Network Measures

Variable	Definitions
N	Number of nodes
E	Number of edges
L	Average shortest path length
D	Diameter
C_i	Clustering Coefficient
k, k^{in}, k^{out}	Degree, in-degree, and out-degree
b	Betweeness centrality
c	Closeness centrality
x	Eigenvector centrality
r	Assortativity
Q	Modularity

There are so many more (e.g., http://schochastics.net/sna/periodic.html).

G(N,E)

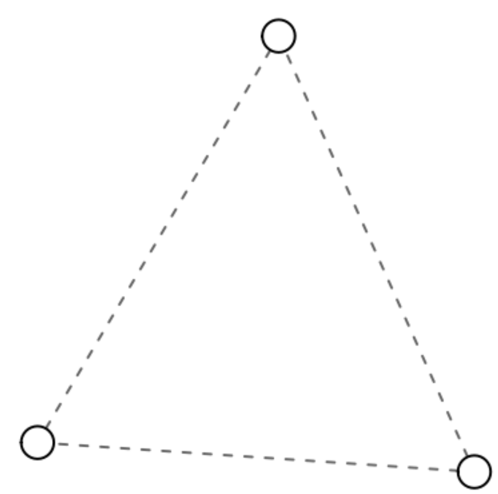
- How many edges in a simple network of N nodes?

	1	2	3	4	5	6
1	0	0	0	1	1	1
2	0	0	0	0	0	1
3	0	0	0	1	0	1
4	1	0	1	0	1	1
5	1	0	0	1	0	0
6	1	1	1	1	0	0

Can you work out a general rule?

Birthday paradox

- What’s the probability at least two people in this room share a birthday?
- Why is this a network problem?



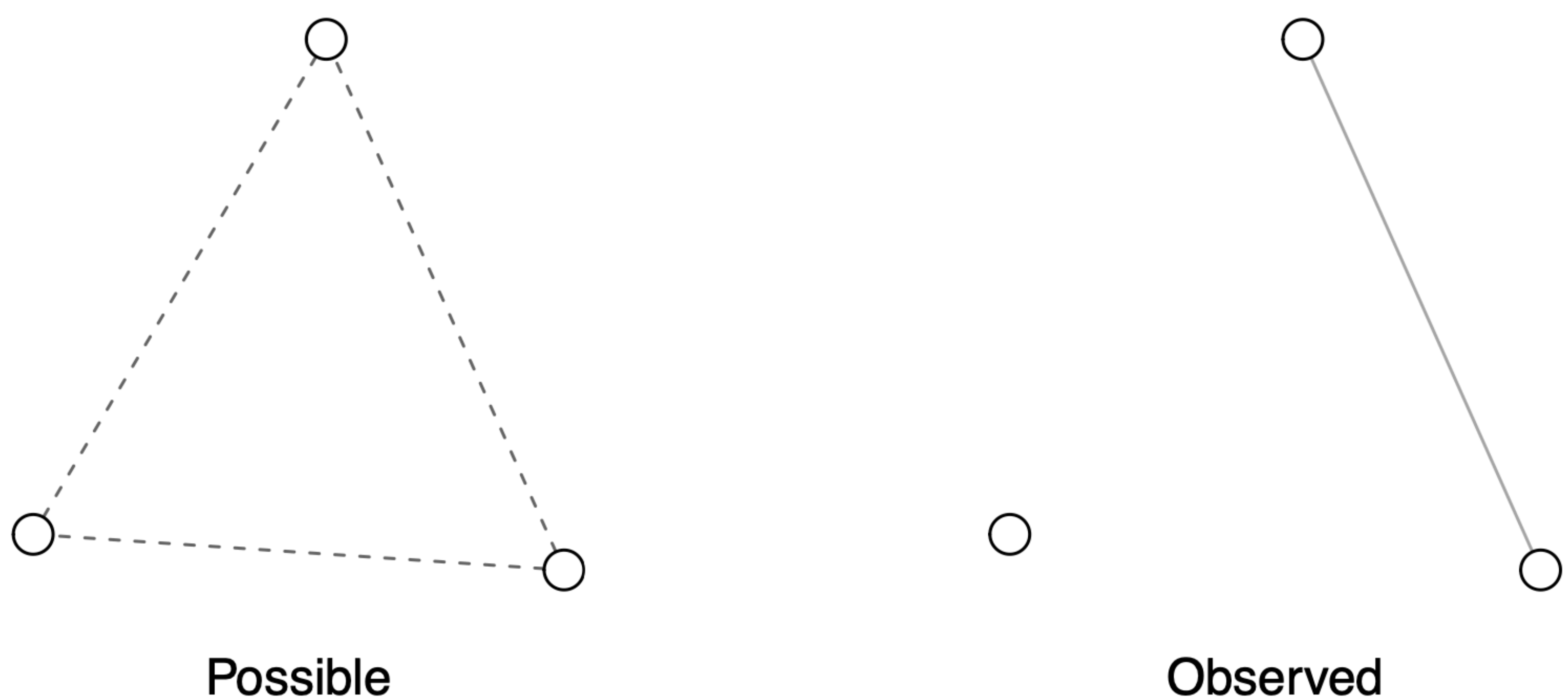
Number of possible edges:
E=N(N-1)/2

Probability of not sharing an edge:
(1-1/365)^E

Density

$$\rho = \frac{2E}{N(N-1)}$$

- The number of observed edges over the number of possible edges.



What does density tell us?

- What kinds of networks are likely to be low density?
- High density?

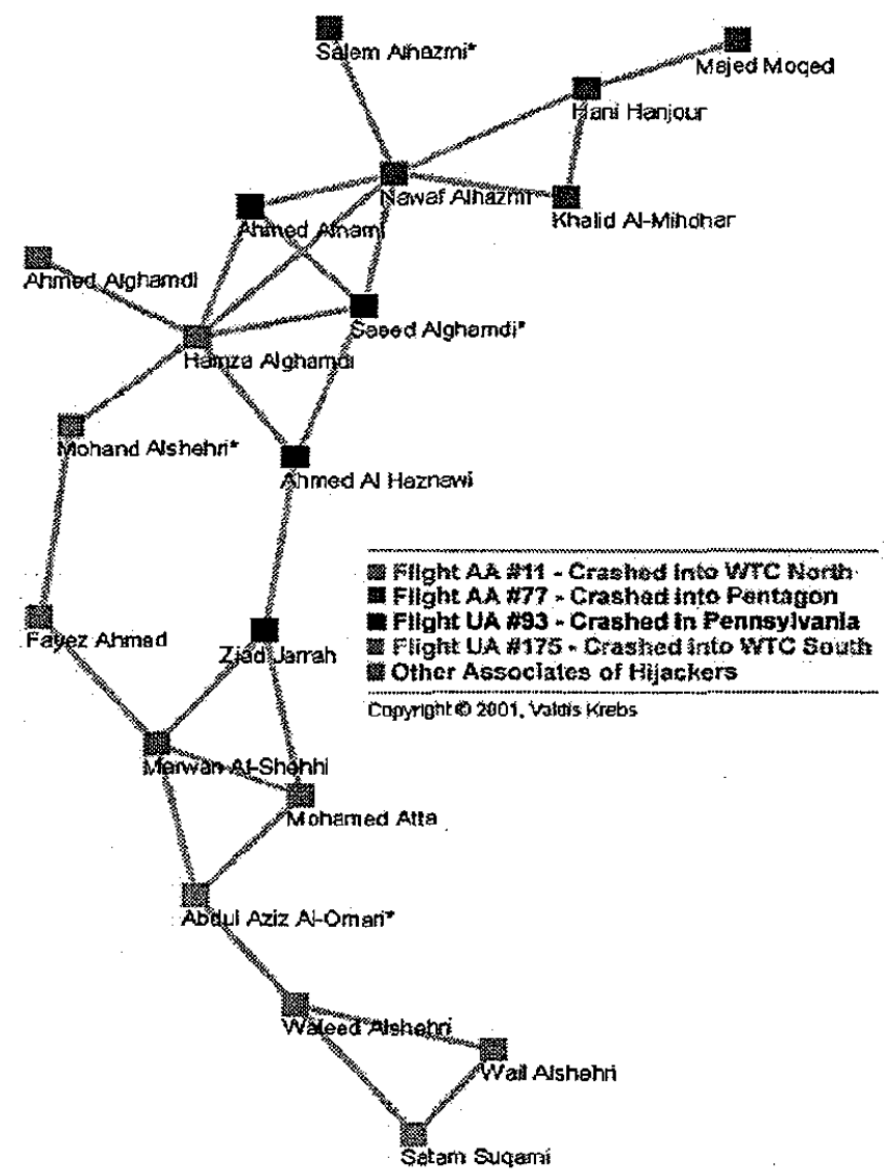
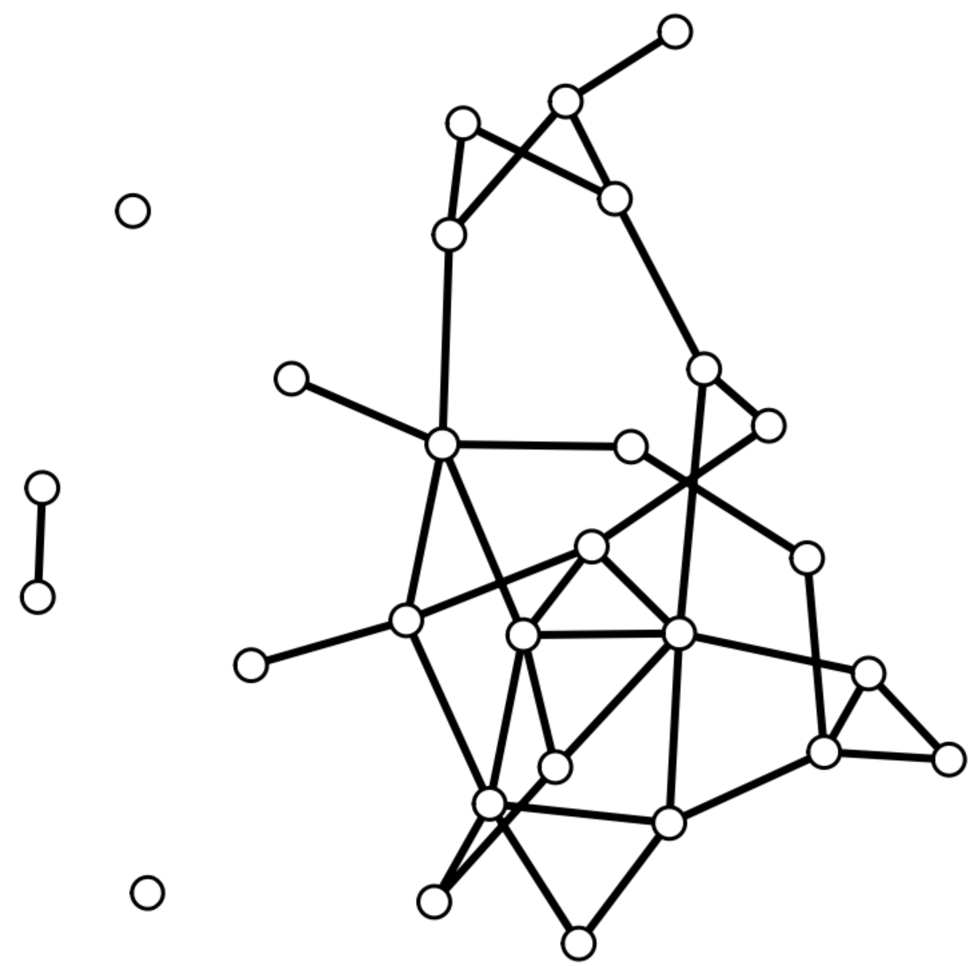


Figure 2 Trusted Prior Contacts

Krebs (2002) Mapping Terrorist Cells

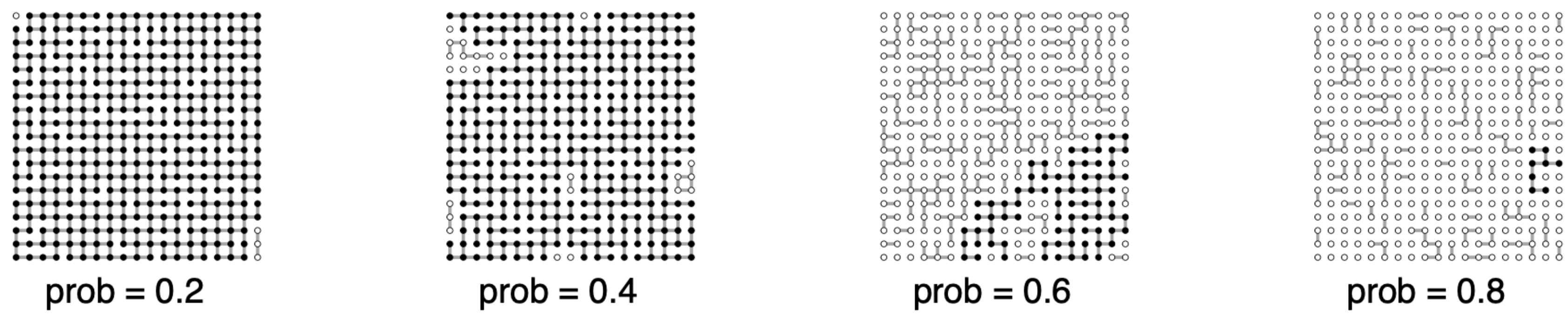
Components

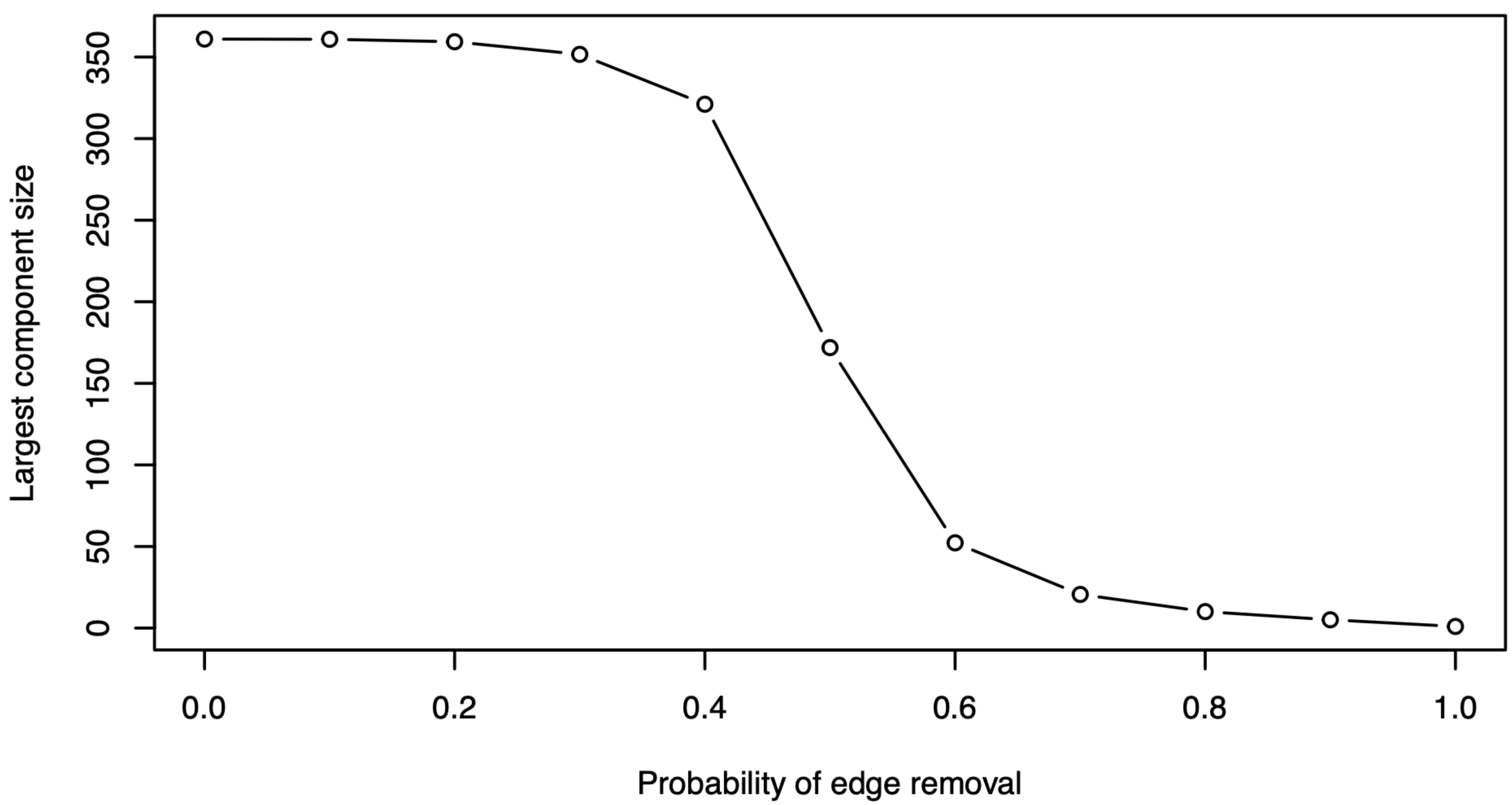
- Component: Collection of nodes that are all ‘reachable’ via a path of edges.



Percolation analysis

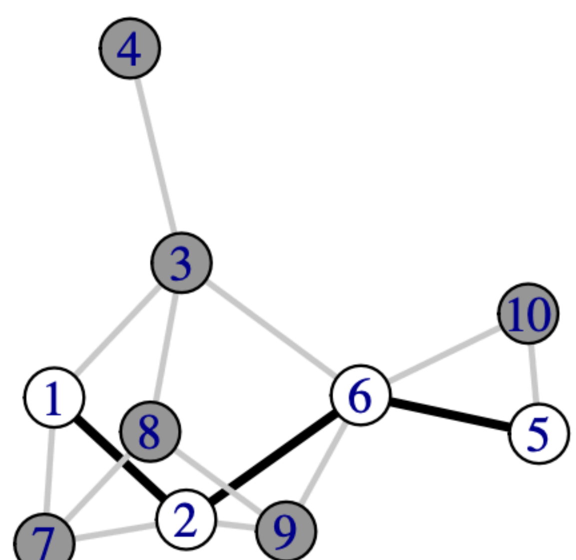
Probability of killing edges in a lattice—how many components?



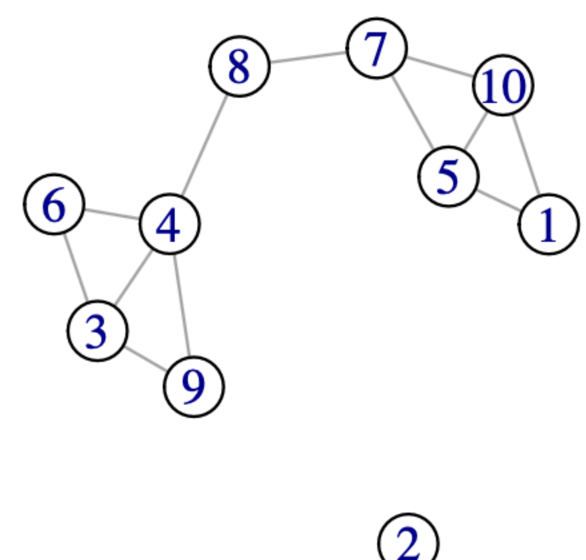


Path Lengths

- Path—a series of contiguous edges.
- Shortest path length (geodesic)
- Diameter—longest shortest path length in a network

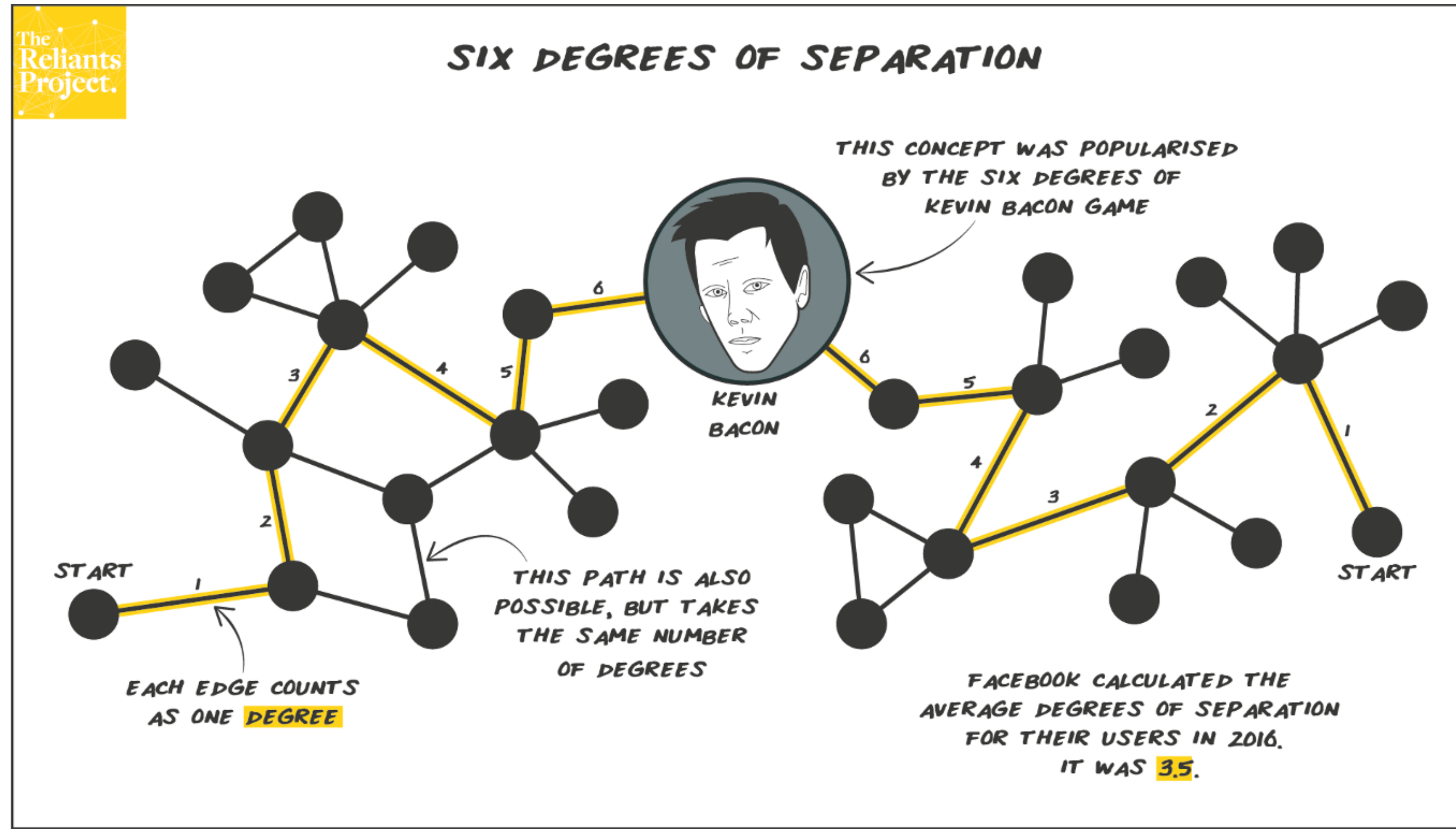


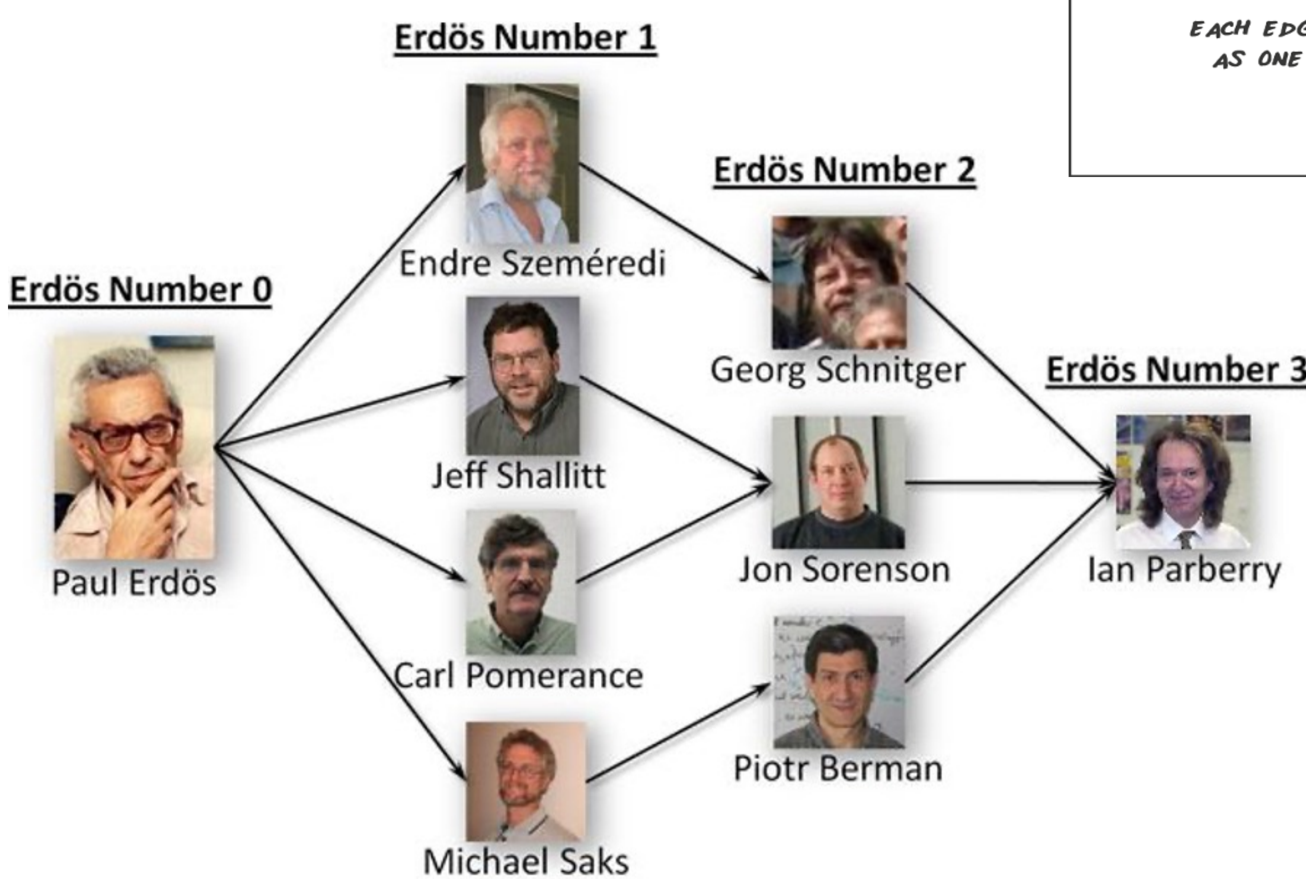
$d_{1,5} = 3$



$d_{4,2} = \infty$

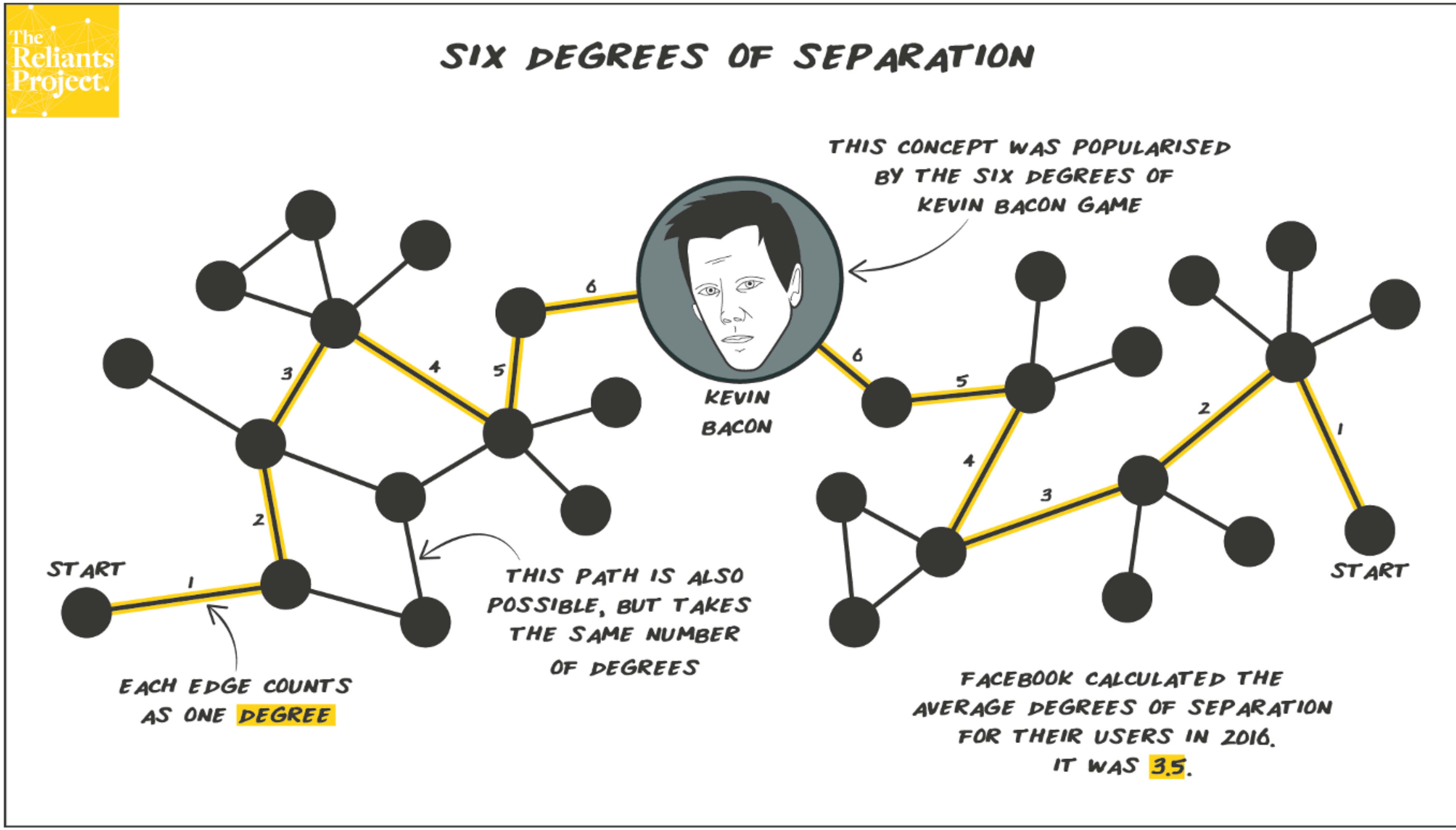
Six Degrees of Separation



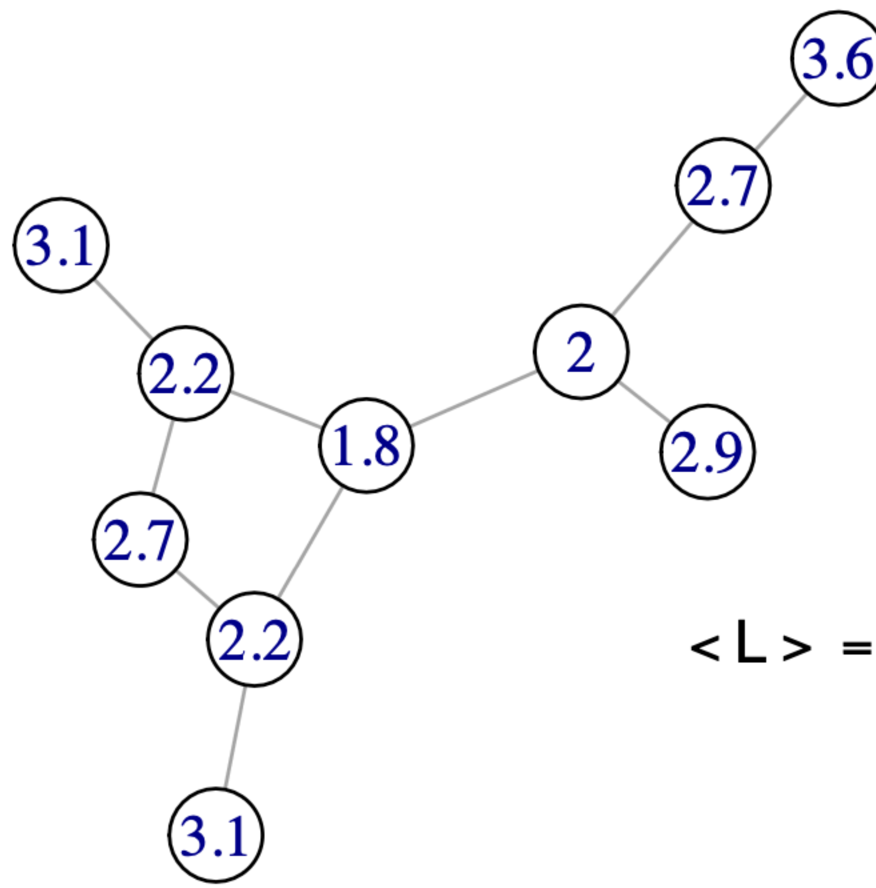


Hills:Adler:Levin:Durrett:Chung:Erdös

Oracle of Bacon

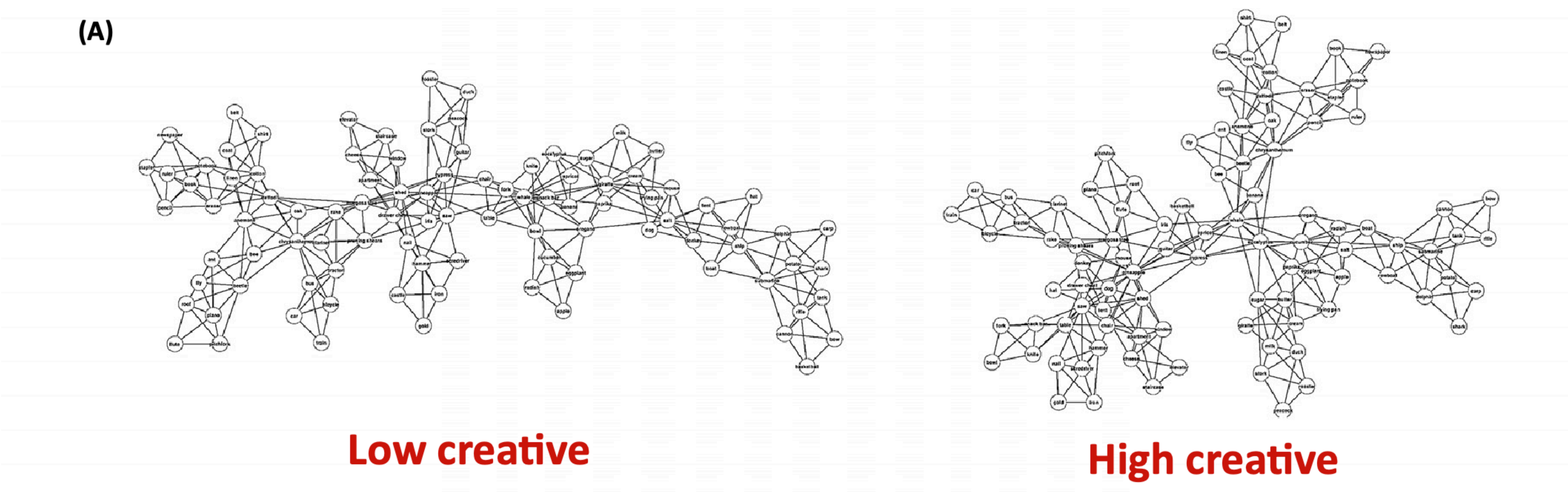


Average shortest path length



$< L > = 2.62$

Average shortest path lengths are shorter among ideas in more creative people

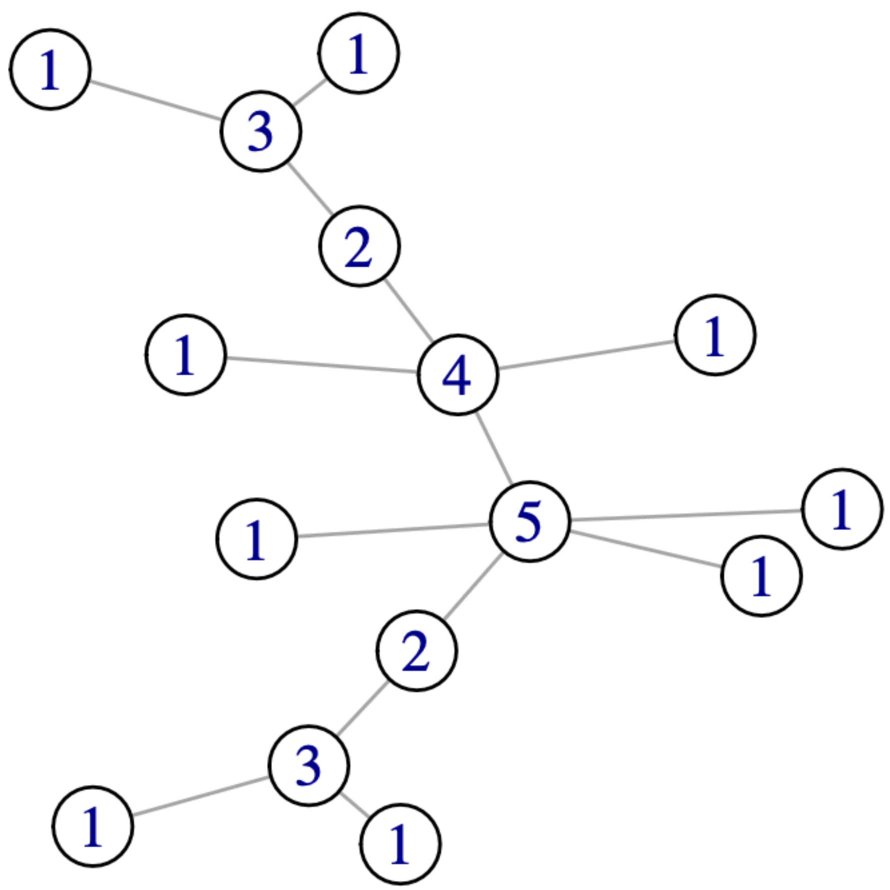


Kenett & Faust (2019)

Centrality

- Centrality is a measure of node importance.
- There are many centrality measures:
- Degree
- Betweenness
- Closeness
- Eigenvector centrality/PageRank
- And many more

Degree centrality



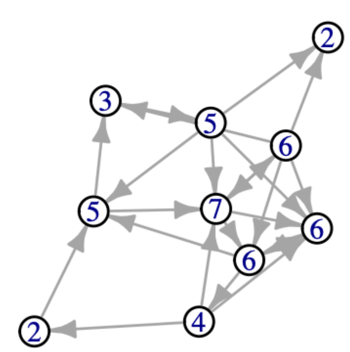
$$k_i = \sum_j A_{i,j} = \sum_j A_{j,i}$$

$$k_i^{in} = \sum_j A_{j,i}$$

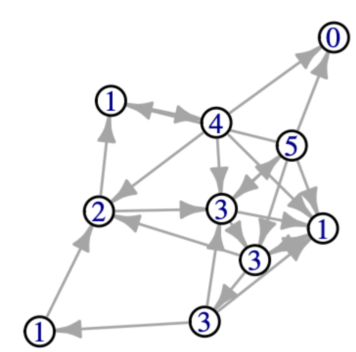
$$k_i^{out} = \sum_j A_{i,j}$$

$$< k >= \frac{1}{N} \sum_{i=1}^{N} k_i$$

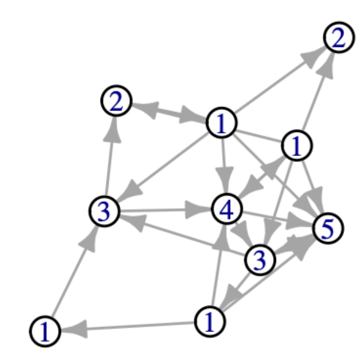
Average degree



Total



Outdegree



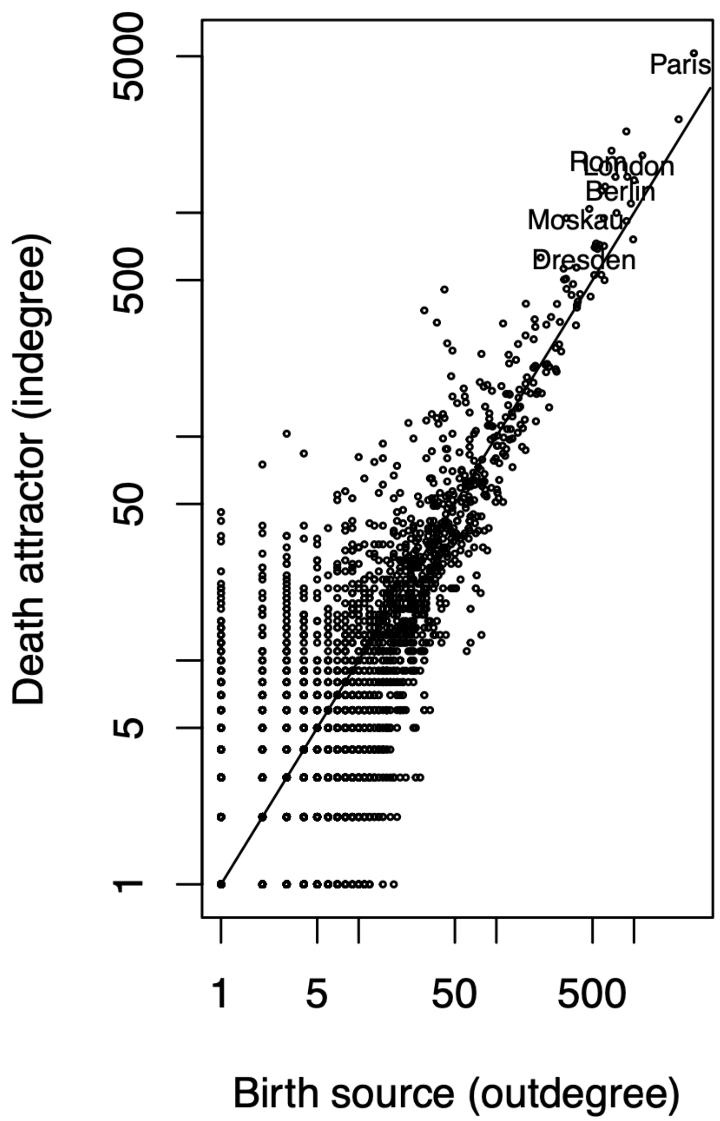
Indegree

QUANTITATIVE SOCIAL SCIENCE

A network framework of cultural history

Maximilian Schich,[1,2,3*] Chaoming Song,[4] Yong-Yeol Ahn,[5] Alexander Mirsky,[2] Mauro Martino,[3] Albert-László Barabási,[3,6,7] Dirk Helbing[2]

The emergent processes driving cultural history are a product of complex interactions among large numbers of individuals, determined by difficult-to-quantify historical conditions. To characterize these processes, we have reconstructed aggregate intellectual mobility over two millennia through the birth and death locations of more than 150,000 notable individuals. The tools of network and complexity theory were then used to identify characteristic statistical patterns and determine the cultural and historical relevance of deviations. The resulting network of locations provides a macroscopic perspective of cultural history, which helps us to retrace cultural narratives of Europe and North America using large-scale visualization and quantitative dynamical tools and to derive historical trends of cultural centers beyond the scope of specific events or narrow time intervals.



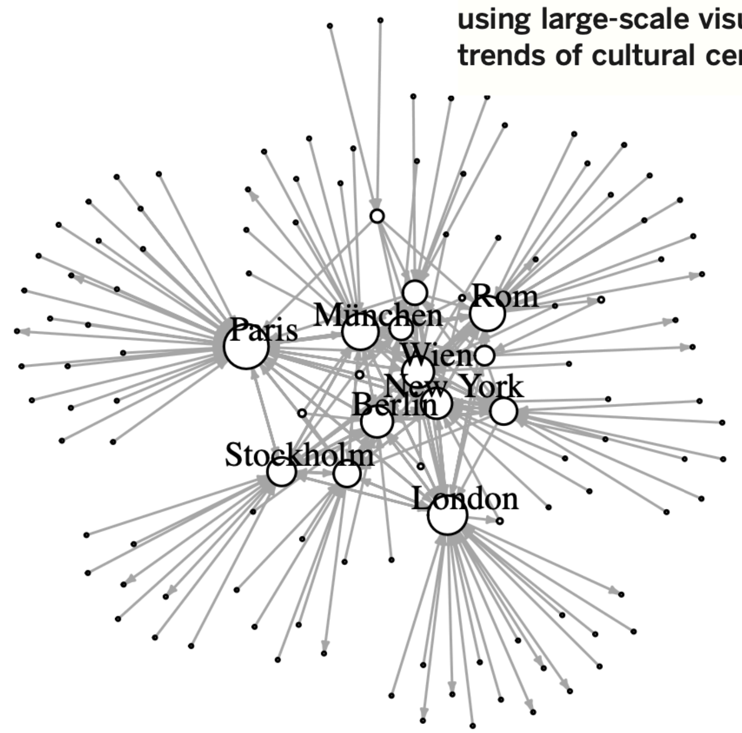


Figure 1: Indegree and Outdegree as birth sources and death attractors for 112,276 notable historical fine artists, c.480bc to 2010ad. There are $n = 46189$ nodes (i.e., places) in the network and 112,276 edges. (Left) Dots represent places. Places above the line are cultural centres where artists tended to be attracted. (Right) A representative subset of the network, showing 13 places with a total degree (in + out) of more than 500 plus a sampling of 117 additional nodes that are connected to them. Data is from the General Artist Lexicon (Beyer, Savoy, and Tegethoff 2016) and the figure is after Schich et al., 2014.

Weighted degree (strength)

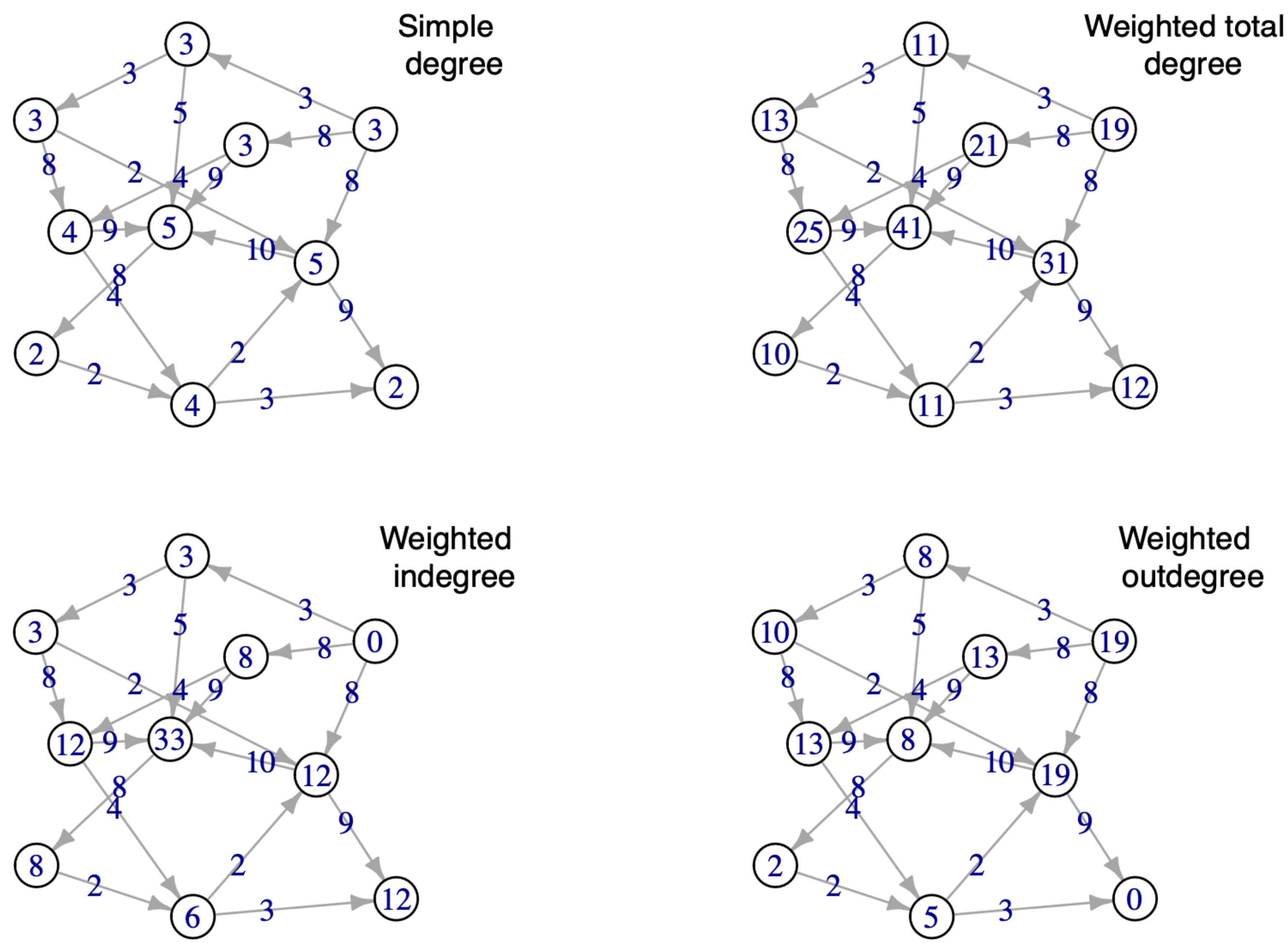
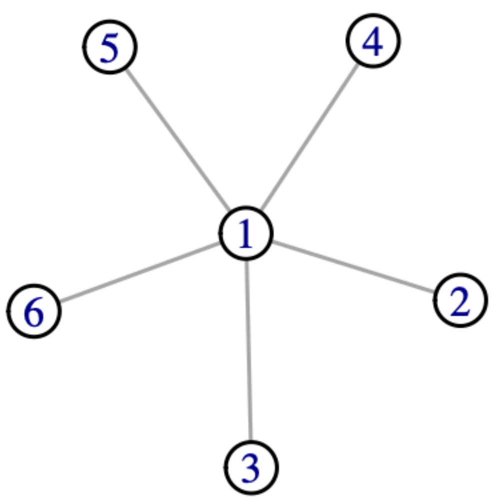


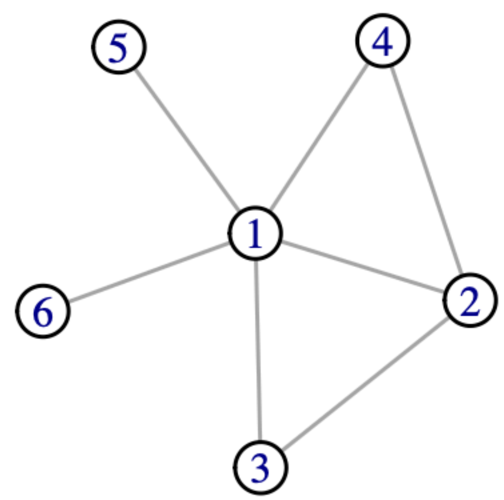
Figure 9: Various ways of computing the degree in directed weighted networks. Nodes are labeled with their relevant degree.

Clustering Coefficient

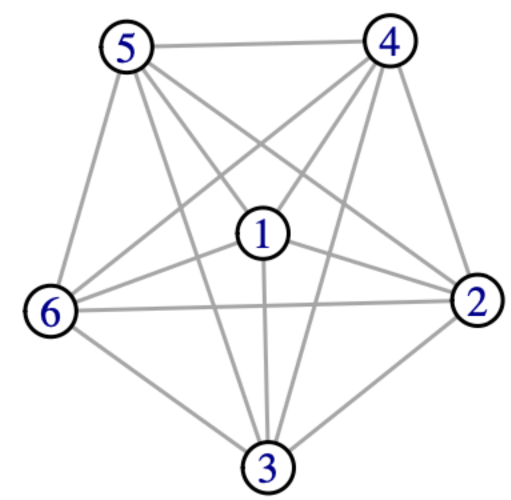
The clustering coefficient has two forms. The first is a node-level or local clustering coefficient. This measures the proportion of a node's neighbors that are connected by an edge.



C=0



C=.2



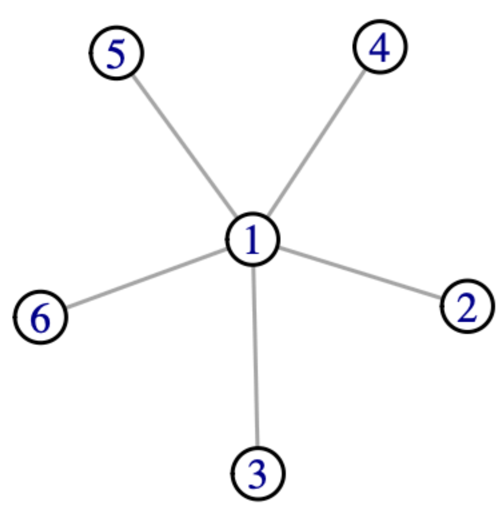
C=1

$$C_i = \frac{2e}{k_i(k_i - 1)}$$

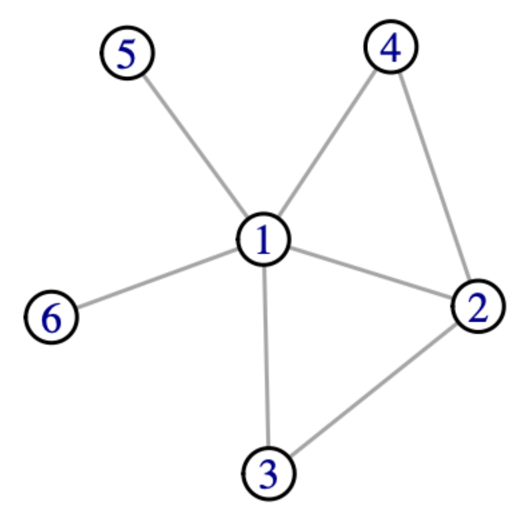
Clustering Coefficient

(Node level)

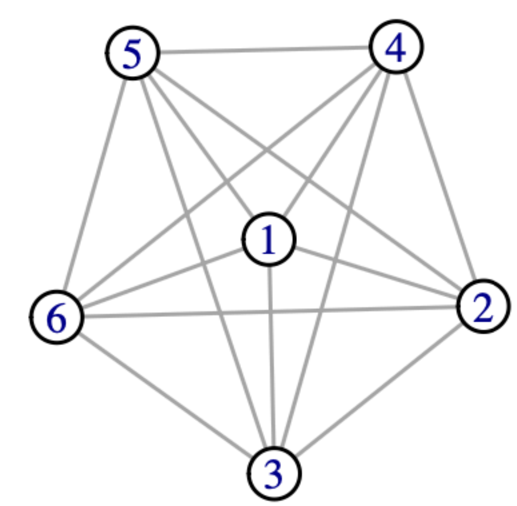
The clustering coefficient has two forms. The first is a node-level or local clustering coefficient. This measures the proportion of a node's neighbors that are connected by an edge.



C=0



C=.2



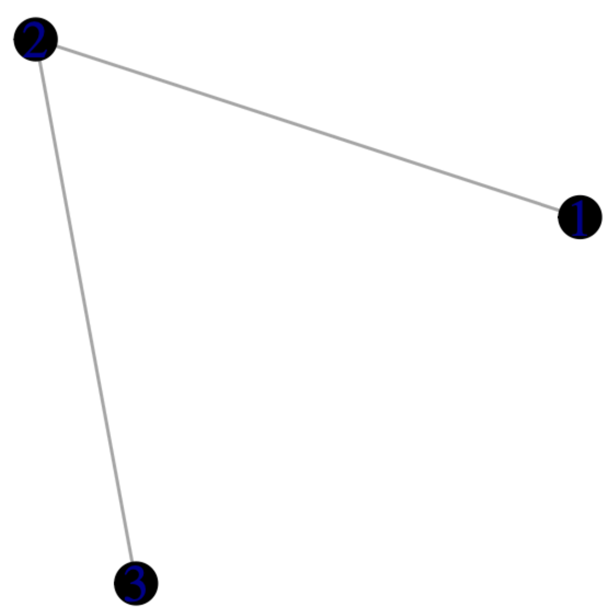
C=1

$$C_i = \frac{2e}{k_i(k_i - 1)}$$

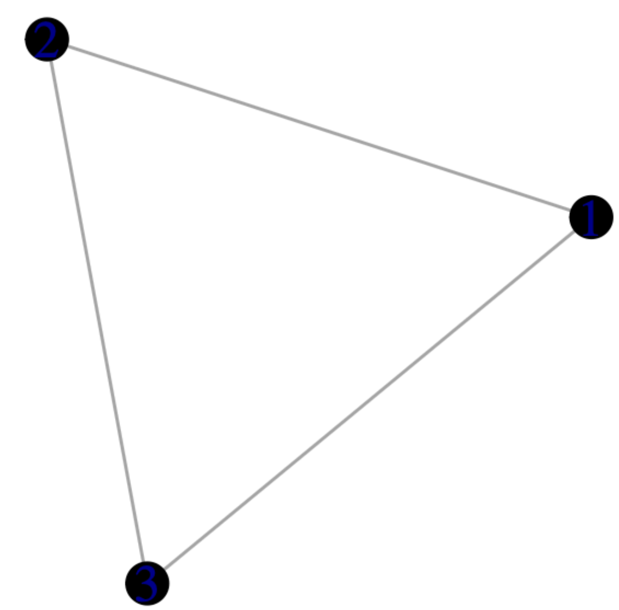
One can compute the average C over all nodes

Clustering Coefficient
(Transitivity: graph level)

Transitivity measures the proportion of triplets in the network that are transitive (i.e. a triangle).



Intransitive triplet



Transitive triplet

$$T = \frac{3\Delta}{\Lambda}$$

Transitivity is a graph level metric

Clustering coefficient and transitivity can diverge (node vs. graph level view)

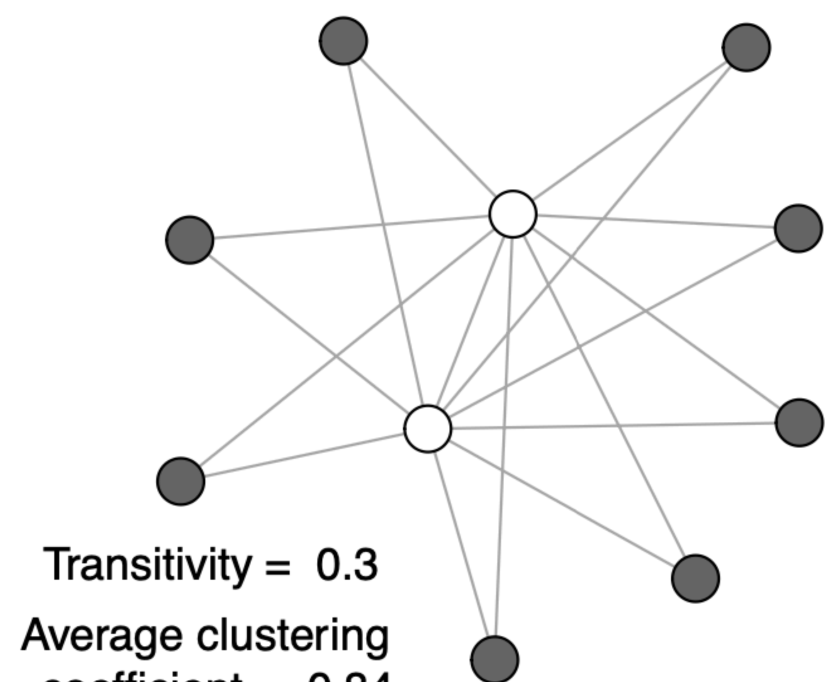
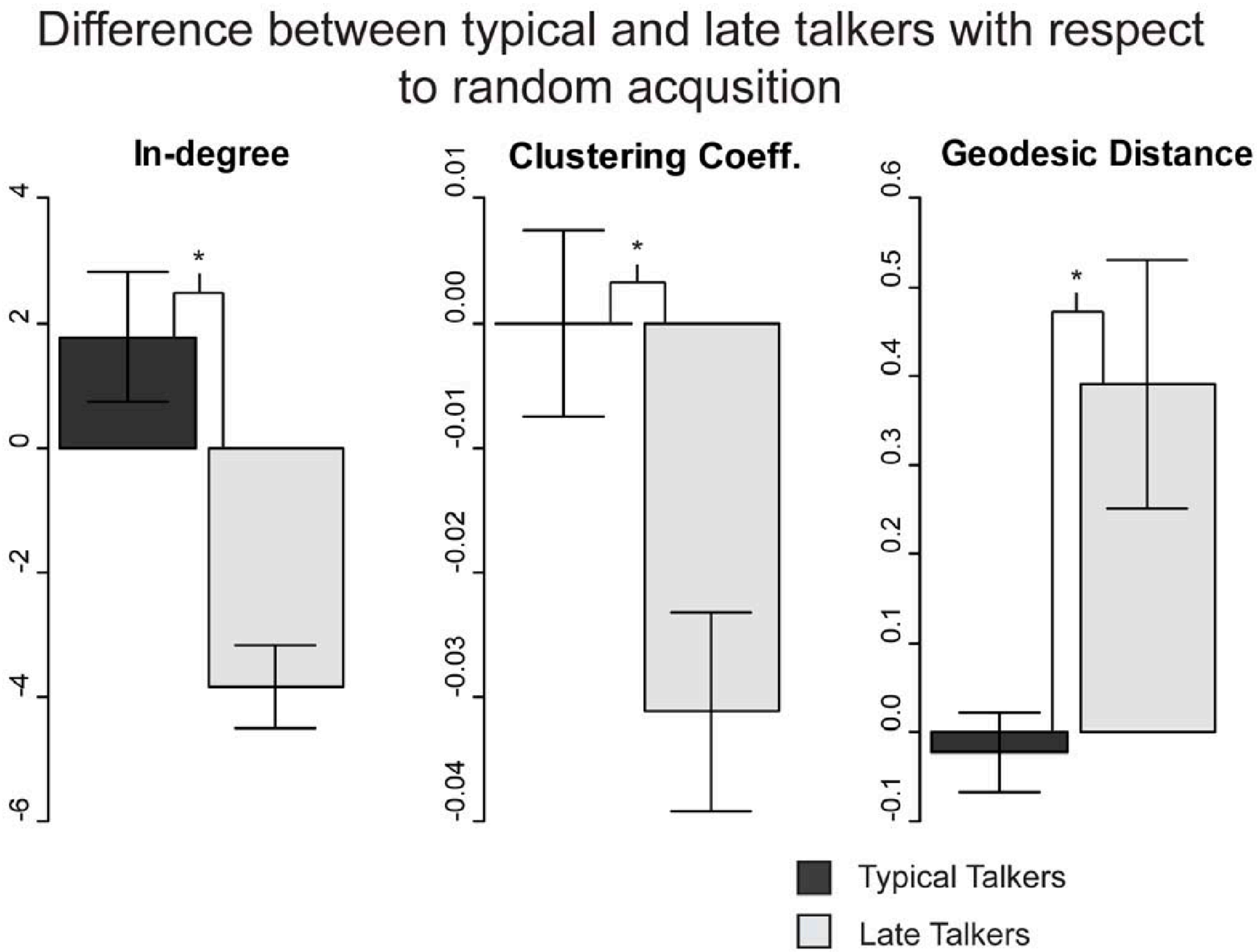


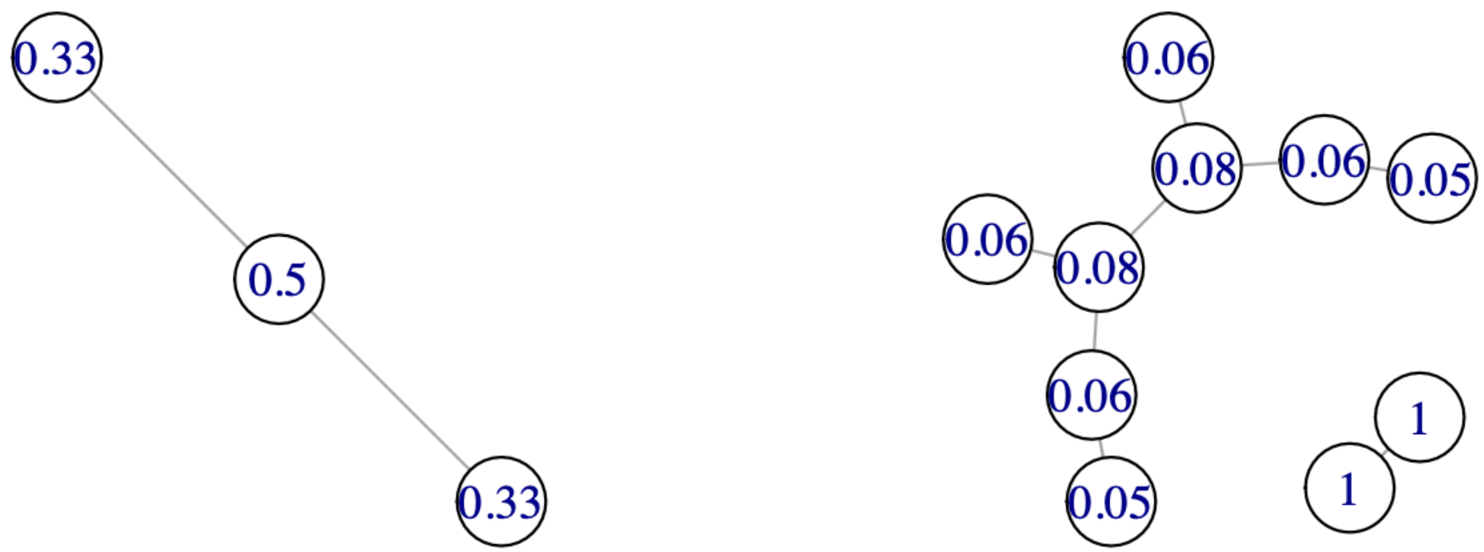
Figure 12: The wheel network demonstrates the difference between transitivity and average clustering coefficient. As the outer nodes increase, the average clustering coefficient approaches 1 and the transitivity approaches 0.

Clustering coefficient and the words children learn



Late talkers have lower degree and lower clustering coefficient and have average shortest path length (ASPL = geodesic distance)

Closeness centrality

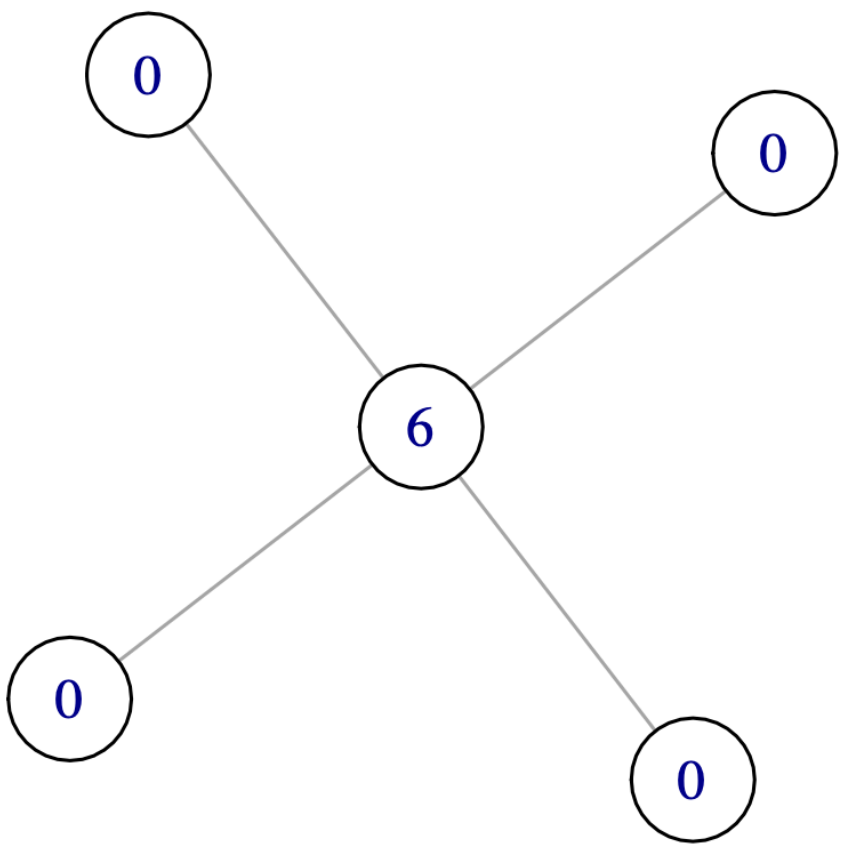


Each node is labeled with its closeness centrality

$$c_i = \frac{1}{\sum_{j=i}^{N} d_{ij}}$$

Betweenness centrality

The betweenness centrality for a node i is the number of shortest paths between all other pairs of nodes that pass through node i.

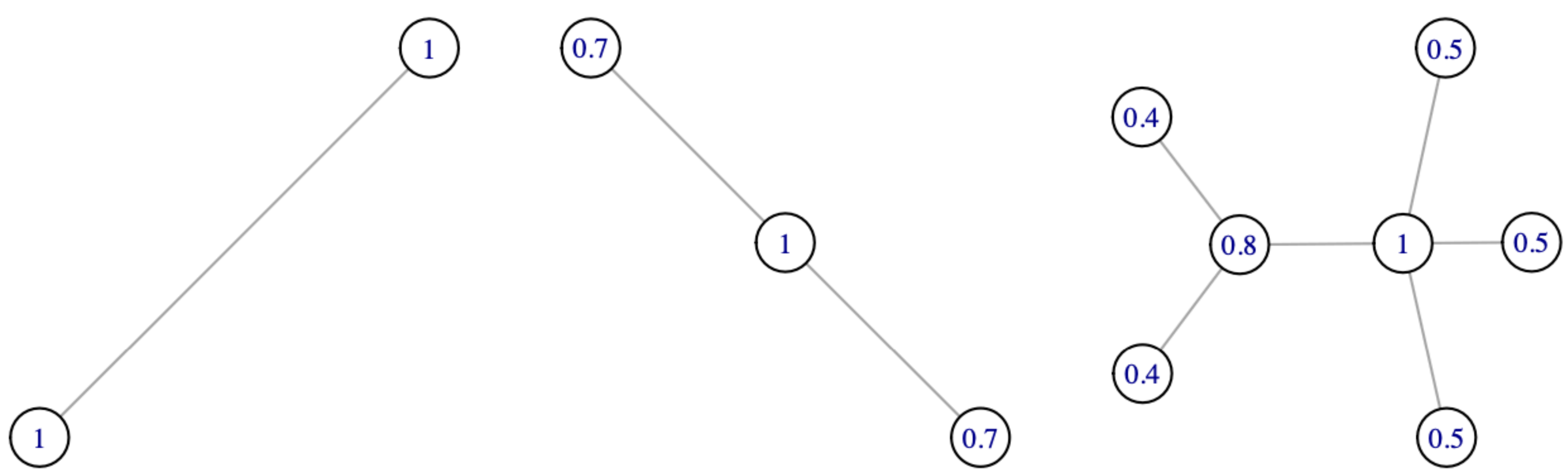


$$b_i = \sum_{i \neq j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

6 possible paths, all pass through the central node.

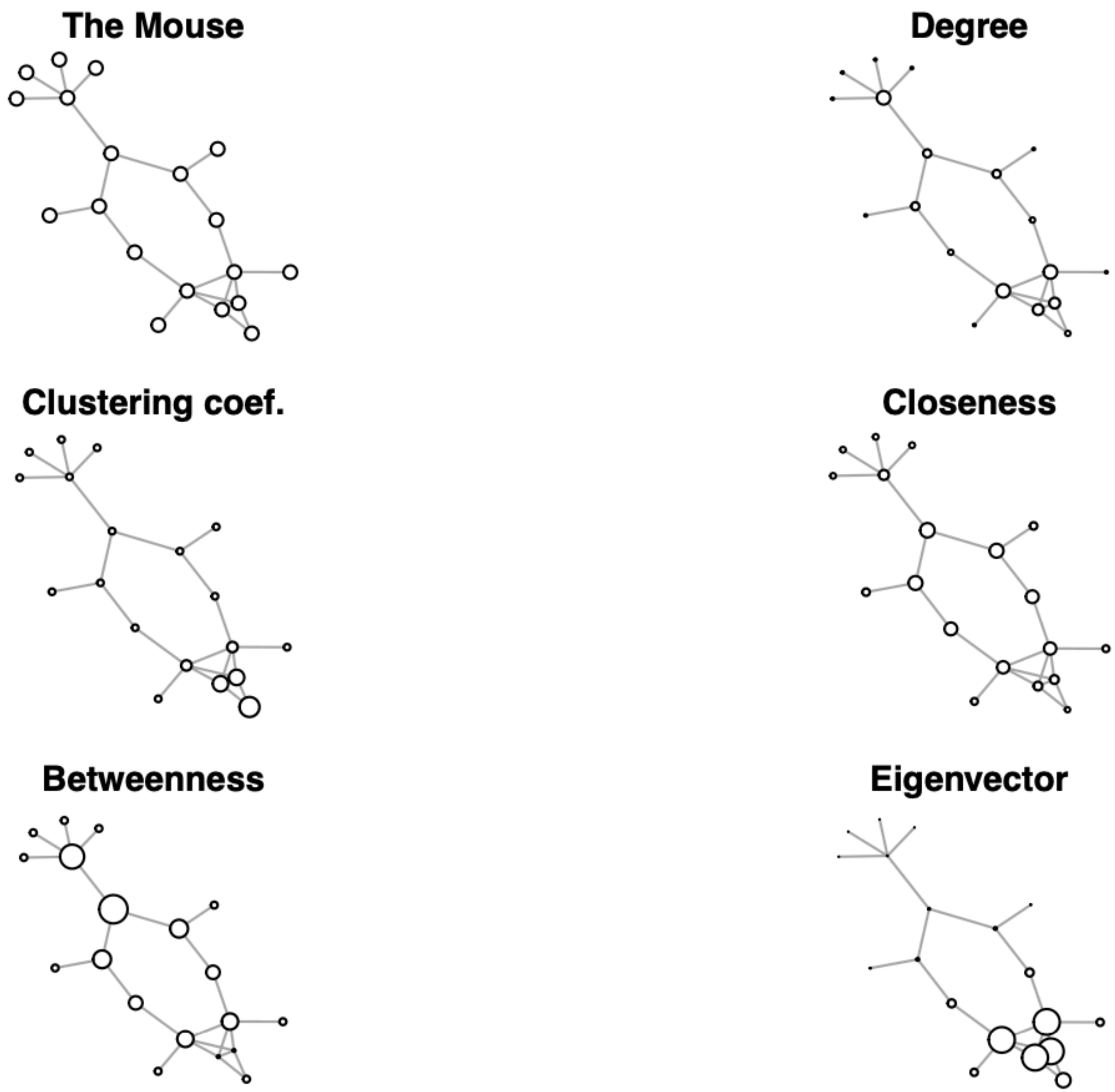
Eigenvector centrality

Eigenvector centrality is analogous to prestige. To be prestigious, one must receive prestige from other nodes. The more prestigious the nodes one receives prestige from, the more prestige one receives. The definition is recursive: It requires that we know how prestigious each node is before we can compute the prestige of any node.



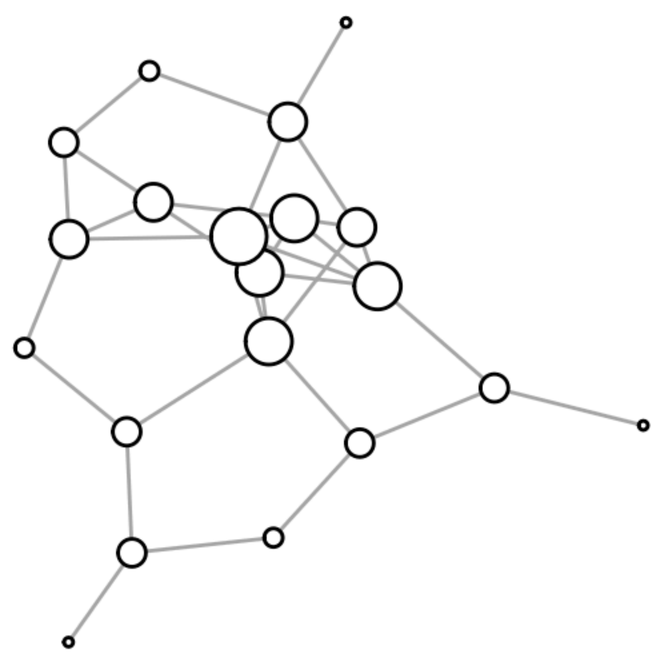
This measure is the basis of PageRank and Katz centrality—both look at how nodes recursively give and receive 'value' to their neighbours.

Measures of centrality

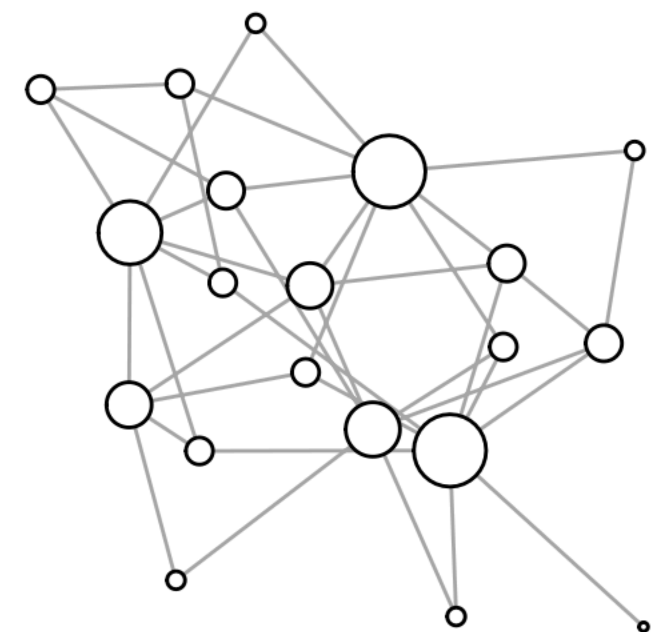


Assortatitivity

Assortativity evaluates the degree to which nodes with similar properties connect with each other. In social networks, this is known as homophily: “birds of a feather flock together.”



r=.45

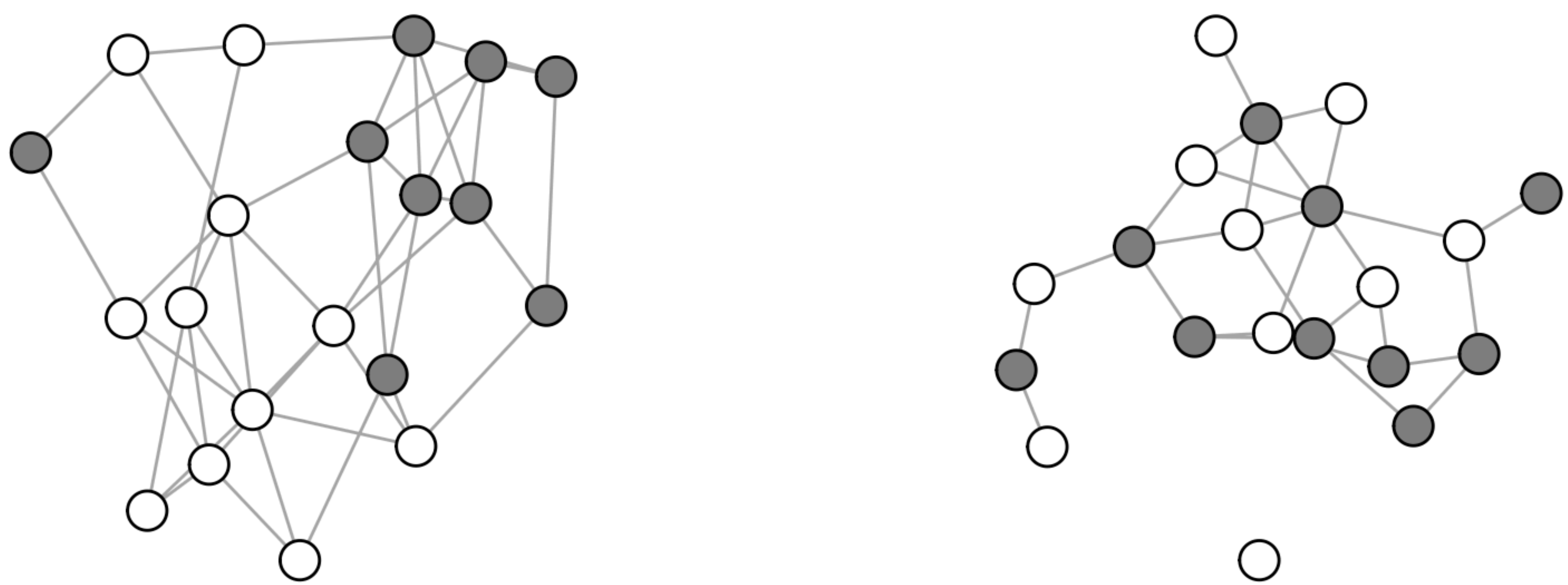


r=-.62

To evaluate assortativity we compute an assortativity coefficient: the Pearson correlation between pairs of connected nodes in the network with respect to the value in question. To do this, generate an edge list from the network, replace the node labels with the value for each node, and take the correlation of the two columns of values.

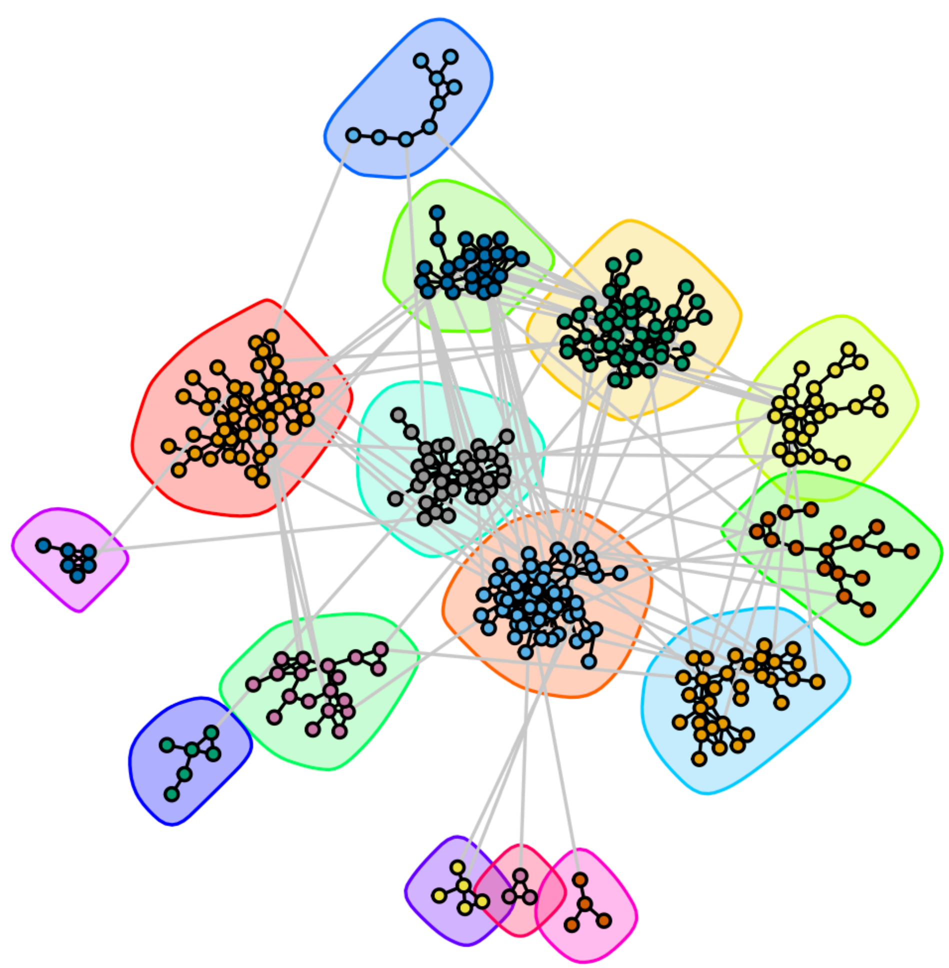
Assortatitivity

Assortativity by colour (a node attribute)



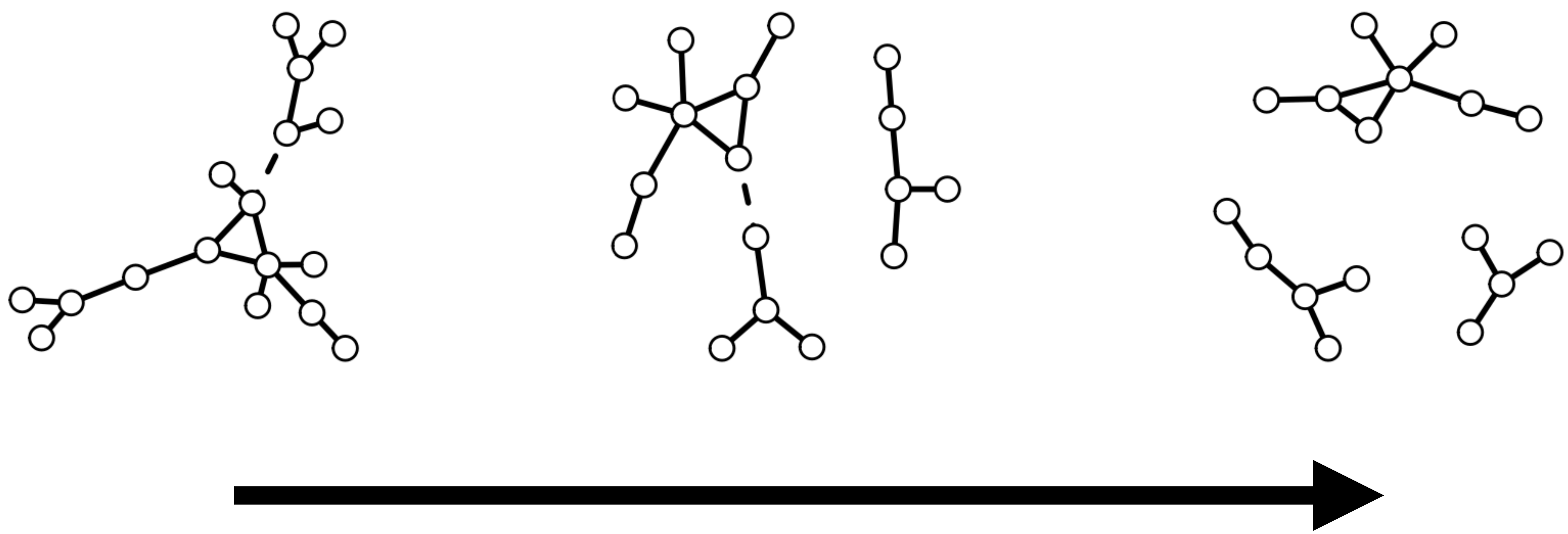
Community Detection

Communities can be detected by identifying clusters of nodes that are more well connected to one another than they are to members of other communities. A division of the network into a set of communities is called a *partition*.

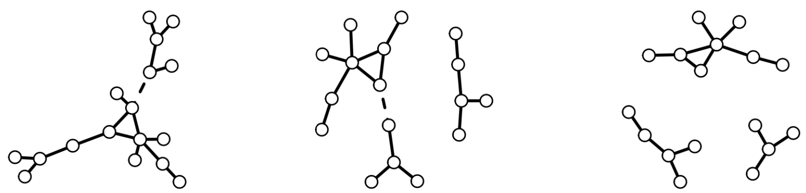


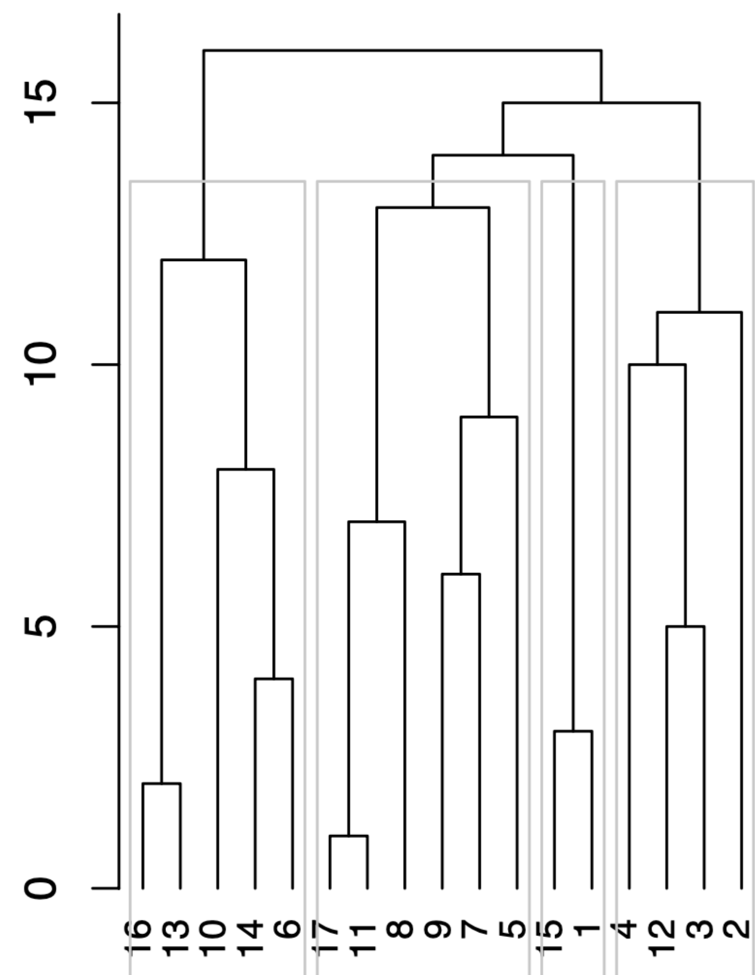
Girvan-Newman Method

The Girvan Newman Method (Girvan and Newman 2002) (or edge betweenness method) is based on the observation that edges connecting separate communities have high edge betweenness: shortest paths between members of different communities will pass through edges with high edge betweenness.



Girvan-Newman Method





How to pick the best partition?

Modularity

Modularity, Q, is a measure of the difference between the observed links within communities and the expected links within the same communities if all edges were distributed at random.

$$Q = \frac{1}{2m}\sum_{i.j}[A_{ij} - \frac{k_i * k_j}{2m}]\delta(c_i, c_j)$$

Observed **Expected**

High modularity means more observed than expected.
Choose partition with highest modularity.

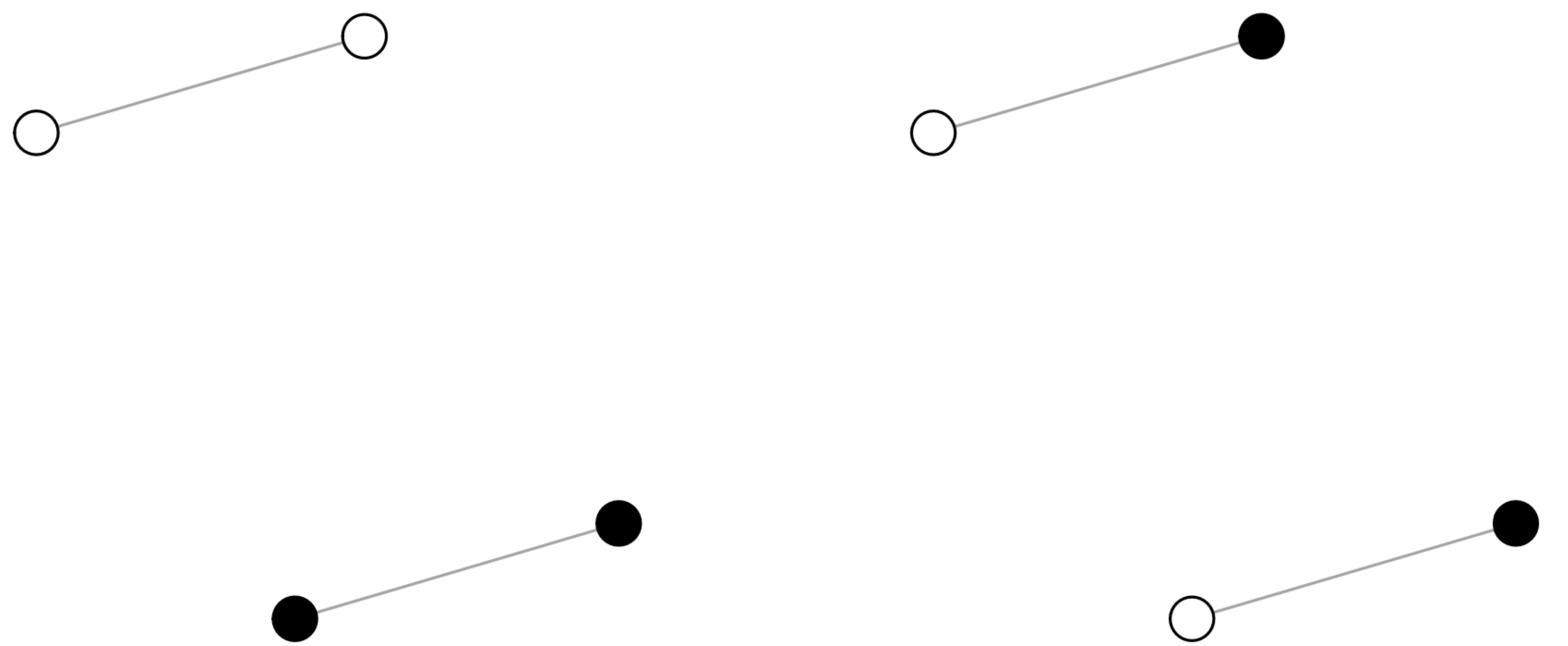
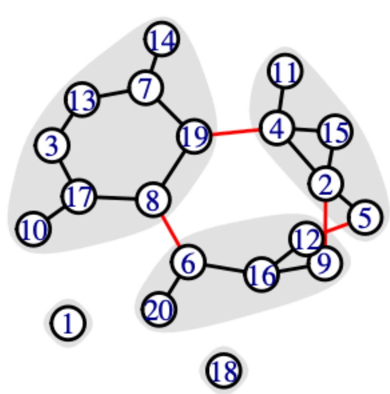


Figure 21: Modularity: Two possible network architectures for 4 nodes with 2 links, each node with degree 1, and 2 communities. The network on the left has a Q=0.5. The network on the right has Q=-0.5.

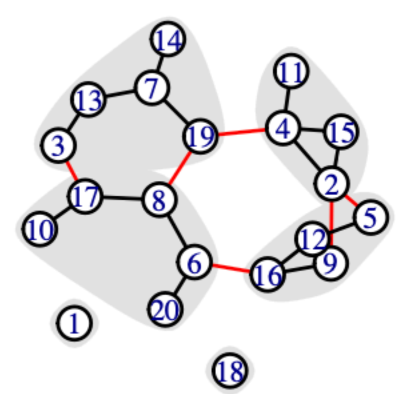
Community detection

(Many methods)

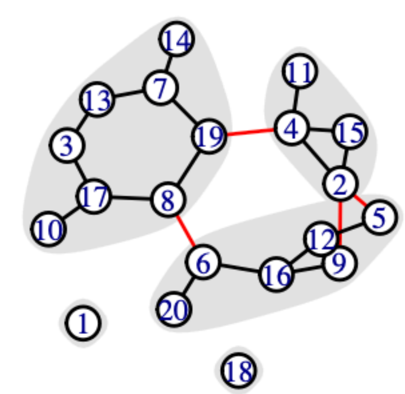
Girvan Newman



Louvain



Walktrap



Clique Percolation