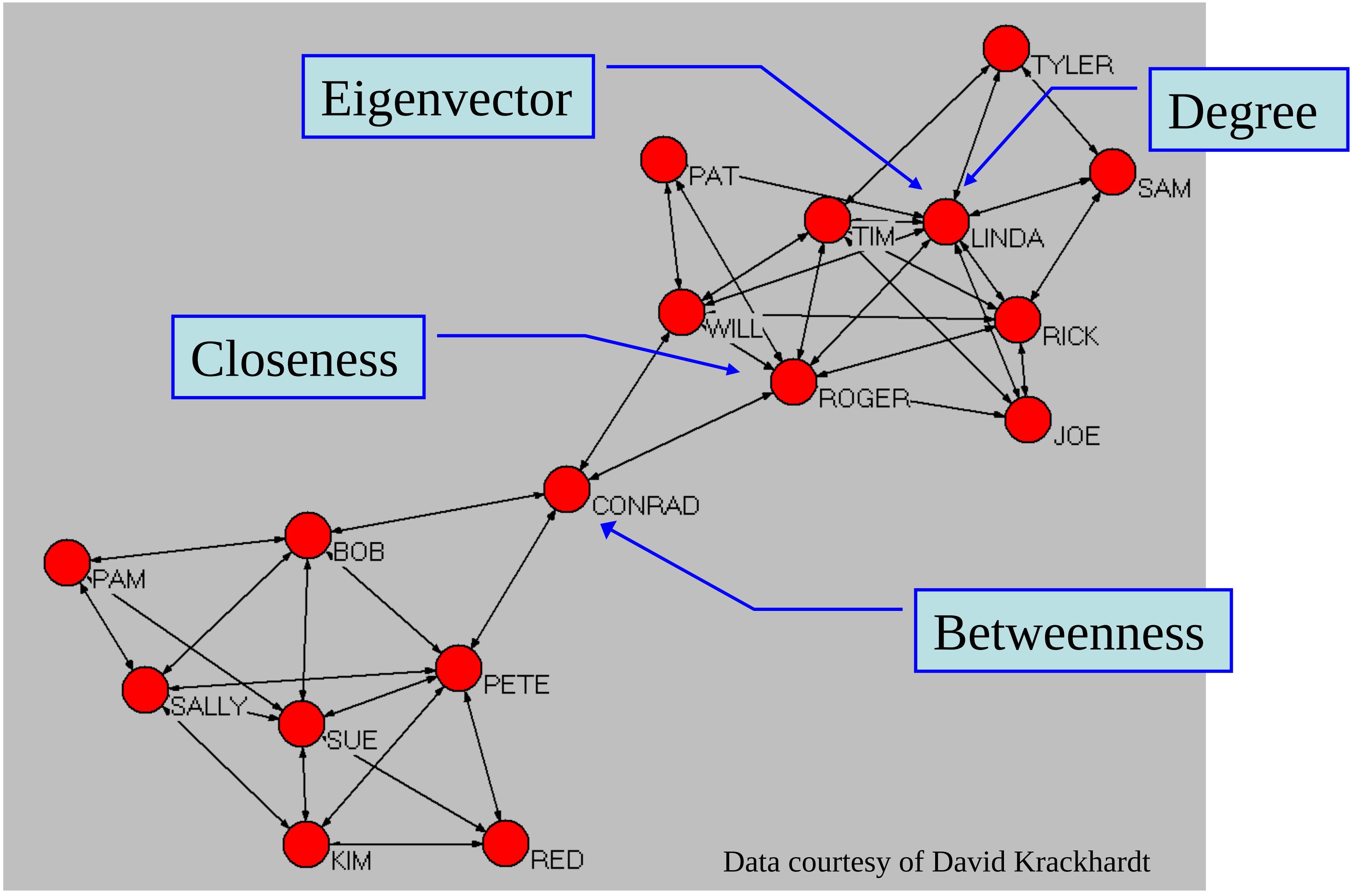
Centrality

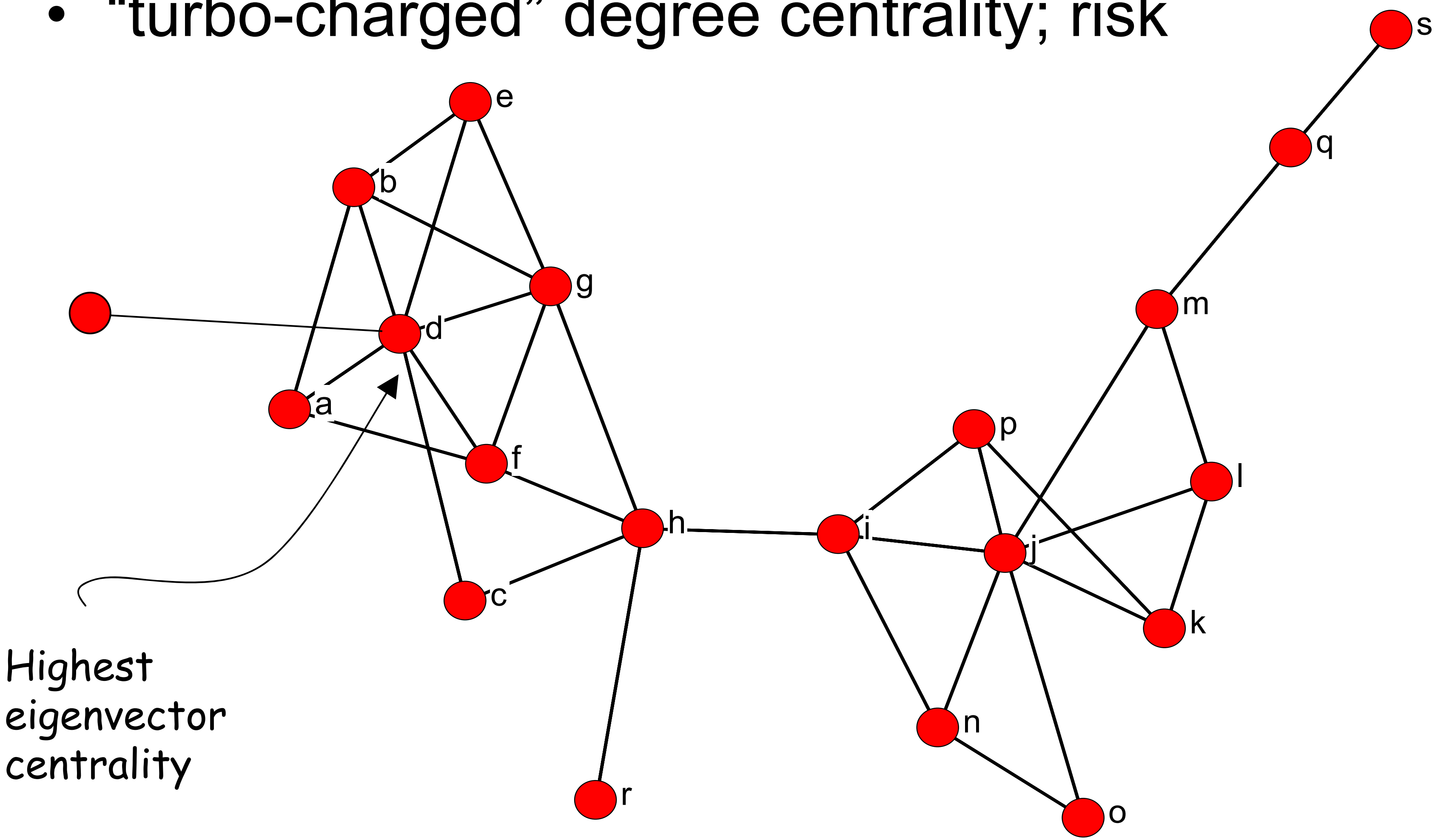
Steve Borgatti

Four Aspects of Centrality



Eigenvector Centrality

- “turbo-charged” degree centrality; risk

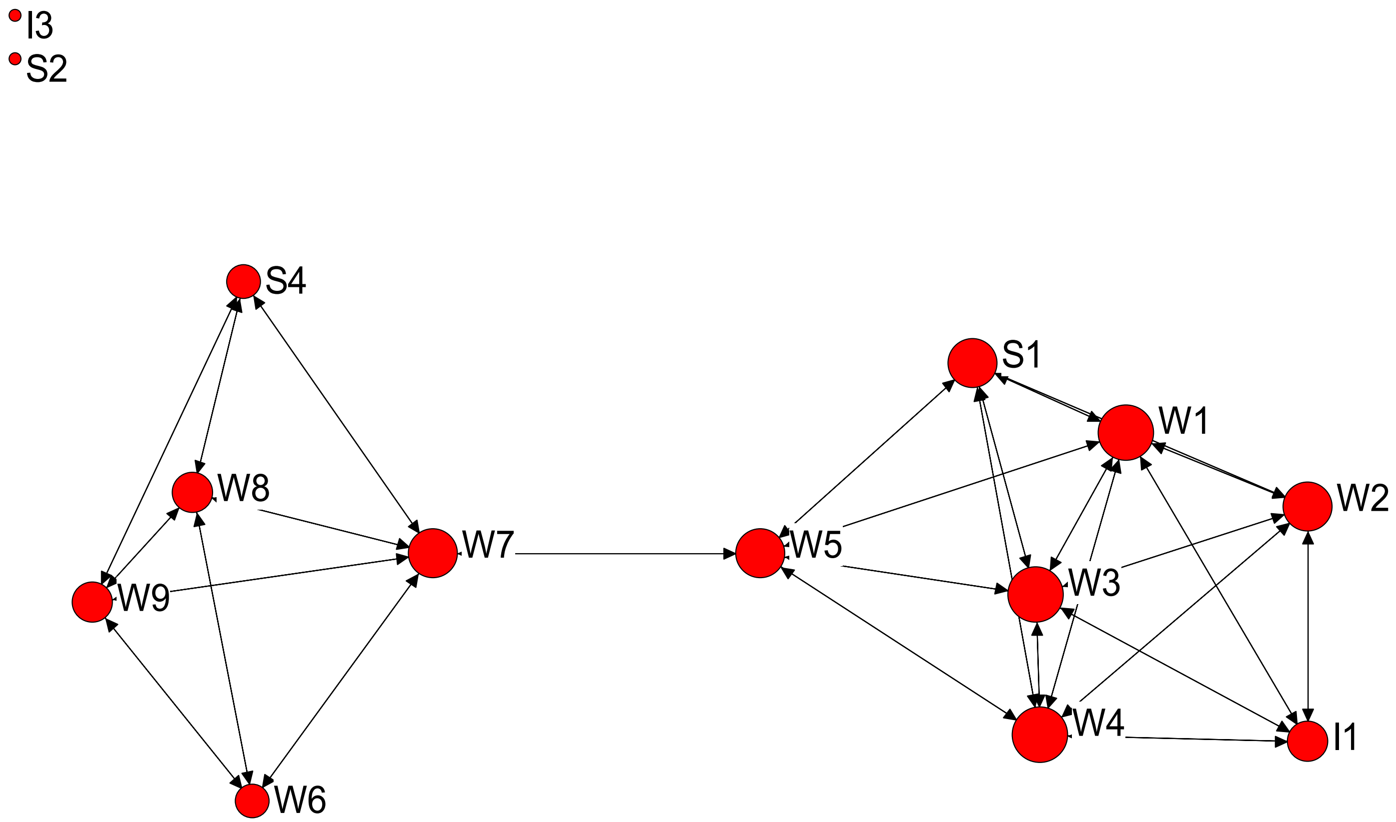


Degree Centrality

- Number of ties that involve a given node
  - Marginals of symmetric adjacency matrix

	I1	I3	W1	W2	W3	W4	W5	W6	W7	W8	W9	S1	S2	S4	Deg
I1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	4
I3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W1	1	0	0	1	1	1	1	0	0	0	0	1	0	0	6
W2	1	0	1	0	1	1	0	0	0	0	0	1	0	0	5
W3	1	0	1	1	0	1	1	0	0	0	0	1	0	0	6
W4	1	0	1	1	1	0	1	0	0	0	0	1	0	0	6
W5	0	0	1	0	1	1	0	0	1	0	0	1	0	0	5
W6	0	0	0	0	0	0	0	0	1	1	1	0	0	0	3
W7	0	0	0	0	0	0	1	1	0	1	1	0	0	1	5
W8	0	0	0	0	0	0	0	1	1	0	1	0	0	1	4
W9	0	0	0	0	0	0	0	1	1	1	0	0	0	1	4
S1	0	0	1	1	1	1	1	0	0	0	0	0	0	0	5
S2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S4	0	0	0	0	0	0	0	0	1	1	1	0	0	0	3

Wiring/Games Degree



	Deg
I1	4
I3	0
W1	6
W2	5
W3	6
W4	6
W5	5
W6	3
W7	5
W8	4
W9	4
S1	5
S2	0
S4	3

Degree Centrality

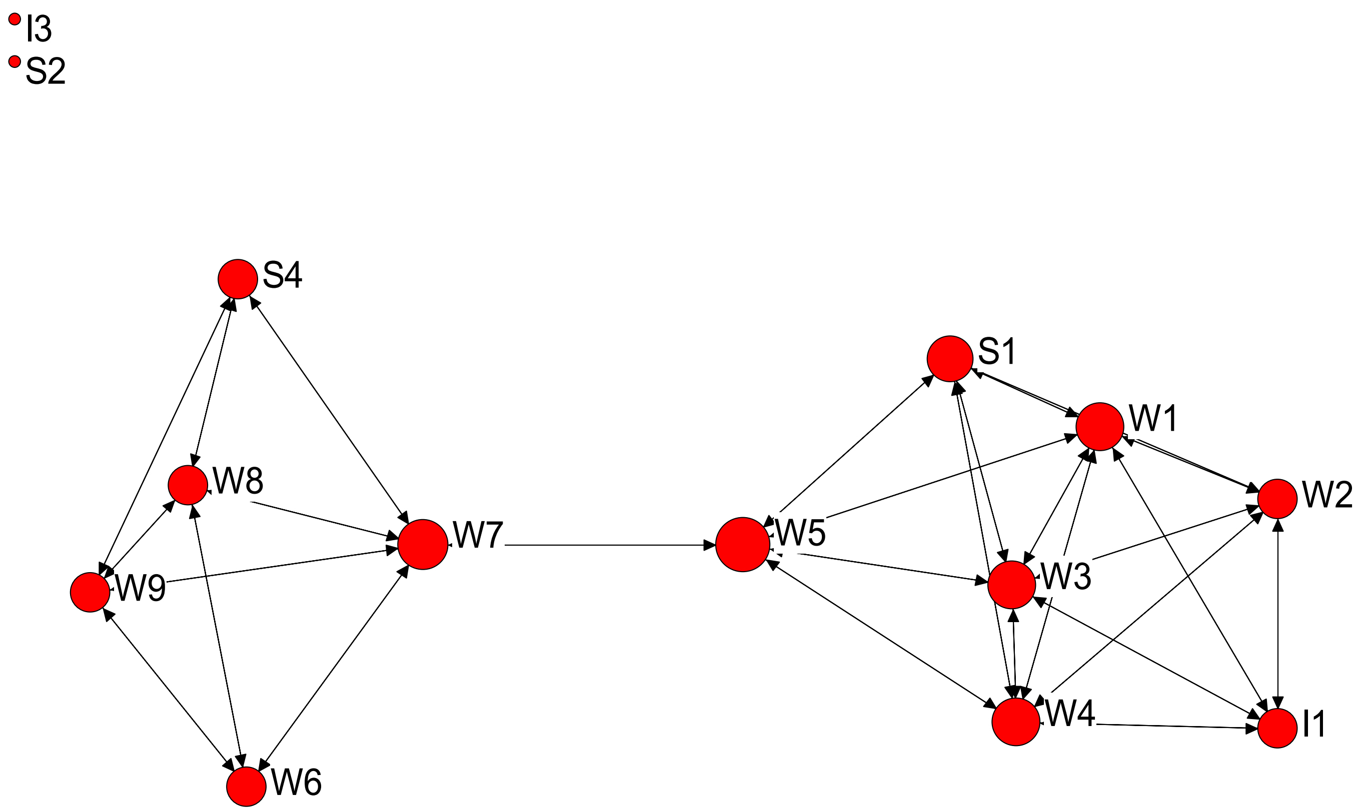
- Index of exposure to what is flowing through the network
  - Gossip network: central actor more likely to hear a given bit of gossip
- Interpreted as opportunity to influence & be influenced directly
- Predicts variety of outcomes from virus resistance to power & leadership to job satisfaction to knowledge

Closeness Centrality

- Sum of distances to all other nodes
  - Computed as marginals of symmetric geodesic distance matrix

	I1	I3	W1	W2	W3	W4	W5	W6	W7	W8	W9	S1	S2	S4	Clo
I1	0	?	1	1	1	1	2	4	3	4	4	2	?	4	27
I3	?	0	?	?	?	?	?	?	?	?	?	?	?	?	0
W1	1	?	0	1	1	1	1	3	2	3	3	1	?	3	20
W2	1	?	1	0	1	1	2	4	3	4	4	1	?	4	26
W3	1	?	1	1	0	1	1	3	2	3	3	1	?	3	20
W4	1	?	1	1	1	0	1	3	2	3	3	1	?	3	20
W5	2	?	1	2	1	1	0	2	1	2	2	1	?	2	17
W6	4	?	3	4	3	3	2	0	1	1	1	3	?	2	27
W7	3	?	2	3	2	2	1	1	0	1	1	2	?	1	19
W8	4	?	3	4	3	3	2	1	1	0	1	3	?	1	26
W9	4	?	3	4	3	3	2	1	1	1	0	3	?	1	26
S1	2	?	1	1	1	1	1	3	2	3	3	0	?	3	21
S2	?	?	?	?	?	?	?	?	?	?	?	?	0	?	0
S4	4	?	3	4	3	3	2	2	1	1	1	3	?	0	27

Wiring/Games Closeness



	Clo
I1	27
I3	
W1	20
W2	26
W3	20
W4	20
W5	17
W6	27
W7	19
W8	26
W9	26
S1	21
S2	
S4	27

Closeness Centrality

- Is an inverse measure of centrality
- Index of expected time until arrival for given node of whatever is flowing through the network
  - Gossip network: central player hears things first

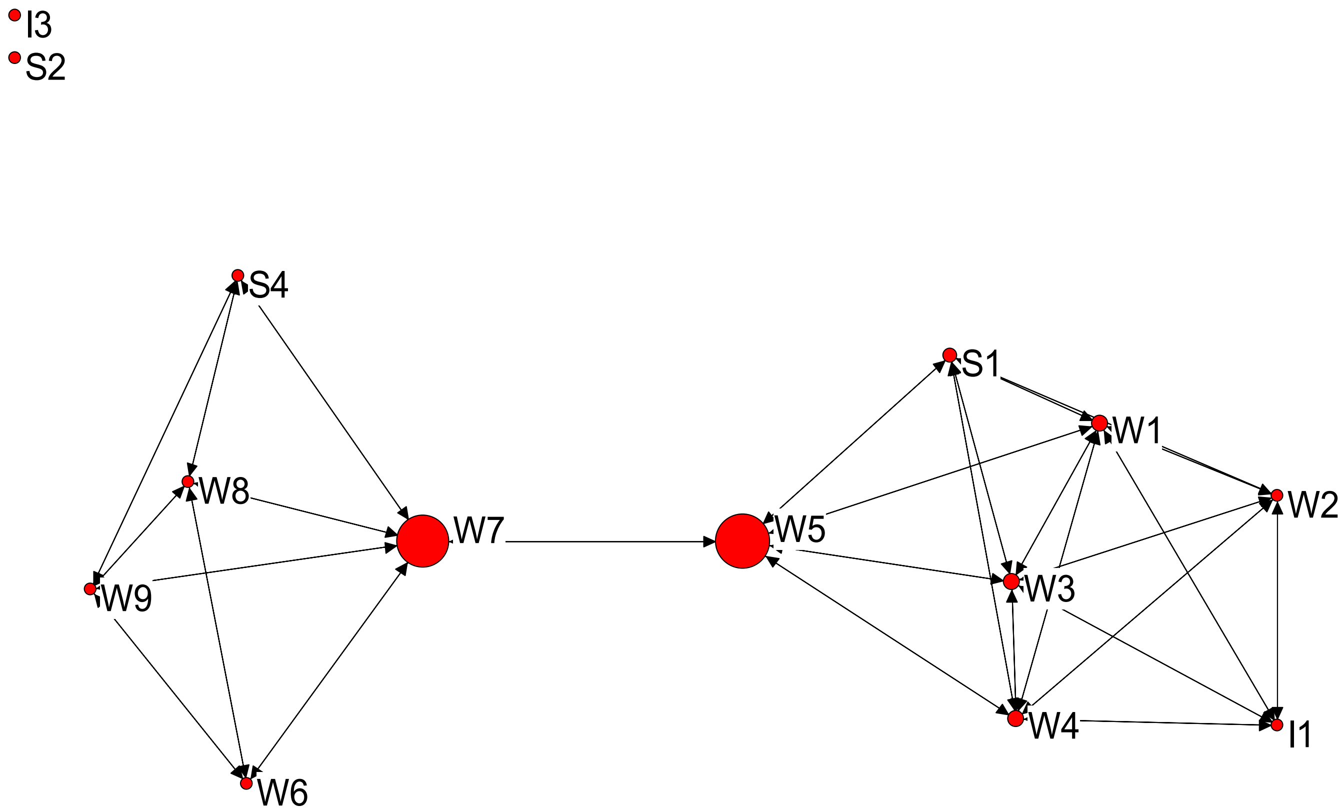
Betweenness Centrality

- How often a node lies along the shortest path between two other nodes
  - Computed as:

$$b_k = \sum_{i,j} \frac{g_{ikj}}{g_{ij}}$$

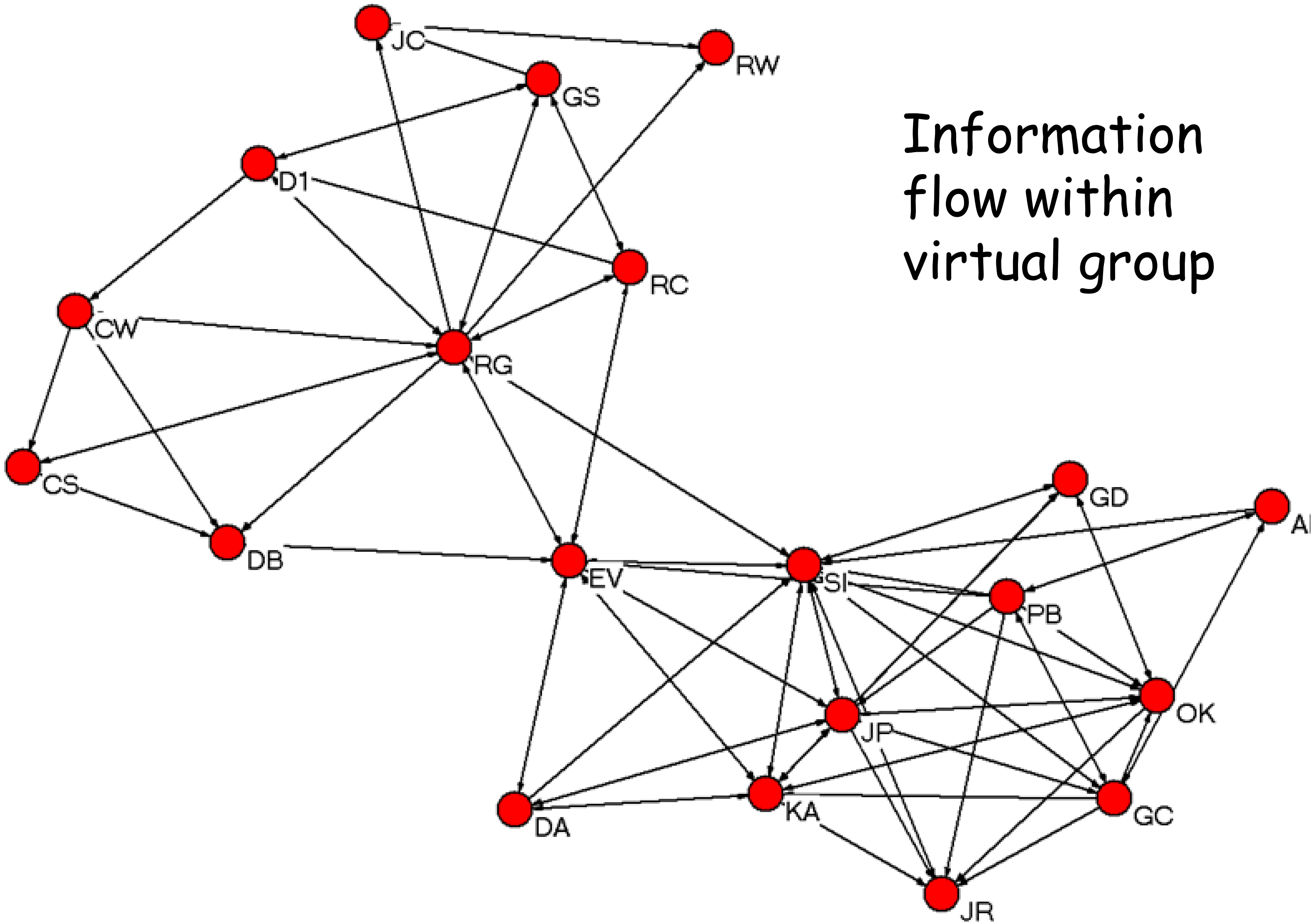
    where gij is number of geodesic paths from i to j and gikj is number of those paths that pass through k
- Index of potential for gatekeeping, brokering, controlling the flow, and also of liaising otherwise separate parts of the network;
- Interpreted as indicating power and access to diversity of what flows; potential for synthesizing

Wiring/Games Betweenness



	Bet
I1	0.00
I3	
W1	6.82
W2	0.45
W3	6.82
W4	6.82
W5	54.55
W6	0.00
W7	51.52
W8	0.61
W9	0.61
S1	2.73
S2	
S4	0.00

Local Gain is Global Pain



Information flow within virtual group

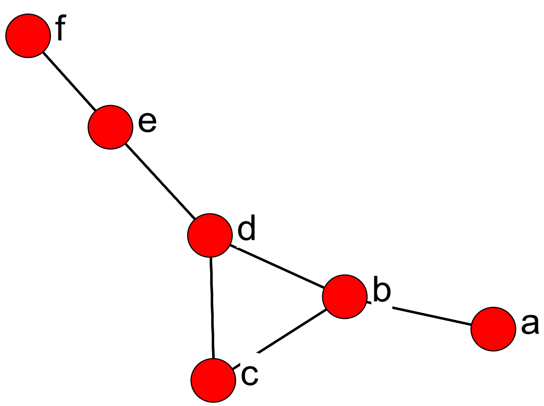
Data collected by Cross

Eigenvector Centrality

- Node has high score if connected to many nodes are themselves well connected
  – Computed as: $$\lambda v = Av$$

    where A is adjacency matrix and V is eigenvector centrality. V is the principal eigenvector of A.
- Indicator of popularity, “in the know”
- Like degree, is index of exposure, risk
- Tends to identify centers of large cliques

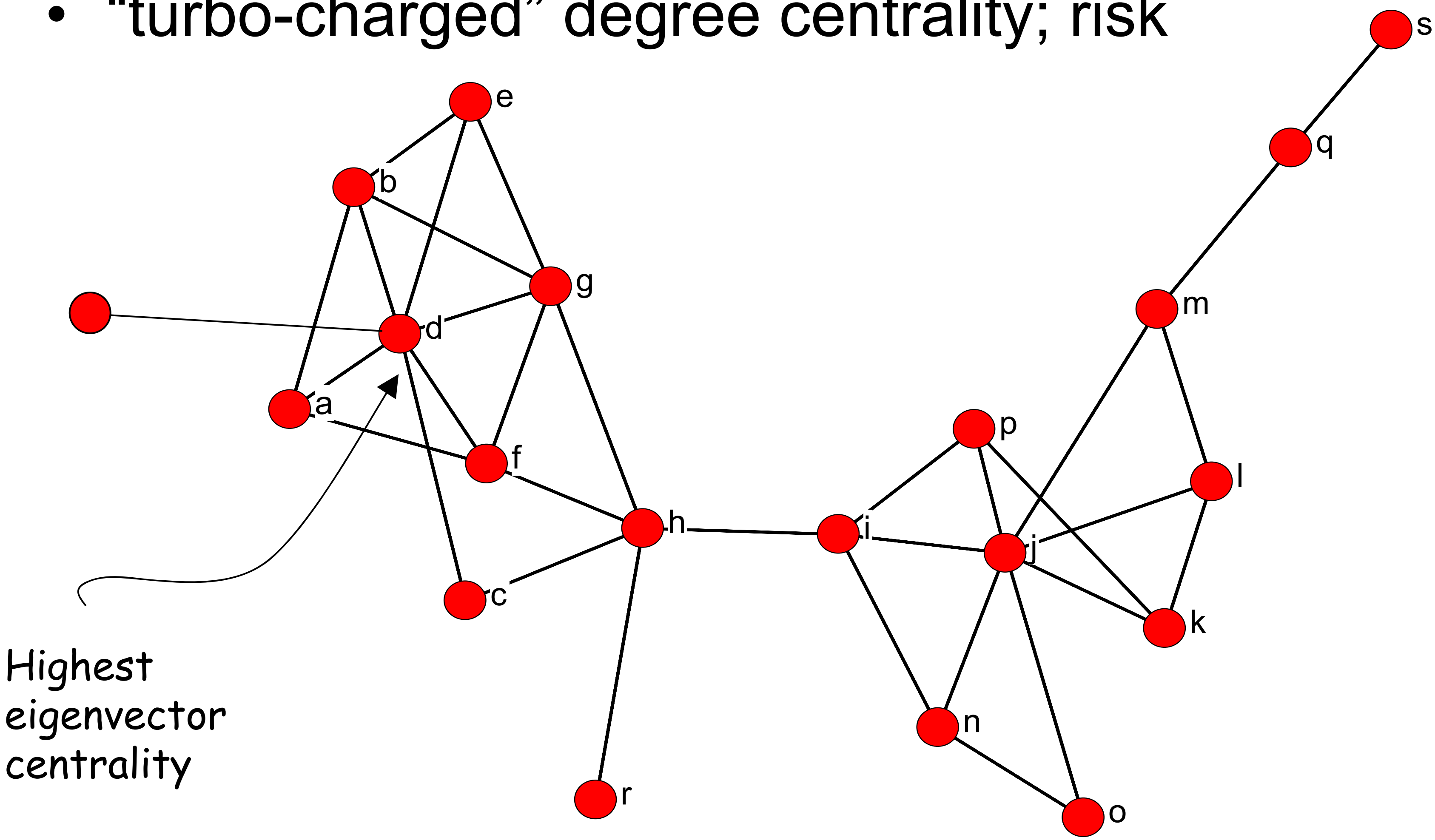
	a	b	c	d	e	f	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
a	0	1	0	0	0	0	1	3	6	16	35	86	195	465	1071	2524
b	1	0	1	1	0	0	3	6	16	35	86	195	465	1071	2524	5854
c	0	1	0	1	0	0	2	6	13	32	73	173	401	940	2190	5117
d	0	1	1	0	1	0	3	7	16	38	87	206	475	1119	2593	6086
e	0	0	0	1	0	1	2	4	9	20	47	107	253	582	1372	3175
f	0	0	0	0	1	0	1	2	4	9	20	47	107	253	582	1372



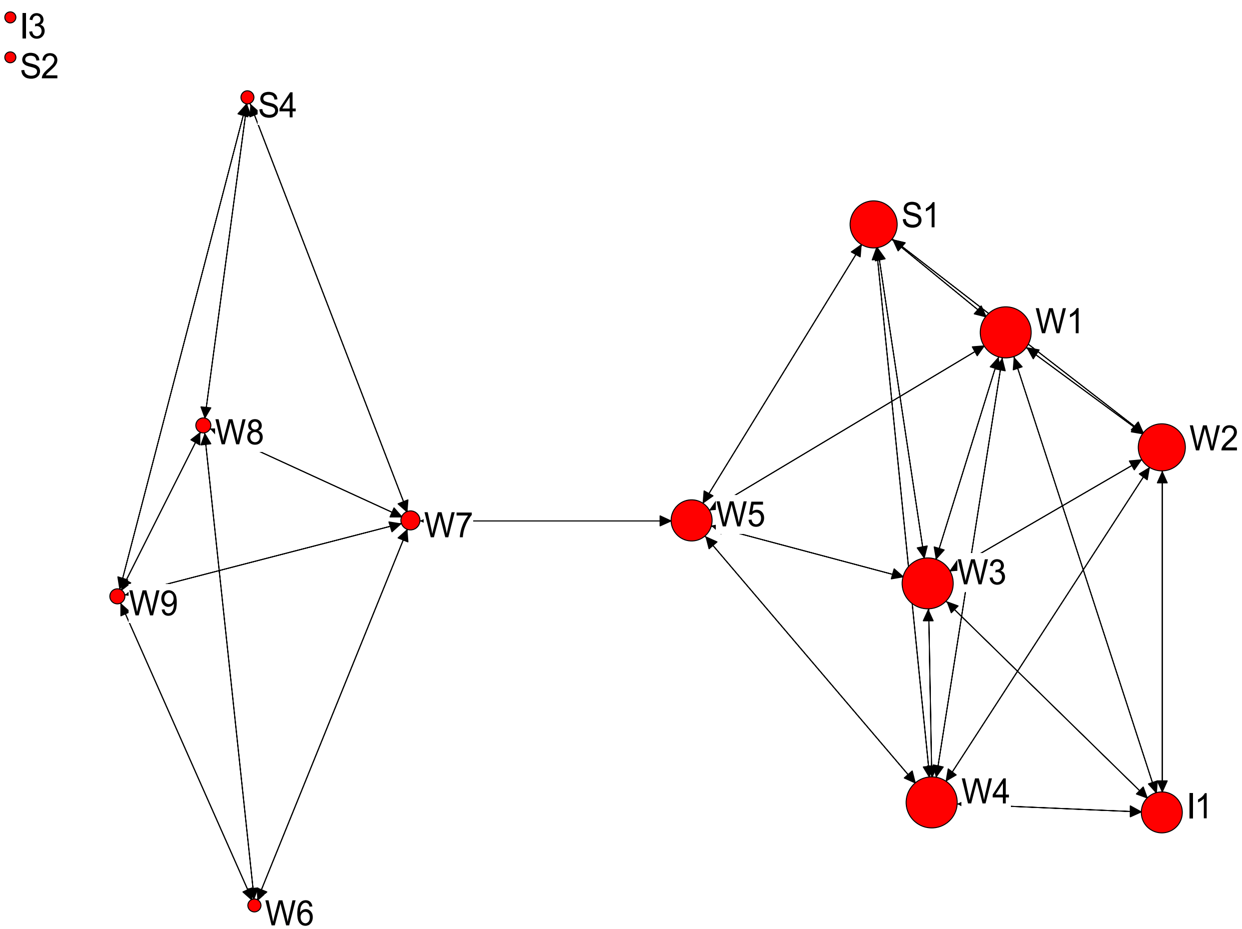
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
A	8.3	10.7	9.4	10.7	10.1	10.6	10.3	10.5	10.4	10.5
B	25.0	21.4	25.0	23.3	24.7	24.0	24.5	24.2	24.4	24.3
C	16.7	21.4	20.3	21.3	21.0	21.3	21.1	21.2	21.2	21.2
D	25.0	25.0	25.0	25.3	25.0	25.3	25.1	25.3	25.1	25.2
E	16.7	14.3	14.1	13.3	13.5	13.1	13.3	13.1	13.3	13.2
F	8.3	7.1	6.3	6.0	5.7	5.8	5.6	5.7	5.6	5.7

Eigenvector Centrality

- "turbo-charged" degree centrality; risk

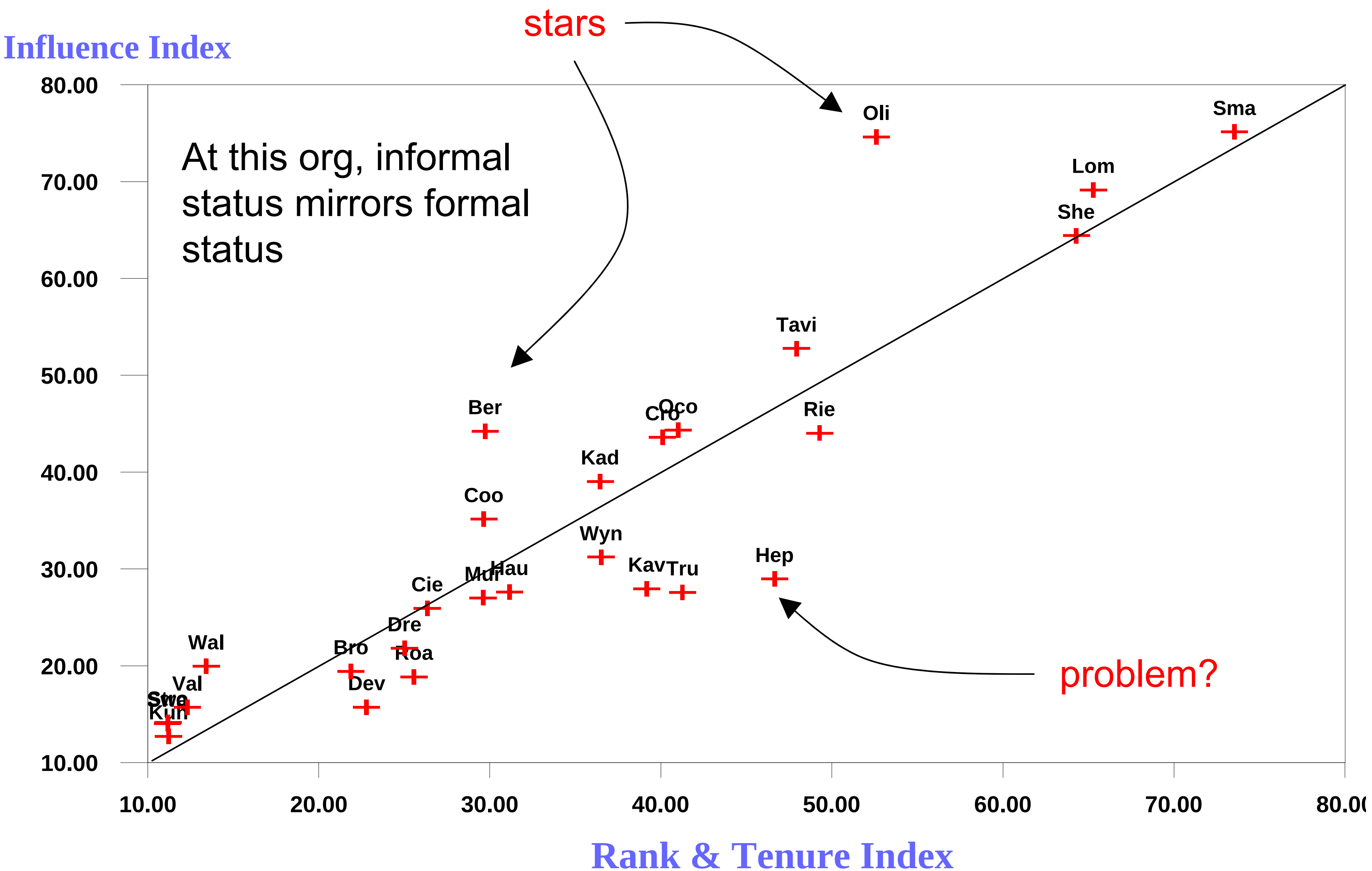


Wiring/Games Eigenvector



	Eig
I1	43
W1	59
W2	52
W3	59
W4	59
W5	46
W6	4
W7	12
W8	5
W9	5
S1	52
S4	4

Influence Network



Walk-Based Measures

- Multiple motivations
  - actor's status is function of not only the number of people who choose them, but their status
  - in an influence process, an actor's impact on another is function of all sequences (walks) linking them
- Resulting measures are similar / related

Influence Approach

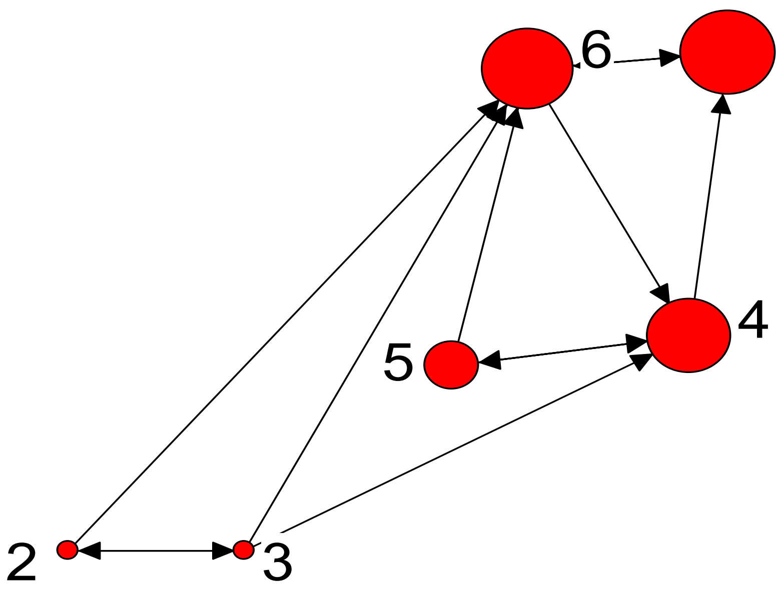
- Variations by Katz, Friedkin, Taylor, etc.
- Generic approach
  - R is network matrix, α is attenuation parameter
  - $\mathbf{Q} = \alpha^0\mathbf{R}^0 + \alpha^1\mathbf{R}^1 + \alpha^2\mathbf{R}^2 + \alpha^3\mathbf{R}^3 + \ldots \alpha^\infty\mathbf{R}^\infty$
  - $\mathbf{Q} = (\mathbf{I}-\alpha\mathbf{R})^{-1}$, assuming $\alpha^{-1} > \lambda_1$
  - $\mathbf{s} = (\mathbf{I}-\alpha\mathbf{R})^{-1}\mathbf{1} = \mathbf{Q1}$ (row sums of **Q**)

Recursive Status Approach

- Hubbell
  - **s** = **Ws** + **e**, where **W** is adj matrix w/ equal col sums < 1, **s** is vector representing status, **e** is vector of exogeneous inputs (usually 1s)
  - $\mathbf{s} = (\mathbf{I}-\mathbf{W})^{-1}\mathbf{e}$
- Bonacich, Coleman, Burt, etc.
  - Principal eigenvector of **W**
  - $\lambda\mathbf{c} = \mathbf{Wc}$ (or **W'c** if appropriate)

Katz example

Who really knows what's going on?



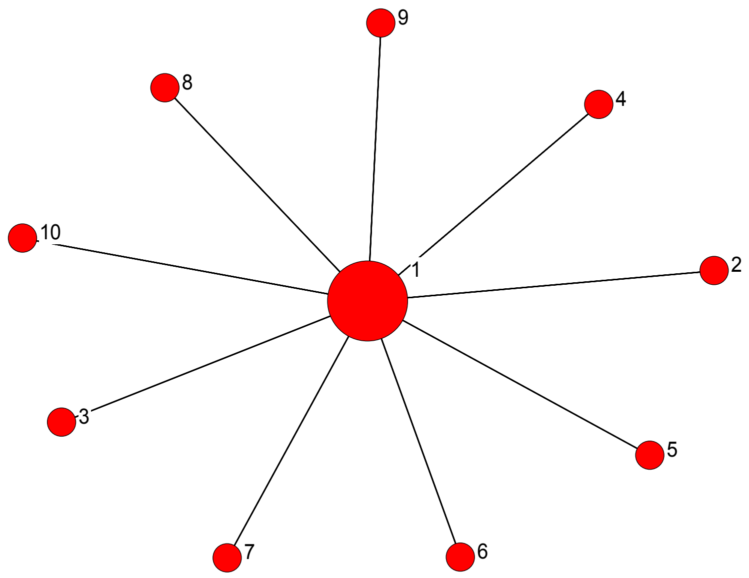
Node	Out	In	Katz
1	1	2	13.0
2	2	1	1.0
3	3	1	1.0
4	2	3	11.4
5	2	1	6.2
6	2	4	12.6

Indegree gives same score to 5 as to 2 and 3. But 5 is chosen by 4, who is chosen by popular nodes like 6. Katz score gives 5 much higher score than 2 or 3. Similarly node 1 has only two incoming choices, but they are from the most sought-after players, so 1 must be even more knowledgeable than they.

Centralization

Definition

- Extent to which network revolves around a single node
- Extent to which the network resembles star shape
- Difference between each node's centrality score and that of the most central node
  - A kind of variance



Calculation

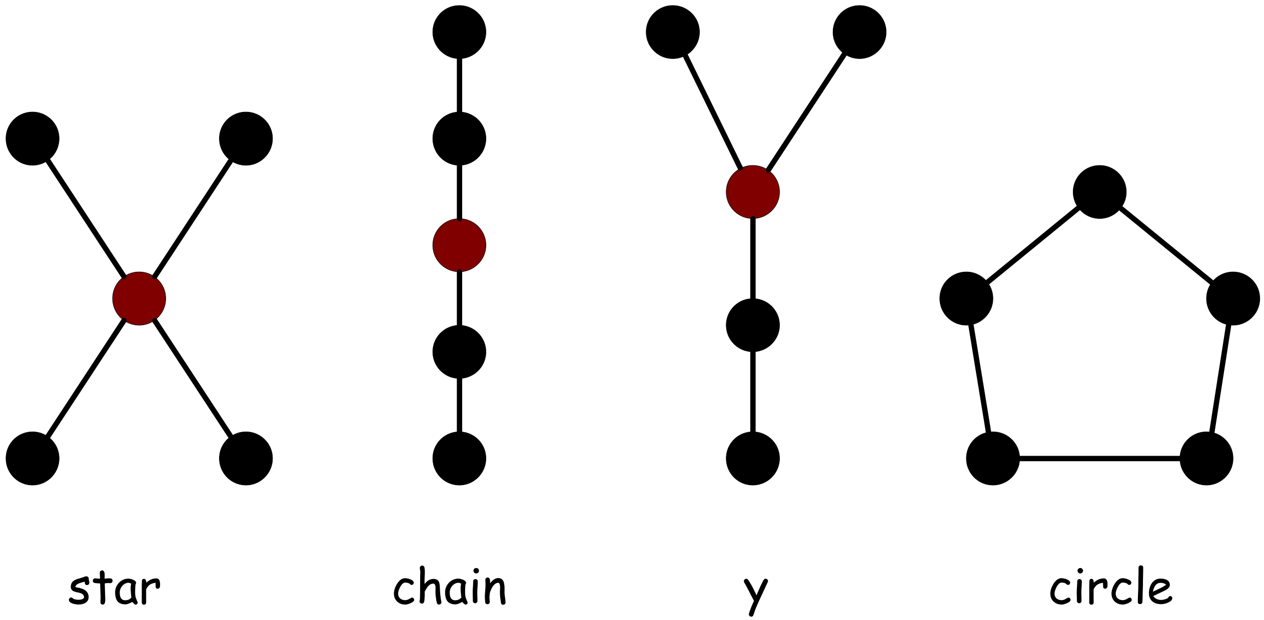
Your graph

$$Centralization = \frac{\sum_i | c_{MAX} - c_i |}{\sum_i | \chi_{MAX} - \chi_i |}$$

Star graph

- C_{MAX} is centrality of the most central node in the observed graph
  - C_i is the centrality of the ith node in the observe graph
- χ_{MAX} is the centrality of the most central node in the star graph
  - χ_i is the centrality of the ith node in the star graph

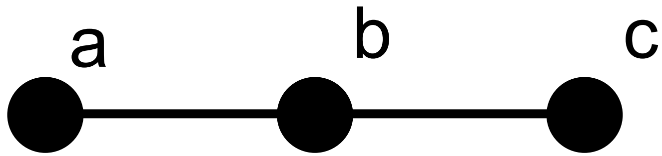
Bavelas/Leavitt Experiments

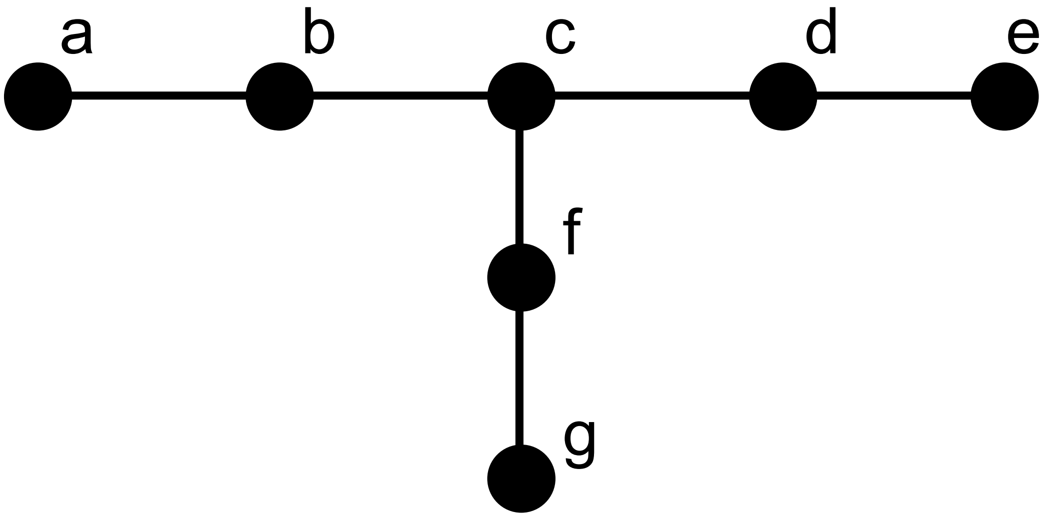


Performance a function of (short) distances from the "information integrator" (typically the node least distant from all others).

Experimental Exchange Nets

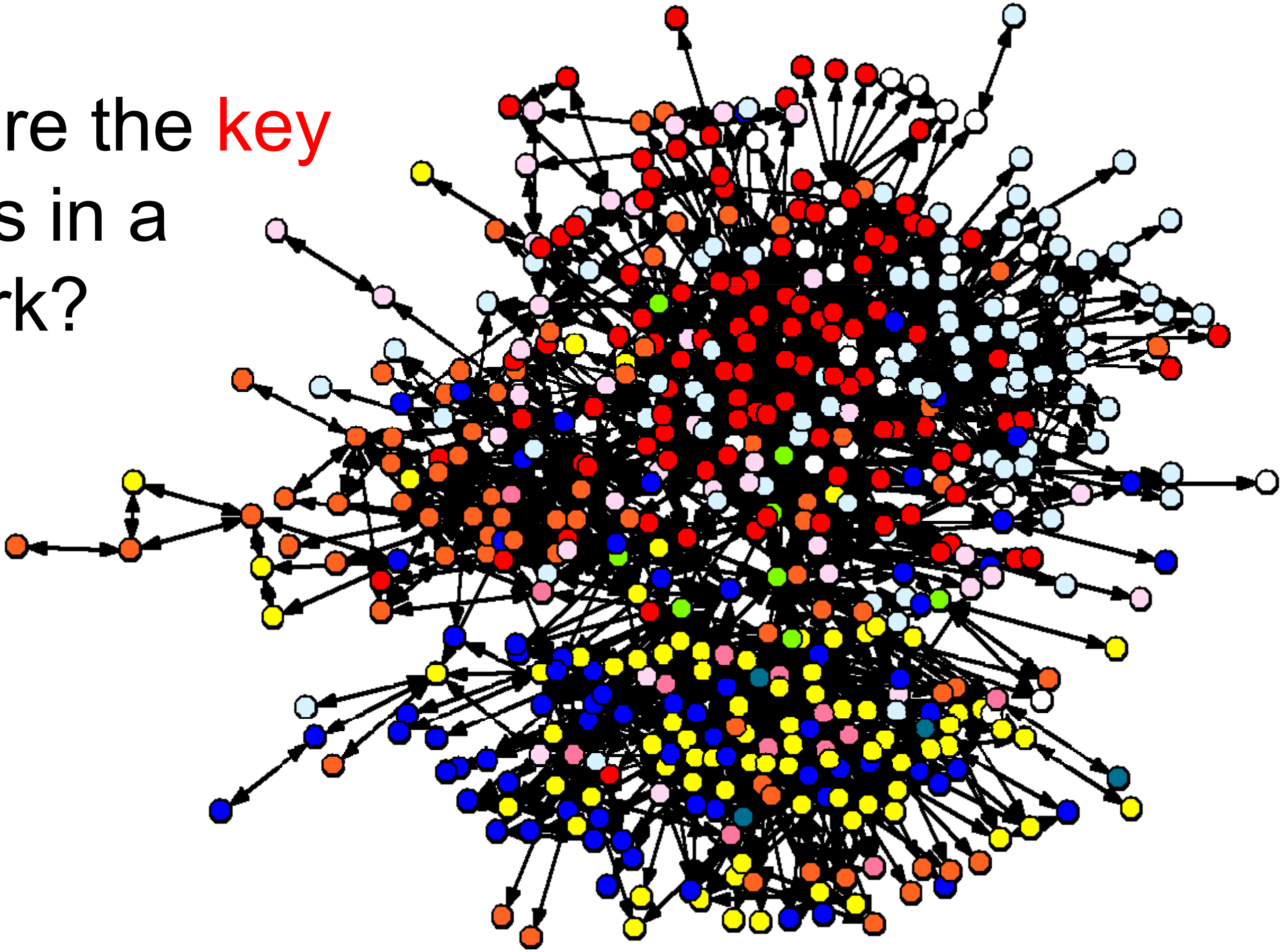
- Divvy up 24 points
- Who has what kinds of outcomes?





Research Question

- Who are the key players in a network?



Initial Answer (from Phase 0)

- It depends on whether we need <u>individual</u> key players or a <u>set</u>
  - I'll focus on the set problem today
- It also depends on the purpose
  - What are the key players <u>for</u>?

What specific problems do we need to solve?

- Network Disruption problem
  - How to maximally disrupt the functioning of a network by intervening with the key players
    - E.g., removing them
- Network Influence problem
  - How to maximally spread ideas, misinforma-tion, materials, diseases, etc. by seeding key players
- Network Surveillance problem
  - How to efficiently learn what the network knows by surveilling key players.

Same under certain conditions

Applications

DISRUPTION		INFLUENCE
Who/how many to **immunize or quarantine** in order to slow spread of infectious disease	Health Care ↔	Selecting peer health advocates for **diffusing** safe practices (e.g. bleaching) and material
Who to **arrest or discredit** to disrupt criminal networks	Criminal Justice ↔	Who to “turn” or feed **false information** to
Where is an organization most vulnerable to **turnover**?	Management ↔	Select subset of employees for **intervention** prior to change initiative

The Naïve Approach

- Open the SNA toolbox and pull out node centrality
- Specifically
  1. Measure the individual network centrality of each node
     - Choose appropriate centrality measure for each problem, e.g.:
       - Betweenness for DISRUPTION problem
       - Katz's measure for INFLUENCE problem
       - In-Closeness for SURVEILLANCE problem
  2. Then select the *k* nodes that are most central

The naïve approach fails for 2 reasons

The **design** issue

Centrality measures not specifically designed for our specific problems, so are sub-optimal

The **ensemble** issue

Centrality measures are node-level, not group-level concepts.

The optimal set of players is not the same as the set of players that are individually optimal

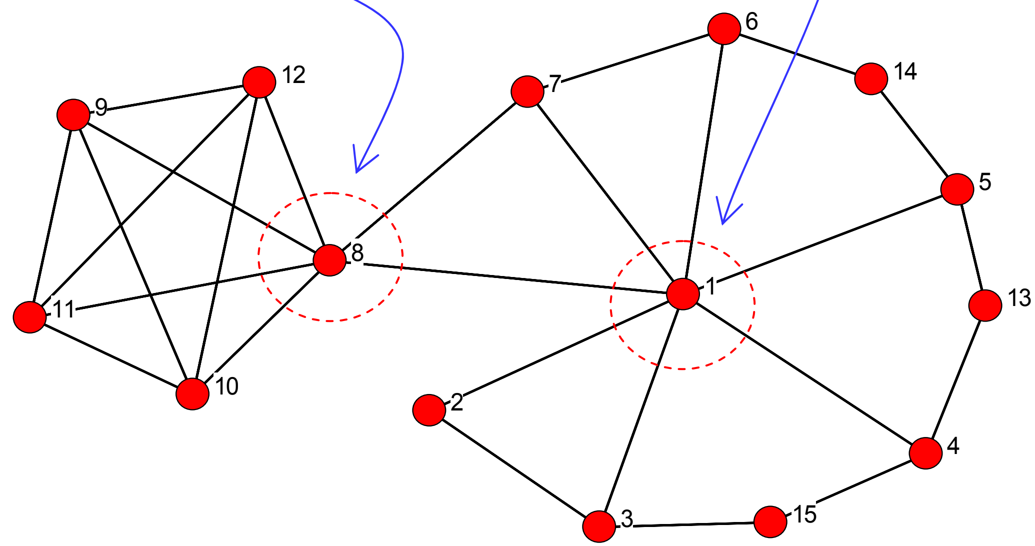
Detroit Pistons vs. the U.S. Men's Olympic Basketball "team"

Illustrating the issues using the DISRUPTION problem

Which nodes to remove from network in order to maximally fragment the network?

The Design Issue

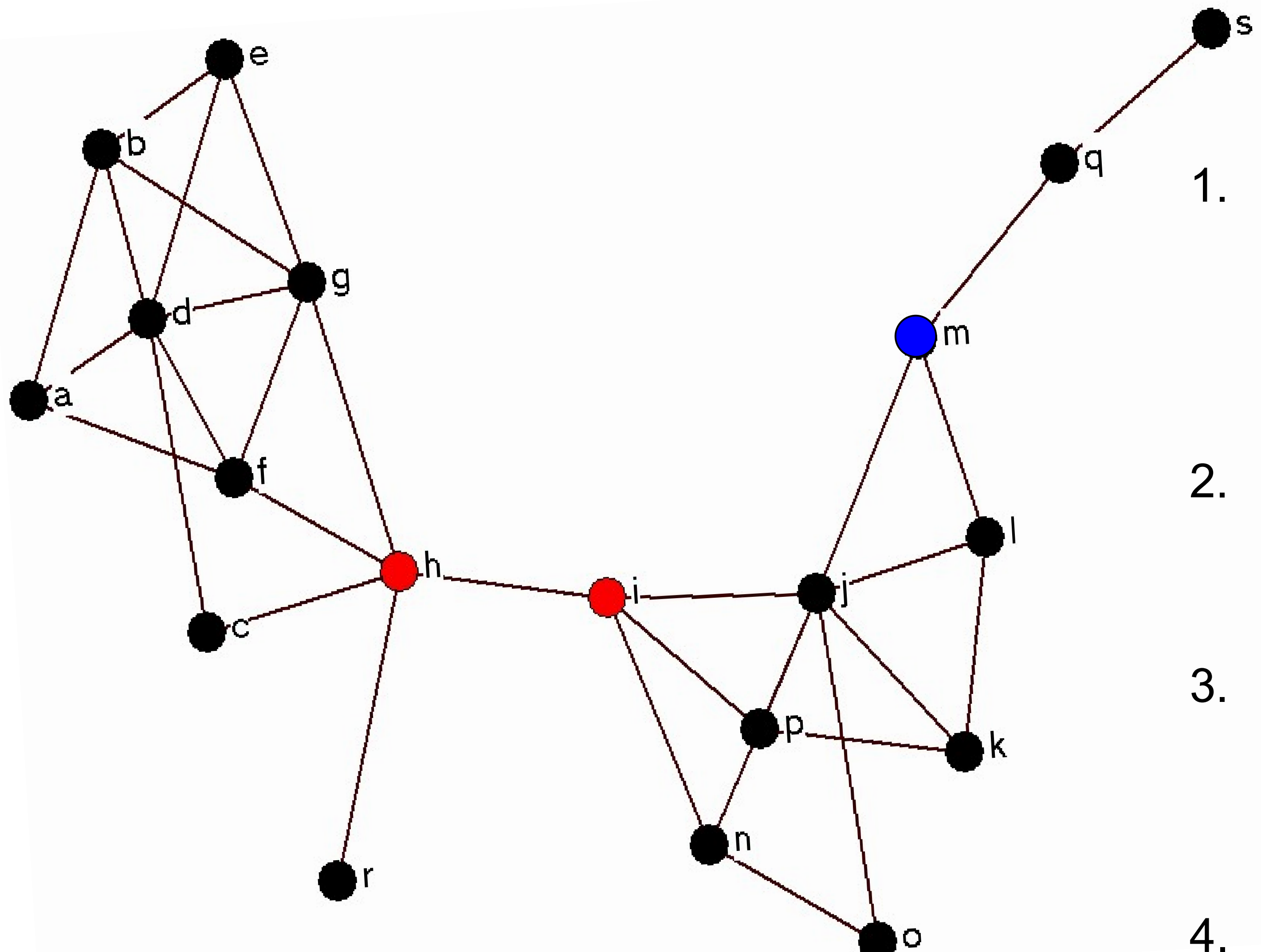
- Node 1 has highest betweenness centrality, but deleting it …
  - does not disconnect the network
  - And no other existing measure is any better
- In contrast, deleting n
  2 components
  - Yet node 8 is not highest in centrality
  - So centrality is not optimal



DISRUPTION

The Ensemble Issue

(Disrupt network by fragmenting)



1. Nodes *h* and *i* are individually optimal-- deleting either will fragment the graph

2. But deleting {*h*,*i*} is no better than deleting {*h*} alone

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

4. Problem is *h* and *i* are redundant

Solve two issues: design & ensemble

Design

Develop measures of node suitability specifically designed for the DISRUPTION, INFLUENCE & SURVEILLANCE problems

Ensemble

Generalize the measures to apply to sets as well as individual nodes

Employ combinatorial optimization algorithm for selecting set of nodes that would maximize the new

A measure for network disruption

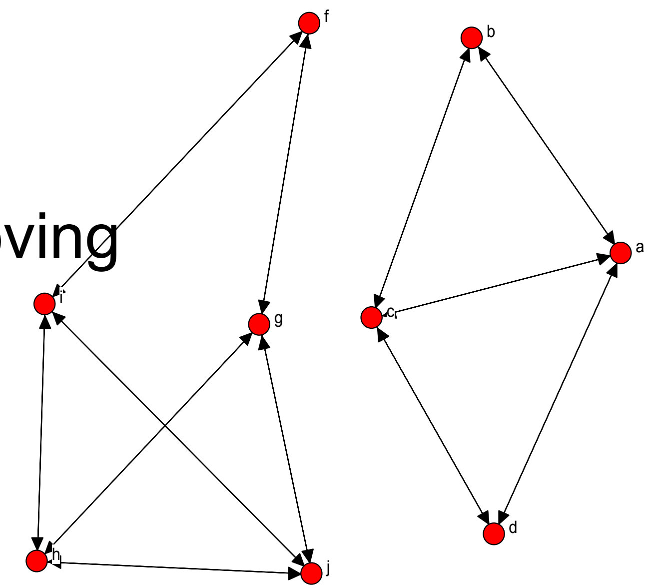
DISRUPTION

- Goal: measure fragmentation of network that remains after removing key player set
- Simplest measure is the component ratio, which counts the number of fragments in the network [after removing key players]
  - divided by number of nodes

$$C = \frac{\text{\# of components}}{n}$$

  - Where n is num. of nodes in network
- Problems:

2 components in this network

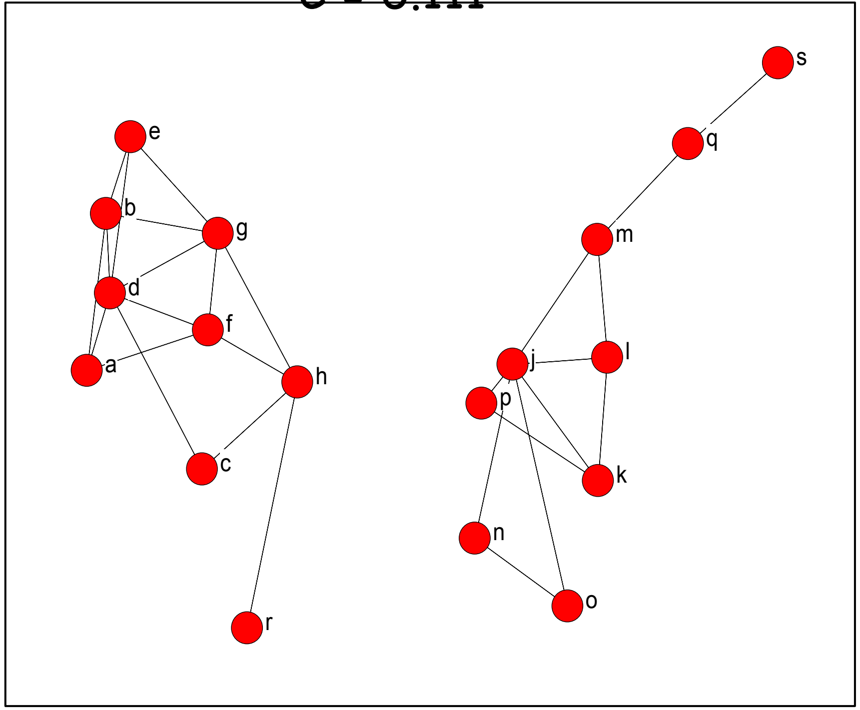


C = 2/9 = 0.22

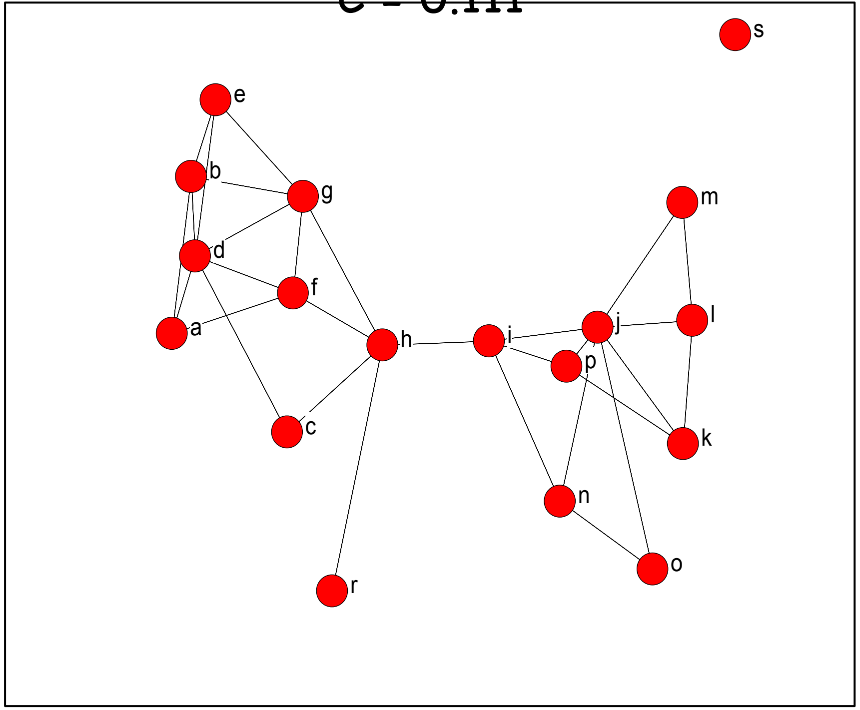
Problems w/ component ratio:

- In this measure, the two networks below are considered equally fragmented

C = 0.111



C = 0.111



- Yet intuitively the one on the left seems more fragmented

DISRUPTION

"Fragmentation" Measure

- Defined as proportion of pairs of nodes that are disconnected from each other
- Given matrix R such that r_{ij} = 1 if node *i* can reach node *j* via a path of any length and r_{ij} = 0 otherwise

Formalization step: (ivory tower)

$$F = 1 - \frac{2\sum_{j<i} r_{ij}}{n(n-1)}$$

= "Fragmentation"

- Since all pairs within a component are mutually reachable, a more economical computational formula is possible:

Implementation step: (real world)

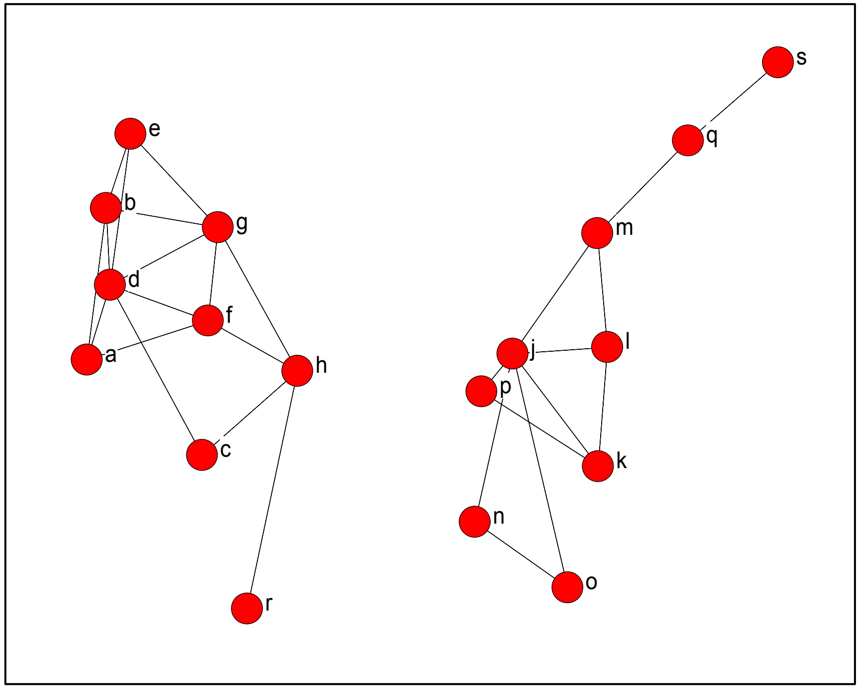
$$F = 1 - \frac{\sum_k s_k(s_k - 1)}{n(n-1)}$$

where s_k is number of nodes in *k*th component

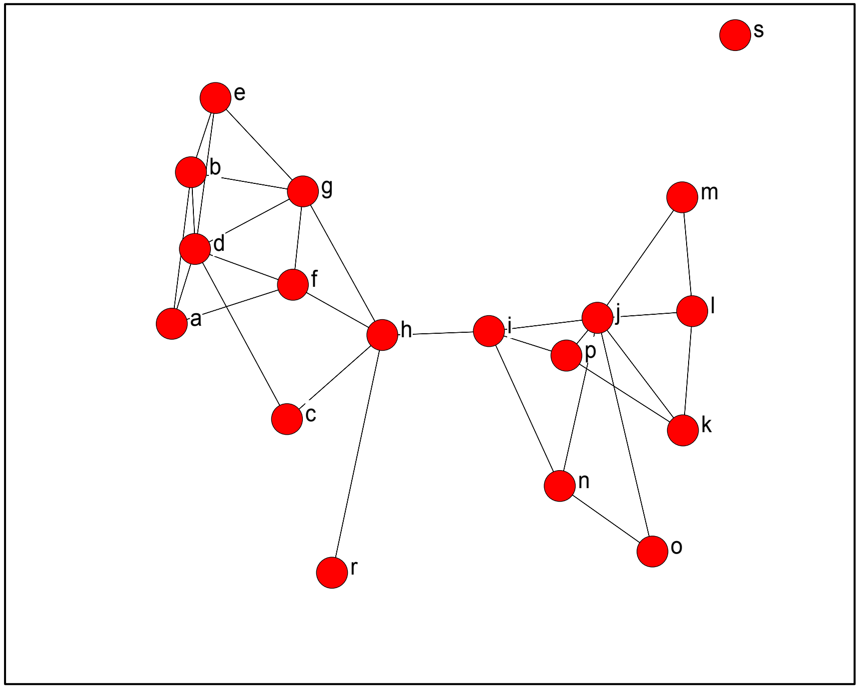
Features of fragmentation measure

- Yields higher value for two large components than for one large component and one small

F = 0.529



F = 0.111

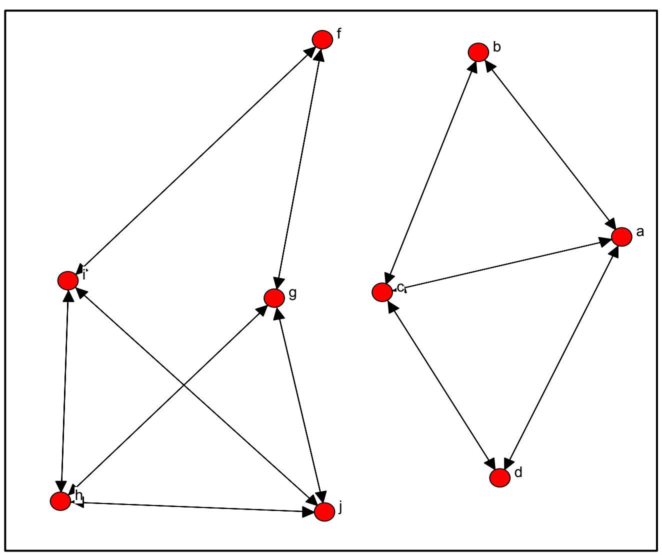


DISRUPTION

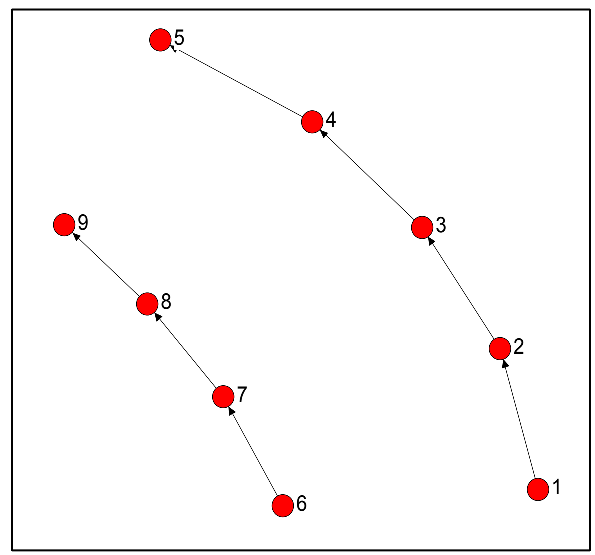
Features of distance-weighted fragmentation measure

- Yields higher value for two "stringy" components than for two well connected components

DF = 0.556



DF = 0.851



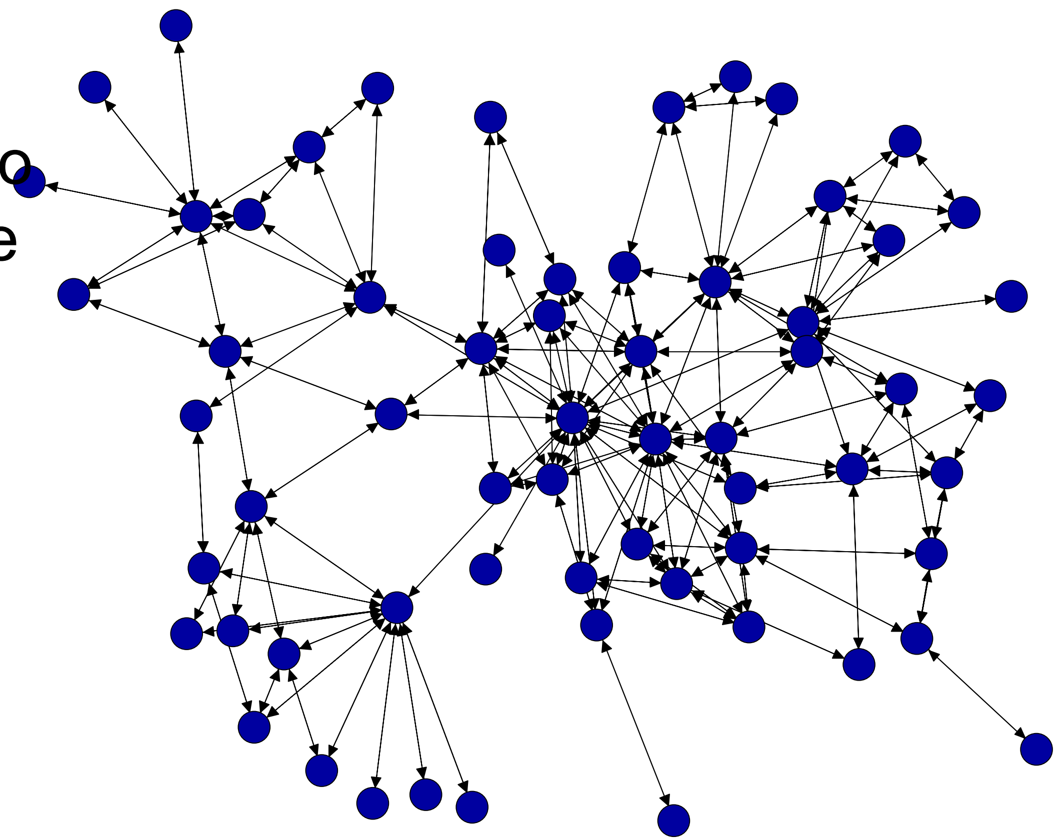
more fragmented

Optimization Algorithm

- Several appropriate choices
  - Genetic algorithm, tabu search, simulated annealing, Kernighan-Lin & variants
- Simple greedy algorithm works well
  - Exhaustive search of neighborhood of current solution (swap each $s \in S$ with every $t \in G-S$)
  - If no swap improves current solution,
    - then quit
    - Else accept best swap
  - Repeat

Empirical Example #1
Disrupt Terrorist Network

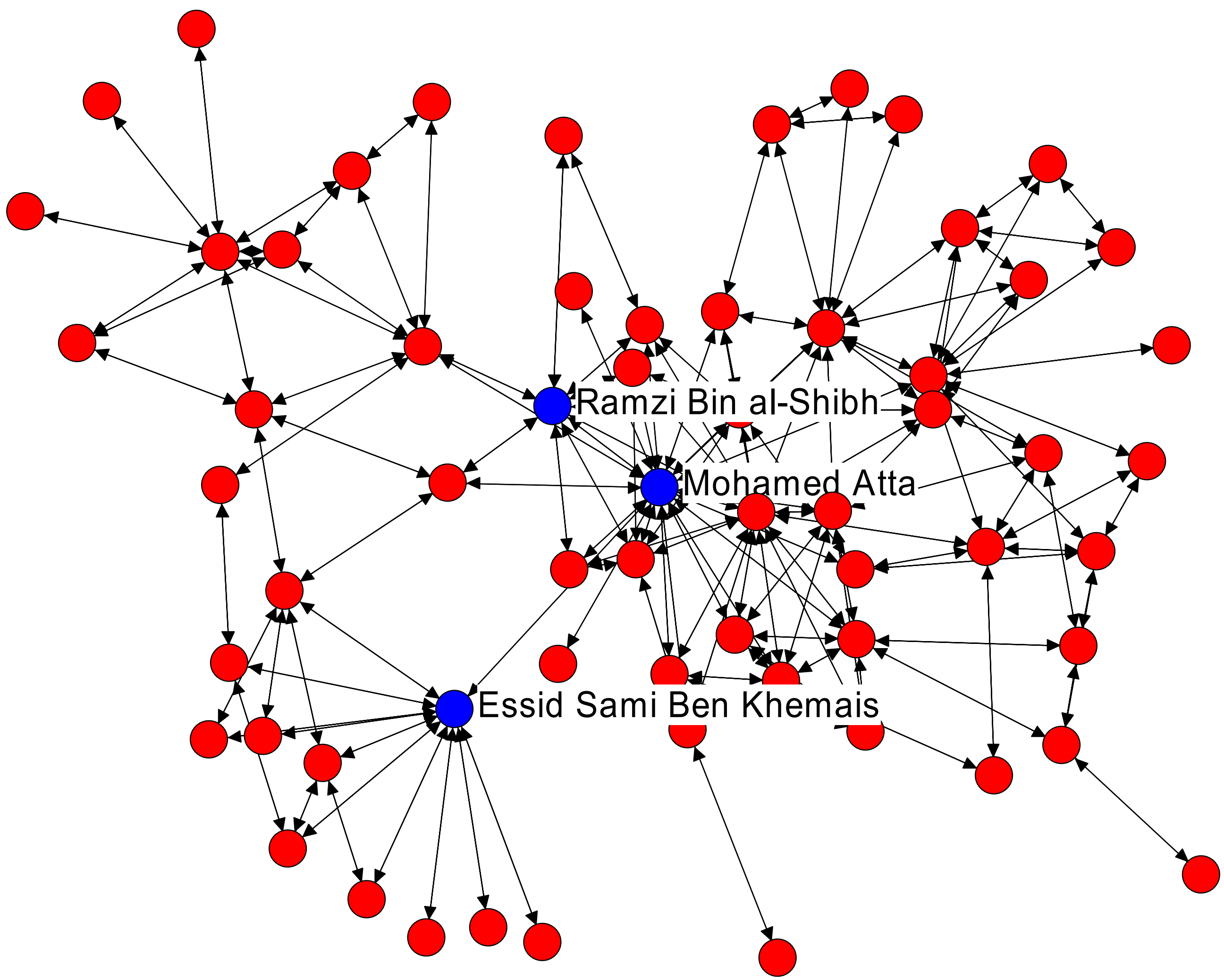
- Which three nodes should
  be isolated in order to maximally disrupt the network?



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

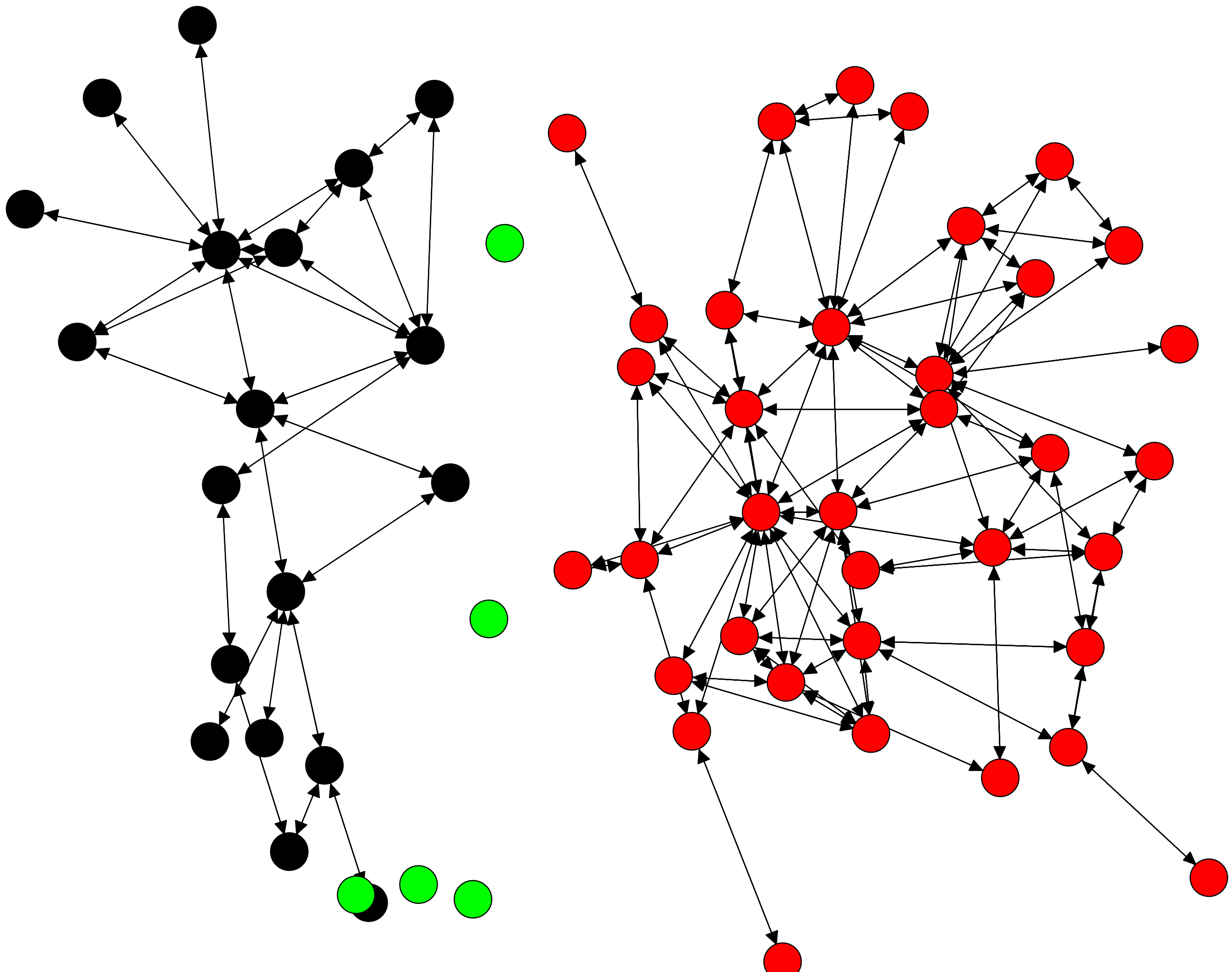
DISRUPTION

KeyPlayer Solution



DISRUPTION

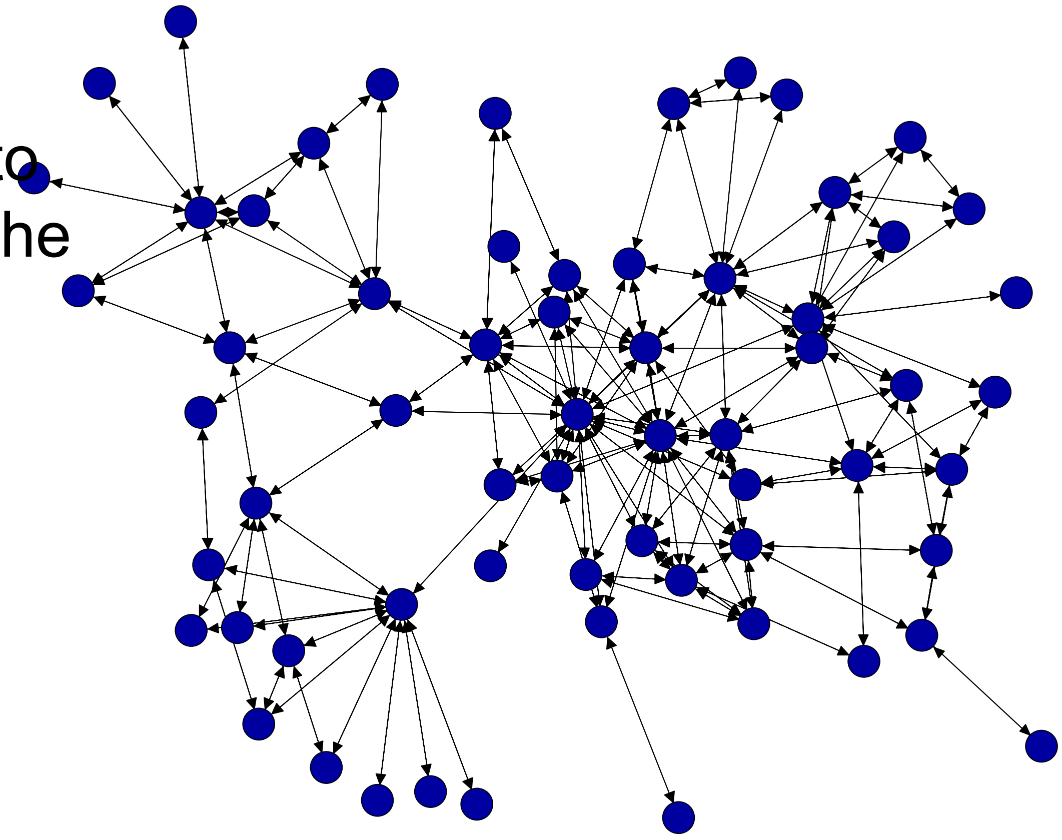
KeyPlayer Solution (key players removed)



INFLUENCE

Empirical Example #2 Influence Terrorist Network

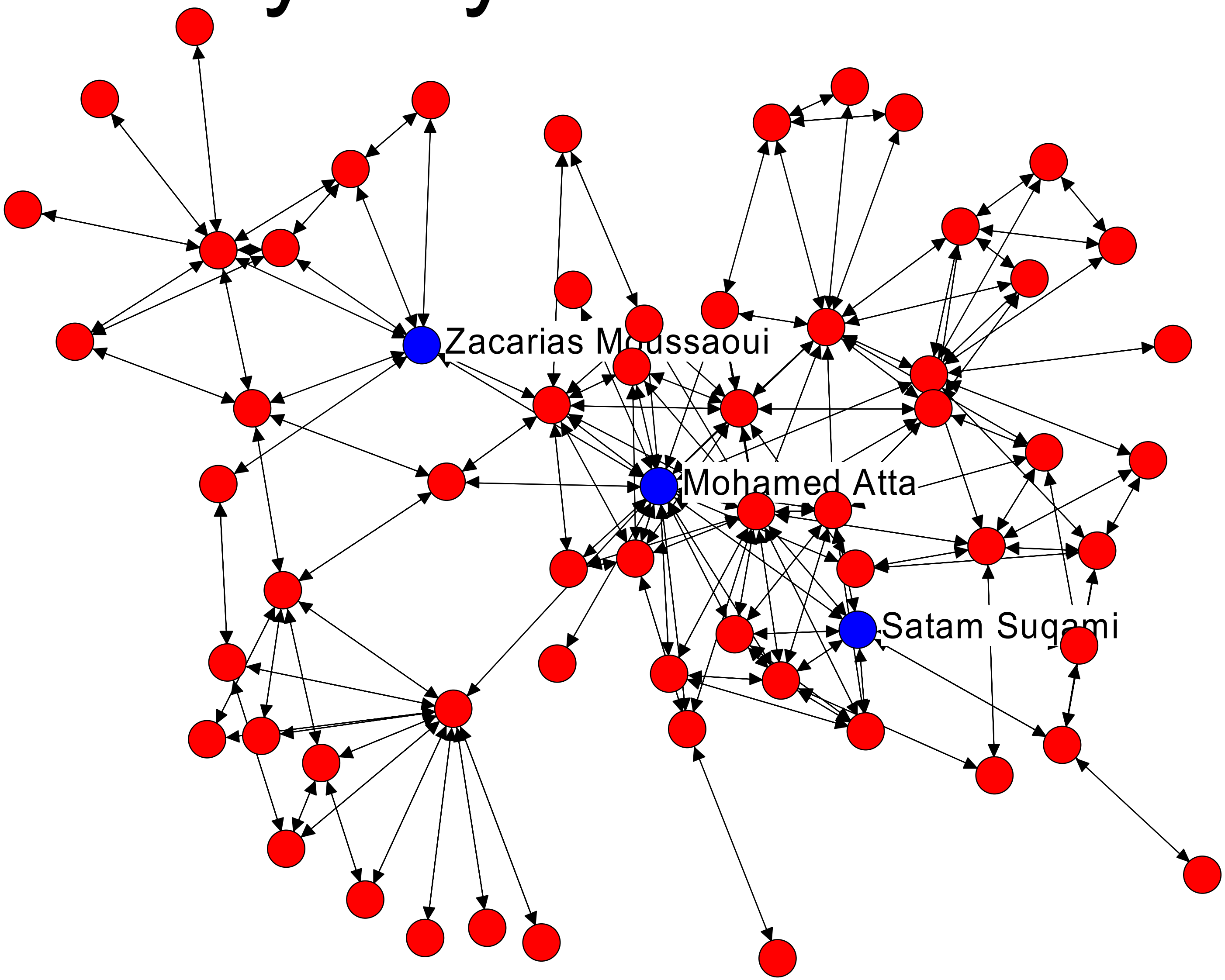
- Which three nodes should
  be selected in order to maximally influence the network by turning / planting information, etc.?



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

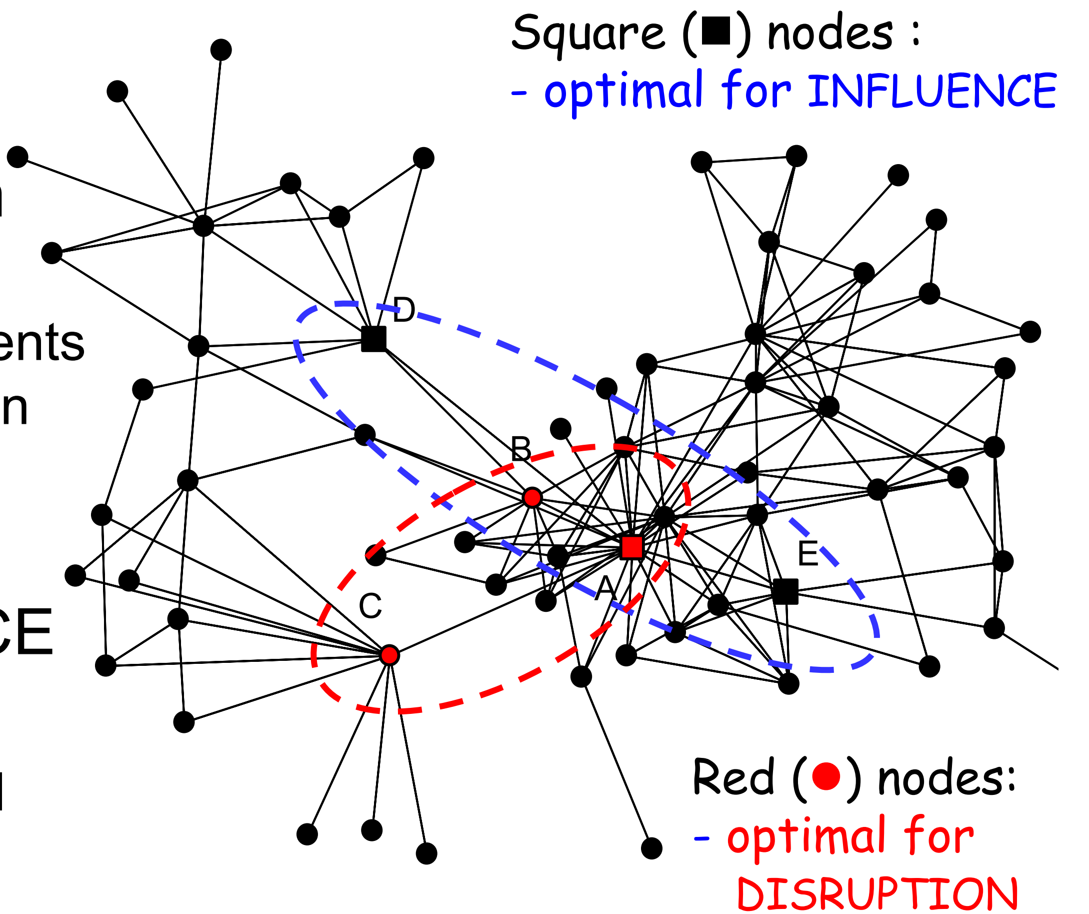
INFLUENCE

KeyPlayer Solution



Terrorist Network

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



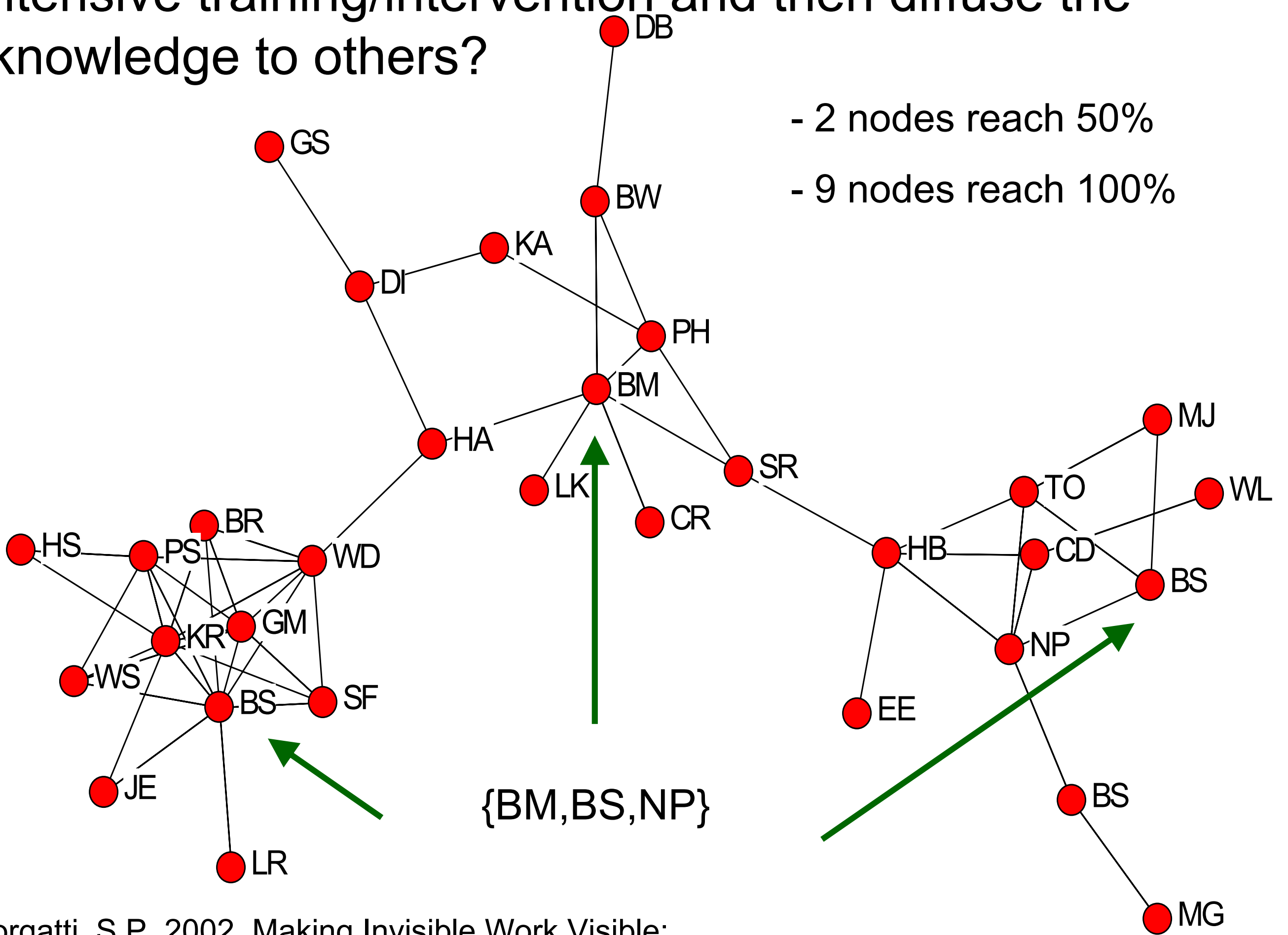
Data from: Krebs, V. 2002. Uncloaking terrorist networks.
First Monday 7(4): April. http://www.firstmonday.dk/issues/issue7_4/krebs/index.html

Empirical Example #2
Global Consulting Firm

- A major change initiative is planned. Which small set of actors can be selected for intensive training/intervention and then diffuse the new attitudes & 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}

- 2 nodes reach 50%
- 9 nodes reach 100%

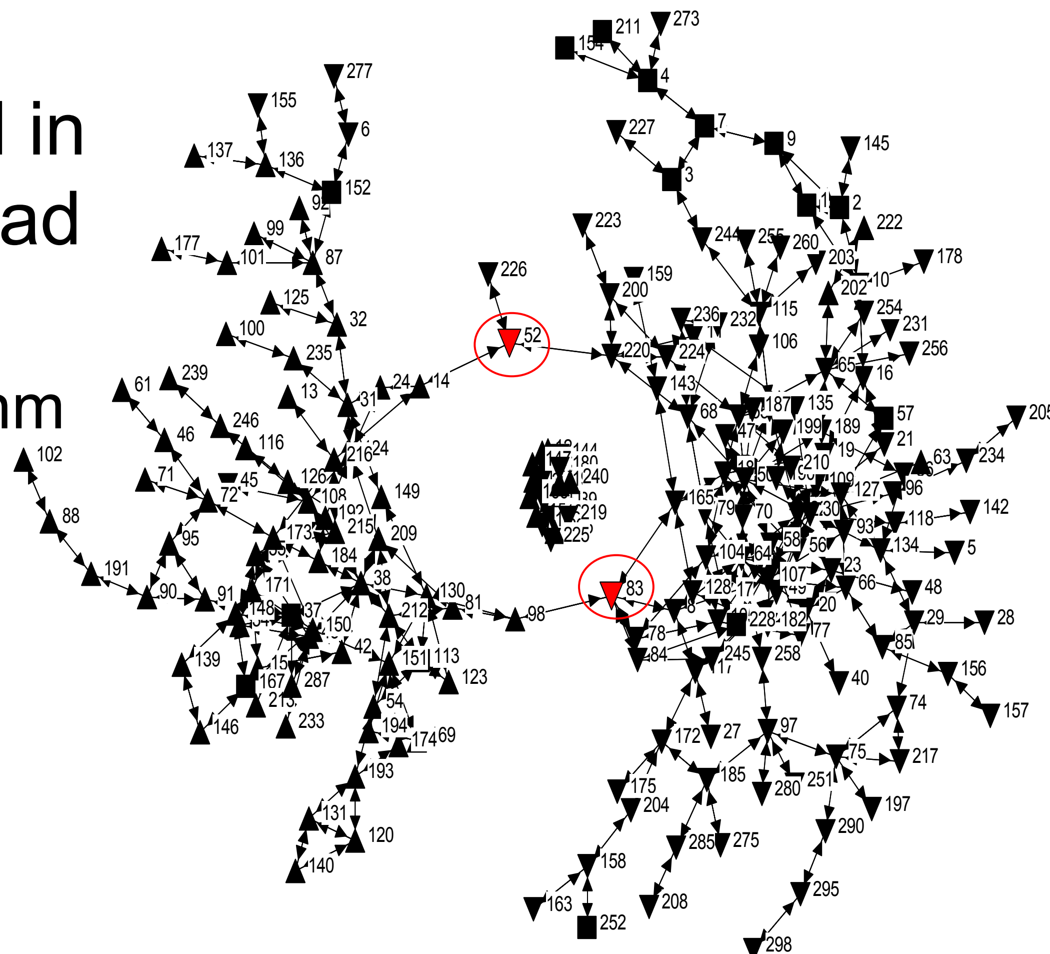


Data from: Cross, R., Parker, A., & Borgatti, S.P. 2002. Making Invisible Work Visible: Using Social Network Analysis to Support Strategic Collaboration. *California Management Review*. 44(2): 25-46

DISRUPTION

Example #3: Relations among drug injectors

- Which two people should be isolated in order to slow spread of HIV?
  - KeyPlayer algorithm identifies the two red nodes



- ■ whites
- ▲ african-american
- ▼ puerto-rican

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