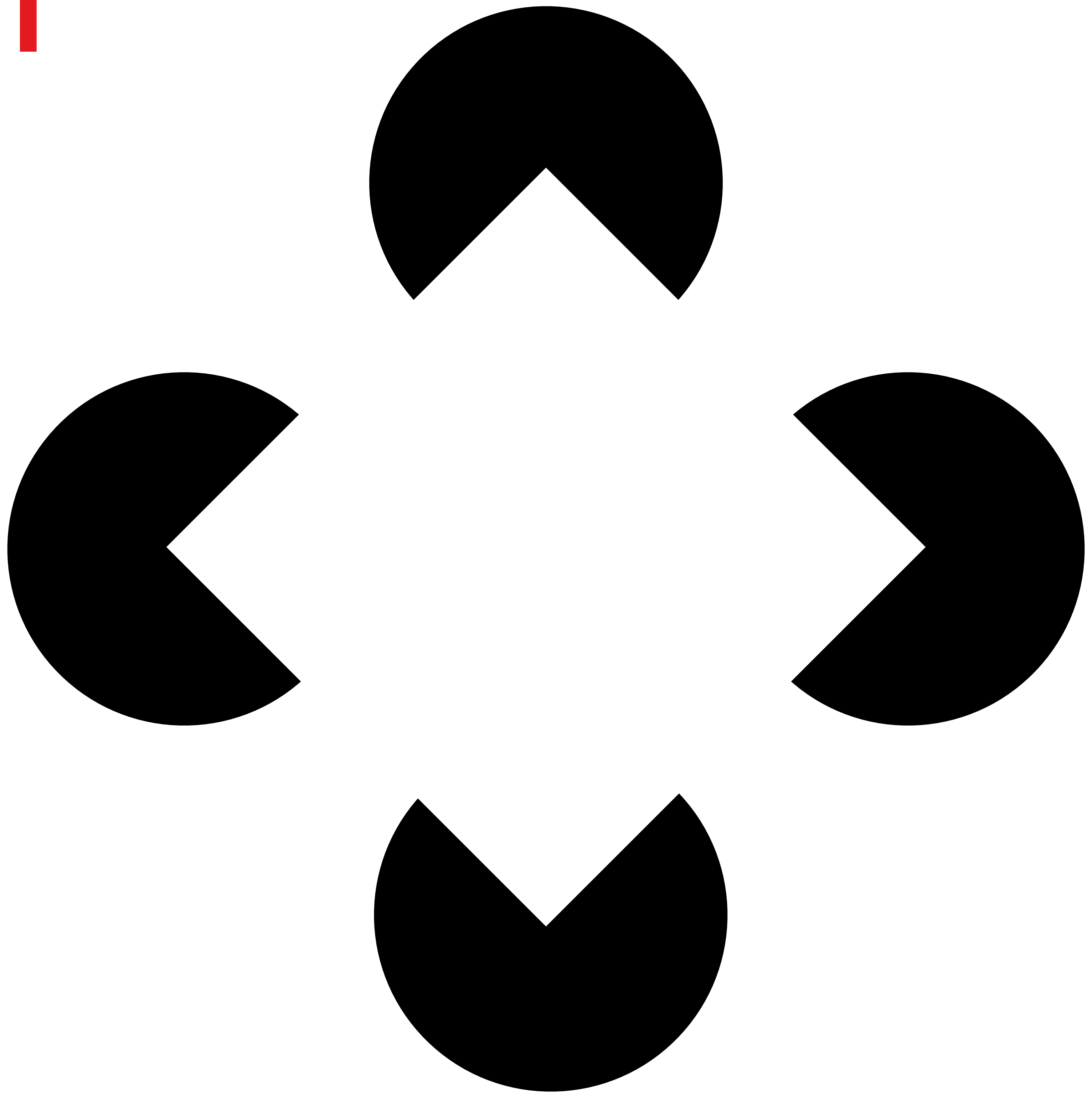
Community, or how to stand together



James
Hollway

GENEVA
GRADUATE
INSTITUTE

INSTITUT DE HAUTES
ÉTUDES INTERNATIONALES
ET DU DÉVELOPPEMENT

GRADUATE INSTITUTE
OF INTERNATIONAL AND
DEVELOPMENT STUDIES

Basic psychological needs

Effectance

- To learn new things, be autonomous, recognised
  - Centrality
  - Brokerage

Safety

- To fulfil emotional needs, trust, reputation
  - Cohesion and community
  - Embeddedness

Kadushin 2011

Generalised cohesion

- Density
  - Basically # ties over # possible ties
    - Ranges between 0 and 1
    - Tends to be low in social networks
    - Tends to be lower in large networks

$$D_u = \frac{2|E|}{|V|(|V|-1)}$$

$$D_d = \frac{|E|}{|V|(|V|-1)}$$

$$D_{2M} = \frac{|E|}{|V||W|}$$

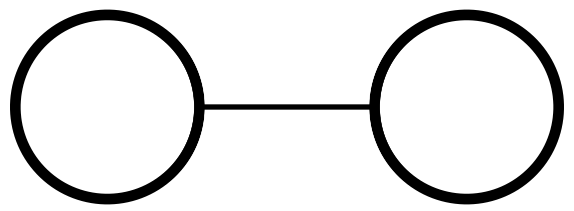
Localised cohesion

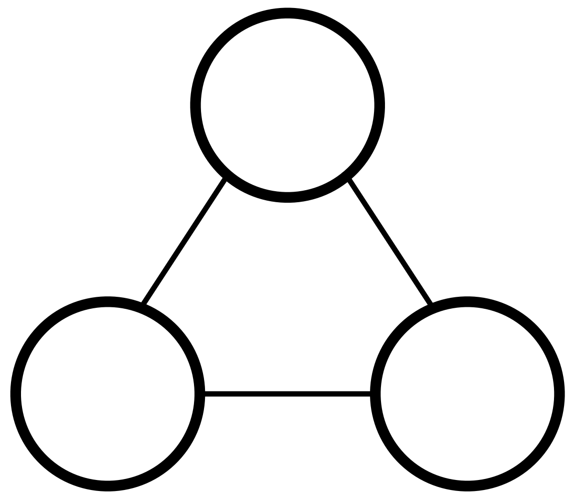
- Reciprocity
  - Main method of calculation: $$R = \sum_{ij} (AA')_{ij}$$
  - Ranges between 0 and 1
    - Asymmetric ties thought to be unstable (or hierarchical), so equilibrium of null or reciprocated dyads
    - Tends to be high in *social* networks
    - Multiplex reciprocity suggests *exchange*

Georg Simmel

- Born 1 March 1858, Berlin, Kingdom of Prussia
  - Bourgeois elite
  - Academic outsider
- Died 28 September 1918, Strassburg, German Empire
- "Quantitative Aspects of the Group"

Dyads and Triads





Dyad = 2 person *clique* (completely connected subgroup)

- Safety: easier to exert control through application of power asymmetries

- Effectance: no other person to shift balance of the group, therefore each able to retain their individuality

Triad = 3 person *clique* (completely connected subgroup)

- Safety: possibility of a dyad forming within the triad, threatening exclusion of individual unless they subordinate to group norms

- Effectance: harder to exert individual control, feels more impersonal and distant, anomie

“”The social structure [of the dyad] rests immediately on the one and on the other of the two, and the secession of either would destroy the whole…. As soon, however, as there is a sociation of three, a group continues to exist even in case one of the members drops out.”

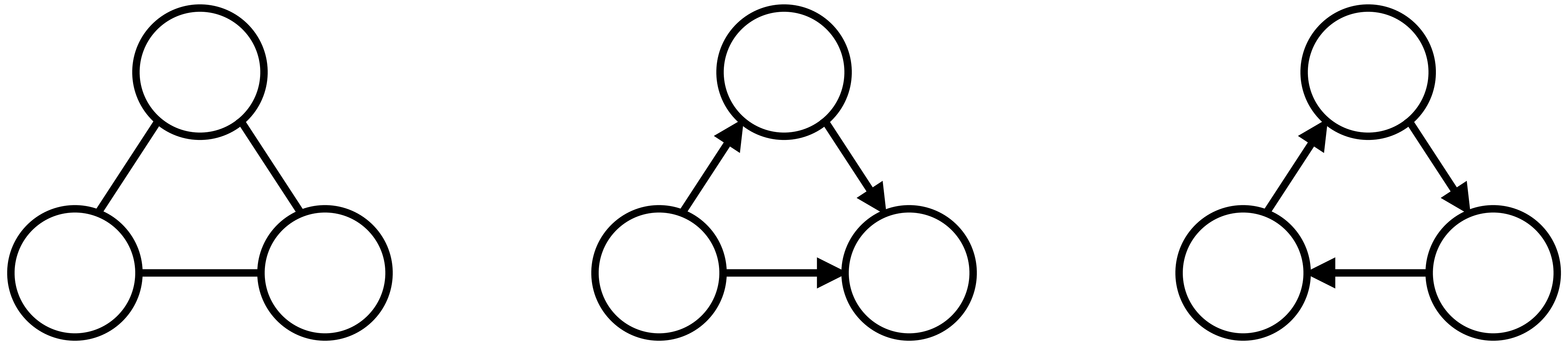
–Simmel ([1908] 1950:123)

Of course, Hollywood reads social networks.

That's why you see so many love triangles.

Cos then strategies like competition, alliances, mediation, etc come into play…

Triangles, Transitivity, and Cycles



- Triangles
  - When A–B, and B–C, then A–C: *"connected to connections' connections"*
  - Various causal arguments, including opportunity for introduction…
  - Social networks typically between 0.3 and 0.6

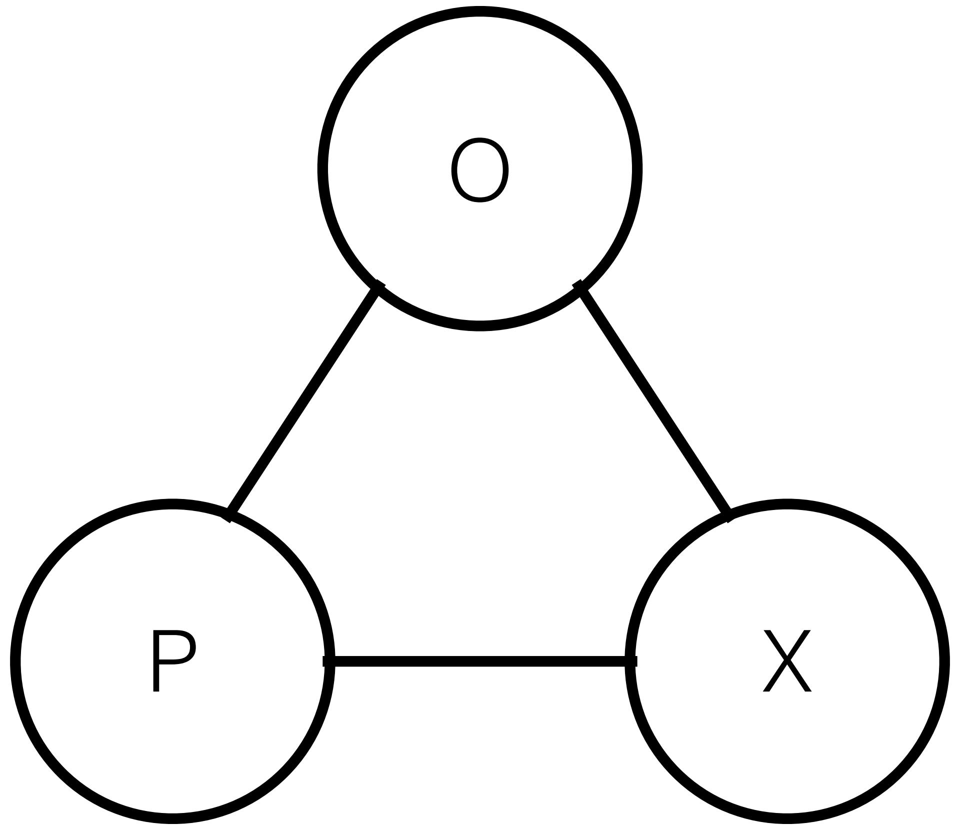
- Transitivity
  - When A→B, and B→C, then A→C: *"friend of my friend is my friend"*
  - Index from 0 to 1, where 1 is a fully transitive graph
  - Supposedly indicative of hierarchy

- Cycles
  - When A→B, and B→C, then C→A: *"I am the friend of a friend of my friend"*
  - Generally argued to be demonstrative of a <u>lack</u> of hierarchy in the network, but see <u>Block 2015</u>…

Signed networks

- So far we have only explored positive relationships (e.g. friends, fans, followers)
- But negative relationships too (e.g. antagonism, controversy, disagreement, conflict)
- Theories of structural balance help us understand the relationship between the two are key in creating cohesion and community

Heider's balance theory

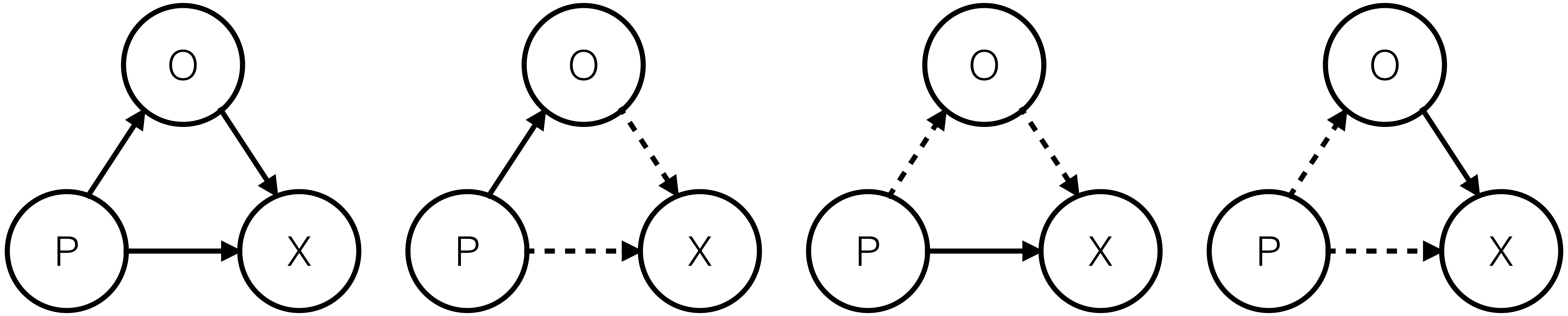


- Motivational theory of attitude change based on cognitive consistency
- Based on cliques (everyone knows everyone else)
- Actor nodes labelled $\{P, O, Q\}$, object node labelled X
- Each edge labelled with + or -, solid or dashed, green/blue or red

What folk proverbs relating to friends and enemies have you heard?

Structural (im)balance

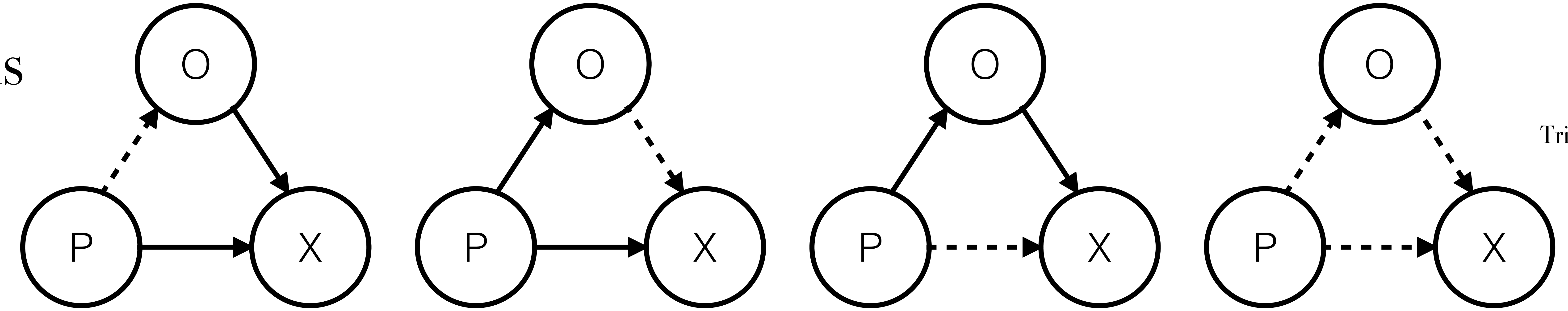
Balanced triads



Triad balanced if product of ties positive
e.g. +++, + - -

I agree with my friends and disagree with my enemies

Imbalanced triads



Triad imbalanced if product of ties negative
e.g. ++-, - - -

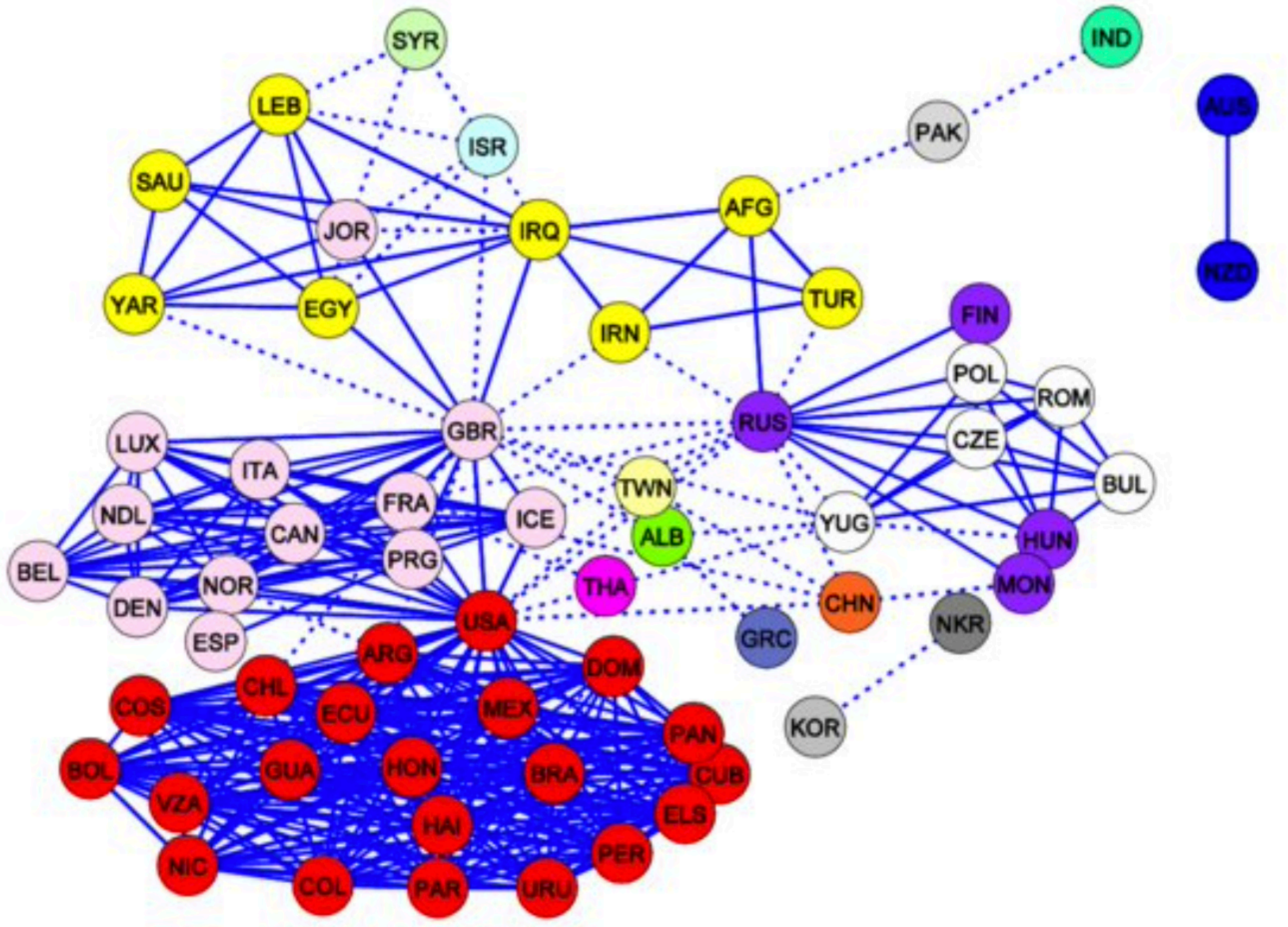
I agree with my enemies and disagree with my friends…

Heider

- Imbalanced triads unstable because they induce “stress” or cognitive dissonance in actors
- Mechanism assumes people prefer (cognitive) consistency and acts to reduce tension move imbalanced states toward balance

Harary and Cartwright

- Generalises balance theory to signed graphs of any size
    - all triads have balanced property
- Theorem: a complete signed graph is balanced if and only if
    - the nodes can be partitioned into two sets so that
    - all ties within sets are positive *and*
    - all ties between sets are negative

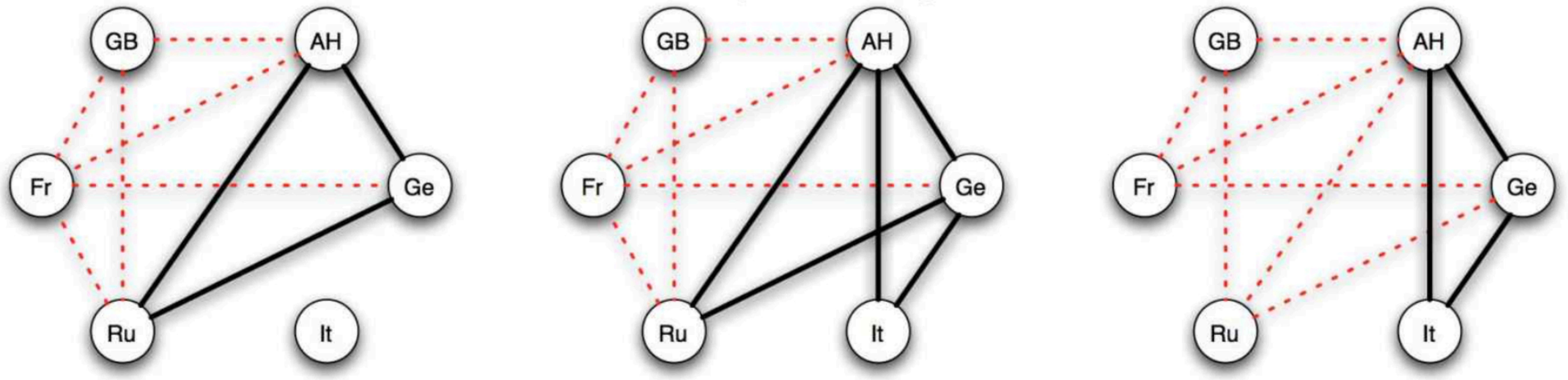


Applications in IR

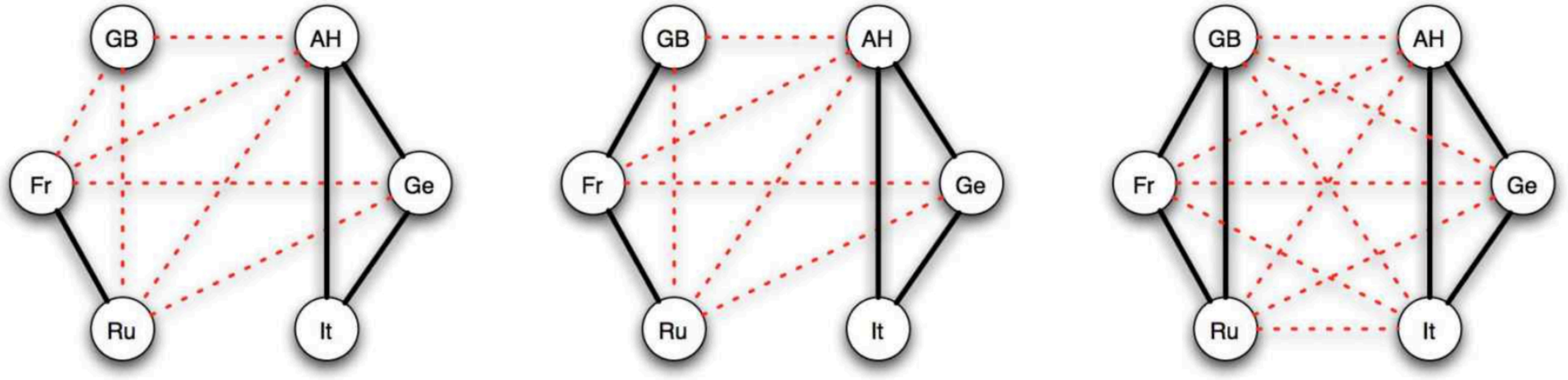
- *Which graphs are balanced?*
- *How many tie changes to balance?*
- *Is structural balance always a good thing?*

Predicting WWI

[Easley and Kleinberg]



(a) *Three Emperors' League 1872–81*　(b) *Triple Alliance 1882*　(c) *German-Russian Lapse 1890*

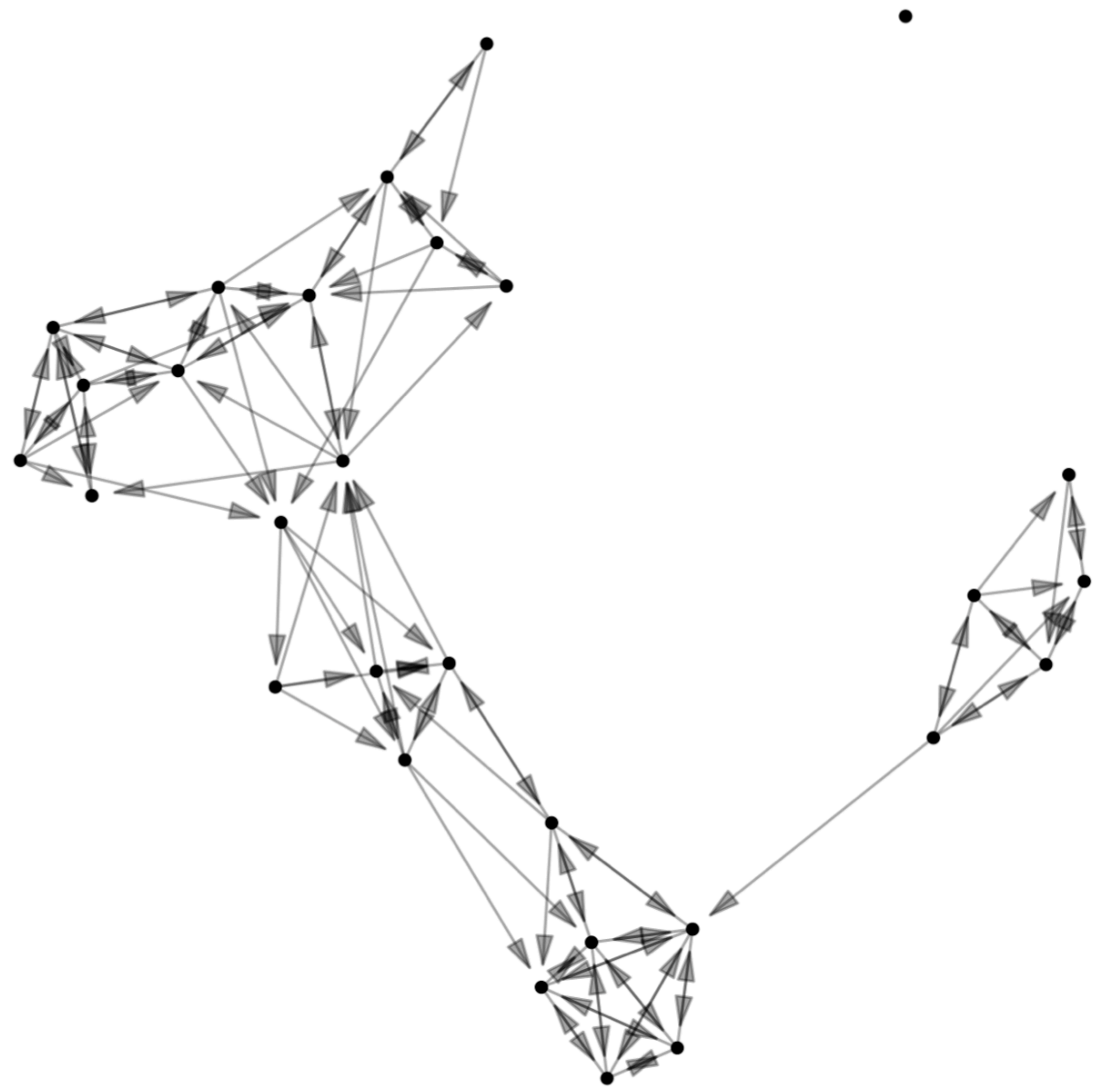


(d) *French-Russian Alliance 1891–94*　(e) *Entente Cordiale 1904*　(f) *British Russian Alliance 1907*

Concluding comments

- Structural balance theory suggests an important micro-macro link between actor-level processes and group structure
- Structural balance theory is dynamic because it predicts change over time, *but* either teleological or balance not an equilibrium
- Empirical research shows balance messy…
  - ++- and - - - not uncommon…
- Plenty of signed data available, e.g. slashdot and epinions and gossip!

How many ‘groups’ in this network?



- components
  - 2 (weak) components
  - 3 (strong) components
- more ‘clusters’ here?

Two strategies

- Faction analysis
  - Deductive: predetermine number of subgroups
  - Algorithm tries to maximise density within subgroups by permuting observed matrix
  - Compare to perfectly maximised ideal type

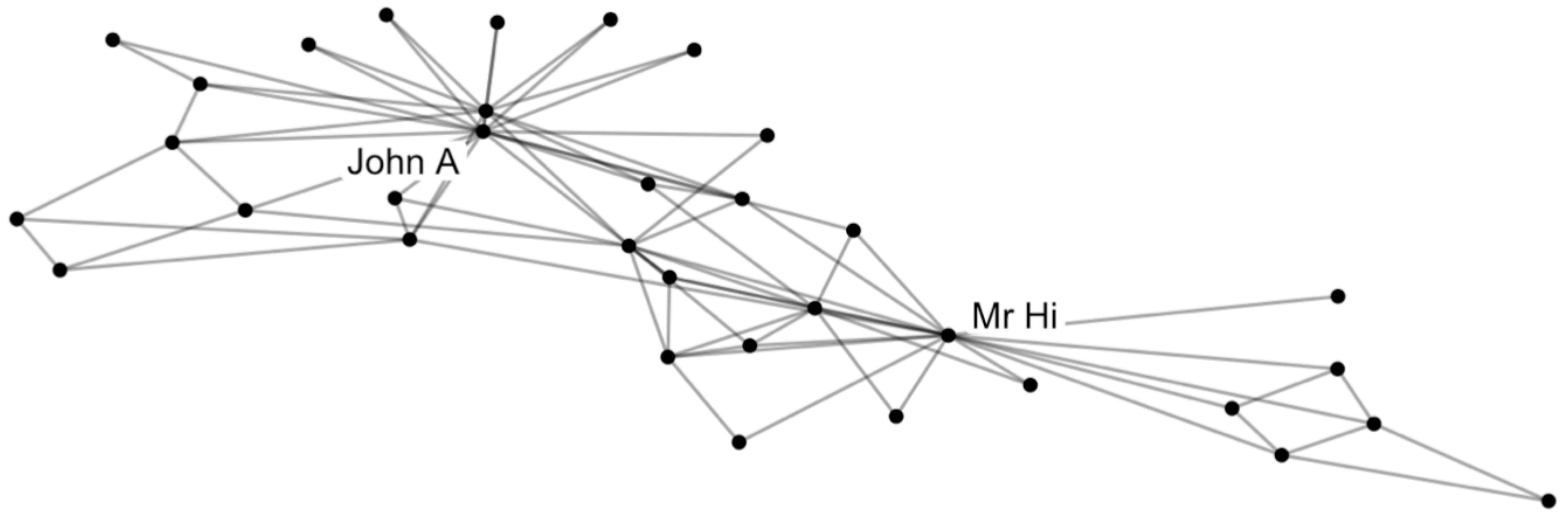
- Community detection
  - Inductive: identify an appropriate number and membership of groups from observed network
  - Try to maximise “modularity” criterion
  - By undertaking a certain method/ algorithm…

What are communities?

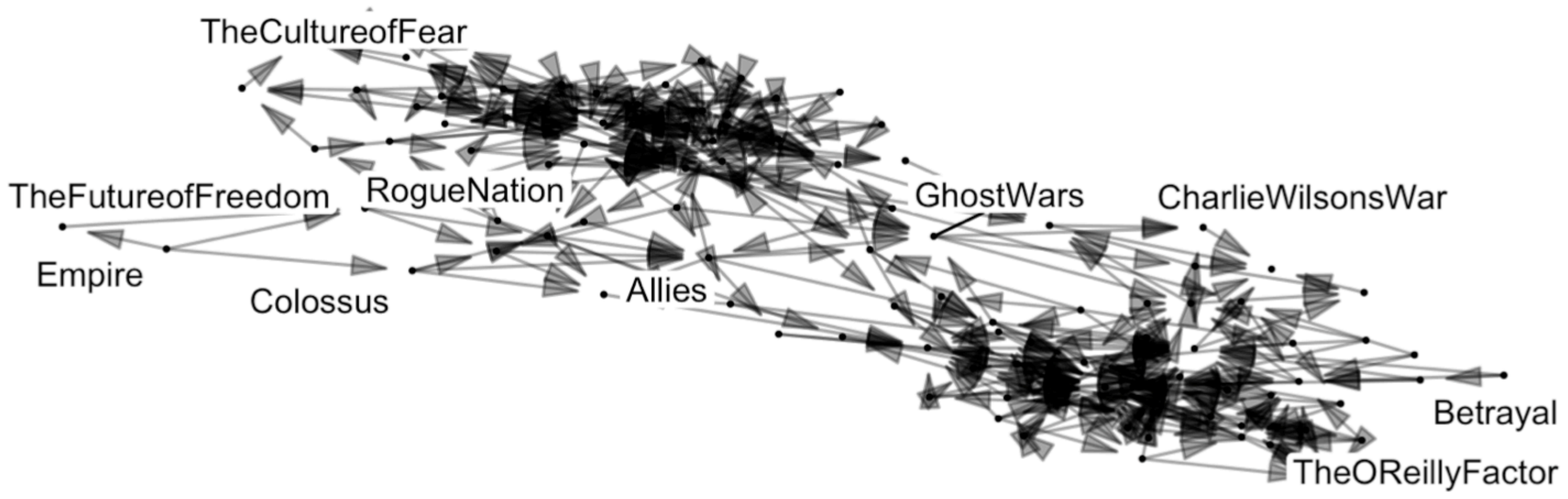
- Communities are densely connected (+ within-group) and well separated (- between-group)
- Includes info about internal structure (like cliques) and external structure (like components)
- They define and affect groups, networks, and individuals

Examples

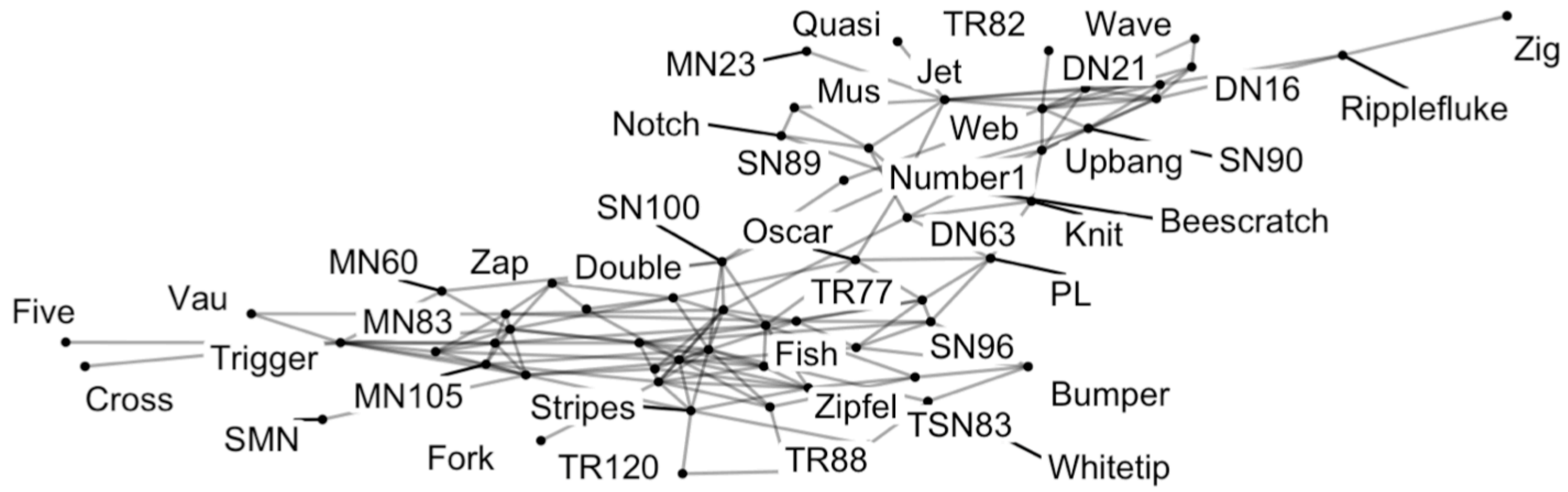
Karateka



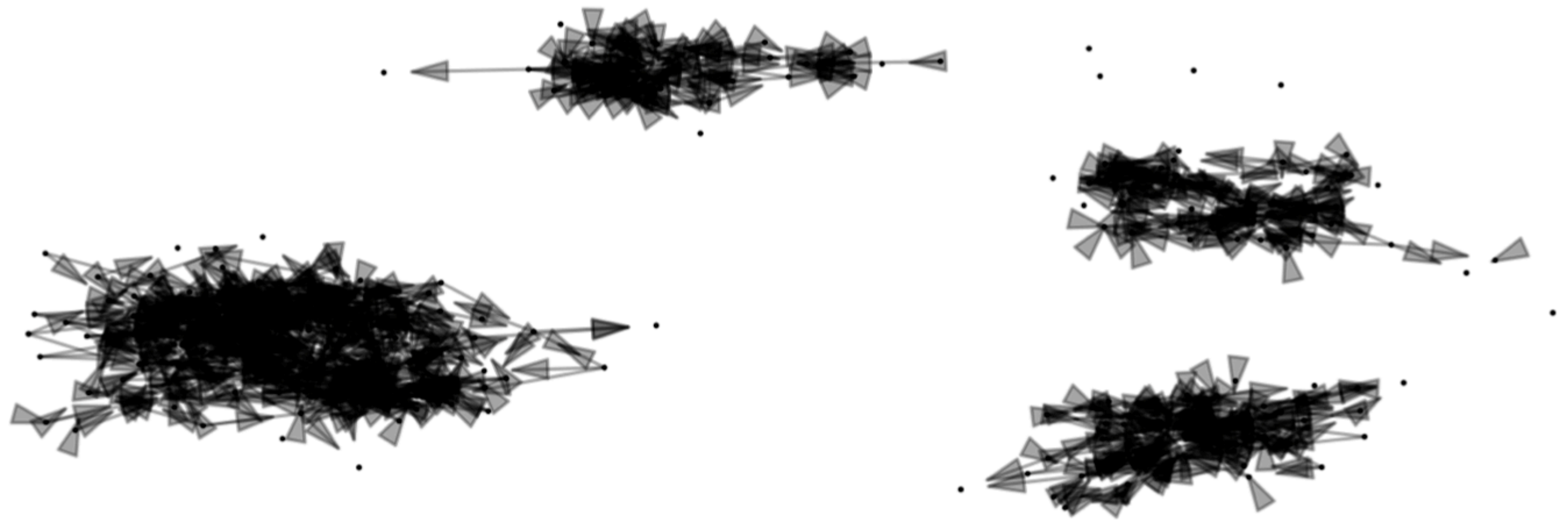
Political Books



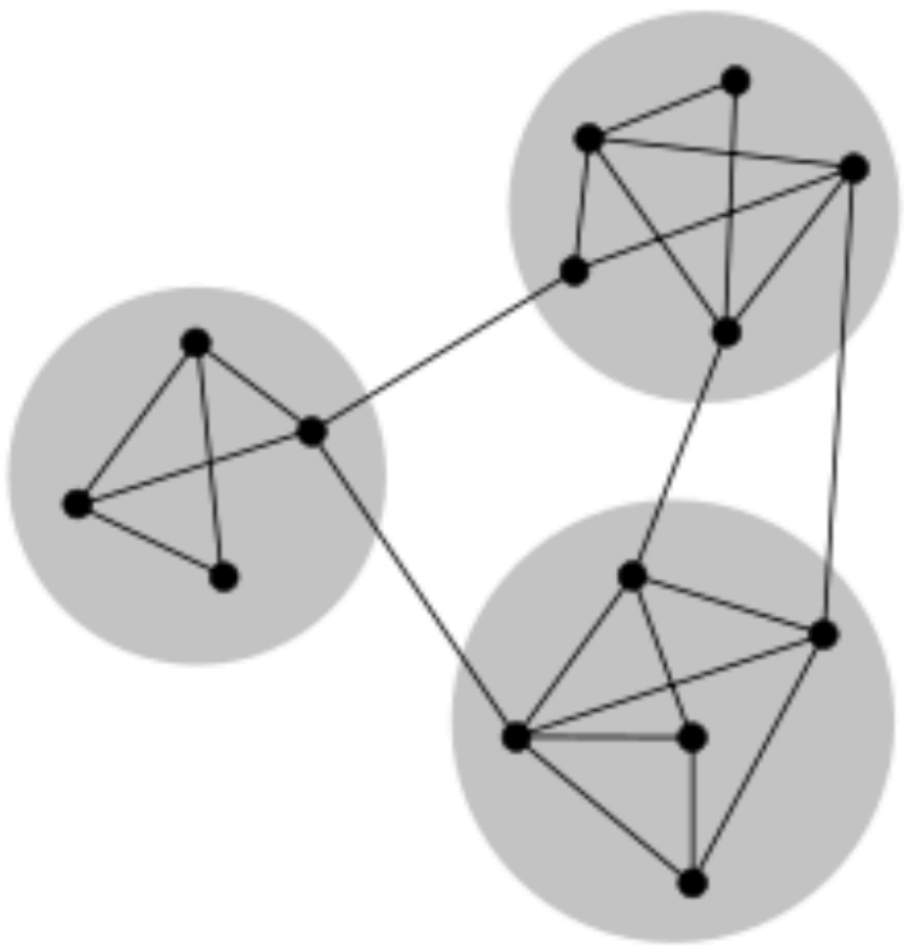
Dolphins



Physicians



Modularity



$$Q = \sum_{c=1}^{k} (e_{cc} - a_c^2)$$

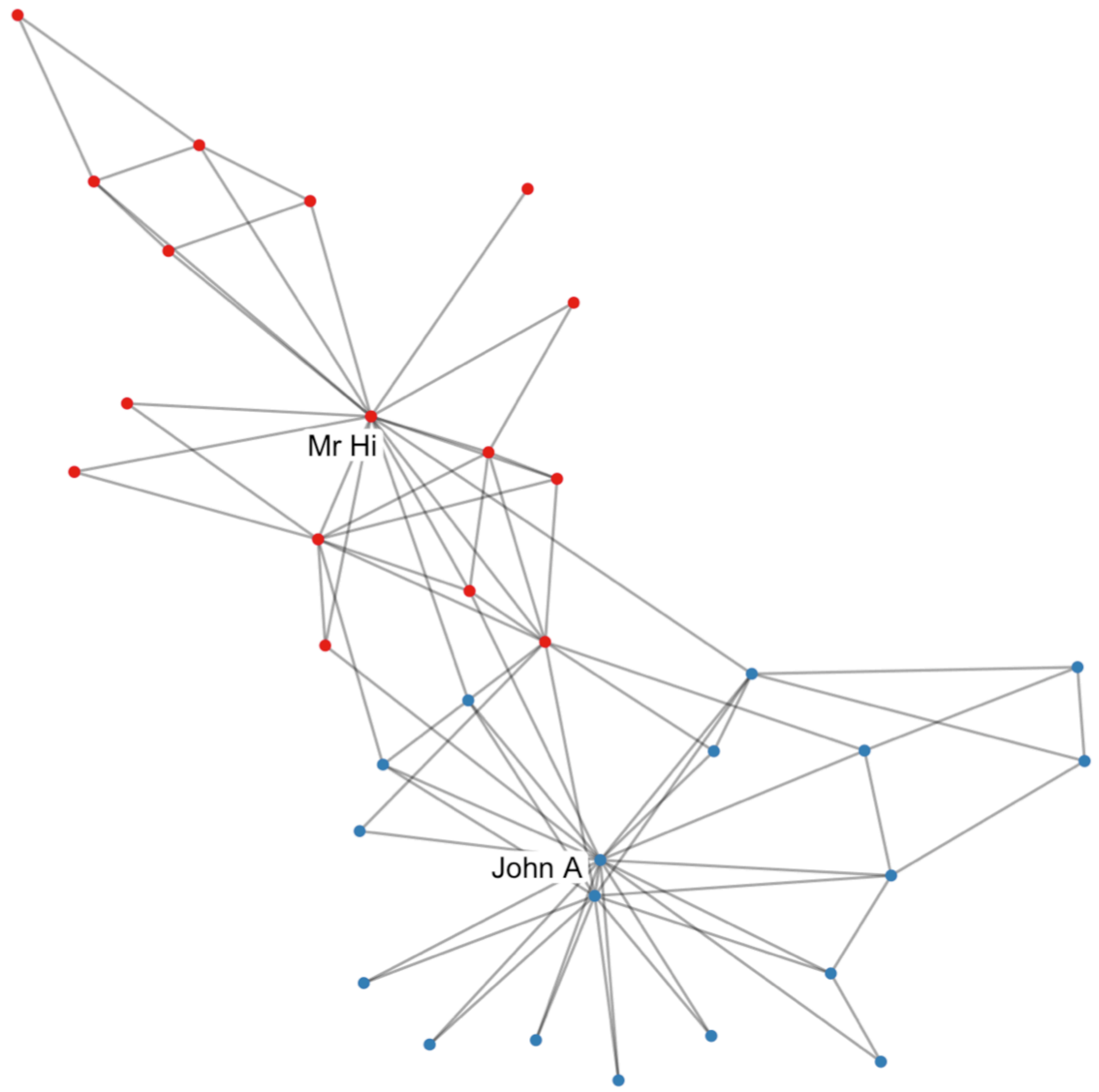
- Simplified expression (see Fortunato 2010: 89) for the modularity criterion Q, where:
  - e_{cc} is % edges within a cluster c
  - a_c^2 is the expected % of ties within the cluster in a random graph with the same degrees

- Always < 1 but can be negative
  - 0 when whole graph is taken as a single community, i.e. edges randomly distributed
  - Negative modularity means less community structure than random, e.g. multipartite structure

- Modularity grows larger if better separated groups or graph gets larger, so do not use to compare graphs of different sizes...

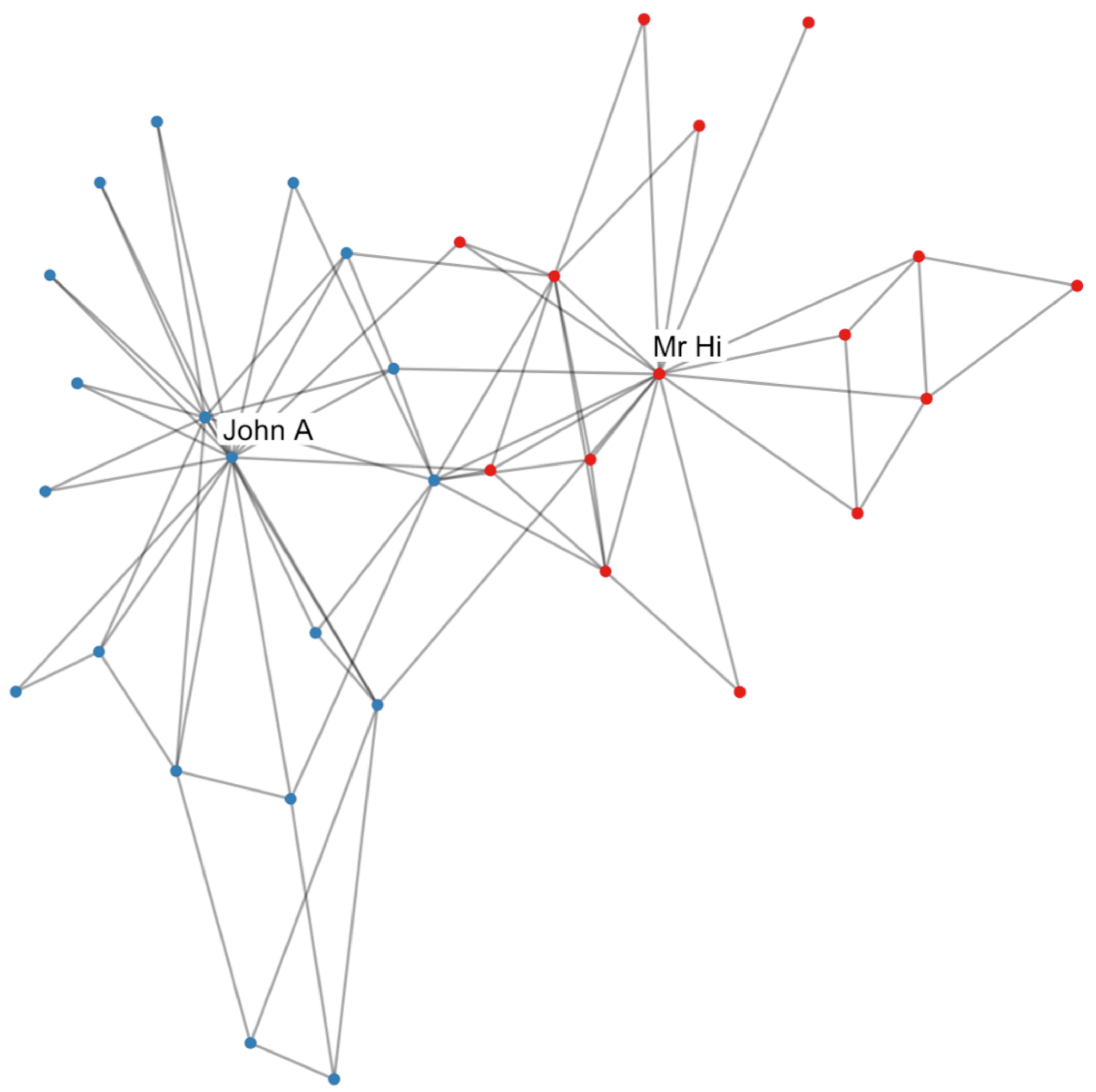
- Partitioning graphs into c components NP-hard (computationally difficult) problem
  - Finding optimal solution not always possible
  - Need an algorithm (a set of rules) by which we can identify community memberships until we find the allocation that maximises Q

Data and algorithms



- Random
  - Label assignment e.g. Kernighan-Lin
  - Walks e.g. Walktrap
- Divisive
  - Hierarchical decomposition e.g. Edge-Betweenness
- Agglomerative
  - Hierarchical composition e.g. Fast-Greedy

Kernighan-Lin algorithm



Intuition: swap group label assignment

1. Start: full network with random (usually two) group assignment
2. Identify: node-group-switch would increase Q the most
3. Modify: that node's group assignment
4. Repeat: steps 2-3
5. Stop: when no switch increases Q

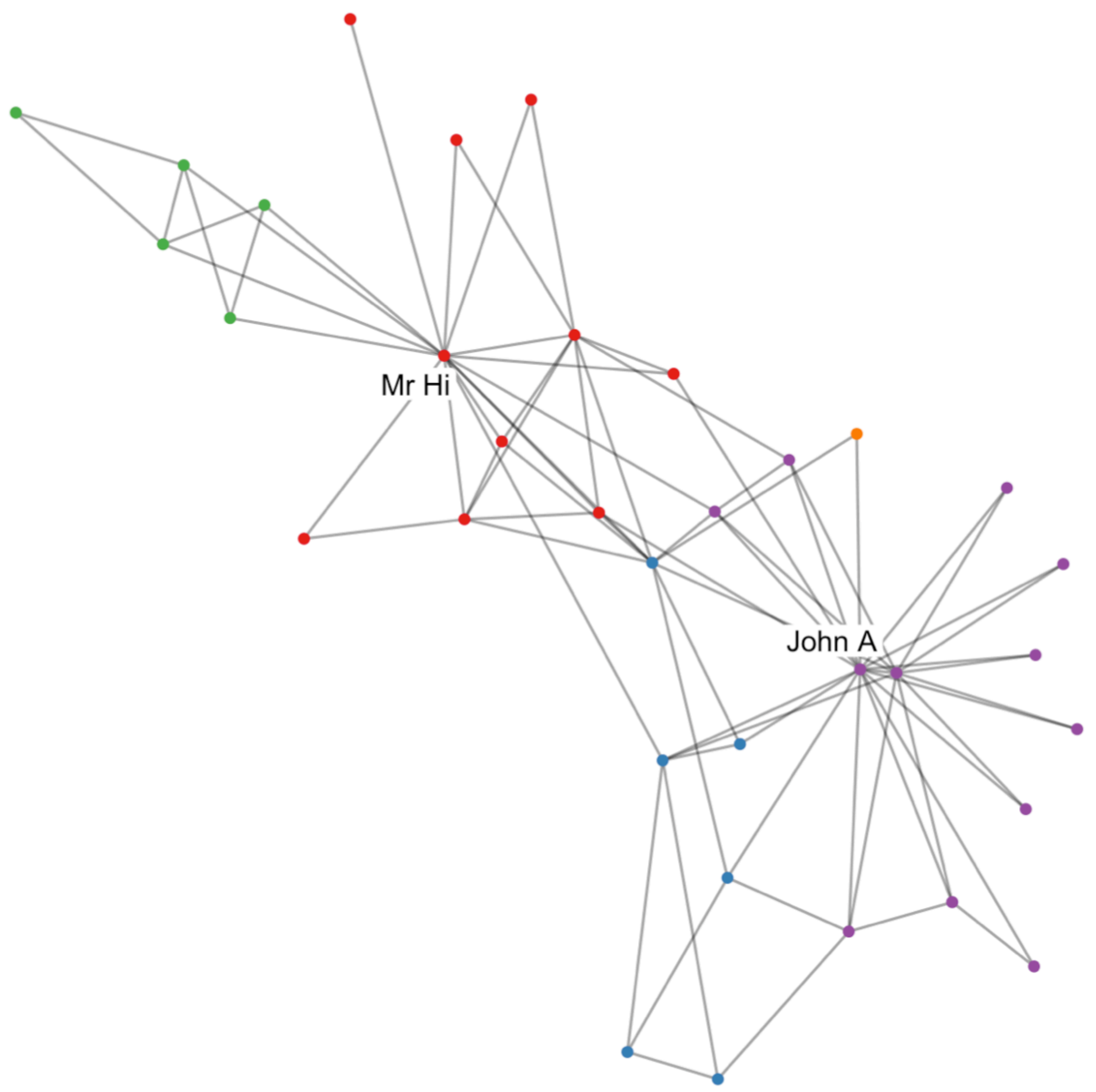
Pros

- Quick and easy
- Similar to faction analysis

Cons

- Depends on initial distribution
- Requires predetermined number of communities
- Does not always find the best solution

Edge-betweenness algorithm



Intuition: full start and divide

1. Start: with full network and a single community
2. Identify: edge with highest edge-betweenness (bridge)
3. Modify: delete that edge and identify any resulting components as a community
4. Repeat: steps 2-3
5. Stop: when no better Q is obtained by any edge deletion

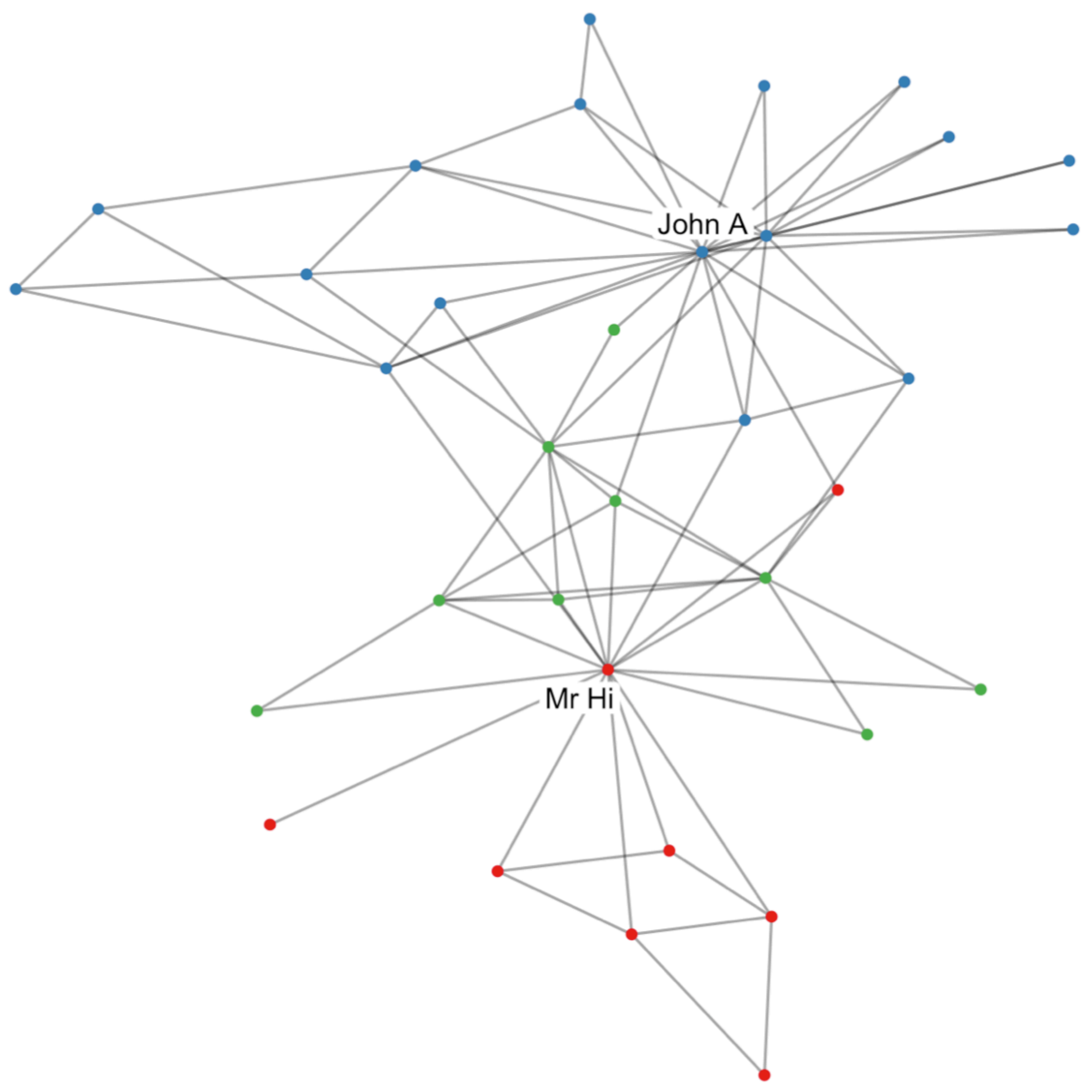
Pros

- Clear logic based on information propagation
- No random variation

Cons

- Does not necessarily maximise modularity
- Calculation of edge betweenness slow
- Does not always find the best solution

Fast-greedy algorithm



Intuition: empty start and agglomerate

1. Start: empty graph and each node its own community
2. Identify: evaluate which merger of two adjacent nodes would increase Q the most
3. Modify: assign same community to all merged nodes
4. Repeat: steps 2-3
5. Stop: when merging decreases Q

Pros

- Quick and possible to use on larger graphs
- Gives a dendrogram of the hierarchical clustering
- Most fun to watch at work

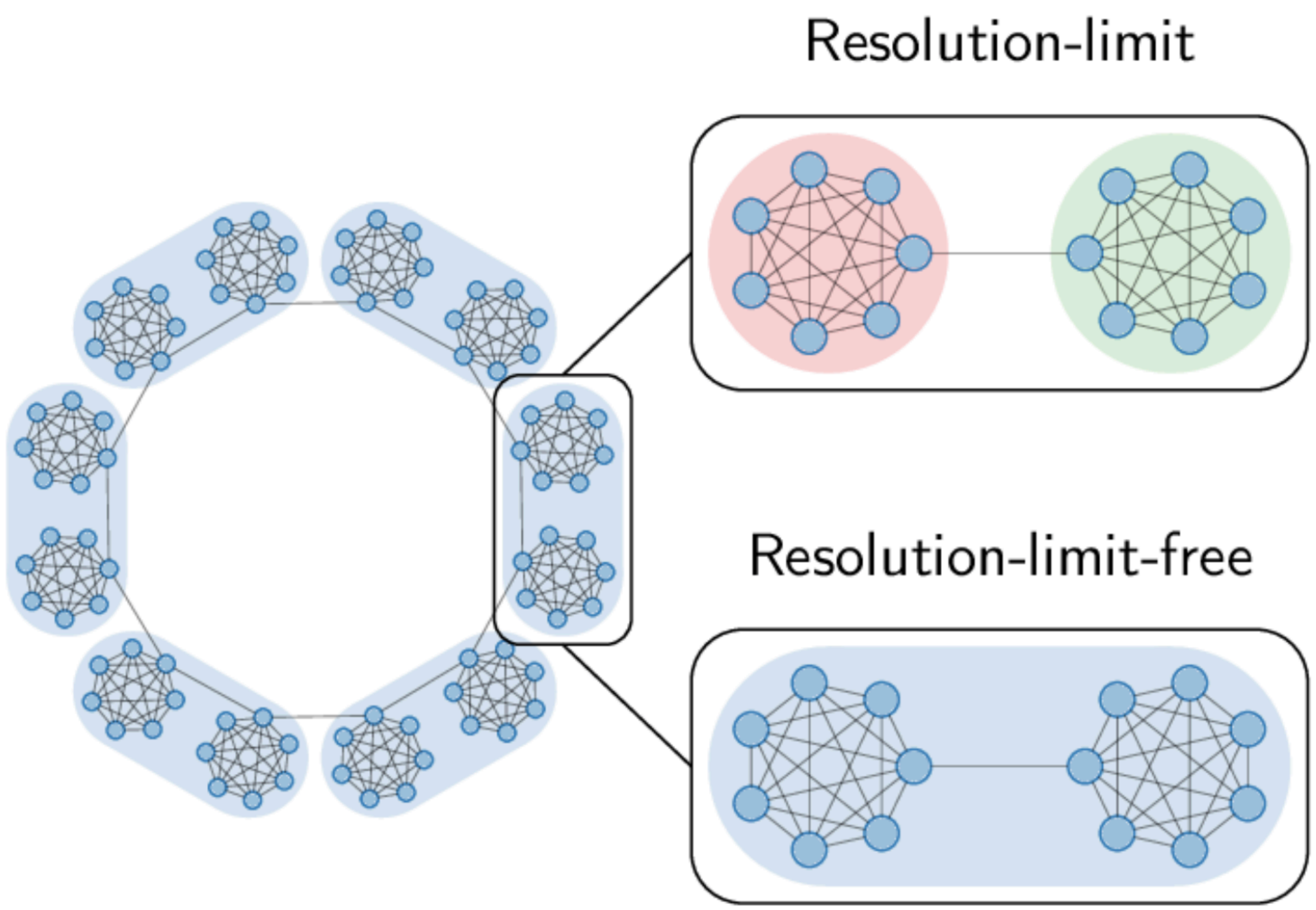
Cons

- Not good at detecting smaller communities, tends to quickly form larger communities

Community detection algorithms

Algorithm	Reference	Type	W	D	M	H
Label propagation	Raghavan et al. 2007	Assignment	X			
Spin glass	Reichhardt & Bornholdt 2006	Assignment	X		X	
Walk trap	Pons & Latapy 2005	Random walks	X			X
Info map	Rosvall & Bergstrom 2008	Random walks	X	X		X
Leading eigenvector	Newman 2006	Hier. Decomposition			X	
Edge betweenness	Newman & Girvan 2004	Hier. Decomposition	X	X	X	X
Fast greedy	Clauset et al. 2004	Hier. Composition	X		X	X
Multilevel	Blondel et al. 2008	Hier. Composition	X		X	
Optimal	Brandes et al. 2008	Problem solving	X	X	X	

Resolution limits



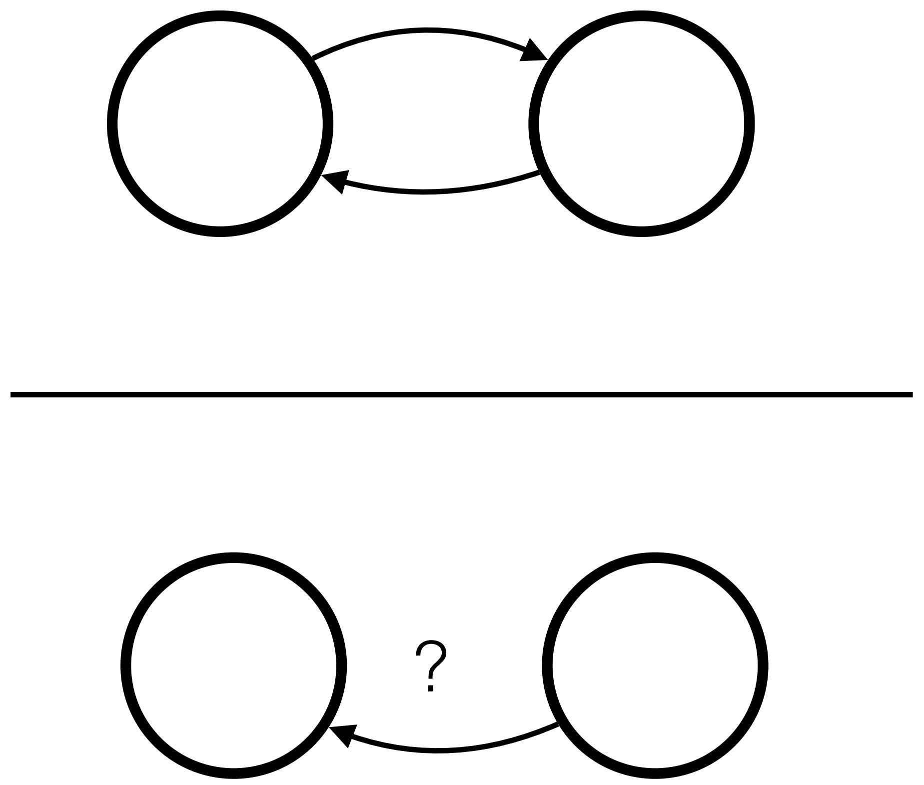
- Possibility of
  - breaking up large communities
  - missing smaller communities
- May depend on network size
  - Larger the network, more likely that modularity increased while ignoring smaller communities
  - Called a resolution limit and is a *problem*

Community detection summary

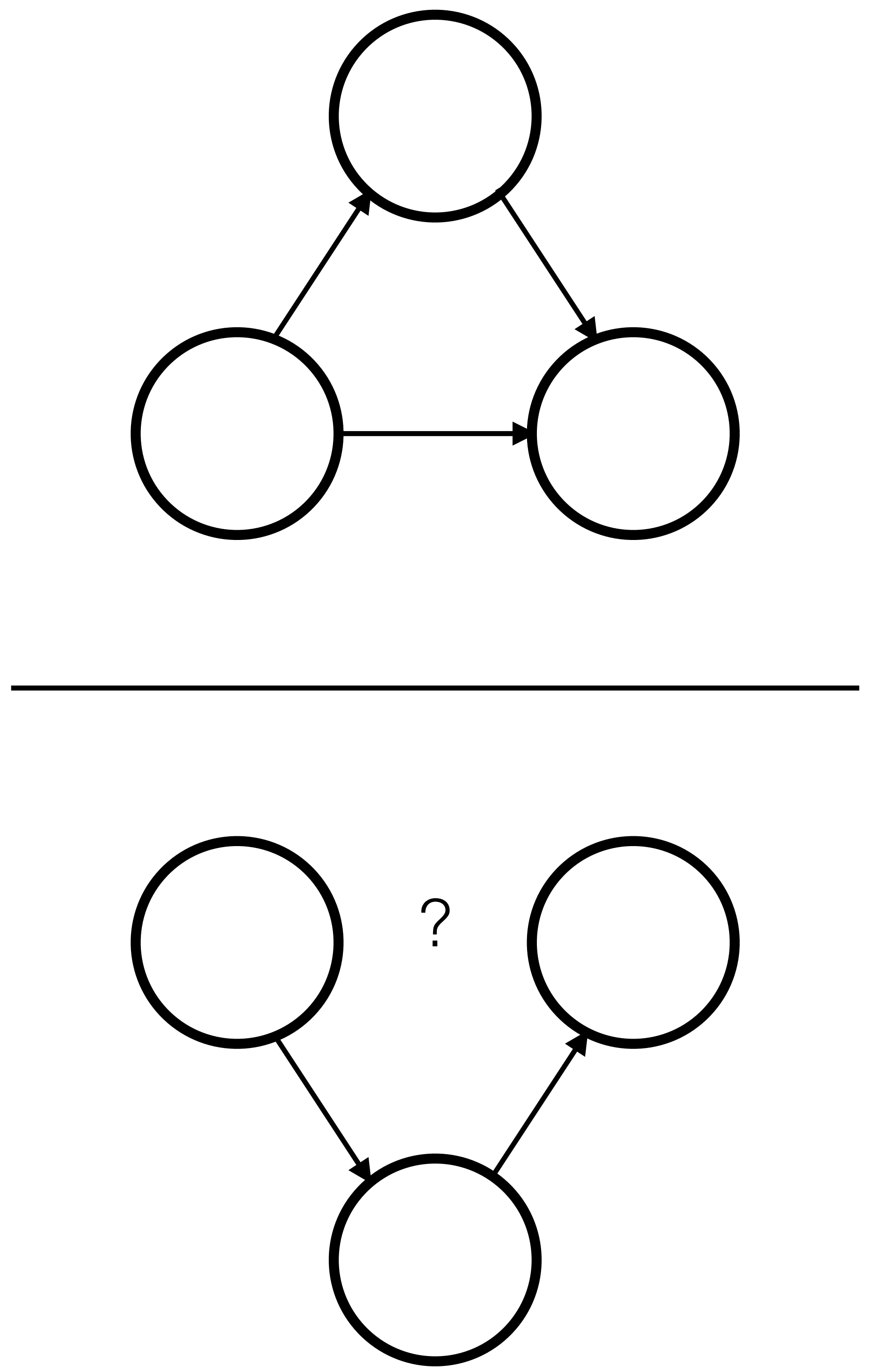
- Difficult to find the “right” answer
  - Many possible, sensible approaches, but hard to know what will be best for your purpose
- A weakness is that many tools lack rigorous theoretical underpinnings
  - A fun line of research, because anything that works is a valid solution
  - ...and we’re still waiting for the ‘one algorithm to rule them all”!

Measures of Cohesion

Reciprocity



Transitivity



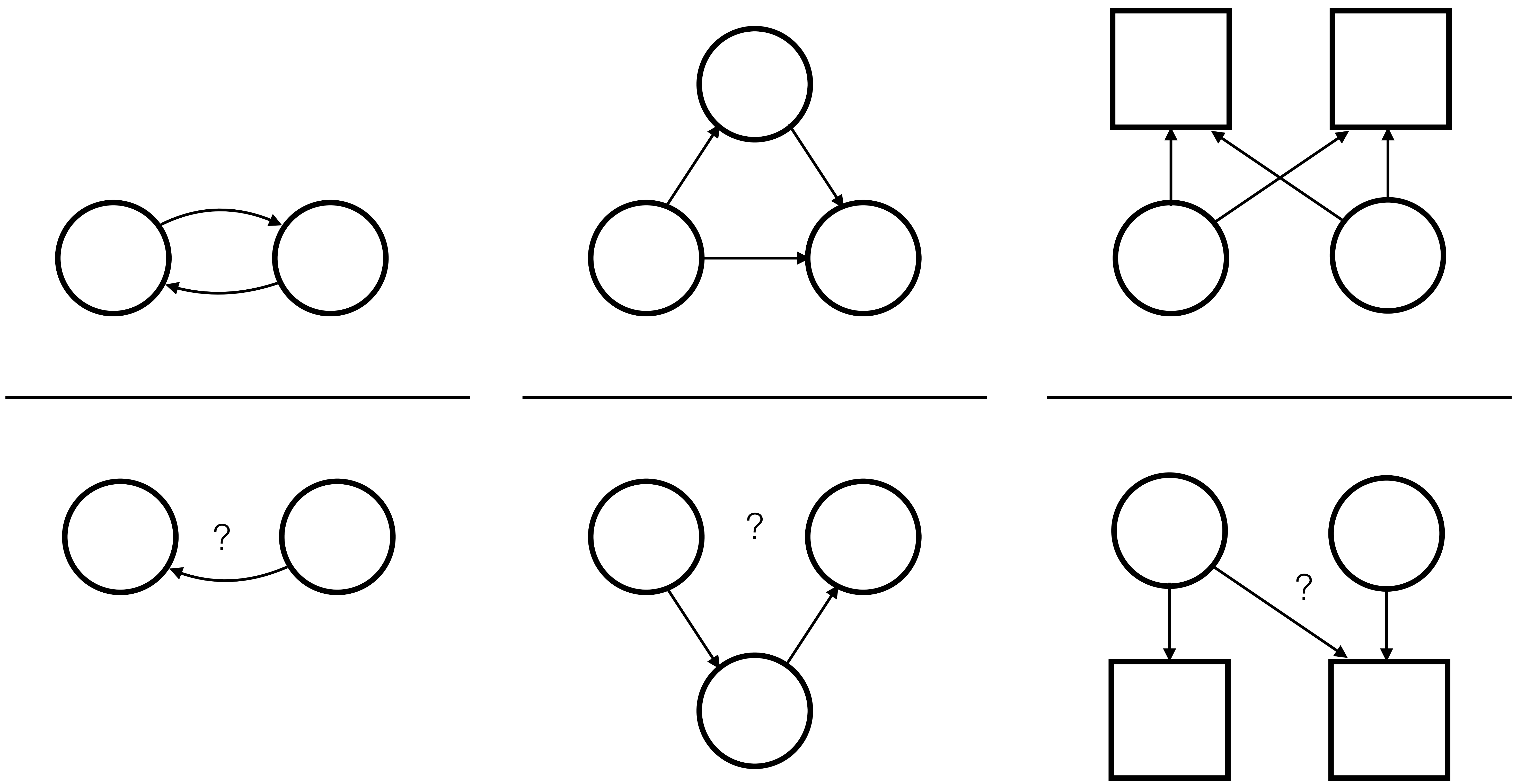
?

Measures of Cohesion

Reciprocity

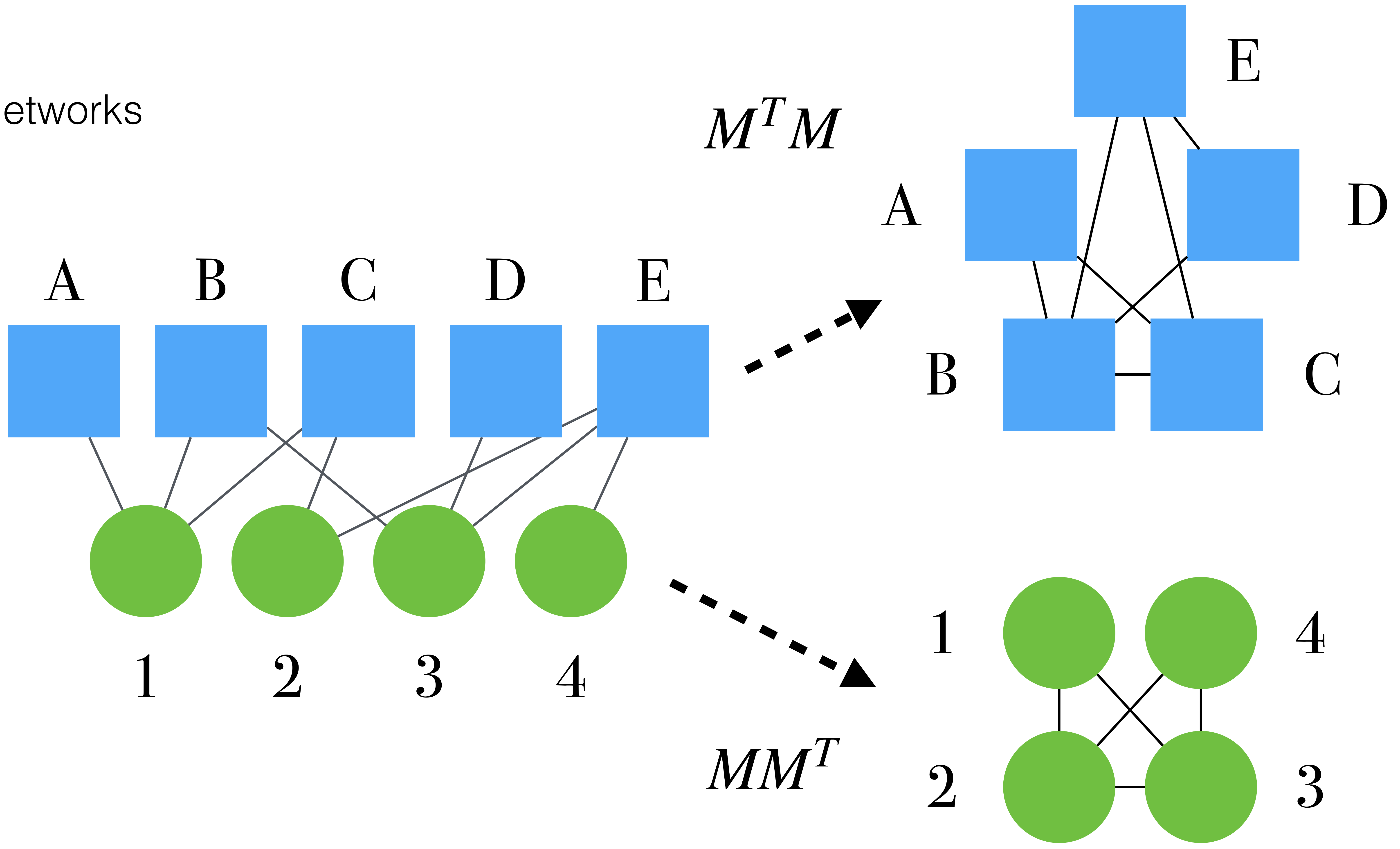
Transitivity

Equivalency
(4-cycles)



Projection

The construction of 1-2 one-mode networks
from a two-mode network



Assumption, sometimes warranted, is that if you are interested in influence across shared relationships among one set of nodes only, then analytic simplification worth the information less

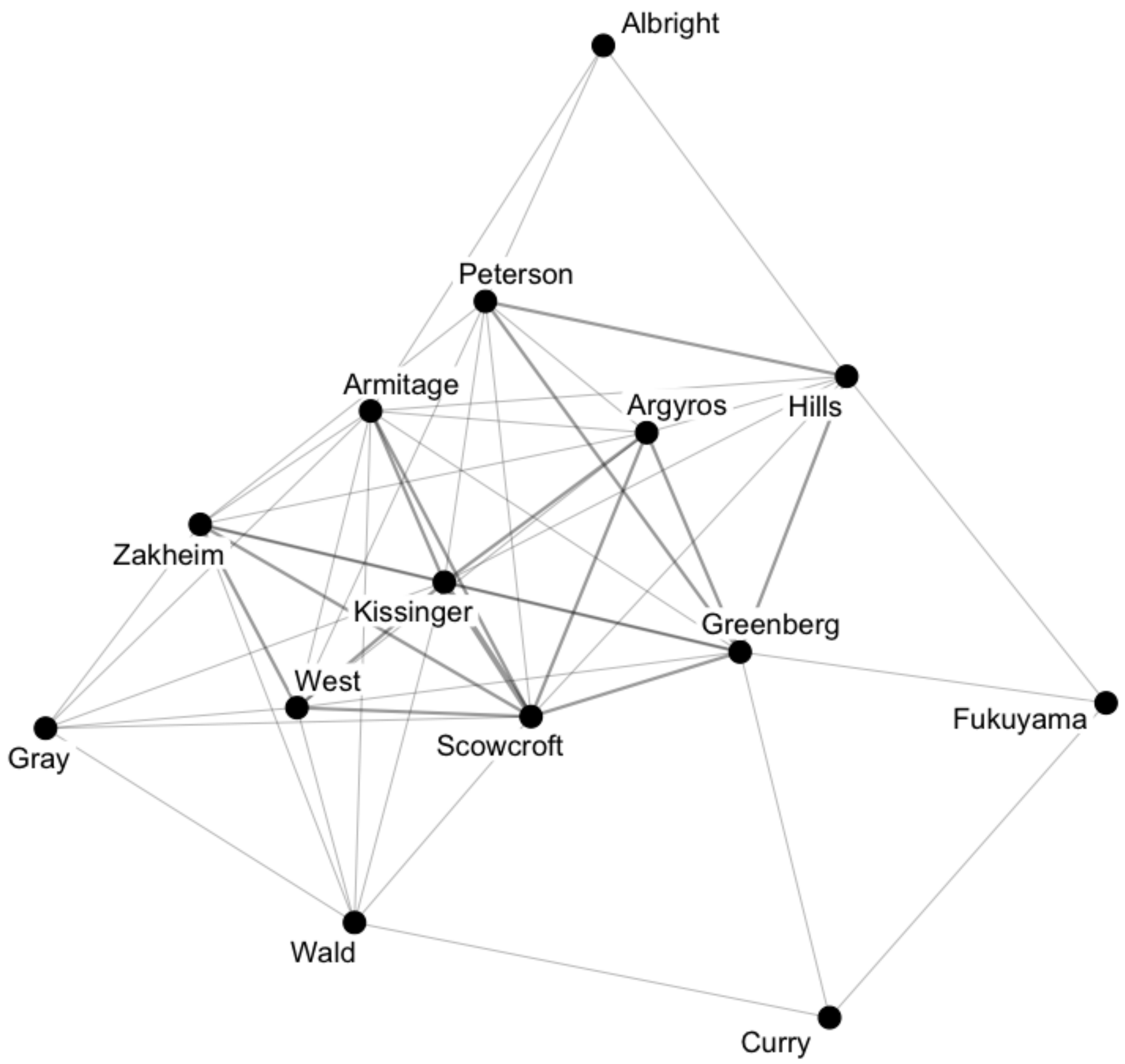
Which information is lost?

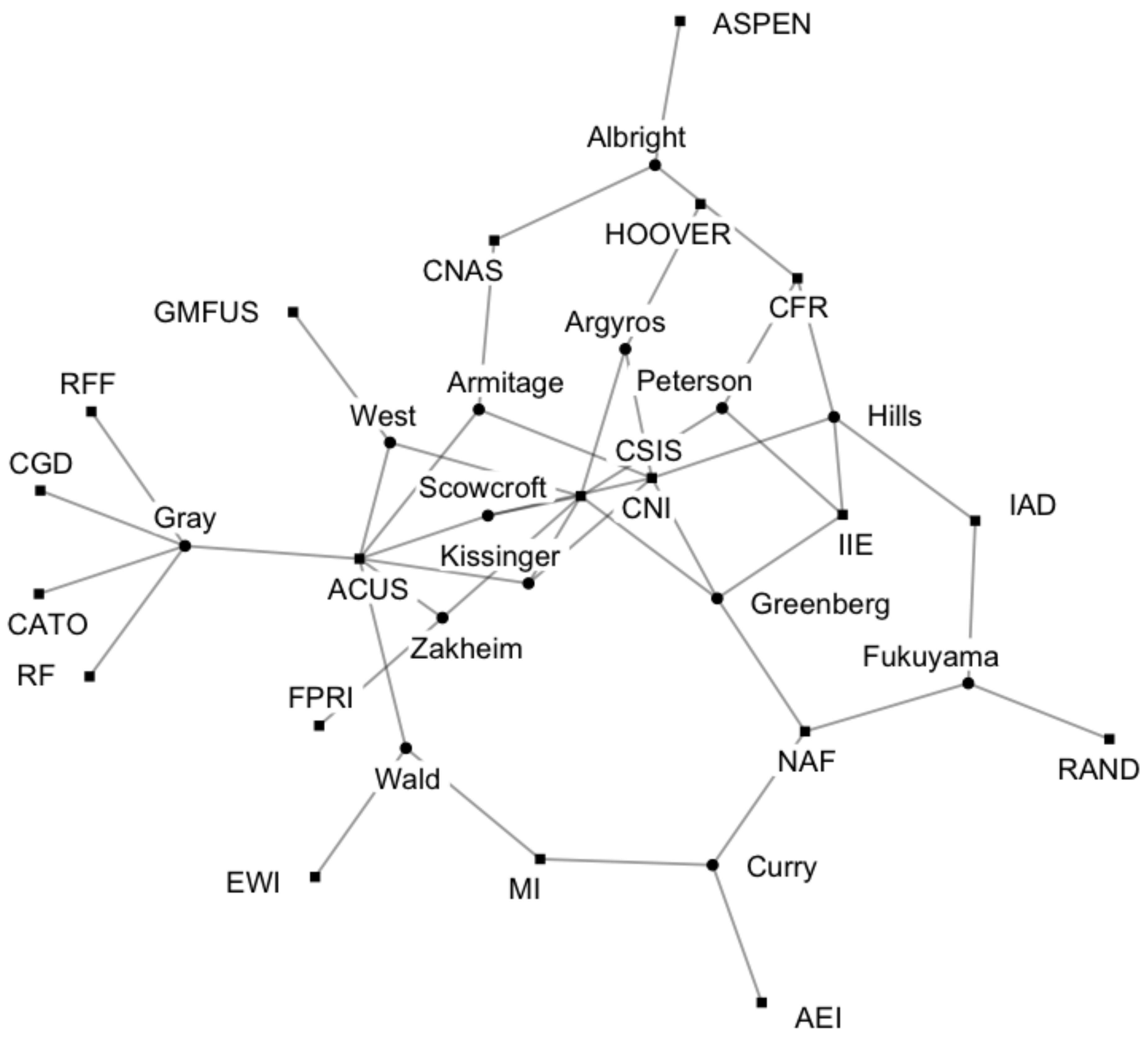
Identity

Configurational

Processual

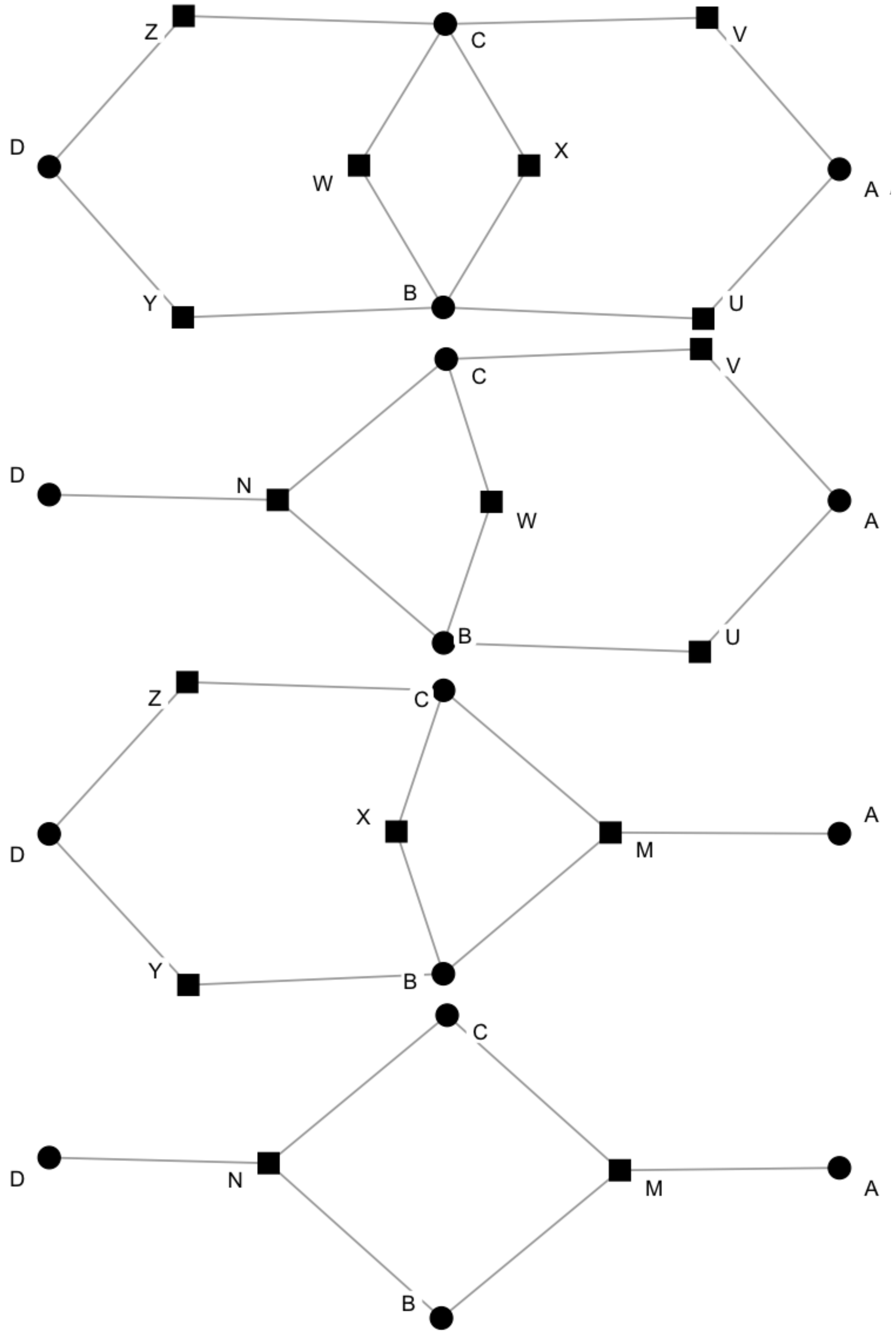
Identity information

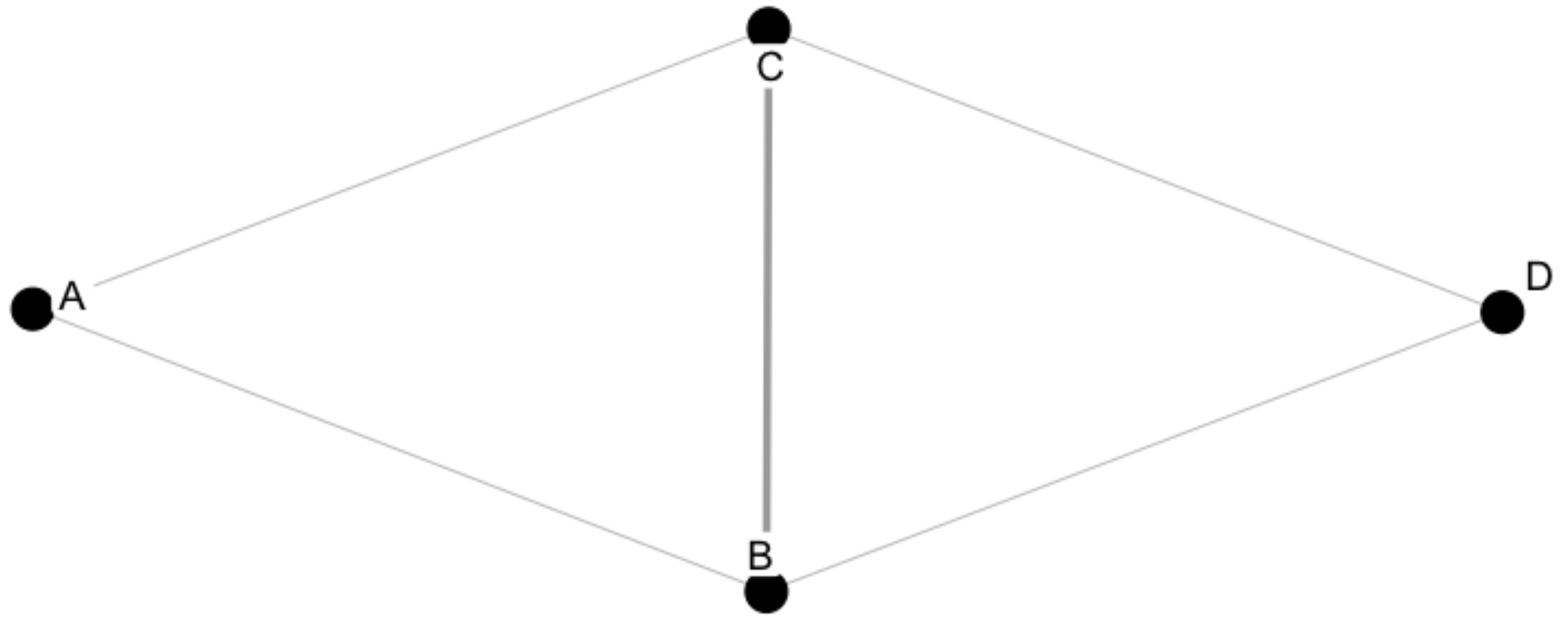


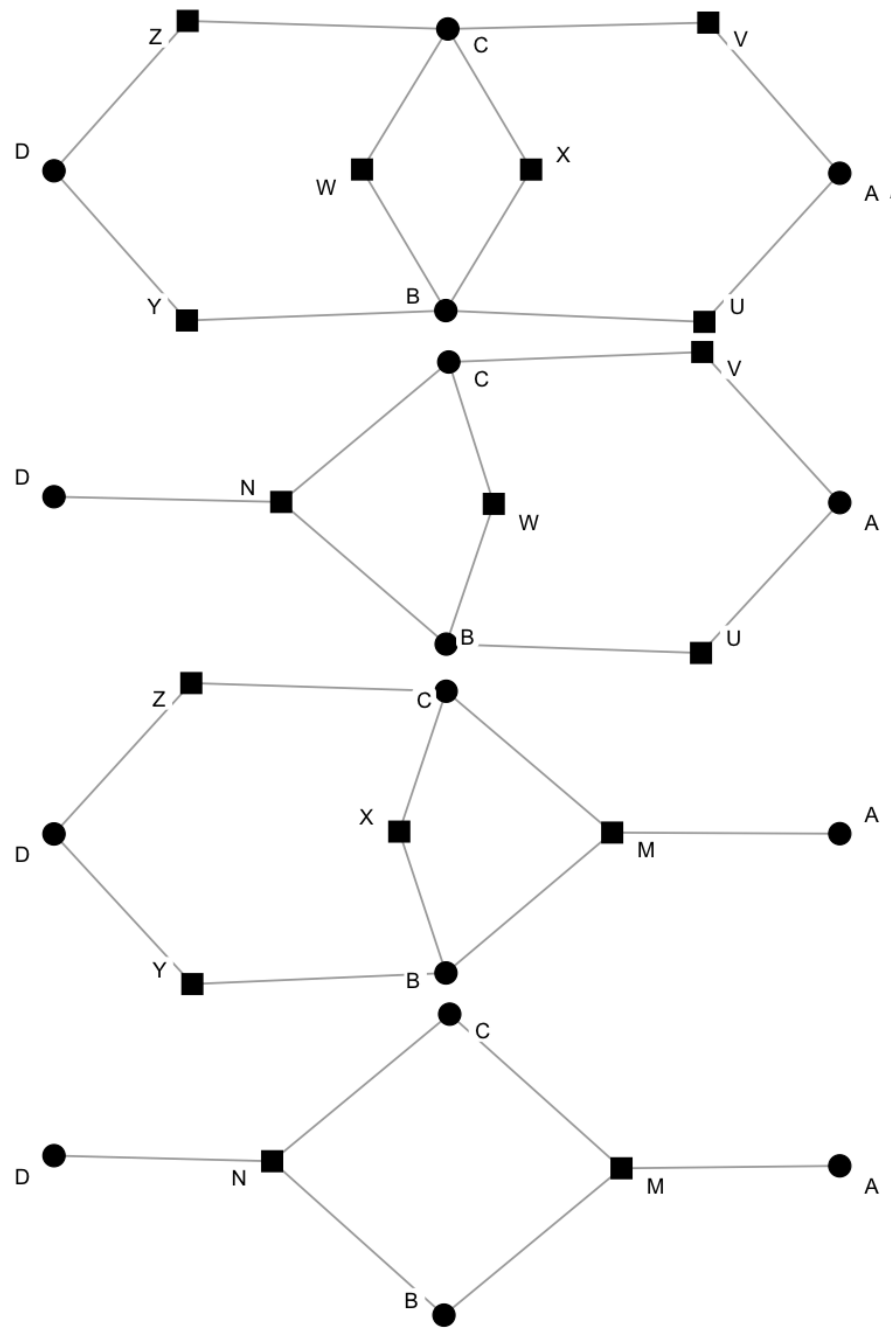


How can you tell, looking only at the graph on the left, which nodes responsible for connecting Wald and Curry or Curry and Fukuyama?

Configurational information







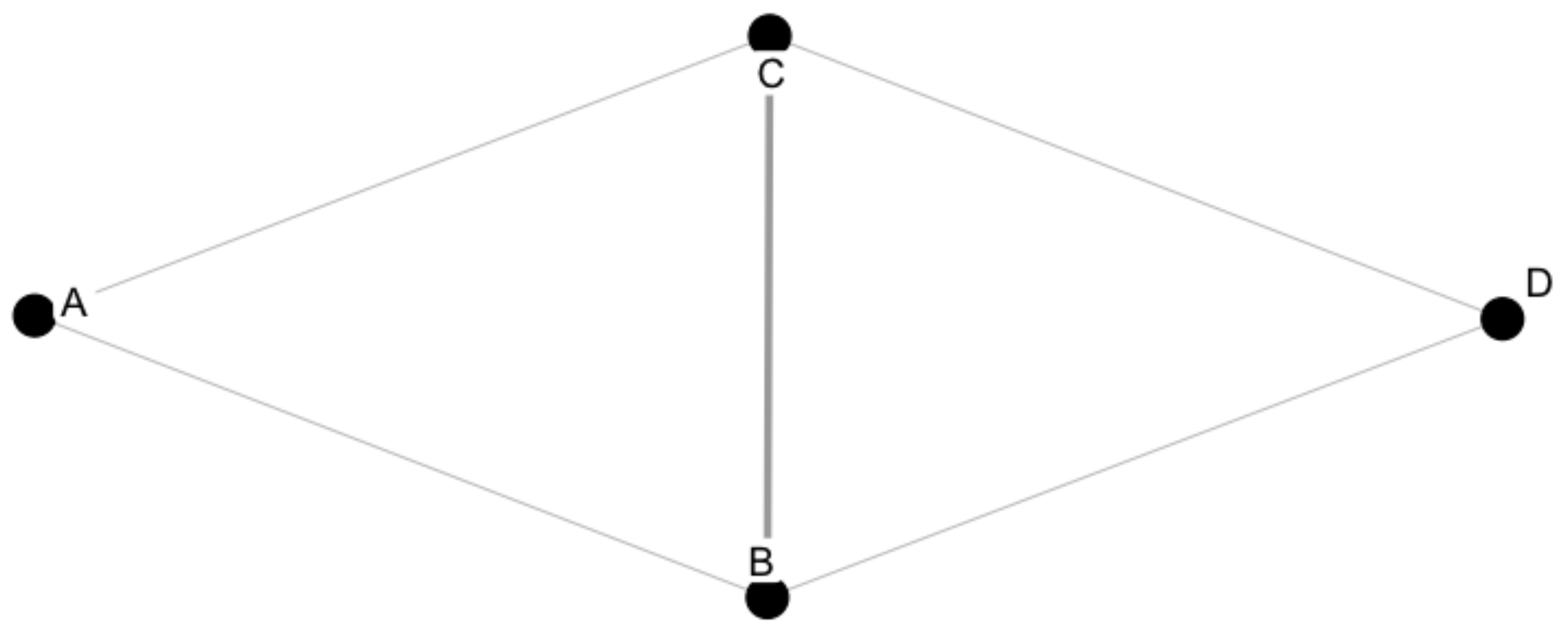
Configurational information

- High levels of triangles (Opsahl 2011)

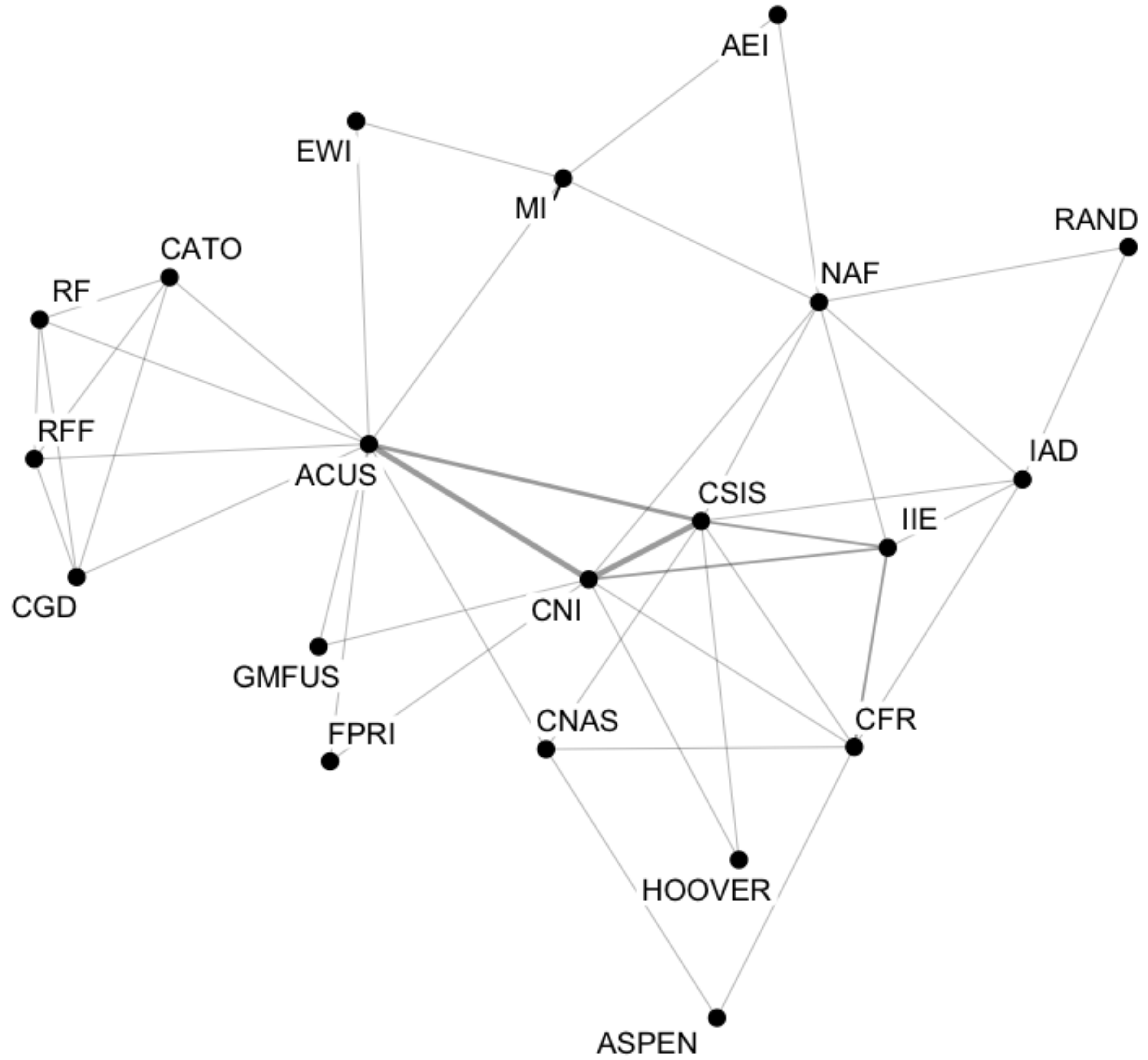
- Network measures the rely on triangles affected

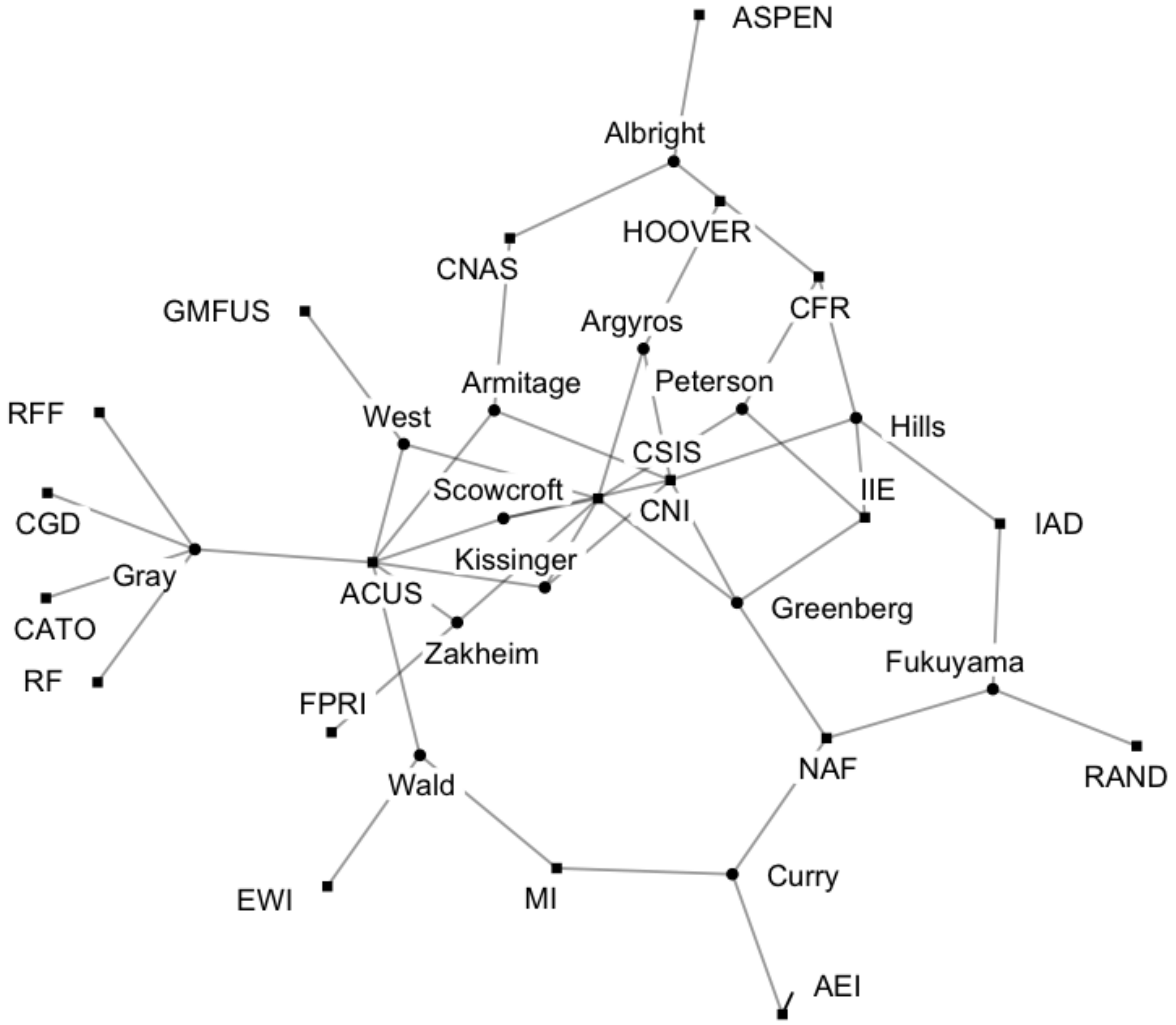
  - Clustering coefficients (Opsahl 2013), structural holes (Burt 1998)

  - Degree, radial, and medial centrality can also be sensitive



Processual information





Projections difficult to interpret, because 'shared'/joint/co- obscures choices behind ties and can result from others' choices (Hollway 2015)

Which information is lost?

Identity

Configurational

Processual

Measures of Cohesion

Reciprocity

Transitivity

Equivalency
(4-cycles)