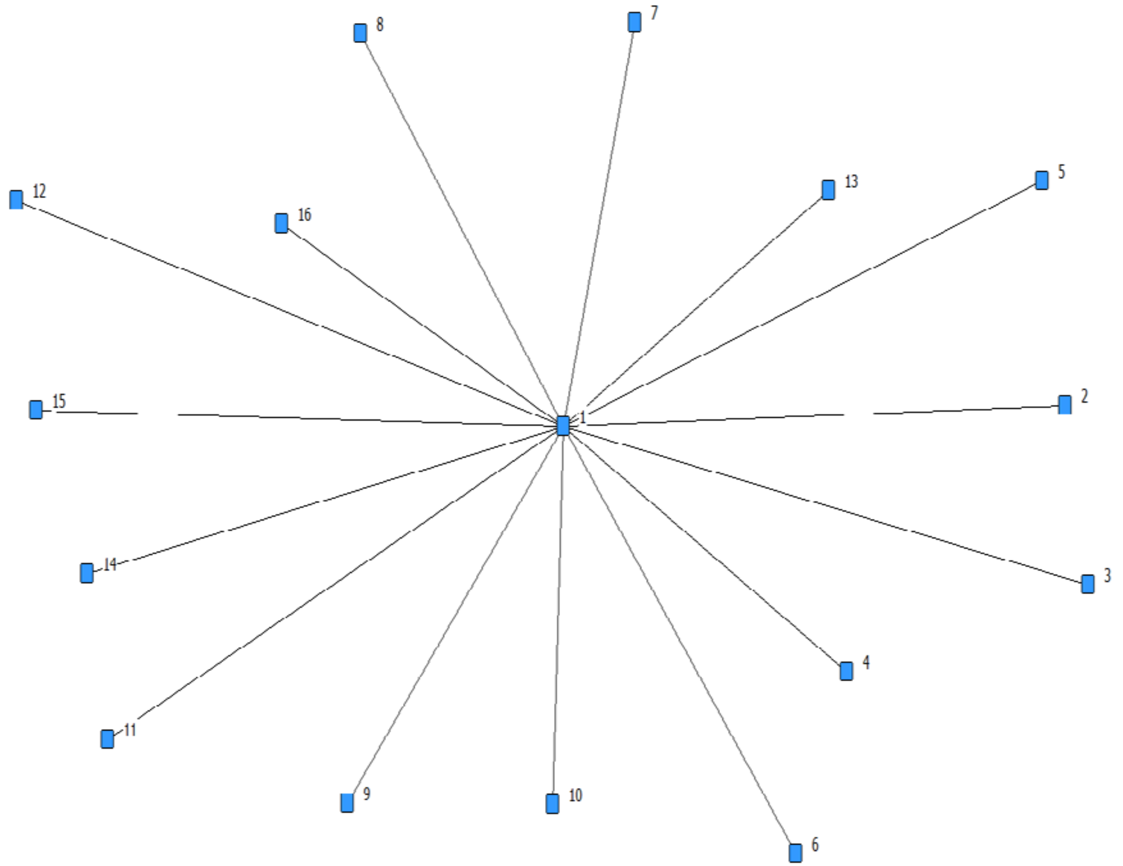
Degree variance and centralization

- **Variance** in degree (or any node level measure) indicates some people are much more central than others.
- Centralization is a kind of variance: the extent to which one person has all of the centrality
  - **Normal variance is variation around the mean**
    - i.e. sum of differences from the mean
  - **Centralization is variation around the maximum**
    - i.e. sum of differences (*squared*) from the maximum

id	Degree
17	14
16	11
7	5
15	5
13	4
1	3
8	3
9	3
11	3
14	3
2	2
3	2
4	2
5	2
6	2
10	2
12	2

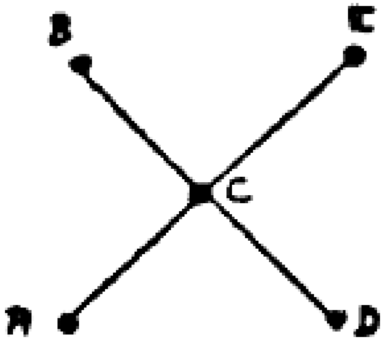
Centralization

- A network is maximally centralized with respect to any given node-level measure if the difference between the centrality of the most central node and that of all others is at a maximum
- For degree, it means the center is connected to all others, and they are only connected to the center

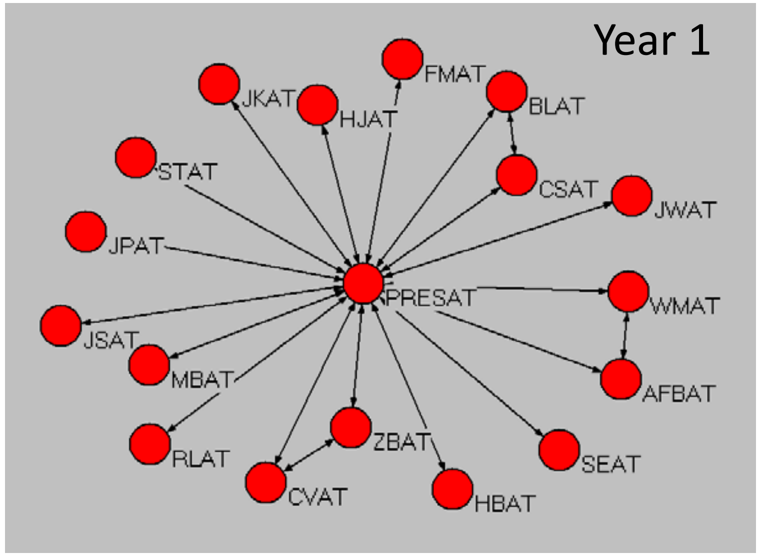


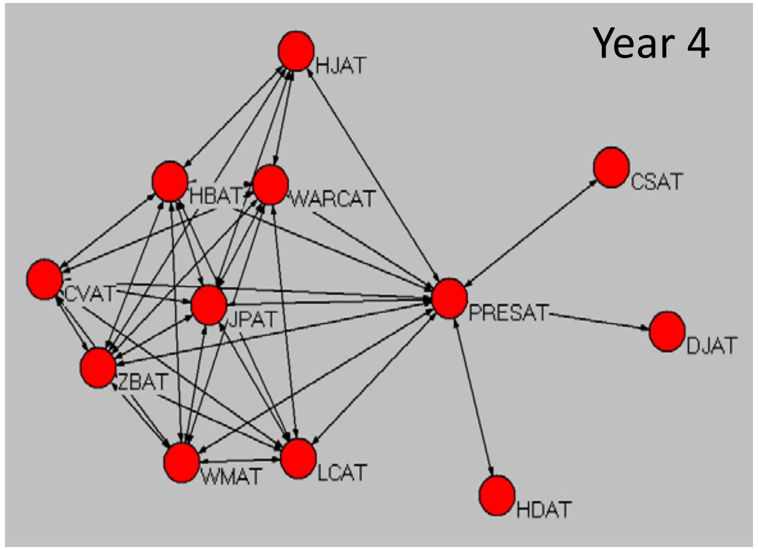
Calculating centralization

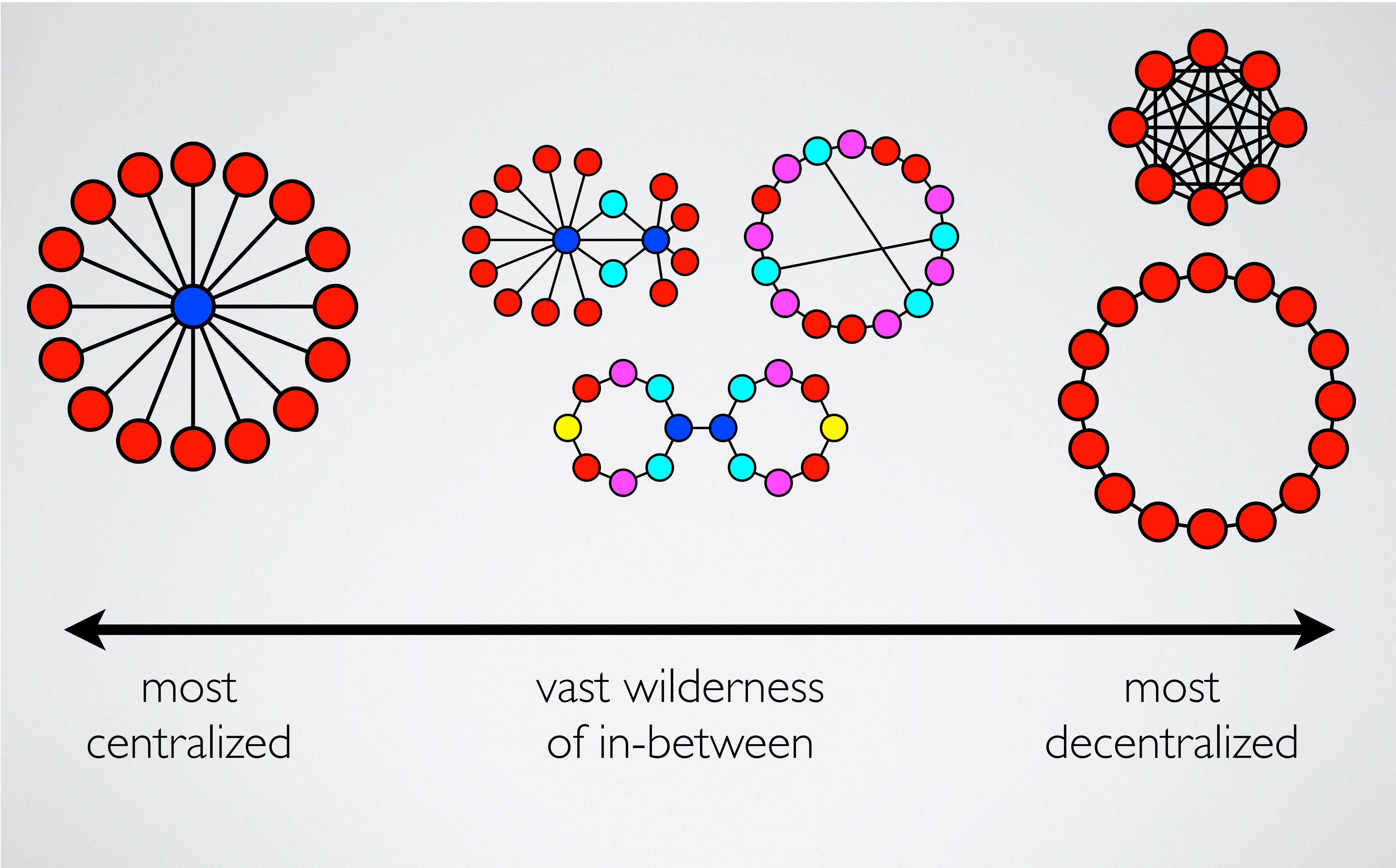
- Extent to which network revolves around a single node
- Sum of differences between the centrality of the most central node, and the centrality of every other node, divided by normalizing constant to make it run between 0 and 1
- Degree centralization:
  - $C = \frac{\sum_i d_{max} - d_i}{(n-1)(n-2)}$
  - (0+3+3+3+3)/(4*3) = 1.0



Carter admin.






most
centralized
vast wilderness
of in-between
most
decentralized

what have we learnt from it…

Baker & Faulkner (1993): Social Organization of conspiracy

(reconstructs communication networks in three well-known price-fixing conspiracies in the heavy electrical equipment industry to study social organization)

Questions: How are relations organized to facilitate illegal behavior?

Pattern of communication maximizes concealment, and predicts the criminal verdict.

Inter-organizational cooperation is common, but too much 'cooperation' can thwart market competition, leading to (illegal) market failure.

Illegal networks differ from legal networks, in that they must conceal their activity from outside agents. A "Secret society" should be organized to (a) remain concealed and (b) if discovered make it difficult to identify who is involved in the activity

The need for secrecy should lead conspirators to conceal their activities by creating ***sparse*** and ***decentralized*** networks.

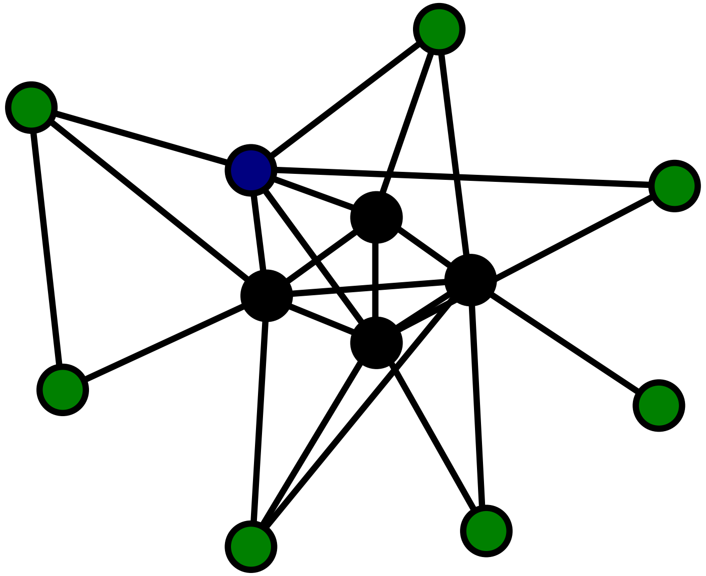
- reconstructs communication networks in three well-known price-fixing conspiracies in the heavy electrical equipment industry to study social organization;
- findings:
    - structure of illegal networks is driven by need to maximize concealment, rather than efficiency;
    - structure also contingent on information-processing requirements;
    - person centrality in networks predicts *verdict, sentence* and *fine.*

Organization Objective	Information-Processing Requirement	
	High	Low
Concealment	Centralized networks	Decentralized networks
Coordination	Decentralized networks	Centralized networks

Figure 1. Concealment Versus Coordination: Theoretical Expectations and experimental results

Core/Periphery

- Extent to which there is a “core” of people that holds the network together, such that
  - Core people are well connected to other core people, in general
  - Periphery people are connected to core people
  - Periphery people are NOT connected to other periphery people

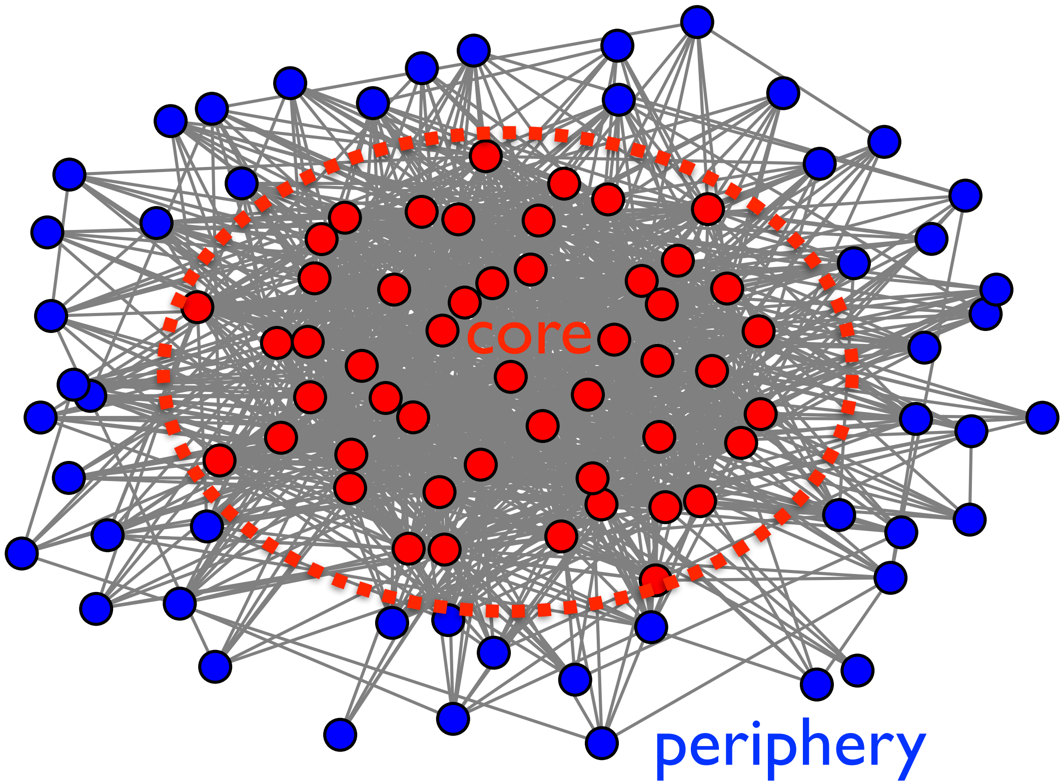


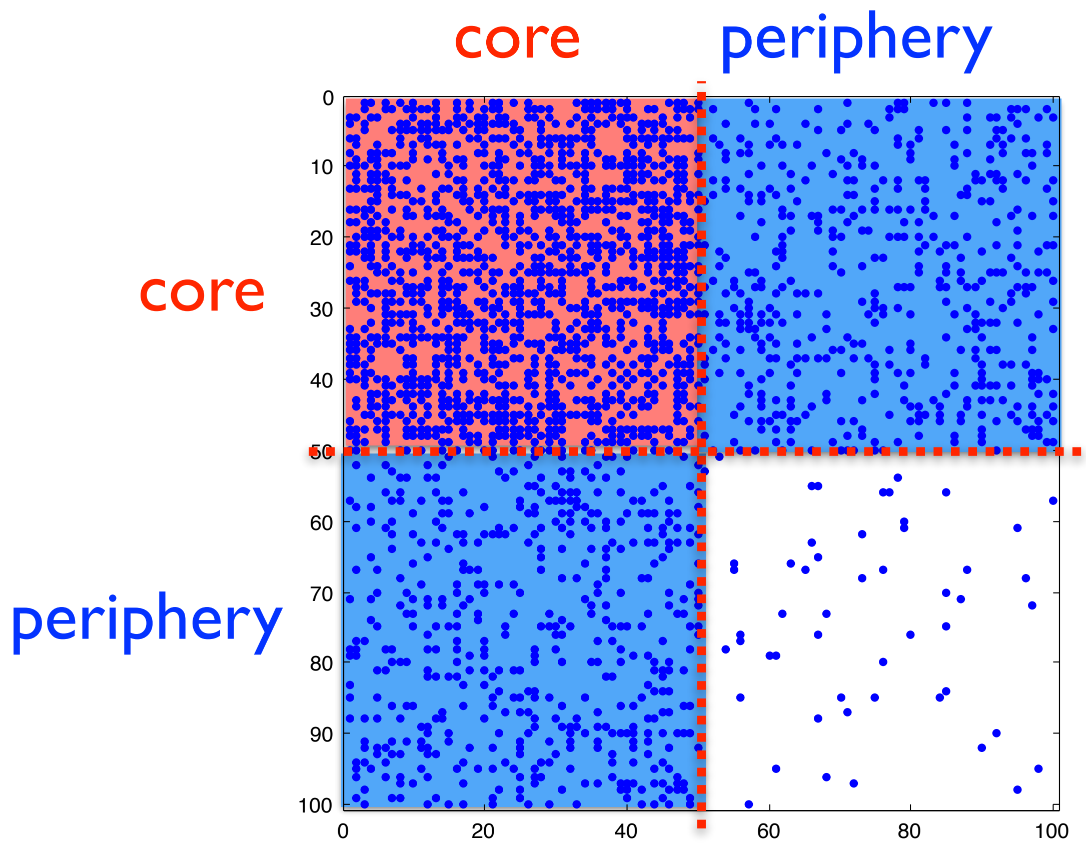
Core Periphery Block Model

Basic Idea:

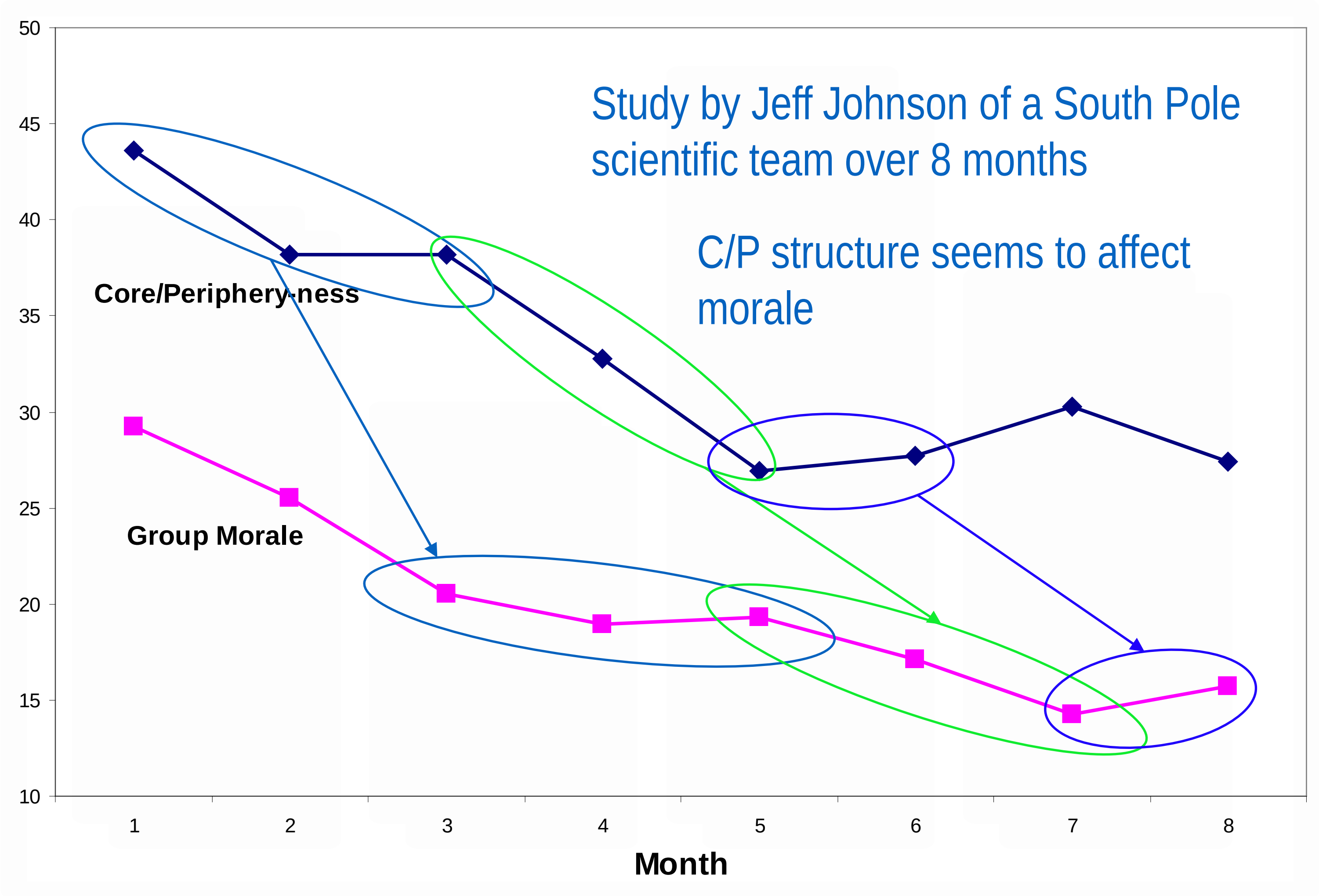
- A *module* or *community* is a collection of nodes defined by how its *edges* behave:
  - **Edge Density:** For social networks, we expect edge density to be greater within a community than without. (Assortative Community)
  - **Edge Weight:** For coexpression networks, we expect the correlations to be higher within a functional module than without.
  - Etc.

Finding Core/Periphery Structures





C/P Structures & Morale



Kapferer tailor shop data

Bruce Kapferer (1972) observed interactions in a tailor shop in Zambia (then Northern Rhodesia) over a period of ten months. His focus was the changing patterns of alliance among workers during extended negotiations for higher wages.

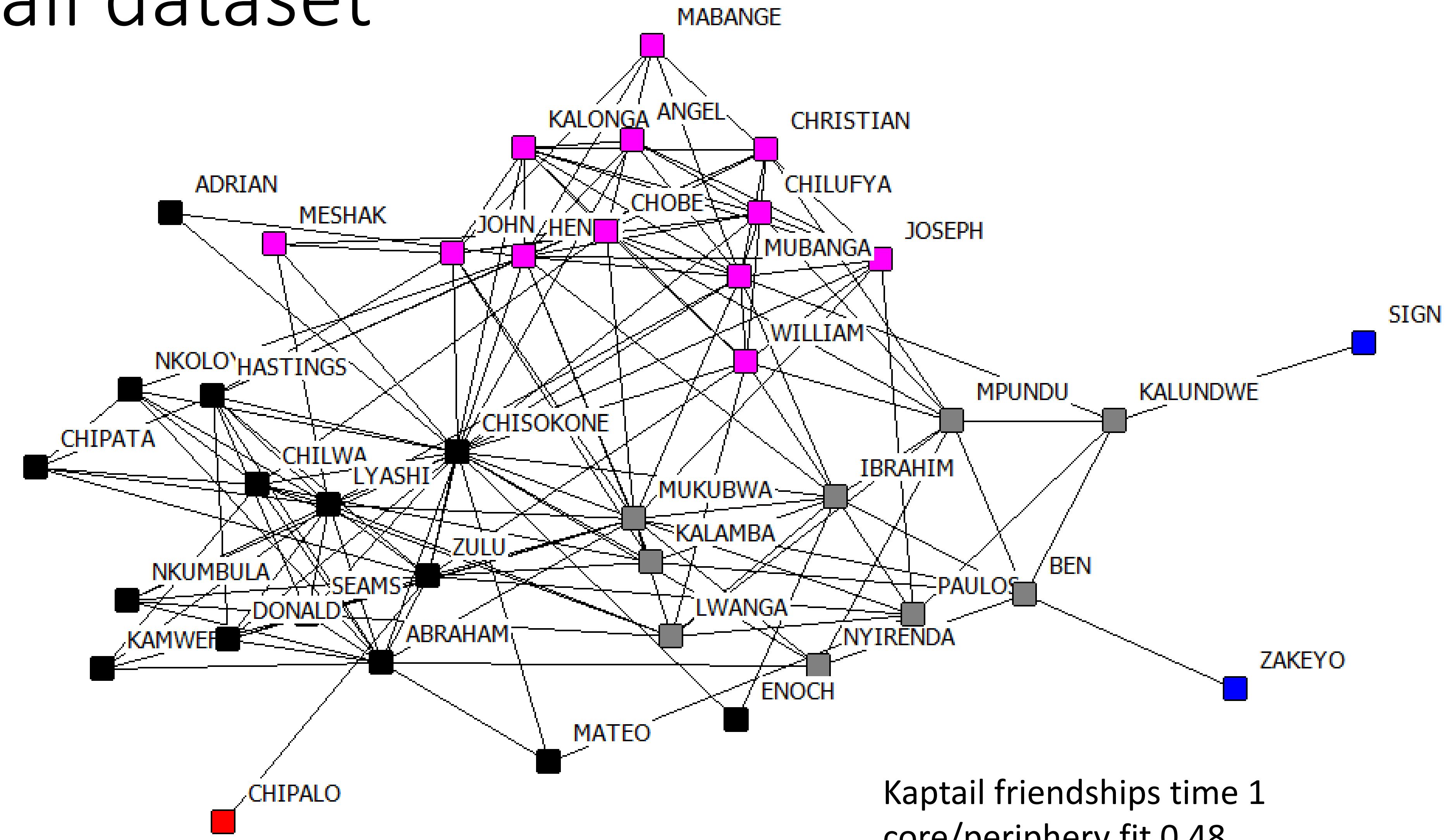
The matrices represent two different types of interaction, recorded at two different times (seven months apart) over a period of one month. TI1 and TI2 record the "instrumental" (work- and assistance-related) interactions at the two times; TS1 and TS2 the "sociational" (friendship, socioemotional) interactions.

The data are particularly interesting since an abortive strike occurred after the first set of observations, and a successful strike took place after the second.

-- UCINET help

Network|Core-periph|Categorical

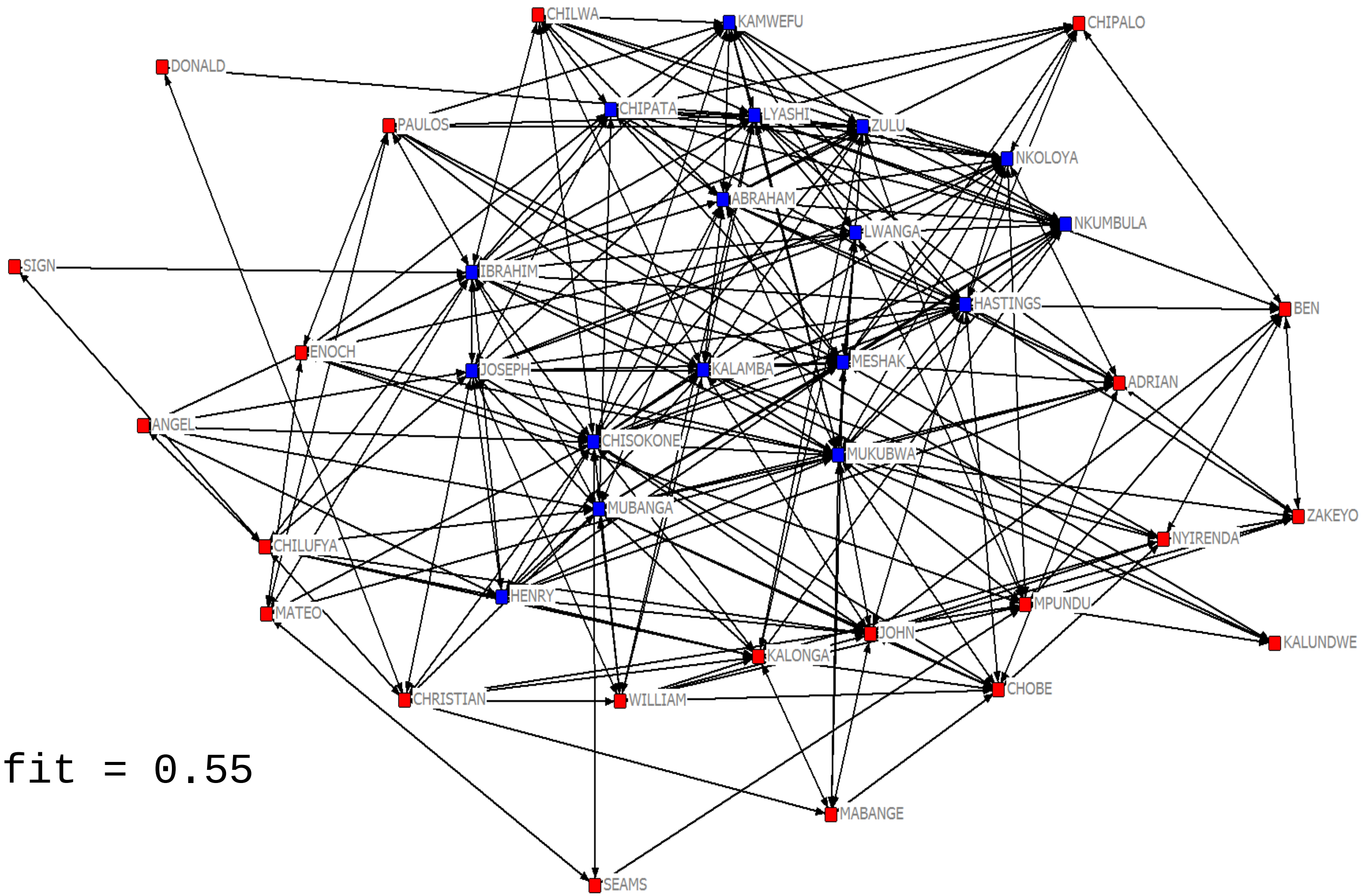
Kaptail dataset



Kaptail friendships time 1
core/periphery fit 0.48

Try cp on event by event matrix
Run Network|Cohesion|multiple measures ~kaptail

Kaptail time 2



Core/Periphery fit = 0.55

Finding Core/Periphery Structures

- Two approaches
  - Discrete/blockmodeling
    - Use combinatorial optimization to partition nodes into core and periphery sets such that core-core ties are maximized and periphery-periphery ties are minimized
  - Continuous
    - Calculate coreness of each node by modeling existence/strength of ties between pair of nodes as function of coreness of each

Categorical Approach

- Use combinatorial optimization to partition nodes into core and periphery sets such that
  - core-core ties are maximized
  - periphery-periphery ties are minimized
  - Core to Periphery: unspecified, but normally expect in-between value

$$\begin{pmatrix}
* & 1 & 1 & 1 & & & & & & \\
1 & * & 1 & 1 & & & & & & \\
1 & 1 & * & 1 & & & & & & \\
1 & 1 & 1 & * & 0 & 0 & 0 & 0 & 0 & 0 \\
 & & & 0 & * & 0 & 0 & 0 & 0 & 0 \\
 & & & 0 & 0 & * & 0 & 0 & 0 & 0 \\
 & & & 0 & 0 & 0 & * & 0 & 0 & 0 \\
 & & & 0 & 0 & 0 & 0 & * & 0 & 0 \\
 & & & 0 & 0 & 0 & 0 & 0 & * & 0 \\
 & & & 0 & 0 & 0 & 0 & 0 & 0 & *
\end{pmatrix}$$

Categorical Results

		1 K	2 N	3 A	3 H	5 C	5 M	7 N	2 Z	1 K	9 M	1 L	6 C	4 M	4 L	2 H	4 I	0 J	9 C	4 S	8 M	7 E	8 P	0 C	0 S	6 D	2 Z	7 K	8 M	9 J	6 A	1 W	3 B	3 C	5 N	5 C	6 K	7 A	8 C	9 M
1	KAMWEFU		1	1	1	1			1		1	1	1		1		1		1				1																	
2	NKUMBULA	1		1	1	1	1		1		1	1	1	1	1																						1			
3	ABRAHAM	1	1		1	1	1	1	1	1	1	1	1	1	1		1		1					1							1									
13	HASTINGS	1	1	1		1	1	1		1	1	1	1				1	1						1			1			1	1		1	1						
5	CHIPATA	1	1	1	1		1	1	1			1		1			1	1	1			1		1																
25	MESHAK		1	1	1	1		1		1	1		1	1	1	1	1	1					1	1							1				1					1
7	NKOLOYA			1	1	1	1		1		1	1	1		1			1	1					1					1		1									
12	ZULU	1	1	1		1		1		1	1	1	1				1		1				1						1								1			
21	KALAMBA			1	1		1		1		1	1	1			1	1	1	1				1					1		1		1			1					
19	MUKUBWA	1	1	1	1		1	1	1	1		1	1	1	1	1	1	1			1	1					1	1	1	1	1			1	1					1
11	LYASHI	1	1	1	1	1		1	1	1	1				1		1	1	1				1	1		1					1	1	1							
16	CHISOKONE	1	1	1	1		1	1	1	1	1					1	1	1	1	1	1	1								1		1		1		1	1	1		
34	MUBANGA		1	1		1	1				1					1	1	1												1	1	1		1		1	1	1	1	
14	LWANGA	1	1	1			1	1			1	1					1	1				1							1								1			
32	HENRY						1			1	1		1	1			1	1												1	1						1	1	1	
24	IBRAHIM	1		1	1	1	1		1	1	1	1	1	1	1	1		1	1		1	1	1		1													1	1	
30	JOSEPH				1	1	1	1		1	1	1	1	1	1	1	1															1				1		1	1	
9	CHILWA	1		1		1		1	1	1		1	1				1																							
4	SEAMS												1								1								1											
8	MATEO										1		1				1			1		1	1																	
17	ENOCH					1					1		1		1		1				1		1						1											
18	PAULOS	1					1		1	1		1					1				1	1						1												
10	CHIPALO			1	1	1	1	1				1																					1							
20	SIGN																1																					1		
6	DONALD											1																							1					
22	ZAKEYO				1						1																		1	1	1		1		1					
27	KALUNDWE									1	1												1						1											
28	MPUNDU							1	1		1				1					1		1					1	1				1	1			1				
29	JOHN				1					1	1		1	1		1											1					1	1	1	1				1	1
26	ADRIAN			1	1		1	1			1	1		1		1											1							1						
31	WILLIAM									1		1	1	1				1											1	1				1		1	1			
23	BEN				1							1												1			1		1	1					1					
33	CHOBE				1						1		1	1																1	1	1			1		1			1
15	NYIRENDA						1			1	1																1			1			1	1			1			
35	CHRISTIAN												1	1				1								1			1			1					1		1	1
36	KALONGA		1						1				1	1	1	1																1		1	1	1			1	1
37	ANGEL												1	1		1	1	1																				1		
38	CHILUFYA													1		1	1	1							1					1						1	1	1		
39	MABANGE						1				1																			1				1		1	1			

Density matrix

	1	2
1	0.699	0.235
2	0.235	0.173

Kaptail-kapfts2

Continuous approach

- Discrete model effectively creates binary coreness variable such that ties between i and j are given by product of coreness of each
  - If ci and cj = 1 then Xij = 1
  - If ci = 1 and cj = 0, then Xij = 0
  - if ci and cj = 0 then Xij = 0
- So this could be generalized to real-valued coreness vector

core ness		1	1	1	0	0	0	0
		a	b	c	d	e	f	g
1	a	1	1	1	0	0	0	0
1	b	1	1	1	0	0	0	0
1	c	1	1	1	0	0	0	0
0	d	0	0	0	0	0	0	0
0	e	0	0	0	0	0	0	0
0	f	0	0	0	0	0	0	0
0	g	0	0	0	0	0	0	0

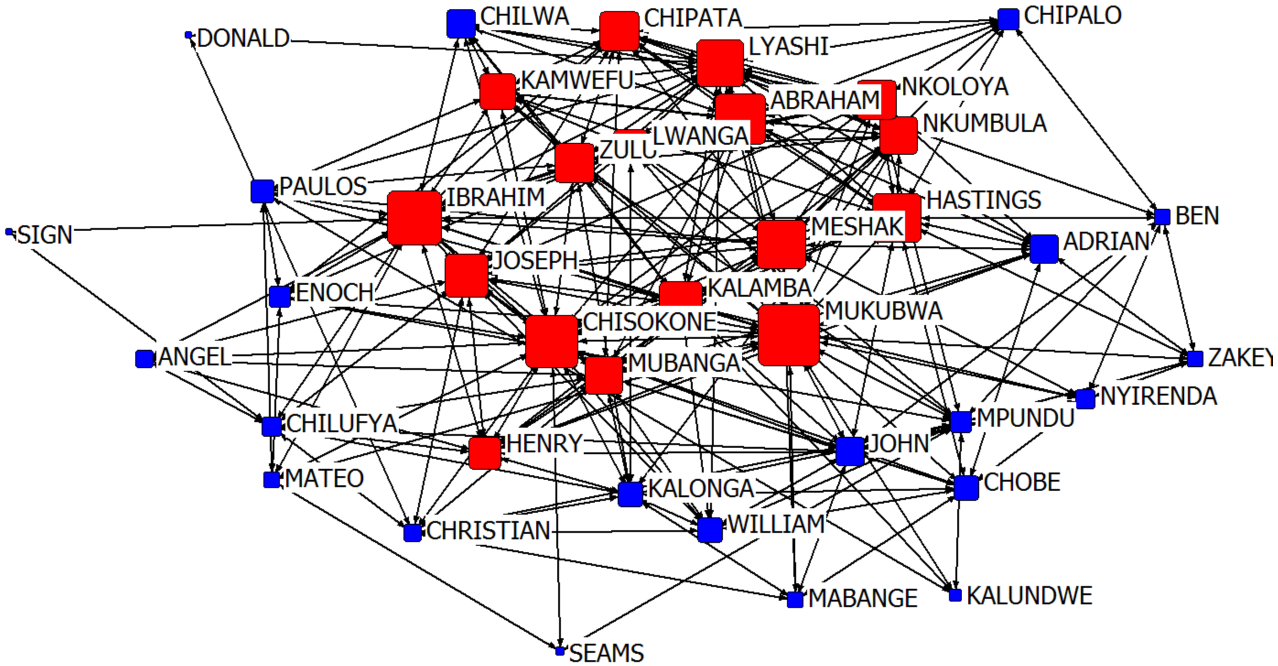
Continuous approach

- We generalize to continuous coreness scores such that prob/strength of a tie between I and j is a function of the coreness of each
  - Xij = f(ci*cj)
  - If both have high coreness, then tied to each other
  - If both have low coreness, then not tied
- We use a least-squares type procedure to find scores c to minimize

$$\sum_{i,j}(x_{ij} - c_i c_j)^2$$

- Fitting a model of ties
  - Could use r-square to measure fit of model

Continuous coreness



Colors based on the discrete model. Sizes based on continuous model

		1 Corene
16	CHISOKONE	0.406
19	MUKUBWA	0.304
11	LYASHI	0.249
34	MUBANGA	0.242
32	HENRY	0.233
12	ZULU	0.232
3	ABRAHAM	0.213
13	HASTINGS	0.184
30	JOSEPH	0.182
24	IBRAHIM	0.181
31	WILLIAM	0.174
4	SEAMS	0.173
36	KALONGA	0.160
21	KALAMBA	0.157
38	CHILUFYA	0.157
29	JOHN	0.152
6	DONALD	0.143
33	CHOBE	0.142
9	CHILWA	0.141
14	LWANGA	0.128
35	CHRISTIAN	0.125
37	ANGEL	0.124
7	NKOLOYA	0.119
2	NKUMBULA	0.114
18	PAULOS	0.102
28	MPUNDU	0.101
15	NYIRENDA	0.099
39	MABANGE	0.085
25	MESHAK	0.085
5	CHIPATA	0.082
23	BEN	0.080
1	KAMWEFU	0.069

Measure cpness

- Both discrete and continuous approaches fit a model to the data, i.,e., predict ties
  - Discrete
    - If ci = 1 and cj = 1 then xij = 1
    - If ci = 0 and cj = 0 then xij = 0
  - Continuous
    - Prob(xij) = f(ci*cj)
- So in both cases we can measure goodness of fit
  - Degree to which data conforms to idealized cp structure

Reciprocity

MAN convention:
- Mutuals
- Asymmetrics
- Nulls

- Let R = number of reciprocated arcs, U = number of unreciprocated arcs
- Arc reciprocity
  - Proportion of outgoing ties that are answered with an incoming tie
  - R/(R+U)
- Dyad reciprocity
  - Proportion of non-null dyads that are symmetric ("mutuals")
  - R/(R+2U)

Reciprocity measures CAMPNET

1	Recip Arcs	38
2	Unrecip Arcs	16
3	All Arcs	54
4	Arc Reciprocity	0.704
5	Sym Dyads	19
6	Asym Dyads	16
7	All ~null Dyads	35
8	Dyad Reciprocity	0.543

Calculating Reciprocity

- Dyad Method

$$\frac{\#Reciprocated\ Dyads}{\#Adjacent\ Dyads}$$

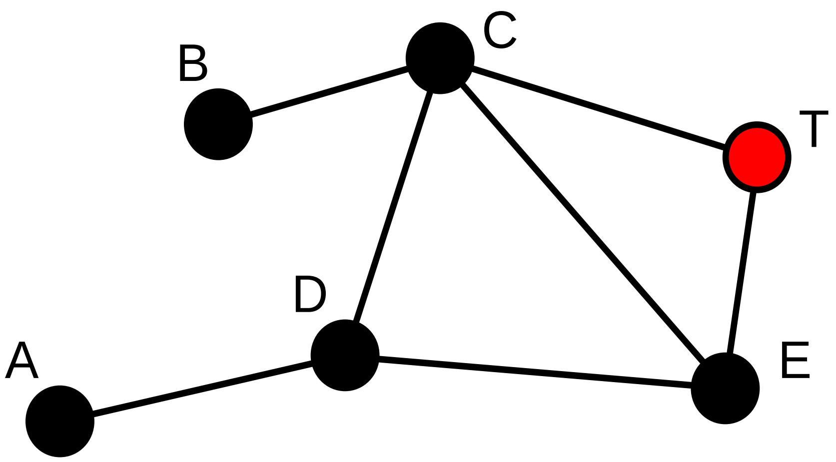
- Arc Method

$$\frac{\#Reciprocated\ Arcs}{\#Total\ Arcs}$$

- Hybrid methods
  - When partitioned: uses Arc Method between groups and Dyad Method within groups
    - When not partitioned, same as Dyad Method

Transitivity

- Proportion of triples with 3 ties as a proportion of triples with 2 or more ties
  - Aka the wtd clustering coefficient
- A clumpiness measure?



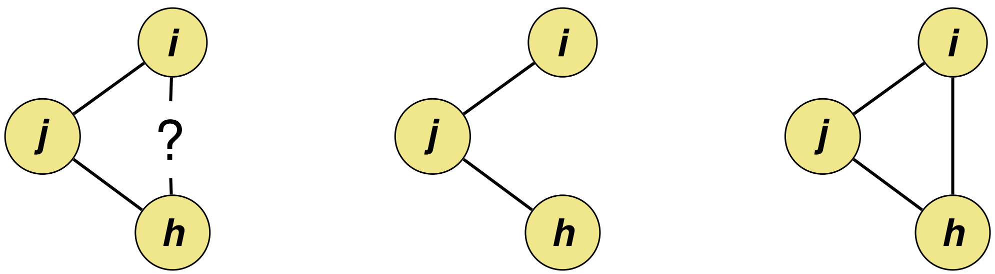
cc = 12/26 = 46.15%

{C,T,E} is a transitive triple, but {B,C,D} is not. {A,D,T} is not counted at all.

Transitivity

The tendency for a tie from i to k to occur at greater than chance frequencies if there are ties from i to j and from j to k – the i to j tie completes "transitively" the triple consisting of the tie from i to j and the tie from j to k.

Transitivity depends on *triads*, subgraphs formed by 3 nodes



Potentially transitive

Intransitive

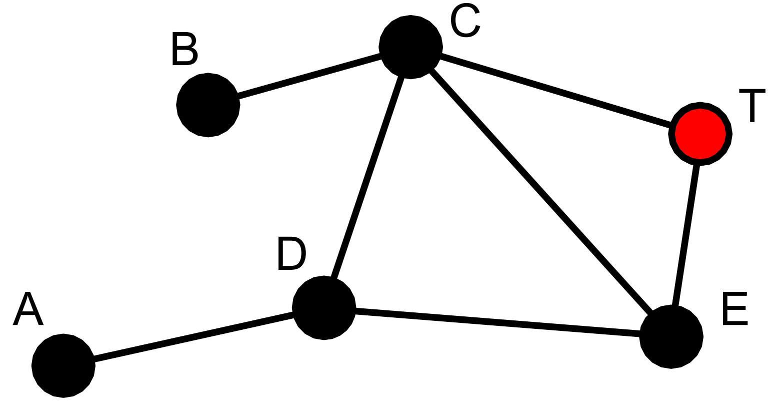
Transitive

Simmelian Ties

- These are the ties that make up transitivity
- Simmelian ties are [reciprocated] transitive triples

  CD,DE,EC is one set of simmelian ties
  CT,TE,EC is another set

  All other sets are not simmelian



measuring transitivity – **clustering index**

A measure for transitivity is the (global) transitivity index, defined as the ratio

$$\text{Transitivity Index} = \frac{\sharp\text{Transitive triads}}{\sharp \text{ Potentially transitive triads}} .$$

(Note that "$\sharp A$" means the number of elements in the set A.)
This also is sometimes called a *clustering* index.

This is between 0 and 1; it is 1 for a transitive graph.
For random graphs, the expected value of the transitivity index is close to the density of the graph (why?);
for actual social networks,
values between 0.3 and 0.6 are quite usual.

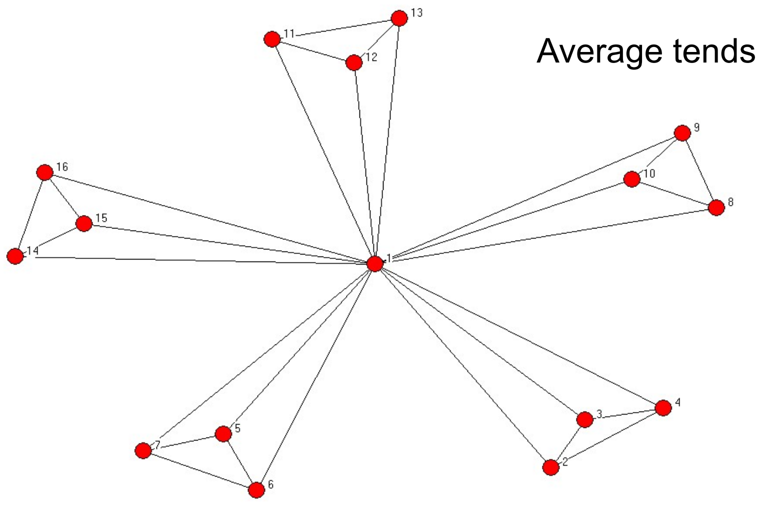
Clustering

What fraction of my friends are friends of each other?

(1) Calculate clustering for a particular node;

(1) Average individual clustering coefficients across the network (it weights clustering node by node)

(2) Overall clustering: out of all possible triplets in the network, what the frequency with which it i

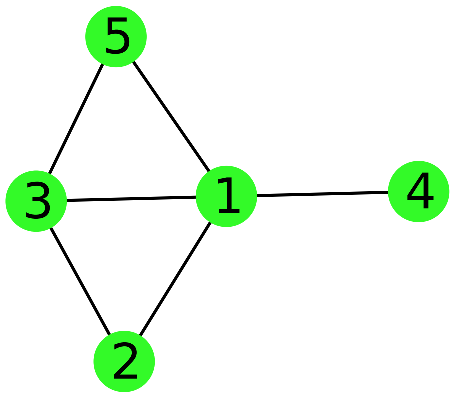


local clustering coefficient

If i is a node with $k_i \geq 2$ then its *local clustering coefficient* is defined as:

$$C_i = \frac{\text{Number of triangles containing } i}{\text{Number of pairs of neighbours of } i},$$
$$= \frac{t_i}{\frac{1}{2}k_i(k_i - 1)},$$

where $t_i = [A^3]_{ii}$.



Possible triangles including node 1:

$$\{(1-2-3), (1-3-5), (1-2-5), (1-5-4), (1-2-4), (1-3-4)\}.$$

Actual triangles:

$$\{(1-2-3), (1-3-5)\}.$$

$$C_1 = \frac{1}{3}.$$

global clustering coefficient

There are two alternative definitions of the global clustering coefficient:

Version 1: Average Clustering Coefficient

$$C = \langle C_i \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i.$$

Version 2: Overall Clustering Coefficient

$$C = \frac{3 \times t}{\text{number of connected triples}}$$

where t is the total number of triangles. If there are no self-loops then $t = \frac{1}{3}\text{trace}(A^3)$.

In adjacency matrix notation,

$$C(v) = \frac{\sum_{u,w \in V} a_{u,v} a_{w,v} a_{u,w}}{\sum_{u,w \in V} a_{u,v} a_{w,v}}.$$

The *(average) clustering coefficient* is defined as

$$C = \frac{1}{|V|} \sum_{v \in V} C(v).$$

Note that

$$\sum_{u,w \in V} a_{u,v} a_{w,v} a_{u,w}$$

is the number of triangles involving v in the graph. Similarly,

$$\sum_{u,w \in V} a_{u,v} a_{w,v}$$

is the number of *2-stars* centred around v in the graph. The clustering coefficient is thus the ratio between the number of triangles and the number of 2-stars. The clustering coefficient describes how "locally dense" a graph is. Sometimes the clustering coefficient is also called the *transitivity*.

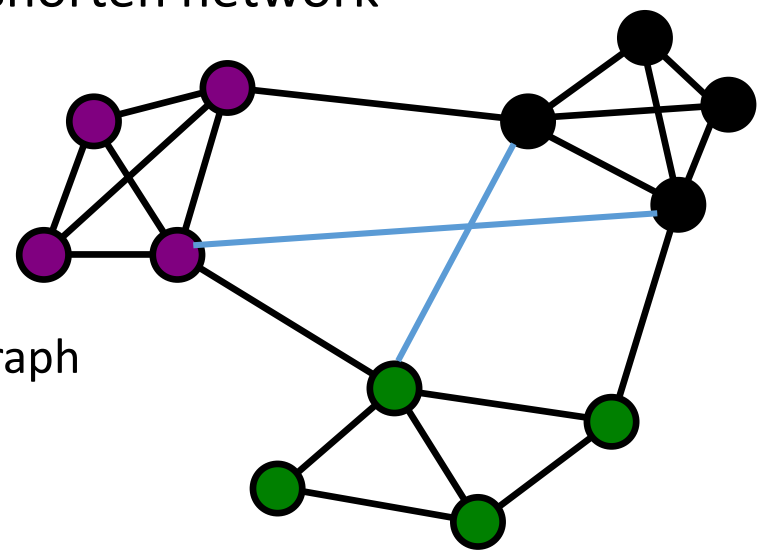
Notes on Clustering Coef

- Unweighted measure
  - Node level clustering coefficient (cc_i) For each node, measure density of their ego network (not including ego)
  - Average cc_i for all *i* to get overall network-level clustering coef
  - Seen as a measure of clumpiness
- Weighted measure
  - When averaging, weight each node by the number of pairs of alters in neighborhood
  - This value is precisely equal to transitivity

Small Worldness

$$\sigma = \frac{C(A)/C(R)}{L(A)/L(R)}$$

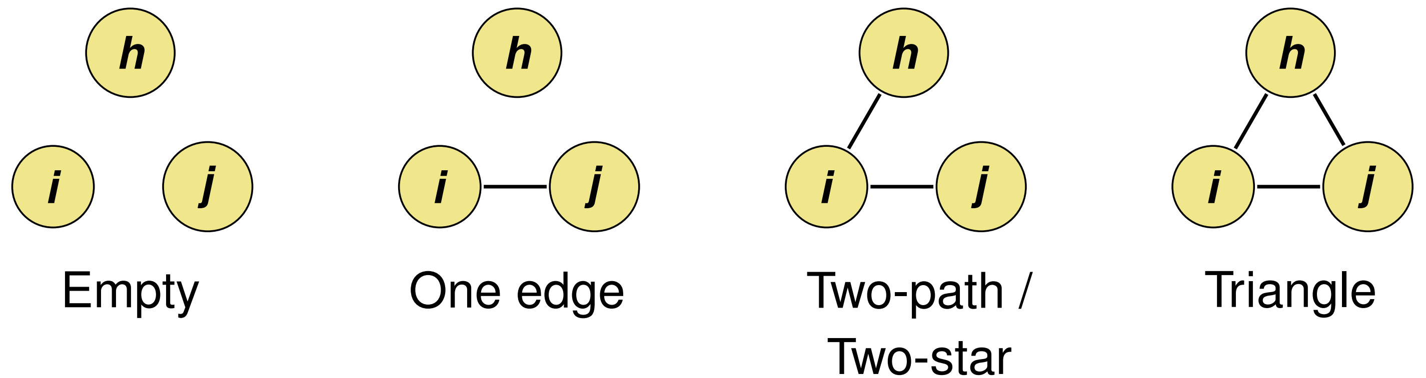
- Theory
  - Human networks typically clumpy
    - Homophily, balance theory, temporal-spatial opportunities
  - In the space of all possible graphs, clumpy graphs tend to have longer distances
    - But as milgram seemed to show, human networks have short distances
  - Watts and Strogatz: a very few random ties will radically shorten network
- Method
  - A network is a small world if it is both clumpy and has short distances
  - How clumpy is clumpy? How short is short? Comparison with random graphs
    - C(A) = clust coef of actual graph; C(R) = clus coef of random graph
    - L(A) = avg dist in actual graph; L(R) = avg dist in random graph
  - Small worldness indices such as σ



local structure and triad counts

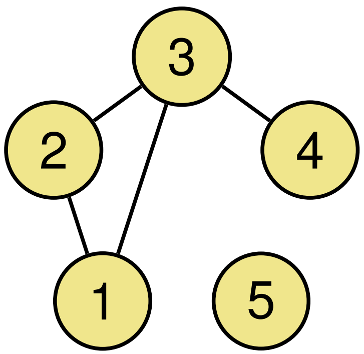
The studies about transitivity in social networks
led Holland and Leinhardt (1975) to propose that
the *local structure* in social networks can be expressed
by the *triad census* or *triad count*, the numbers of triads of any kinds.

For (nondirected) graphs, there are four triad types:



local structure and triad counts

A simple example graph
with 5 nodes.



i	j	h	triad type
1	2	3	triangle
1	2	4	one edge
1	2	5	one edge
1	3	4	two-star
1	3	5	one edge
1	4	5	empty
2	3	4	two-star
2	3	5	one edge
3	4	5	one edge

In this graph, the triad census is $(1, 5, 2, 1)$
(ordered as: empty – one edge – two-star – triangle).

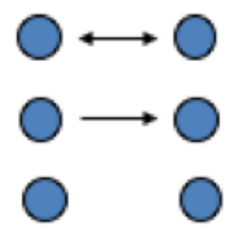
MAN coding for triad census

Holland and Leinhardt (1975) proposed the following MAS coding.

(1) Mutual

(2) Asymmetric

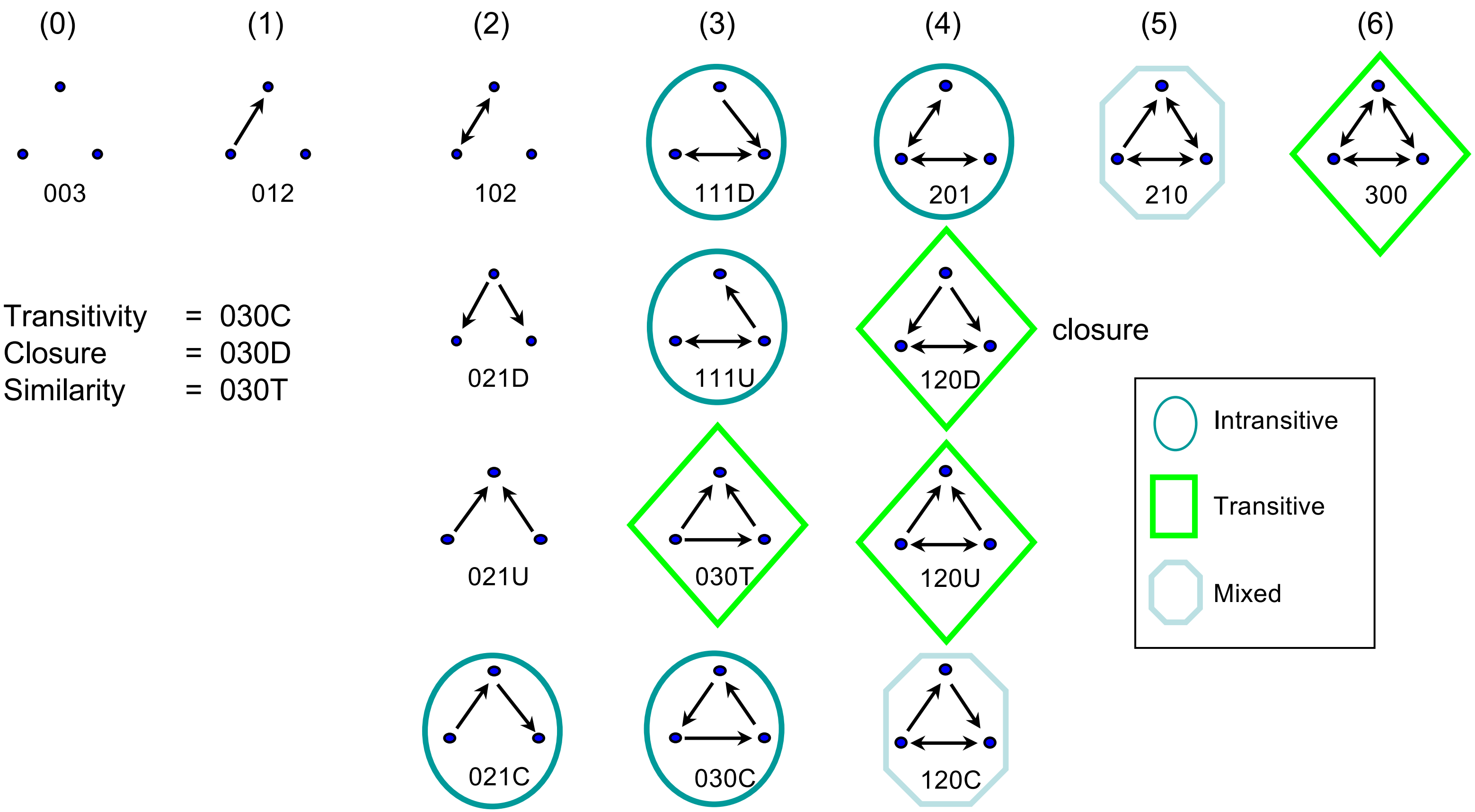
(3) Null



the scheme a further identifying letter: Up, Down, Cyclical, Transitive.

E.g. 120 has 1 mutual, 2 asymmetric, 0 null dyads and the Down orientation

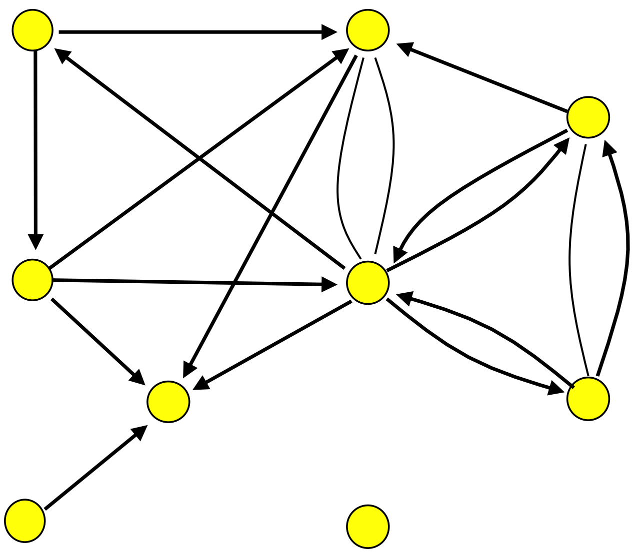
triad census



Transitivity	=	030C
Closure	=	030D
Similarity	=	030T

Transitivity: tie **i** to **k** to occur if ties from **i** to **j** and **j** to **k** exist;
Closure: tie **i** to **j** to occur if persons **k** with ties to both **i** and **j** exist;
Similarity: tie **i** to **j** to occur if persons **k** to whom **i** and **j** have ties exist;

triad census - example



Type	Number of triads
1 - 003	21
2 - 012	26
3 - 102	11
4 - 021D	1
5 - 021U	5
6 - 021C	3
7 - 111D	2
8 - 111U	5
9 - 030T	3
10 - 030C	1
11 - 201	1
12 - 120D	1
13 - 120U	1
14 - 120C	1
15 - 210	1
16 - 300	1
Sum (2 - 16):	63

- triads define behavioral mechanisms: we can leverage the distribution of triads in a network to test whether the hypothesized mechanism is active.

- How?

(1) Count the number of each triad type in a given network

(2) Compare to the expected number, given some (random) distribution of ties in the network;

- Statistical approach proposed by Holland and Leinhardt is now obsolete. Statistical methods have been proposed for probability distributions of graphs depending primarily on triad counts, but complemented with stat counts and nodal variables, along with some higher-order configurations essential for adequate modeling of empirical network data.

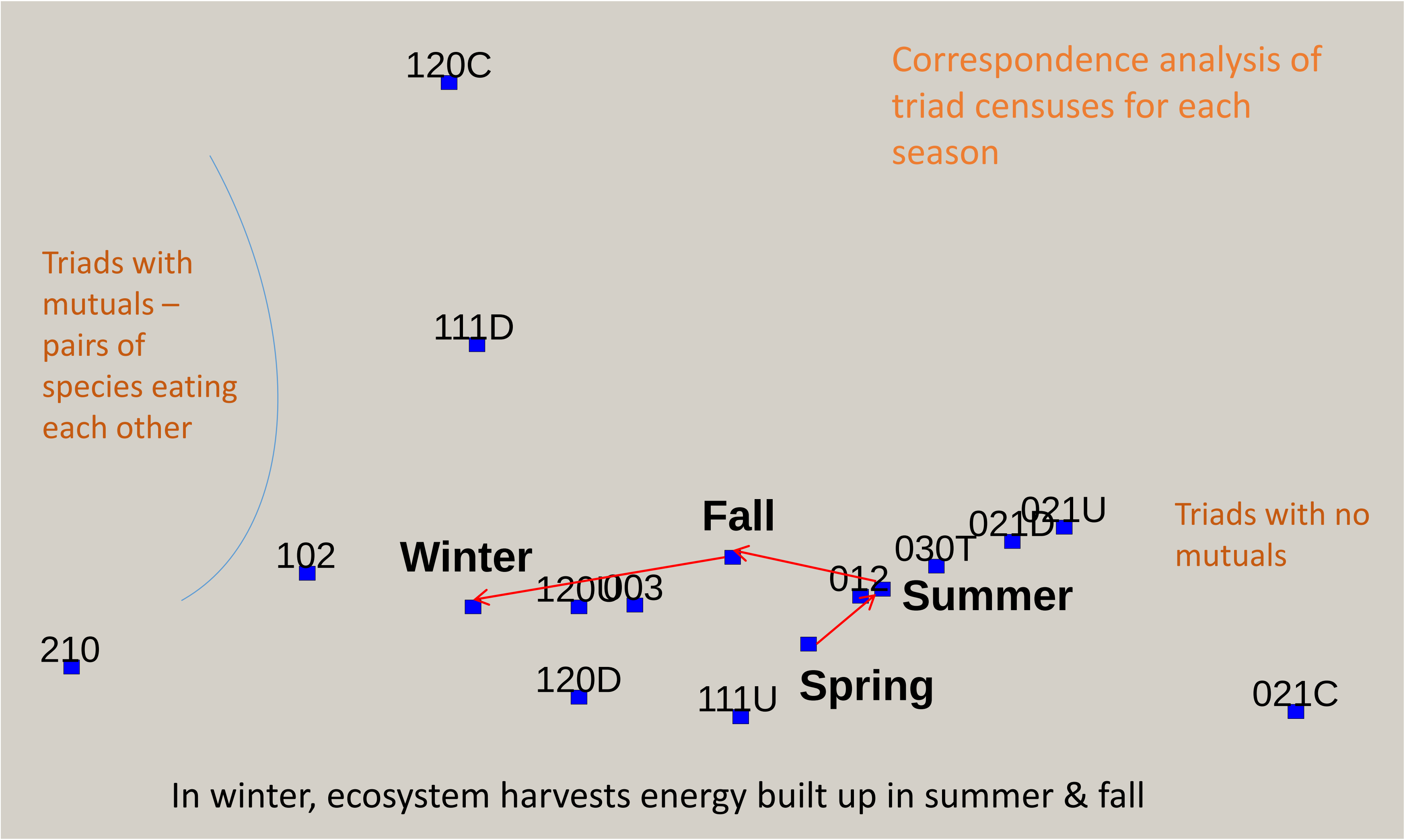
Triad census at multiple time periods

- Chesapeake Bay ecosystem Baird and Ulanowicz (1989)
  - Nodes are species-compartments
- Who eats whom collected once each season for one year
- Table shows change in frequencies of each kind of triad

Triad	Spring	Summer	Fall	Winter
003	4487	4359	4539	4906
012	1937	2001	1884	1663
102	75	71	88	118
021D	115	136	119	88
021U	259	300	273	180
021C	156	153	113	67
111D	25	27	44	37
111U	14	13	11	13
030T	46	54	46	39
030C	7	4	0	0
201	0	0	1	1
120D	8	6	7	8
120U	7	8	7	9
120C	1	5	5	5
210	3	3	3	6
300	0	0	0	0

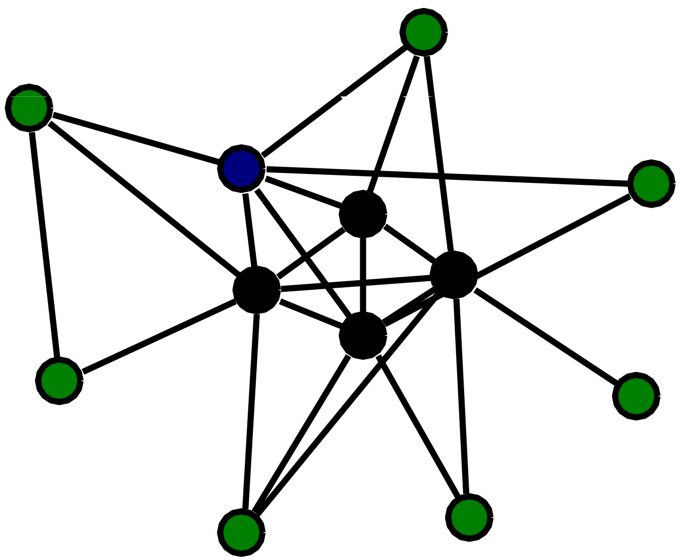
Johnson, JC Luczkovich, JJ, Borgatti, SP, Snijders, TAB. 20 alysis tools in ecology: Markov process transition models applied to the seasonal trophic network dynamics of the Chesapeake Bay. *Ecological Modelling* 220 (2009) 3133-3140

Visualizing change in frequencies

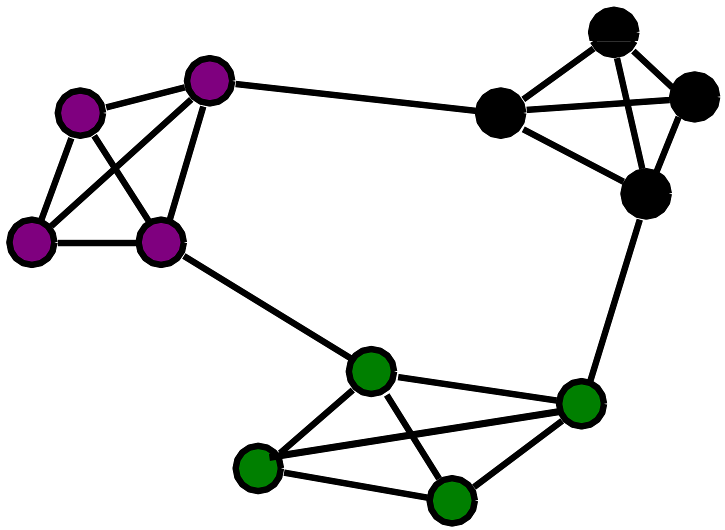


Average Distance

- Average geodesic distance between all pairs of nodes



avg. dist. = 1.9



avg. dist. = 2.4

Average geodesic distance

Only makes sense in connected graphs

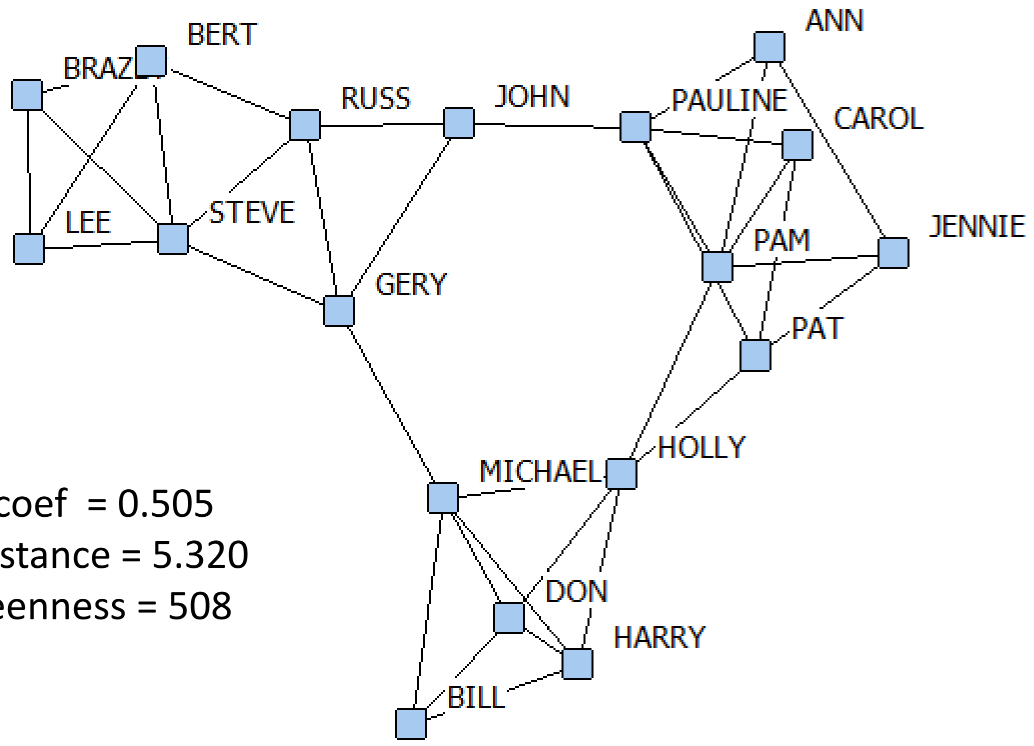
	HOLLY	BRAZEY	CAROL	PAM	PAT	JENNIE	PAULINE	ANN	MICHAEL	BILL	LEE	DON	JOHN	HARRY	GERY	STEVE	BERT	RUSS
HOLLY	0	4	2	1	1	2	2	2	1	2	4	1	3	1	2	3	4	3
BRAZEY	4	0	5	5	5	6	4	5	3	4	1	4	3	4	2	1	1	2
CAROL	2	5	0	1	1	2	1	2	3	4	5	3	2	3	3	4	4	3
PAM	1	5	1	0	2	1	1	1	2	3	5	2	2	2	3	4	4	3
PAT	1	5	1	2	0	1	1	2	2	3	5	2	2	2	3	4	4	3
JENNIE	2	6	2	1	1	0	2	1	3	4	6	3	3	3	4	5	5	4
PAULINE	2	4	1	1	1	2	0	1	3	4	4	3	1	3	2	3	3	2
ANN	2	5	2	1	2	1	1	0	3	4	5	3	2	3	3	4	4	3
MICHAEL	1	3	3	2	2	3	3	3	0	1	3	1	2	1	1	2	3	2
BILL	2	4	4	3	3	4	4	4	1	0	4	1	3	1	2	3	4	3
LEE	4	1	5	5	5	6	4	5	3	4	0	4	3	4	2	1	1	2
DON	1	4	3	2	2	3	3	3	1	1	4	0	3	1	2	3	4	3
JOHN	3	3	2	2	2	3	1	2	2	3	3	3	0	3	1	2	2	1
HARRY	1	4	3	2	2	3	3	3	1	1	4	1	3	0	2	3	4	3
GERY	2	2	3	3	3	4	2	3	1	2	2	2	1	2	0	1	2	1
STEVE	3	1	4	4	4	5	3	4	2	3	1	3	2	3	1	0	1	1
BERT	4	1	4	4	4	5	3	4	3	4	1	4	2	4	2	1	0	1
RUSS	3	2	3	3	3	4	2	3	2	3	2	3	1	3	1	1	1	0

	Geo Dist
Mean	2.66
Std Dev	1.26
Sum	814
Variance	1.60
SSQ	2654
MCSSQ	488.65
Euc Norm	51.52
Minimum	1
Maximum	6
N of Obs	306

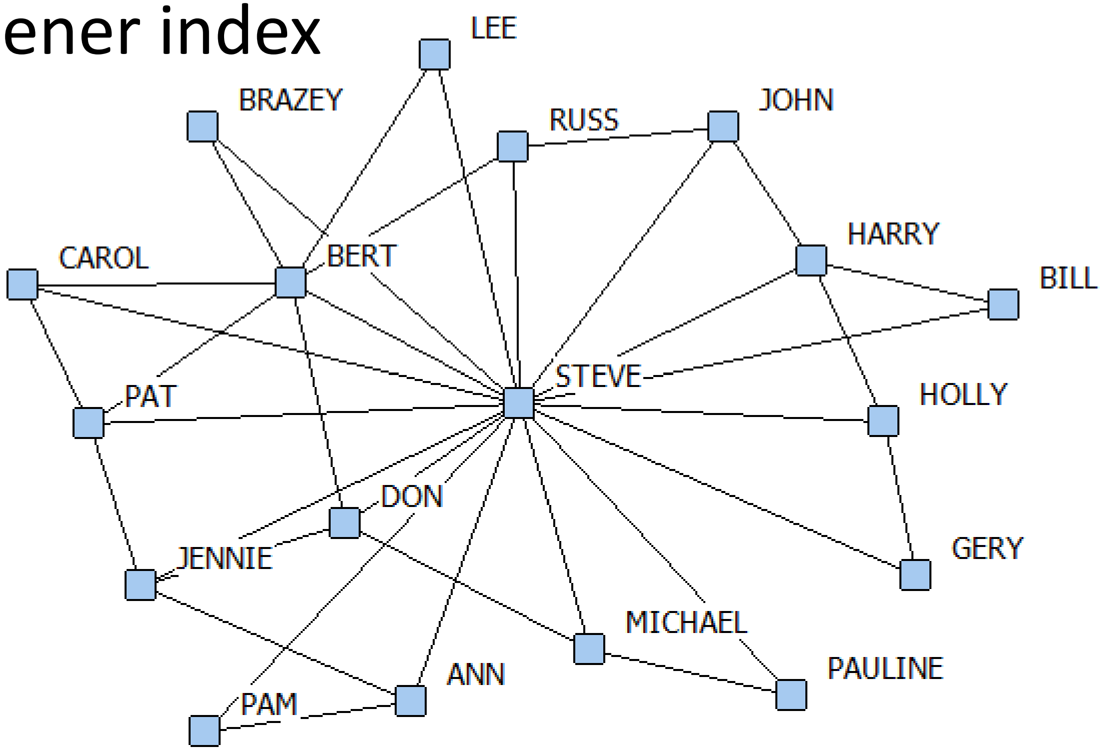
Average Distance

Clumpy networks tend to have longer distances

- Average geodesic distance between all pairs of nodes
- Sum of distances is known as the Wiener index



Clustering coef = 0.505
Avg Geo Distance = 5.320
Sum Betweenness = 508

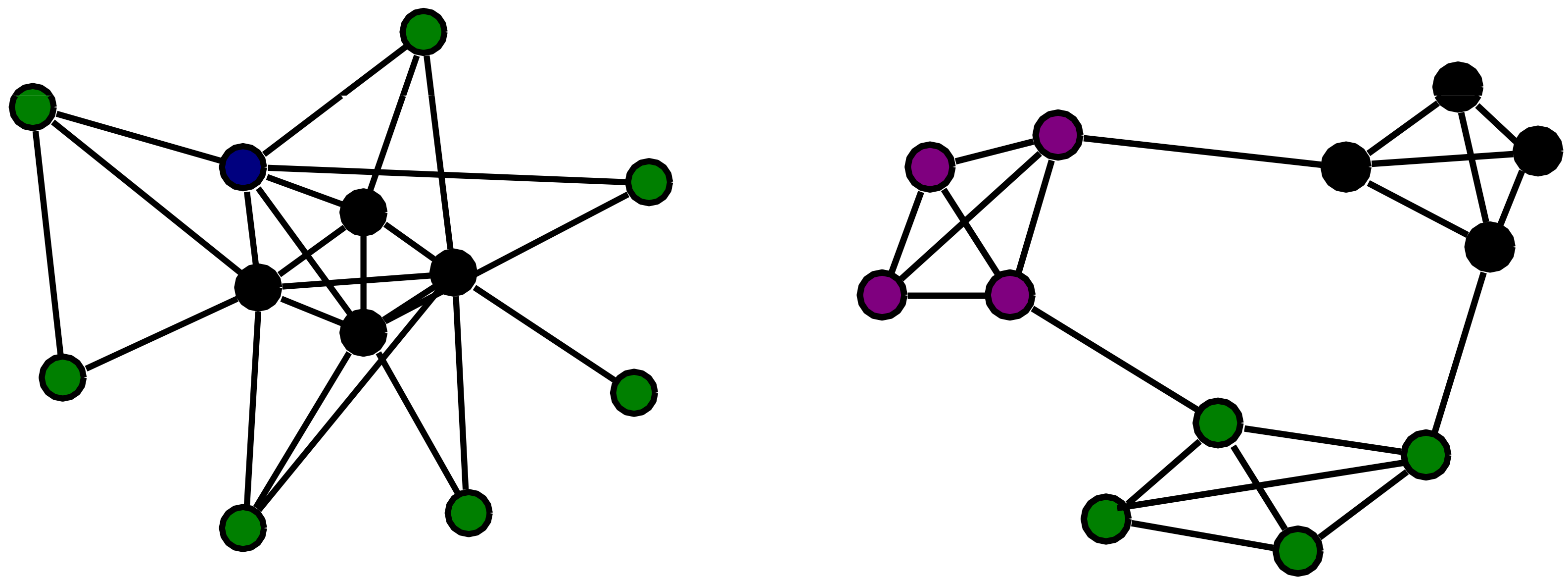


Clustering coef = 0.278
Avg Geo Distance = 3.542
Sum of betweenness = 236

Note that the number of nodes and ties is th works

Diameter

- Maximum distance



Diameter = 3

Diameter = 3

Fragmentation Measures

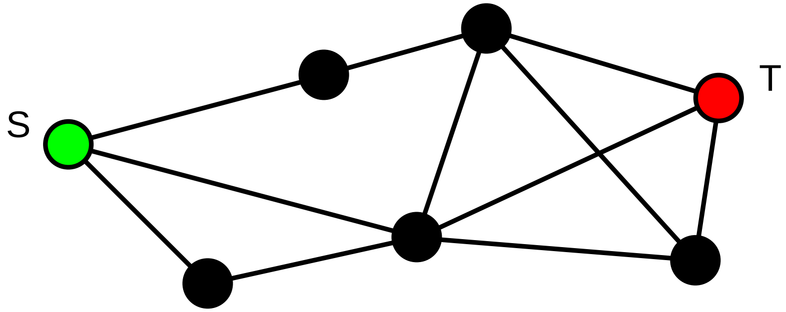
- Component ratio
- F measure of fragmentation
- Distance-weighted fragmentation ^{D}F

Compactness

[useful given inability to compute average distance for disconnected networks]

- Average reciprocal of distance
  - letting $1/\infty = 0$

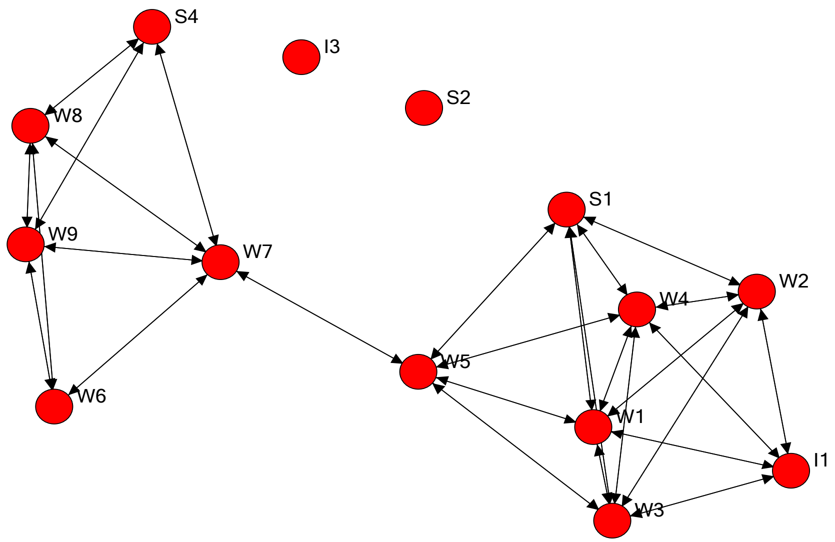
$$C = \frac{\sum_{i \neq j} {}^{1}/_{d_{ij}}}{n(n-1)}$$



- Bounds
  - C = 1 when density = 1. I.e.,. When everyone directly connected to every other
  - C = 0 when density = 0. i.e., all nodes are isolates (and distances are treated as infinite)
- Breadth = 1 - C

Component Ratio (CR)

- No. of components minus 1 divided by number of nodes minus 1



Measure of disconnectedness

CR = (3-1)/(14-1) = 0.154

CR is 1 when all nodes are isolates
CR is 0 when all nodes in one component

F Measure of Fragmentation

- Proportion of pairs of nodes that are unreachable from each other

$$F = 1 - \frac{\sum_{i \neq j} r_{ij}}{n(n-1)}$$

proportion of pairs of nodes that can reach each other via path.

Subtract from 1 to get proportion of pairs that **cannot** reach each other

r_{ij} = 1 if node i can reach node j by a path of any length
r_{ij} = 0 otherwise

- If all nodes reachable from all others (i.e., one component), then F = 0
- If graph is all isolates, then F = 1
- Connectedness = 1 - F

Shortcut Formula for F Measure

- No ties across components, and all reachable within components, hence can express in terms of size of components

$$F = 1 - \frac{\sum_k s_k(s_k - 1)}{n(n-1)}$$

S_k = size of kth component

Distance-Weighted Fragmentation

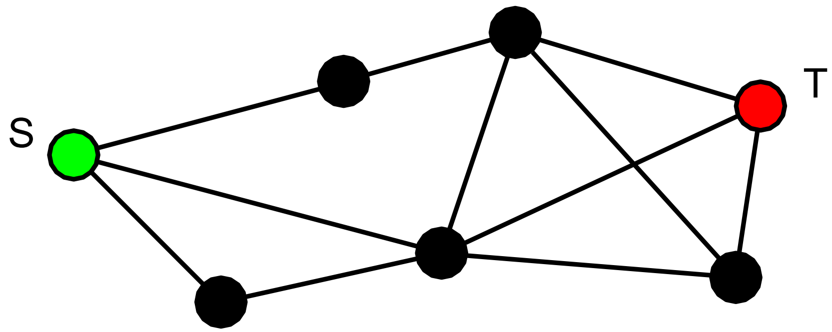
- Use the reciprocal of distance
  - letting $1/\infty = 0$

$$^{D}F = 1 - \frac{\sum_{i \neq j} \frac{1}{d_{ij}}}{n(n-1)}$$

- Bounds
  - lower bound of 0 when every pair is adjacent to every other (entire network is a clique)
  - upper bound of 1 when graph is all isolates

Connectivity

- Line connectivity λ is the minimum number of lines that must be removed to discon- nect network
- Node/point connectivity κ is minimum number of nodes that must be removed to discon- nect network



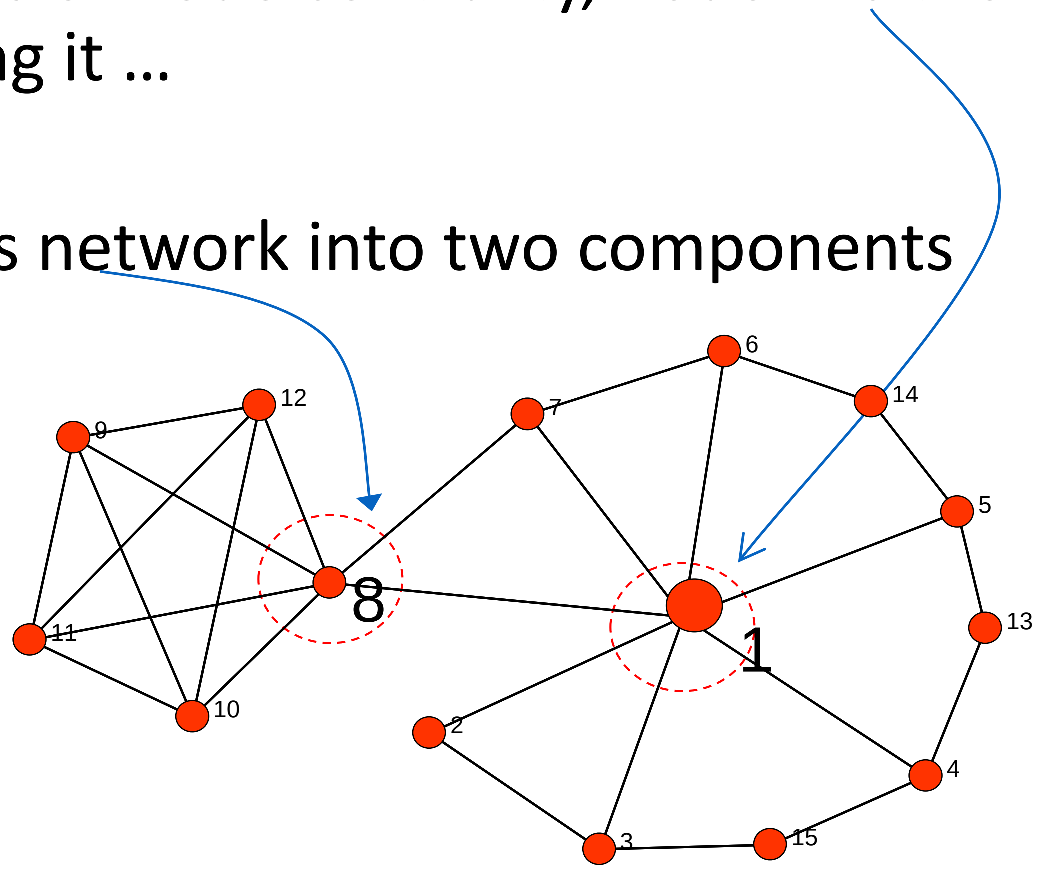
KeyPlayer application

- Suppose you want to disrupt a network
  - E.g., stop epidemic by immunizing/quarantining an affordable # of people
  - Disrupt terrorist group's ability to coordinate
- You have the resources to neutralize just k nodes. Which ones do you pick?
- Obvious solution is the pick the k most central nodes
- Two problems
  - Off-the-shelf measures are not designed for this specific purpose (but we can improvise) *Design Problem*
  - Picking an optimal set of k nodes is not the same thing as picking the k nodes that individually most optimal *Ensemble Problem*

DISRUPTION

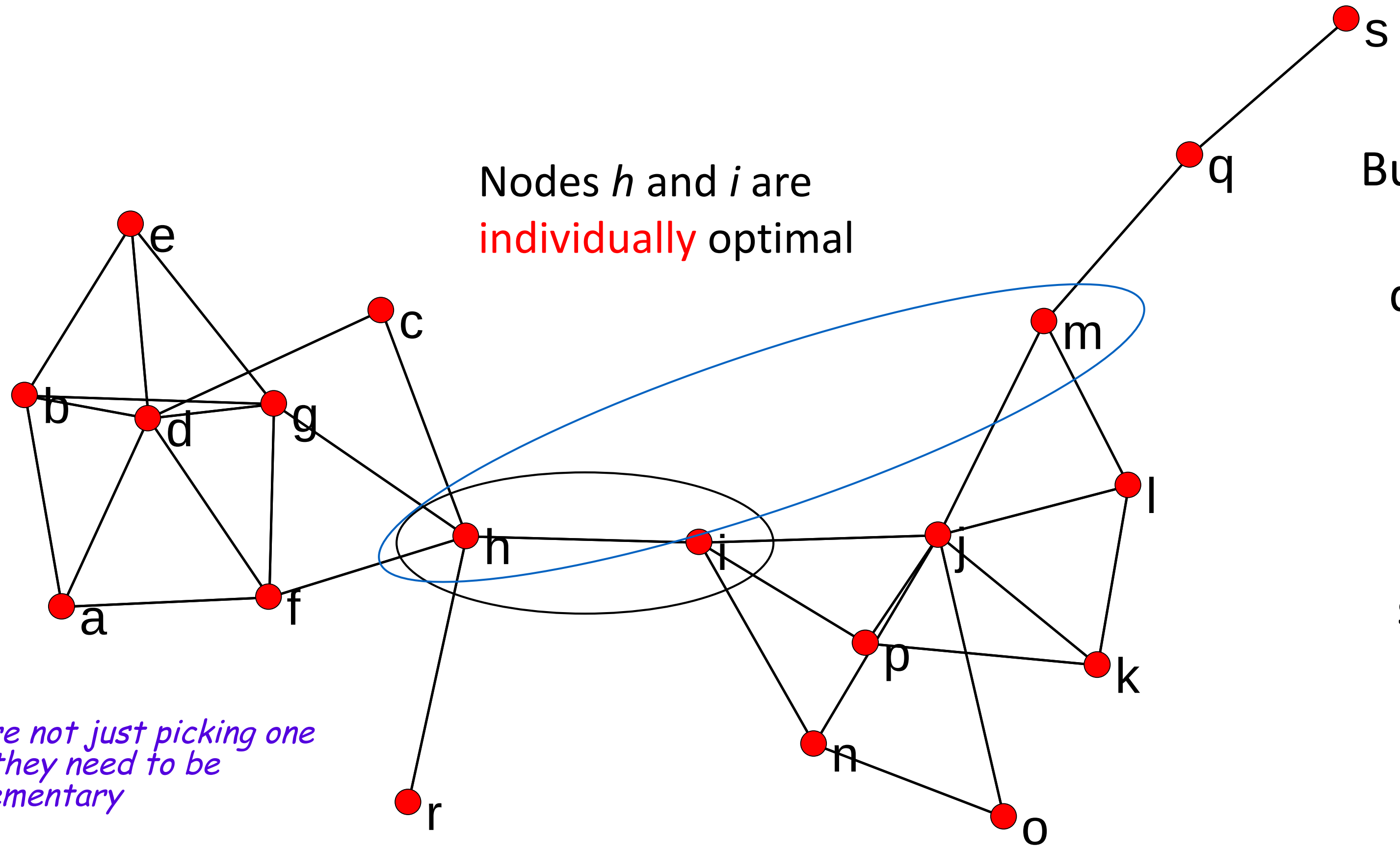
The Design Issue

- By standard off-the-shelf measures of node centrality, node 1 is the most important player, but deleting it …
  - does not disconnect the network
- In contrast, deleting node 8 breaks network into two components
  - Yet node 8 is not highest in centrality
- Standard off-the-shelf centrality measures not optimal for the purpose of disrupting networks
  - Nor many other specific purposes



The Ensemble Issue

Structural redundancy creates need for choosing complementary nodes



Nodes *h* and *i* are individually optimal

But deleting both is no better than deleting *h* alone -- *h* and *i* are redundant

In contrast, {*h*,*m*} splits graph into 4 fragments (is optimal)

if we're not just picking one node, they need to be complementary

- Choosing optimal **set** of *k* players is not same as choosing the *k* best players

KeyPlayer – cont.

- Use a combinatorial optimization algorithm to identify the best combination of k nodes to remove
- Measure "bestness" of a particular combination by the amount of increase in fragmentation as measured by F or breadth

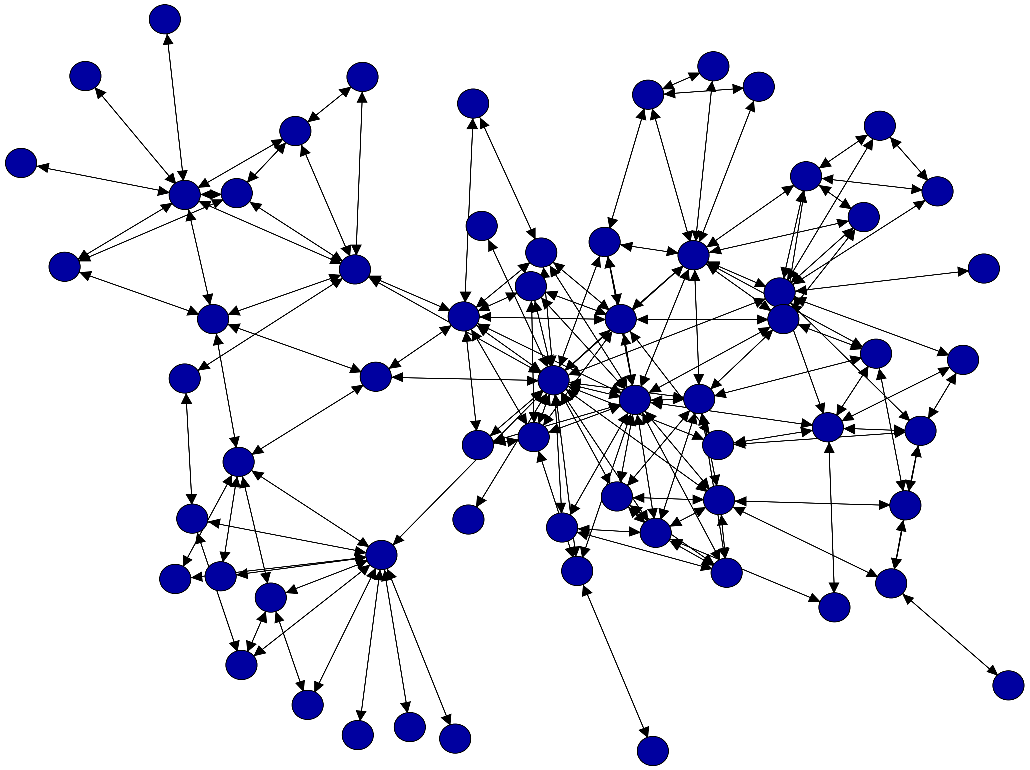
$$F = 1 - \frac{\sum_{i \neq j} r_{ij}}{n(n-1)}$$

r_{ij} = 1 if node i can reach node j by a path of any length
r_{ij} = 0 otherwise

Empirical Example #1
Disrupt Terrorist Network

DISRUPTION

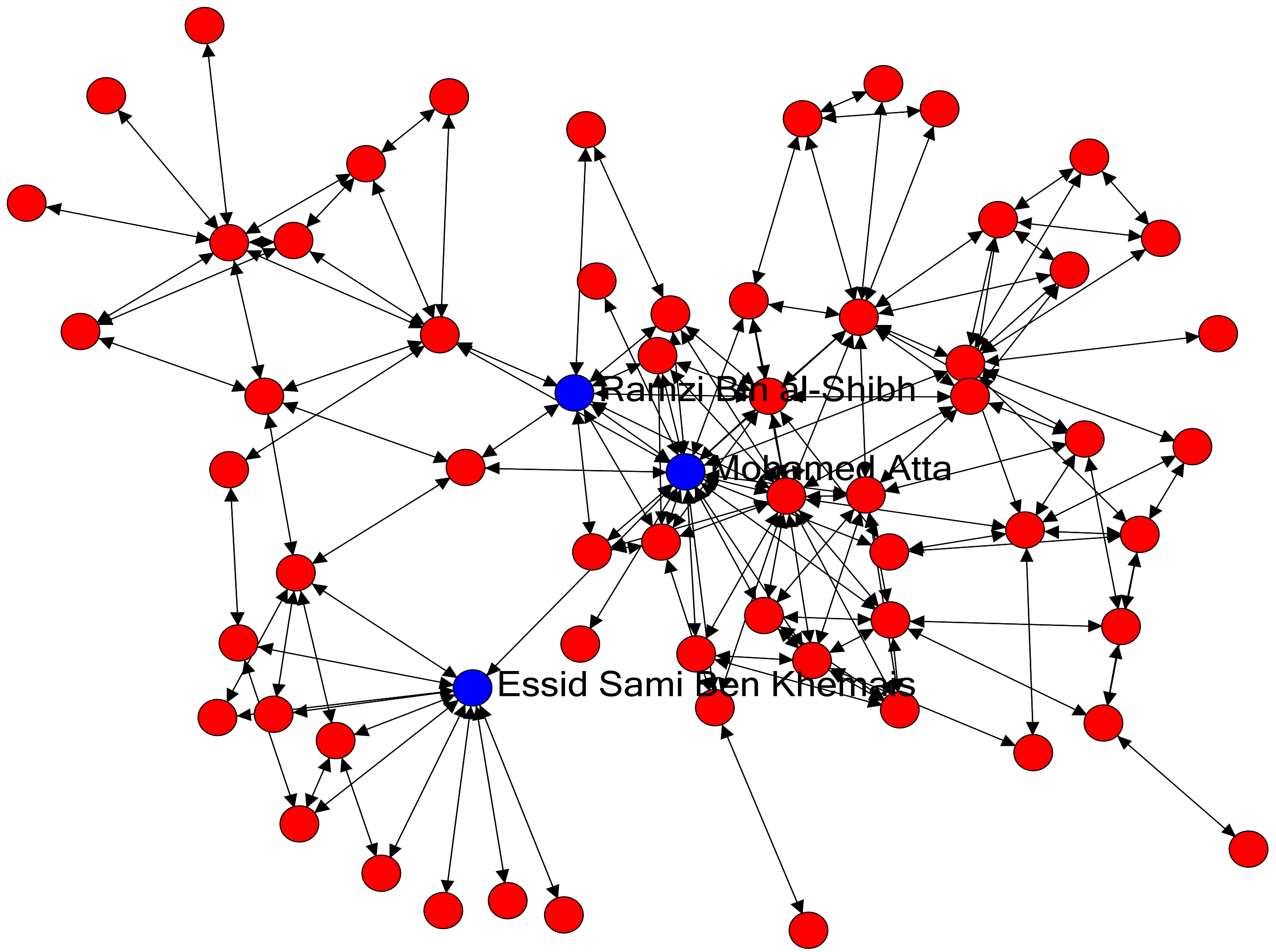
- Which three nodes should be isolated in order to maximally disrupt the network?



Data from: Krebs, V. 2002. Uncloaking terrorist networks.
First Monday 7(4): April. http://www.firstmonday.dk/issues/issue7_4/krebs/index.html

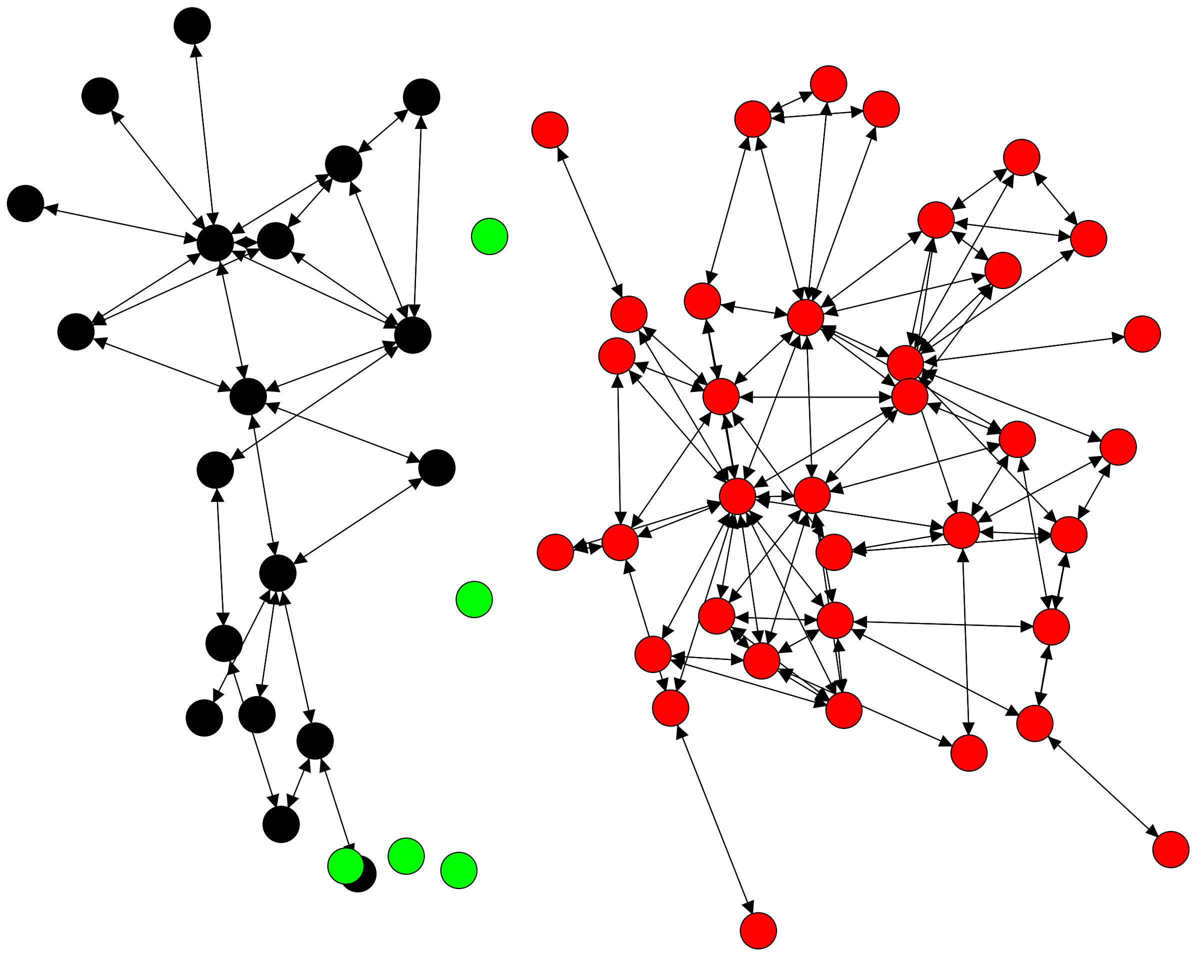
DISRUPTION

KeyPlayer Solution



DISRUPTION

KeyPlayer Solution
(key players removed)



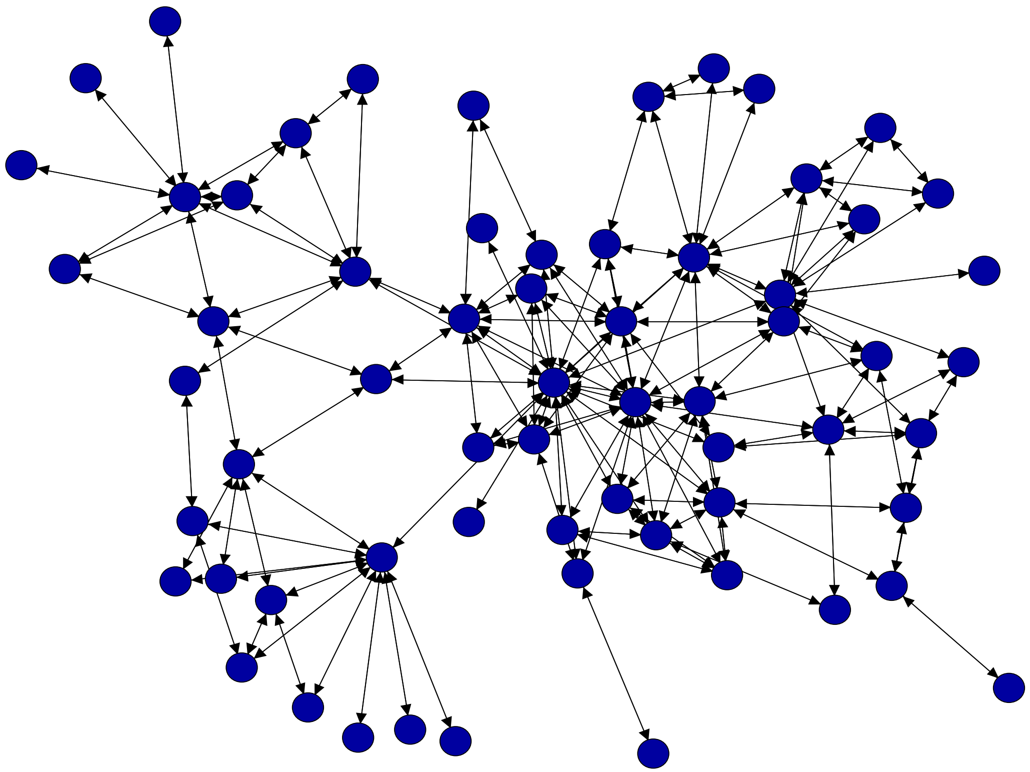
Why do we want to know who the key players are?

DISRUPT	We want to **remove them** – to maximally **disrupt** the network
ENHANCE	We want to **help** them – in order to make network as a whole **function better** (diffuse info; coordinate well)
INFLUENCE	We want to identify **key opinion** leaders – to **influence** the network
LEARN	We want to know who **is in the know** – so we can question or **surveil** them
REDIRECT	We want to remove/prune them – to **redirect flows** in the network toward our preferred players

INFLUENCE

Empirical Example #2
Influence Terrorist Network

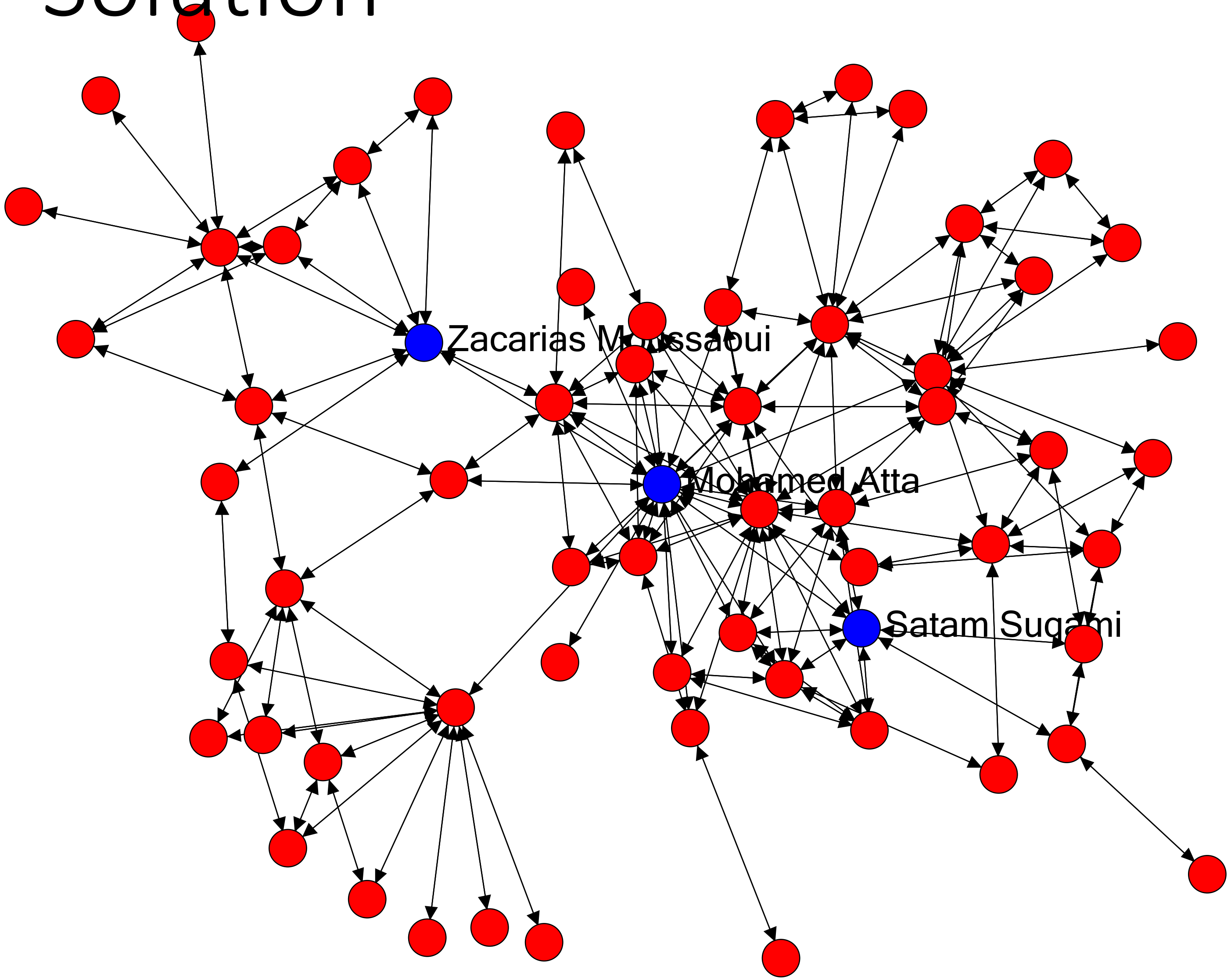
- Which three nodes should be selected in order to maximally influence the network by turning / planting information, etc.?



Data from: Krebs, V. 2002. Uncloaking terrorist networks.
First M w.firstmonday.dk/issues/issue7_4/krebs/index.html

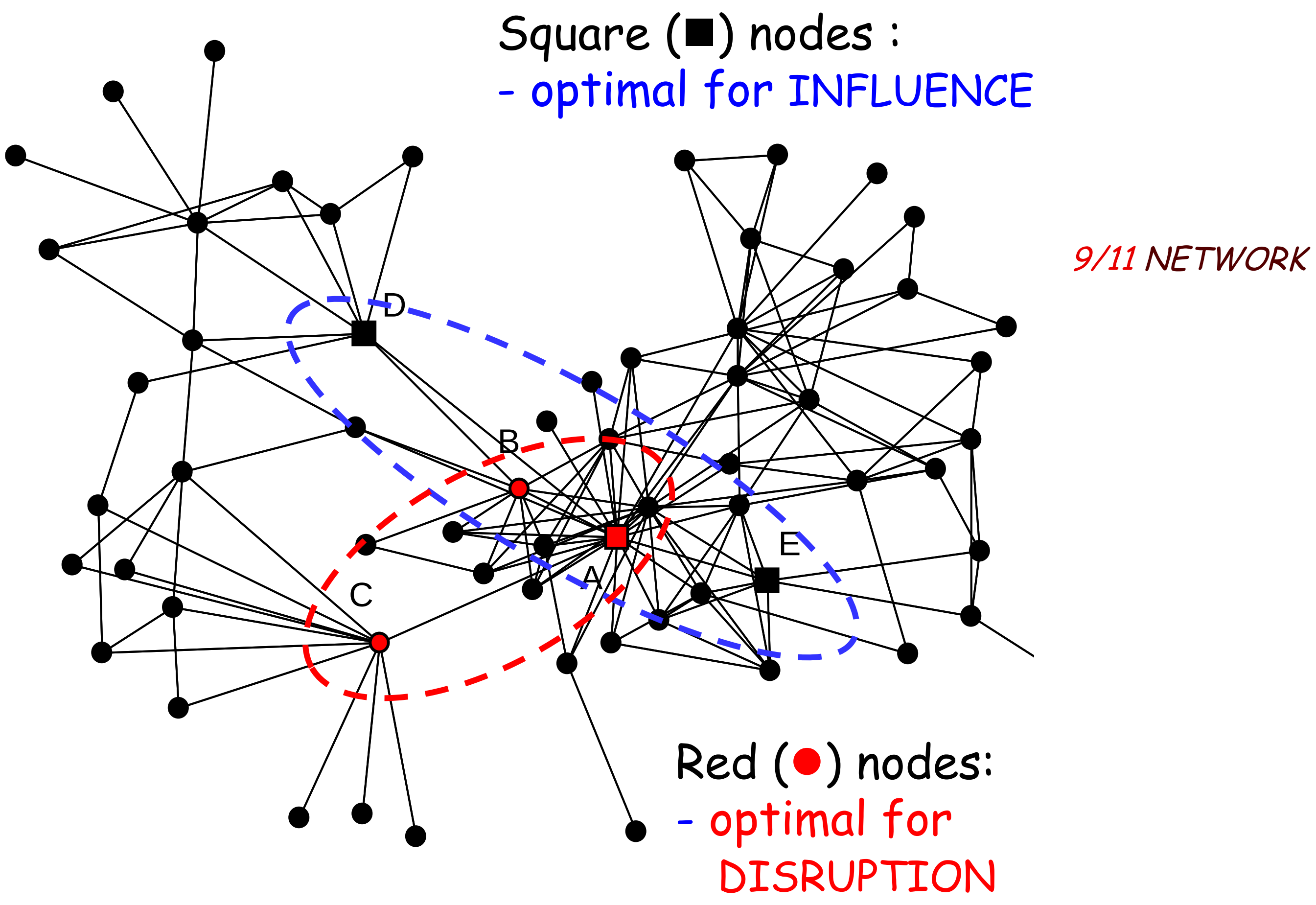
INFLUENCE

KeyPlayer Solution



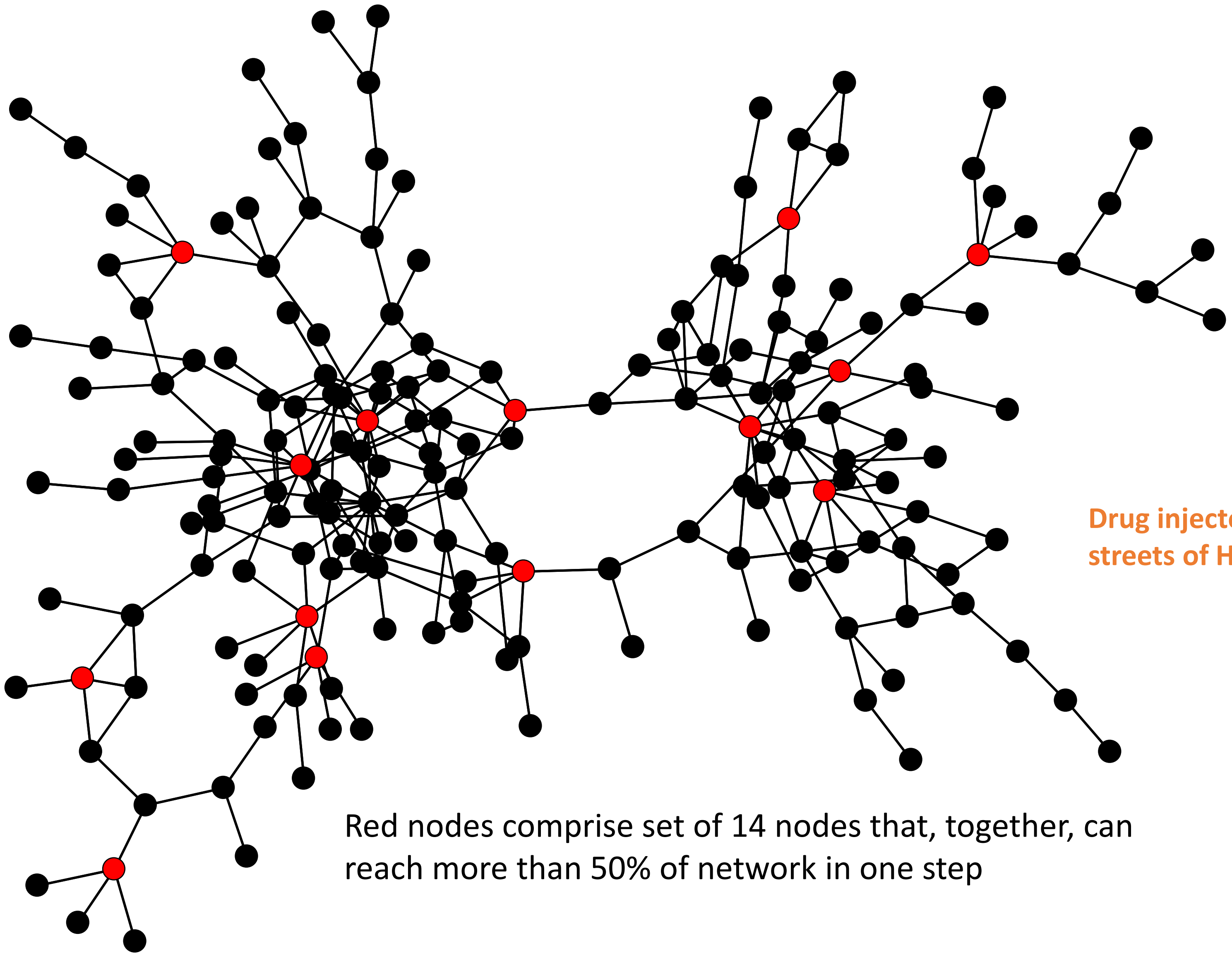
Terrorist Network

- Red nodes identify optimal cho
  for DISRUPTION problem
  - Removing them splits network in
    7 components and yields
    fragmentation metric of 0.59
- Square nodes identify solution
  INFLUENCE problem
  - The best nodes to seed with
    disinformation



Data from: Krebs, V. 2002. Uncloaking terrorist networks.
First M stmonday.dk/issues/issue7_4/krebs/index.html

IDENTIFYING SETS OF KEY PLAYERS FOR INTERVENTION



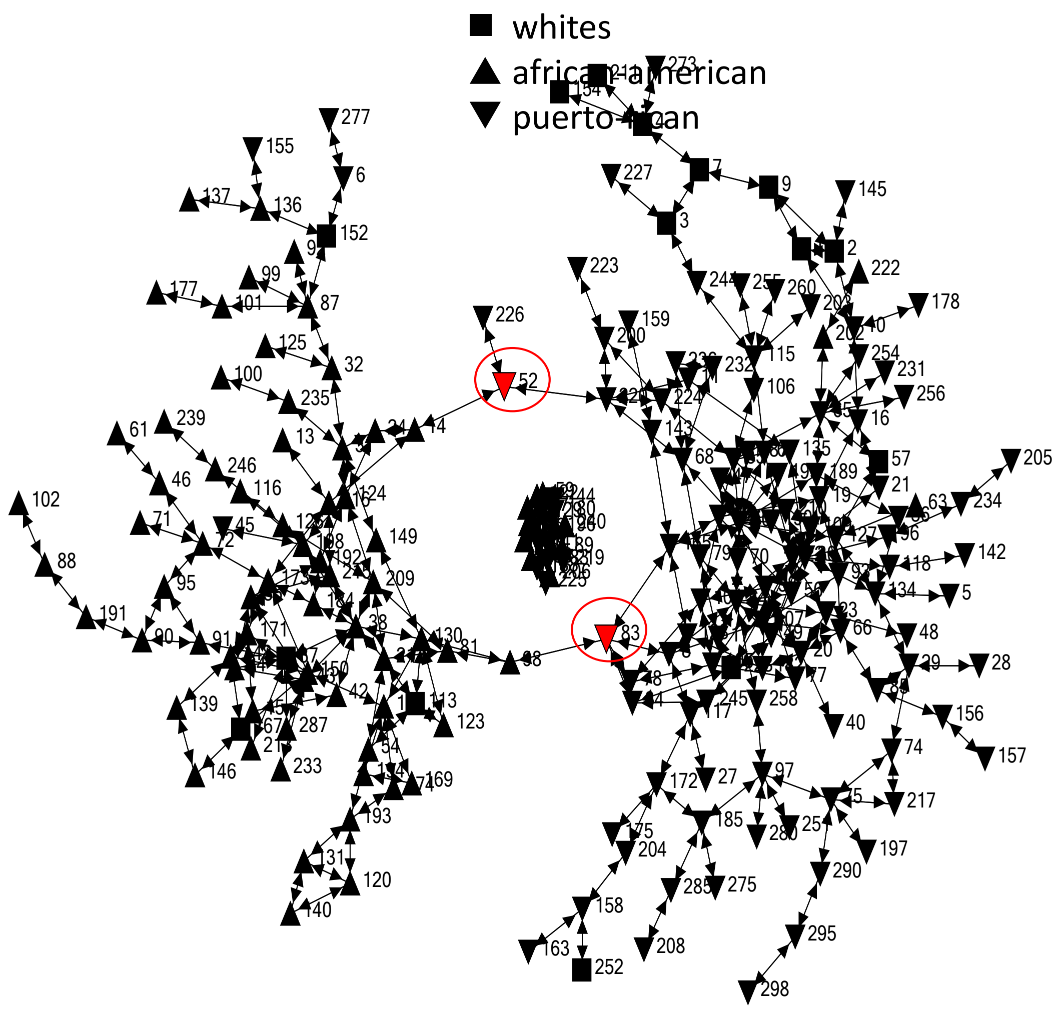
Drug injectors on streets of Hartford

Red nodes comprise set of 14 nodes that, together, can reach more than 50% of network in one step

Disruption Example – health context

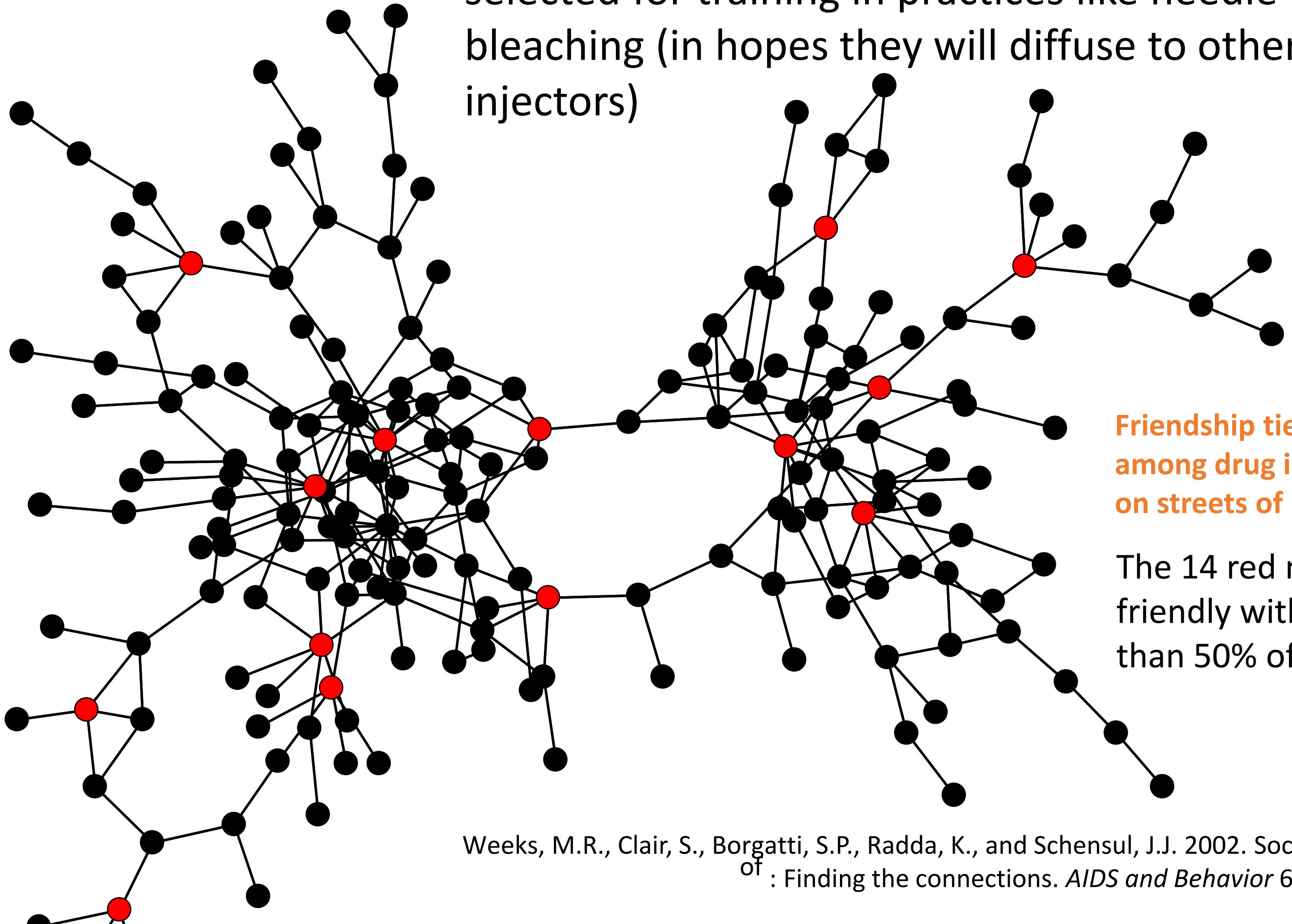
- Which two people should be isolate slow the spread of HIV?
  - KeyPlayer algorithm dc identifies the two red nodes

Friendship ties among drug injectors on streets of Hartford



Influence Example – health context

Which small set of drug injectors should be selected for training in practices like needle bleaching (in hopes they will diffuse to other injectors)



Friendship ties among drug injectors on streets of Hartford

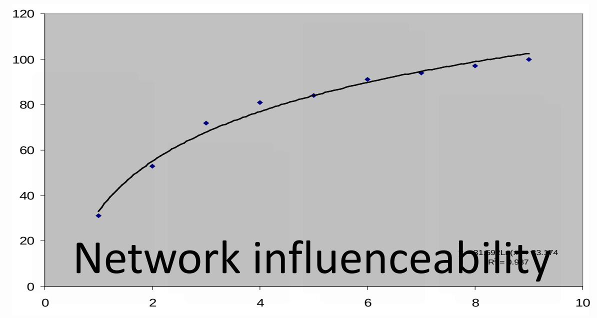
The 14 red nodes are friendly with more than 50% of network

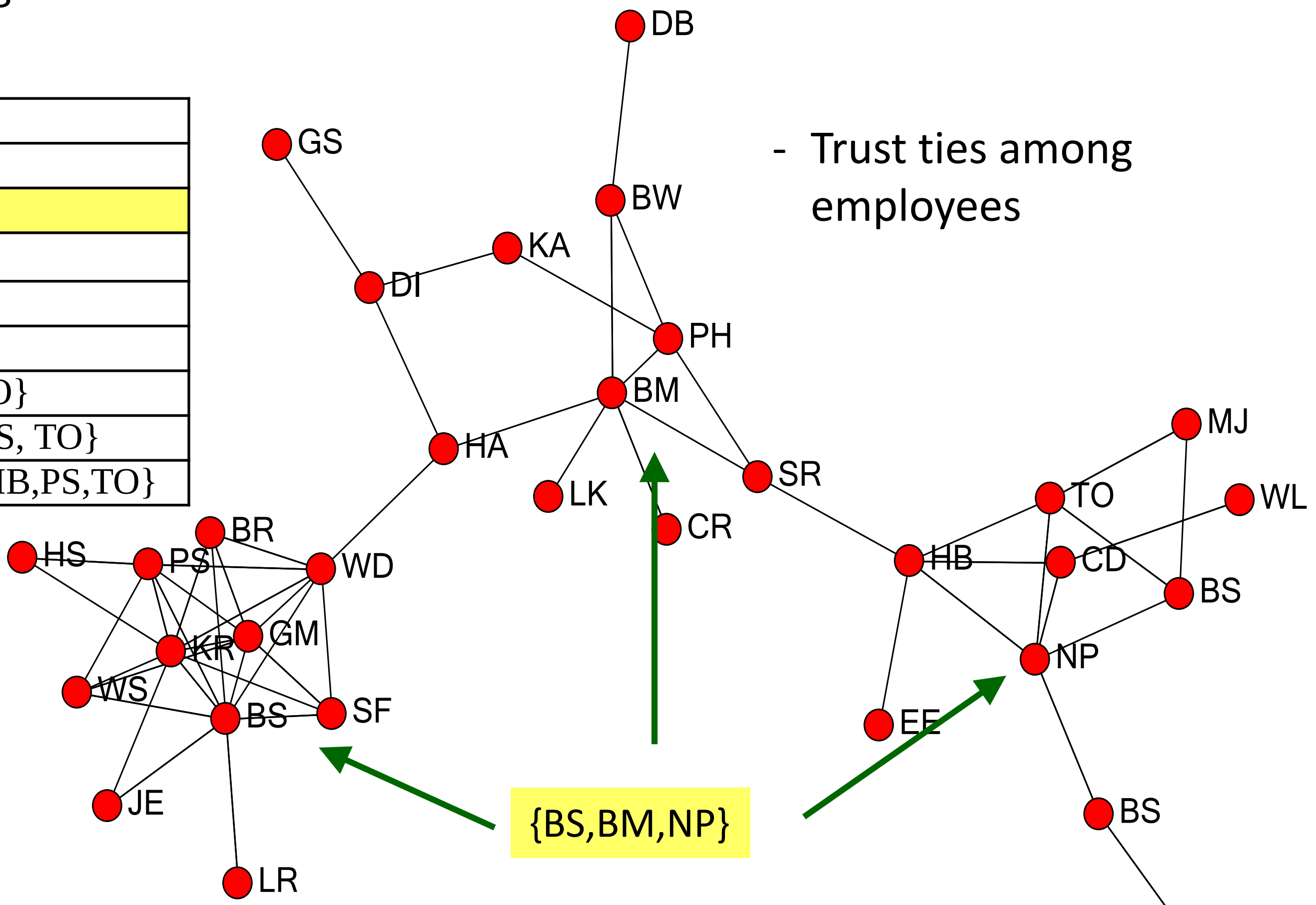
Weeks, M.R., Clair, S., Borgatti, S.P., Radda, K., and Schensul, J.J. 2002. Social networks of : Finding the connections. *AIDS and Behavior* 6(2): 193-206

Influence Example – mgmt context

- Major change initiative is planned. Which small set of employees should we select for intensive indoctrination? in hopes they will diffuse positive attitude/knowledge to others

K	%	KP-Set
1	31	{KR}
2	53	{BM,BS}
3	72	{BM,BS,NP}
4	81	{BM,BS,DI,NP}
5	84	{BM,BS,DI,KR,NP}
6	91	{BM,BS,DI,HB,KR,TO}
7	94	{BM,BS,BS2,DI,HB,PS,TO}
8	97	{BM,BS,BS2,CD,DI,HB,PS, TO}
9	100	{BM,BS,BW,BS2,CD,DI,HB,PS,TO}





a from: Cross, R., Parker, A., & Borgatti, S.P. 2002. Making Invisible Work Visible: Using Social Network Analysis to Support Strategic Collaboration. *California Management Review*. 44(2): 25-46

Dyadic Cohesion

- Adjacency
  - Strength of tie ← Average is density
  - Reciprocity
- Reachability ← 1- f(Average) is fragmentation Or distance weighted fragmentation
  - A path exists or does not (usually as $1/d_{ij}$)
- Distance ← Average is average distance
  - Length of shortest path between two nodes
  - # Geodesics (how many paths of this length)
- Multiplexity
  - Number of ties of different relations linking two nodes
- Number of paths linking two nodes
  - Edge independent ← Minimum is line connectivity
  - Node independent ← Minimum is point connectivity