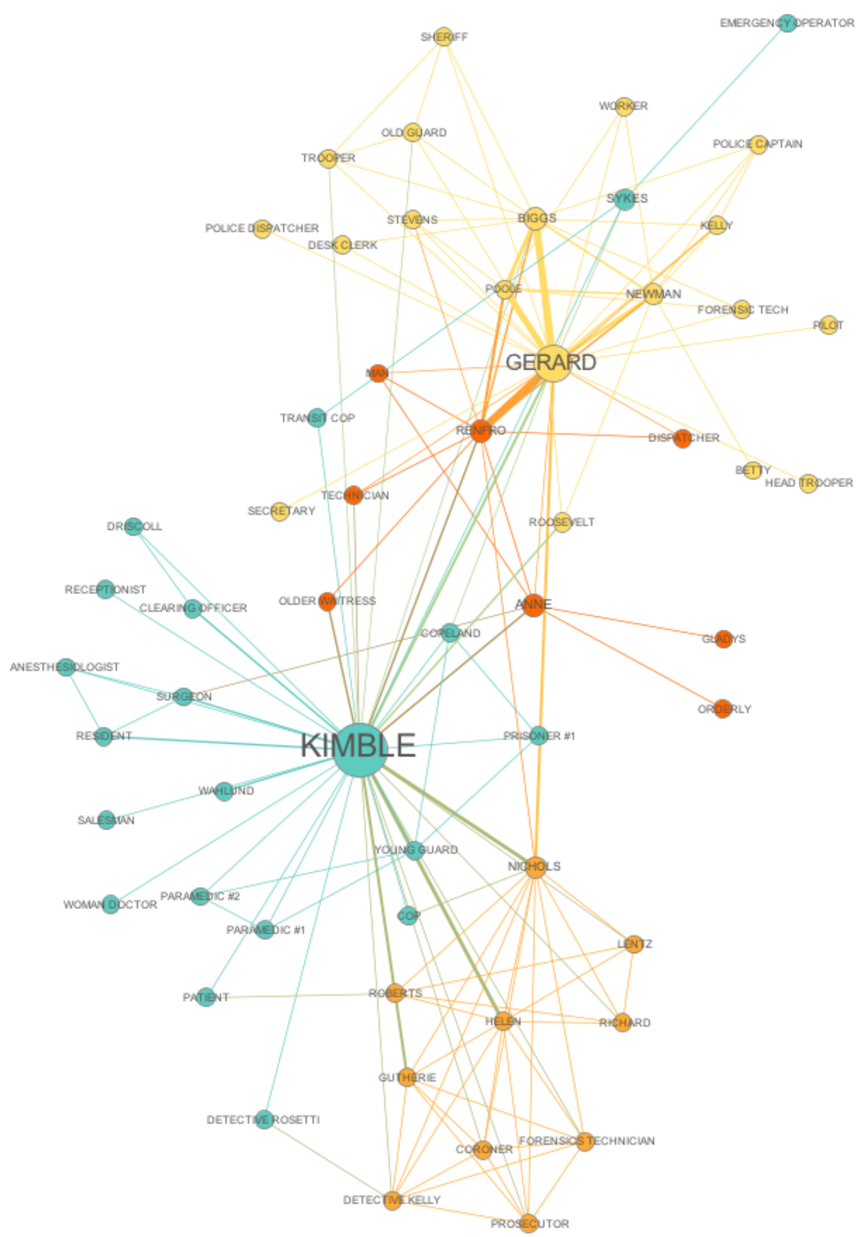
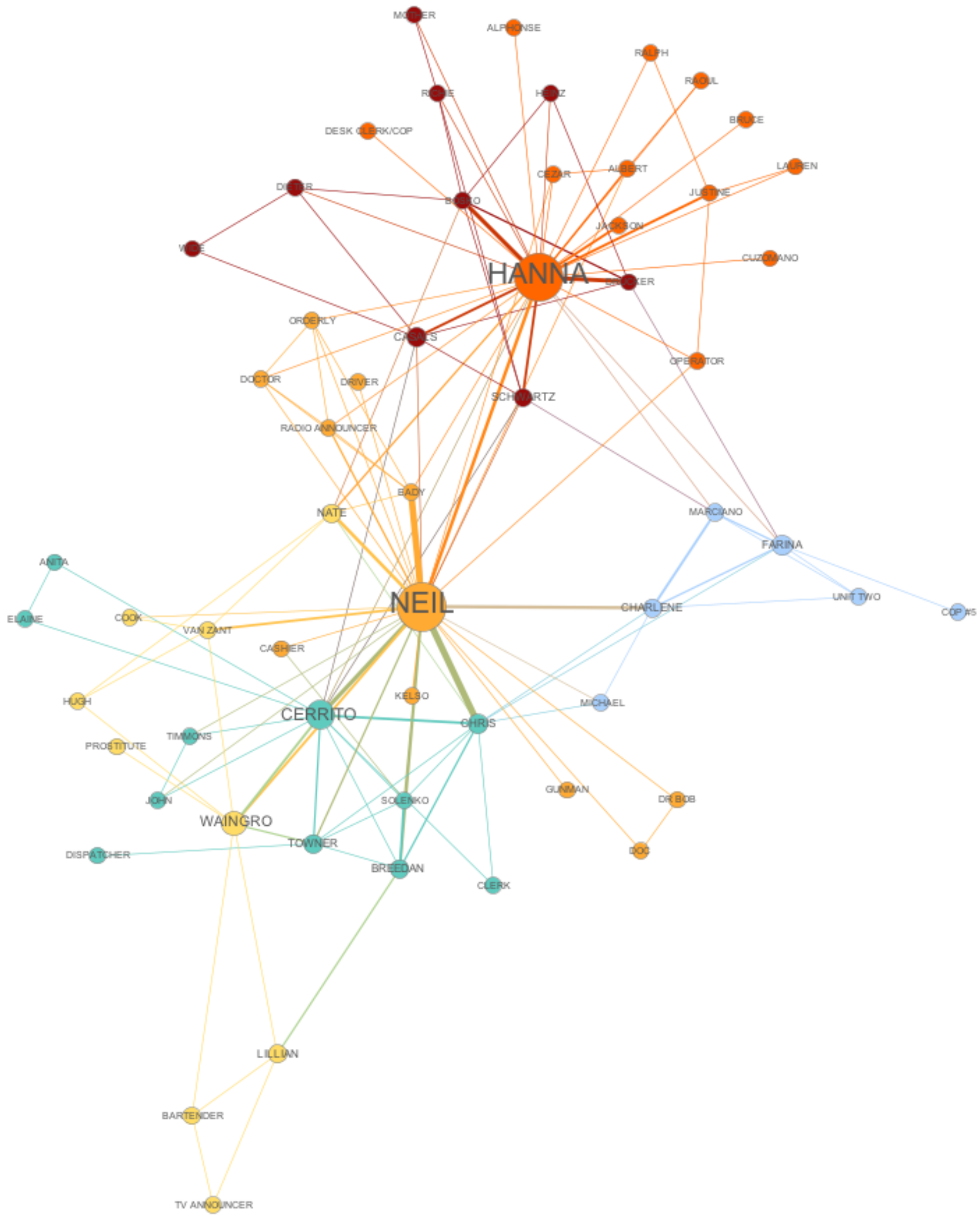
Centrality

CS224W: Social and Information Network Analysis
Lada Adamic
http://cs224w.stanford.edu

The Fugitive (1993)

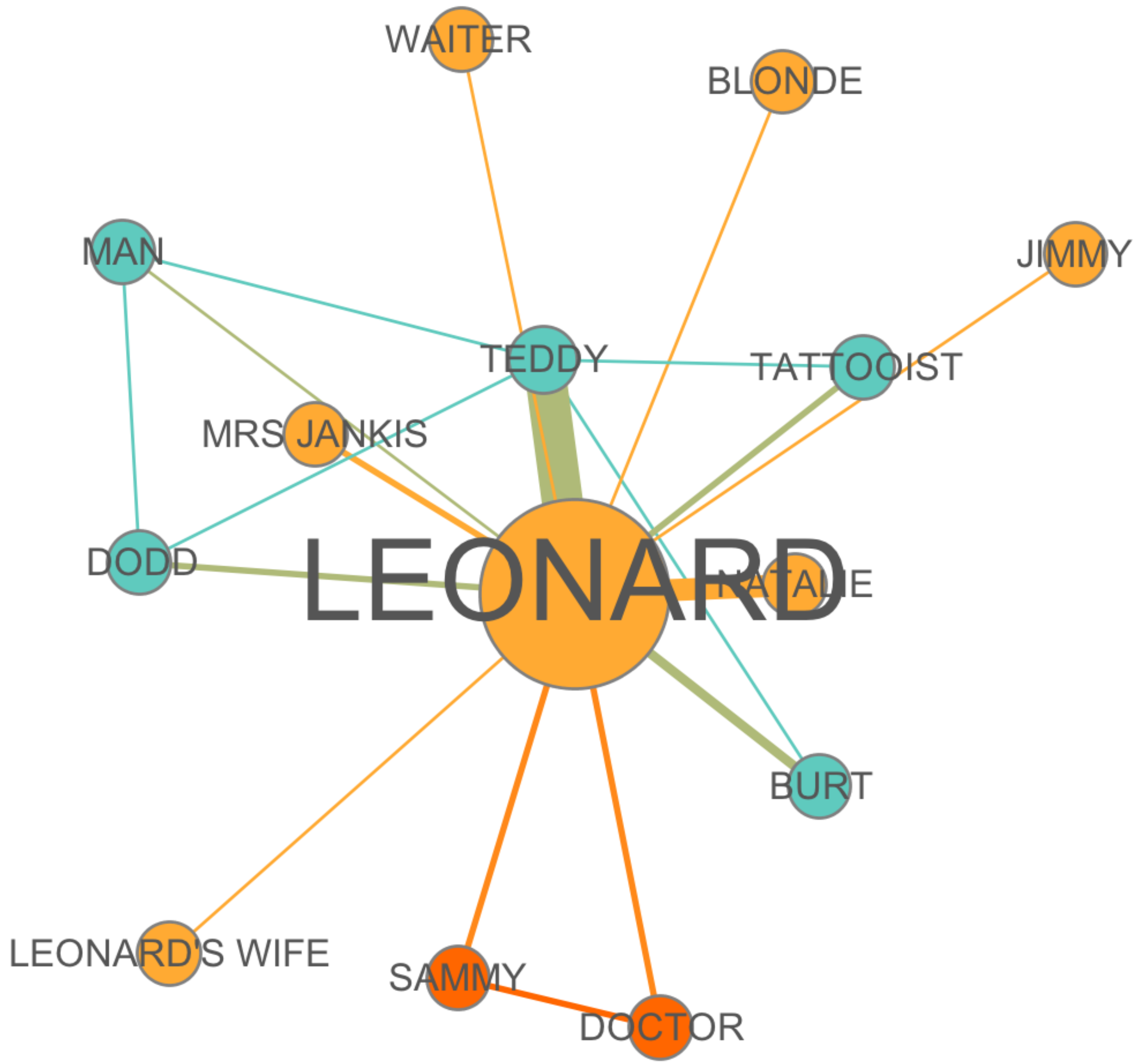


Heat (1995)



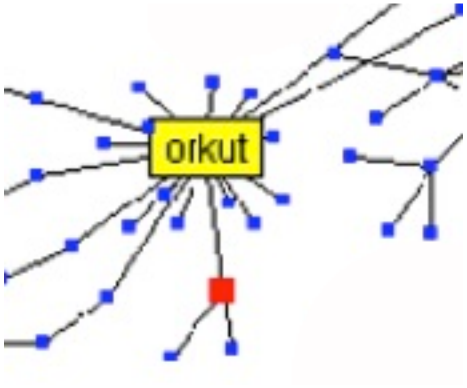
moviegalaxies.com

Memento (2000)

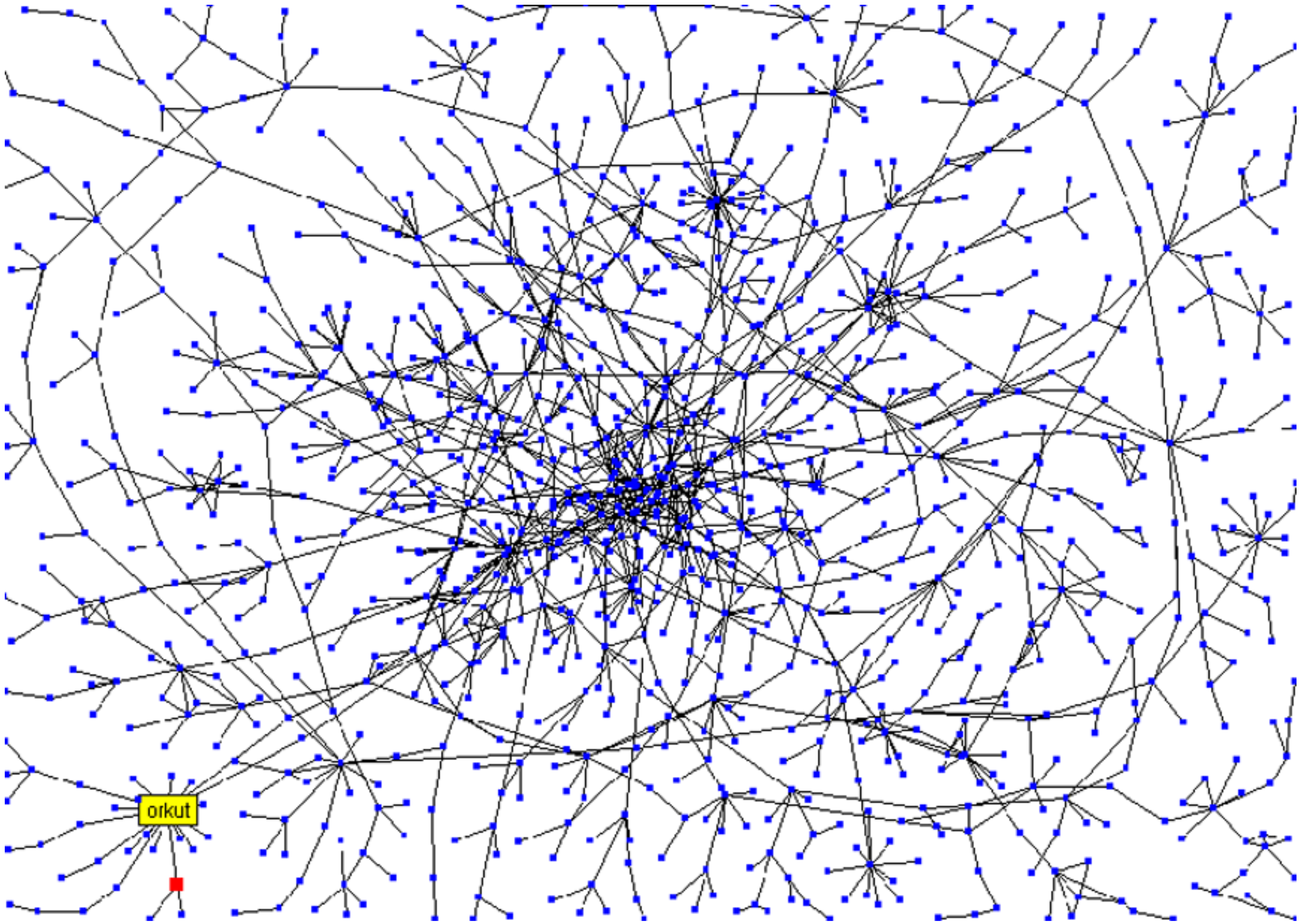


moviegalaxies.com

Is counting the edges enough?



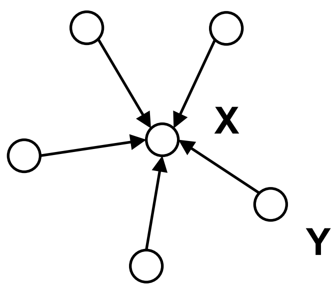
Stanford Social Web (ca. 1999)



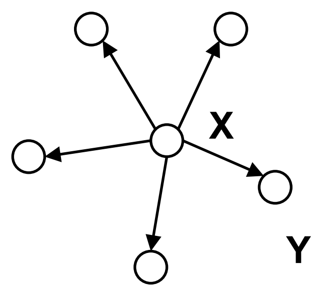
network of personal homepages at Stanford

different notions of centrality

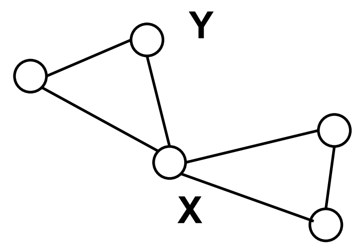
In each of the following networks, X has higher centrality than Y according to a particular measure



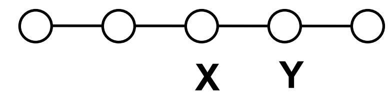
indegree



outdegree



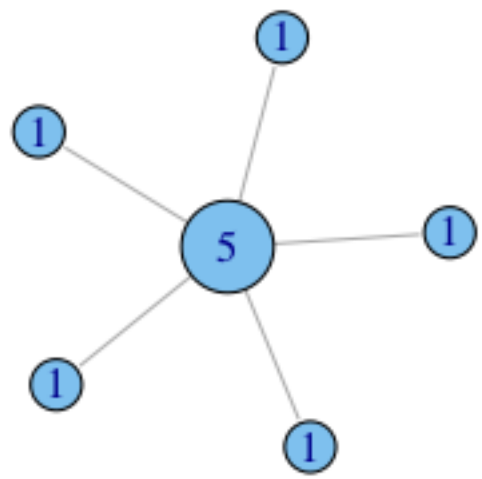
betweenness



closeness

putting numbers to it

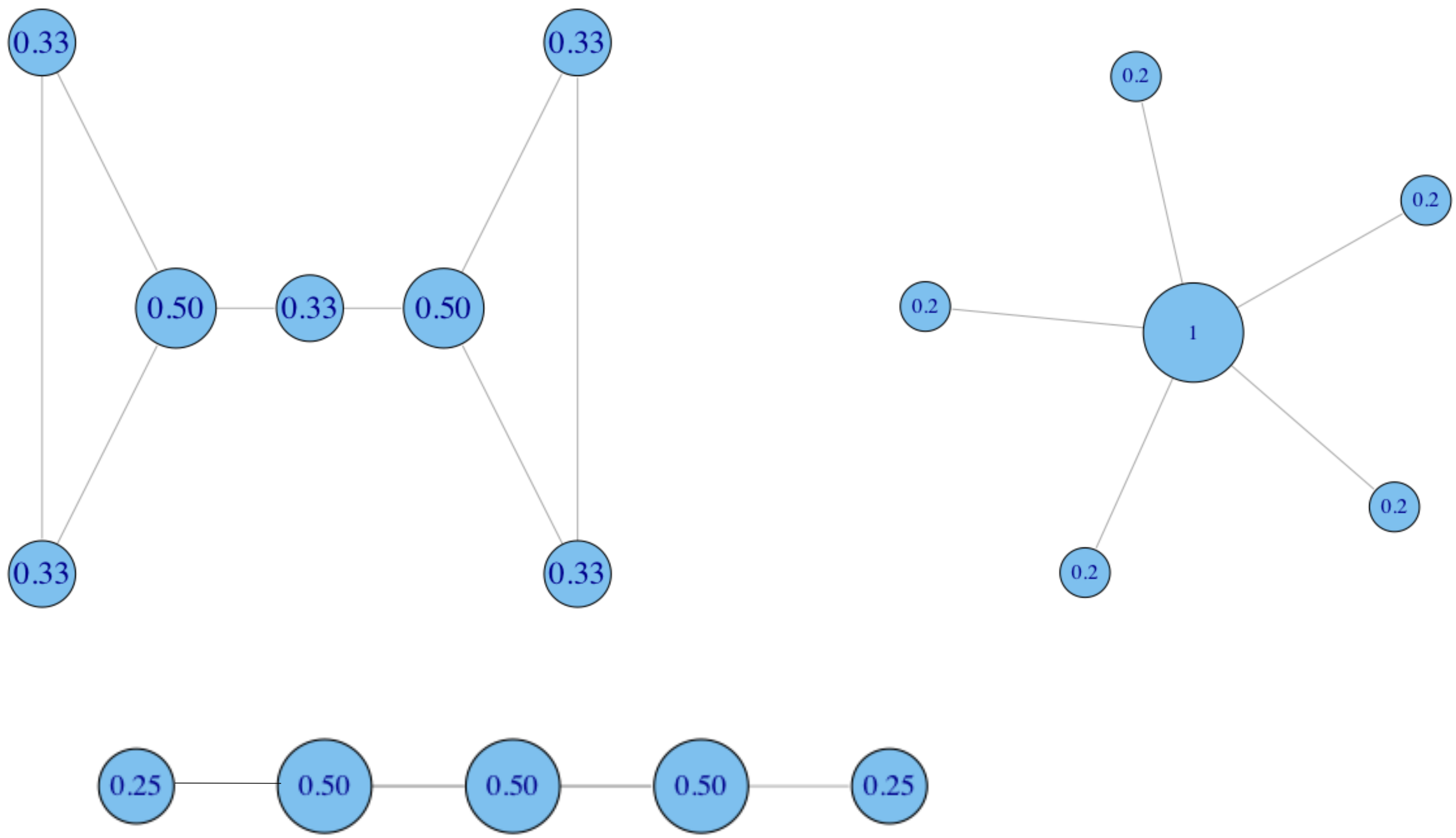
Undirected degree, e.g. nodes with more friends are more central.



Assumption: the connections that your friend has don't matter, it is what they can do directly that does (e.g. go have a beer with you, help you build a deck...)

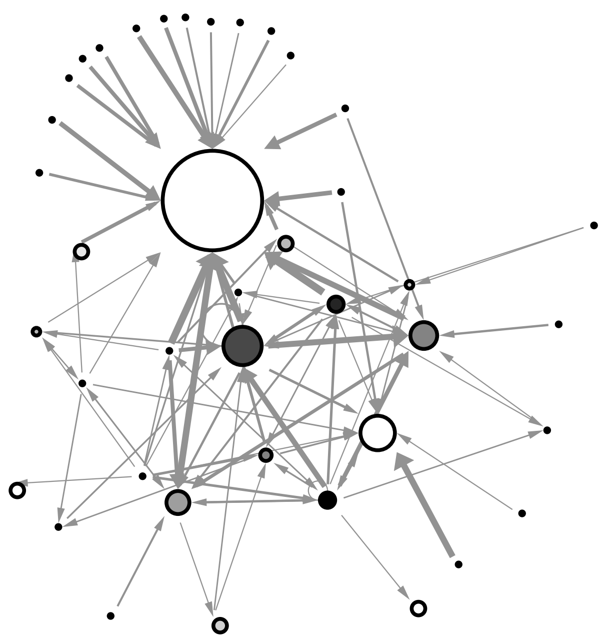
normalization

divide degree by the max. possible, i.e. (N-1)

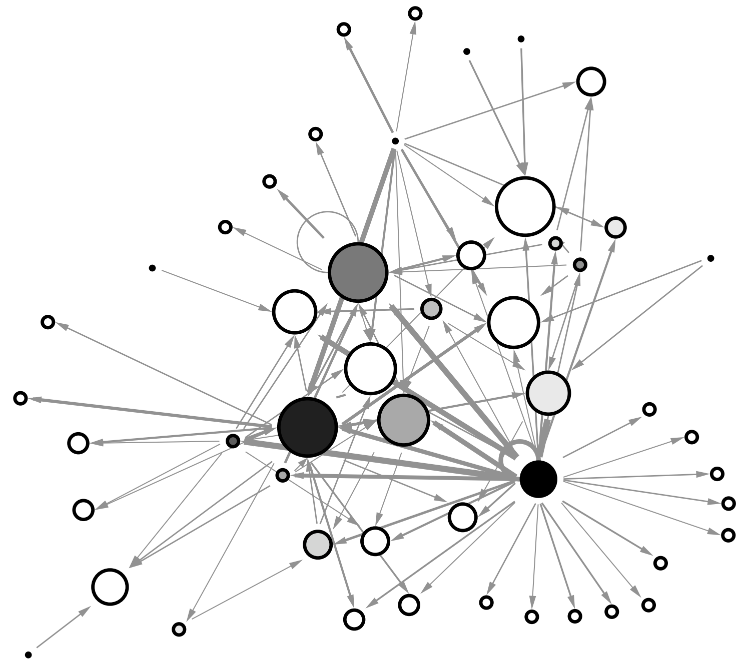


real-world examples

example financial trading networks



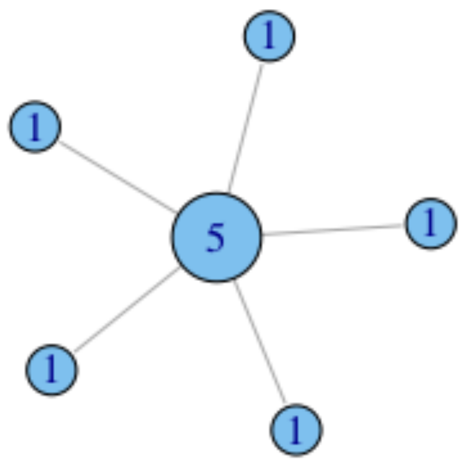
high in-centralization:
one node buying from
many others

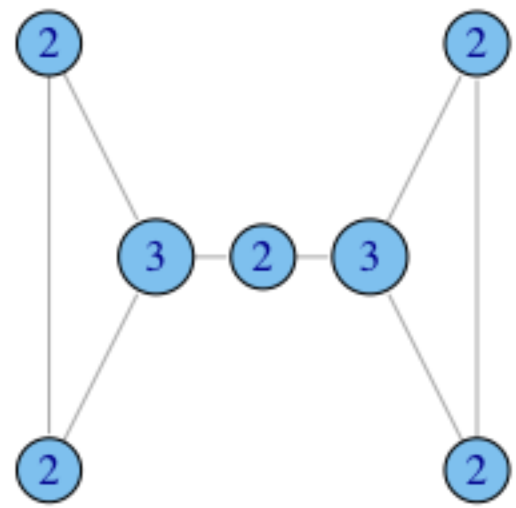


low in-centralization:
buying is more evenly
distributed

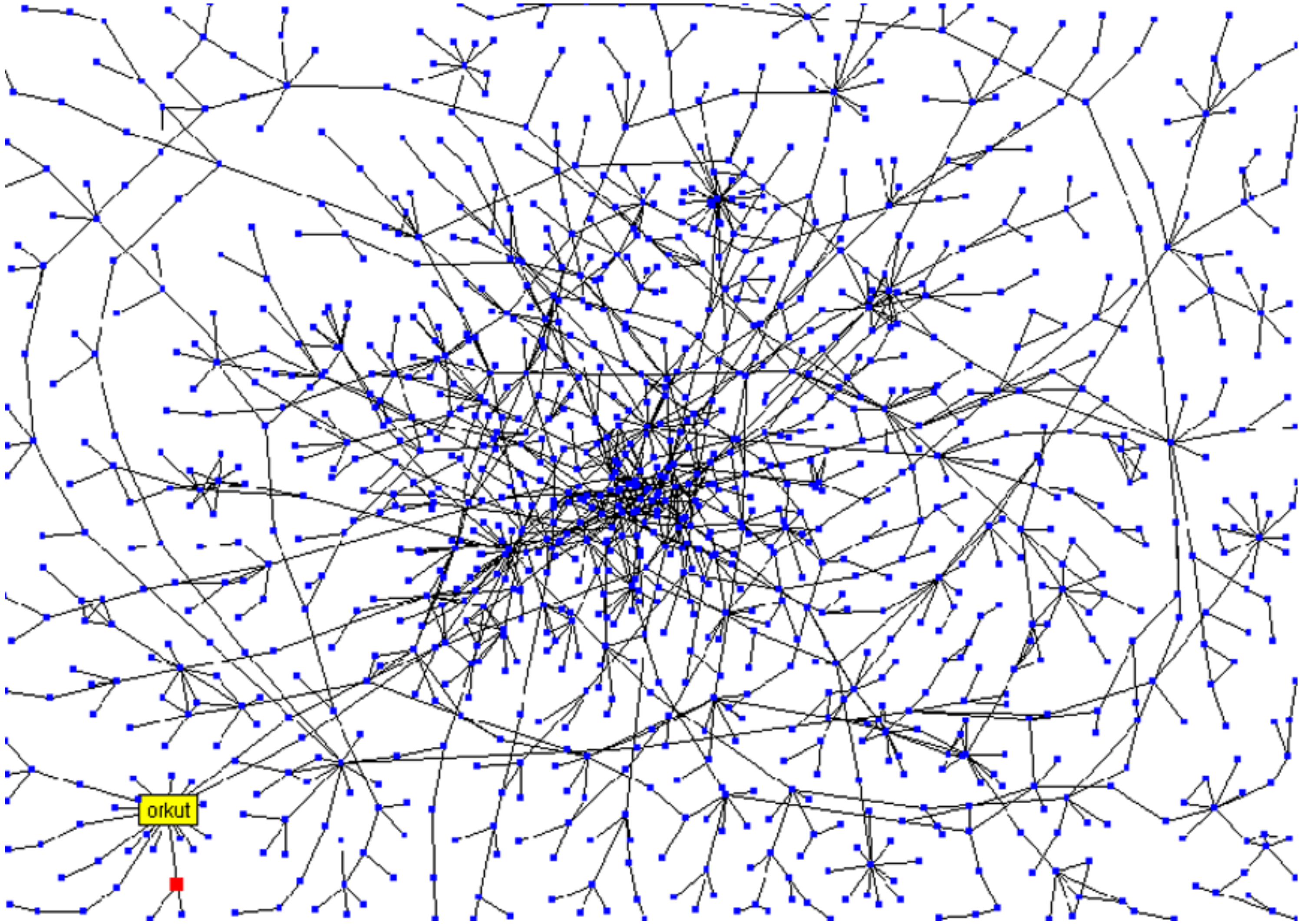
what does degree not capture?

In what ways does degree fail to capture centrality in the following graphs?



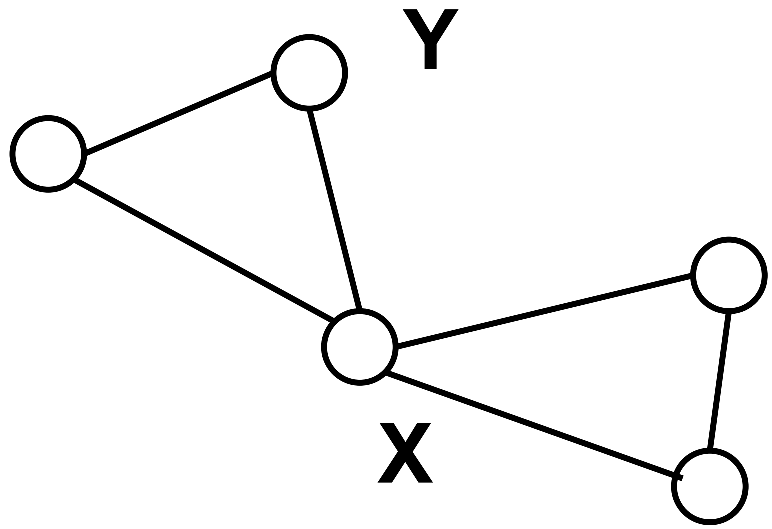


Stanford Social Web (ca. 1999)



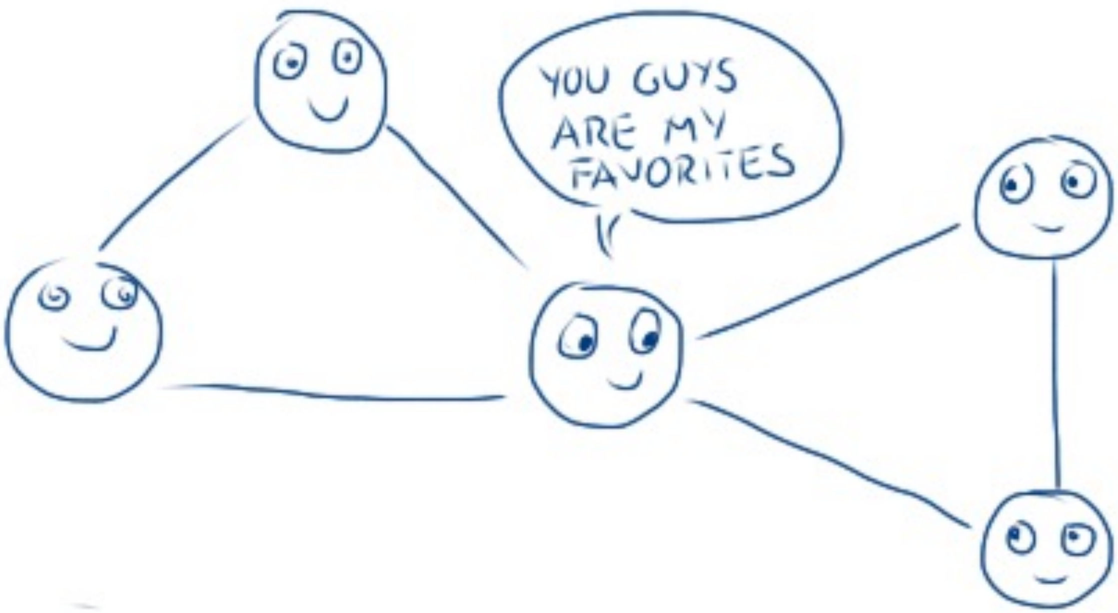
network of personal homepages at Stanford

Brokerage not captured by degree



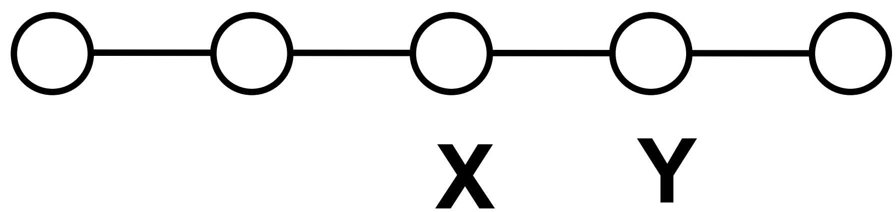
Constraint

constraint



Betweenness: capturing brokerage

- intuition: how many pairs of individuals would have to go through you in order to reach one another in the minimum number of hops?



Betweenness: definition

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

Where g_{jk} = the number of shortest paths connecting *jk*
$g_{jk}(i)$ = the number that actor *i* is on.

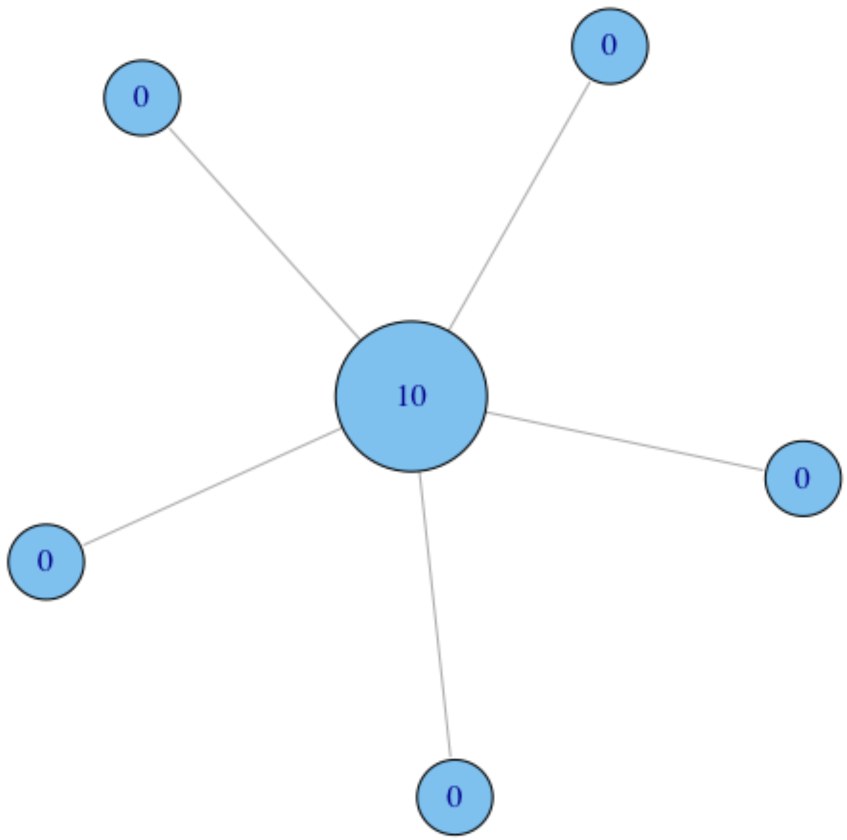
Usually normalized by:

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

number of pairs of vertices excluding the vertex itself

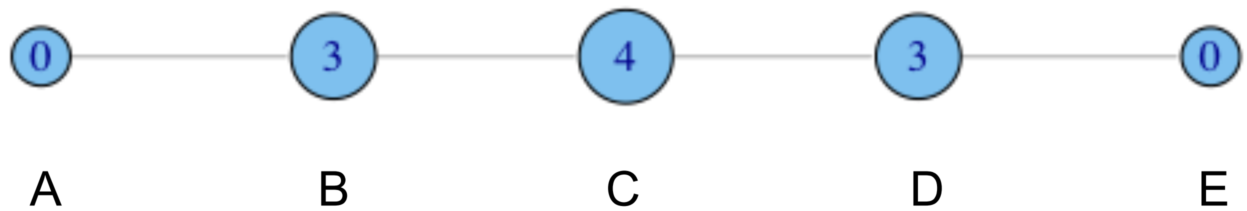
Betweenness on toy networks

- non-normalized version:



Betweenness on toy networks

- non-normalized version:

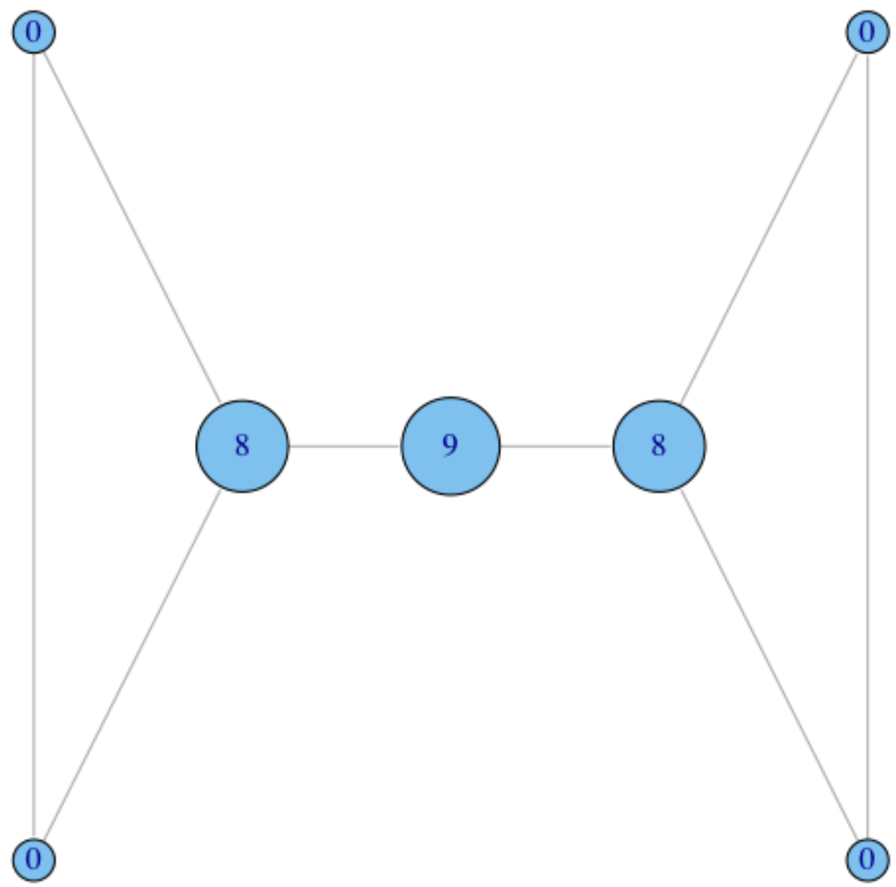


- A lies between no two other vertices
- B lies between A and 3 other vertices: C, D, and E
- C lies between 4 pairs of vertices (A,D),(A,E),(B,D),(B,E)

- note that there are no alternate paths for these pairs to take, so C gets full credit

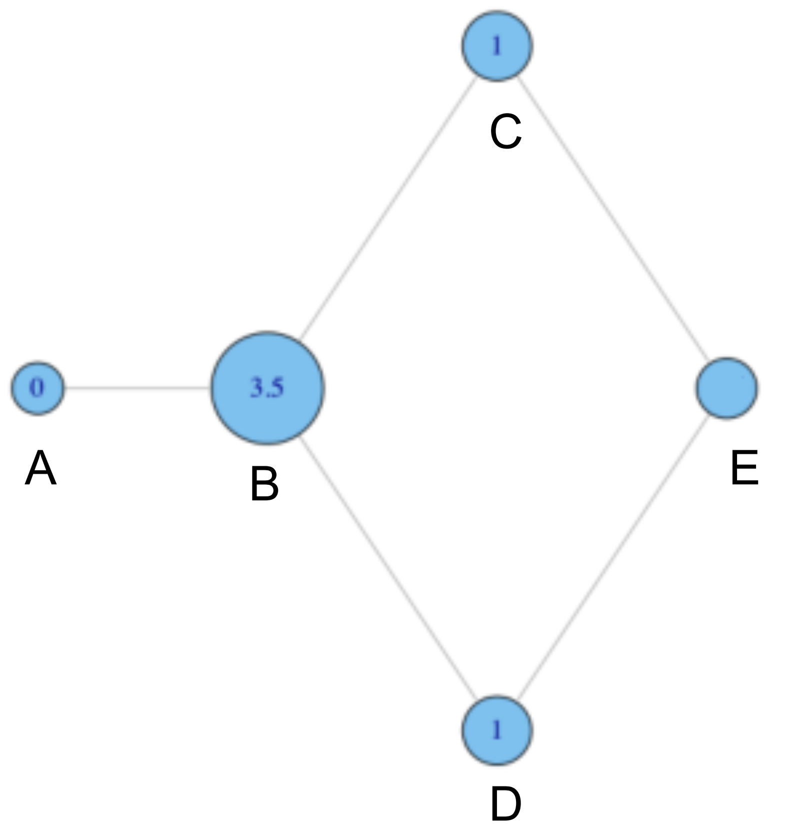
Betweenness on toy networks

- non-normalized version:



Betweenness on toy networks

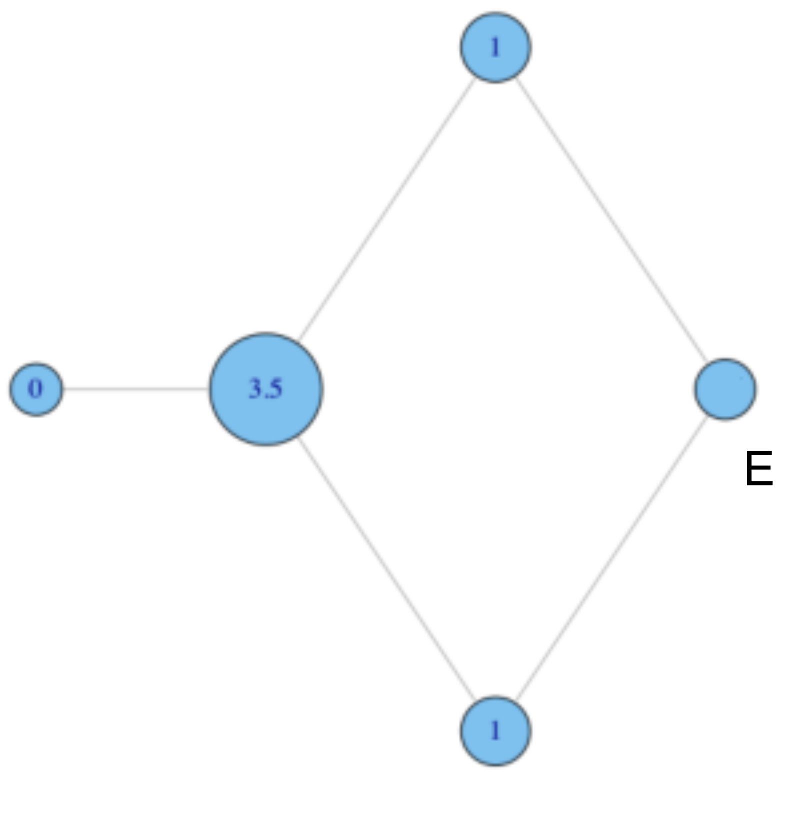
- non-normalized version:



- 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

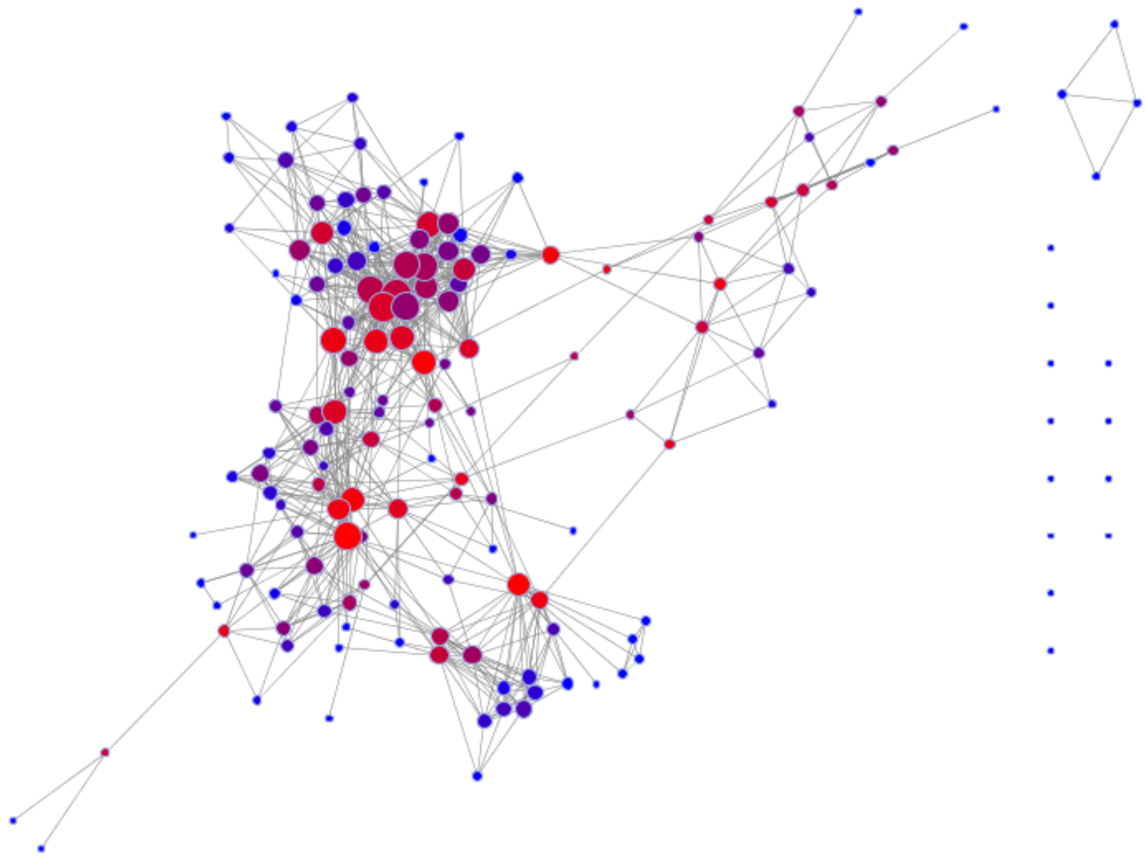
Q: betweenness

- What is the betweenness of node E?



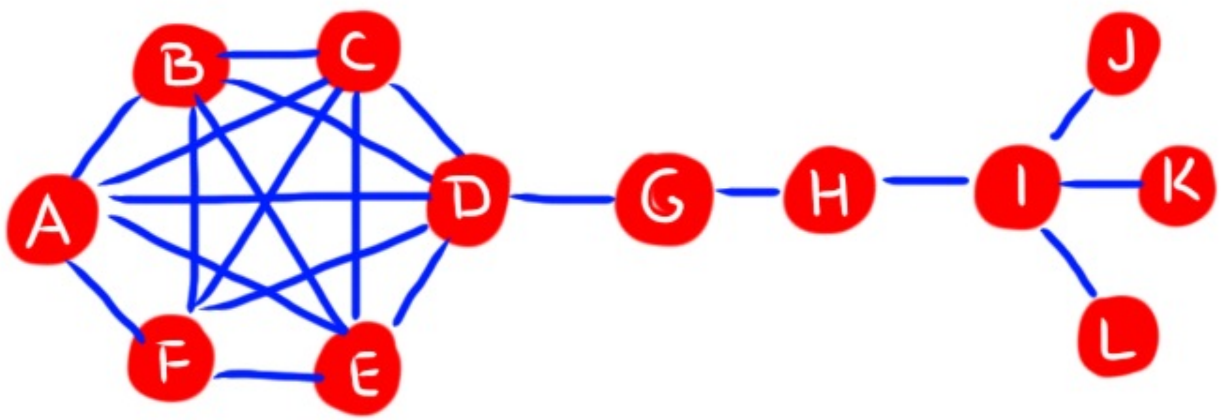
betweenness: example

Lada's old Facebook network: nodes are sized by degree, and colored by betweenness.



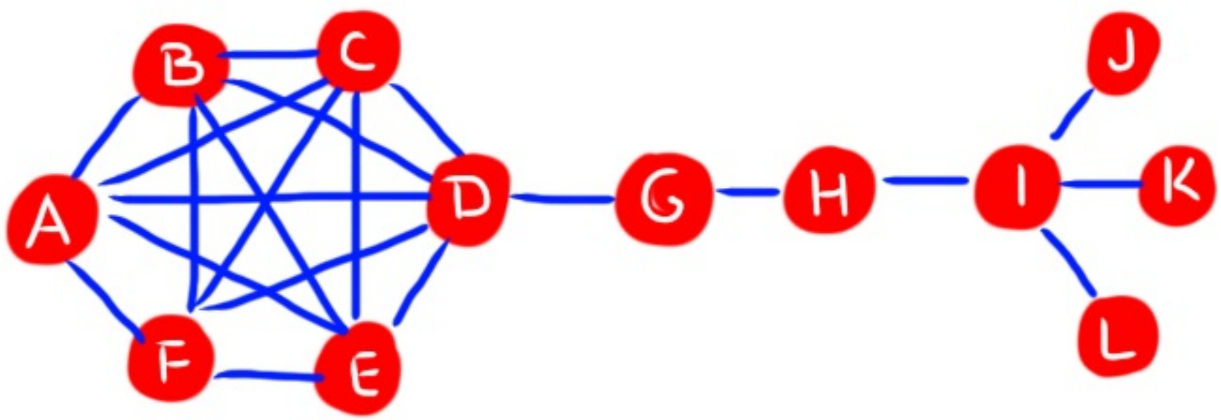
Q: high betweenness, low degree

- Find a node that has high betweenness but low degree



Q: low betweenness, high degree

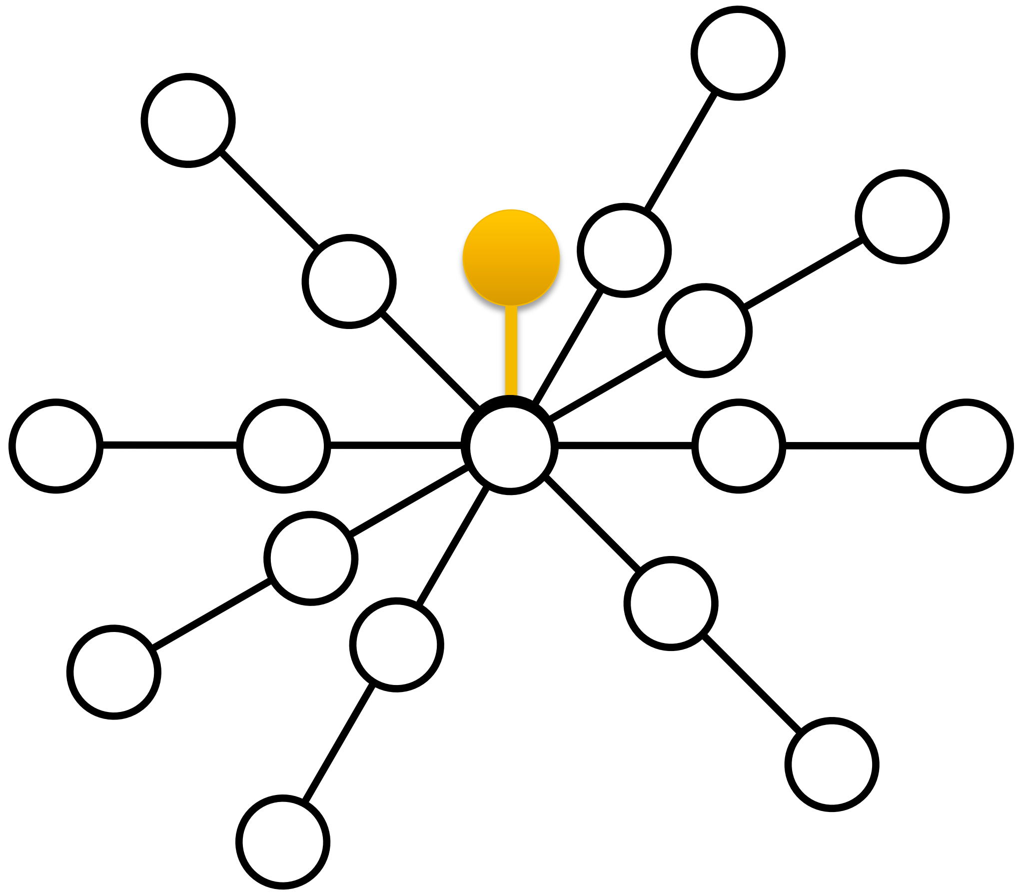
- Find a node that has low betweenness but high degree



closeness

- What if it's 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

need not be in a brokerage position



closeness: definition

Closeness is based on the length of the average shortest path between a node and all other nodes in the network

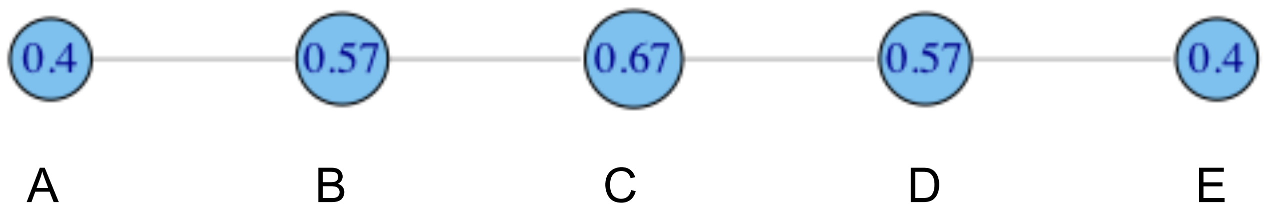
Closeness Centrality:

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

Normalized Closeness Centrality

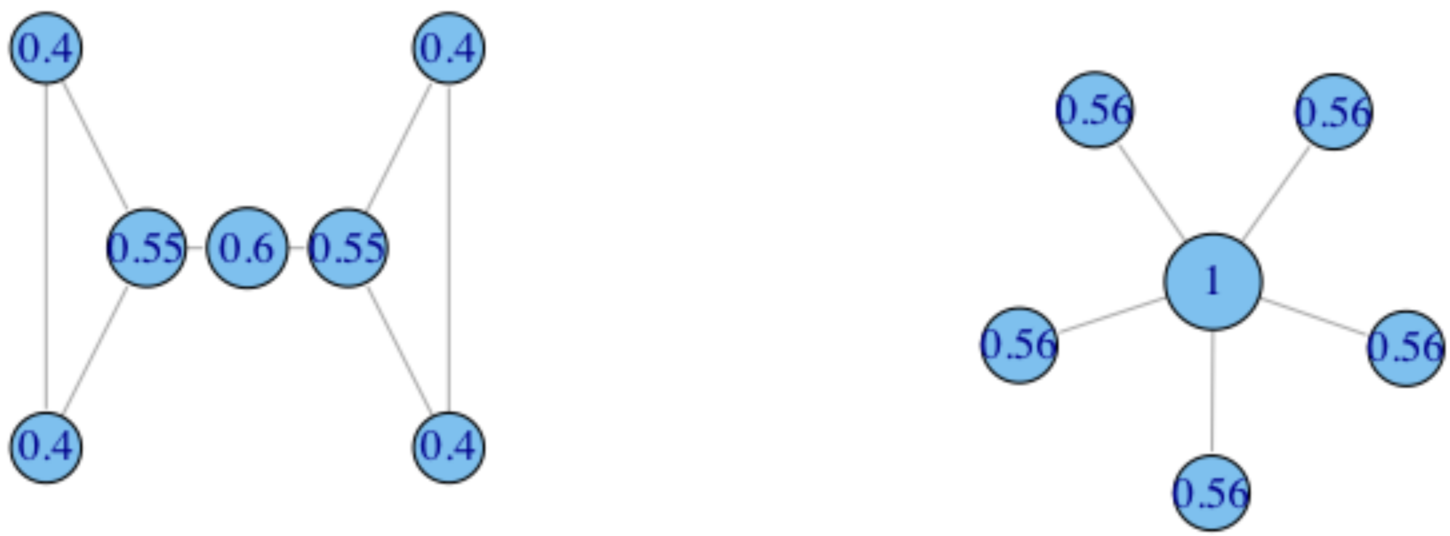
$$C'_C(i) = (C_C(i))/(N-1)$$

Closeness: toy example



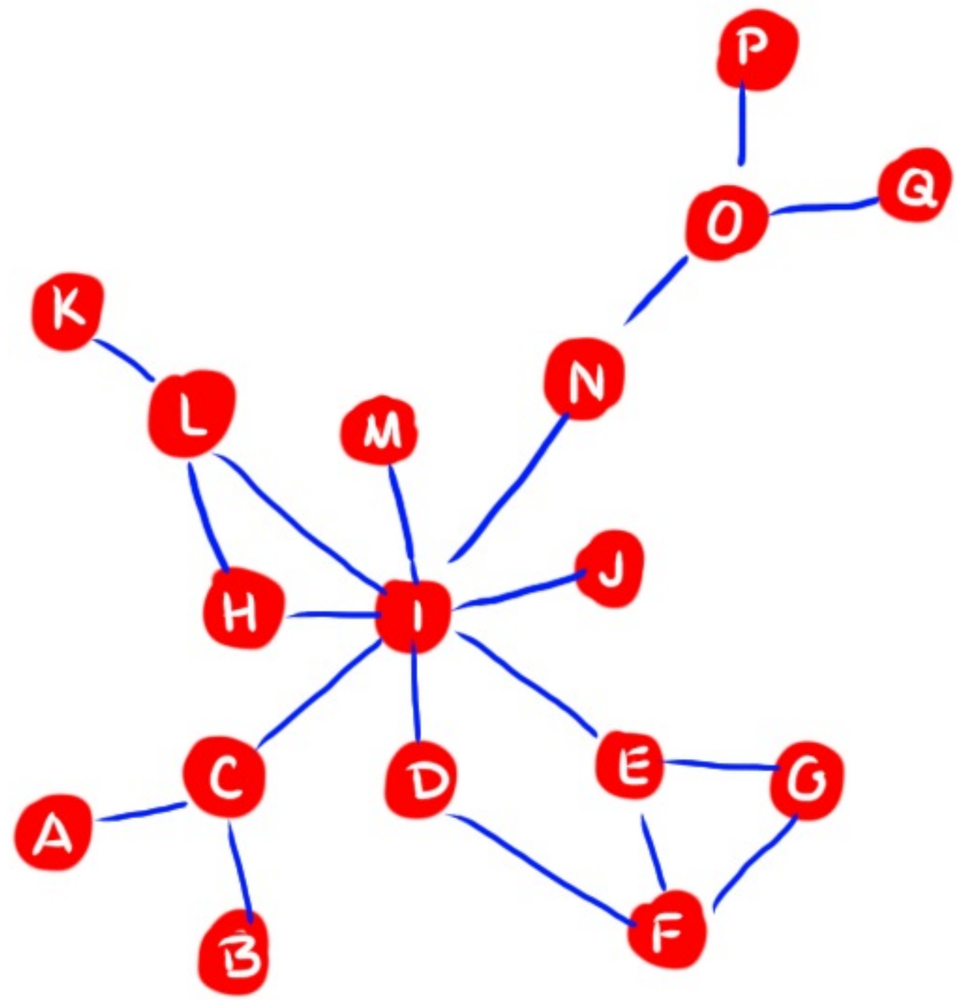
$$C_c'(A) = \left[\frac{\sum_{j=1}^{N} d(A,j)}{N-1}\right]^{-1} = \left[\frac{1+2+3+4}{4}\right]^{-1} = \left[\frac{10}{4}\right]^{-1} = 0.4$$

Closeness: more toy examples



Q:high degree, low closeness

Which node has relatively high degree but low closeness?



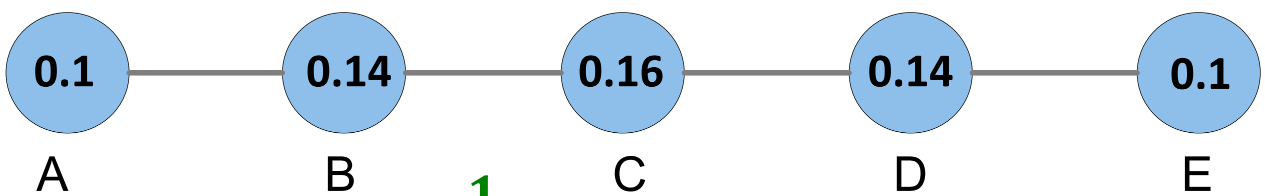
Closeness Centrality

- **Geometric measures**
  - **Closeness Centrality**:

$$c_{\text{clos}}(x) = \frac{1}{\sum_y d(y,x)}$$

length of the shortest path from x to y

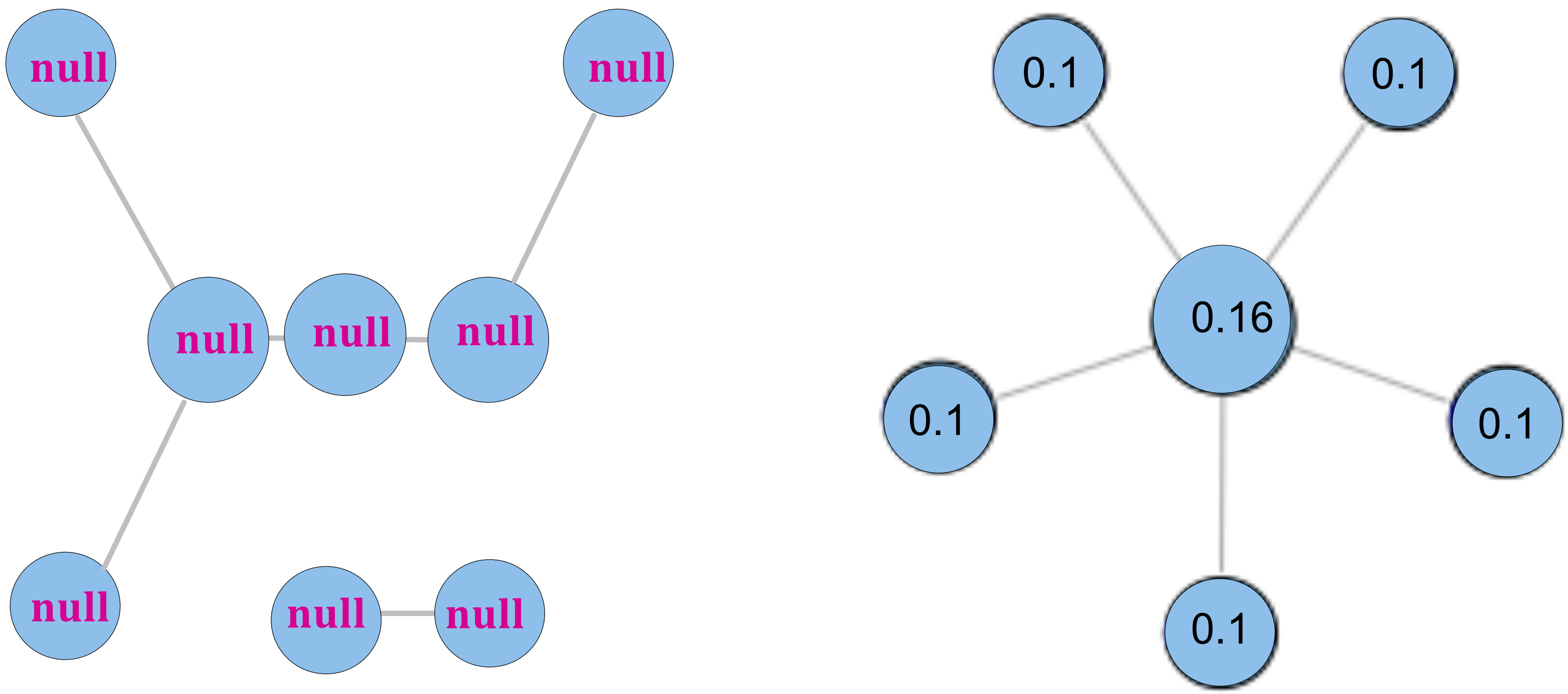
  - How much a vertex can communicate without relying on third parties for his messages to be delivered



$$c_{clos}(A) = \frac{1}{1+2+3+4} = 0.1$$

- **Problem**: The graph must be **(strongly) connected!**

Closeness centrality: Example



We get null score for all nodes, if the graph is not connected!

$$c_{\text{clos}}(x) = \frac{1}{\sum_y d(y,x)}$$

Centrality Measures

- **Geometric measures**
  - **Harmonic Centrality**: Who are the **bridges**?
    - Replace the average distance with the harmonic mean of all distances.
    - The $n(n-1)$ distances between every pair of distinct nodes:

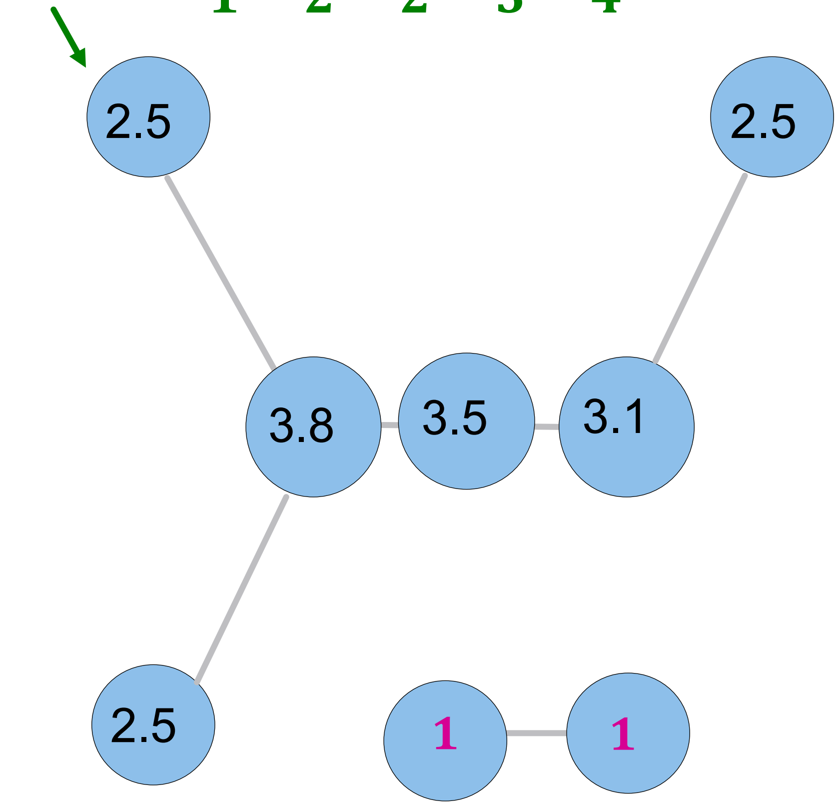
$$c_{\text{har}}(x) = \boxed{\sum_{y \neq x} \frac{1}{d(y,x)}} = \sum_{d(y,x)<\infty, y \neq x} \frac{1}{d(y,x)}$$

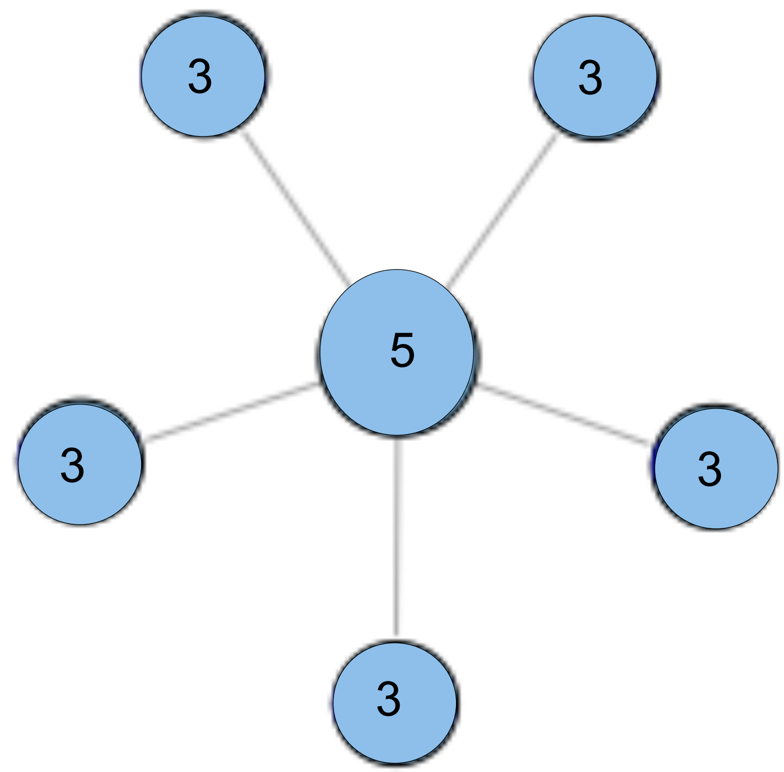
Harmonic mean

    - Strongly correlated to closeness centrality
    - Naturally also accounts for nodes y that cannot reach x
    - Can be applied to graphs that are **not strongly connected**

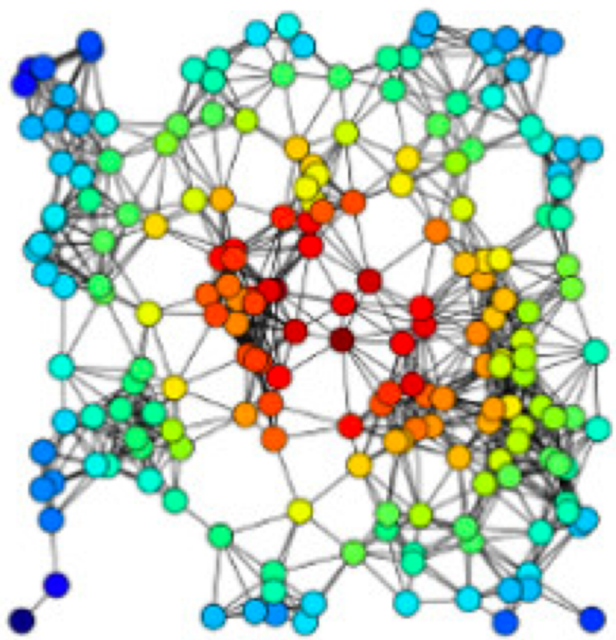
Harmonic centrality: Example

$$c_{harm} = \frac{1}{1} + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = 2.5$$



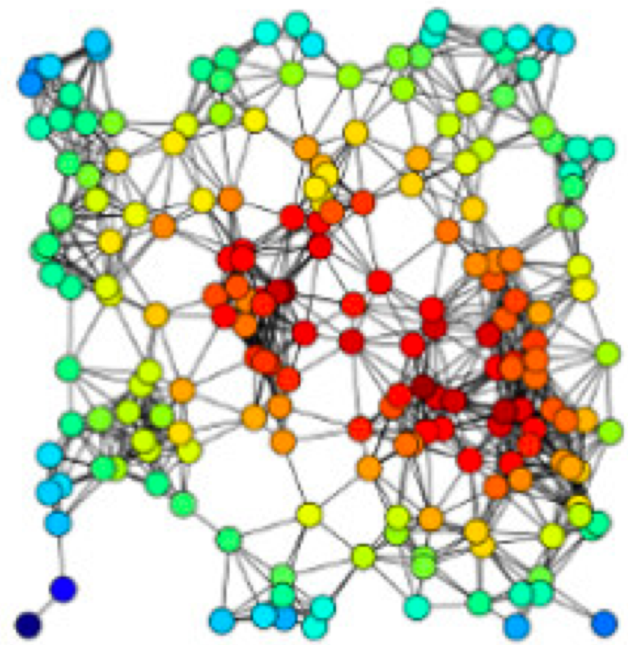


Closeness vs Harmonic centrality



Closeness

Red nodes are closer to all the other nodes



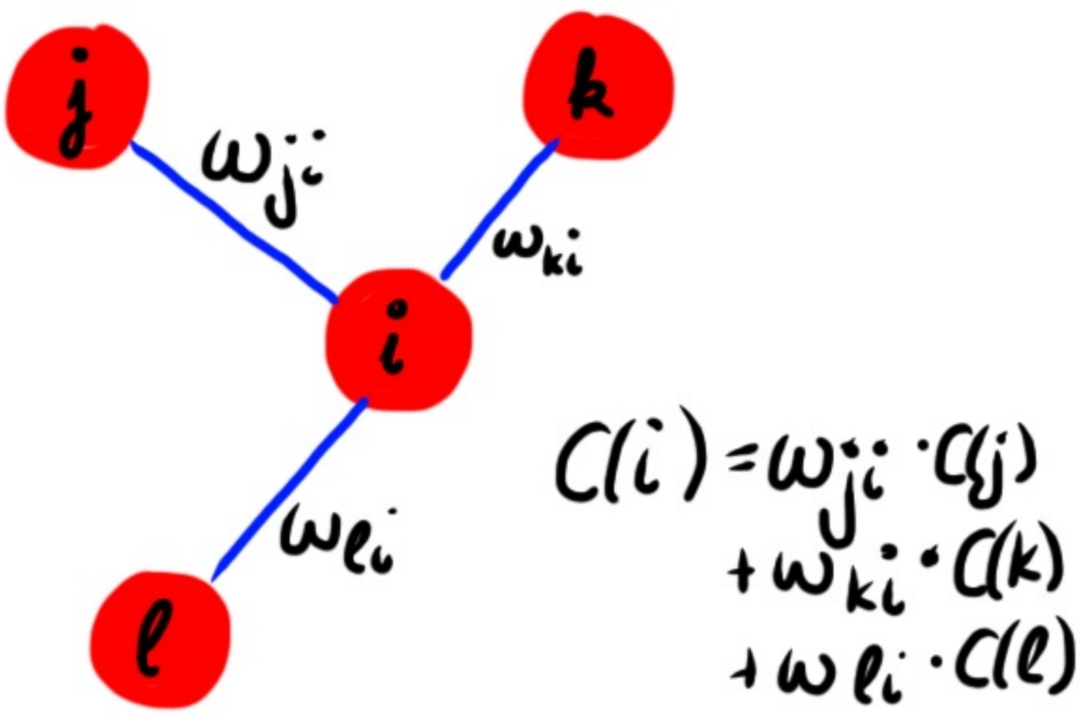
Harmonic

Red nodes are closer to all the other nodes, and have larger degrees

Examples of Closeness centrality, and Harmonic Centrality of the same graph.

Eigenvector centrality

- How central you are depends on how central your neighbors are



Bonacich eigenvector centrality

$$c_i(\beta) = \sum_j (\alpha + \beta c_j) A_{ji}$$

$$c(\beta) = \alpha (I - \beta A)^{-1} A\mathbf{1}$$

- α is a normalization constant
- β determines how important the centrality of your neighbors is
- $\mathbf{A}$ is the adjacency matrix (can be weighted)
- $\mathbf{I}$ is the identity matrix (1s down the diagonal, 0 off-diagonal)
- $\mathbf{1}$ is a matrix of all ones.

Bonacich Power Centrality: attenuation factor β

small β ➔ high attenuation

only your immediate friends matter, and their importance is factored in only a bit

high β ➔ low attenuation

global network structure matters (your friends, your friends' of friends etc.)

β = 0 yields simple degree centrality

$$c_i(\beta) = \sum_j (\alpha \quad\quad) A_{ji}$$

Centrality in directed networks

- WWW
- food webs
- population dynamics
- influence
- hereditary
- citation
- transcription regulation networks
- neural networks

Betweenness centrality in directed networks

- 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

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

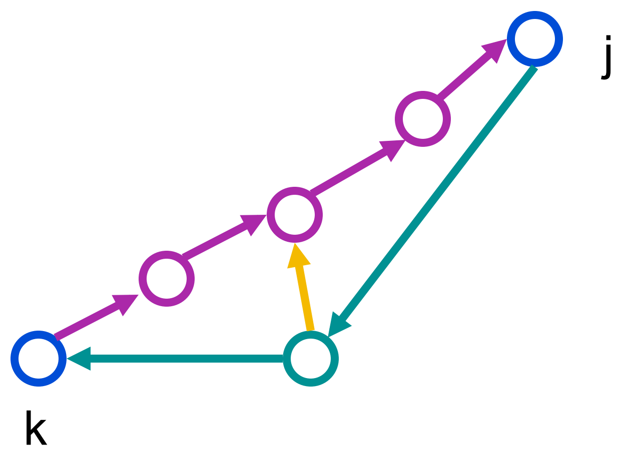
all paths between j and k

- 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 geodesics

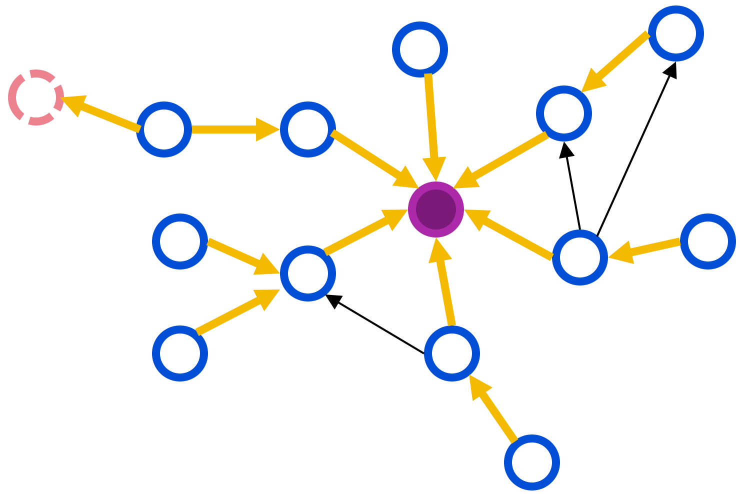
- A node does not necessarily lie on a geodesic (shortest path) from *j* to *k* if it lies on a geodesic from *k* to *j*



Directed closeness centrality

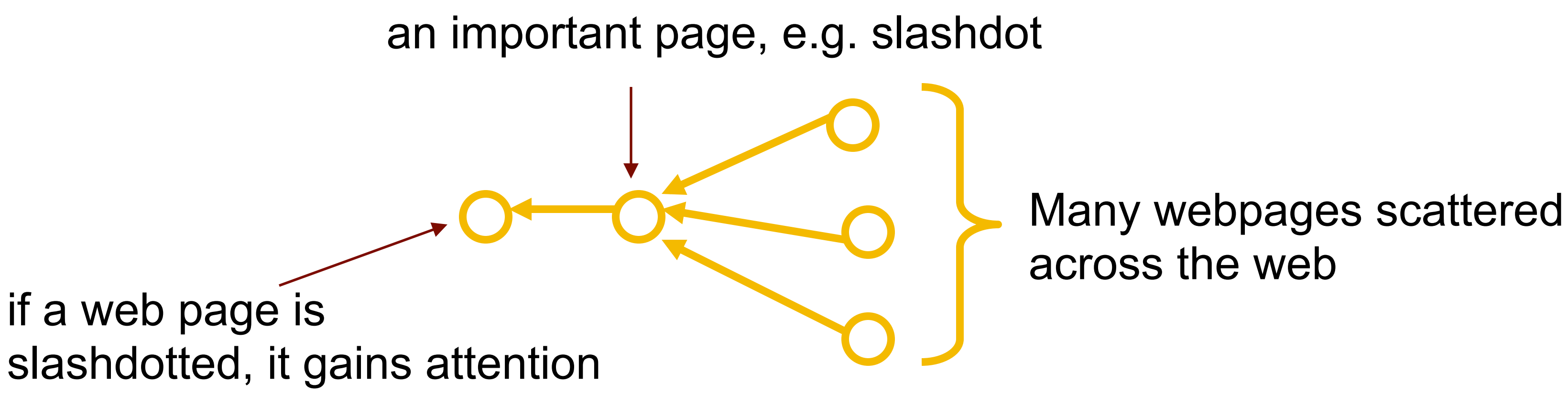
- choose a direction
  - in-closeness (e.g. prestige in citation networks)
  - out-closeness

- usually consider only vertices from which the node *i* in question can be reached



Eigenvector centrality in directed networks

- PageRank (centrality) brings order to the Web:
  - it's not just the pages that point to you, but how many pages point to those pages, etc.
  - more difficult to artificially inflate centrality with a recursive definition



Ranking pages by tracking a drunk

- A random walker following edges in a network for a very long time will spend a proportion of time at each node which can be used as a measure of importance