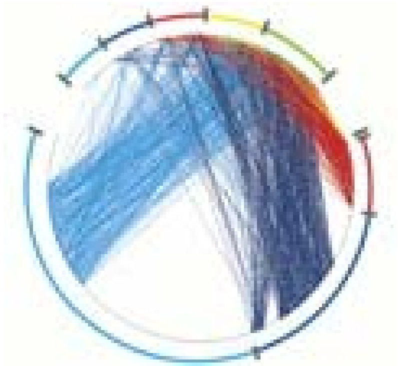
Lecture 6: Centrality Measures



Why a centrality measure?

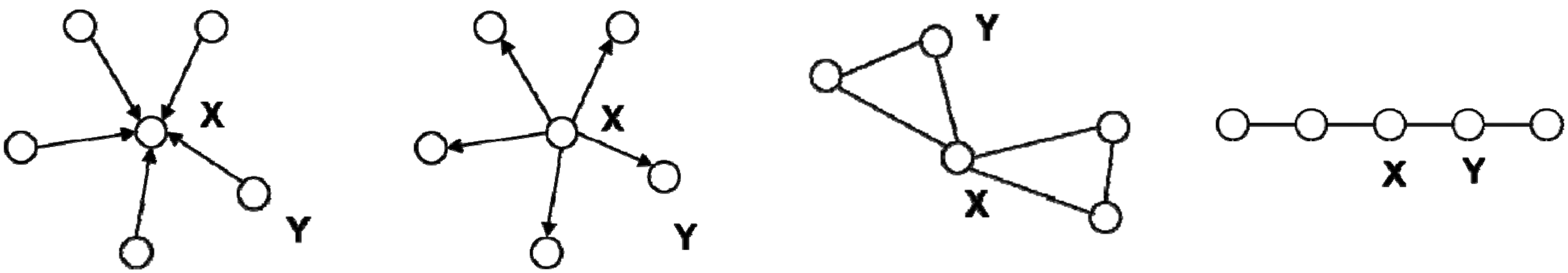
- Often we are interested in identifying IMPORTANT network components
  - Nodes
  - Edges
- Central components may play critical role in network functions
  - Robustness
  - Collective behavior
  - Synchronization
  - Information spreading
  - Social dynamics
  - ...

Centrality

- Which nodes are most ‘central’?
- Definition of ‘central’ varies by context/purpose.
- Local measure:
  - Degree
- Relative to the rest of the network:
  - Closeness
  - Betweenness
  - Eigenvector (Bonacich power centrality)
- How evenly is centrality distributed among nodes?
  - Centralization
  - ...

Centrality

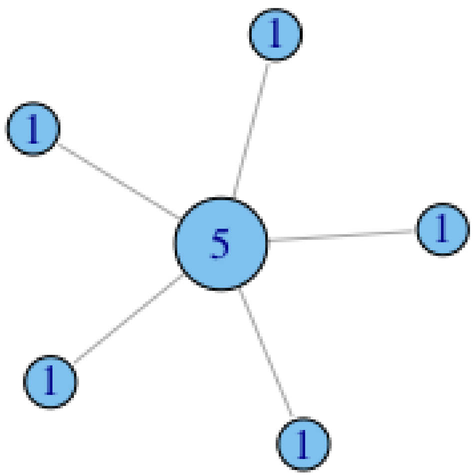
In each of the following networks, X has higher centrality than Y according to a particular measure



indegree outdegree betweenness closeness

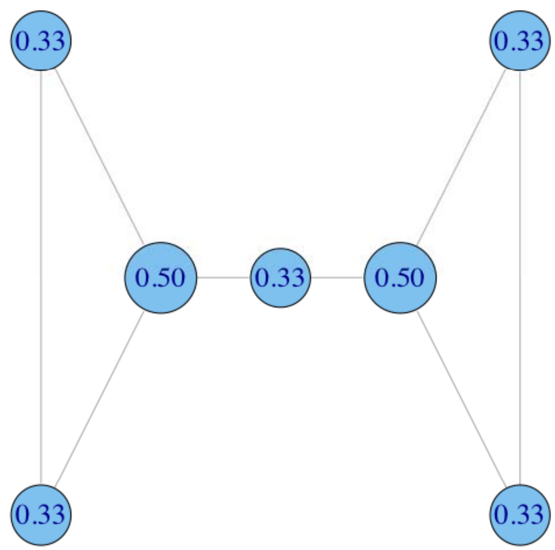
Degree centrality (undirected)

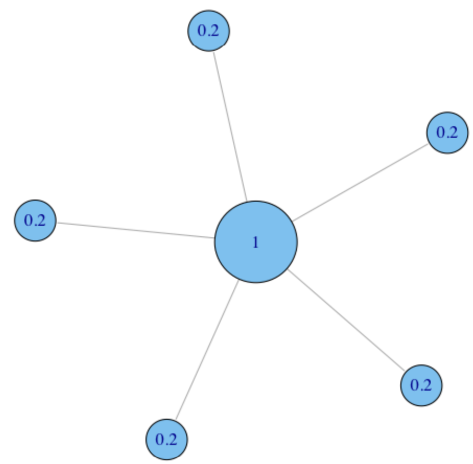
The more the friends the more the importance (the richer the better)



When is the number of connections the best centrality measure?

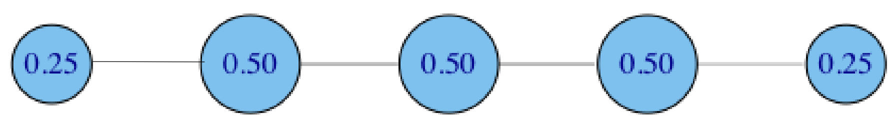
- o people who will do favors for you
- o people you can talk to / have a drink with





Normalized degree centrality:

Degree is divided by the max. possible, i.e. (N-1)

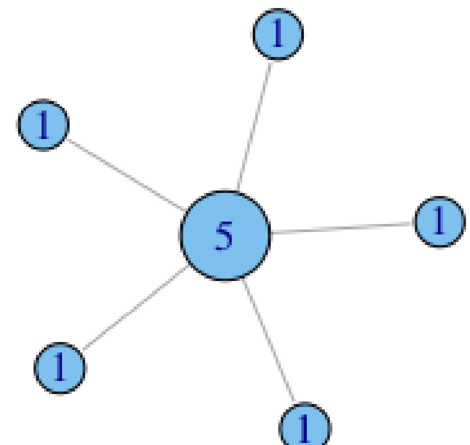


How equal are the nodes?

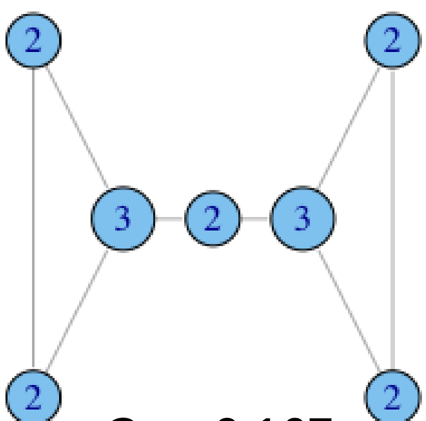
- How much variation is there in the centrality scores among the nodes?
- Freeman's general formula for centralization (can use other metrics, e.g. Gini coefficient or standard deviation):

maximum value in the network

$$C_D = \frac{\sum_{i=1}^{g}\left[\boxed{C_D(n^*)} - C_D(i)\right]}{[(N-1)(N-2)]}$$



$C_D = 1.0$

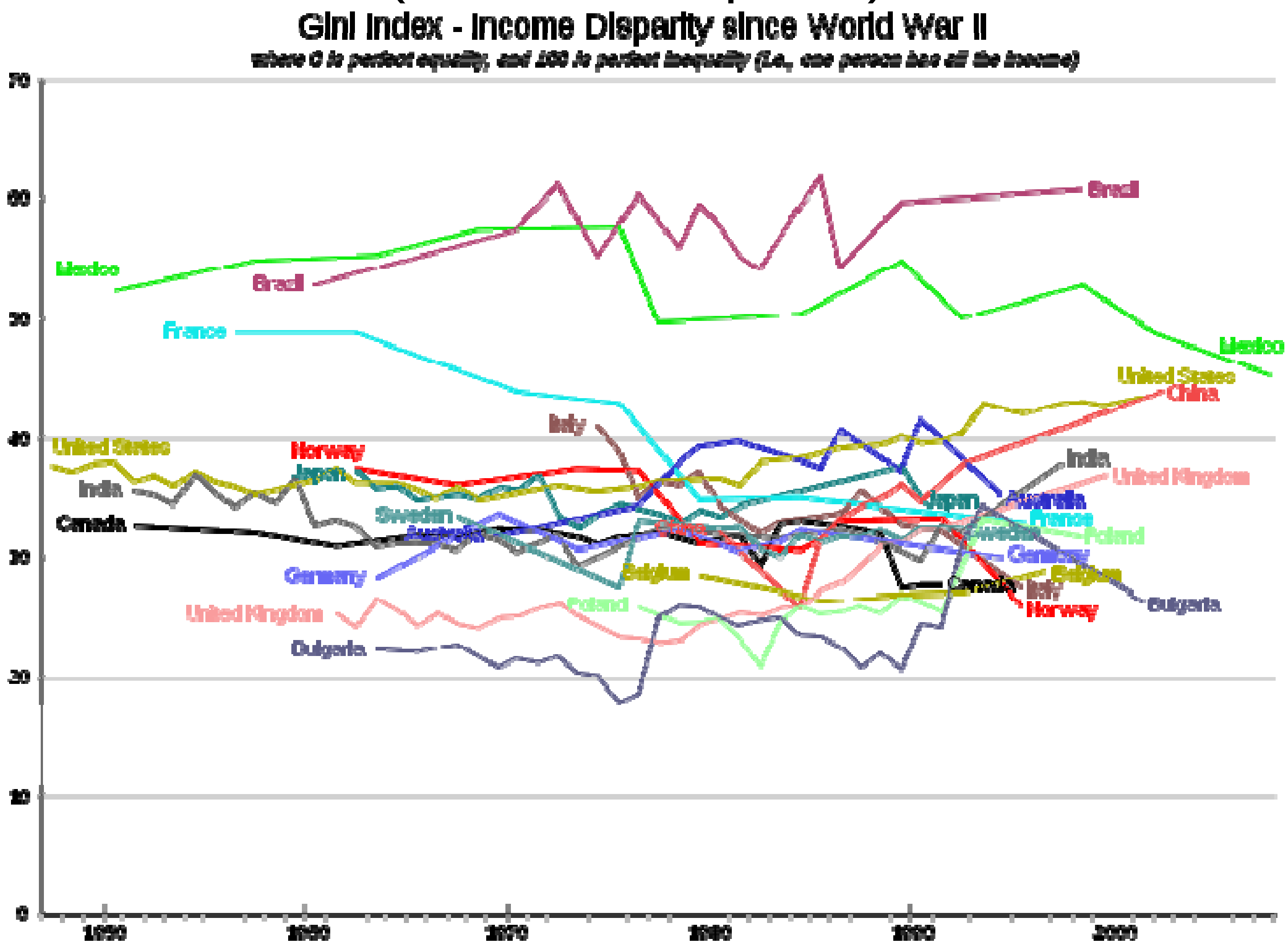


$C_D = 0.167$

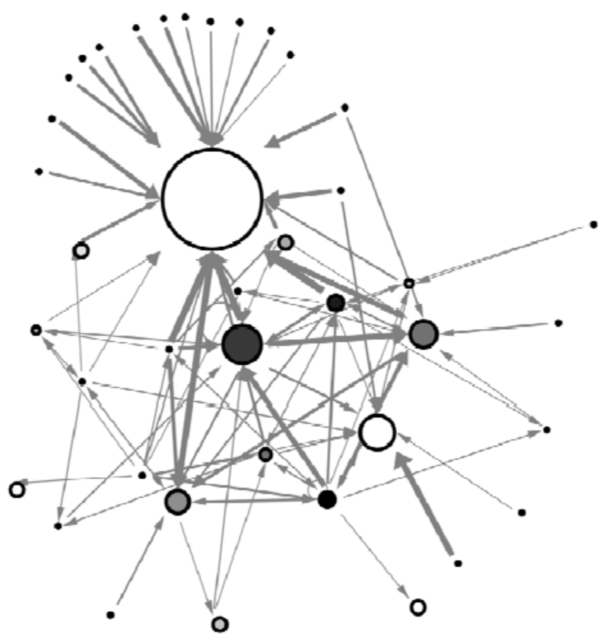
$C_D = 0.167$

How equal are the nodes?

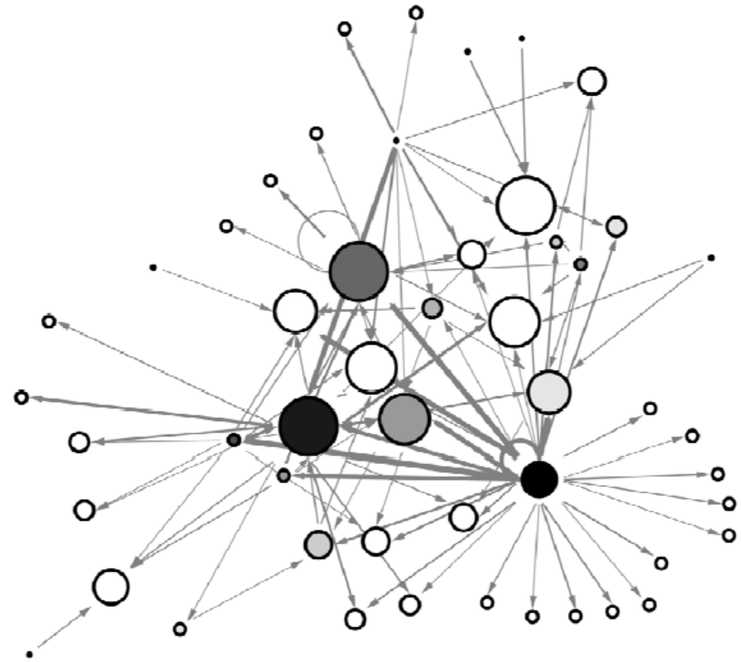
- Gini coefficient? (source: Wikipedia)



Example: financial trading networks



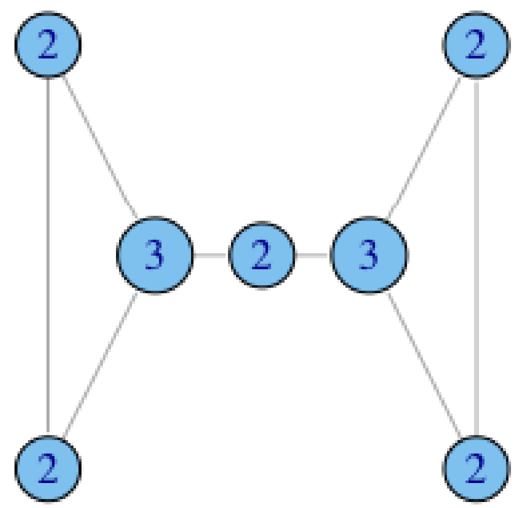
high centralization: one node trading with many others



low centralization: trades are more evenly distributed

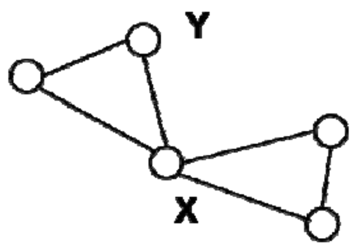
Is degree everything?

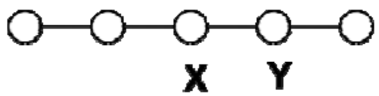
- Nodes with the same degree might have different properties
- In what ways does degree fail to capture centrality in the following graphs?
  - ability to broker between groups
  - likelihood that information originating anywhere in the network reaches you...

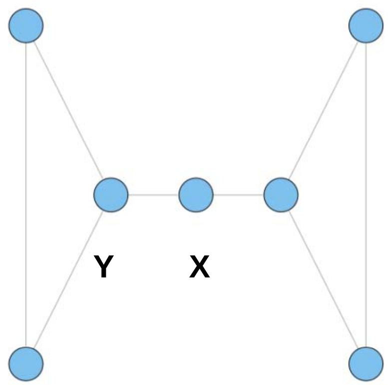


Betweenness centrality

- Intuition: how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?
- who has higher betweenness, X or Y?







Betweenness centrality

- The betweenness of a node i is defined as

$$C_B(i) = \sum_{j<k} g_{jk}(i) / g_{jk}$$

- Where
  - g_{jk} = the number of shortest paths connecting the nodes j and k
  - $g_{jk}(i)$ = the number of these paths making use of i (those paths originating from or ending at i are not counted)
- Usually the betweenness is normalized by

$$C'_B(i) = C_B(i) / \boxed{[(n-1)(n-2)/2]}$$

number of pairs of vertices
excluding the vertex itself

Betweenness centrality

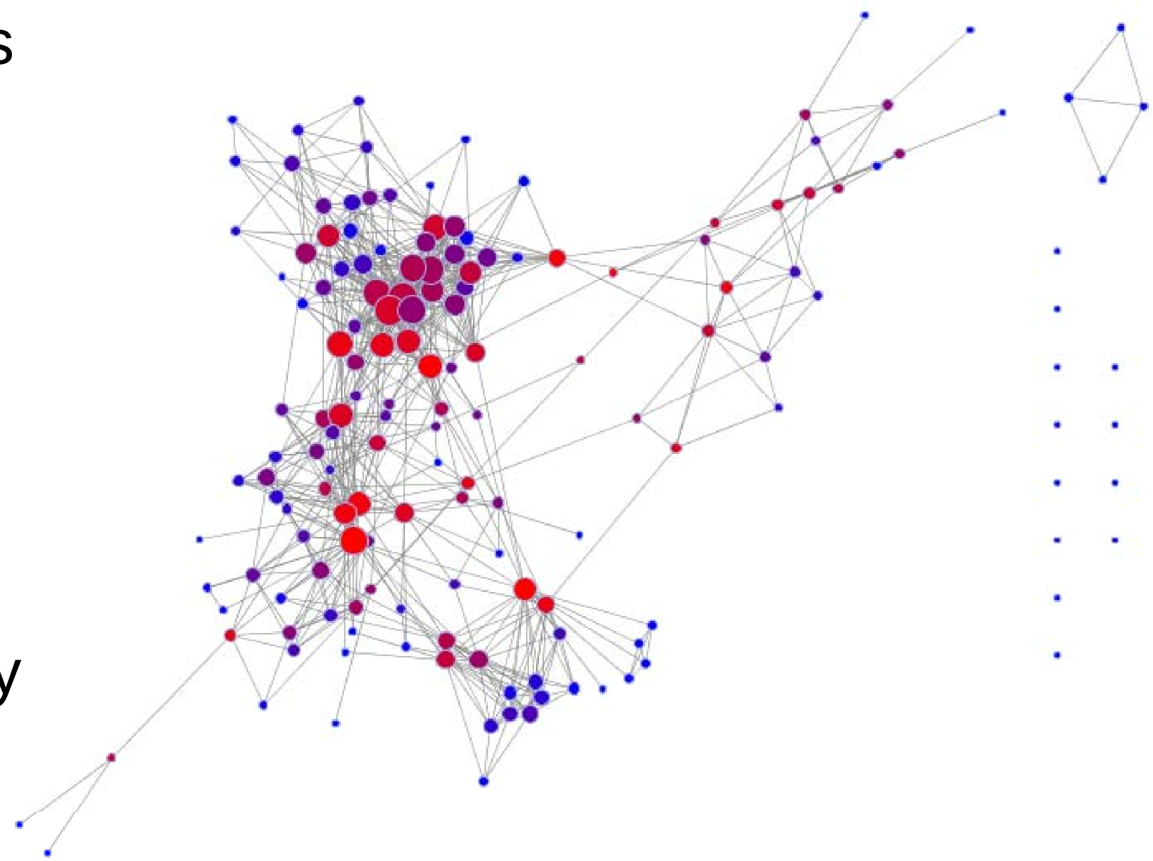
- The same concept of betweenness centrality can be defined on edges
- The betweenness of an edge (i,j) between nodes i and j is defined as $C_B(i,j) = \sum_{h<k} g_{jk}(i,j) / g_{hk}$
- Where
  - g_{hk} = the number of shortest paths connecting the nodes h and k
  - $g_{jk}(i,j)$ = the number of these paths making use of the edge (i,j)
- Usually the betweenness is normalized by

$$C'_B(i,j) = C_B(i,j) \boxed{/ [(n-1)(n-2)/2]}$$

Number of possible edges

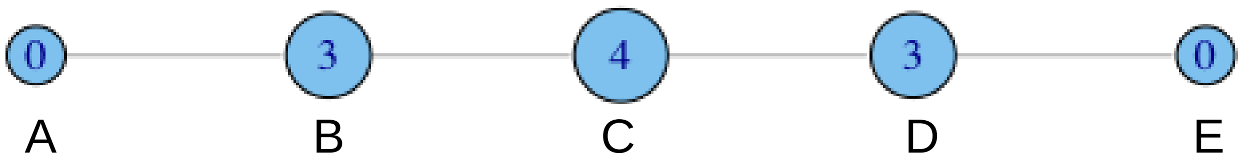
Betweenness centrality: an example

- Nodes are sized by degree, and colored by betweenness.
- Can you spot nodes with high betweenness but relatively low degree? Explain how this might arise.
- What about high degree but relatively low betweenness?

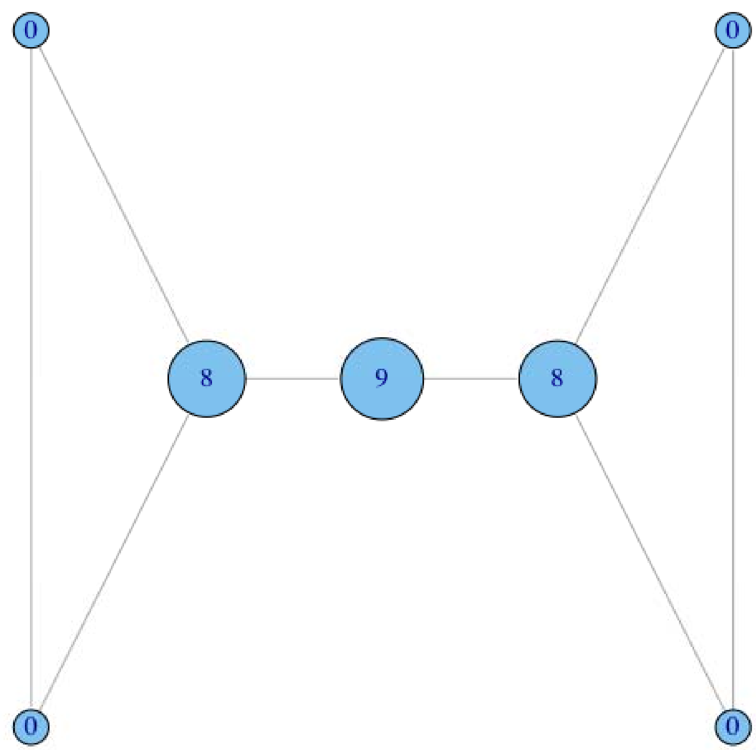


Betweenness centrality

- Non-normalized version:

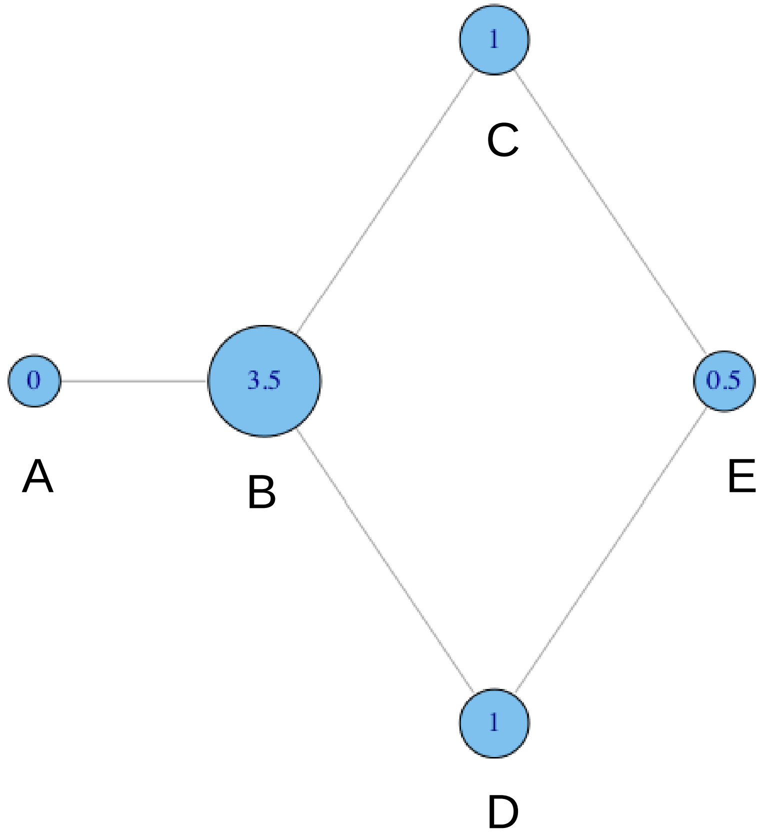


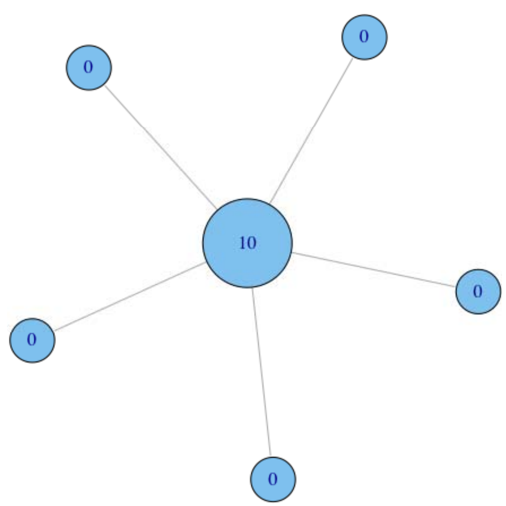
- A lies between no two other vertices
- B lies between A and 3 other vertices: C, D, and E
- C lies between 4 pairs of vertices (A,D),(A,E),(B,D),(B,E)
- note that there are no alternate paths for these pairs to take, so C gets full credit



Betweenness centrality

- Why do C and D each have betweenness 1?
- They are both on shortest paths for pairs (A,E), and (B,E), and so must share credit: ½+½ = 1
- Can you figure out why B has betweenness 3.5 while E has betweenness 0.5?





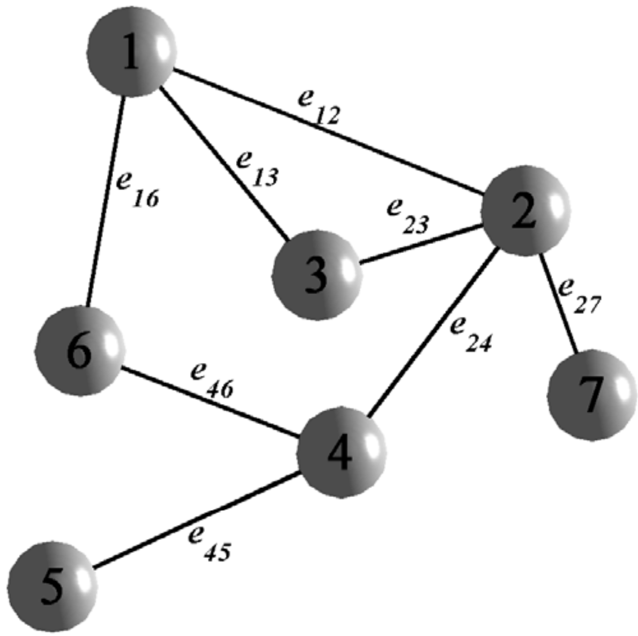
Connection graph stability scores

- In some applications the importance of the shortest paths are also important
- The importance of the shortest paths might be different
- The more important paths making use of an edge the more its importance
- A simple measure of importance would be the length
- The connection graph stability (CGS) method takes into account this issue
- It has application in synchronization analysis
- The CGS-score b_{ij} for the link between the nodes i and j is defined as

$$b_{ij} = \sum_{u=1}^{n-1} \sum_{v>u;e_{ij} \in P_{uv}}^{n} | P_{uv} |$$

- $|P_{uv}|$: length of path P_{uv} between the nodes u and v

Connection graph stability scores



$P_{12} = e_{12}$, $P_{13} = e_{13}$, $P_{14} = e_{12}e_{24}$,
$P_{15} = e_{16}e_{46}e_{45}$, $P_{16} = e_{16}$, $P_{17} = e_{12}e_{77}$,
$P_{23} = e_{23}$, $P_{24} = e_{24}$, $P_{25} = e_{24}e_{45}$,
$P_{26} = e_{24}e_{46}$, $P_{27} = e_{27}$, $P_{34} = e_{23}e_{24}$,
$P_{35} = e_{23}e_{24}e_{45}$, $P_{36} = e_{13}e_{16}$, $P_{37} = e_{23}e_{27}$,
$P_{45} = e_{45}$, $P_{46} = e_{46}$, $P_{47} = e_{24}e_{27}$, $P_{56} = e_{45}e_{46}$, $P_{57} = e_{45}e_{24}e_{27}$, $P_{67} = e_{16}e_{12}e_{27}$

$$b_{12} = |P_{12}| + |P_{14}| + |P_{17}| + |P_{67}| = 1 + 2 + 2 + 3 = 8,$$

$$b_{13} = |P_{13}| + |P_{36}| = 1 + 2 = 3,$$

$$b_{16} = |P_{15}| + |P_{16}| + |P_{36}| + |P_{67}| = 3 + 1 + 2 + 3 = 9,$$

$$b_{23} = |P_{23}| + |P_{34}| + |P_{35}| + |P_{37}| = 1 + 2 + 3 + 2 = 8,$$

$$b_{24} = |P_{14}| + |P_{24}| + |P_{25}| + |P_{26}| + |P_{34}| + |P_{35}| + |P_{47}| + |P_{57}| = 2 + 1 + 2 + 2 + 2 + 3 + 2 + 3 = 17,$$

$$b_{27} = |P_{17}| + |P_{27}| + |P_{37}| + |P_{47}| + |P_{57}| + |P_{67}| = 2 + 1 + 2 + 2 + 3 + 3 = 13,$$

$$b_{45} = |P_{15}| + |P_{25}| + |P_{35}| + |P_{45}| + |P_{56}| + |P_{57}| = 3 + 2 + 3 + 1 + 2 + 3 = 14,$$

$$b_{46} = |P_{15}| + |P_{26}| + |P_{46}| + |P_{56}| = 3 + 2 + 1 + 2 = 8.$$

Closeness centrality

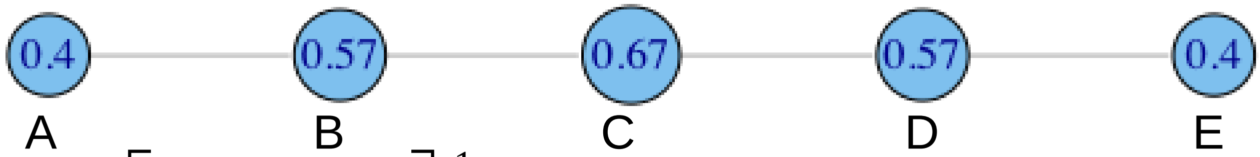
- What if it is not so important to have many direct friends?
- Or be "between" others
- But one still wants to be in the "middle" of things, not too far from the center
- So, let us define closeness centrality
- Closeness is based on the length of the average shortest path between a vertex and all vertices in the graph as

$$C_c(i) = \left[\sum_{j=1}^{N} P(i,j) \right]^{-1}$$

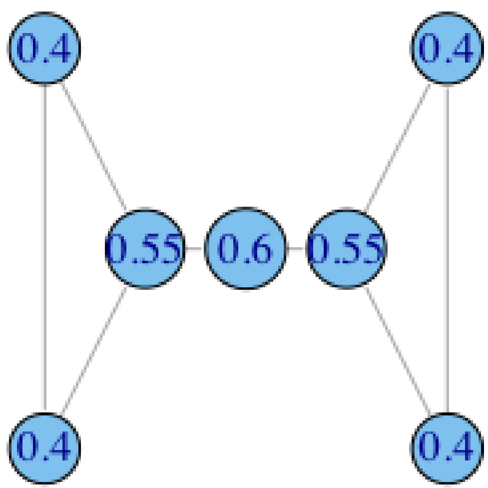
- Normalized Closeness Centrality

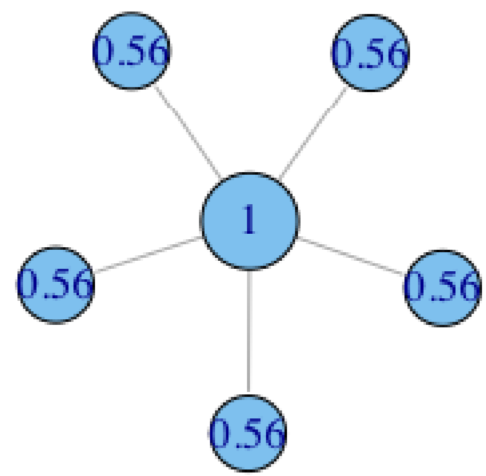
$$C_C^{'}(i) = \left[\sum_{j=1}^{N} P(i,j) / (N-1) \right]^{-1}$$

Closeness centrality: toy example



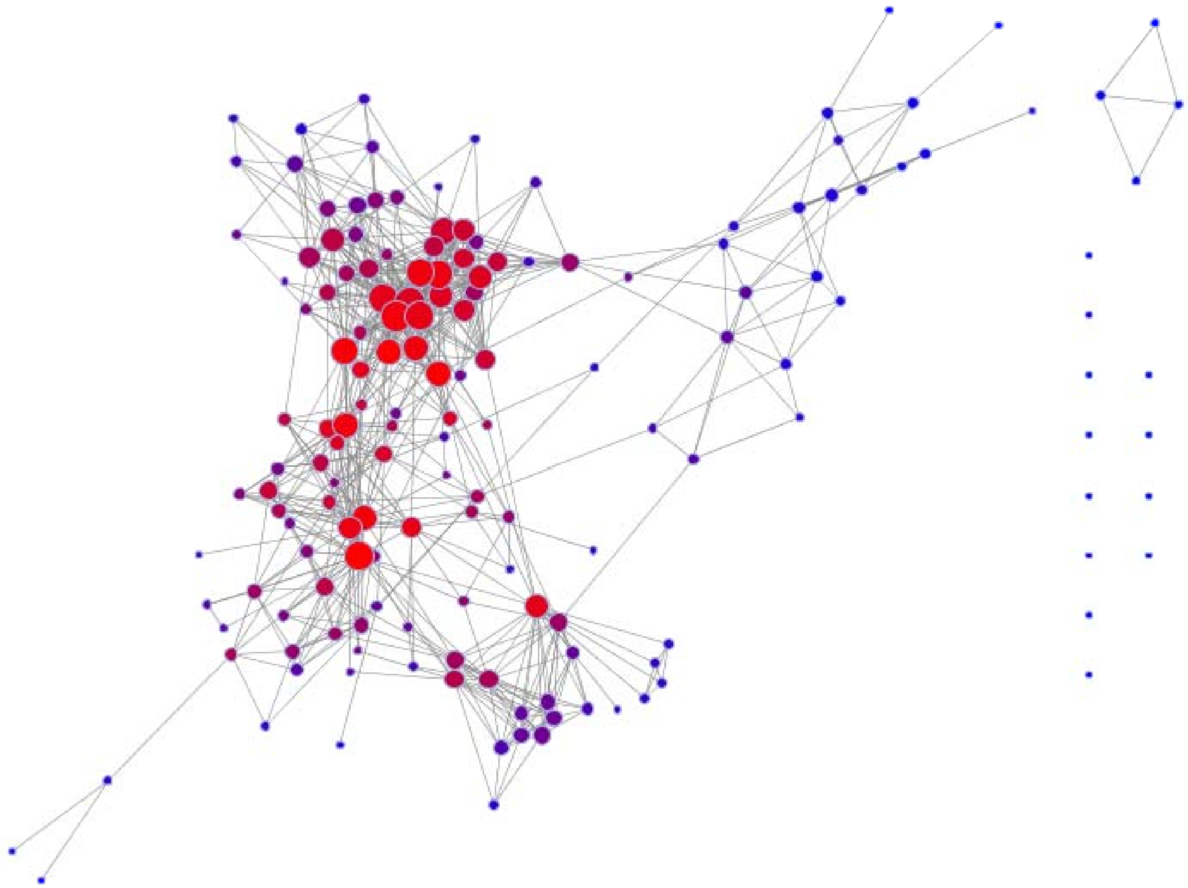
$$C_c^{'}(A) = \left[\frac{\sum_{j=1}^{N} d(A,j)}{N-1}\right]^{-1} = \left[\frac{1+2+3+4}{4}\right]^{-1} = \left[\frac{10}{4}\right]^{-1} = 0.4$$





Closeness and degree

- **degree** (number of connections) denoted by size
- **closeness** (length of shortest path to all others) denoted by color



Centrality ...

- generally different centrality metrics will be positively correlated
- when they are not, there is likely something interesting about the network
- suggest possible topologies and node positions to fit each square
- Note that CSG and closeness are closely related
- There are various extensions to these centrality measures

Bonacich power centrality

- The centrality might depend on neighbors' centrality
- Consider an eigenvector measure

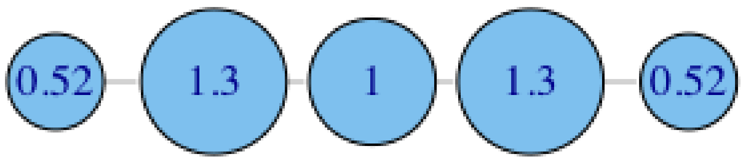
$$C(\alpha, \beta) = \alpha(I - \beta R)^{-1} R1$$

  - α is a scaling vector, which is set to normalize the score
  - β reflects the extent to which you weight the centrality of a node's neighbors
  - **R** is the adjacency matrix (can be valued)
  - **I** is the identity matrix (1s down the diagonal)
  - **1** is a matrix of all ones
- The magnitude of β reflects the radius of power. Small values of β weight local structure, larger values weight global structure
  - If $\beta > 0$, the node has higher centrality when tied to people who are central
  - If $\beta < 0$, the node has higher centrality when tied to people who are not central
  - With $\beta = 0$, you get degree centrality

Bonacich power centrality

β=.25

β=-.25



Why does the middle node have lower centrality than its neighbors when β is negative?

Centrality in directed networks

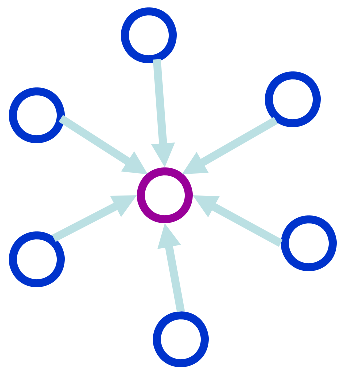
- Many of real-world networks are directed such as
  - WWW
  - food webs
  - population dynamics
  - influence
  - hereditary
  - citation
  - transcription regulation networks
  - neural networks
  - ...

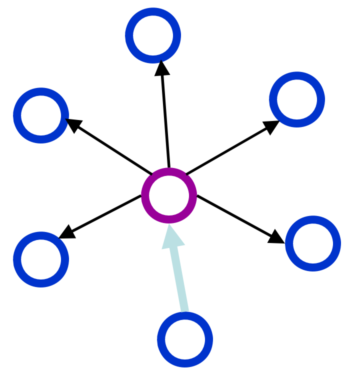
Prestige in directed social networks

- when ‘prestige’ may be the right word
  - admiration
  - influence
  - gift-giving
  - Trust
- directionality especially important in instances where ties may not be reciprocated (e.g. trust (or influence) networks)
- when ‘prestige’ may not be the right word
  - gives advice to (can reverse direction)
  - gives orders to (- ” -)
  - lends money to (- ” -)
  - dislikes
  - distrusts

Prestige in directed social networks

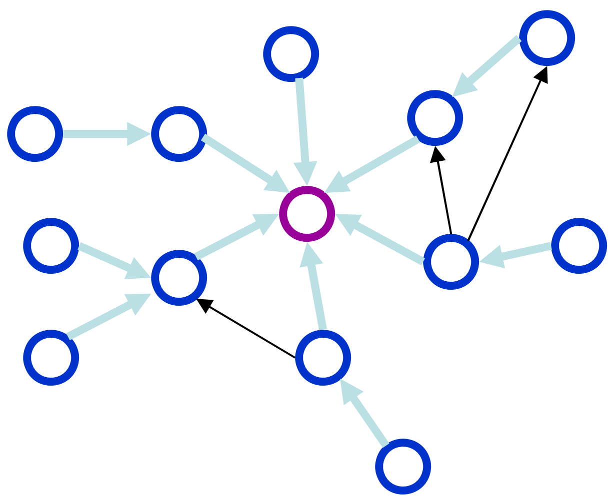
- degree centrality
  - indegree centrality
    - a paper that is cited by many others has high prestige
    - a person nominated by many others for a reward has high prestige





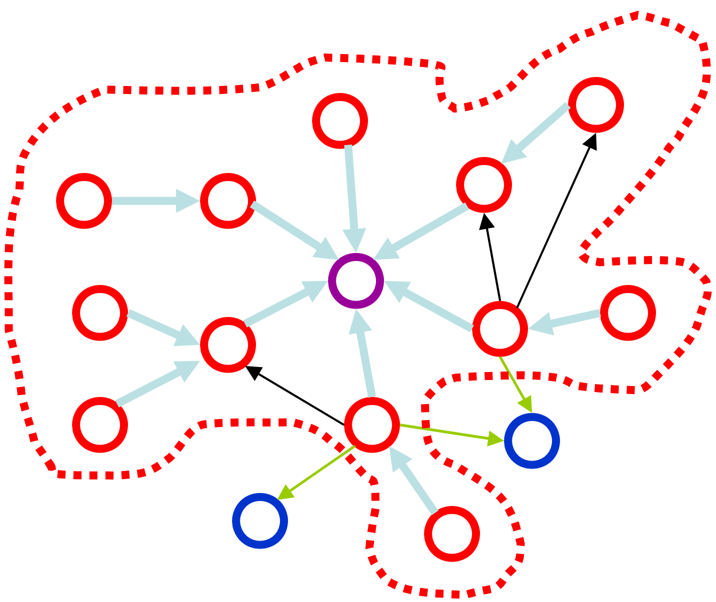
Prestige in directed social networks

- closeness centrality usually implies
  - all paths should lead to you and
  - unusually not paths should lead from you to everywhere else
- usually consider only vertices from which the node i in question can be reached



Influence range

- The influence range of i is the set of vertices who are reachable from the node i
- Alternatively, we can also consider the influential range of node i as a set of the nodes with a path to i



Extended betweenness centrality

- We now consider the fraction of all directed paths between any two vertices that pass through a node

betweenness of vertex i

paths between j and k that pass through i

$$C_B(i) = \sum_{j,k} g_{jk}(i) / g_{jk}$$

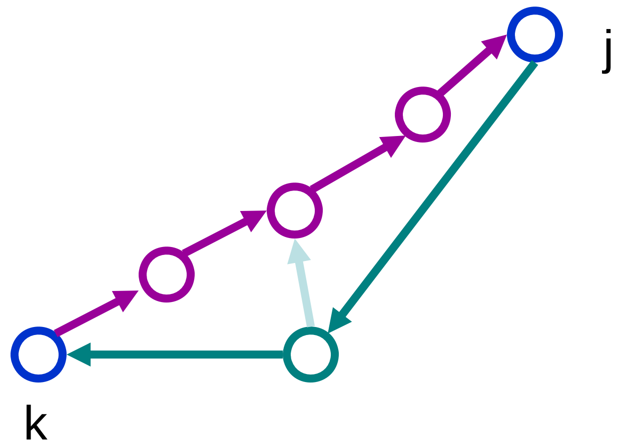
all paths between j and k

- Only modification: when normalizing, we have (N-1)*(N-2) instead of (N-1)*(N-2)/2, because we have twice as many ordered pairs as unordered pairs

$$C'_B(i) = C_B(i) / [(N-1)(N-2)]$$

Directed shortest paths

- A node does not necessarily lie on a geodesic from j to k if it lies on a geodesic from k to j



Structural bases of influence

- <u>Centrality</u>
  - Central actors are likely more influential. They have greater access to information and can communicate their opinions to others more efficiently. Research shows they are also more likely to use the communication channels than are periphery actors.
- <u>Structural Similarity</u>
  - Two people may not be directly connected, but occupy a similar position in the structure. As such, they have similar interests in outcomes that relate to positions in the structure.
  - Similarity must be conditioned on visibility. P must know that O is in the same position, which means that the effect of similarity might be conditional on communication frequency.
- <u>Cohesion</u>
  - Members of a cohesive group are likely to be aware of each others' opinions, because information diffuses quickly within the group.
  - Groups encourage reciprocity and compromise. This likely increases the salience of opinions of other group members, over non-group members.

Social organization of conspiracy

- Questions: How are relations organized to facilitate illegal behavior?
  - Pattern of communication maximizes concealment, and predicts the criminal verdict.
  - Inter-organizational cooperation is common, but too much ‘cooperation’ can thwart (or deviate) market competition, leading to (illegal) market failure.
  - Illegal networks differ from legal networks, in that they must conceal their activity from outside agents. A “Secret society” should be organized to (a) remain concealed and (b) if discovered make it difficult to identify who is involved in the activity
  - The need for secrecy should lead conspirators to conceal their activities by creating **sparse** and **decentralized** networks.

Social organization of conspiracy

- center: good for reaping (or gaining) the benefits
- periphery: good for remaining concealed
  - They examine the effect of Degree, Betweenness and Closeness centrality on the criminal outcomes, based on reconstruction of the communication networks involved.
  - At the **organizational level**, low information-processing conspiracies are decentralized; high information processing load leads to centralization
  - At the **individual level**, degree centrality (net of other factors) predicts judgment.

Readings

- L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. Advances in Physics, 56(1):167 – 242, 2007.
- “Networks, Crowds, and Markets” by Easley and Kleinberg (Chapter 3)
- Ulrik Brandes, A Faster Algorithm for Betweenness Centrality, Journal of Mathematical Sociology 25(2):163-177, 2001.
- Kazuya Okamoto, Wei Chen, and Xiang-Yang Li, Ranking of Closeness Centrality for Large-Scale Social Networks, LNCS, 5059, 186-195,2008.