Social Network Analysis

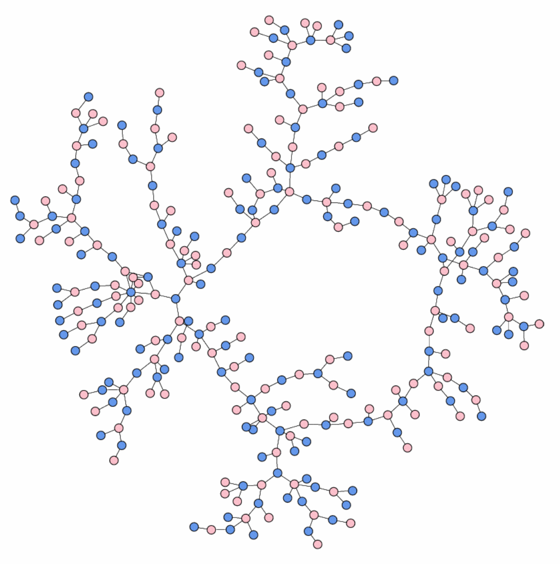
Philip Leifeld

BEAR 2018 Multi-Method Workshop

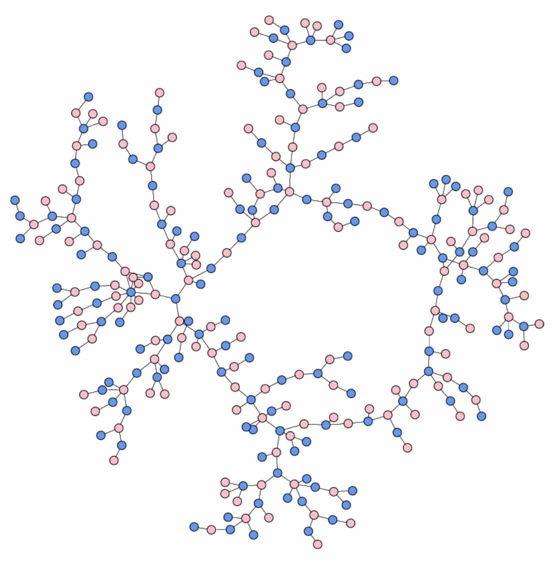
4 October 2018

University of Glasgow

What is This?

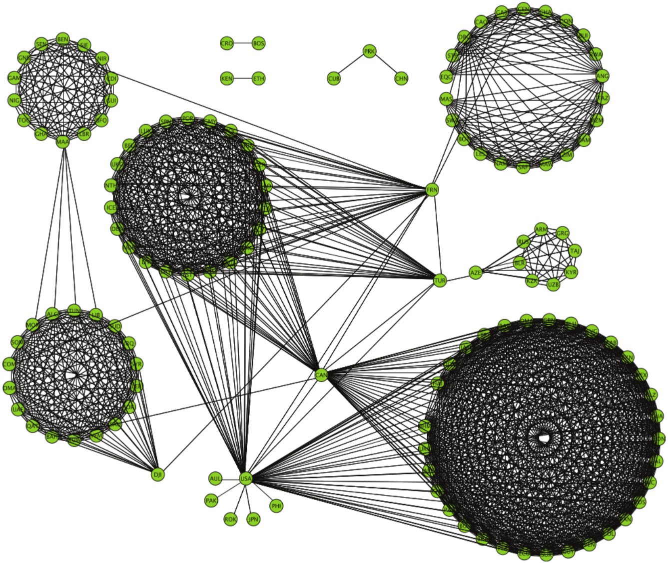


Romantic and sexual relationships at Jefferson High



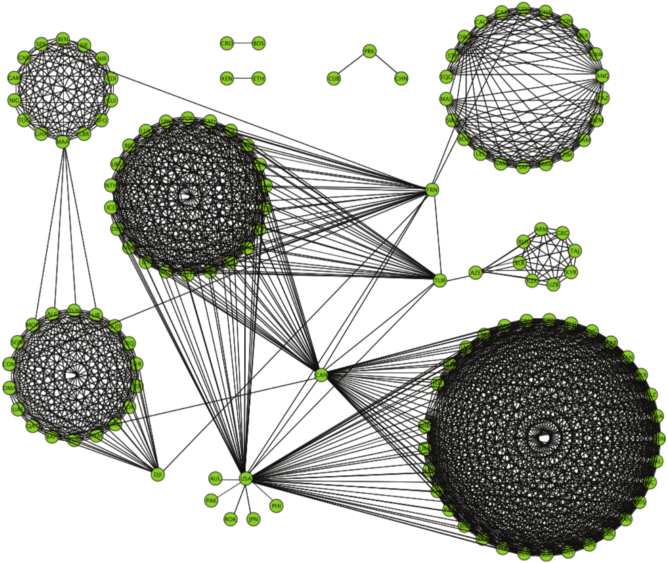
(source: Bearman, Moody, and Stovel, 2004)

What is This?



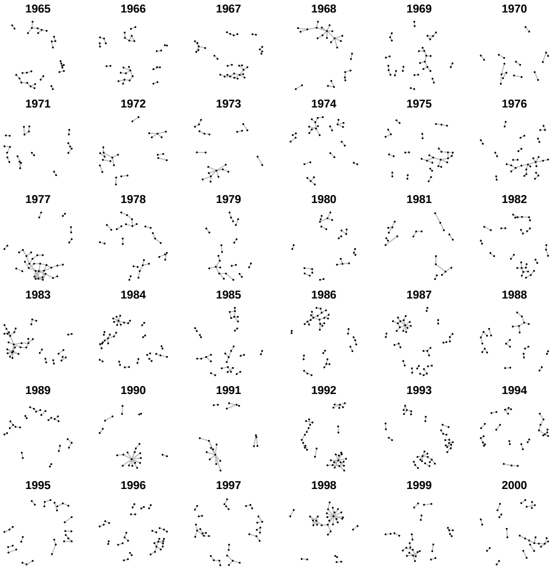
CRO
BOS
KEN
ETH
PRK
CUB
CHN
FRN
TUR
AZE
ARM
GRG
TAJ
RUS
BLR
KYR
KZK
UZB
CAN
USA
AUL
PAK
PHI
ROK
JPN
DJI
BEN
SEN
NIR
GAM
CDI
NIG
GUI
BFO
GHA
LBR
MAA
ANG

International bilateral defensive alliances in 2003



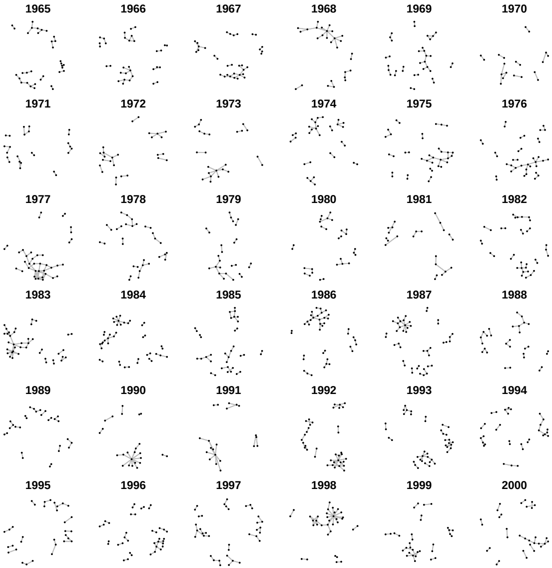
(source: Cranmer, Desmarais, and Menninga, 2012)

What is This?



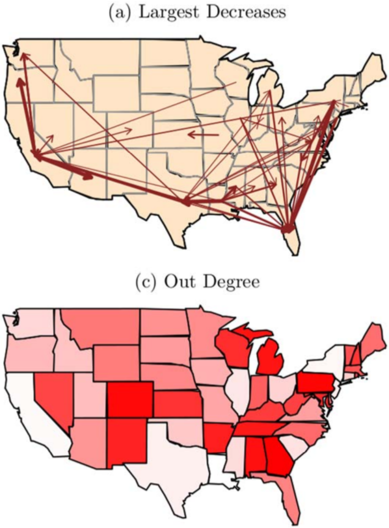
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000

Violent militarized interstate disputes, 1965–2000

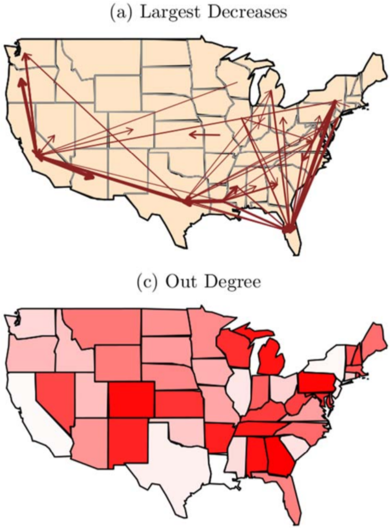


(source: Bradshaw, Leifeld, Li, Clary, and Cranmer, 2017)

What is This?

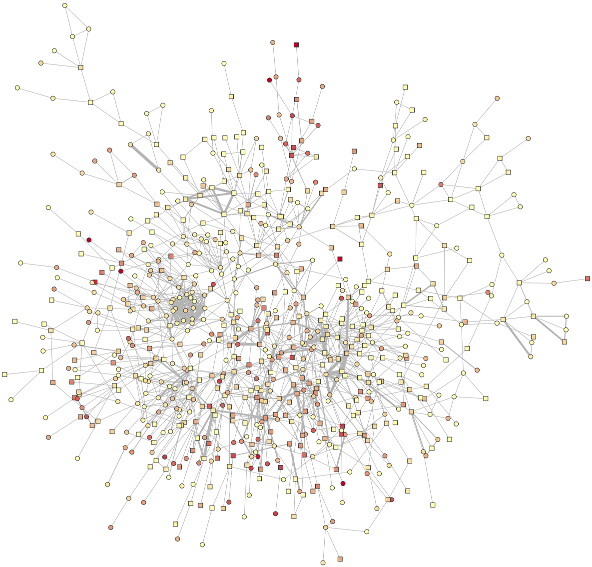

(a) Largest Decreases
(b) Largest Increases
(c) Out Degree
(d) In Degree

Change in interstate migration flows, 2006–2007

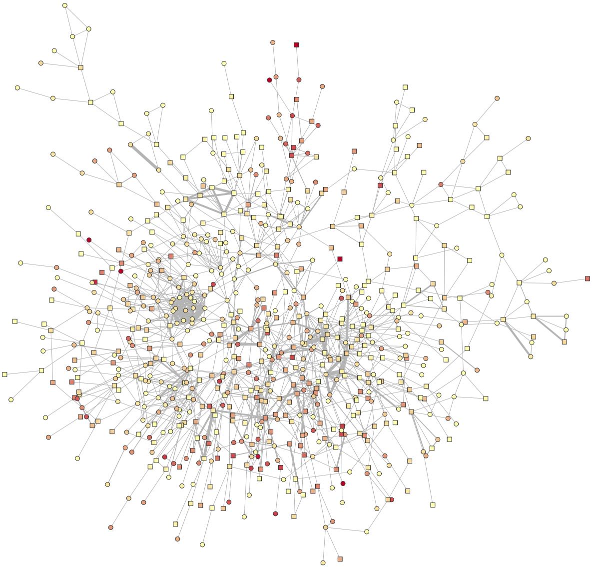


(source: Desmarais and Cranmer, 2012)

What is This?

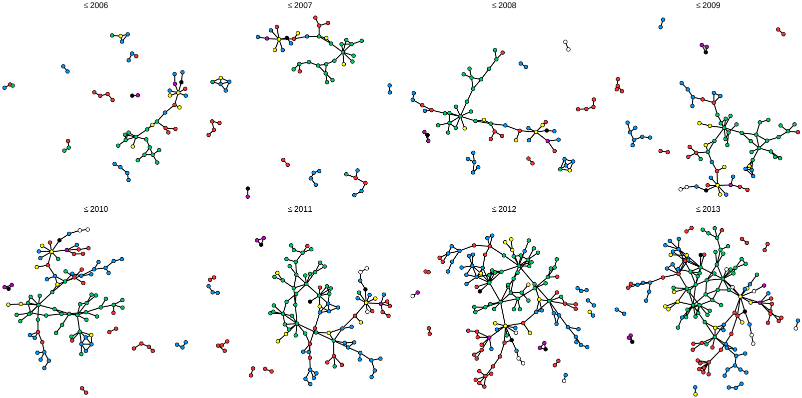


Co-authorship among German political scientists, 2014

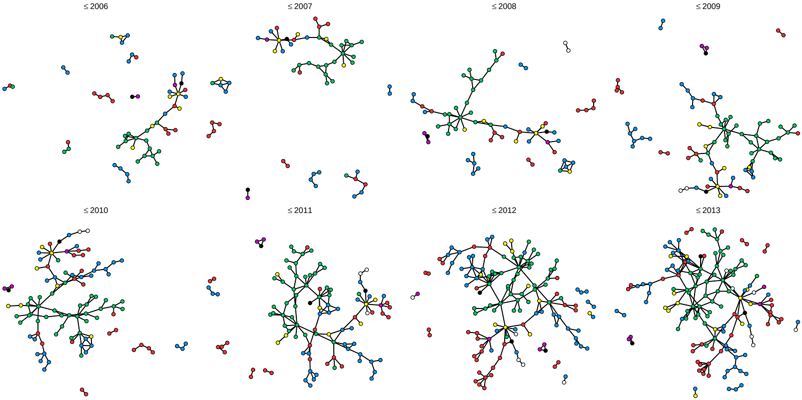


(source: Leifeld 2018)

What is This?

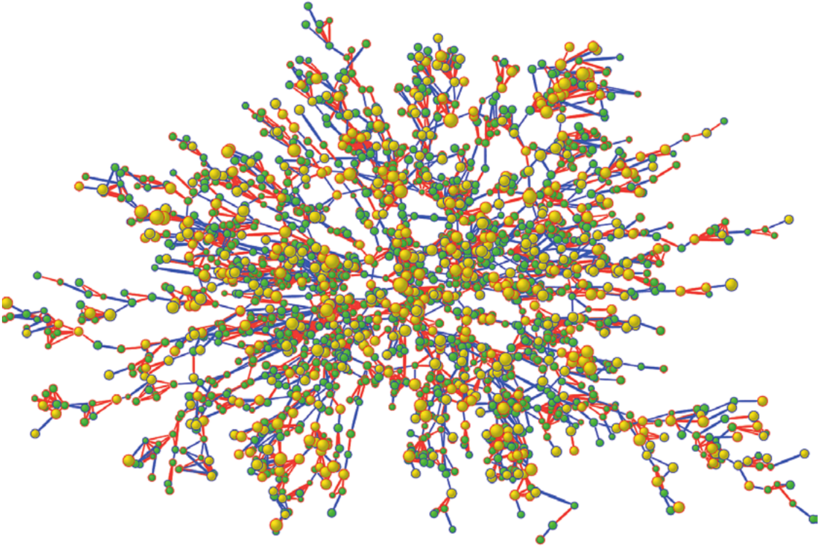


Co-authorship among Swiss political scientists, 2013

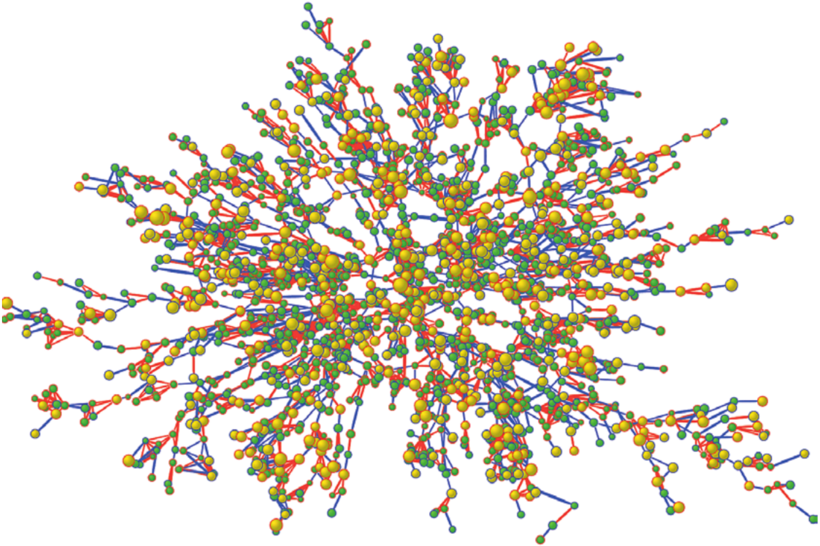


(source: Leifeld and Ingold, 2016)

What is This?



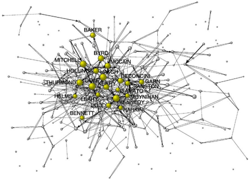
Friendship, kinship, and obesity



(source: Christakis and Fowler, 2007)

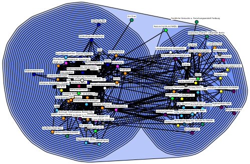
Networks are Ubiquitous in the Study of Politics

Legislative networks



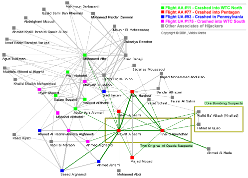
Fowler 2006

Policy processes



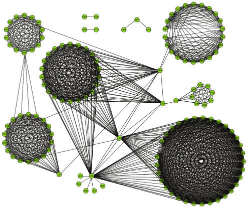
Nagel 2015

Terrorism



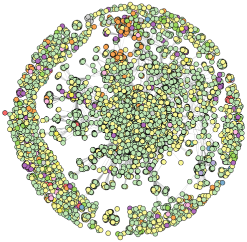
Krebs 2008

International relations



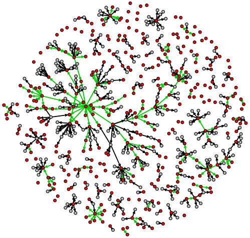
Cranmer/Desmarais/Menninga 2012

Interest groups



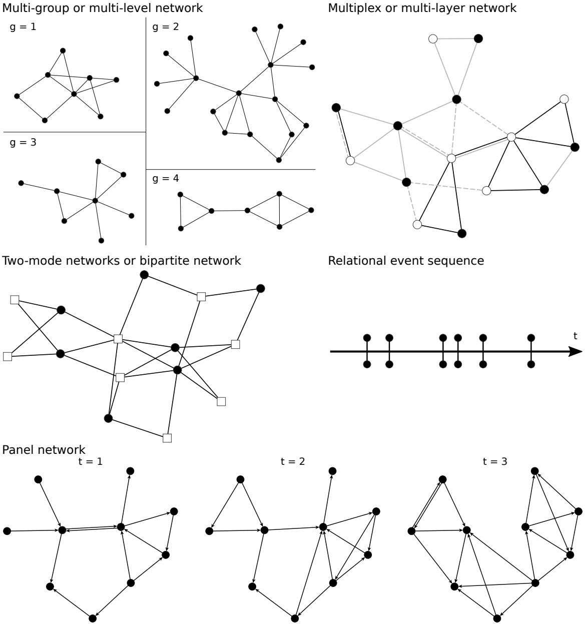
Box-Steffensmeier/Christenson 2014

Epistemic communities



Leifeld/Fisher 2017

Network Types



Basic Methodological Distinction

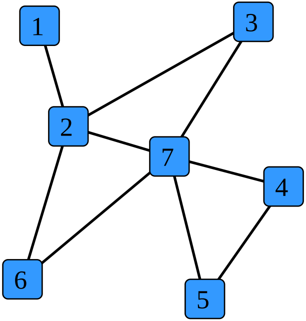
Descriptive Network Analysis

- Node level: centrality, or the importance of nodes.
- Meso level: subgroup analysis, or which clusters or communities is the network composed of?
- Network level: density, centralisation, clustering etc.

Inferential Network Analysis

- Explaining the structure of the network.
- Explaining the attributes of nodes in a network.
- Explaining temporal change of attributes or structure.

Elements of networks



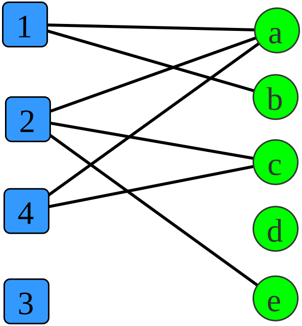
A network N consists of...

- vertices (*nodes*, *points*)
- denoted as i, j, k
- edges (*ties*, *lines*)
- denoted as N_{ij}, N_{ik} etc.

A network...

- is a descriptive model of social reality.
- depicts relations rather than attributes.
- often represents the outcome of a dynamic process.

Two-mode networks



Two-mode networks:

- a.k.a. bipartite graphs
- a.k.a. affiliation networks
- two classes of nodes
- no within-class edges

Examples

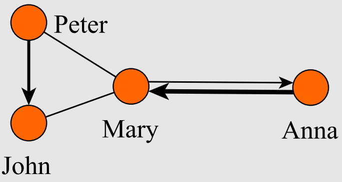
- employees and departments
- organizations and associations
- managers and boards of directors

Data structures for network analysis

Matrix

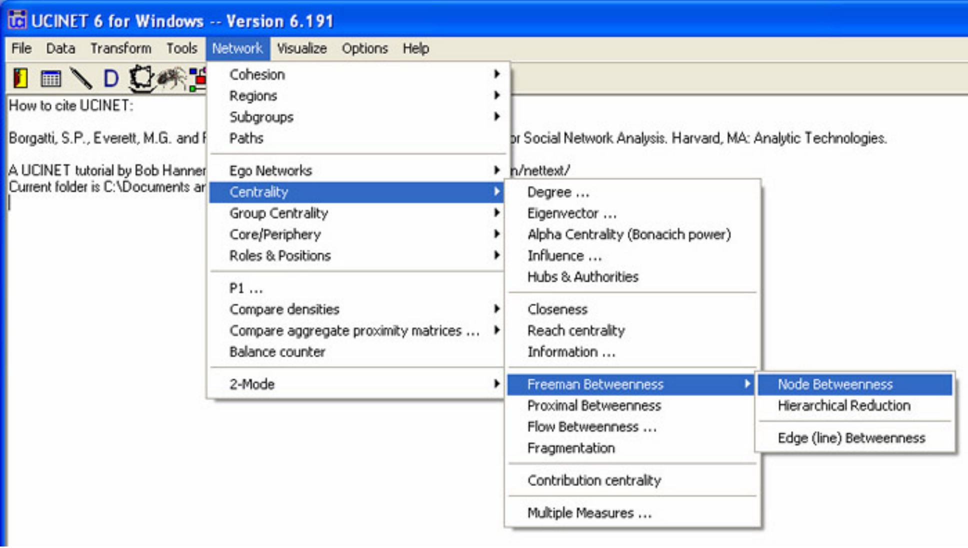
	P	J	A	M
Peter		2	0	1
John	0		0	1
Anna	0	0		3
Mary	1	1	1	

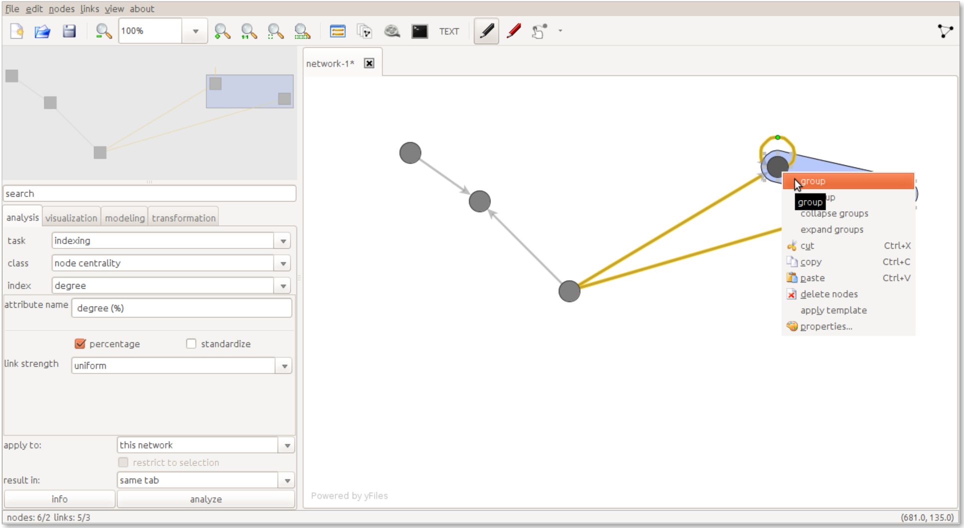
Graph

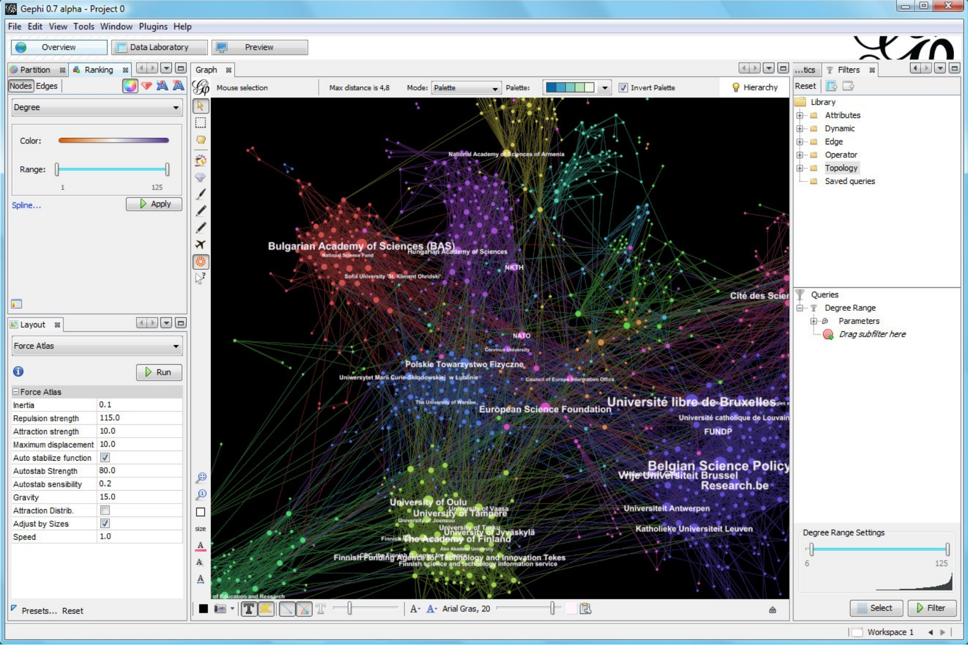


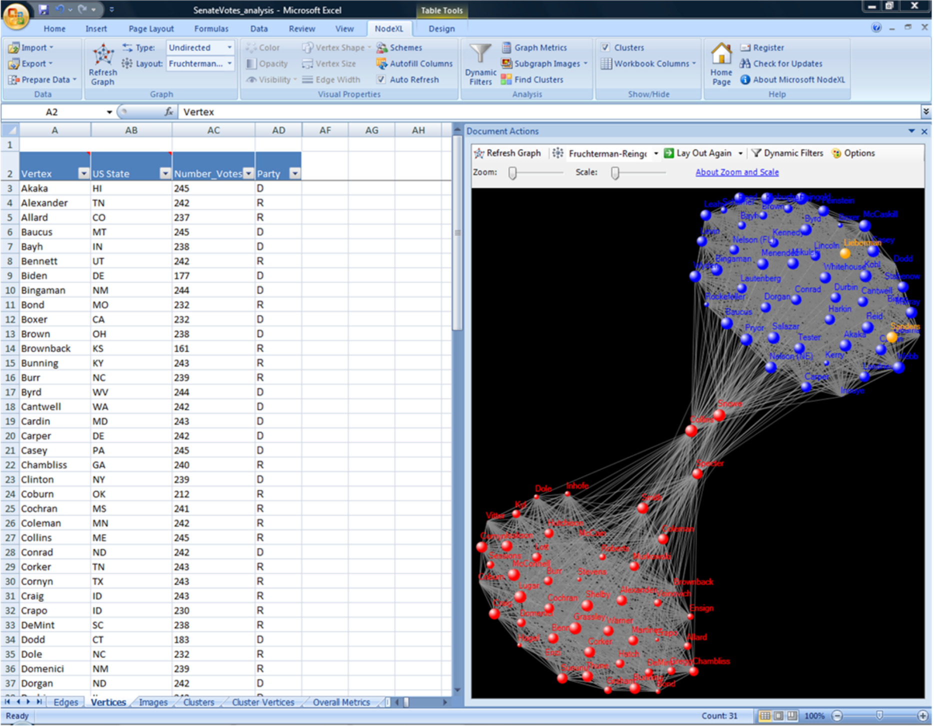
Edge list

Peter	$\rightarrow$	John	2
Peter	$\rightarrow$	Mary	1
John	$\rightarrow$	Mary	1
Anna	$\rightarrow$	Mary	3
Mary	$\rightarrow$	Peter	1
Mary	$\rightarrow$	John	1
Mary	$\rightarrow$	Anna	1


UCINET 6 for Windows -- Version 6.191
File
Data
Transform
Tools
Network
Visualize
Options
Help
How to cite UCINET:
Borgatti, S.P., Everett, M.G. and F
r Social Network Analysis. Harvard, MA: Analytic Technologies.
A UCINET tutorial by Bob Hanner
n/nettext/
Current folder is C:\Documents ar
Cohesion
Regions
Subgroups
Paths
Ego Networks
Centrality
Group Centrality
Core/Periphery
Roles & Positions
P1 ...
Compare densities
Compare aggregate proximity matrices ...
Balance counter
2-Mode
Degree ...
Eigenvector ...
Alpha Centrality (Bonacich power)
Influence ...
Hubs & Authorities
Closeness
Reach centrality
Information ...
Freeman Betweenness
Proximal Betweenness
Flow Betweenness ...
Fragmentation
Contribution centrality
Multiple Measures ...
Node Betweenness
Hierarchical Reduction
Edge (line) Betweenness


file edit nodes links view about
100%
TEXT
network-1*
search
analysis
visualization
modeling
transformation
task
indexing
class
node centrality
index
degree
attribute name
degree (%)
percentage
standardize
link strength
uniform
apply to:
this network
restrict to selection
result in:
same tab
info
analyze
group
group
collapse groups
expand groups
cut Ctrl+X
copy Ctrl+C
paste Ctrl+V
delete nodes
apply template
properties...
Powered by yFiles
nodes: 6/2 links: 5/3
(681.0, 135.0)


Gephi 0.7 alpha - Project 0
File Edit View Tools Window Plugins Help
Overview
Data Laboratory
Preview
Partition
Ranking
Nodes Edges
Degree
Color:
Range:
1
125
Spline...
Apply
Layout
Force Atlas
Run
Force Atlas
Inertia 0.1
Repulsion strength 115.0
Attraction strength 10.0
Maximum displacement 10.0
Auto stabilize function
Autostab Strength 80.0
Autostab sensibility 0.2
Gravity 15.0
Attraction Distrib.
Adjust by Sizes
Speed 1.0
Presets... Reset
Graph
Mouse selection
Max distance is 4,8
Mode: Palette
Palette:
Invert Palette
Hierarchy
National Academy of Sciences of Armenia
Bulgarian Academy of Sciences (BAS)
Hungarian Academy of Sciences
NKTH
Cité des Scier
NATO
Polskie Towarzystwo Fizyczne
European Science Foundation
Université libre de Bruxelles
Université catholique de Louvain
FUNDP
Belgian Science Policy
Vrije Universiteit Brussel
Research.be
Universiteit Antwerpen
Katholieke Universiteit Leuven
University of Oulu
University of Tampere
University of Jyväskylä
The Academy of Finland
Finnish science and technology information service
size
Arial Gras, 20
...tics
Filters
Reset
Library
Attributes
Dynamic
Edge
Operator
Topology
Saved queries
Queries
Degree Range
Parameters
Drag subfilter here
Degree Range Settings
6
125
Select
Filter
Workspace 1



Vertex	US State	Number_Votes	Party
Akaka	HI	245	D
Alexander	TN	242	R
Allard	CO	237	R
Baucus	MT	245	D
Bayh	IN	238	D
Bennett	UT	242	R
Biden	DE	177	D
Bingaman	NM	244	D
Bond	MO	232	R
Boxer	CA	232	D
Brown	OH	238	D
Brownback	KS	161	R
Bunning	KY	243	R
Burr	NC	239	R
Byrd	WV	244	D
Cantwell	WA	242	D
Cardin	MD	243	D
Carper	DE	242	D
Casey	PA	245	D
Chambliss	GA	240	R
Clinton	NY	239	D
Coburn	OK	212	R
Cochran	MS	241	R
Coleman	MN	242	R
Collins	ME	245	R
Conrad	ND	242	D
Corker	TN	243	R
Cornyn	TX	243	R
Craig	ID	243	R
Crapo	ID	230	R
DeMint	SC	238	R
Dodd	CT	183	D
Dole	NC	232	R
Domenici	NM	239	R
Dorgan	ND	242	D

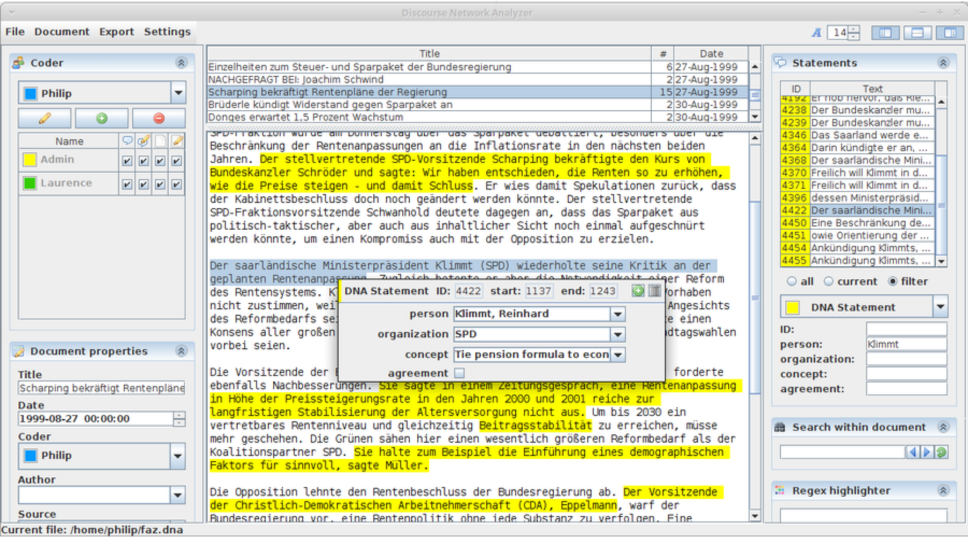
R Packages for Network Analysis

- statnet
- xergm
- RSiena
- igraph

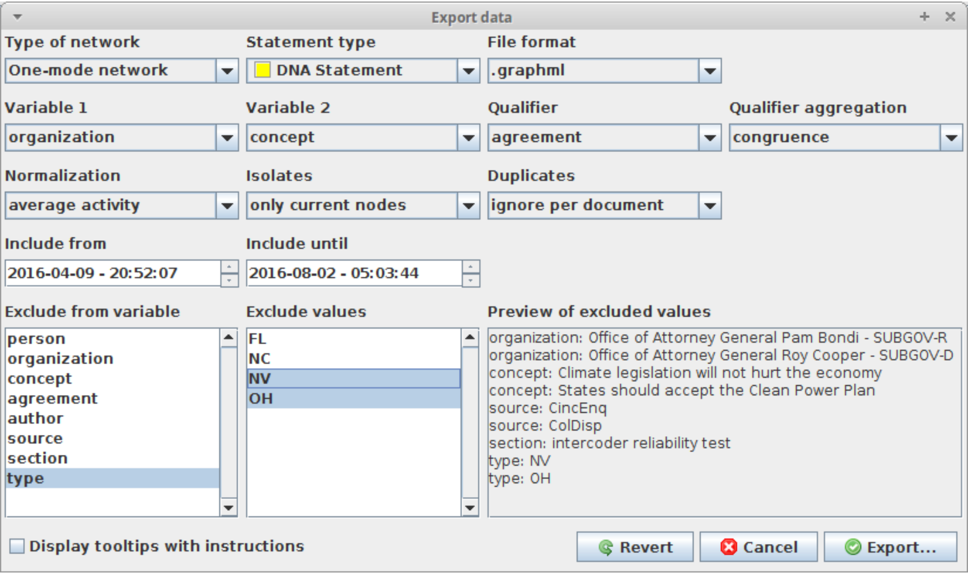
Discourse Network Analyzer

- name: Discourse Network Analyzer (DNA)
- download: `http://www.github.com/leifeld/dna`
- operating system: any (platform-independent!)
- requirements: Java 8
- purpose:
  1. assign tags to text data
  2. convert these structured data into networks

Discourse Network Analyzer: Main Window

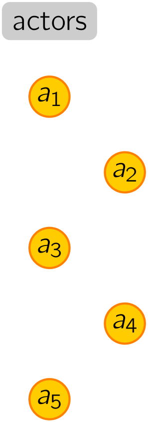


Discourse Network Analyzer: Network Export Window

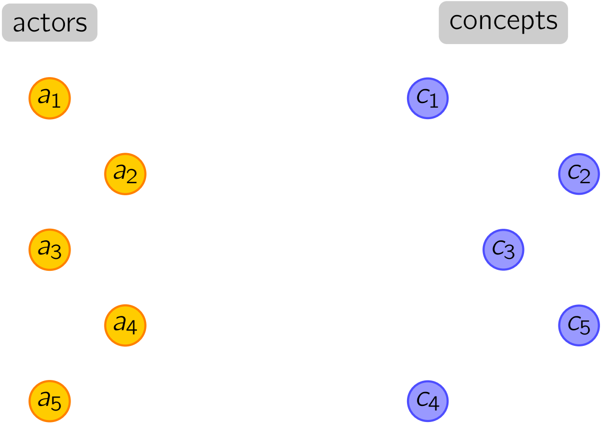


```r
affil <- dna_network(conn,
                     networkType = "twomode",
                     statementType = "DNA Statement",
                     variable1 = "organization",
                     variable2 = "concept",
                     qualifier = "agreement",
                     qualifierAggregation = "combine",
                     duplicates = "document",
                     verbose = FALSE)
plot(nw,
     edge.col = get.edge.attribute(nw, "color"),
     vertex.col = c(rep("white", nrow(affil)),
                    rep("black", ncol(affil))),
     displaylabels = TRUE,
     label.cex = 0.5
     )
```

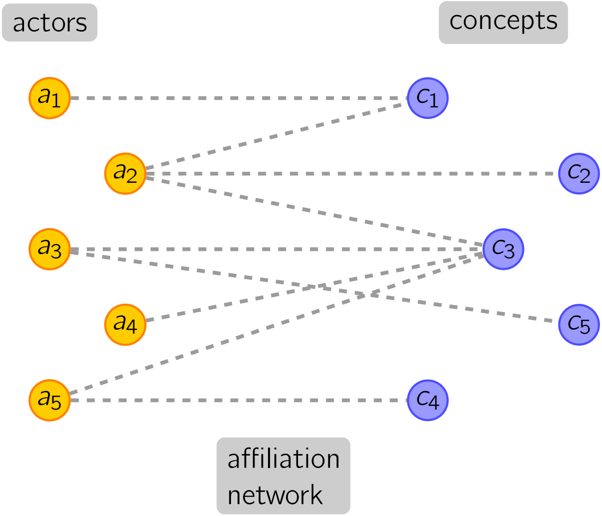
Graphical intuition of discourse networks



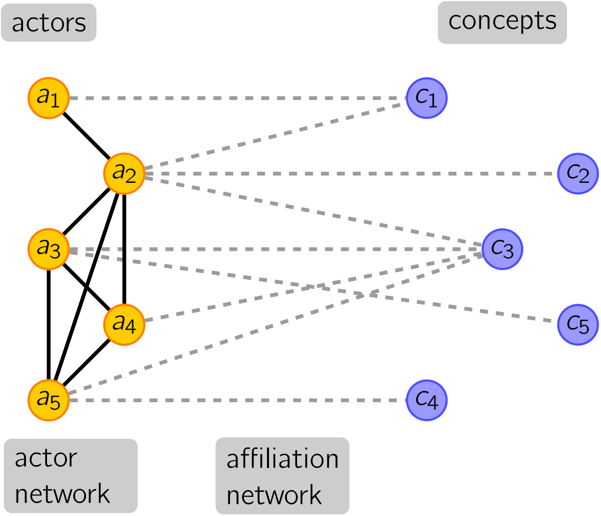
Graphical intuition of discourse networks



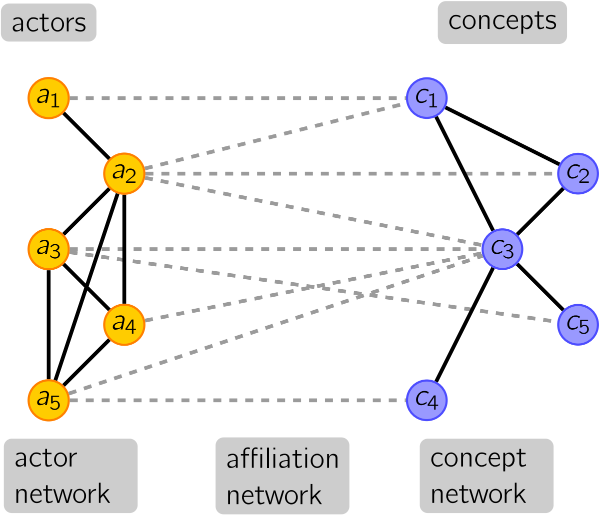
Graphical intuition of discourse networks



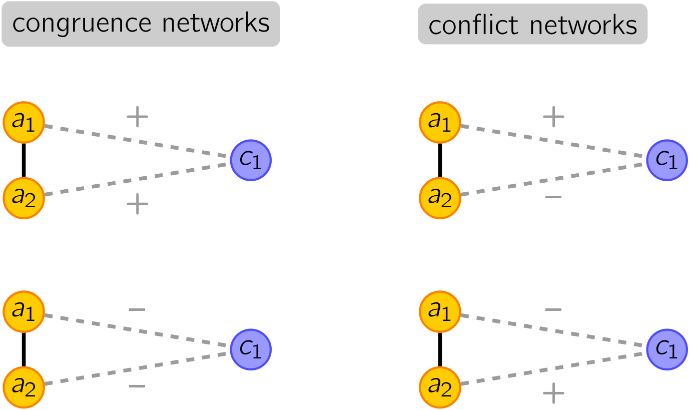
Graphical intuition of discourse networks



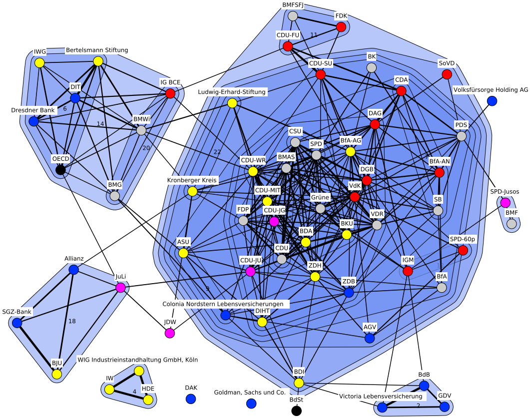
Graphical intuition of discourse networks



Extension: agreement and disagreement

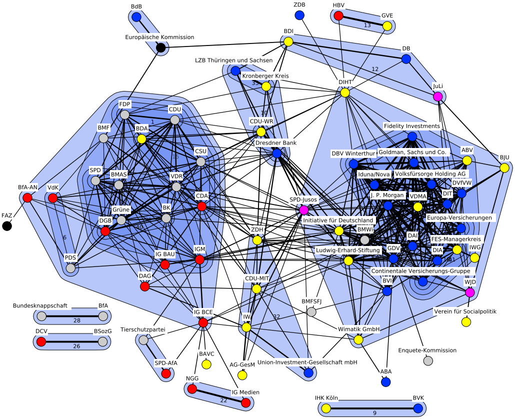


Actor congruence network in 1997 ($w \geq 0.31$)



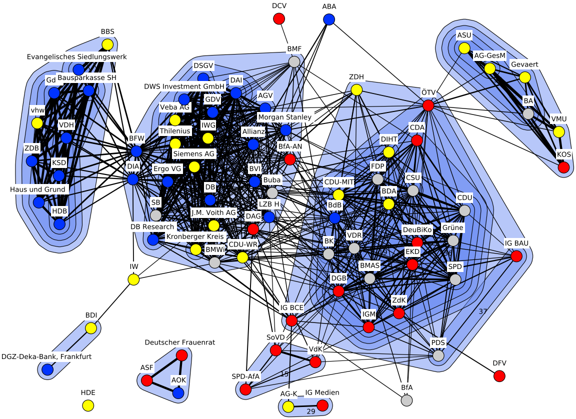
Financial interest groups (= blue nodes) are scattered around a single corporatist community.

Actor congruence network in 1998 ($w \geq 0.29$)



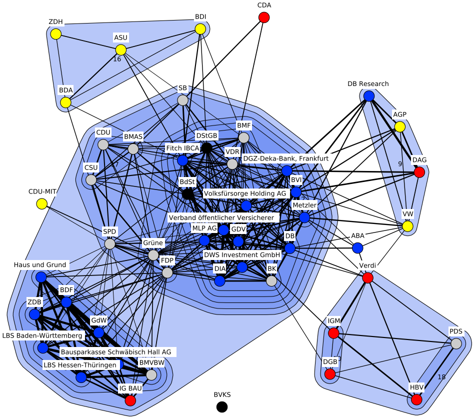
Financial interest groups (= blue nodes) start to make more coherent claims; polarization emerges.

Actor congruence network in 2000 ($w \geq 0.27$)



Polarization becomes more extreme. Some actors leave their coalition and join the new coalition.

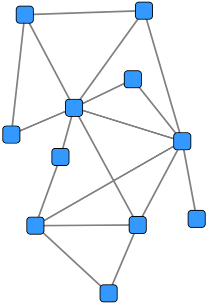
Actor congruence network in 2001 ($w \geq 0.23$)



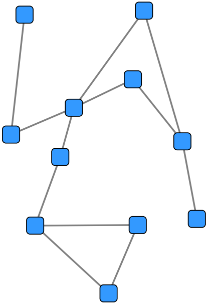
The old coalition erodes. Their actors are now scattered around the new coalition.

Density

- Density measures how many edges are present in a network.
- Equation: $d = \frac{\text{edges present}}{\text{edges possible}}$



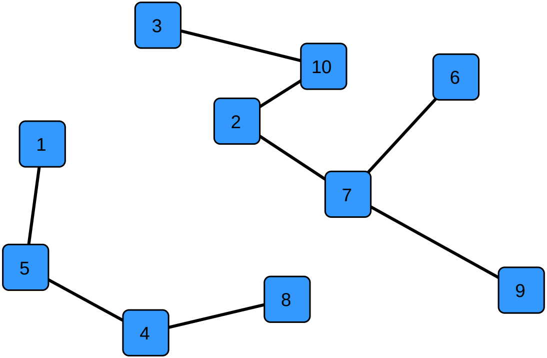
dense graph with $d = 0.33$



sparse graph with $d = 0.22$

Subgraph, component

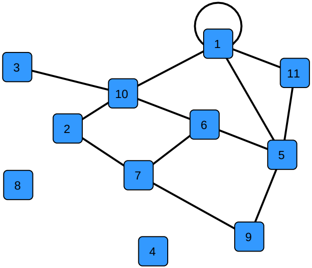
A subgraph is any part of a network (whether connected or not).
A component is a subgraph that is maximally connected.



This graph contains two components.

Walk, path, trail, isolate, pendant

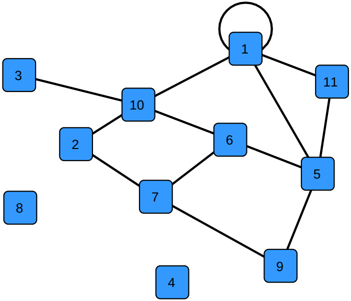
- A walk or chain is a sequence of incident vertices and edges, e. g. 10-6-7-6-5.
- A trail is a walk where an edge is not allowed to appear more than once, e. g. 7-6-5-9-7-2.
- A path is a walk where neither an edge, nor a vertex may appear more than once, e. g. 9-5-6-7-2.



- The degree of a vertex is its number of incident edges.
- An isolate is a vertex with a degree of 0 (e. g. 4 or 8).
- A hanger or pendant is a vertex with a degree of 1 (e. g. 3).

Geodesic, cut vertex, diameter

- The geodesic or geodesic distance or path distance is the shortest path connecting two vertices. In our example, there are two geodesics of length 3 between vertices 3 and 5.



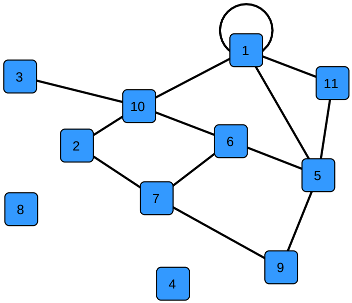
- A cut vertex or bridge is a vertex whose removal would cause the graph to be cut into several components, e. g. vertex 10.
- The diameter of a component is the maximum geodesic observed in the component. Our example has a diameter of 4 (this corresponds to the vertices 3 and 9).

Dr. Stanley Milgram 1933 – 1984, an American social psychologist at Yale, Harvard and the City University of New York, conducted in 1967 the small-world experiment that is the foundation of the six degrees of separation concept.

Milgram sent several packages to random people in the United States, asking them to forward the package, by hand, to someone specific or someone who is more likely to know the target. The average path length for the received packages was around 5.5 or six, resulting in widespread acceptance for the term six degrees of separation.

Dyad, triad, cycle, star

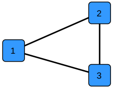
- A dyad is any pair of two vertices. In a stricter definition, a dyad is an adjacent pair of vertices, e g. 3–10 or 6–7.
- A triad or triangle is a completely connected subgraph of three vertices (1–5–11).

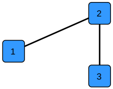


- A cycle is a closed path, e. g. 6–5–9–7.
- In a star, a vertex is connected to all other vertices, but they are not connected with each other (e. g. 1–2–3–6–10)
- A loop is an edge where the source vertex and the target vertex are identical. This corresponds to a diagonal cell entry in a matrix (e. g. 1).

Triplet, clustering coefficient

- A triad is composed of three triplets.
- A closed triplet or triple is a set of three vertices which are all adjacent.
- An open triplet or triple is a set of three vertices which are connected by two edges.

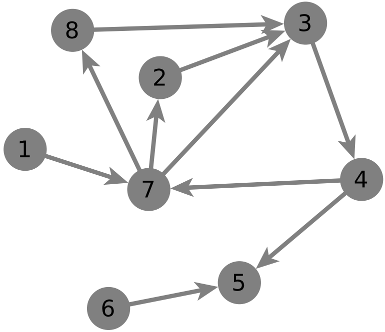




- The global clustering coefficient measures the degree to which vertices tend to cluster together in a graph.
  $C(G) = \frac{\text{closed triplets}}{\text{closed triplets} + \text{open triplets}}$
- The local clustering coefficient measures the degree to which the neighborhood of a certain vertex is clustered:
  $C(v) = \frac{\text{realized edges among vertices adjacent to } v}{\text{possible edges among vertices adjacent to } v}$

Assortativity, shared partners

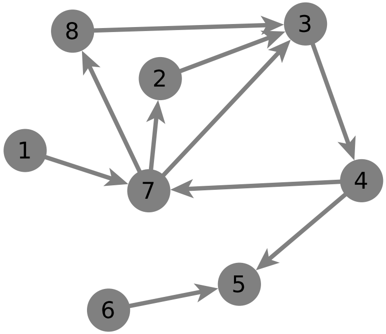
- Assortativity or assortative mixing refers to the tendency of vertices to be connected to other vertices with the same degree or attribute.



Is there a tendency for assortative mixing in this graph?

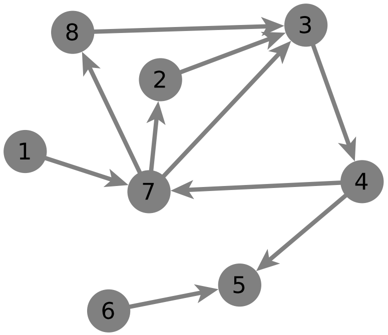
Assortativity, shared partners

- Assortativity or assortative mixing refers to the tendency of vertices to be connected to other vertices with the same degree or attribute.
- Edge-wise shared partners are indirect contacts (twopaths) in the same direction as the direct tie.
- Dyad-wise shared partners are like edge-wise shared partners but a direct tie is not necessary.



Assortativity, shared partners

- Assortativity or assortative mixing refers to the tendency of vertices to be connected to other vertices with the same degree or attribute.
- Edge-wise shared partners are indirect contacts (twopaths) in the same direction as the direct tie.
- Dyad-wise shared partners are like edge-wise shared partners but a direct tie is not necessary.



Find edge- and dyad-wise shared partners here!

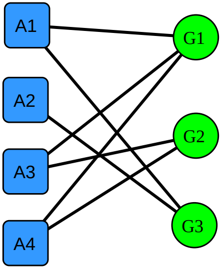
Symmetry, reciprocity, dichotomization

- There is reciprocity if an edge from vertex u to vertex v presupposes an edge from v to u.
- A network is symmetric if all edges are reciprocal. In a symmetric matrix, the upper triangle equals the transposed lower triangle.
- A binary relation is a set of edges that do not have weights.
- A weighted relation can be dichotomized if all weights above 0 are recoded as 1. A weighted relation can also be recoded by imposing a threshold value, e. g. all values above 5 are recoded as 1, all other edge weights as 0.

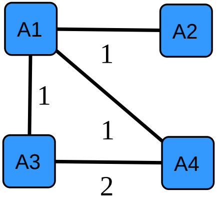
Co-occurrence graphs

also known as one-mode projections

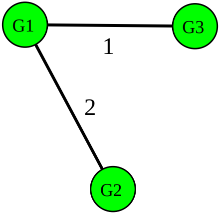
A bipartite graph



→ Actor co-occurrence graph



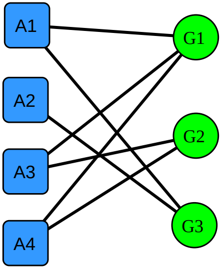
→ Group co-occurrence graph



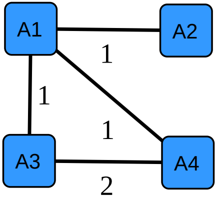
Co-occurrence graphs

also known as one-mode projections

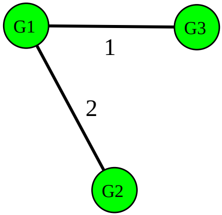
A bipartite graph



→ Actor co-occurrence graph



→ Group co-occurrence graph

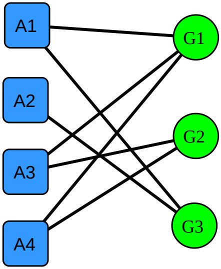


How can this be achieved?

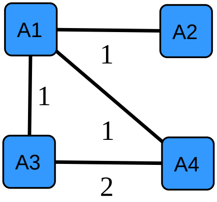
Co-occurrence graphs

also known as one-mode projections

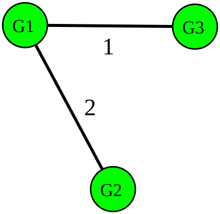
A bipartite graph



→ Actor co-occurrence graph



→ Group co-occurrence graph



How can this be achieved?
Using matrix transposition and matrix multiplication!

Transposing a matrix

The transpose of a matrix is obtained by taking the rows and using them as the columns of a new matrix.

Example: a two-mode network

original matrix:

	g_1	g_2	g_3
a_1	1	0	1
a_2	0	0	1
a_3	1	1	0
a_4	1	1	0

transposed matrix:

	a_1	a_2	a_3	a_4
g_1	1	0	1	1
g_2	0	0	1	1
g_3	1	1	0	0

The transpose of $\mathbf{X}$ is written as $\mathbf{X}^\top$ or $\mathbf{X}'$.

Matrix multiplication

- Example: $\mathbf{X}\mathbf{X}^\top = \mathbf{Z}$
- Multiplication works in a different way than the Hadamard product!
- usually $\mathbf{XY} \neq \mathbf{YX}$
- There is a simple trick called the Falk scheme.

The Falk scheme

			1	0	1	1
			0	0	1	1
			1	1	0	0
1	0	1	2	1	1	1
0	0	1	1	1	0	0
1	1	0	1	0	2	2
1	1	0	1	0	2	2

- For each cell of the new matrix, calculate the dot product of the corresponding row of the first matrix and the column of the second matrix, then add up the values.

Co-occurrence networks

- Why do we need matrix multiplication?
- Answer: For the conversion of two-mode networks into one-mode networks!
- Example: We have a set of actors connected to a set of groups.
- We want to create a network where two actors are connected if they are in the same group.
- Additionally, the edge weight should reflect the number of common groups between the two actors.
- This is called a co-occurrence network because the groups co-occur between the actors.
- Such a network can be obtained by computing $\mathbf{X}\mathbf{X}^\top$ (example on the previous slide!).

Co-occurrence networks (continued)

- At the same time, we can also create a network of groups.
- Two groups are connected if an actor is affiliated with both of them.
- The edge weight between two groups reflects the number of common actors.
- This network can be obtained by calculating $\mathbf{X}^\top\mathbf{X}$.

Example

$$\mathbf{X}^\top\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

What is centrality?

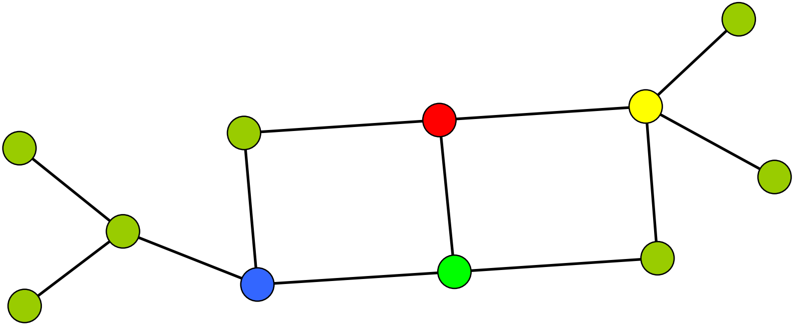


Figure: Example graph

Problem: Which is the most central vertex?

Example 1: Degree centrality

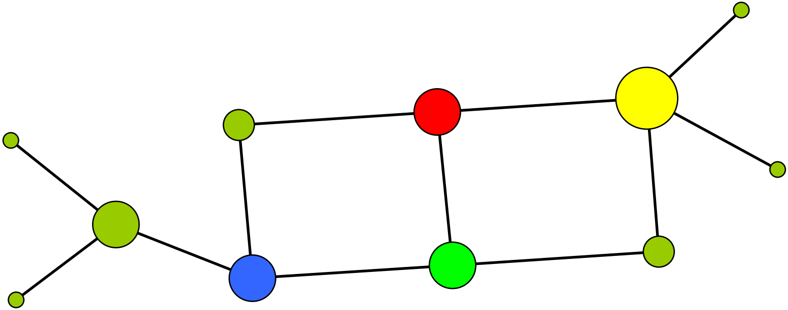


Figure: Degree centrality – the yellow vertex is most central!

Example 2: Betweenness centrality

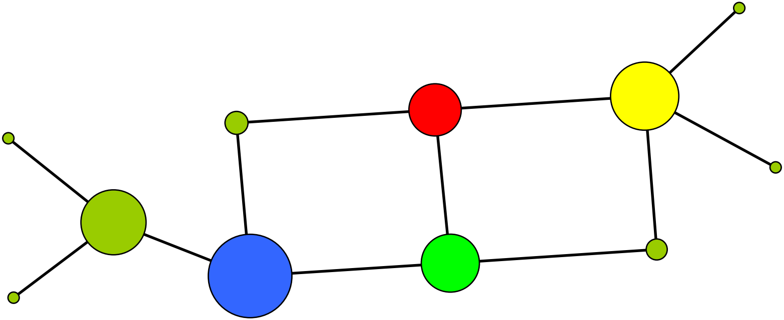


Figure: Betweenness centrality – blue is most central!

Example 3: Closeness centrality

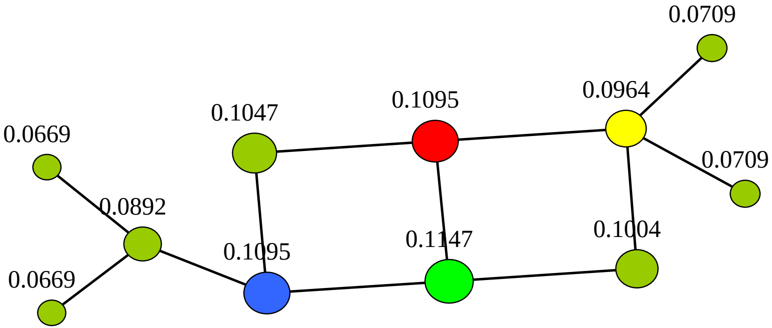
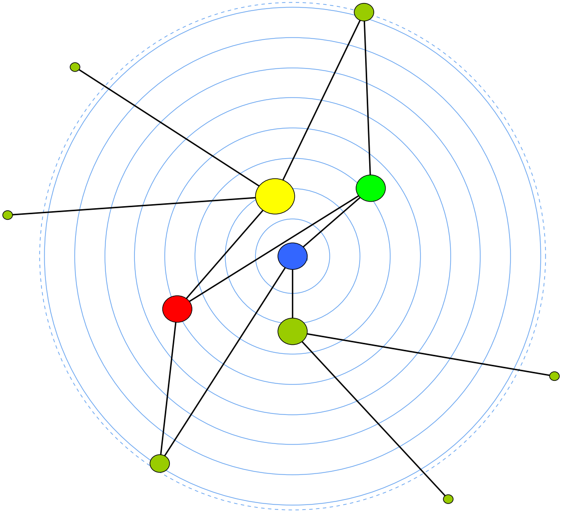


Figure: Closeness centrality – green is most central!

Radial layout (= centrality layout)

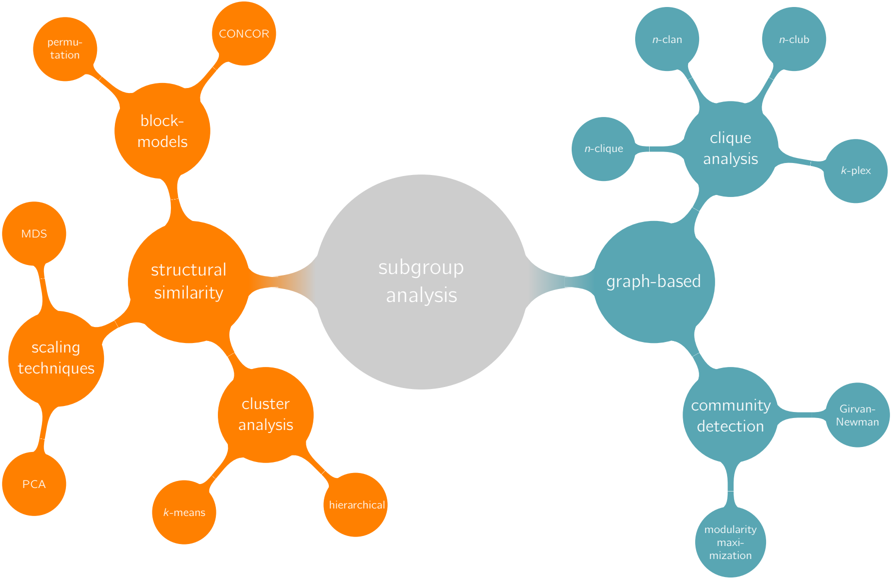
Betweenness (position) and degree (node size) in the same visualization



What is the meaning of centrality?

- Analysis on the level of vertices, not the overall network structure!
- "One of the primary uses of graph theory in social network analysis is the identification of the most important actors in a social network." (Wasserman/Faust 1994)
- But what does importance mean?
- Many different measures yield different types of importance!

Classification of methods for subgroup analysis
(not exhaustive)



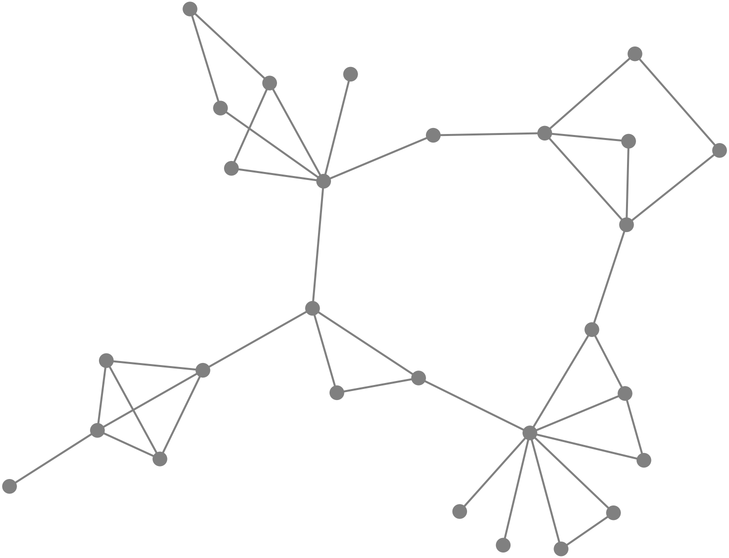
Community detection

Edge betweenness: on how many shortest paths between other edges is an edge located?

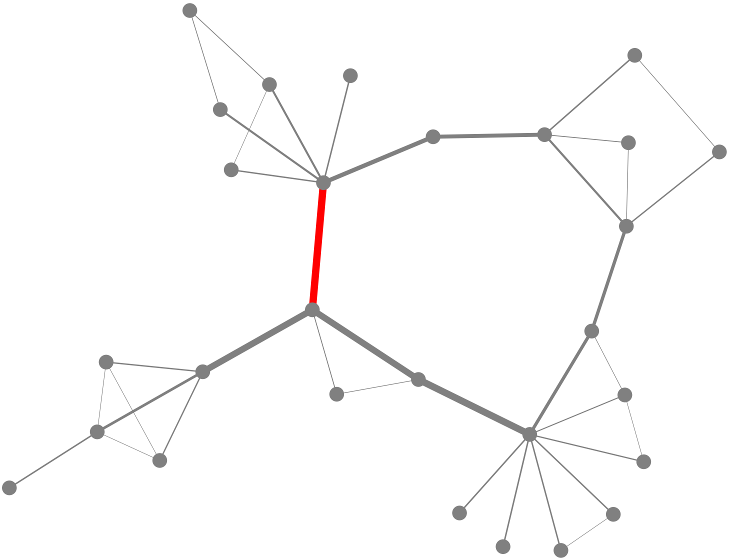
The Girvan-Newman algorithm

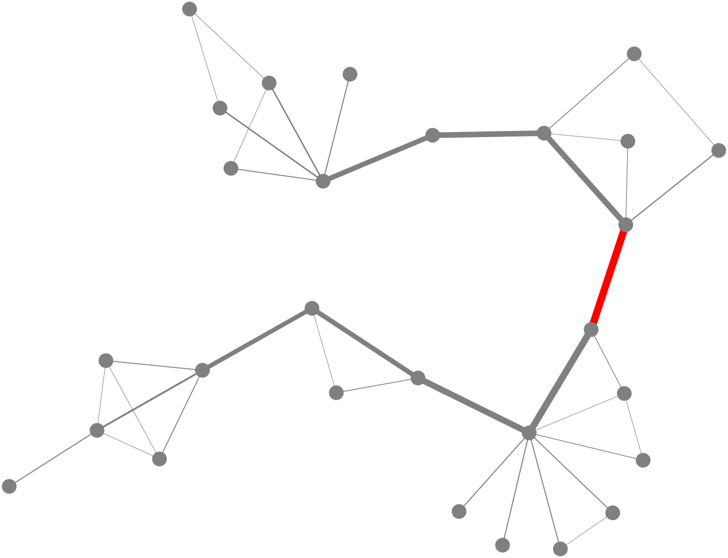
1. Calculate the betweenness for all edges in the network.
2. Remove the edge with the highest betweenness.
3. Recalculate betweennesses for all edges affected by the removal.
4. Repeat from step 2 until no edges remain.

Community detection example

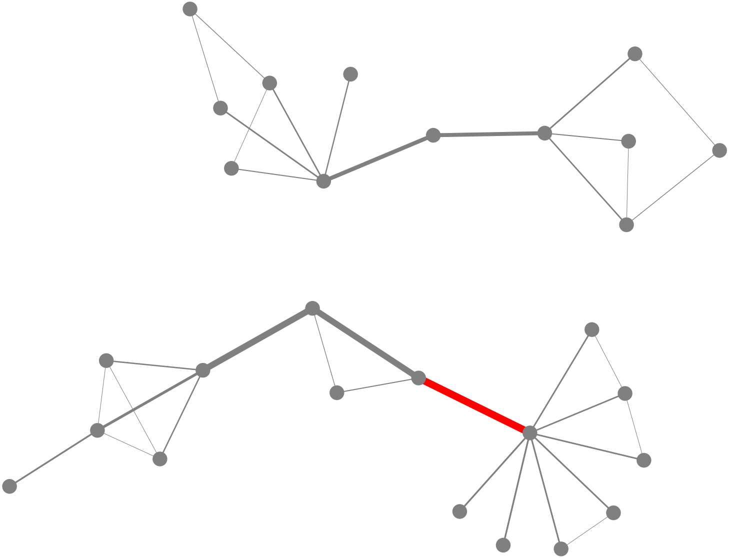


Community detection example

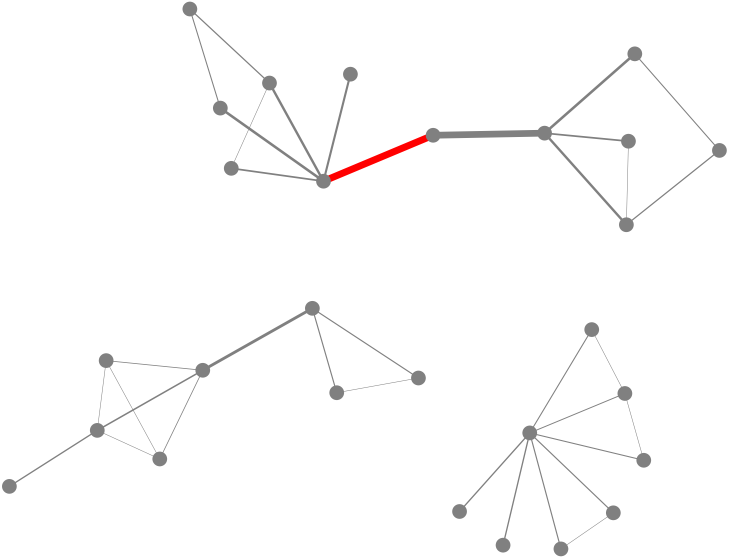




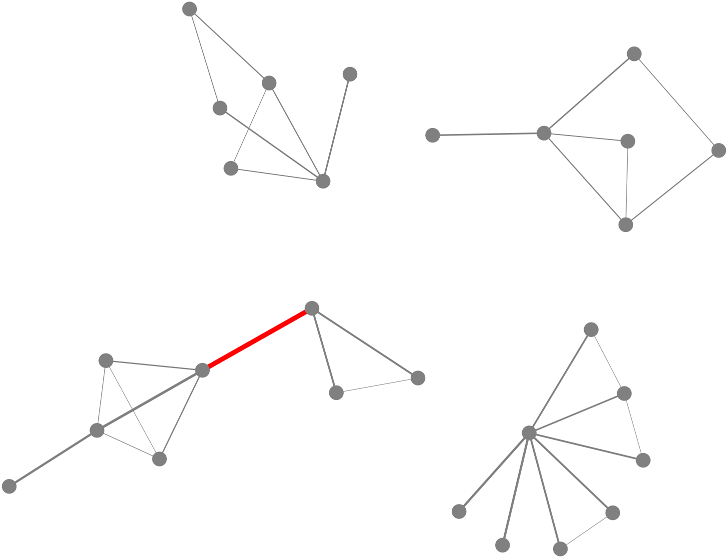
Community detection example



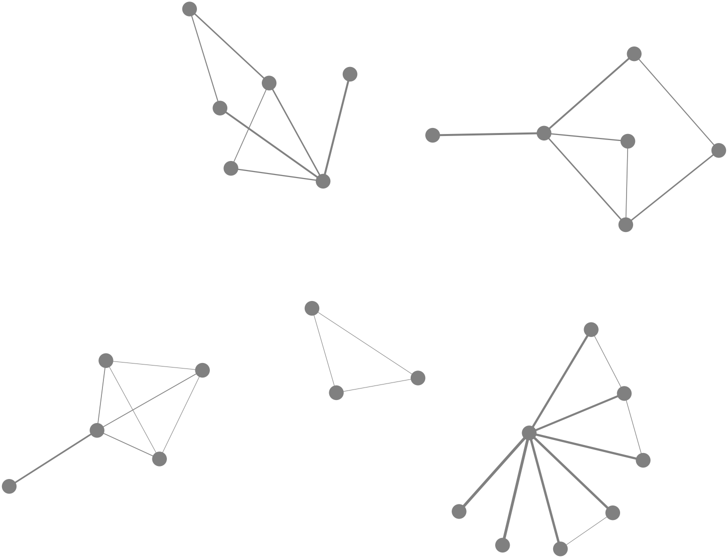
Community detection example



Community detection example



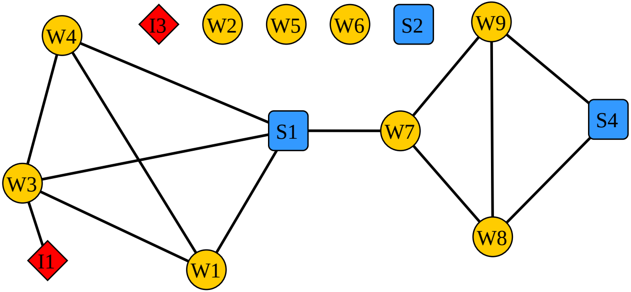
Community detection example



Structural similarity

- Structural similarity: Similarity of tie profiles.
- If two actors have edges to the same actors, they are structurally similar.
- The extreme form is structural equivalence, where two actors have exactly the same neighbors.

Structural similarity: example



W1 and W4 or W8 and W9 are structurally equivalent.

W1 and W3 are structurally rather similar.

W3 and W9 are structurally rather dissimilar.

Similarity and distance

- Similarity between two rows in a matrix can be understood as structural similarity.
- Standardized metrics take values between 0 and 1.
- Standardized similarities s and distances d are actually the same; they can be converted: $d_{ij} = 1 - s_{ij}$
- Dissimilarity measures: geodesic distance, Jaccard coefficient, Euclidean distance.
- Similarity measures: correlation, simple matching coefficient.
- The calculated distances can be saved in a distance matrix.
- Also possible for two-mode networks!

Jaccard coefficient

$d_{pq} = 1 - \frac{a}{a+b+c}$

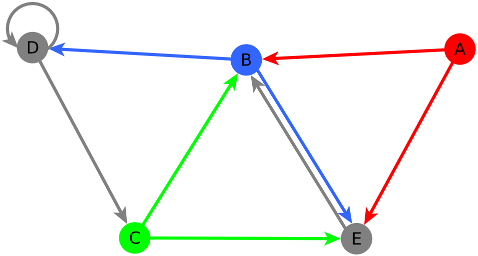
p: a row in the matrix
q: any other row in the matrix
a: number of columns where p and q are both 1
b: number of columns where p is 1 and q is 0
c: number of columns where q is 1 and p is 0

- This results in a quadratic distance matrix!
- Values between 0 and 1.
- Can be converted into similarities by computing $s_{pq} = 1 - d_{pq}$.

Example: Jaccard distances and structural similarity

Consider the following directed network:

	A	B	C	D	E
A	0	1	0	0	1
B	0	0	0	1	1
C	0	1	0	0	1
D	0	0	1	1	0
E	0	1	0	0	0



$d_{AB} = 1 - \frac{1}{1+1+1} = \frac{2}{3}$

for comparison:

$d_{AC} = 1 - \frac{2}{2+0+0} = 0$

Euclidean distance

$d_{pq} = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$

i: a column in the matrix.

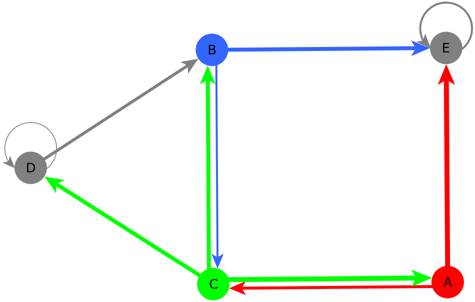
In words: add up the differences between all data points/columns for any two rows p and q.

- Again, this results in a quadratic distance matrix!
- Can take values greater than 1.
- Conversion into similarities: $s_{pq} = max(d) - max_{pq}$.
- Can also be applied to spatial coordinates instead of row profiles!

Example: Euclidean distances and structural similarity

Consider the following weighted network:

	A	B	C	D	E
A	0	0	3	0	5
B	0	0	2	0	4
C	5	4	0	4	0
D	0	3	0	1	0
E	0	0	0	0	2



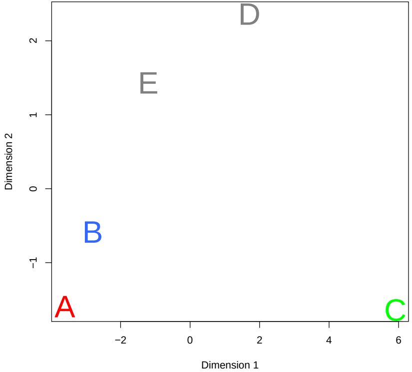
$$d_{AB} = \sqrt{(0-0)^2 + (0-0)^2 + (3-2)^2 + (0-0)^2 + (5-4)^2} = 1.41$$

for comparison:

$$d_{AC} = \sqrt{(0-5)^2 + (0-4)^2 + (3-0)^2 + (0-4)^2 + (5-0)^2} = 9.54$$

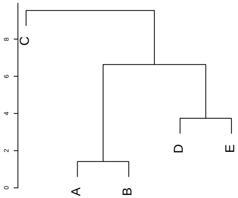
Multidimensional Scaling

- Goal: map the distances in two dimensions.
- Spatial interpretation of distances.
- A and B are close to each other $\rightarrow$ subgroup!
- Problem: higher-dimensional data.
- Approximation is necessary.



Hierarchical cluster analysis

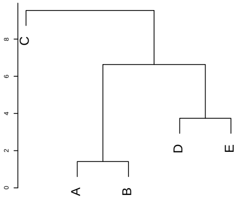
	A	B	C	D	E
A					
B	1.41				
C	9.54	8.77			
D	6.63	5.48	5.92		
E	4.24	2.83	7.81	3.74	



1. Which actors are most similar? Fusion of A and B!
2. Recalculation of the similarity matrix (here: complete linkage).
3. Fusion of D and E to DE and recalculation of distance matrix.
4. Fusion of DE and AB to ABDE.
5. Fusion of ABDE and C.

Hierarchical cluster analysis

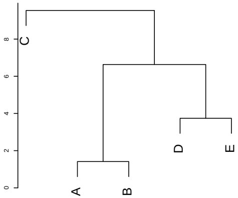
	A	B	C	D	E
A					
B	1.41				
C	9.54	8.77			
D	6.63	5.48	5.92		
E	4.24	2.83	7.81	3.74	



1. Which actors are most similar? Fusion of A and B!
2. Recalculation of the similarity matrix (here: complete linkage).
3. Fusion of D and E to DE and recalculation of distance matrix.
4. Fusion of DE and AB to ABDE.
5. Fusion of ABDE and C.

Hierarchical cluster analysis

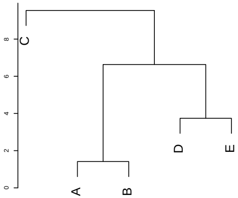
	A	B	C	D	E
A					
B	1.41				
C	9.54	~~8.77~~			
D	6.63	~~5.48~~	5.92		
E	4.24	~~2.83~~	7.81	3.74	



1. Which actors are most similar? Fusion of A and B!
2. Recalculation of the similarity matrix (here: complete linkage).
3. Fusion of D and E to DE and recalculation of distance matrix.
4. Fusion of DE and AB to ABDE.
5. Fusion of ABDE and C.

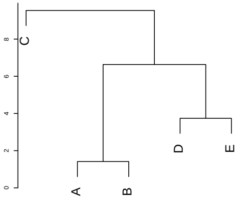
Hierarchical cluster analysis

	AB	C	D	E
AB				
C	9.54			
D	6.63	5.92		
E	4.24	7.81	3.74	



1. Which actors are most similar? Fusion of A and B!
2. Recalculation of the similarity matrix (here: complete linkage).
3. Fusion of D and E to DE and recalculation of distance matrix.
4. Fusion of DE and AB to ABDE.
5. Fusion of ABDE and C.

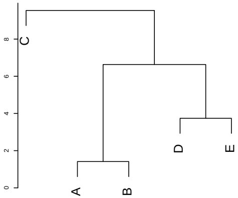
	AB	C	D	E
AB				
C	9.54			
D	6.63	5.92		
E	4.24	7.81	3.74	



1. Which actors are most similar? Fusion of A and B!
2. Recalculation of the similarity matrix (here: complete linkage).
3. Fusion of D and E to DE and recalculation of distance matrix.
4. Fusion of DE and AB to ABDE.
5. Fusion of ABDE and C.

Hierarchical cluster analysis

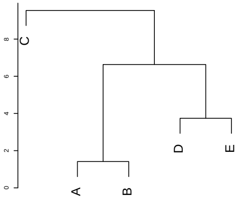
	AB	C	DE
AB			
C	9.54		
DE	6.63	7.81	



1. Which actors are most similar? Fusion of A and B!
2. Recalculation of the similarity matrix (here: complete linkage).
3. Fusion of D and E to DE and recalculation of distance matrix.
4. Fusion of DE and AB to ABDE.
5. Fusion of ABDE and C.

Hierarchical cluster analysis

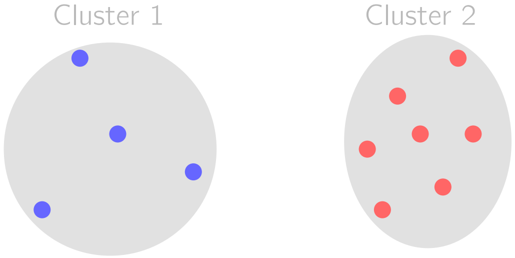
	ABDE	C
ABDE		
C	9.54	



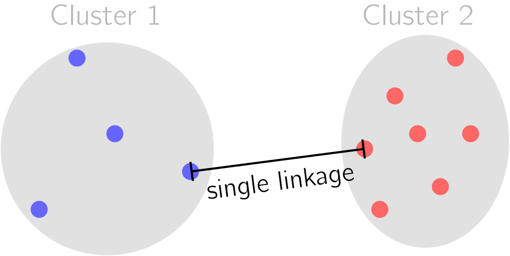
1. Which actors are most similar? Fusion of A and B!
2. Recalculation of the similarity matrix (here: complete linkage).
3. Fusion of D and E to DE and recalculation of distance matrix.
4. Fusion of DE and AB to ABDE.
5. Fusion of ABDE and C.

How are similarities recalculated?

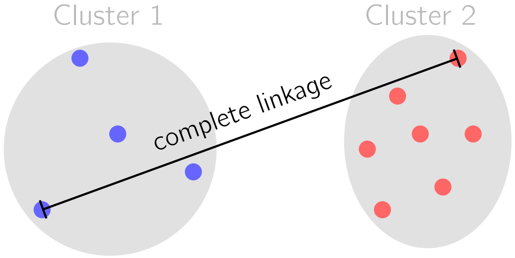
Assume for a moment that similarities can be mapped on a plane.



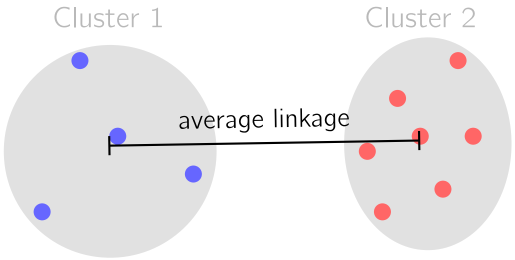
How are similarities recalculated?



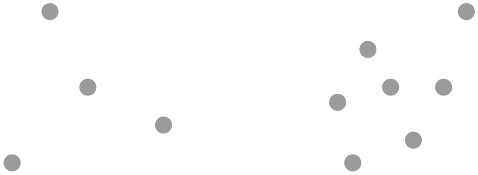
How are similarities recalculated?



How are similarities recalculated?

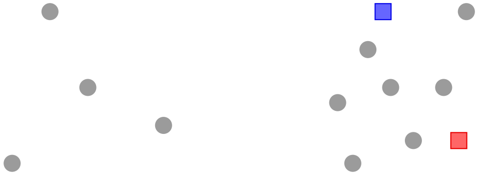


k-means cluster analysis



Assume again that similarities can be mapped on a plane.

k-means cluster analysis

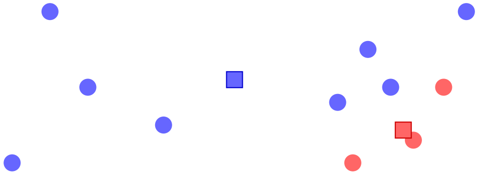


Step 1: add k nodes ("centers") at random coordinates.

k-means cluster analysis

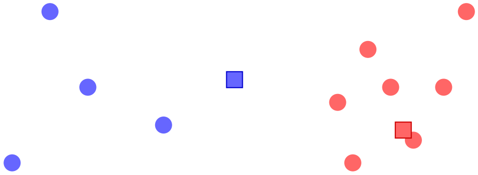
Step 2: classify other nodes according to their distance to the centers.

k-means cluster analysis



Step 3: move the centers to the center of each cluster.

k-means cluster analysis



Step 4: re-classify nodes according to their new distances.

k-means cluster analysis

Step 5: re-move the centers to the center of each cluster.

k-means cluster analysis

Repeat steps 4 and 5 until stable.

Some useful concepts for inferential network modeling

- Topology; structure.
- Exogenous covariate; attribute; exogenous relation.
- Endogeneity; endogenous process; network statistic.
- Data-generating process (DGP).
- Observation.
- Deterministic vs. stochastic processes.
- Local interaction.
- Emergence.
- Parametric model.
- Estimation.

The exponential random graph model

$$P(N, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}^\top \mathbf{h}(N^*)\}}$$

- Probability density function of the cross-sectional ERGM.

The exponential random graph model

$$P(N, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}^\top \mathbf{h}(N^*)\}}$$

- Probability that we observe this particular network.

The exponential random graph model

$$P(N, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}^\top \mathbf{h}(N^*)\}}$$

- $\mathbf{h}(N)$ are network statistics.

The exponential random graph model

$$P(N, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}^\top \mathbf{h}(N^*)\}}$$

- Coefficients (to be estimated).

The exponential random graph model

$$P(N, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}^\top \mathbf{h}(N^*)\}}$$

- Exponential function of the sum of the weighted statistics.

The exponential random graph model

$$P(N, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}^\top \mathbf{h}(N^*)\}}$$

- The sum of the same for all possible topologies.

The exponential random graph model

$$P(N, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}^\top \mathbf{h}(N^*)\}}$$

- Probability of a given network over all networks one could have observed.

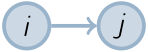
The exponential random graph model

$$P(N, \boldsymbol{\theta}) = \frac{\exp\{\boldsymbol{\theta}^\top \mathbf{h}(N)\}}{\sum_{N^* \in \mathcal{N}} \exp\{\boldsymbol{\theta}^\top \mathbf{h}(N^*)\}}$$

- Task: define $\mathbf{h}(N)$ in order to operationalize theory.

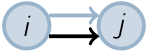
Number of edges

$$h_{\text{edges}} = \sum_{i \neq j} N_{ij}$$



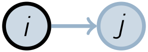
Dyadic covariate

$$h_{\text{edgecov}} = \sum_{i \neq j} N_{ij} X_{ij}$$

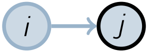


Covariates for sender and receiver

$$h_{\text{nodeocov}} = \sum_{i \neq j} N_{ij} x_i$$

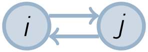


$$h_{\text{nodeicov}} = \sum_{i \neq j} N_{ij} x_j$$



Reciprocity

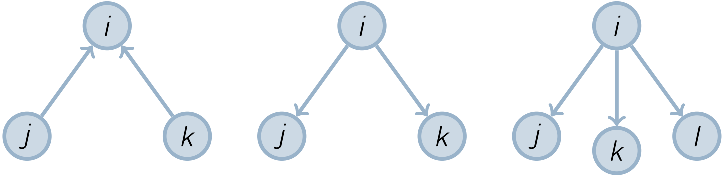
$$h_{\text{reciprocity}} = \sum_{i \neq j} N_{ij} N_{ji}$$



Two-stars and three-stars

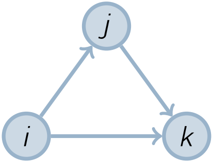
$$h_{\text{in-two-star}} = \sum_{i,j,k} N_{ji} N_{ki} (1 - N_{jk})(1 - N_{kj})$$

Incoming 2-star   Outgoing 2-star   Outgoing 3-star

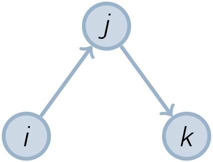


Edge-wise shared partners and two-paths

$$h_{\text{esp}} = \sum_{i,j,k} N_{ij} N_{jk} N_{ik}$$

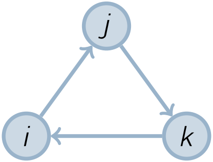


$$h_{\text{twopath}} = \sum_{i \notin \{j;k\}} \sum_{j \notin \{i;k\}} \sum_{k \notin \{i;j\}} N_{ij} N_{jk} (1 - N_{ik})(1 - N_{ki})$$

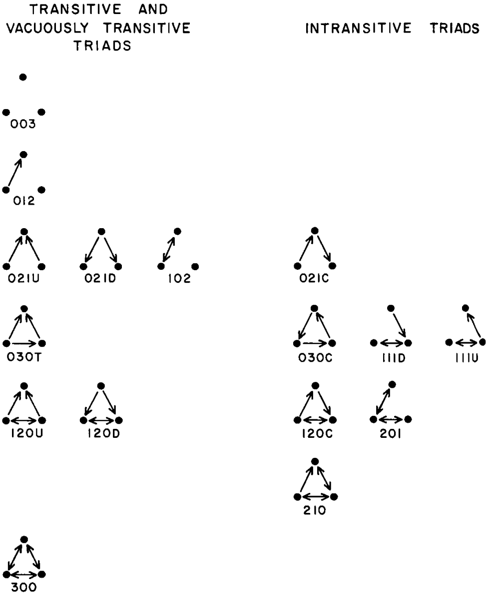


Three-cycles

$$h_{\text{three-cycle}} = \sum_{i,j,k} N_{ij} N_{jk} N_{ki}$$

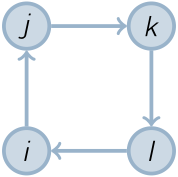


Triad census (Holland and Leinhardt 1971)



Four-cycles

$$h_{\text{four-cycle}} = \sum_{i,j,k,l} N_{ij} N_{jk} N_{kl} N_{li} (1 - N_{ik})(1 - N_{jl})(1 - N_{ki})(1 - N_{lj})$$

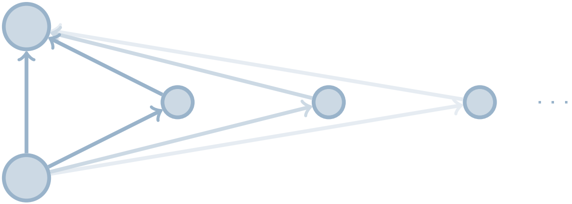


GWESP

Geometrically weighted edge-wise shared partners

$$h_{\mathrm{GWESP}}(N, \alpha) = \mathrm{e}^{\alpha} \sum_{i=1}^{n-2} \left\{1 - \left(1 - \mathrm{e}^{-\alpha}\right)^{i}\right\} \mathrm{ESP}_i(N)$$

where $\mathrm{ESP}_i(N)$ is the number of edges with i shared partners.



ERGM theory building example 1

How can we model visits among inhabitants of a residential care home?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory building example 1

How can we model visits among inhabitants of a residential care home?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory building example 1

How can we model visits among inhabitants of a residential care home?

1. Dyadic covariates.
   - Age difference ($-$).
   - Same gender ($+$).
   - Proximity of apartments ($+$).
   - Similar size of visible families ($+$).
   - Similar profile of medical problems and disabilities ($+$).
   - Apartment of alter is between ego's apartment and the restaurant ($+$).
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory building example 1

How can we model visits among inhabitants of a residential care home?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory building example 1

How can we model visits among inhabitants of a residential care home?

1. Dyadic covariates.
2. Node covariates of ego.
   - Physical and mental fitness $(+)$.
   - Encouragement by family members $(+)$.
   - Owns a TV set $(-)$.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory building example 1

How can we model visits among inhabitants of a residential care home?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory building example 1

How can we model visits among inhabitants of a residential care home?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
   - Spacious balcony (+).
   - Pension level (+).
   - Altruism (+).
   - Physical and mental fitness (+).
   - Apartment is close to the restaurant (+).
4. Endogenous graph statistics.

ERGM theory building example 1

How can we model visits among inhabitants of a residential care home?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory building example 1

How can we model visits among inhabitants of a residential care home?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.
   - Reciprocity.
   - Edge-wise shared partners.
   - Cyclic triads.
   - k-in-stars.
   - k-out-stars.

ERGM theory-building example 2

How can we explain militarized interstate disputes?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory-building example 2

How can we explain militarized interstate disputes?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory-building example 2

How can we explain militarized interstate disputes?

1. Dyadic covariates.
   - Direct contiguity (+).
   - Colonial contiguity (−).
   - Distance (−).
   - Both countries are democracies (−).
   - Military capability ratio (−).
   - Trade dependence (−).
   - Bilateral alliances (−).
   - Joint membership in international organizations (−).
   - Shared allies (−).
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory-building example 2

How can we explain militarized interstate disputes?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory-building example 2

How can we explain militarized interstate disputes?

1. Dyadic covariates.
2. Node covariates of ego.
   - Democracy? No...
   - GDP per capita? No...
   - Demography; share of young men? Maybe...
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory-building example 2

How can we explain militarized interstate disputes?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory-building example 2

How can we explain militarized interstate disputes?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
   - Democracy $(-)$.
   - GDP per capita $(-)$.
   - Natural resources?
   - Has nuclear arms $(-)$.
4. Endogenous graph statistics.

ERGM theory-building example 2

How can we explain militarized interstate disputes?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.

ERGM theory-building example 2

How can we explain militarized interstate disputes?

1. Dyadic covariates.
2. Node covariates of ego.
3. Node covariates of alter.
4. Endogenous graph statistics.
   - Reciprocity $(+)$.
   - Structural balance: closed triangles $(-)$.
   - Structural balance: 4-cycles $(+)$.
   - Structural balance: edge-wise shared partners $(-)$.
   - k-in-stars $(+)$.
   - k-out-stars $(+)$.

ERGM results: the desired output

Leifeld and Schneider (2012), AJPS

	Political inf.ex.	Technical inf.ex.
Edges	-3.63 (0.19)***	-5.86 (0.31)***
Preference similarity	0.07 (0.07)	-0.05 (0.11)
Interest group homophily	1.18 (0.12)***	1.01 (0.32)**
Governmental alter	0.53 (0.06)***	0.41 (0.07)***
Scientific ego	0.05 (0.09)	1.51 (0.10)***
Common committees	0.31 (0.01)***	0.16 (0.01)***
Scientific communication	3.12 (0.38)***	
Political communication		2.75 (0.06)***
Influence attribution	0.84 (0.07)***	0.47 (0.07)***
GWESP: edge-wise shared p.	1.26 (0.03)***	0.43 (0.04)***
GWDSP: dyadic shared p.	-0.15 (0.02)***	-0.23 (0.02)***
Reciprocity	0.82 (0.06)***	1.86 (0.15)***

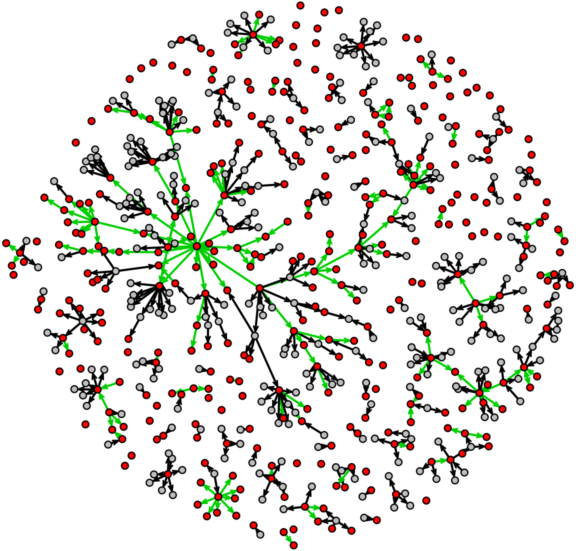
Case study: Nominations in an epistemic community

Leifeld/Fisher (2017), Nature Climate Change 7(10)

- "Millennium Ecosystem Assessment" (2002–2005)
- International scientific assessment.
- Membership recruitment by individual nomination.
- Research question: How do these nominations work?
- By merit/functional requirements or personal affinity?
- 1,360 experts in this policy-relevant network.

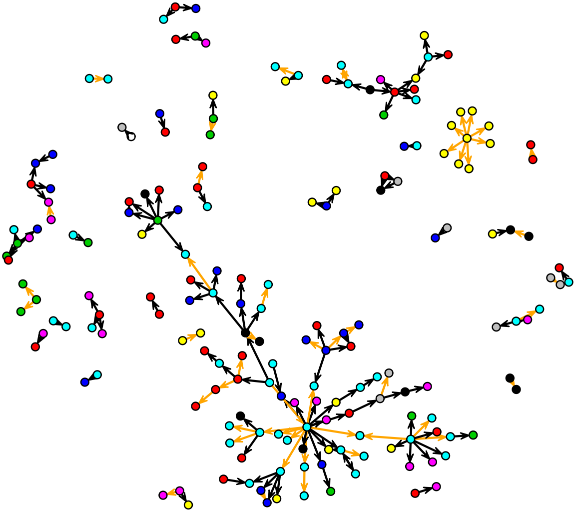
Nominations among members

Red: survey respondents; green: nominations among respondents



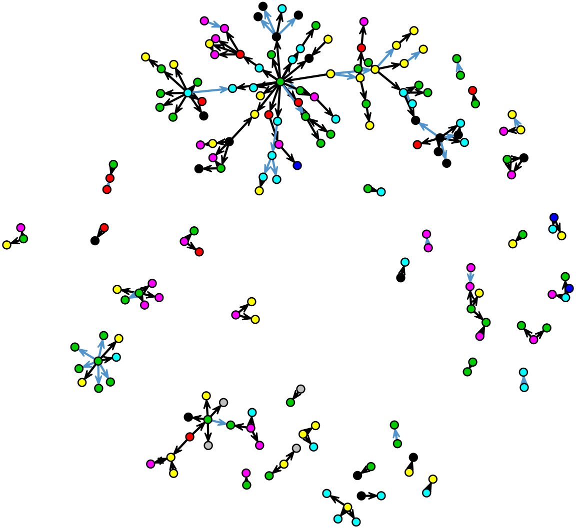
Nominations among survey respondents

Node colors: nationalities; orange: same nationality; no isolates



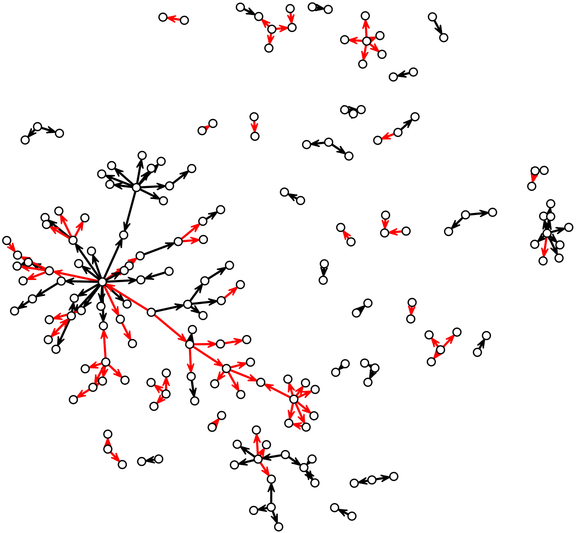
Nominations among survey respondents

Node colors: disciplines; blue: same same discipline; no isolates



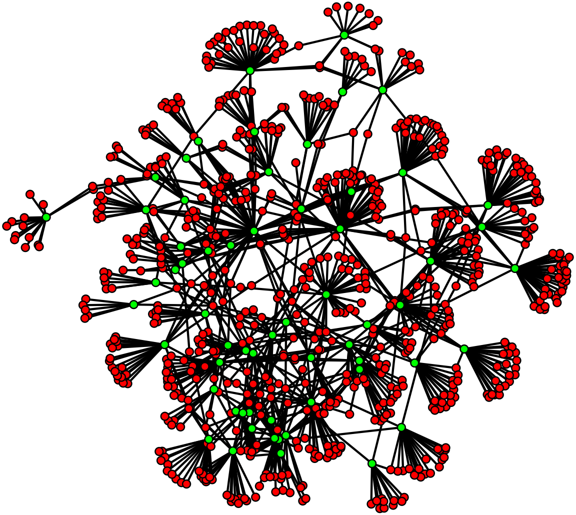
Nominations among survey respondents

Red: co-authorship in the final assessment report



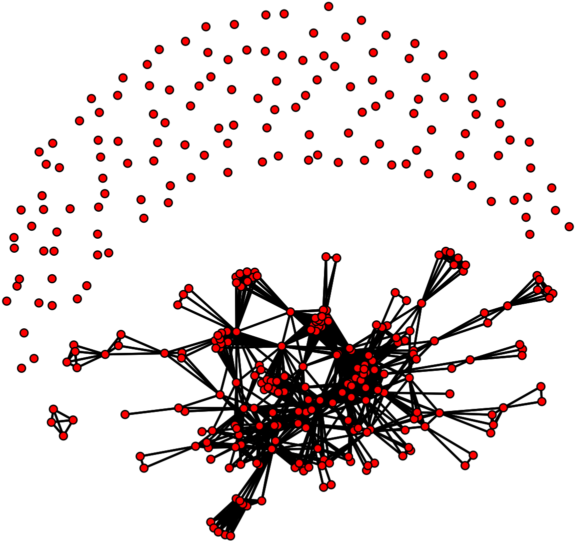
Collaboration on the assessment report

Red: authors; green: chapters; two-mode network

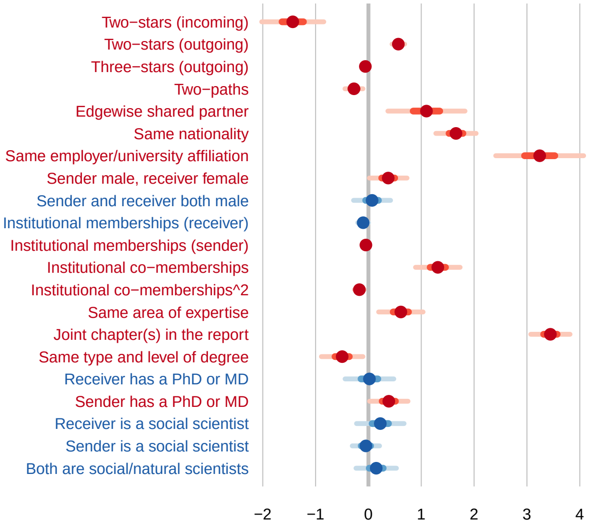


Collaboration on the assessment report

One-mode projection for all survey respondents

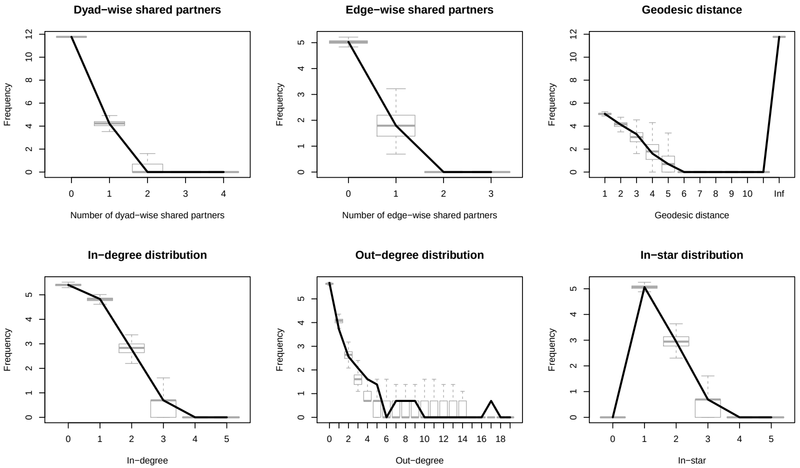


ERGM coefficients and confidence intervals

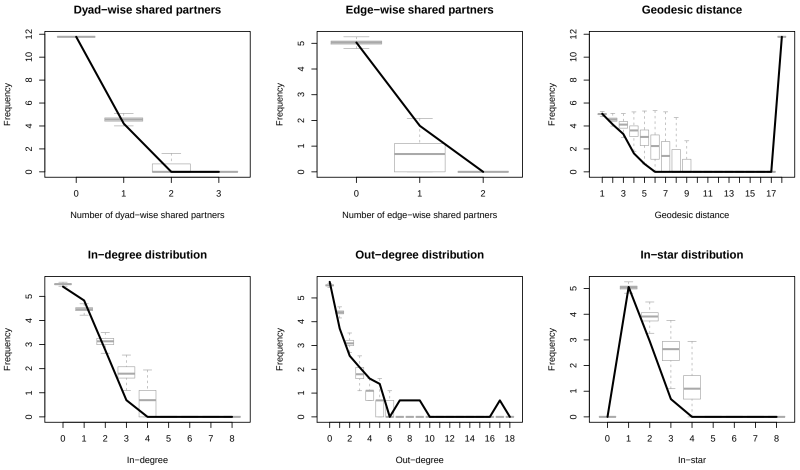


Horizontal bars denote 95% confidence intervals.

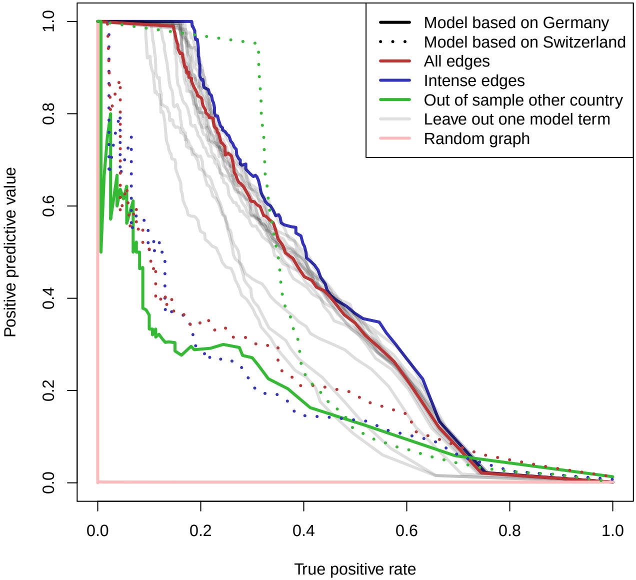
GOF: full model



GOF: model without endogenous processes



Precision–recall curves and out-of-sample prediction



Other Inferential Network Models

- Exponential Random Graph Model (ERGM).
- Temporal Exponential Random Graph Model (TERGM).
- Generalized Exponential Random Graph Model (GERGM).
- Count-ERGM.
- Multiplex/multilayer/multi-level ERGM.
- Quadratic Assignment Procedure.
- Latent Space Models.
- Stochastic Actor-Oriented Model (SAOM).
- Relational Event Model (REM).
- (Temporal) Network Autocorrelation Model (TNAM).

Group Work

Think of research questions and designs suitable for network analysis. Consider the following guiding questions.

1. What are the nodes? Are there one or two types of nodes?
2. What relations are you interested in? Are they binary?
3. Is there one cross-sectional network, panel data, or a relational event sequence?
4. How are you going to collect the data?
5. Do you want to explain the network structure? What theories or covariates are there?
6. Does the network structure explain something else?
7. Do you want to explain the attributes of nodes?
8. What is the added value of the network perspective?