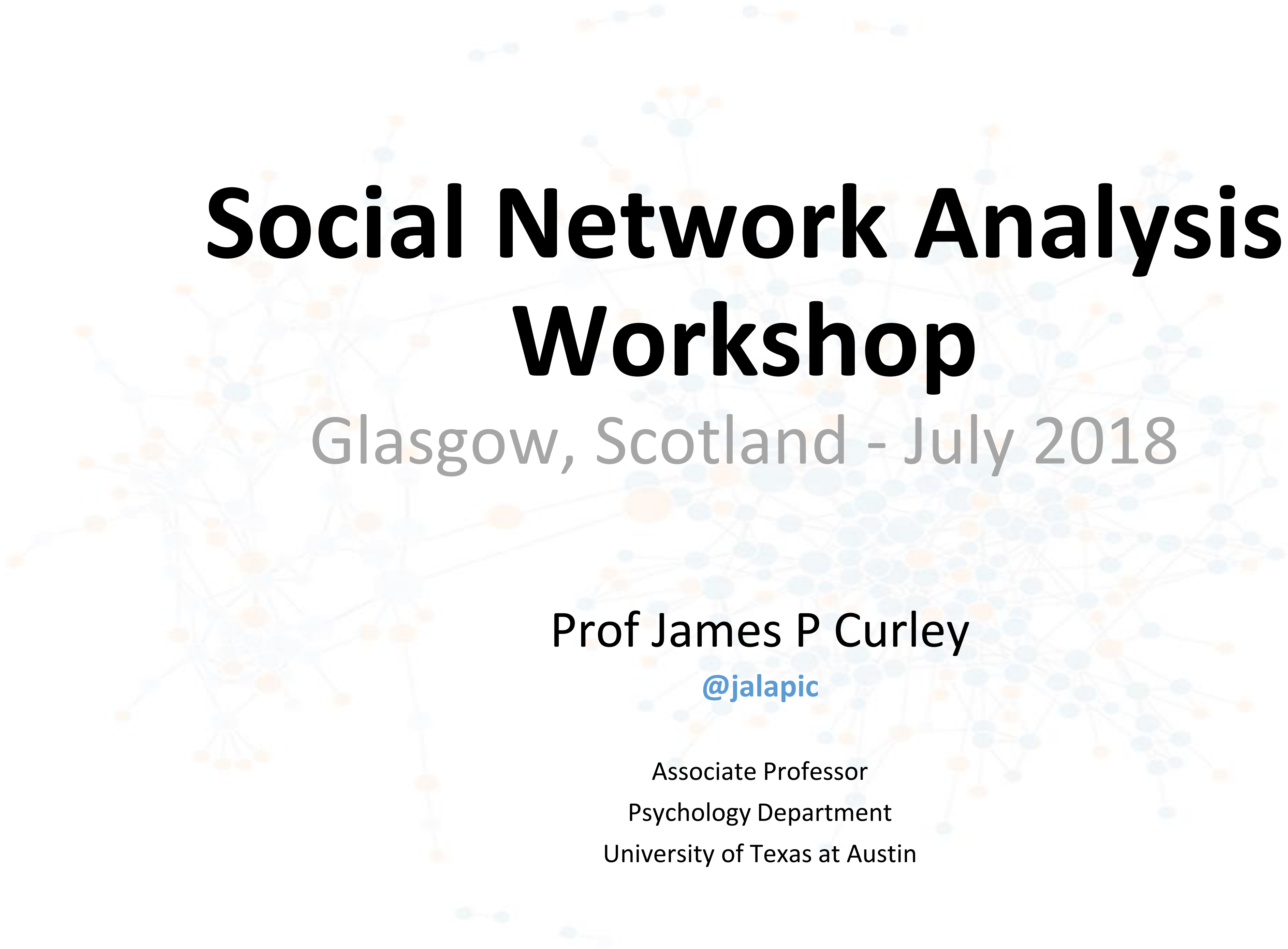


Social Network Analysis Workshop

Glasgow, Scotland - July 2018

Prof James P Curley

@jalapic

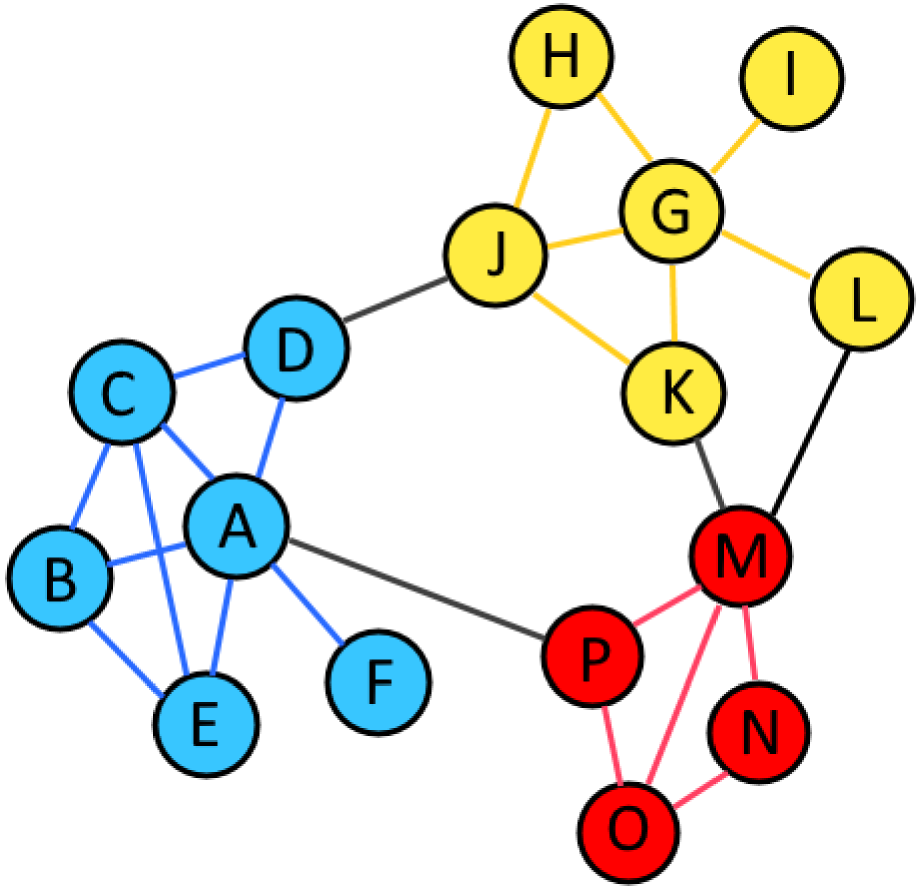
Associate Professor

Psychology Department

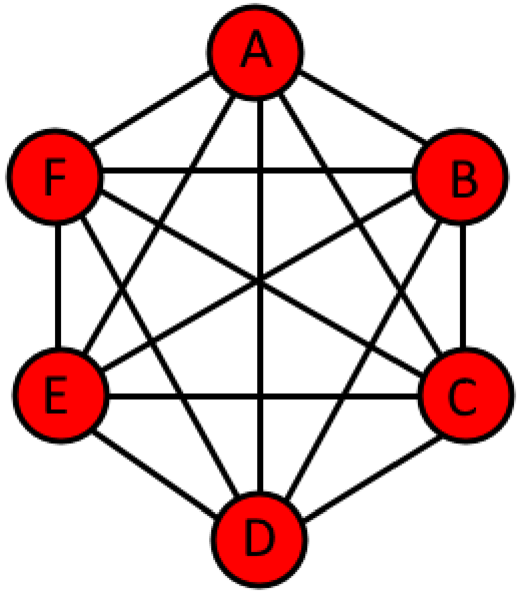
University of Texas at Austin

Subgroups

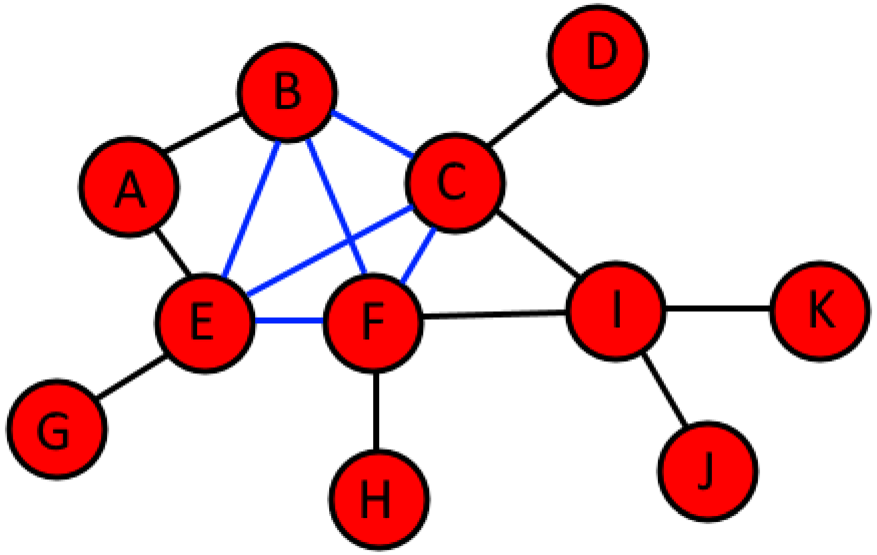
Subgroups



Cliques

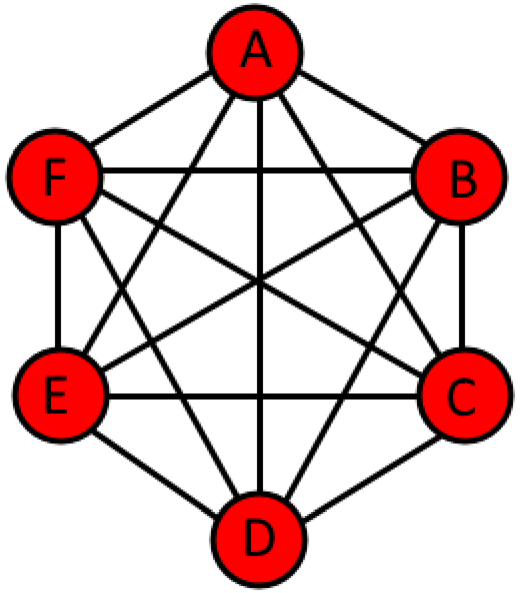


Cliques are maximally connected components – all nodes in a clique share an edge with all other nodes in the clique

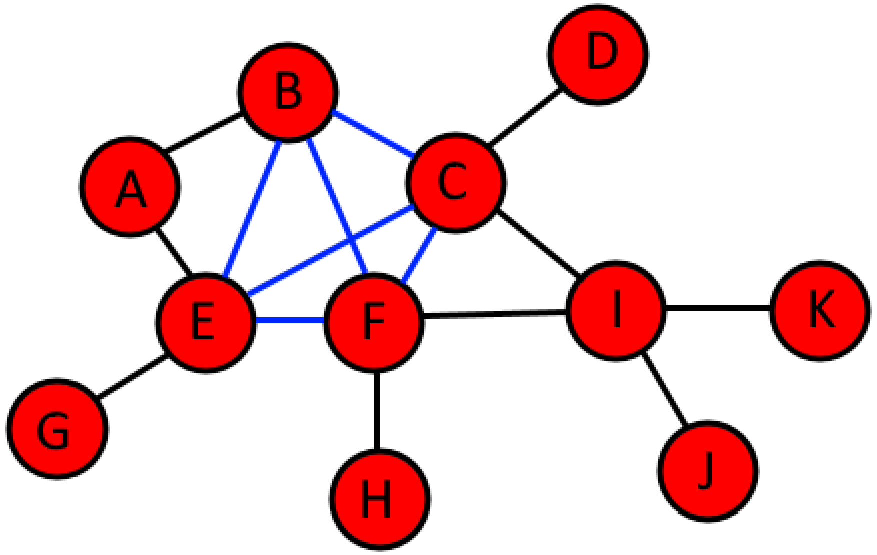


[B,C,E,F] is the largest clique in this network

Independent Vertex Sets



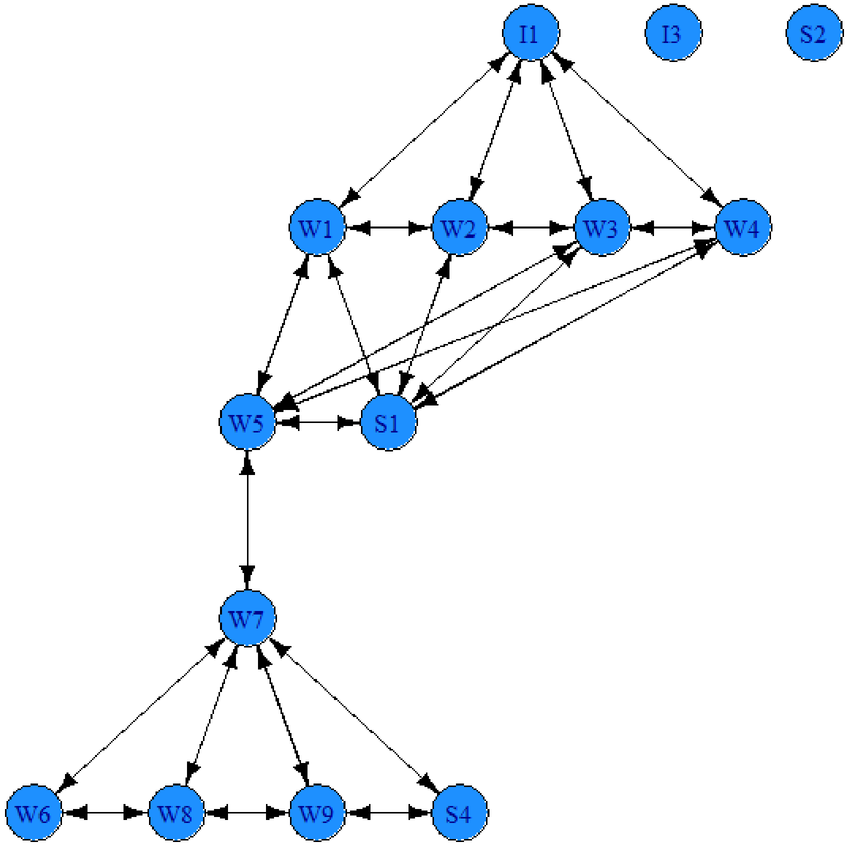
This graph has 0 IVS



The largest IVS for this graph = 6, &
there are four of this size:

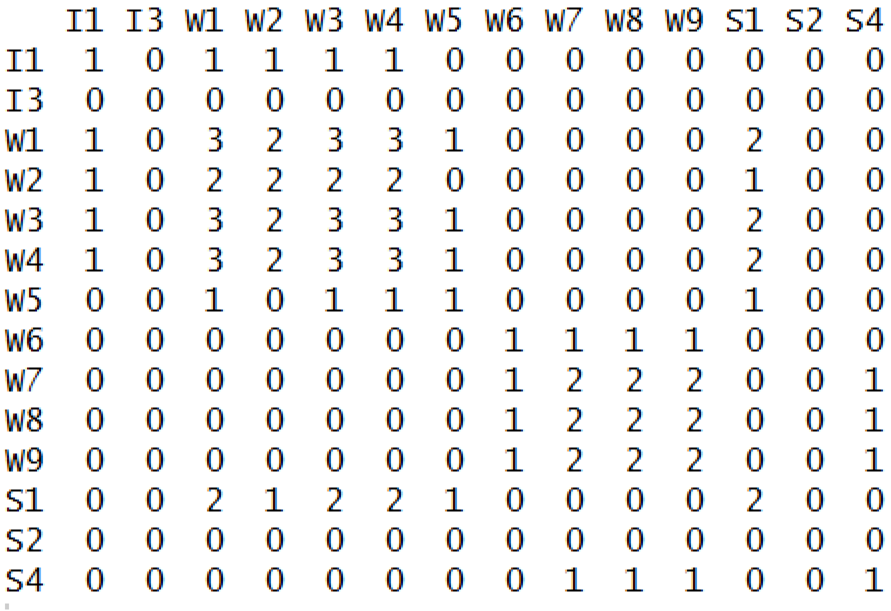
[A,D,F,G,J,K]
[A,D,G,H,J,K]
[A,C,G,H,J,K]
[B,D,G,H,J,K]

Clique Overlap



Roethlisberger & Dickson 1939
Bank wiring games' room data.

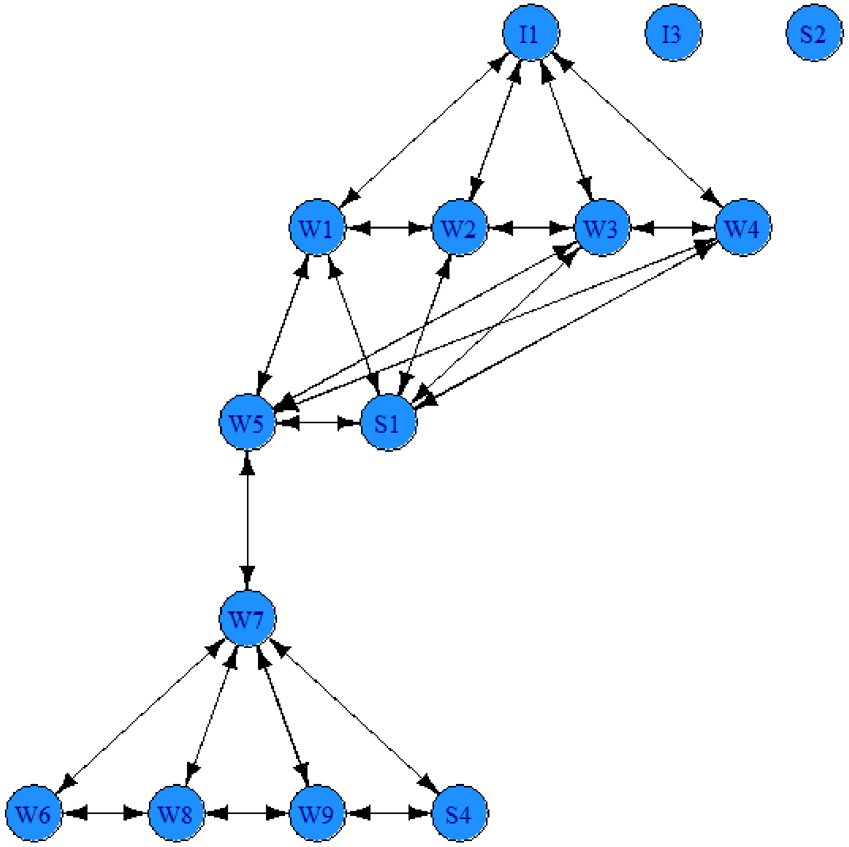
Clique co-membership matrix (min clique size = 4)



```
    I1 I3 W1 W2 W3 W4 W5 W6 W7 W8 W9 S1 S2 S4
I1   1  0  1  1  1  1  0  0  0  0  0  0  0  0
I3   0  0  0  0  0  0  0  0  0  0  0  0  0  0
W1   1  0  3  2  3  3  1  0  0  0  0  2  0  0
W2   1  0  2  2  2  2  0  0  0  0  0  1  0  0
W3   1  0  3  2  3  3  1  0  0  0  0  2  0  0
W4   1  0  3  2  3  3  1  0  0  0  0  2  0  0
W5   0  0  1  0  1  1  1  0  0  0  0  1  0  0
W6   0  0  0  0  0  0  0  1  1  1  1  0  0  0
W7   0  0  0  0  0  0  0  1  2  2  2  0  0  1
W8   0  0  0  0  0  0  0  1  2  2  2  0  0  1
W9   0  0  0  0  0  0  0  1  2  2  2  0  0  1
S1   0  0  2  1  2  2  1  0  0  0  0  2  0  0
S2   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  1
```

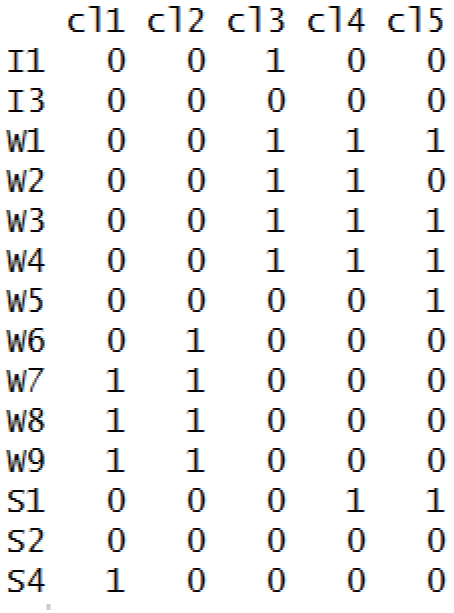
The clique co-membership matrix can be used to identify non-overlapping clusters of cliques using e.g. hierarchical clustering

Clique Overlap



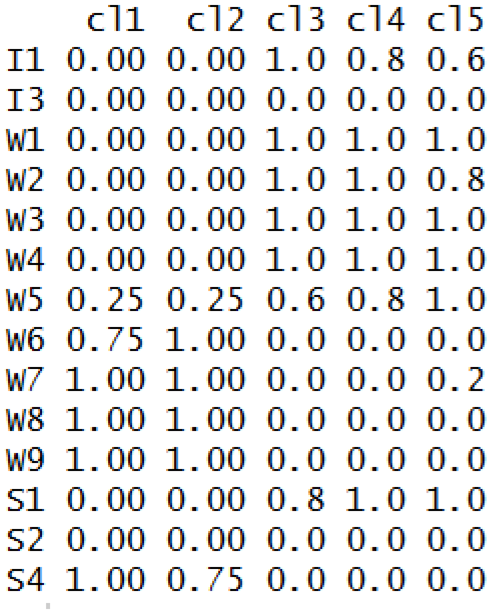
Use correspondence analysis on node by cluster matrix to identify relationships

Clique membership

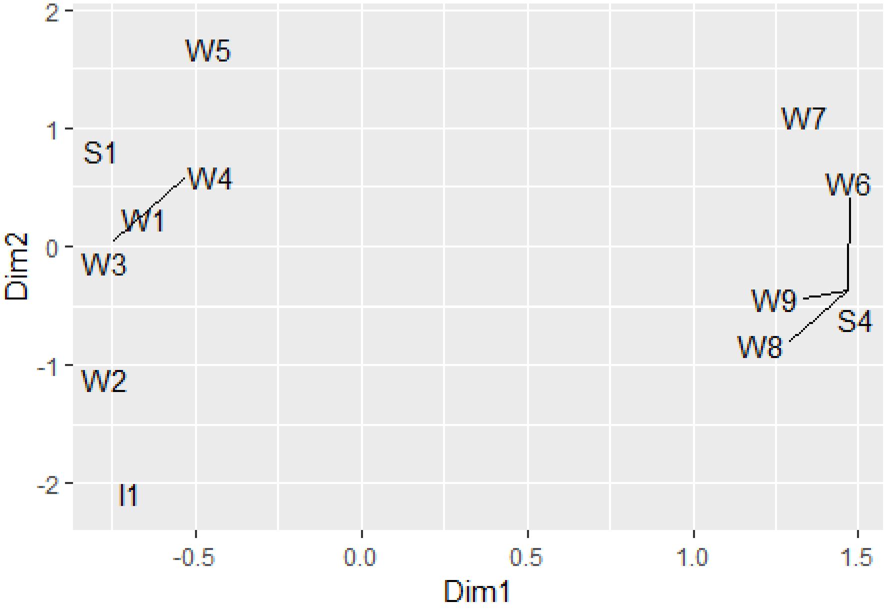


```
   cl1 cl2 cl3 cl4 cl5
I1   0   0   1   0   0
I3   0   0   0   0   0
W1   0   0   1   1   1
W2   0   0   1   1   0
W3   0   0   1   1   1
W4   0   0   1   1   1
W5   0   0   0   0   1
W6   0   1   0   0   0
W7   1   1   0   0   0
W8   1   1   0   0   0
W9   1   1   0   0   0
S1   0   0   0   1   1
S2   0   0   0   0   0
S4   1   0   0   0   0
```

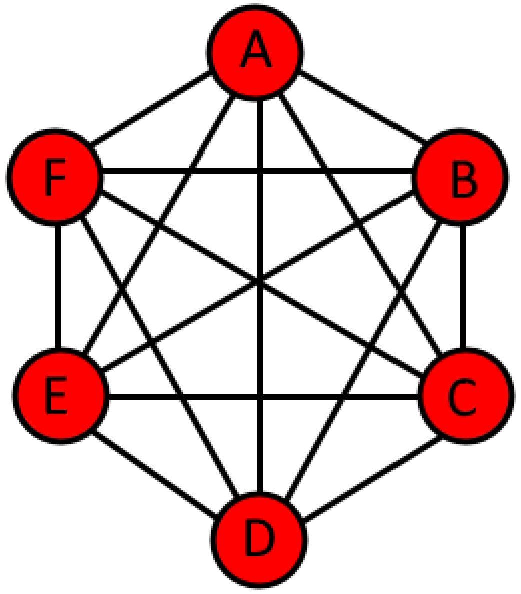
Proportion of ties to clique

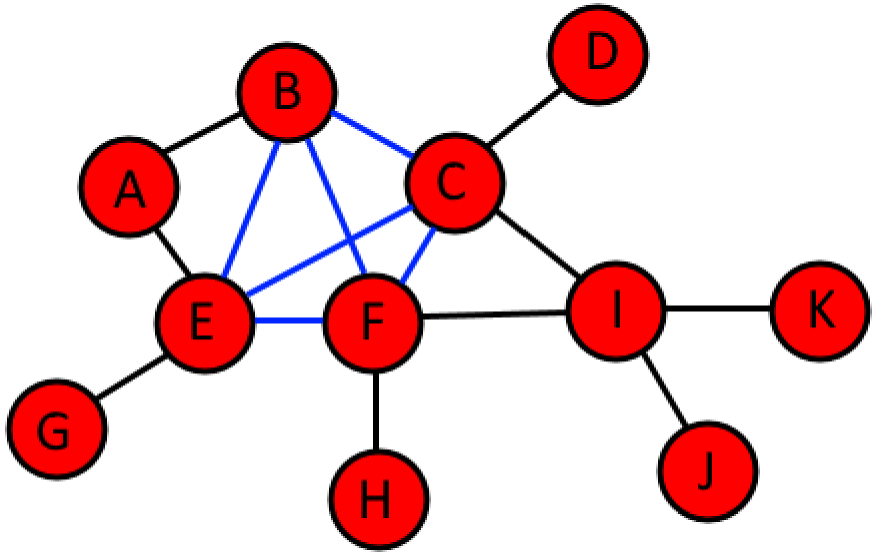


```
    cl1  cl2 cl3 cl4 cl5
I1 0.00 0.00 1.0 0.8 0.6
I3 0.00 0.00 0.0 0.0 0.0
W1 0.00 0.00 1.0 1.0 1.0
W2 0.00 0.00 1.0 1.0 0.8
W3 0.00 0.00 1.0 1.0 1.0
W4 0.00 0.00 1.0 1.0 1.0
W5 0.25 0.25 0.6 0.8 1.0
W6 0.75 1.00 0.0 0.0 0.0
W7 1.00 1.00 0.0 0.0 0.2
W8 1.00 1.00 0.0 0.0 0.0
W9 1.00 1.00 0.0 0.0 0.0
S1 0.00 0.00 0.8 1.0 1.0
S2 0.00 0.00 0.0 0.0 0.0
S4 1.00 0.75 0.0 0.0 0.0
```



Other ways of thinking about ‘maximally connected’





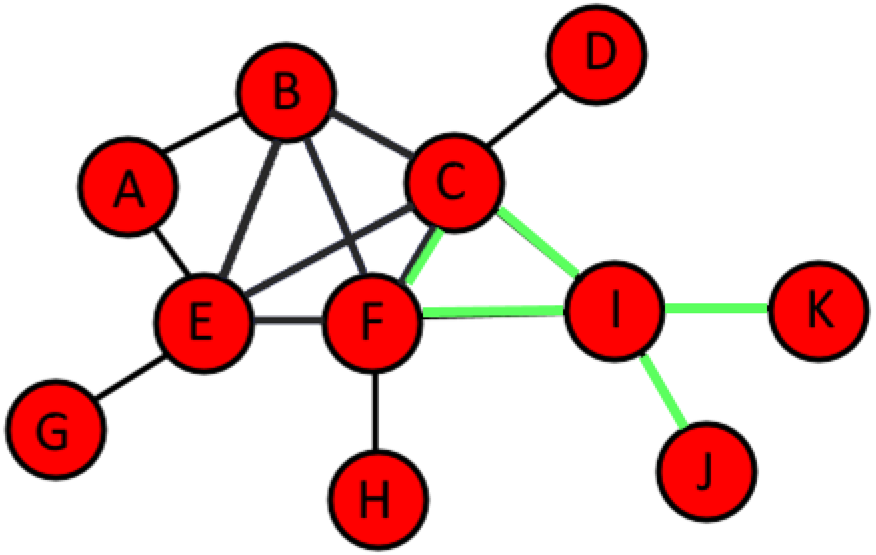
The criterion for all members to share an edge with all other members can be too strict in many real cases

K-cliques / N-cliques

Wasserman & Faust.

All individuals of a clique must connect to all other members within N steps. So nodes may not connect directly to all members, but they are 'friends-of-friends‘. These intermediary nodes may or may not be in the clique themselves.

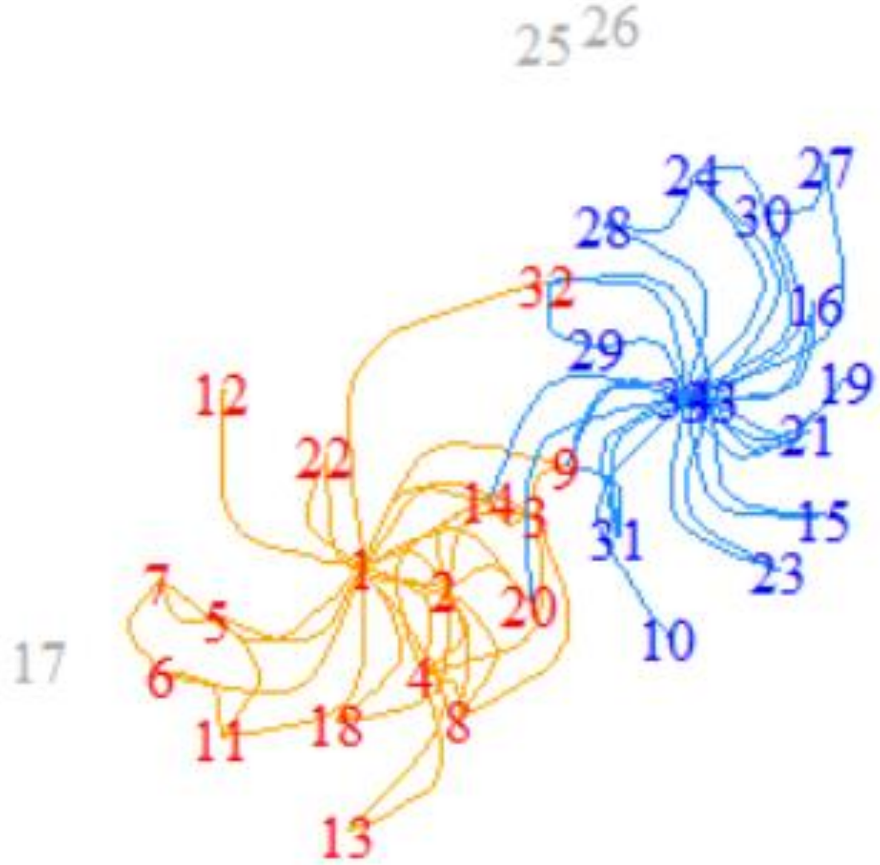
A 1-clique is the same as the original definition of clique.



There are four 2-cliques in this graph:

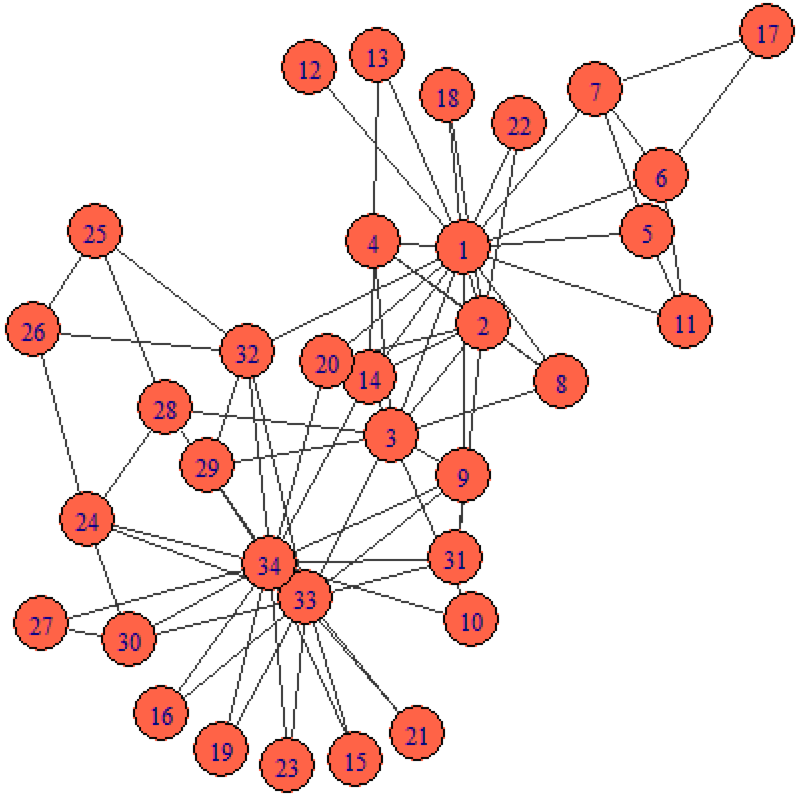
1. "G" "E" "B" "A" "F" "C“
2. "E" "B" "F" "C" "D" "I"
3. "E" "B" "F" "C" "I" "H“
4. "F" "C" "I" "J" "K"

K-clique analysis of Zachary network (k=2)

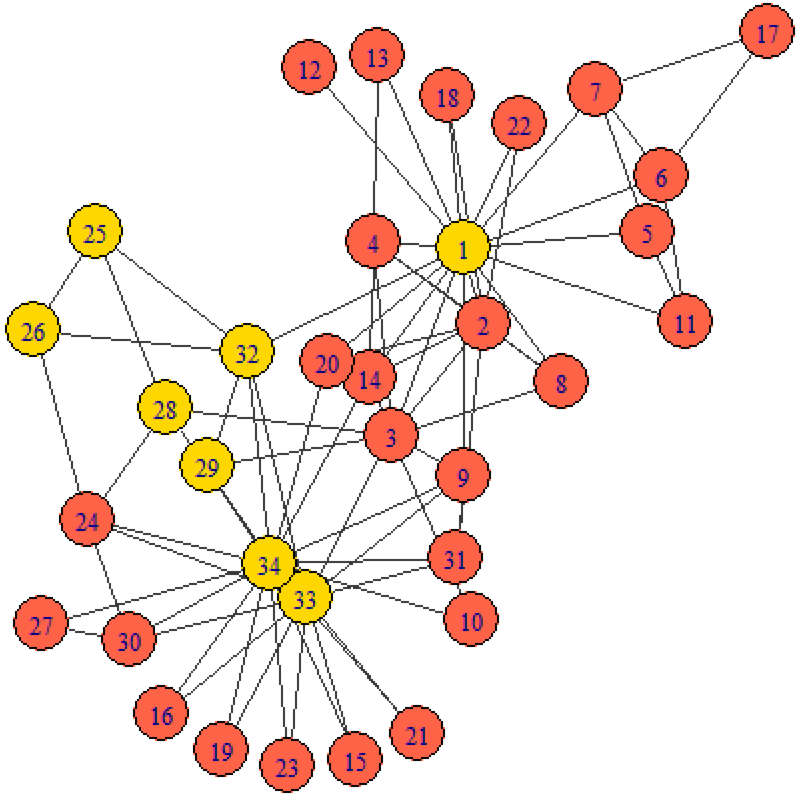


N-clans

Compared to K-cliques, in N-clans, friends-of-friends connections cannot go via nodes who are not members of the clique – so this has a stronger threshold for connectedness

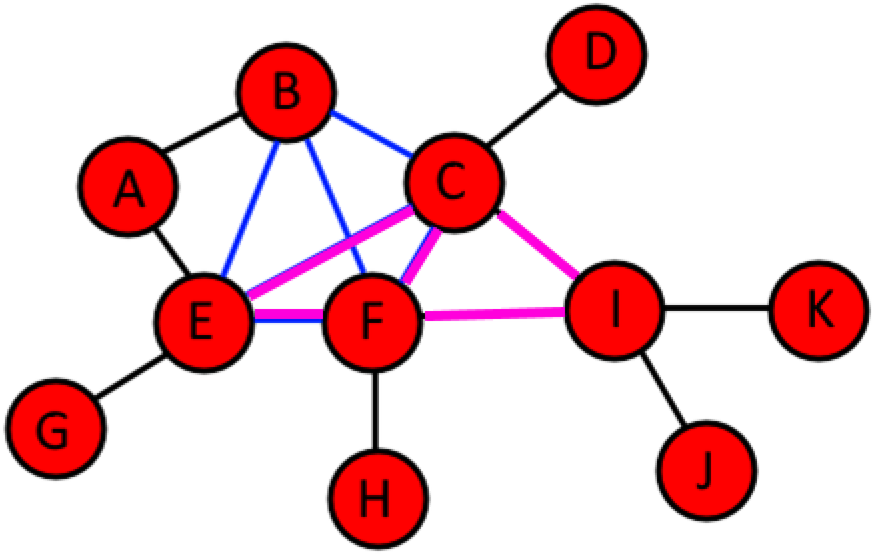


Zachary karate club



This gold group is an n-clique but not an n-clan – 1 & 28 do not connect via a clique member (3)

K-plexes



E could be considered part of the CFIE clique as it connects to N-1 members of the CFI clique.

A cannot be considered part of the BCEF clique as it only connects with N-2 members of that clique.

Consider individuals to be part of a clique if they inter-connect with N-k other individuals in the clique

e.g. if 6 in a clique, a node could connect to 5 of the clique but not the 6th.

K-cliques / N-clans tend to find long 'stringy' cliques

K-plexes tends to find more numerous smaller connected cliques

K-plexes tends to identify overlapping social circles

K-cores

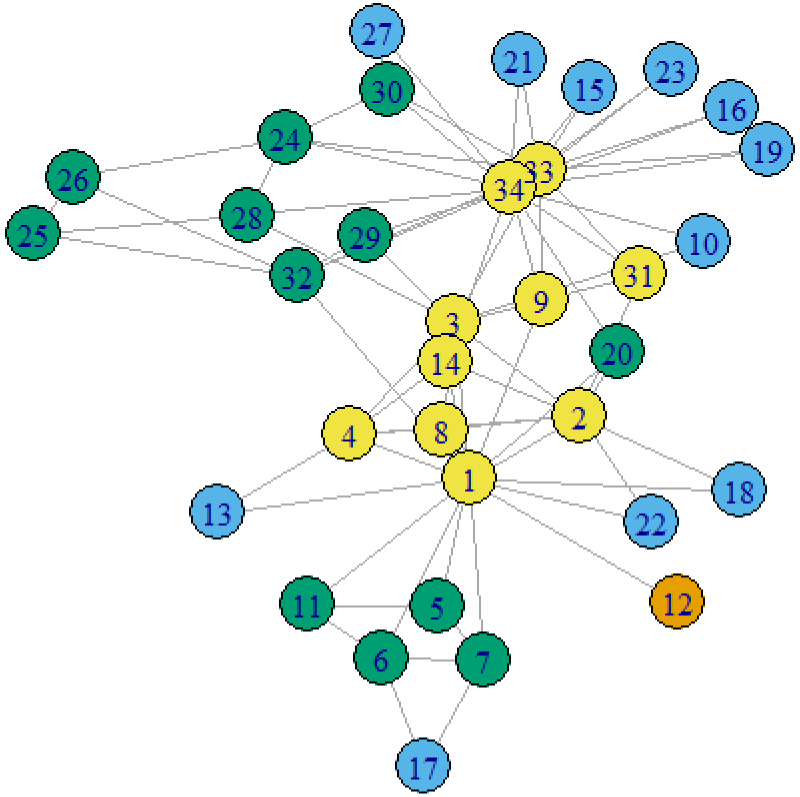
Seidman 1983

A maximal subset of vertices such that each is connected to at least k others in the subset. i.e. in a 4 core, all members of that core are connected with at least four other members.

Typically identify the largest sized K-core for each value of K.

Here the yellow group is K=4. After this, assign nodes to K=3 (green) – these can connect to nodes in higher K cores but K-cores cannot overlap.

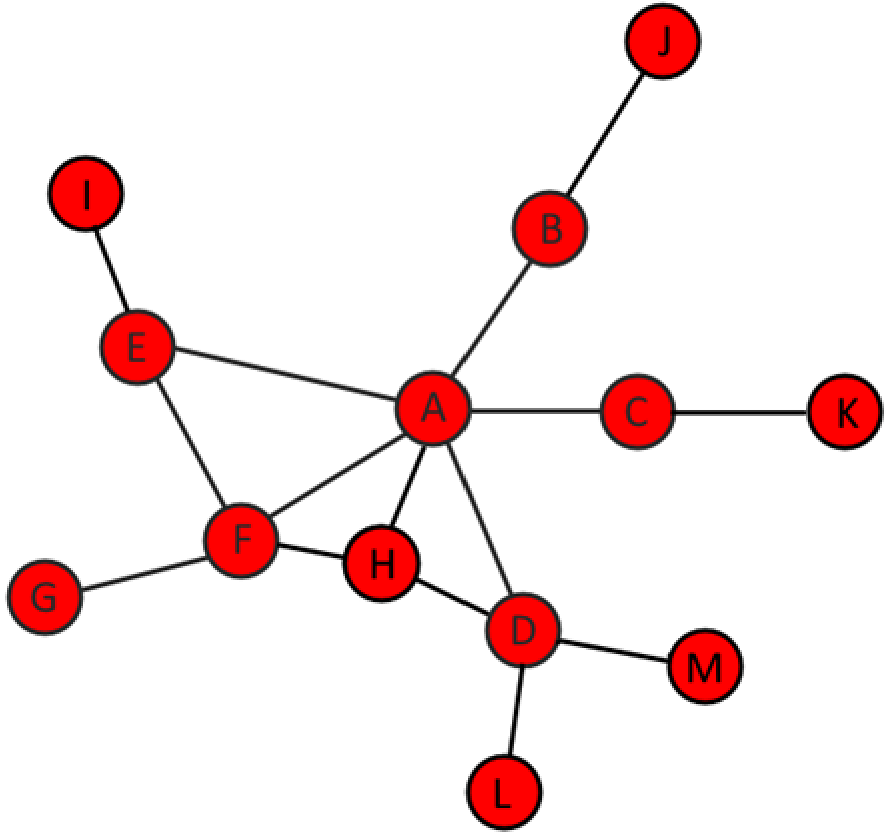
K = 4 yellow
K = 3 green
K = 2 blue
K = 1 orange



Useful identification method to identify key membership groups in networks

Social Roles

Cut-points



A, B, C, D, E, F are all cut-points
H is not a cut-point.

Cutpoints (also called articulation points) = nodes whose removal increases the number of connected components in a graph.

These individuals are often called 'brokers'

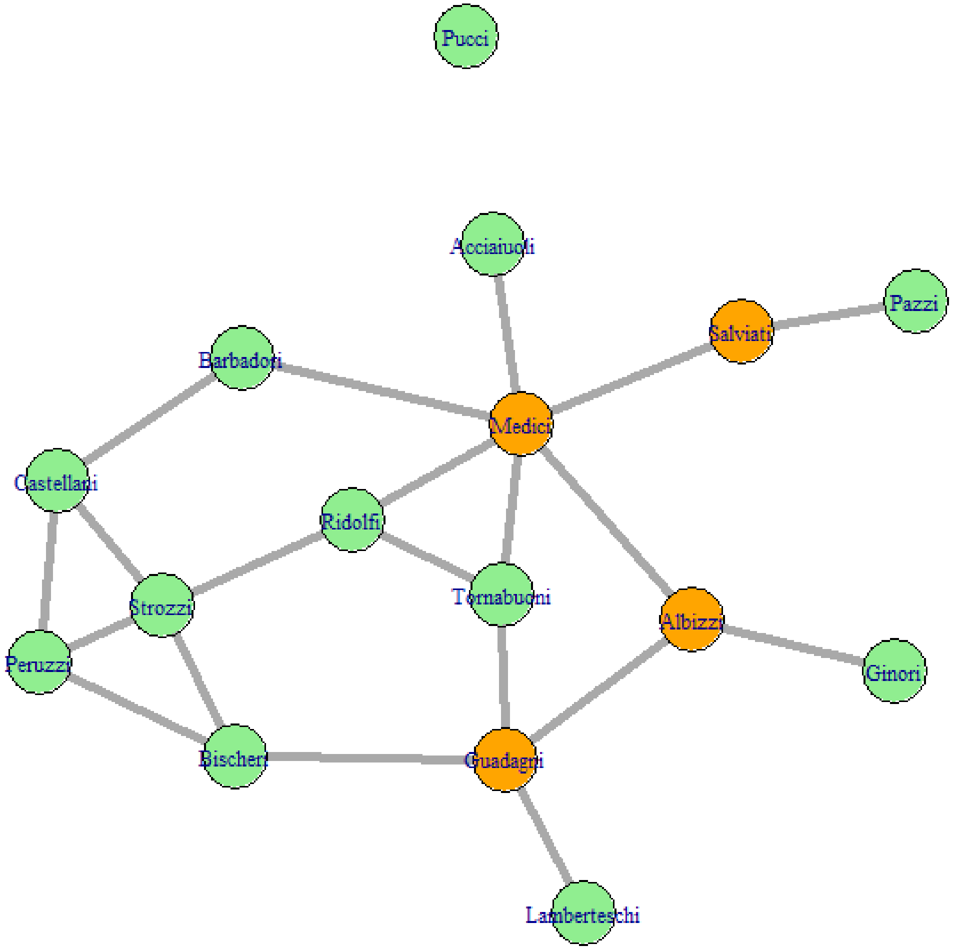
Maximal non-separable sub-graphs are called 'blocks' or 'bi-components'

These are the groups of nodes that keep the graph connected. They do not contain cut-points (cut-vertices).

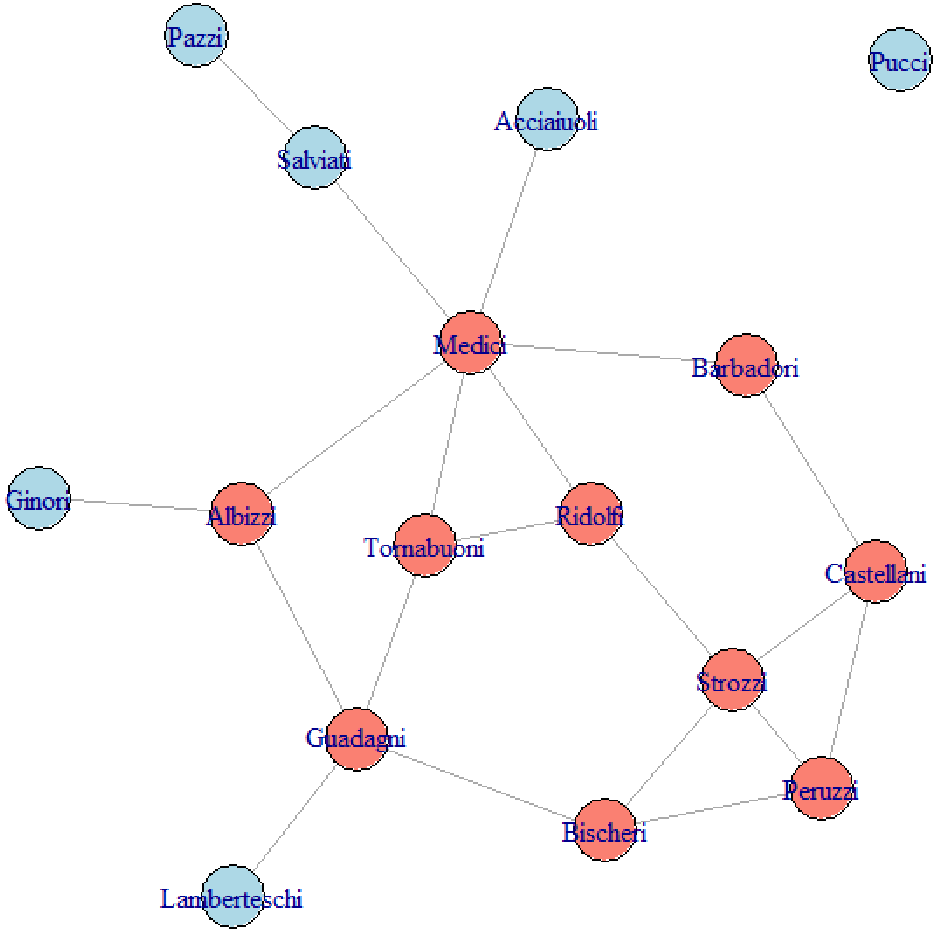
Bridges are edges whose removal increases the components in the network

Cut-points & Bi-connected Components

A graph is bi-connected if the removal of any single vertex (and its adjacent edges) does not disconnect it.



Cut-points (orange) in Padgett’s Florentine Marriage Network



The largest bi-connected component can help identify which cut-points are notable.

Brokerage

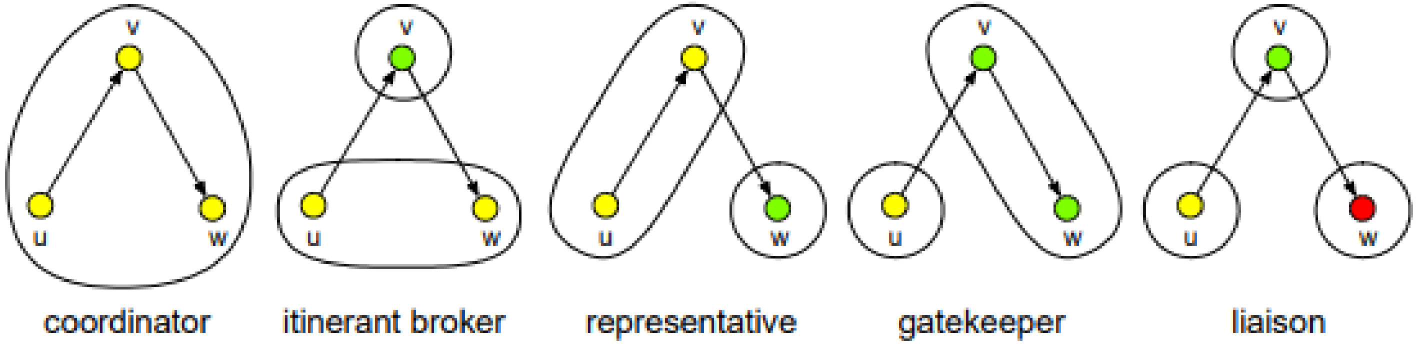


Figure 9 - Five brokerage roles of actor *v*.

http://vlado.fmf.uni-lj.si/pub/networks/course/ch07/Chapter7.pdf

Gould & Fernandez define brokerage as the role played by an actor who mediates contact between two other nodes.

When dealing with class memberships, we can define several 'roles':

Coordinator: actor brokers contact between two members of own group

Itinerant broker: actor brokers contact between two members of another different group (not actor's own group)

Representative: actor mediates incoming tie from out-group indiv to an in-group indiv

Gatekeeper: actor mediates tie from in group-member to an out-group member

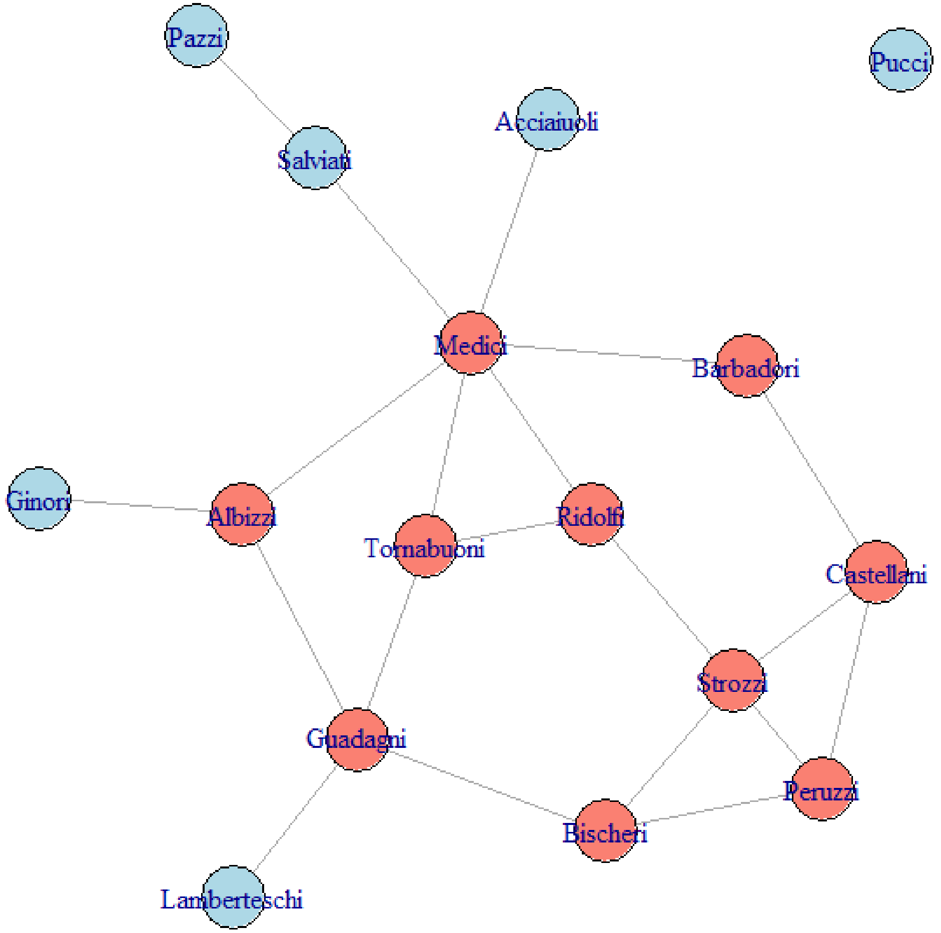
Liaison: actor mediates contact between two individuals from different groups neither of which actor belongs to.

Total: cumulative brokerage role occupancy

Broker in middle (‘sna’ terminology):

Coordinator (w_I): A → A → A
Itinerant broker (w_O): A → B → A
Representative (b_{IO}): A → B → B
Gatekeeper (b_{OI}): A → A → B
Liaison (b_O): A → B → C

Brokerage

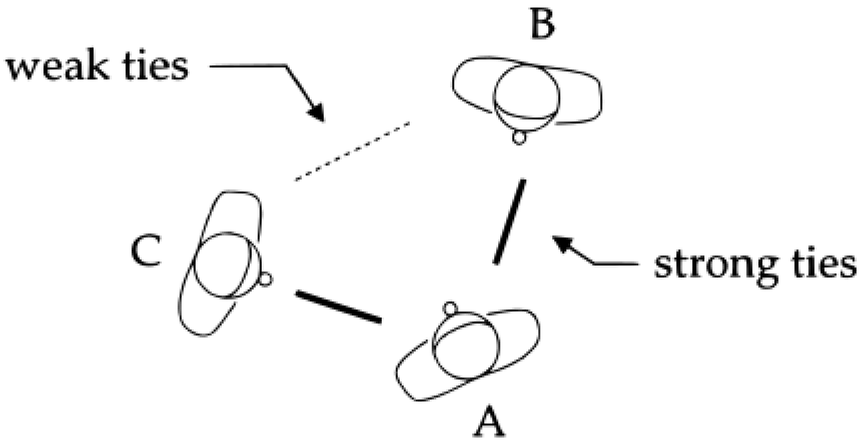


Broker in middle ('sna' terminology):

Coordinator (w_I): A → A → A
Itinerant broker (w_O): A → B → A
Representative (b_{IO}): A → B → B
Gatekeeper (b_{OI}): A → A → B
Liaison (b_O): A → B → C

```
             w_I w_O b_IO b_OI b_O  t
Acciaiuoli     0   0    0    0   0  0
Albizzi        0   2    0    0   4  6
Barbadori      0   0    0    0   2  2
Bischeri       4   0    0    0   0  4
Castellani     0   0    2    2   0  4
Ginori         0   0    0    0   0  0
Guadagni       2   0    4    4   2 12
Lamberteschi   0   0    0    0   0  0
Medici         4   6    9    9   0 28
Pazzi          0   0    0    0   0  0
Peruzzi        2   0    0    0   0  2
Pucci          0   0    0    0   0  0
Ridolfi        0   0    2    2   0  4
Salviati       0   2    0    0   0  2
Strozzi        2   0    3    3   0  8
Tornabuoni     0   0    2    2   0  4
```

Weak vs. Strong Ties



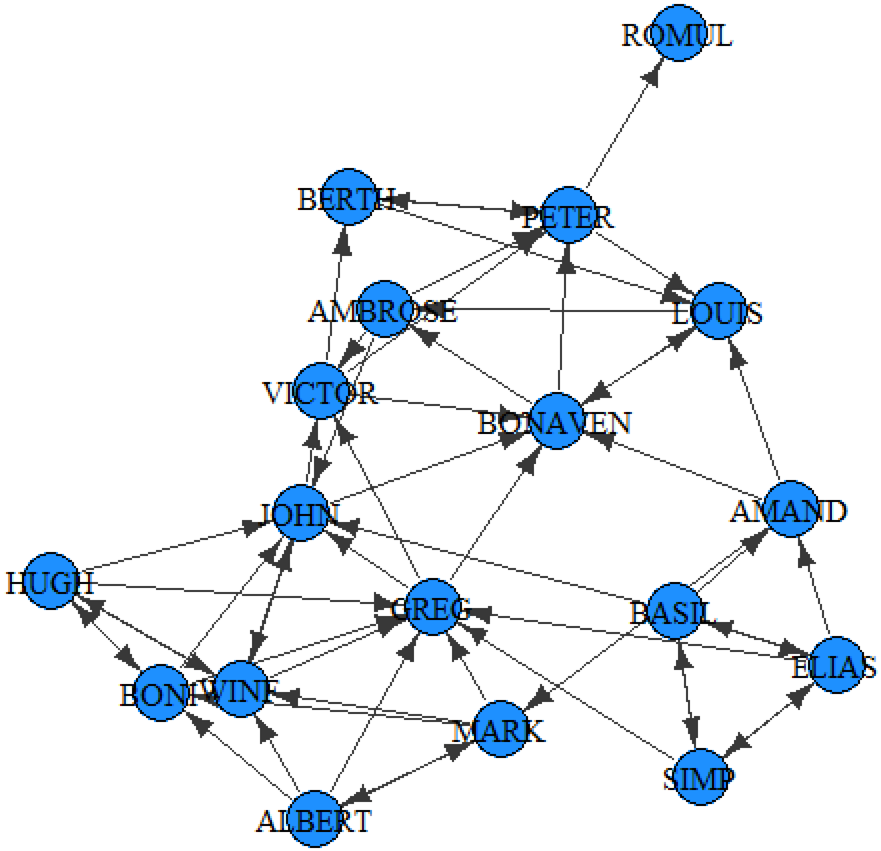
Reach refers to how many ties can be reached in n steps from a node. Subtracting degree from reach gives a measure of the number of 'weak ties' that a node has

Structural Equivalence

Structural Equivalence assesses the similarity of actors in a network.

- For example, if two actors send and receive ties to the same third parties then they are considered to be structurally equivalent
- Structurally equivalent actors tend to show homogeneity in attributes
- This differs from cohesive subgroups, as actors in subgroups tend to interact with each other whereas structurally equivalent actors may not even know each other.
- Same logic for directed graphs: structural equivalence if out-going/incoming ties from 3rd parties are similar. For valued graphs: weights of out-going/incoming edges from 3rd parties should be similar

Comparing individuals within networks



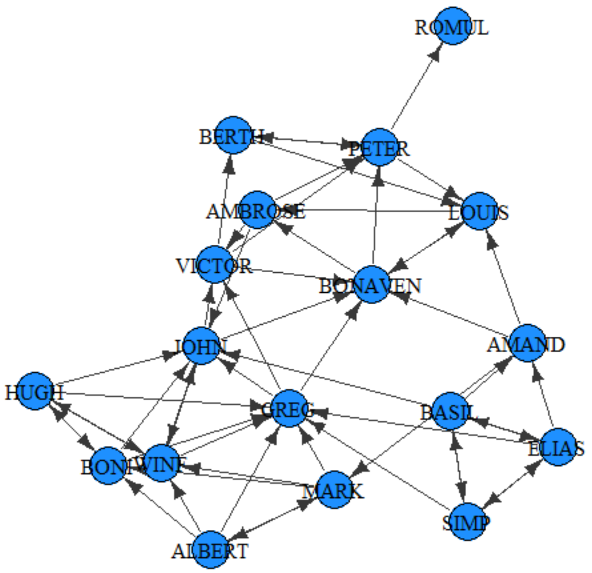
Sampson's 1968 monastery data
Esteem network

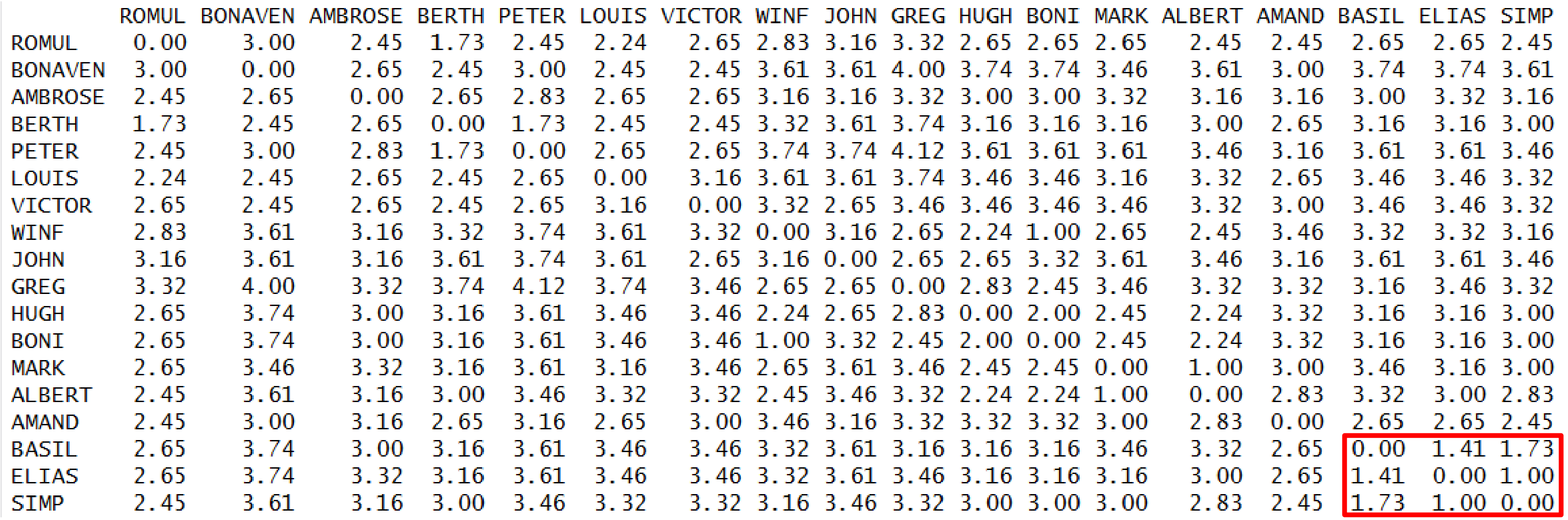
Compare 'similarity' of matrix cells using a number of measures:

- correlation
- Euclidean distance
- Hamming distance
- Jacard's distance

```
ROMUL    . . . . . . . . . . . . . . . . . .
BONAVEN  . . 1 . 1 1 . . . . . . . . . . . .
AMBROSE  . . . . 1 . 1 . 1 . . . . . . . . .
BERTH    . . . . 1 1 . . . . . . . . . . . .
PETER    1 . . 1 . 1 . . . . . . . . . . . .
LOUIS    . 1 1 . . . . . . . . . . . . . . .
VICTOR   . 1 . 1 1 . . . . . . . . . . . . .
WINF     . . . . . . . . 1 1 1 . . . . . . .
JOHN     . 1 . . . . 1 1 . . . . . . . . . .
GREG     . 1 . . . . 1 . 1 . . . . . . . . .
HUGH     . . . . . . . 1 1 1 . 1 . . . . . .
BONI     . . . . . . . . 1 1 1 . . . . . . .
MARK     . . . . . . . 1 . 1 . 1 . 1 . . . .
ALBERT   . . . . . . . 1 . 1 . 1 1 . . . . .
AMAND    . 1 . . . 1 . . . . . . 1 . . . . .
BASIL    . . . . . . . . 1 . . . . . 1 . 1 1
ELIAS    . . . . . . . . . 1 . . . . 1 1 . 1
SIMP     . . . . . . . . . 1 . . . . . 1 1 .
```

Comparing individuals within networks



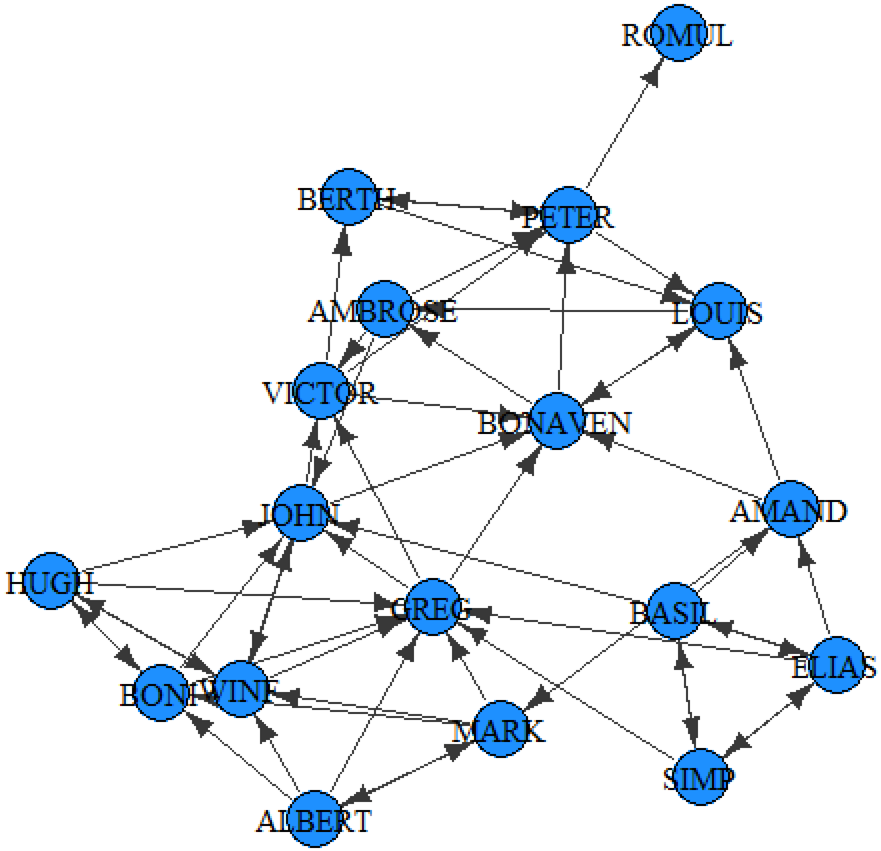


	ROMUL	BONAVEN	AMBROSE	BERTH	PETER	LOUIS	VICTOR	WINF	JOHN	GREG	HUGH	BONI	MARK	ALBERT	AMAND	BASIL	ELIAS	SIMP
ROMUL	0.00	3.00	2.45	1.73	2.45	2.24	2.65	2.83	3.16	3.32	2.65	2.65	2.65	2.45	2.45	2.65	2.65	2.45
BONAVEN	3.00	0.00	2.65	2.45	3.00	2.45	2.45	3.61	3.61	4.00	3.74	3.74	3.46	3.61	3.00	3.74	3.74	3.61
AMBROSE	2.45	2.65	0.00	2.65	2.83	2.65	2.65	3.16	3.16	3.32	3.00	3.00	3.32	3.16	3.16	3.00	3.32	3.16
BERTH	1.73	2.45	2.65	0.00	1.73	2.45	2.45	3.32	3.61	3.74	3.16	3.16	3.16	3.00	2.65	3.16	3.16	3.00
PETER	2.45	3.00	2.83	1.73	0.00	2.65	2.65	3.74	3.74	4.12	3.61	3.61	3.61	3.46	3.16	3.61	3.61	3.46
LOUIS	2.24	2.45	2.65	2.45	2.65	0.00	3.16	3.61	3.61	3.74	3.46	3.46	3.16	3.32	2.65	3.46	3.46	3.32
VICTOR	2.65	2.45	2.65	2.45	2.65	3.16	0.00	3.32	2.65	3.46	3.46	3.46	3.46	3.32	3.00	3.46	3.46	3.32
WINF	2.83	3.61	3.16	3.32	3.74	3.61	3.32	0.00	3.16	2.65	2.24	1.00	2.65	2.45	3.46	3.32	3.32	3.16
JOHN	3.16	3.61	3.16	3.61	3.74	3.61	2.65	3.16	0.00	2.65	2.65	3.32	3.61	3.46	3.16	3.61	3.61	3.46
GREG	3.32	4.00	3.32	3.74	4.12	3.74	3.46	2.65	2.65	0.00	2.83	2.45	3.46	3.32	3.32	3.16	3.46	3.32
HUGH	2.65	3.74	3.00	3.16	3.61	3.46	3.46	2.24	2.65	2.83	0.00	2.00	2.45	2.24	3.32	3.16	3.16	3.00
BONI	2.65	3.74	3.00	3.16	3.61	3.46	3.46	1.00	3.32	2.45	2.00	0.00	2.45	2.24	3.32	3.16	3.16	3.00
MARK	2.65	3.46	3.32	3.16	3.61	3.16	3.46	2.65	3.61	3.46	2.45	2.45	0.00	1.00	3.00	3.46	3.16	3.00
ALBERT	2.45	3.61	3.16	3.00	3.46	3.32	3.32	2.45	3.46	3.32	2.24	2.24	1.00	0.00	2.83	3.32	3.00	2.83
AMAND	2.45	3.00	3.16	2.65	3.16	2.65	3.00	3.46	3.16	3.32	3.32	3.32	3.00	2.83	0.00	2.65	2.65	2.45
BASIL	2.65	3.74	3.00	3.16	3.61	3.46	3.46	3.32	3.61	3.16	3.16	3.16	3.46	3.32	2.65	0.00	1.41	1.73
ELIAS	2.65	3.74	3.32	3.16	3.61	3.46	3.46	3.32	3.61	3.46	3.16	3.16	3.16	3.00	2.65	1.41	0.00	1.00
SIMP	2.45	3.61	3.16	3.00	3.46	3.32	3.32	3.16	3.46	3.32	3.00	3.00	3.00	2.83	2.45	1.73	1.00	0.00

e.g. Matrix of Euclidian distance. Notice the distances for Basil-Simp-Elias are all low relative to other individuals

Sampson's 1968 monastery data
Esteem network

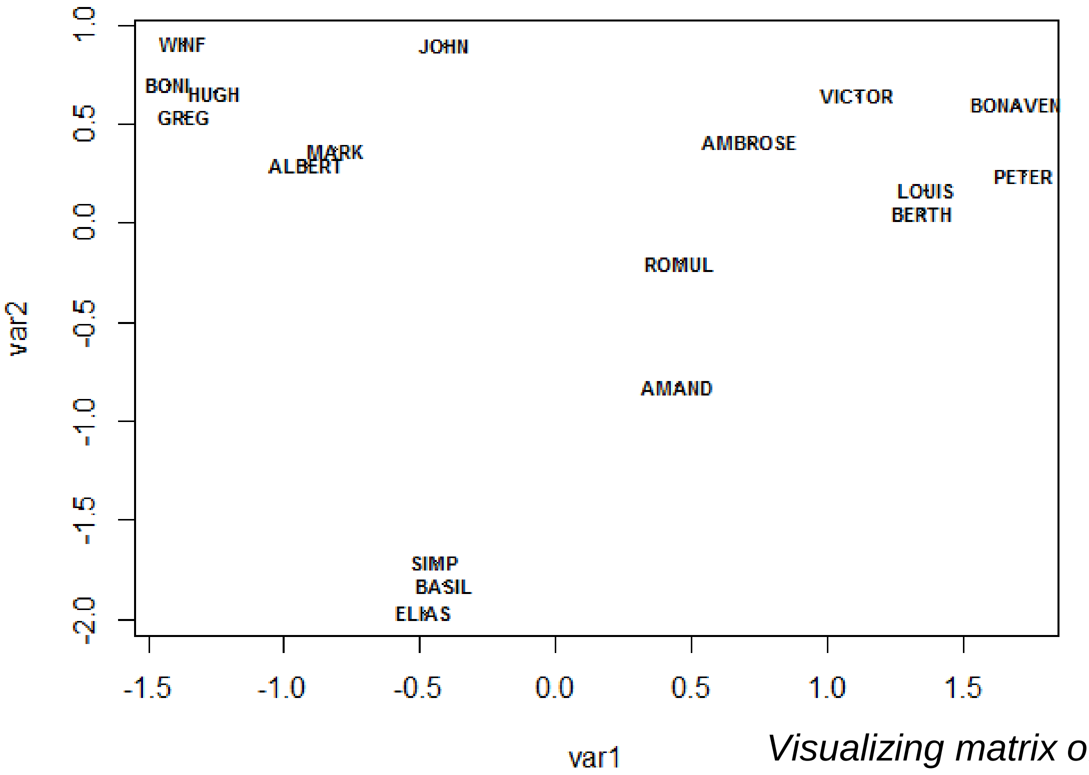
Comparing individuals within networks



Sampson's 1968 monastery data
Esteem network

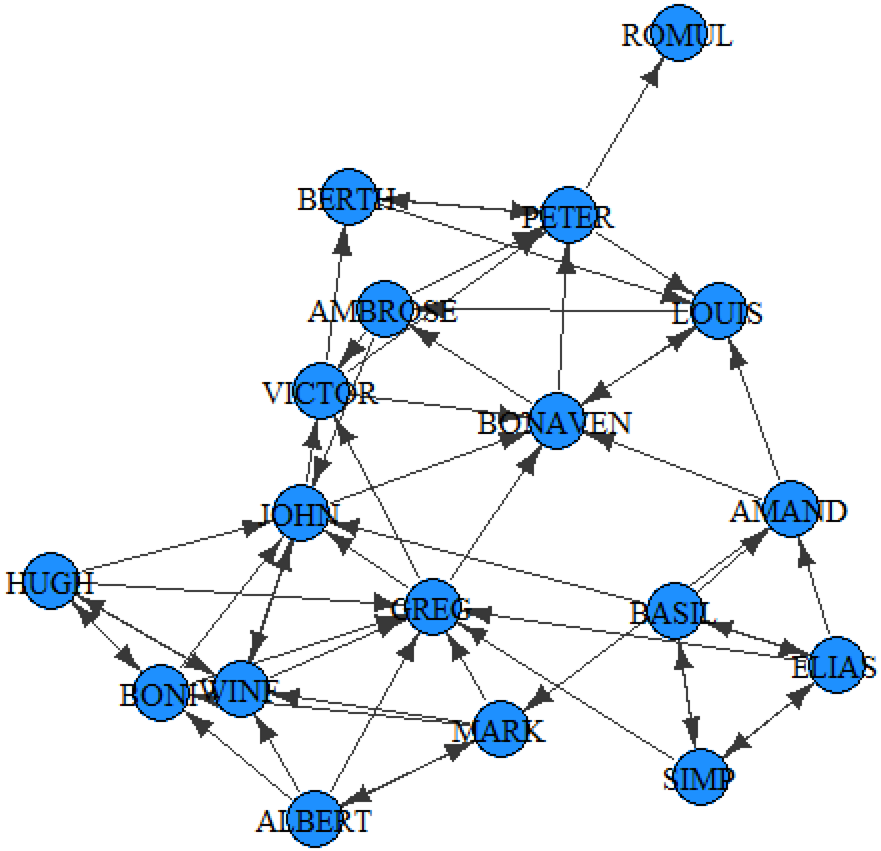
Compare 'similarity' of matrix cells using a number of measures:

- correlation
- Euclidean distance
- Hamming distance
- Jacard's distance



Visualizing matrix of Euclidian distance using MDS

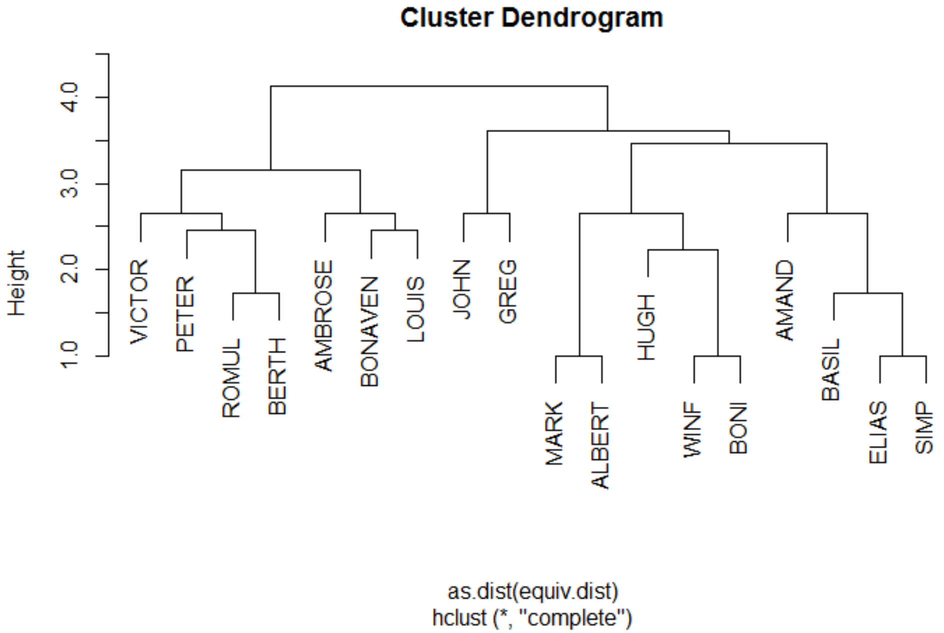
Comparing individuals within networks



Sampson's 1968 monastery data
Esteem network

Compare 'similarity' of matrix cells using a number of measures:

- correlation
- Euclidean distance
- Hamming distance
- Jacard's distance



Hierarchical clustering of similarity/dissimilarity matrix

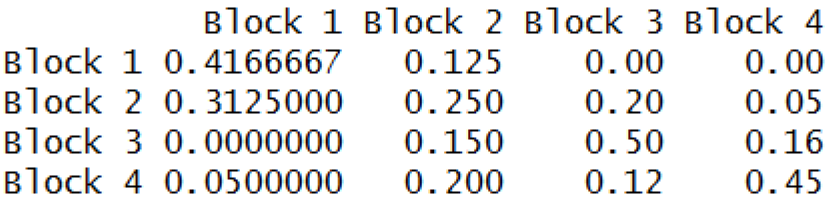
Blockmodels

An exploratory method to identify role similarity.

- Blockmodels are typically used to identify common characteristics of actors within a block or to describe individual positions/social groups

- Several blockmodel algorithms have been developed to assign individual nodes into ‘blocks’ based on their similarity/dissimilarity to other nodes

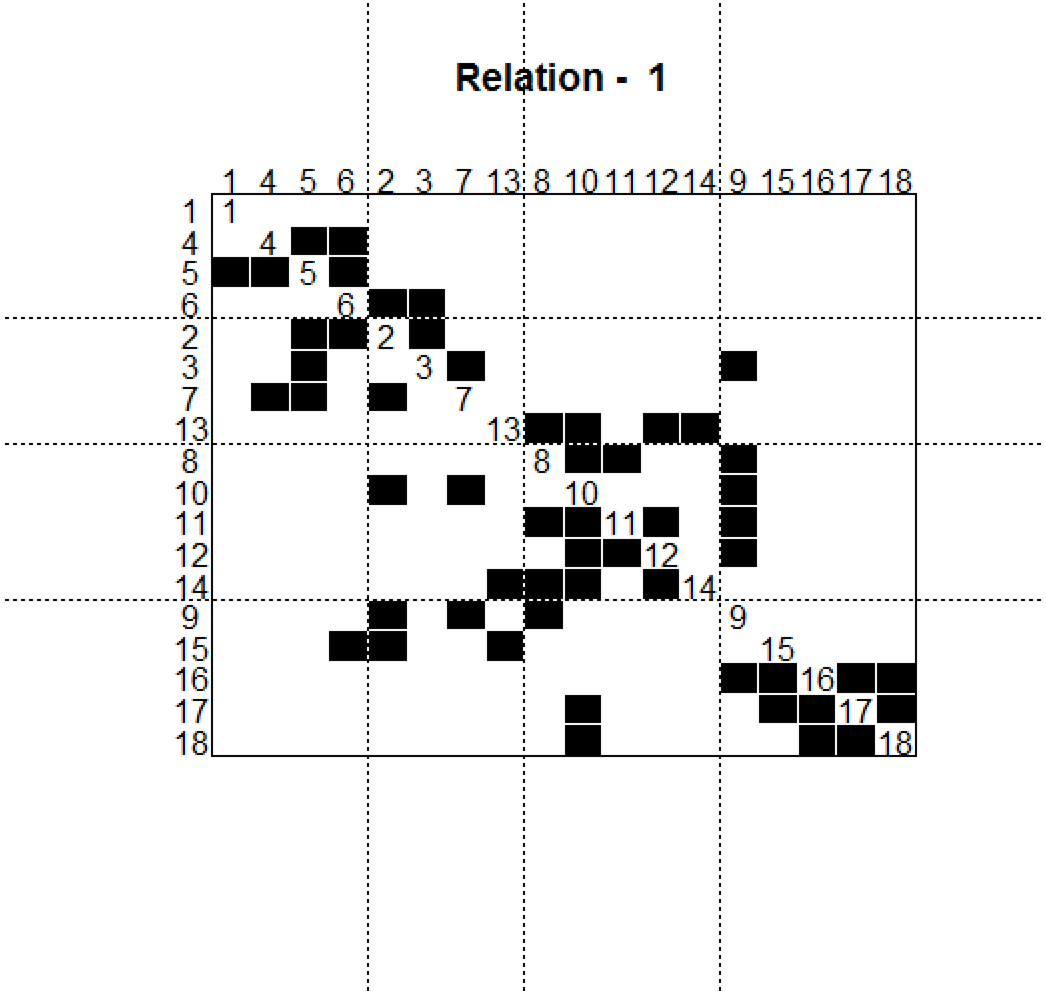
e.g. Breiger, R.L., Boorman, S.A., and Arabie, P. 1975. An Algorithm for Clustering Relational Data with Applications to Social Network Analysis #and Comparison with Multidimensional Scaling. Journal of Mathematical Psychology 12: 328--383.



```
          Block 1 Block 2 Block 3 Block 4
Block 1 0.4166667   0.125    0.00    0.00
Block 2 0.3125000   0.250    0.20    0.05
Block 3 0.0000000   0.150    0.50    0.16
Block 4 0.0500000   0.200    0.12    0.45
```

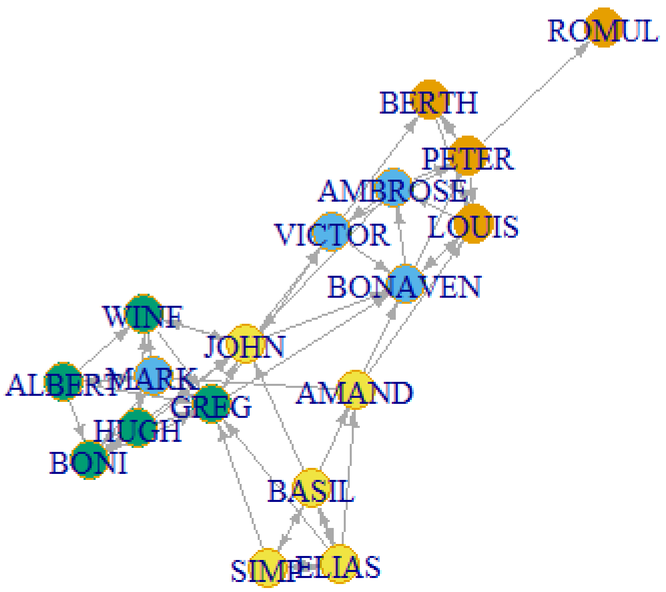
Blockmodeling helps identify groups of nodes that show high density within their ‘block’. It can also identify higher than average interactions between blocks.

Graph density = 0.176



Relation - 1

	1	4	5	6	2	3	7	13	8	10	11	12	14	9	15	16	17	18
1	1																	
4		4	■	■														
5	■	■	5	■														
6				6	■	■												
2			■	■	2	■												
3			■			3	■							■				
7		■	■		■		7											
13								13	■	■		■	■					
8									8	■	■			■				
10					■		■			10				■				
11									■	■	11	■		■				
12										■	■	12		■				
14								■	■	■		■	14					
9					■		■		■					9				
15				■	■			■							15			
16														■	■	16	■	■
17										■					■	■	17	■
18										■						■	■	18



- e.g. Amand, Basil, Elias and Simplius are outcasts.

Regular Equivalence

A methodology for finding individuals that have similar social roles – defined by identifying nodes that are similar in the types and number of edges that they have. These do not have to be edges shared with the same individuals.

e.g. two school-teachers will likely have the same types of edges to school children in their class and to other teachers. They won’t necessarily share an edge themselves or to each other’s class members.

e.g. The REGE Algorithm (Profile Similarity Measure)

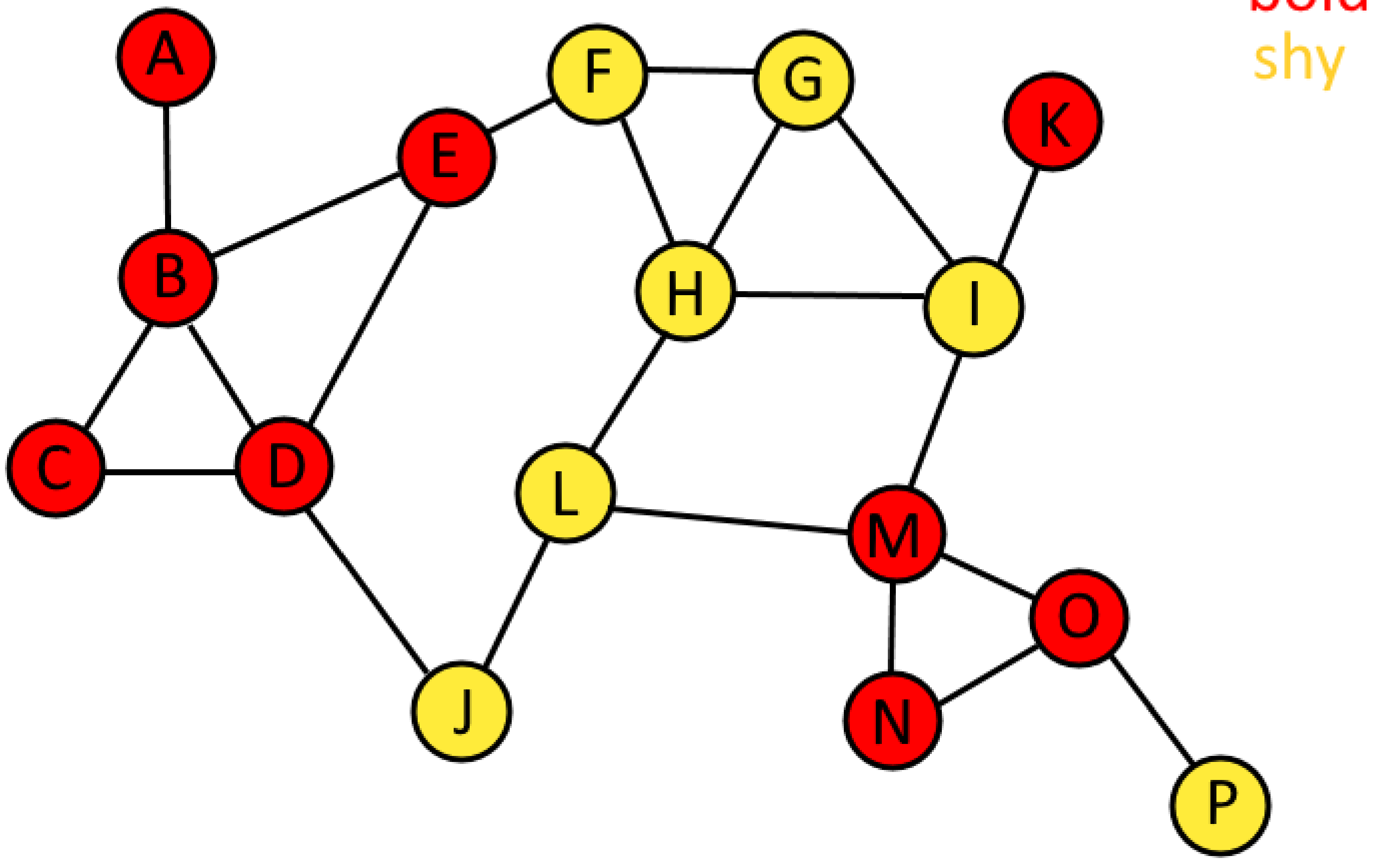
This only works on directed data, and there must be at least one actor with either zero out-degree (a sink) or zero-indegree (a source)

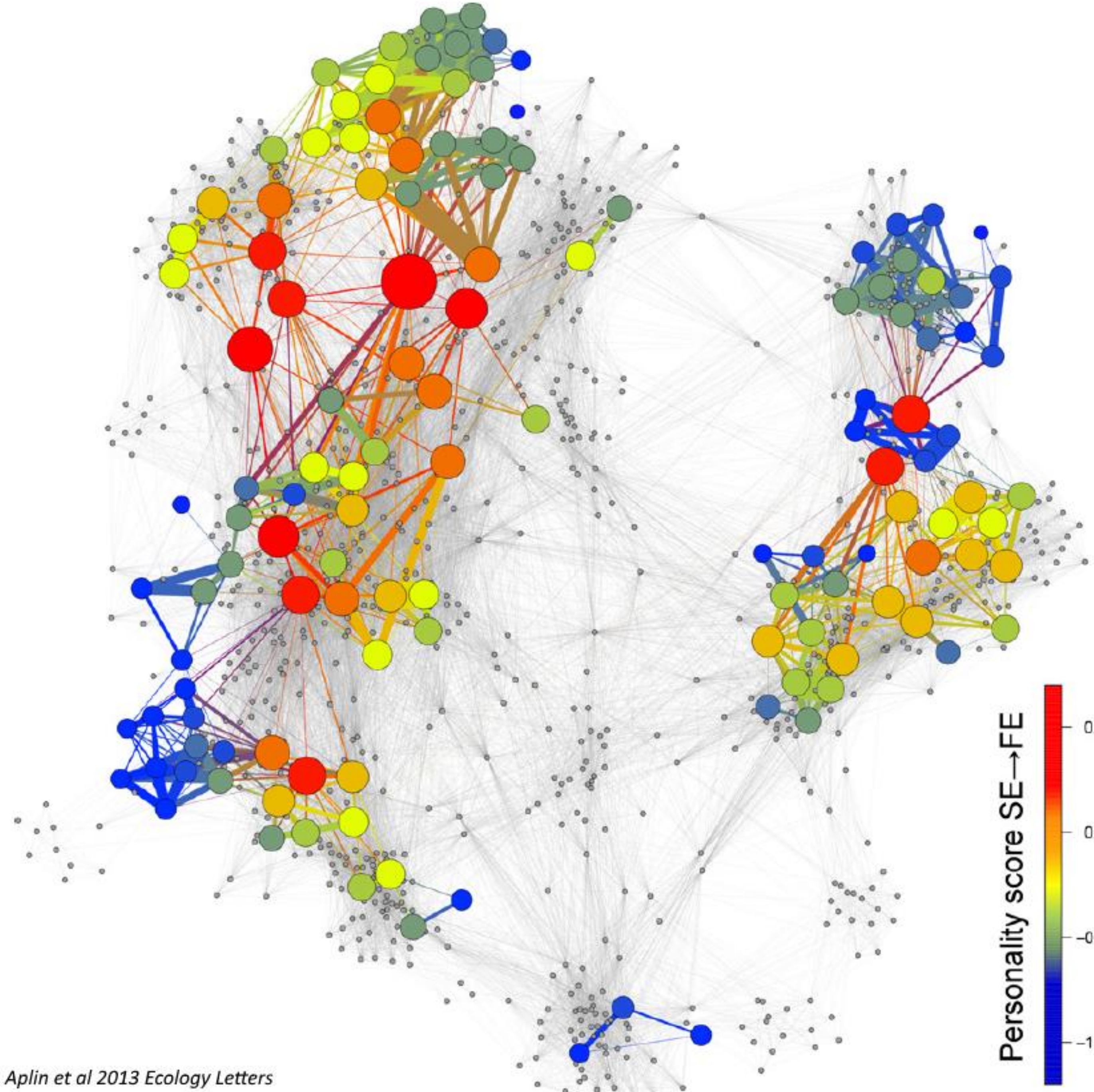
For REGE: scores of 1 equals perfect equivalence.
For REGD: scores of 1 equals perfect lack of equivalence.

Žiberna, A. (2008). Direct and indirect approaches to blockmodeling of valued networks in terms of regular equivalence. Journal of Mathematical Sociology, 32(1), 57-84.

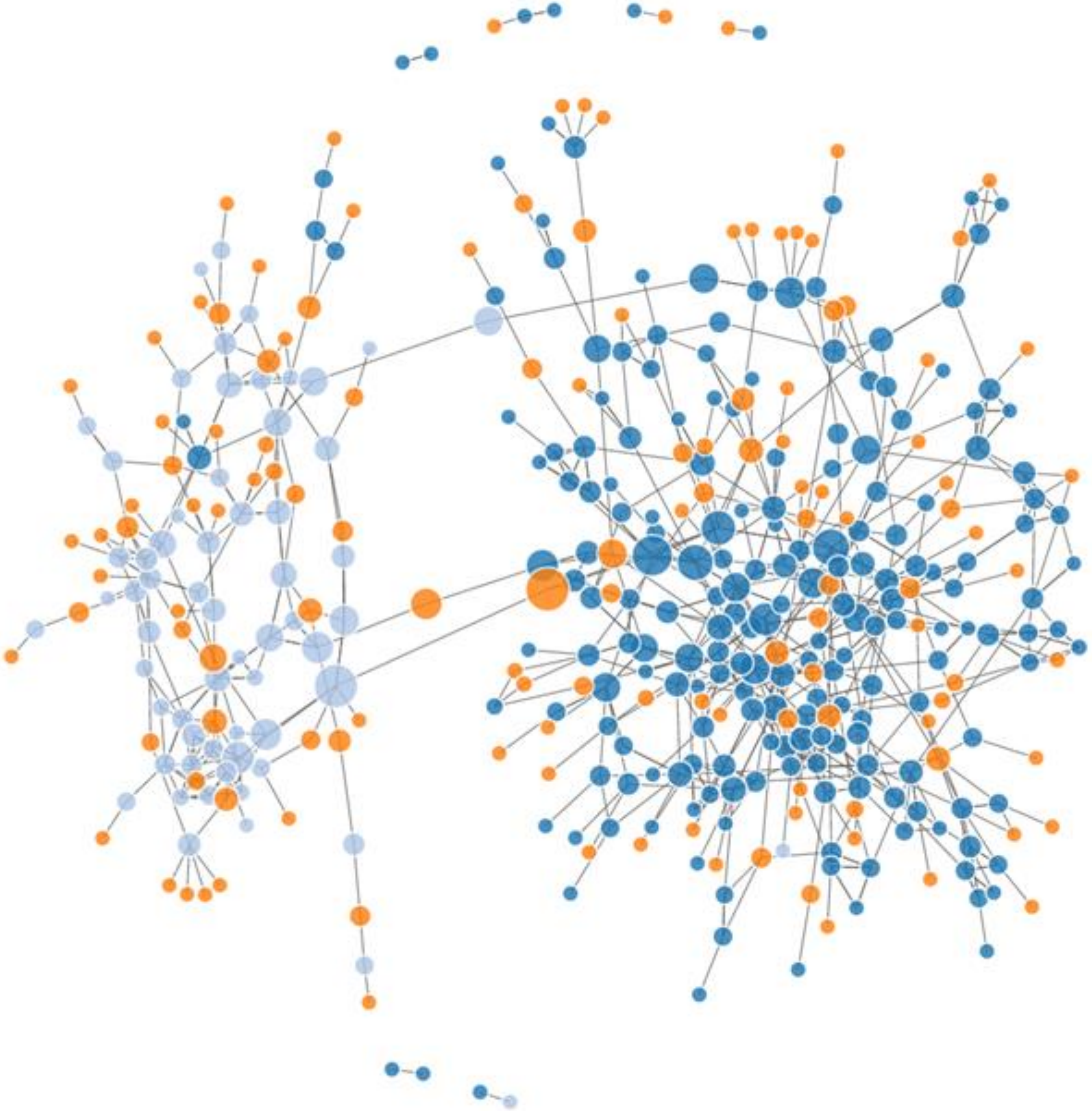
Preferential Attachments - Birds of a feather ?

Assortativity / Homophily





Aplin et al 2013 Ecology Letters



The assortativity coefficient ranges from -1 (nodes fully disassorted) to +1 (nodes fully assorted).

In a fully assorted network all ties would be between individuals that share attributes

In a fully disassorted network all ties would be between individuals that do not share attributes.

Attributes may be categorical or continuous.

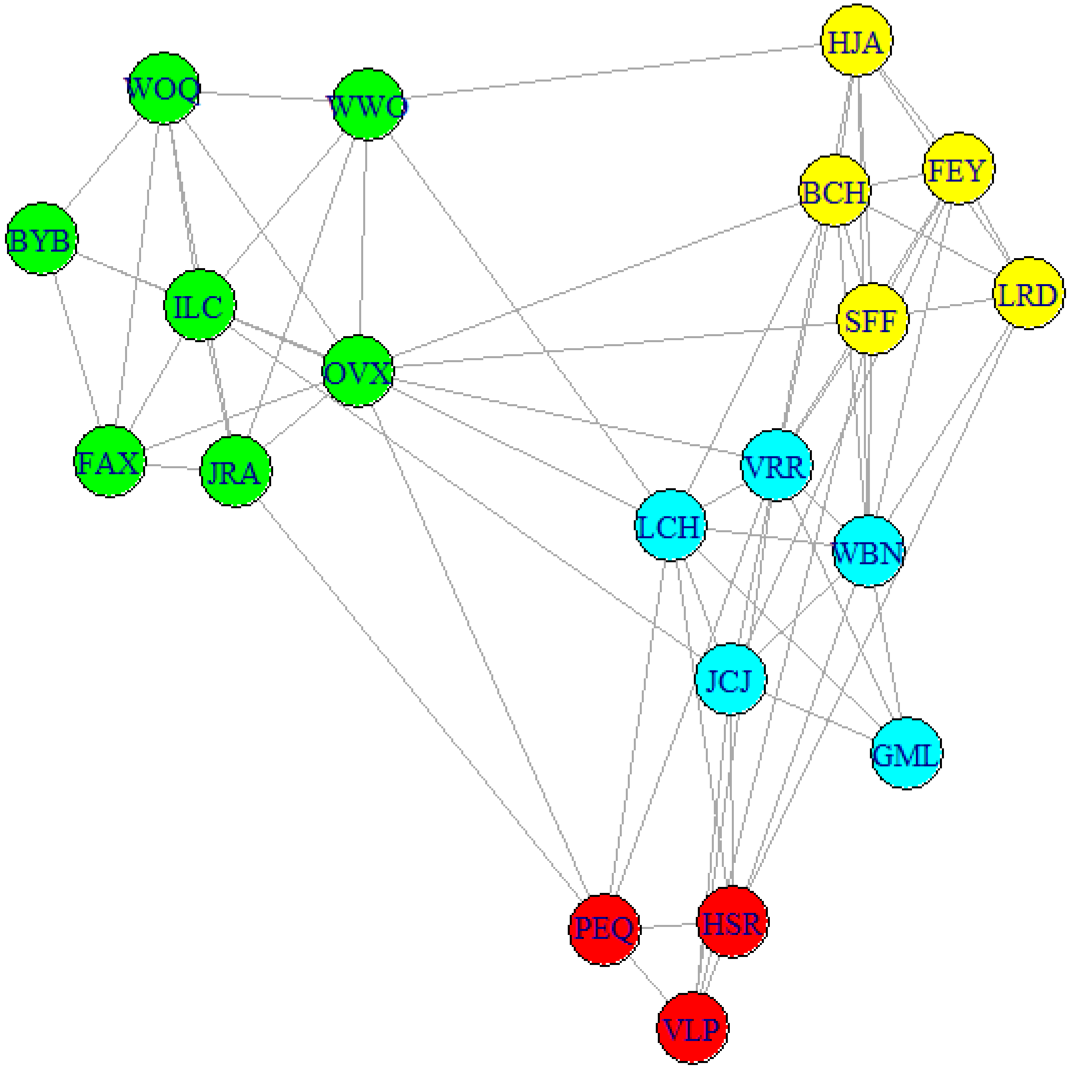
Newman’s 2003 assortativity coefficient was originally described only for undirected and unweighted networks.

Assortativity Functions in R
- assortativity_nominal() is for categorical variables (labels)
- assortativity() is for ordinal and above variables
- assortativity_degree() checks assortativity in node degrees

Farine’s 2014 paper extended the view of assortativity coefficient to enable it to be used for directed and weighted networks. The assortativity coefficients generated from these networks are much more robust than using the original method.

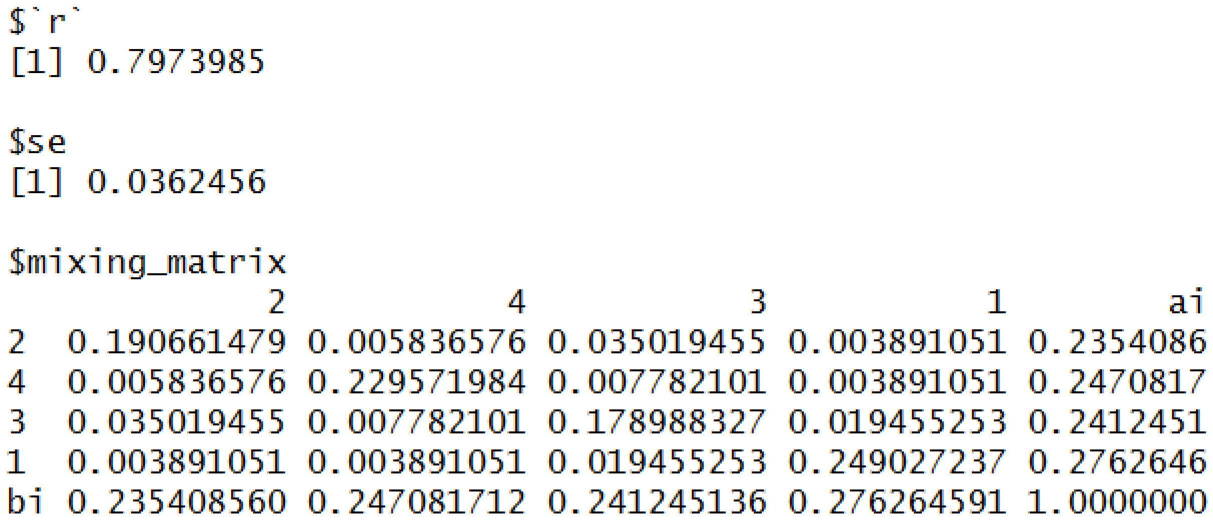
Assortativity functions in 'assortnet'
- assortment.discrete()
- assortment.continuous()

Farine, D.R. (2014) Measuring phenotypic assortment in animal social networks: weighted associations are more robust than binary edges. *Animal Behaviour* 89: 141-153



Assortativity of this weighted network is 0.797

The standard error of assortativity is 0.036 (using the jackknifing method)



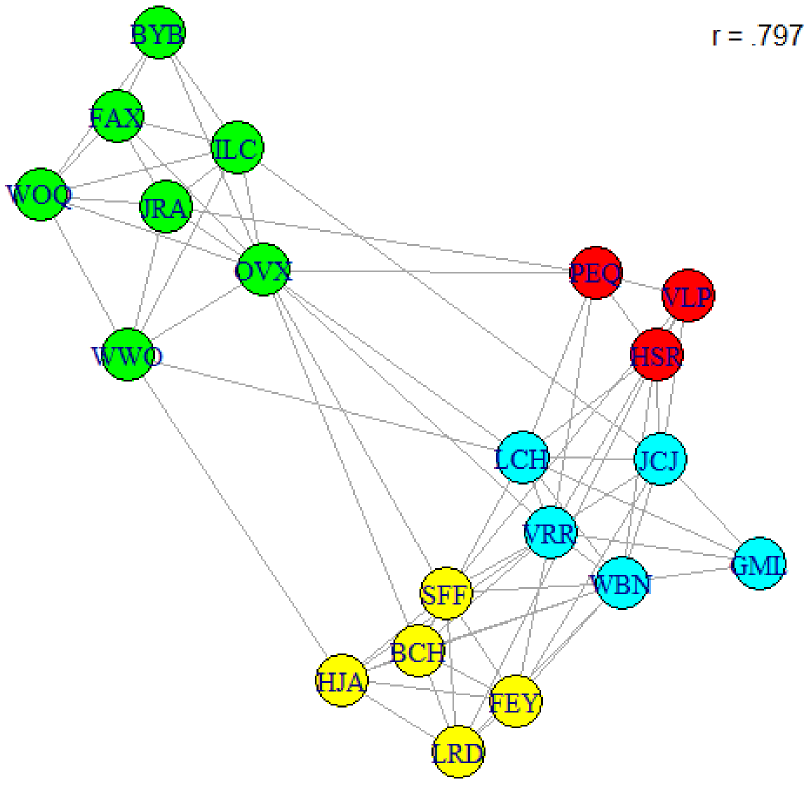
```
$`r`
[1] 0.7973985

$se
[1] 0.0362456

$mixing_matrix
             2           4           3           1        ai
2  0.190661479 0.005836576 0.035019455 0.003891051 0.2354086
4  0.005836576 0.229571984 0.007782101 0.003891051 0.2470817
3  0.035019455 0.007782101 0.178988327 0.019455253 0.2412451
1  0.003891051 0.003891051 0.019455253 0.249027237 0.2762646
bi 0.235408560 0.247081712 0.241245136 0.276264591 1.0000000
```

The mixing matrix represents the distribution of edge-weights (or edges in unweighted networks) between each group.

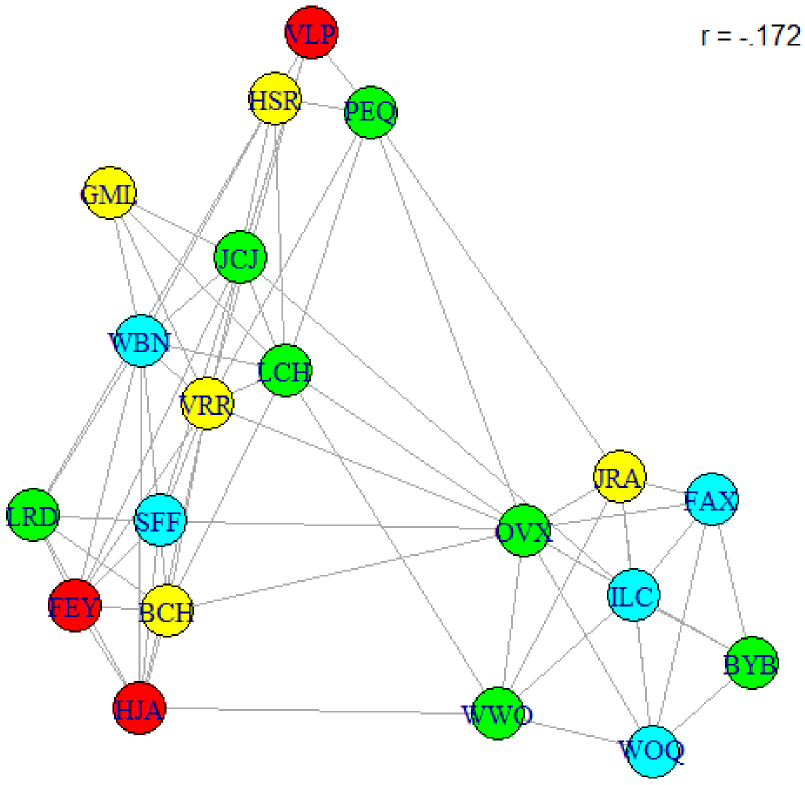
Significance testing assortativity: Node Based Permuations

In node-based permutations we can simply permute the node-level attribute data and recalculate assortativity. Illustrated here is one such permutation.

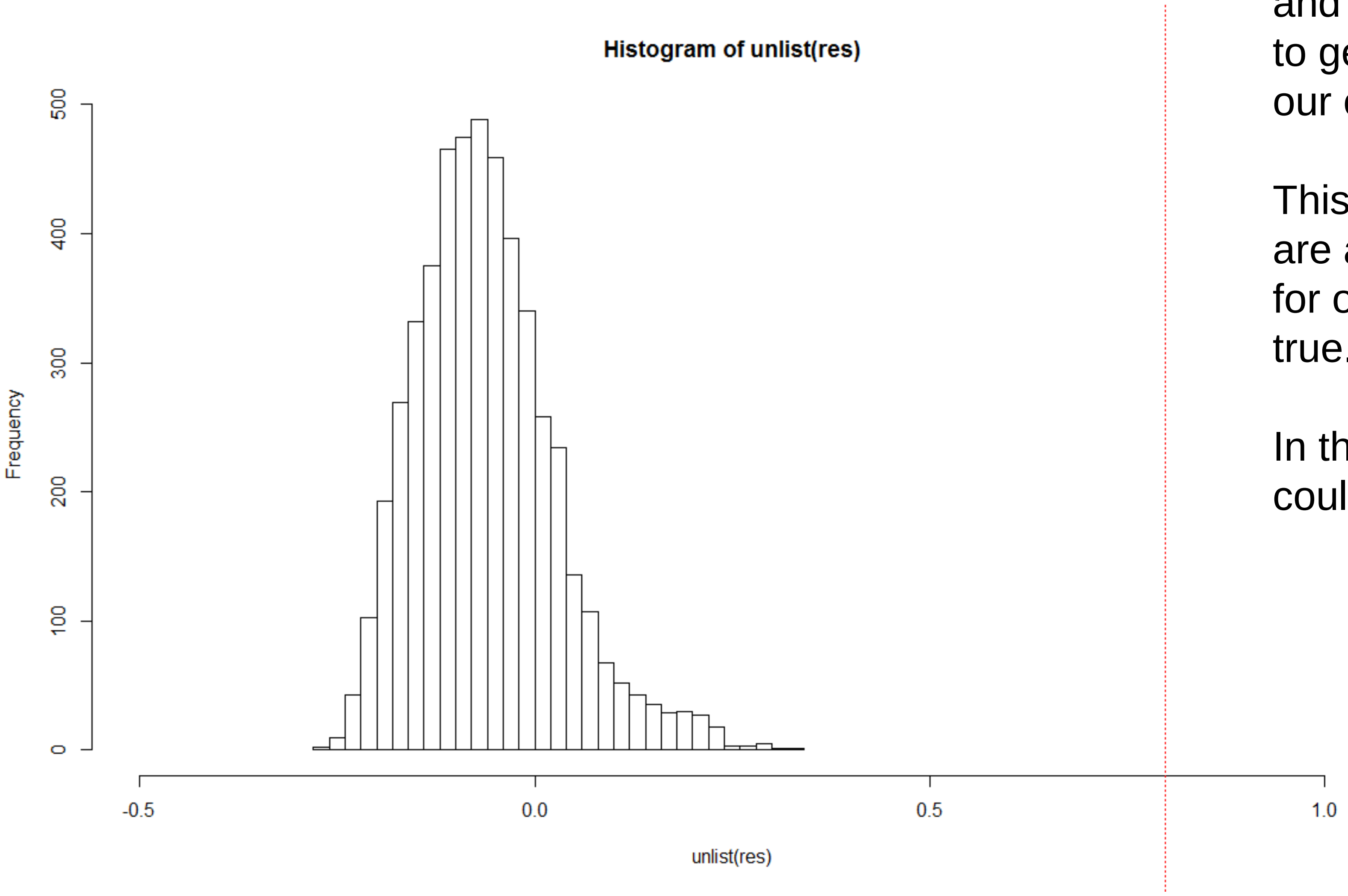
original data



permuted data



Significance testing assortativity: Node Based Permuations

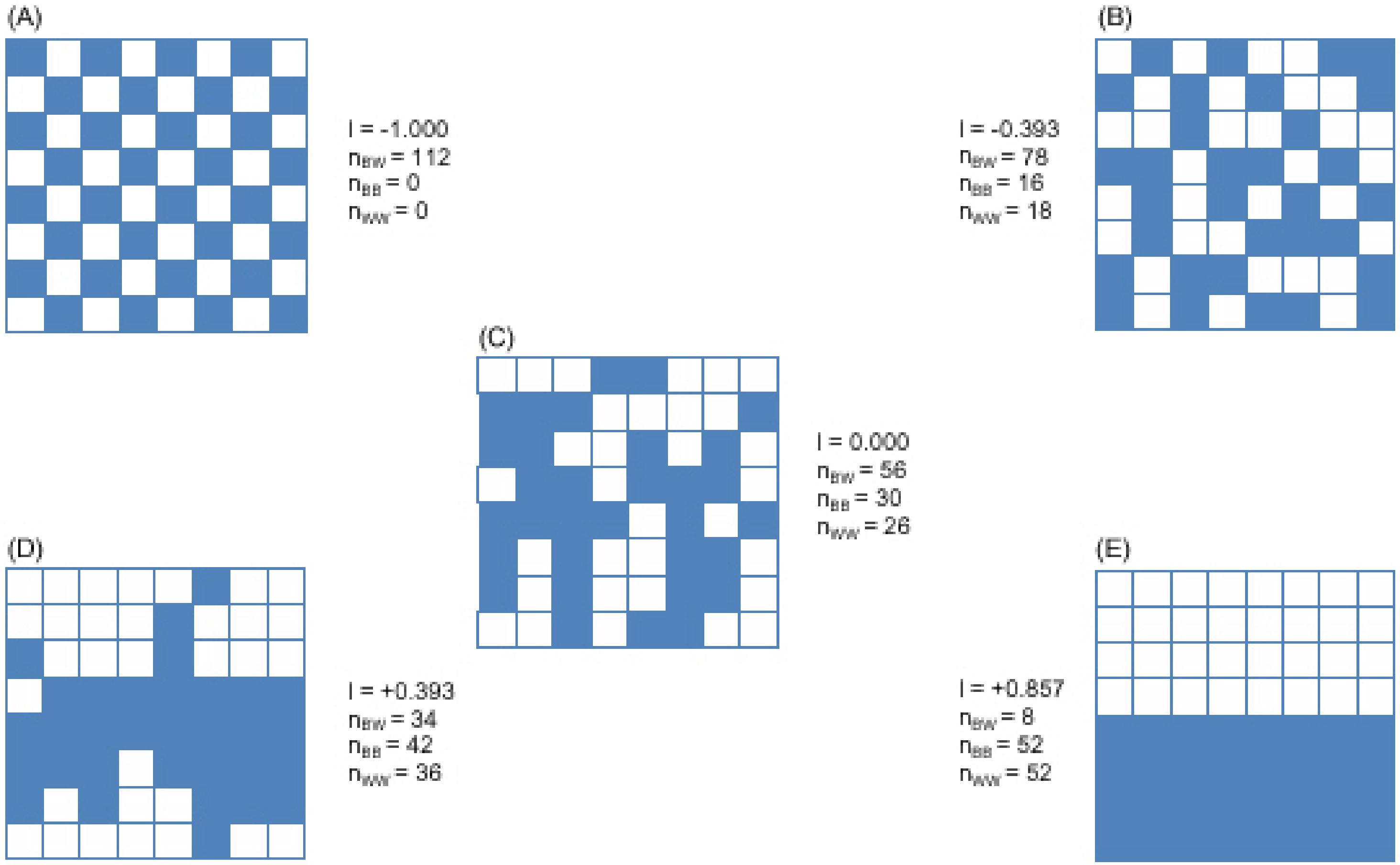


Repeating this process 1000s of times and recalculating this value enables us to generate a p-value of how unexpected our observed value is.

This assumes that node permutations are an unbiased choice of permutation for our data – this may not always be true.

In this case a pre-network randomization could be done.

Spatial Autocorrelation – Moran’s I



Moran's I (ranges from -1 to 1 , similar to Pearson's r)

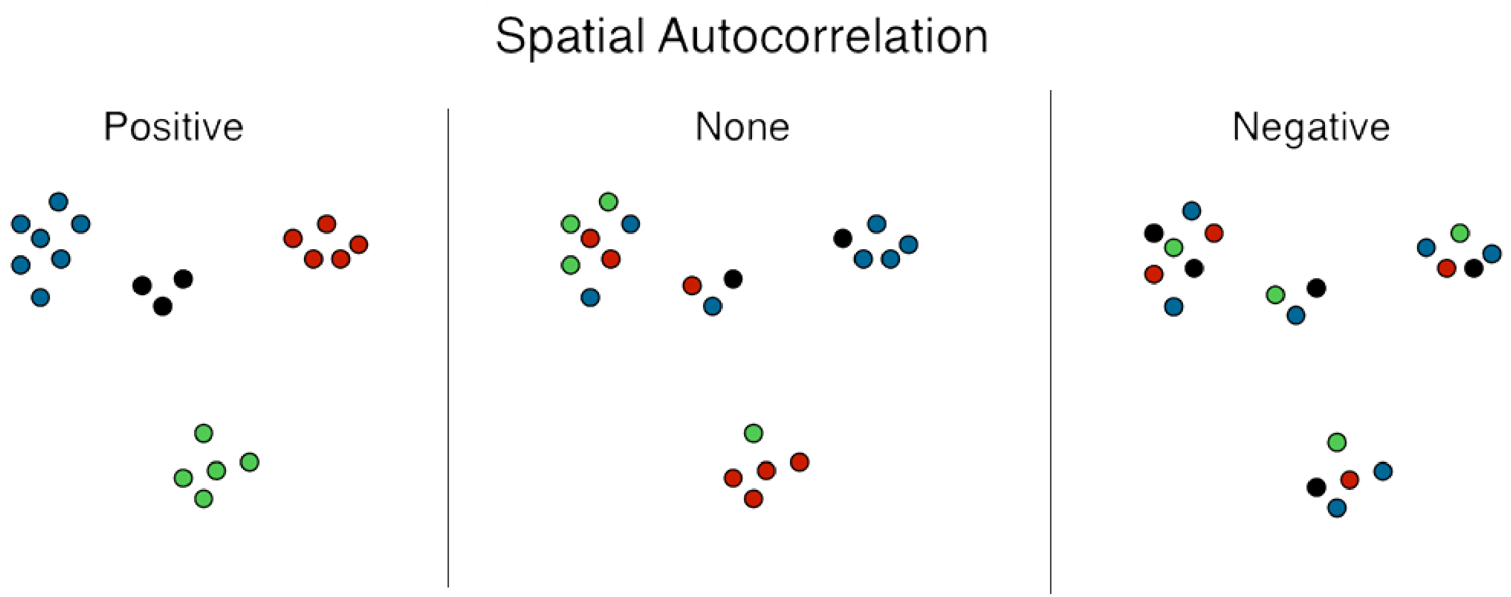
-1 = negative autocorrelation

+1 = positive autocorrelation

0 = no correlation (independent of each other)

In networks we can assess Moran’s I at different geodesic distances from each node.

Spatial Autocorrelation – Geary's C



Geary's C ranges from 0 to 2.

<1 = positive autocorrelation

~1 = no autocorrelation (network/attributes are independent)

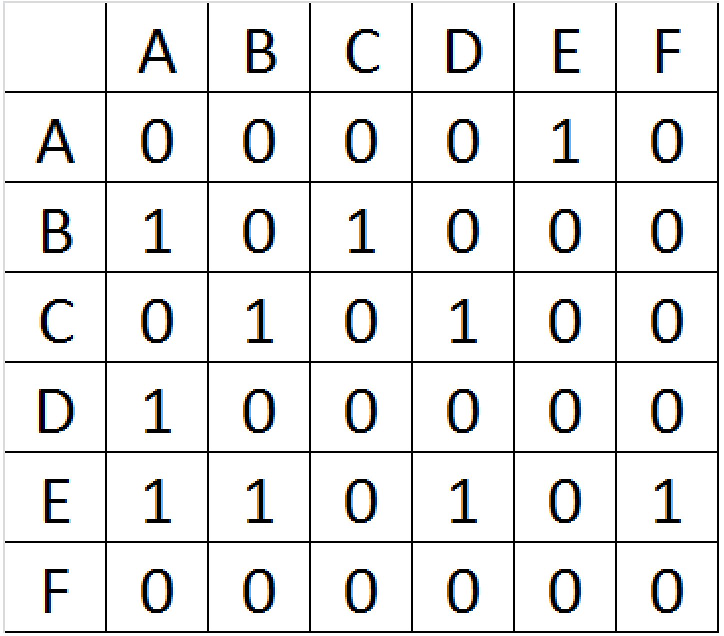
>1 = negative autocorrelation

The main difference between Moran's I and Geary's C is that Geary's C emphasizes autocorrelation at a more local level than Moran's I

In networks we can assess Geary's C at different geodesic distances from each node.

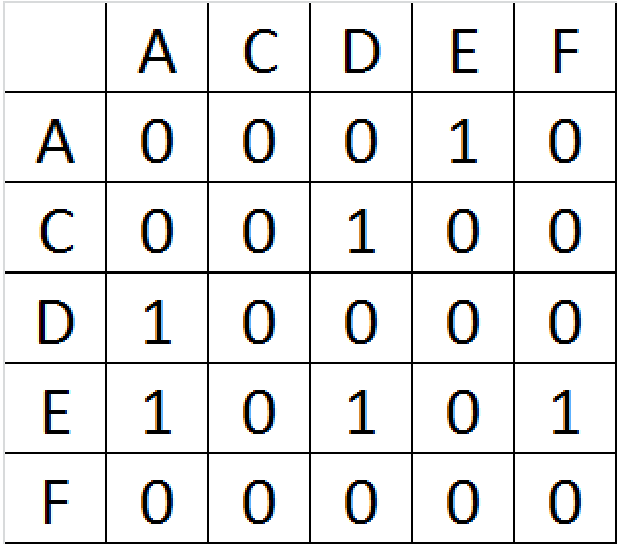
Random Models - Intro

Jackknifing



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

Original



	A	C	D	E	F
A	0	0	0	1	0
C	0	0	1	0	0
D	1	0	0	0	0
E	1	0	1	0	1
F	0	0	0	0	0

Jackknifed

- Jackkifing involves removal of one node at a time, from which the graph is reconstructed

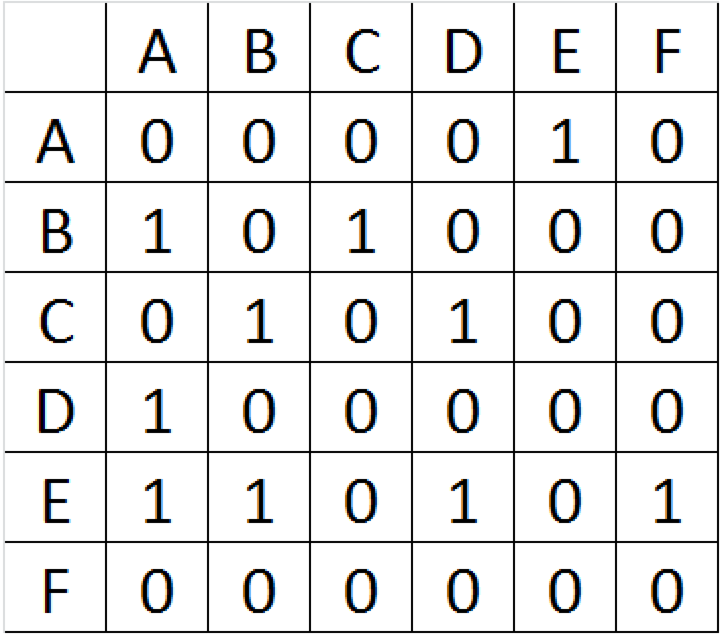
- Generating sampling distributions of our observed network

- Use these to calculate standard errors of observed descriptive statistics

- Can also use these to generate p-values for how ‘unexpected’ our observed descriptive statistics are (but not recommended)

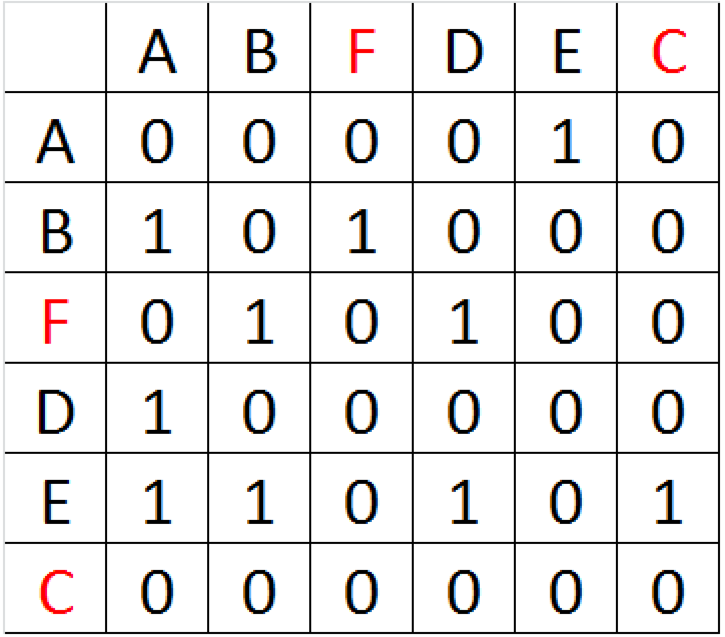
Tom A.B. Snijders & Stephen P. Borgatti, 1999

Node Permutations



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

Original



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

Permuted

- Such permutations would be repeated 1000s of times

- Generating sampling distributions of our observed network

- Use these to calculate standard errors of observed descriptive statistics (can also be used to calculate t-statistics)

- Can also use these to generate p-values for how ‘unexpected’ our observed descriptive statistics are

Tom A.B. Snijders & Stephen P. Borgatti, 1999

One common method is to compare our observed findings to those from standard random graphs:

- Conditional Uniform Graphs (CUGs)
- Classic Random graphs
- ERGMs

Conditional Uniform Graphs (CUGs)

CUGs can be used to test how typical some observed network metric is for a given set of networks. The general routine is as follows:

1. Calculate an observed value of some network metric.

2. Generate many networks that share some characteristic in common with the original network.

3. Compare the observed network metric with the same metric recalculate over all generated networks.

The key question is what characteristics to simulate networks on? There are three default options using the ‘sna’ package:

a) size (number of nodes)
b) number of edges (density)
c) the distribution of dyads (size + density + reciprocity)

As we cannot possibly produce all possible permutations of the graph, we randomly sample from a uniform distribution of graphs that share these properties.

Random Graphs

1. Classic Random Graphs - e.g. Erdos-Renyi (constant probability of an edge between any pair of nodes):

    - large component
    - Poisson degree distribution
    - Low average path length
    - Low clustering

2. Small World Graphs – e.g. Watts & Strogatz (one end of each edge is independently and with probability p rewired to another node)

    - high clustering
    - small distances between nodes

3. Preferential Attachment Models – e.g. Barabasi-Albert model (network grows over time with edges connecting to individuals preferentially, e.g. well connected nodes are preferentially attached to)

    - Discrete time step model
    - Common in many large networks

Randomization Basic Method:

1. Generate the social network from the observed data

2. Calculate and record the test statistic, using conventional statistics such as linear (mixed effect) models on the data from the observed network

3. Randomize the observed data and generate a ‘random’ social network

4. Calculate and record the test statistic, using the exact same model as in 2, but on the random social network

Farine 2014

Why null models ?

- Network data violate independence of parametric statistics by their very nature

- Often what we are studying is the population – not a sample

- Randomization can be used to account for this non-independence

- There are usually other sources of non-independence in data also (e.g. space, time) and these should be accounted for in the randomizations.

- Essentially, the researcher is aiming to produce randomized blocks of pseudo-replicated data. This will enable them to generate a realistic null distribution for use in significance testing.

Null models

- Data sets that are based on observed data in some way

- They may be randomizations of original data (2 main types for static networks: node & data-stream).

- Or they may be based on models based on properties of the original data

- They should strive to maintain constant all other aspects of the data that are not directly relevant to the hypothesis.

Critical issue regarding null models is whether they accurately reflect the structure of the original data – or are they biased in some way?

2 types of Data Randomizations in Static Networks based on original data:

1. Node-based randomizations (Network randomization)

e.g. randomizing attributes of nodes, but maintain the same number of each class. An example would be randomizing gender/sex among nodes. This does assume the observed network is a very good representation of the true network. Farine 2014 has identified that violation of this assumption can lead to higher type I and type II errors.

2. Data Stream-based randomizations (Pre-network randomization)

Sequential swaps between individuals constrained by e.g. time or space.   These swaps can occur at the individual level but may also be at the group level.