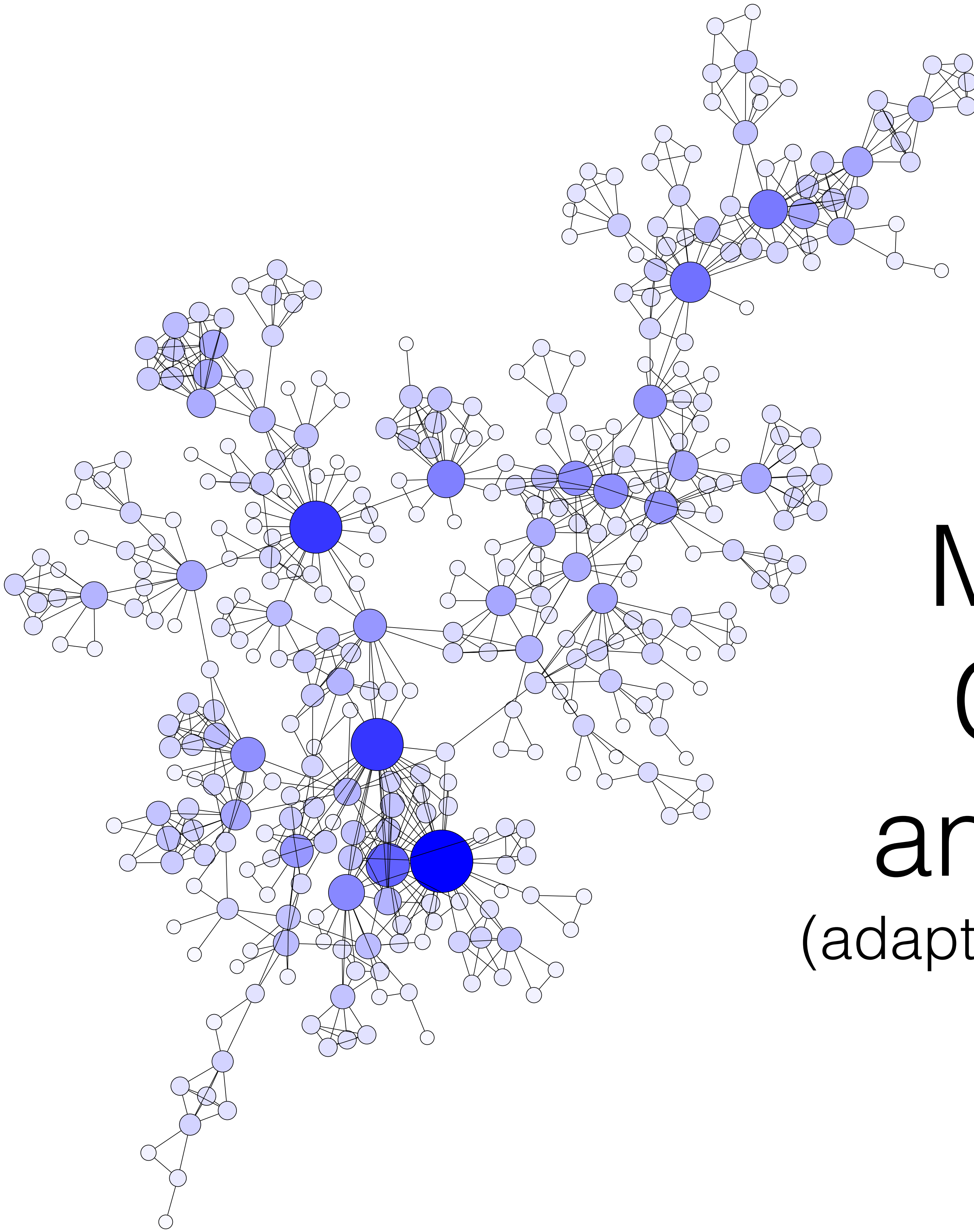
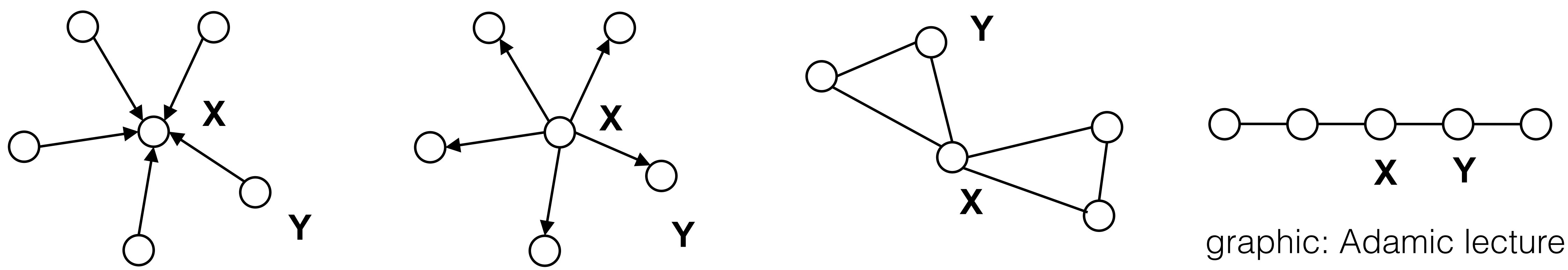


Network Measures: Centrality and Prestige

(adapted from Lada Adamic)

Centrality and Prestige

Some nodes are more important than others



graphic: Adamic lecture

But what it means to be "important" depends on the context: exchange, spread of information, brokerage opportunities, etc.

Centrality measures give us a way to quantify the different ways that a node can be important

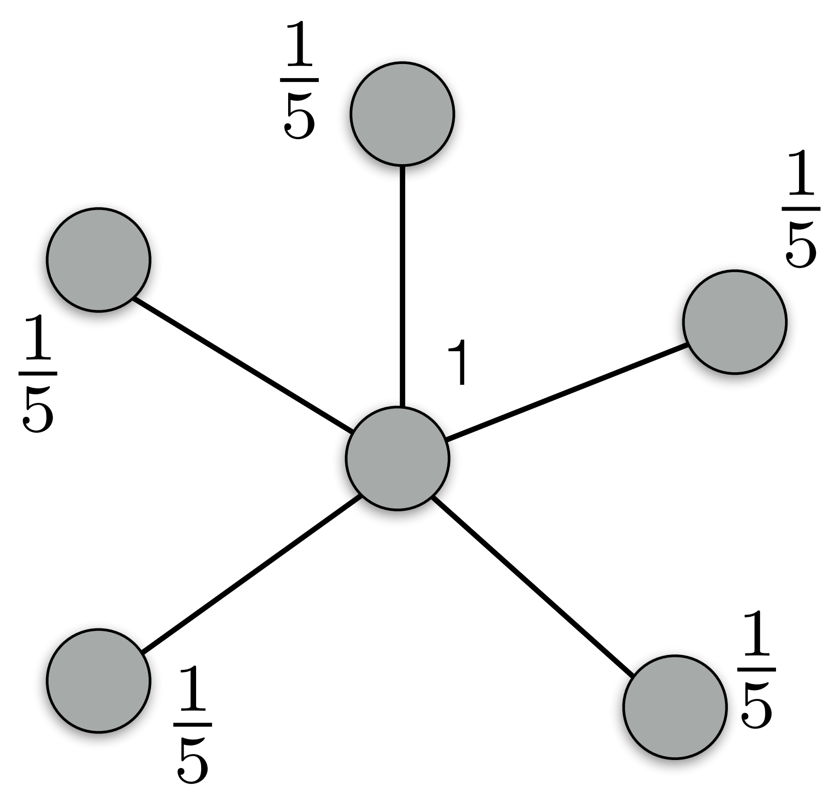
Centrality and Prestige

Today:

- Tour through a variety of centrality measures:
    - Degree
    - Betweenness
    - Closeness
    - Eigenvector
- Look at how centrality is distributed: centralization
- Centrality on a directed network: prestige

Degree Centrality

- First notion: the person with the most connections is most important

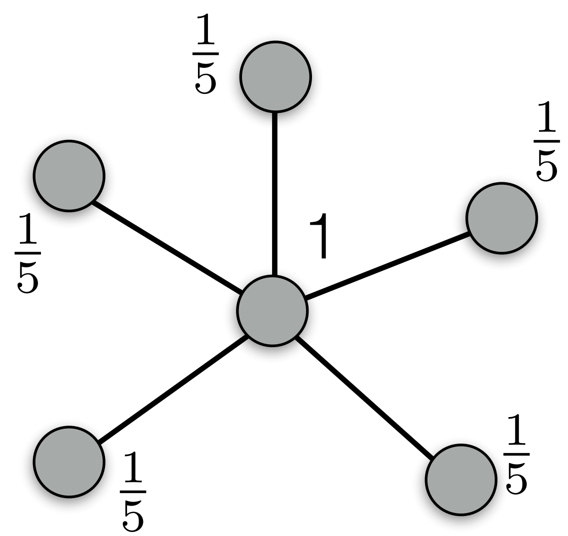


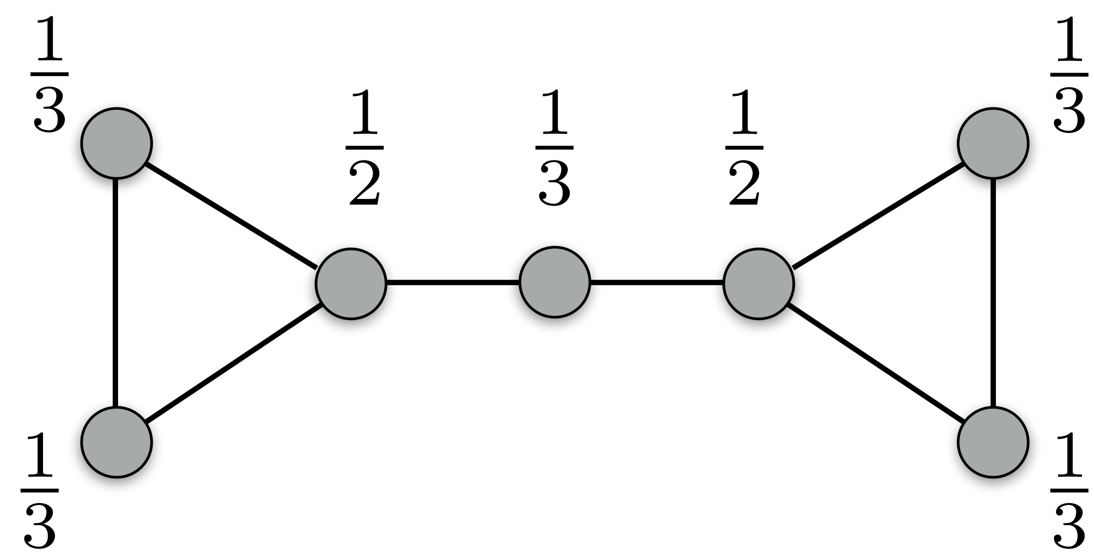
$$C_d(v_i) = \frac{1}{N-1} d_i$$

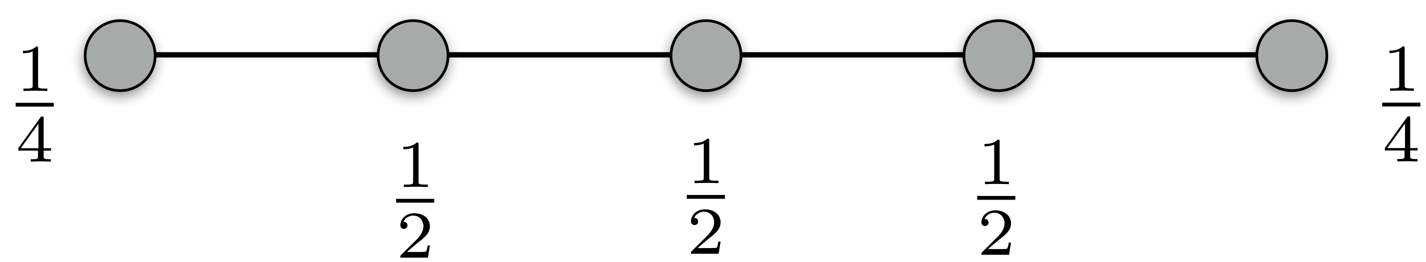
- Normalize by the maximum possible (N-1)

Degree Centrality

$$C_d(v_i) = \frac{1}{N-1} d_i$$

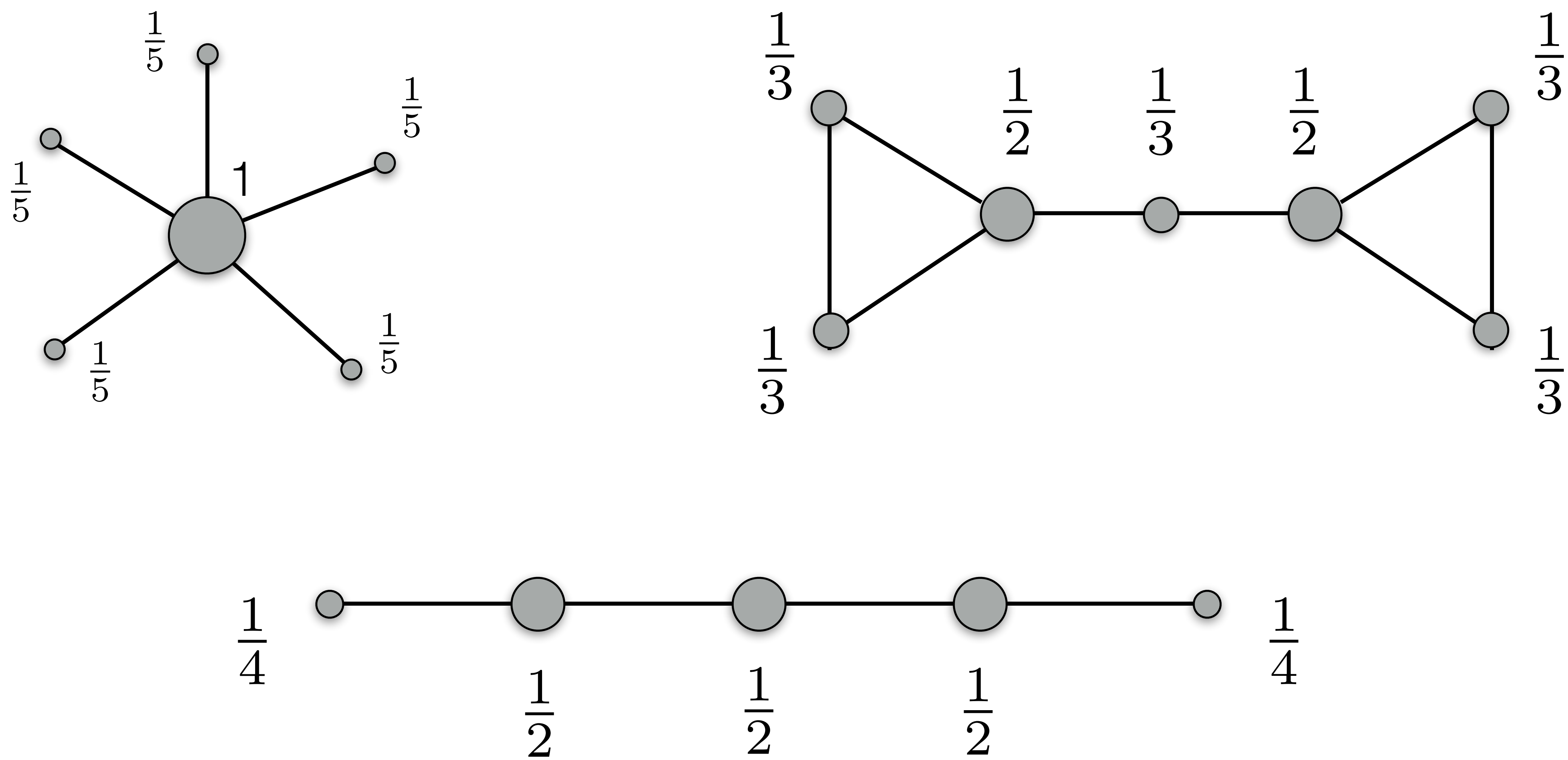






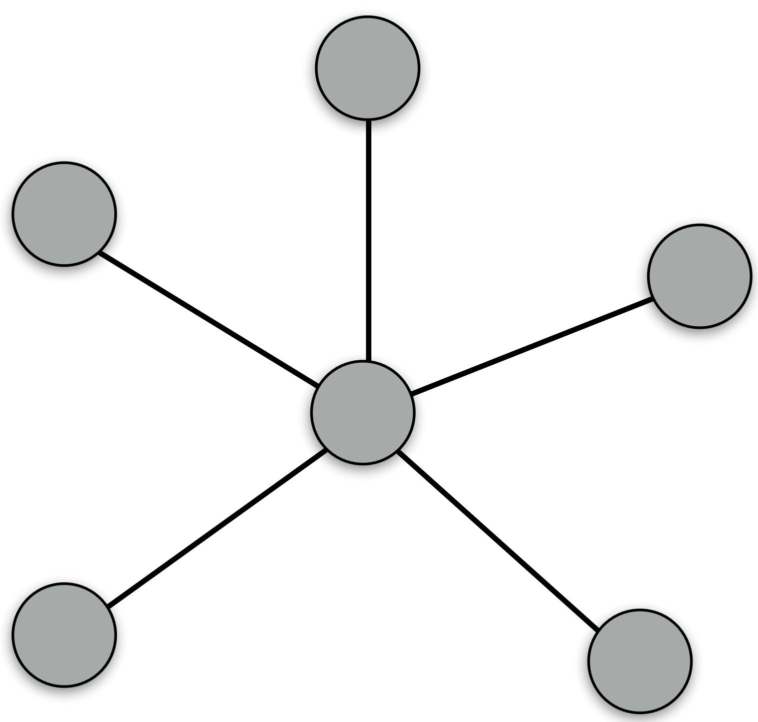
Degree Centrality

$$C_d(v_i) = \frac{1}{N-1} d_i$$



Degree Centrality

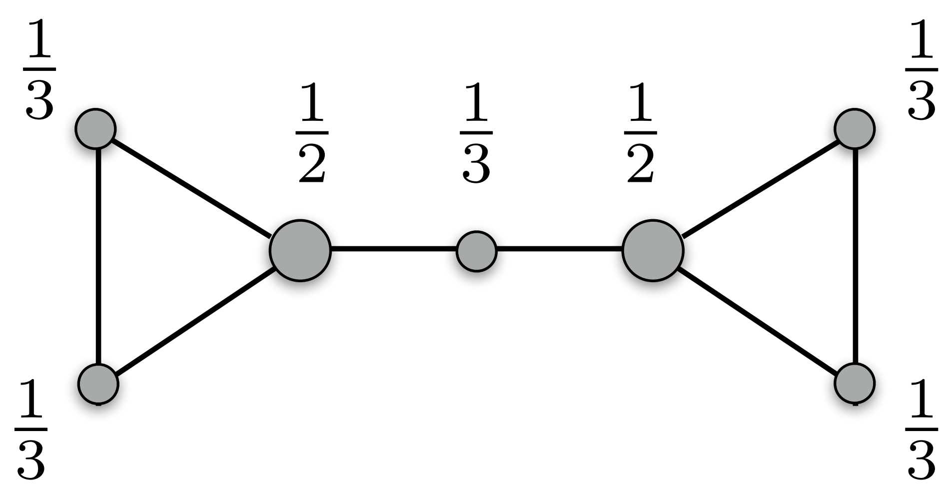
- Degree centrality makes sense when sheer number of contacts is important:

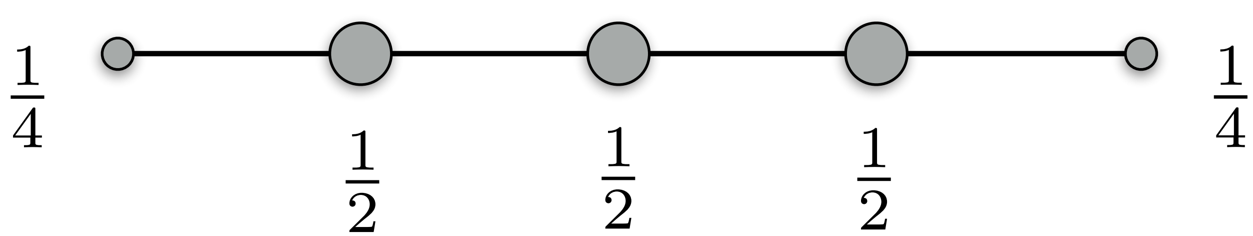


  - Number of supporters
  - Number of confidants
  - Audience size
  - Number of trading partners
  - Number of direct reports

Degree Centrality

- Clearly, there are some contexts where degree isn't exactly what we mean by "centrality"





- Suppose we are interested in who gets access to information?
- Or who can broker between different groups?

Closeness Centrality

- Second notion: the person in the middle of the action is most central

$$C_C(v_i) = \frac{\boxed{(N-1)}}{\boxed{\sum_{v_j \in G} d(v_i, v_j)}}$$

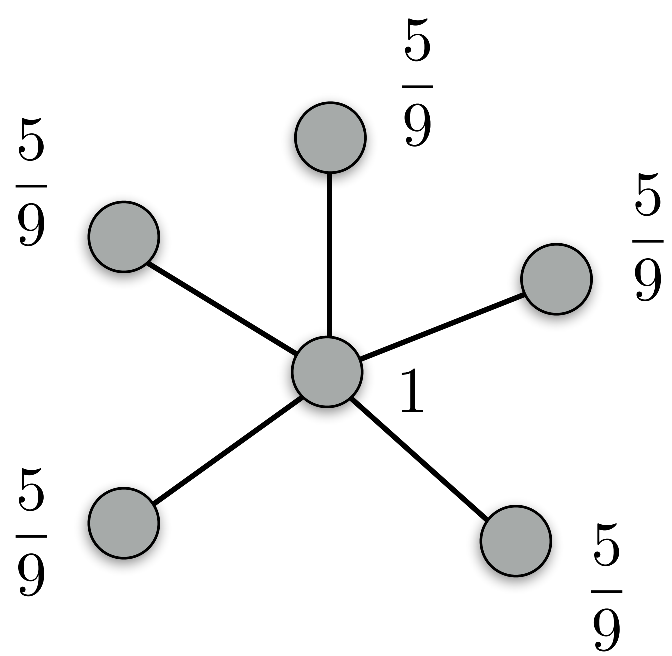
Normalization (min possible distance to the N-1 other nodes)

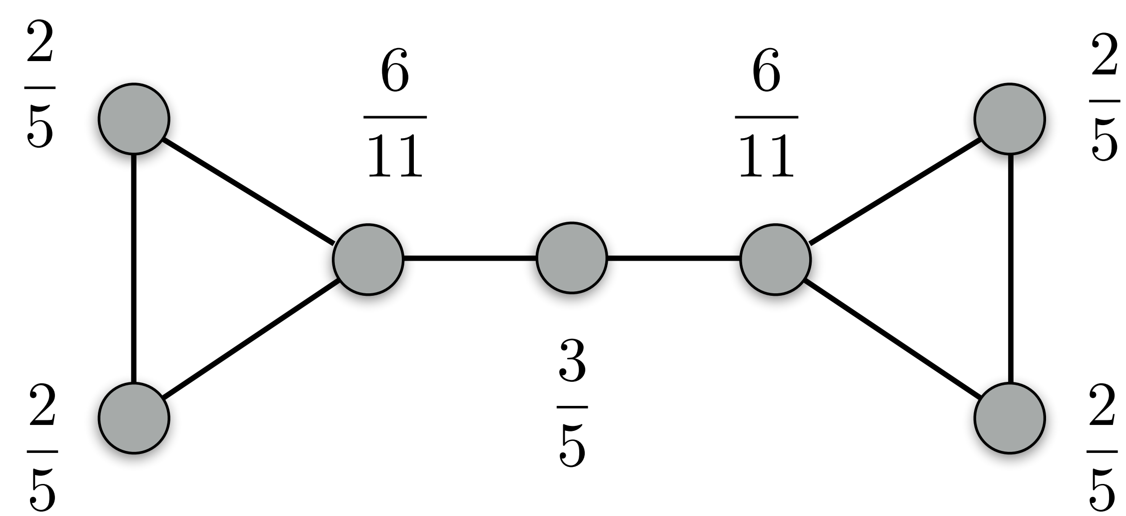
Total distance to the other nodes

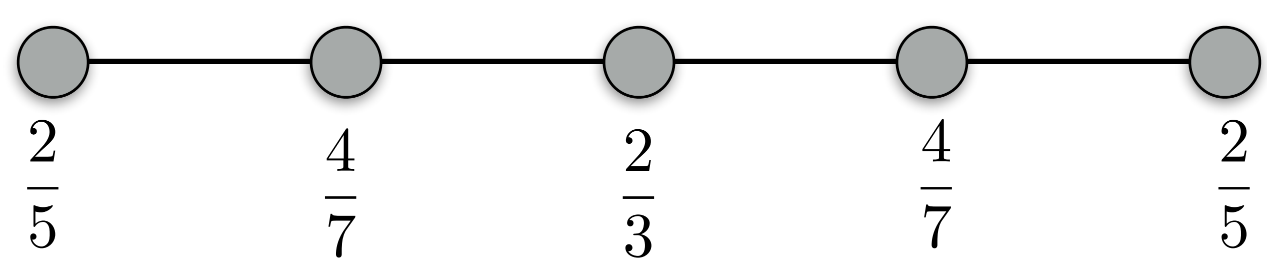
- Person with the highest closeness centrality has the shortest distance to the other nodes, on average

Closeness Centrality

$$C_C(v_i) = \frac{(N-1)}{\sum_{v_j \in G} d(v_i, v_j)}$$

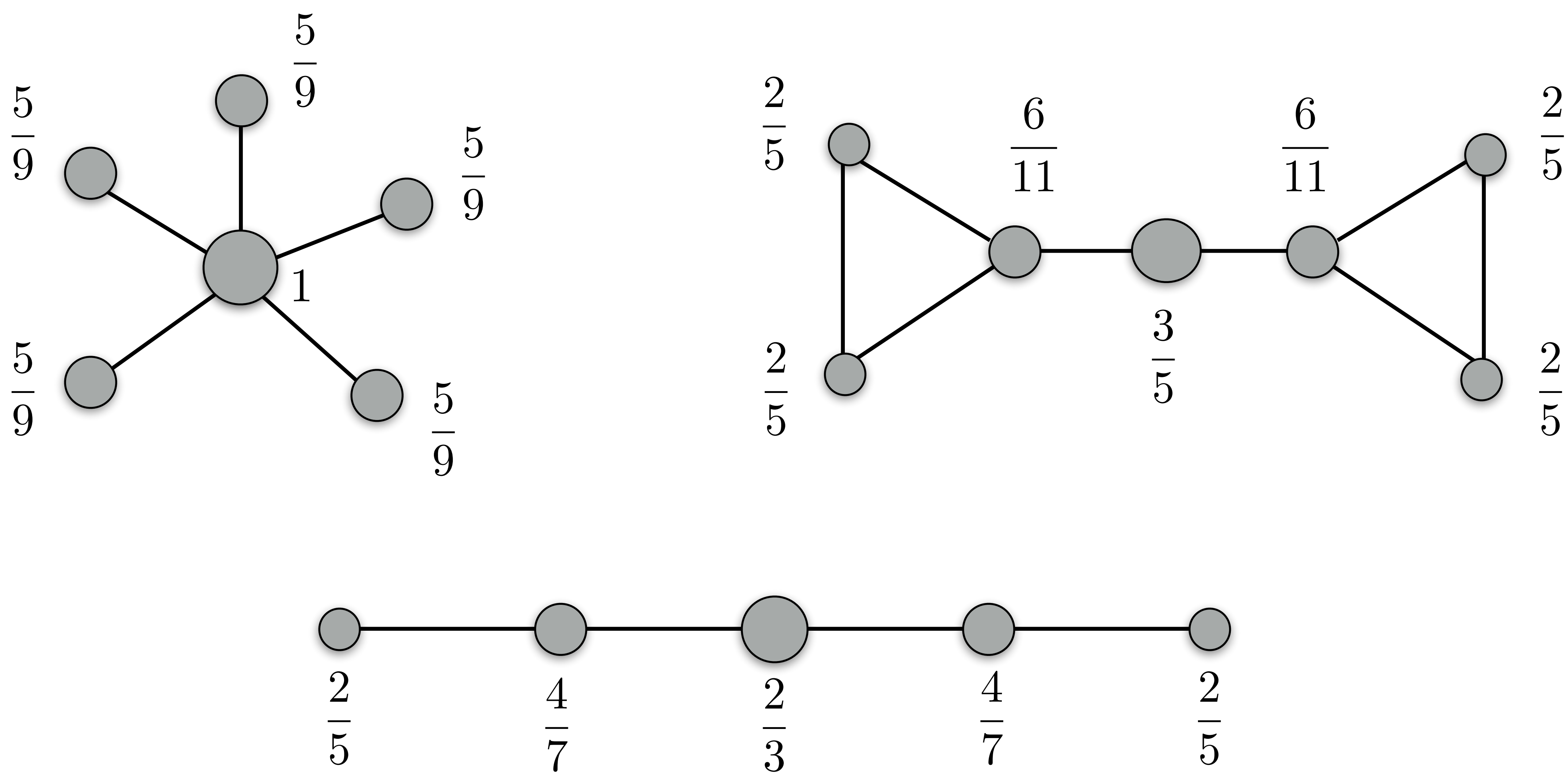






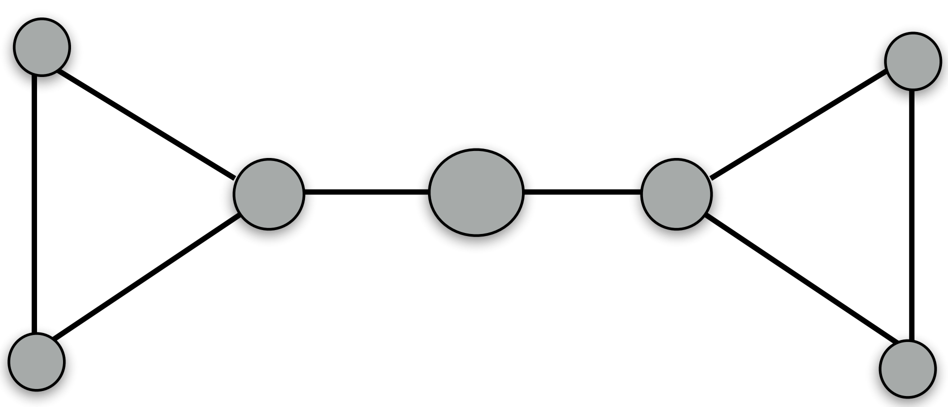
Closeness Centrality

$$C_C(v_i) = \frac{(N-1)}{\sum_{v_j \in G} d(v_i, v_j)}$$



Closeness Centrality

- Closeness centrality makes sense whenever direct access is important



- Access to information
- Opinion formation
- Spread of disease
- Adoption of new technology

Betweenness Centrality

- Third notion: people you have to go through are most central
  - For each pair of nodes in the network, what fraction of shortest paths go through the node?

Fraction of geodesics going through the node

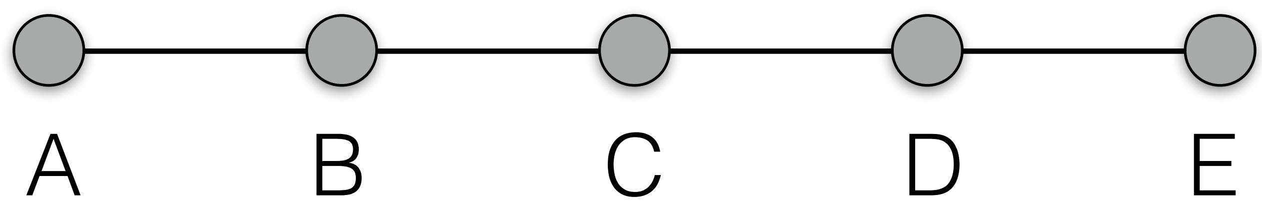
$$C_B(v_i) = \frac{\boxed{\sum_{j<k} \frac{g_{jk}(v_i)}{g_{jk}}}}{\boxed{(N-1)(N-2)/2}}$$

Normalization (number of pairs of nodes)

where g_{jk} is the number of geodesics between j and k and $g_{jk}(v_i)$ is the number that go through i

Betweenness Centrality

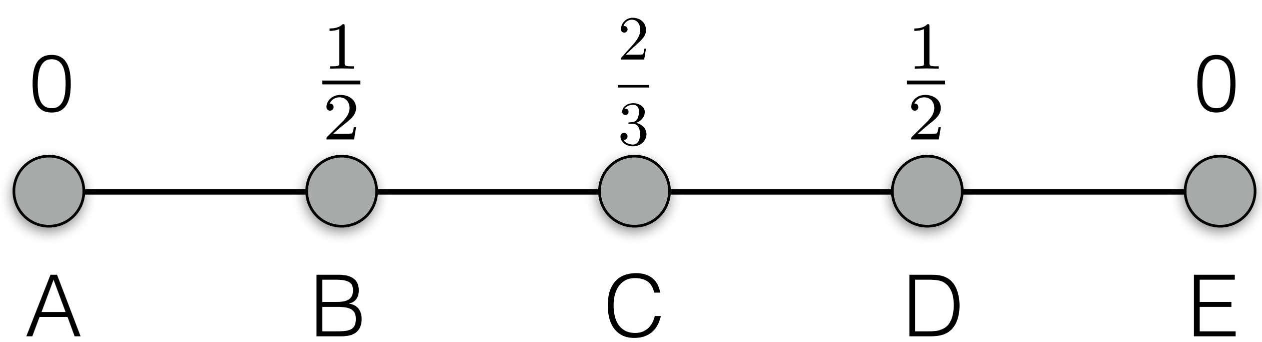
$$C_B(v_i) = \frac{\sum_{j<k} \frac{g_{jk}(v_i)}{g_{jk}}}{(N-1)(N-2)/2}$$



- A and E are not on any shortest paths
- B and D are both on 3 shortest paths
- C is on 4 shortest paths
- There are (N-1)(N-2)/2 = 6 total paths

Betweenness Centrality

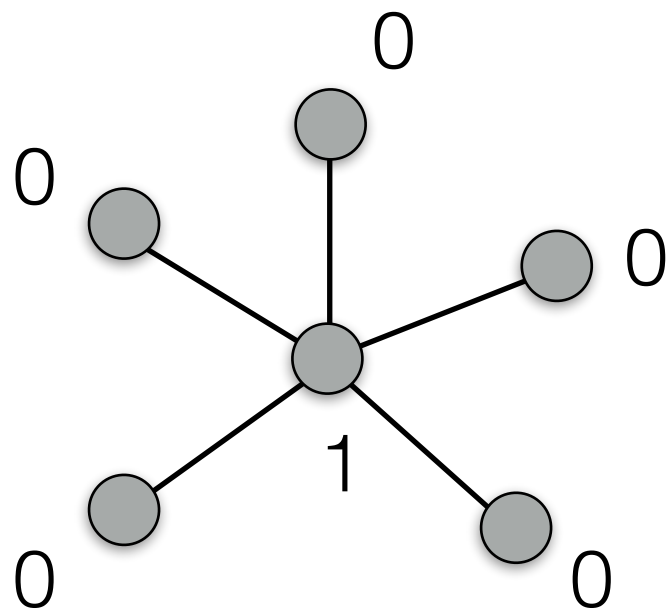
$$C_B(v_i) = \frac{\sum_{j<k} \frac{g_{jk}(v_i)}{g_{jk}}}{(N-1)(N-2)/2}$$

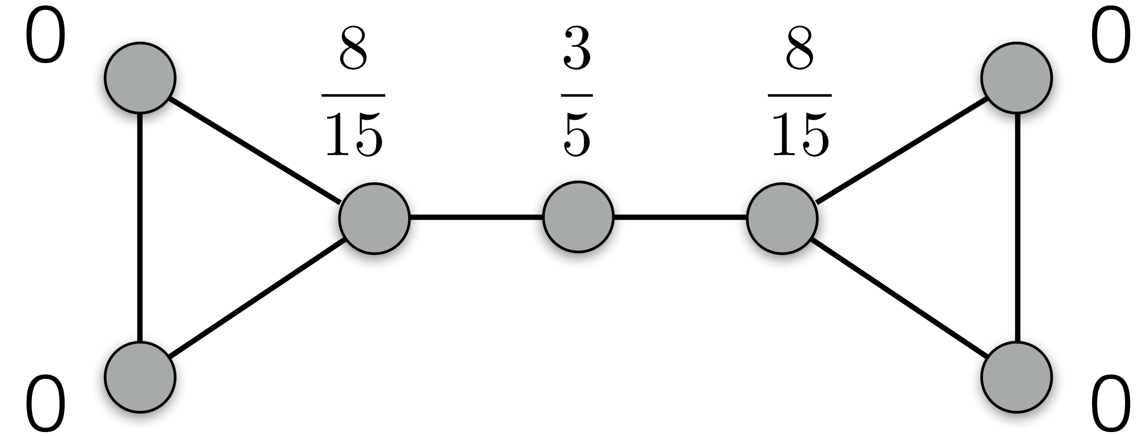


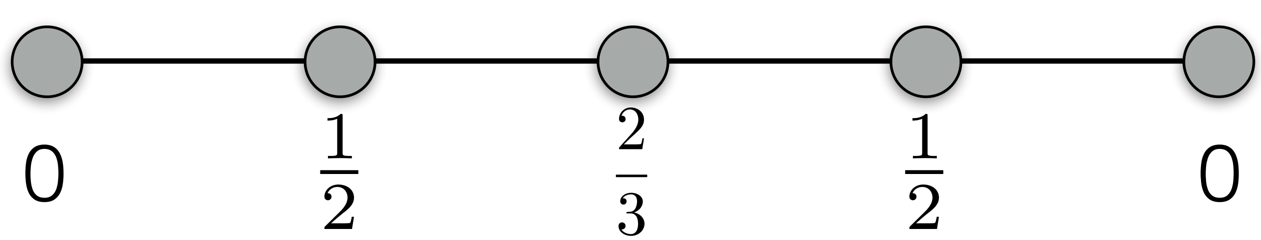
- A and E are not on any shortest paths
- B and D are both on 3 shortest paths
- C is on 4 shortest paths
- There are (N-1)(N-2)/2 = 6 total paths

Betweenness Centrality

$$C_B(v_i) = \frac{\sum_{j<k} \frac{g_{jk}(v_i)}{g_{jk}}}{(N-1)(N-2)/2}$$

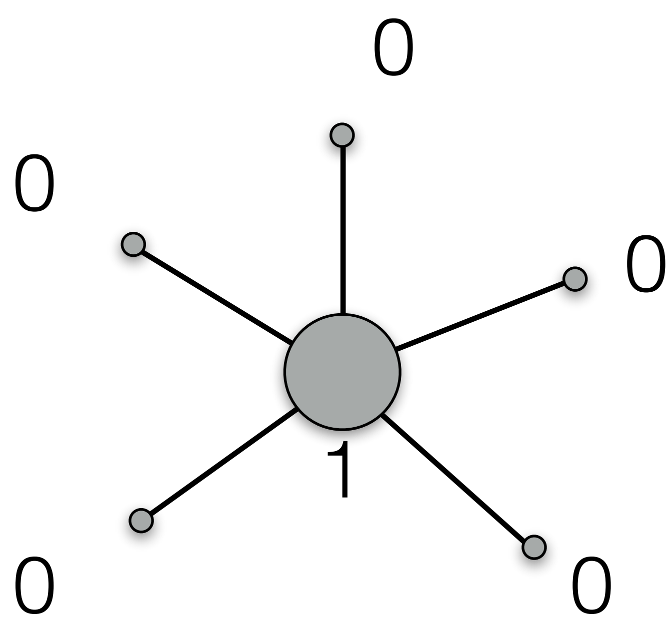


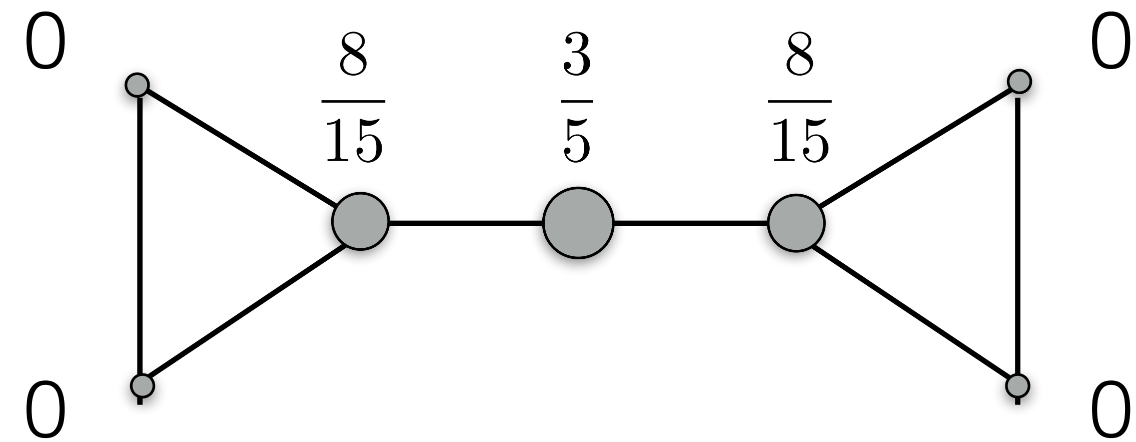


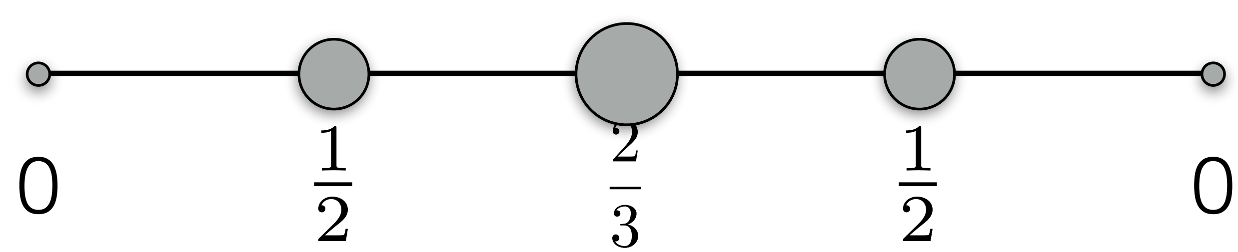


Betweenness Centrality

$$C_B(v_i) = \frac{\sum_{j<k} \frac{g_{jk}(v_i)}{g_{jk}}}{(N-1)(N-2)/2}$$

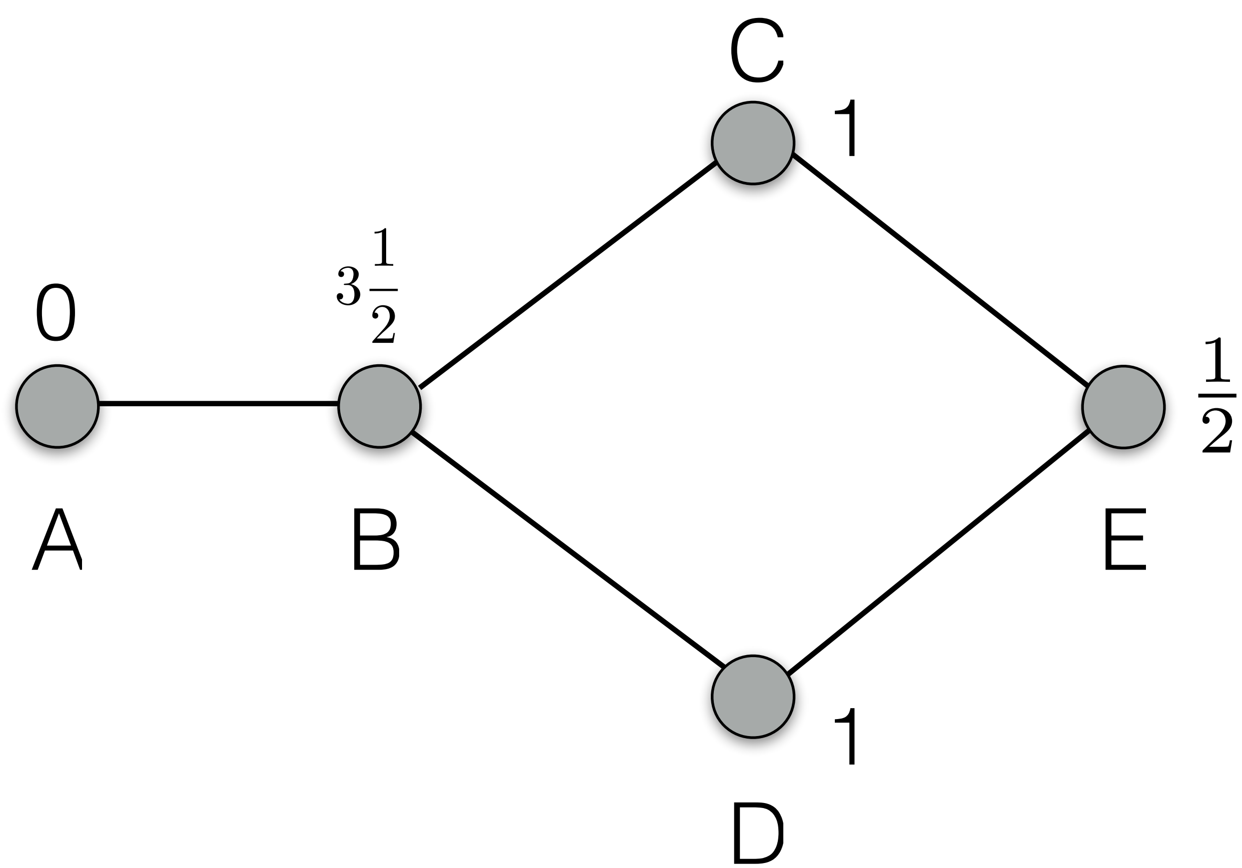






Betweenness Centrality

$$C_B(v_i) = \frac{\sum_{j<k} \frac{g_{jk}(v_i)}{g_{jk}}}{(N-1)(N-2)/2}$$

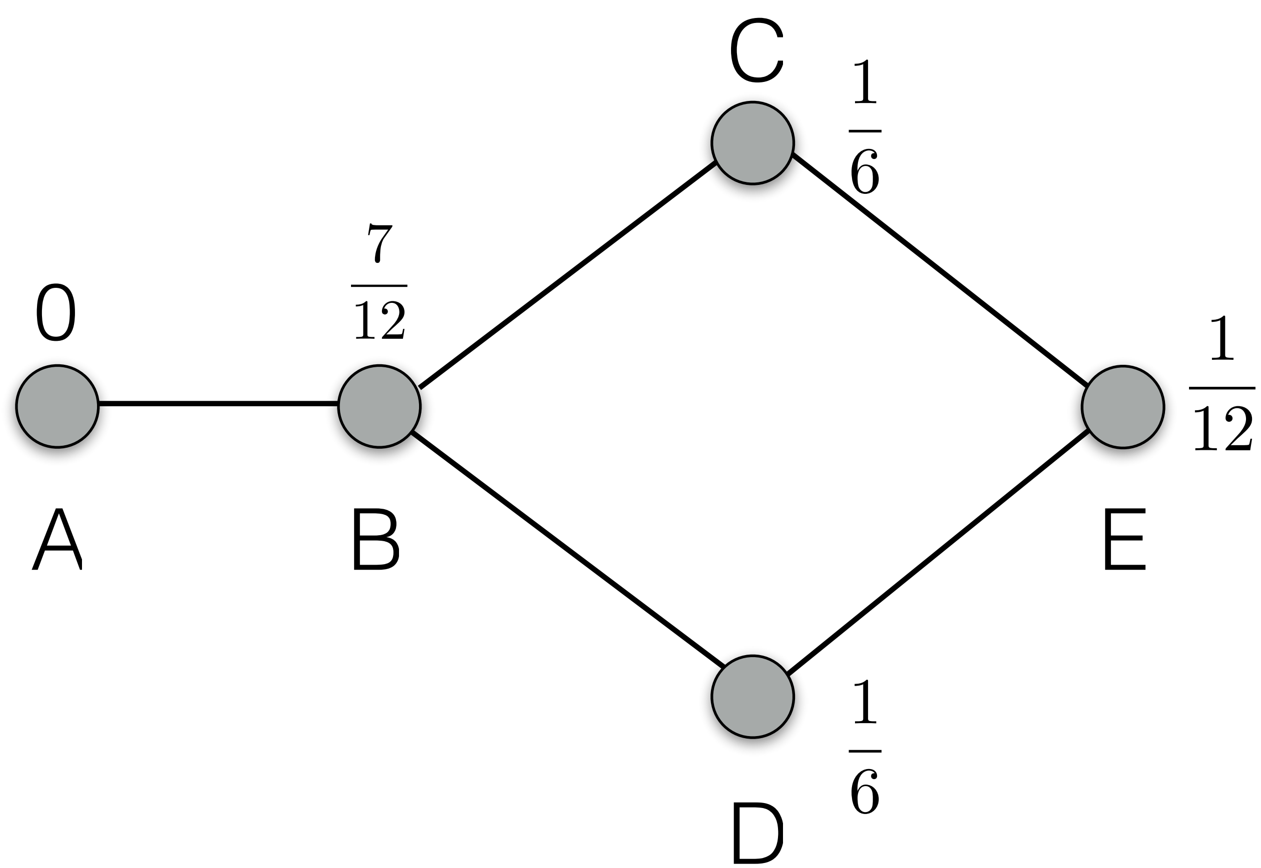


Start with unnormalized centrality

(remember to account for multiple paths)

Betweenness Centrality

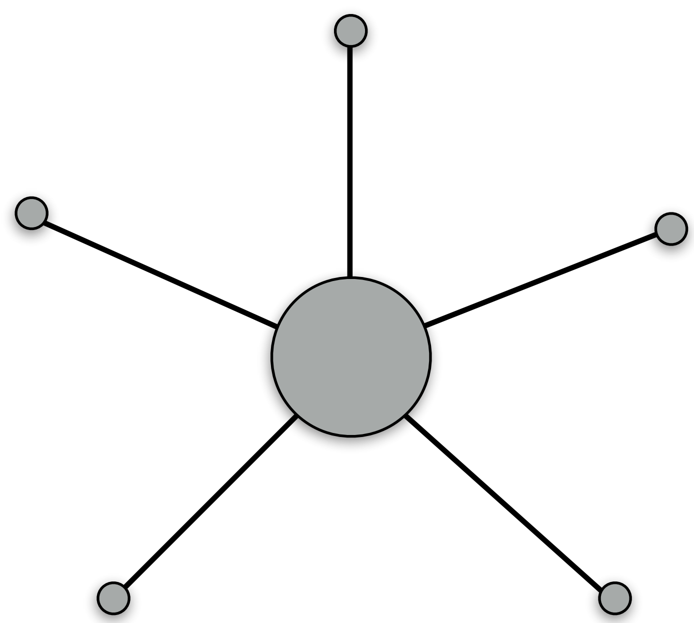
$$C_B(v_i) = \frac{\sum_{j<k} \frac{g_{jk}(v_i)}{g_{jk}}}{(N-1)(N-2)/2}$$



Then normalize

Betweenness Centrality

- Betweenness centrality make sense when you gain from bridging between different groups



- Brokering between groups
- Control of information
- Innovation
- Collaboration

Eigenvector Centrality

- Fourth notion: you are more important if you're connected to important people
- For example:
    - a small twitter account followed by someone with a large audience
    - a entrepreneur who knows Jack Dorsey
    - a senator's barber
- This is harder to calculate (I would not make you calculate it on an exam)

Eigenvector Centrality

Such a centrality measure must satisfy: $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$

the sum of the centralities of you neighbors

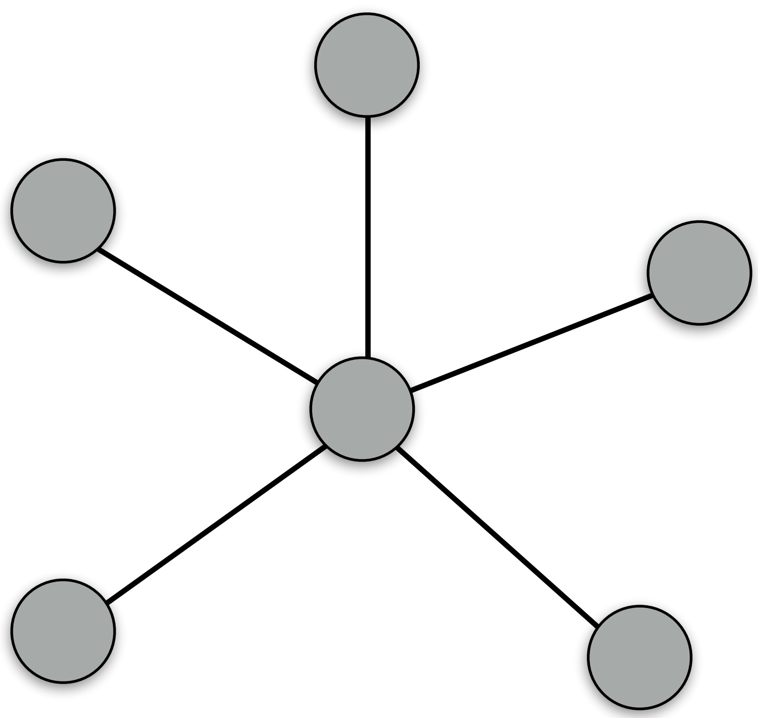
$$C_e\left(v_i\right) = \frac{1}{\boxed{\lambda_1}} \boxed{\sum_j A_{ij} C_E\left(v_j\right)}$$

leading eigenvalue of the matrix A

- A node's eigenvector centrality is proportional to the centrality of it's neighbors
- A node can have higher eigenvector centrality because:
    - They have more connections
    - They have more important connections

Network Centralization

- Centralization: a measure of how centrality is distributed in the network



→An attempt to quantify how centralized the network is as a whole

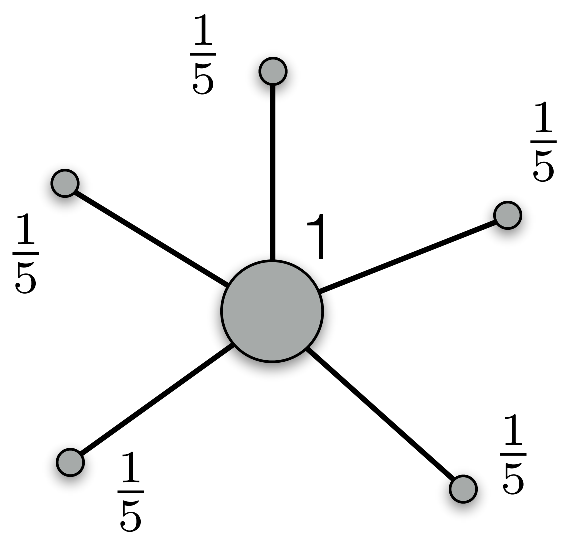
Difference between a node's centrality and the maximum centrality in the network

$$C_D(G) = \frac{\sum_{v_i \in G} \boxed{\left[C_D(v^*) - C_D(v_i)\right]}}{\boxed{(N-1)}}$$

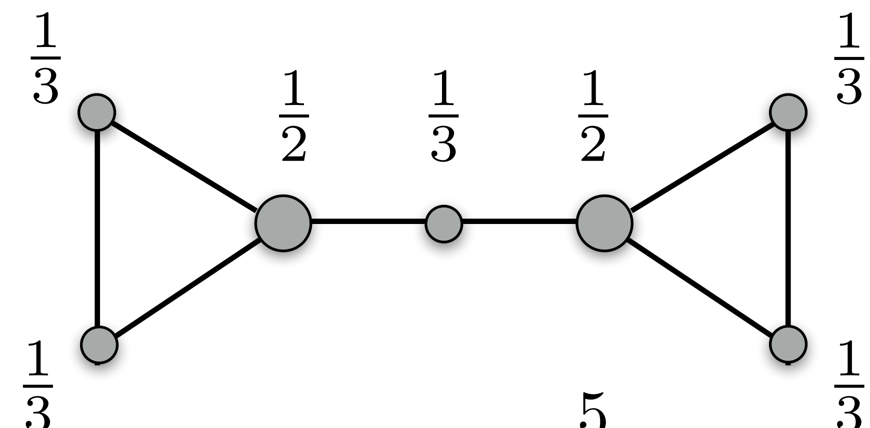
Normalization

Centralization

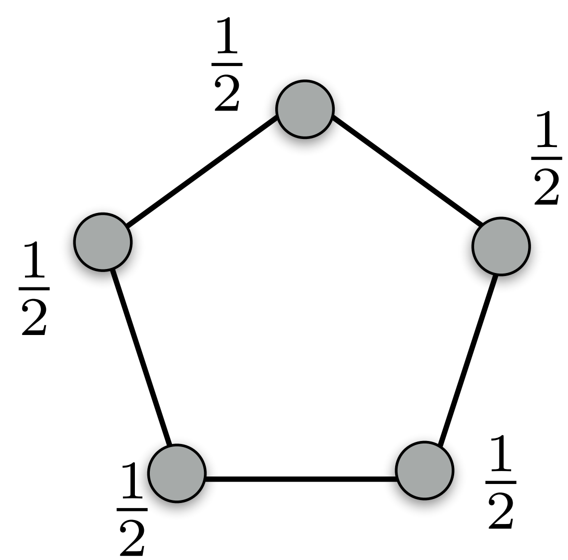
$$C_D(G) = \frac{\sum_{v_i \in G} \left[C_D(v^*) - C_D(v_i) \right]}{(N-1)}$$



$$C_D(G) = \frac{4}{5}$$



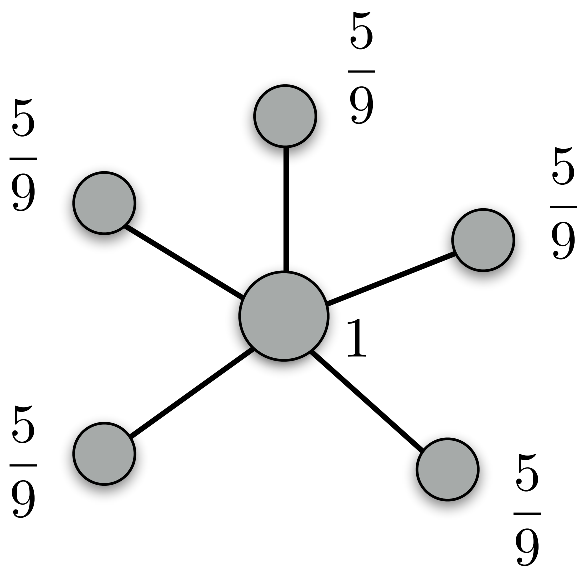
$$C_D(G) = \frac{5}{36}$$



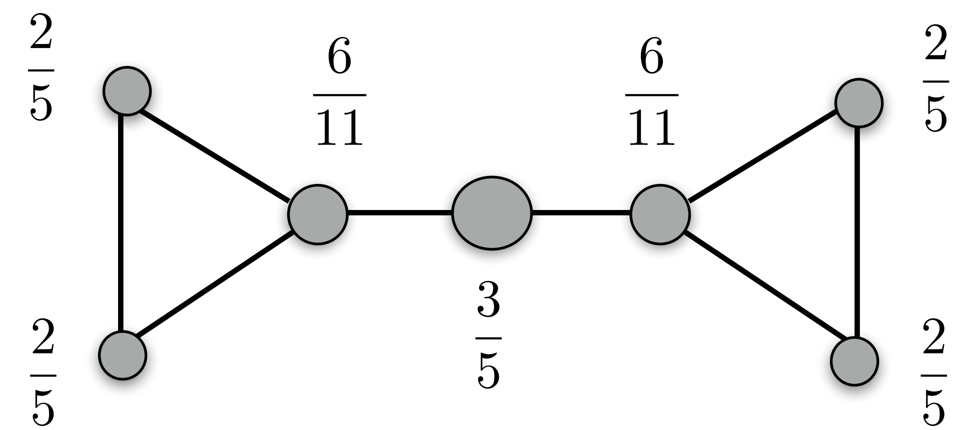
$$C_D(G) = 0$$

Centralization

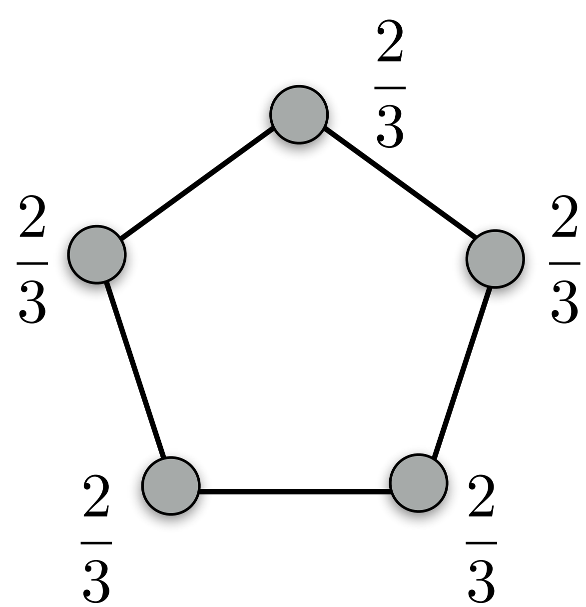
$$C_C(G) = \frac{\sum_{v_i \in G} \left[C_C(v^*) - C_C(v_i)\right]}{(N-1)}$$



$$C_C(G) = \frac{4}{9}$$



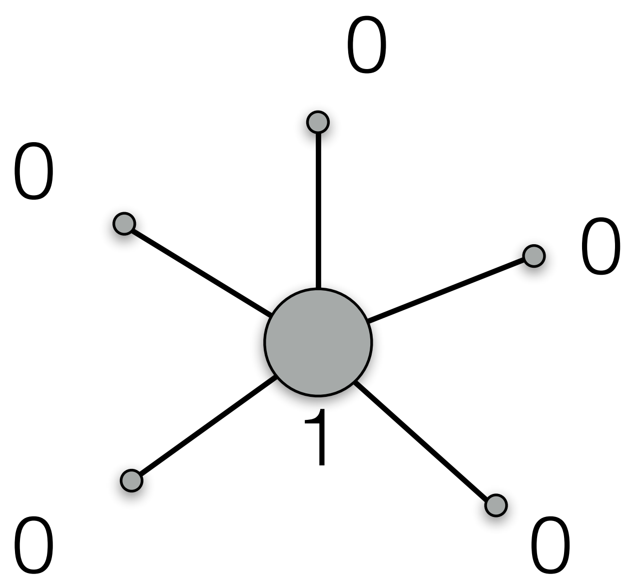
$$C_C(G) = \frac{10}{66}$$



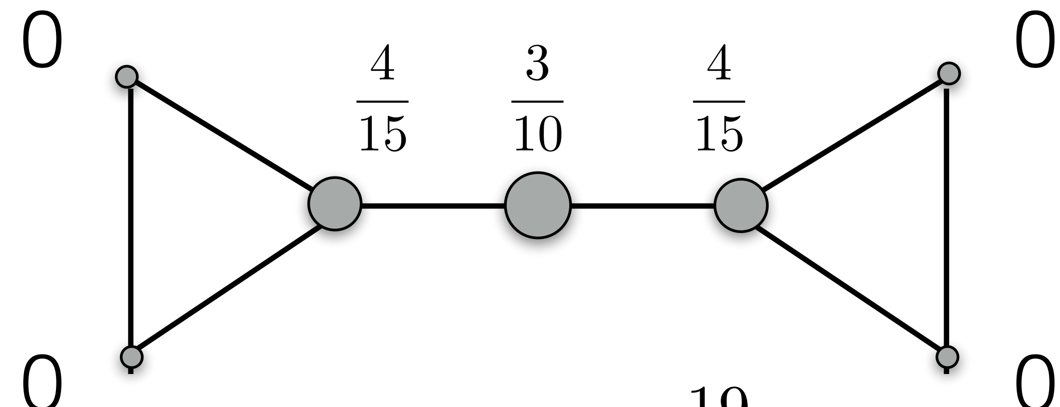
$$C_C(G) = 0$$

Centralization

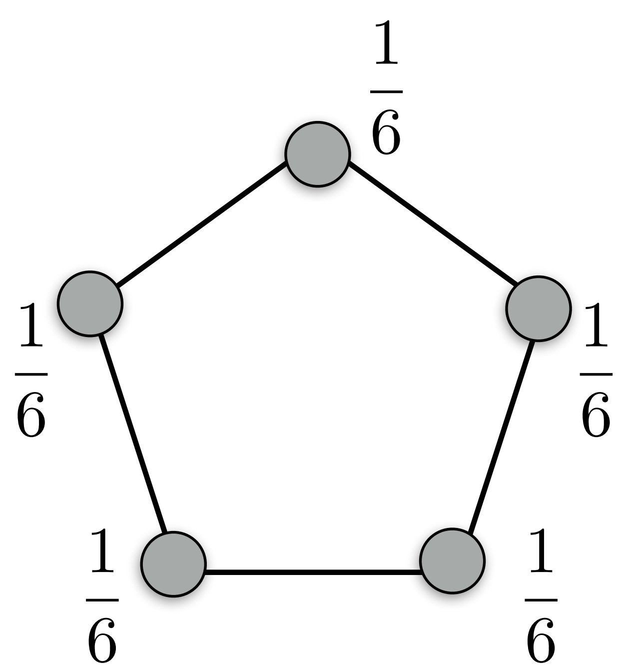
$$C_B(G) = \frac{\sum_{v_i \in G} \left[C_B(v^*) - C_B(v_i) \right]}{(N-1)}$$



$C_B(G) = 1$



$C_B(G) = \frac{19}{90}$

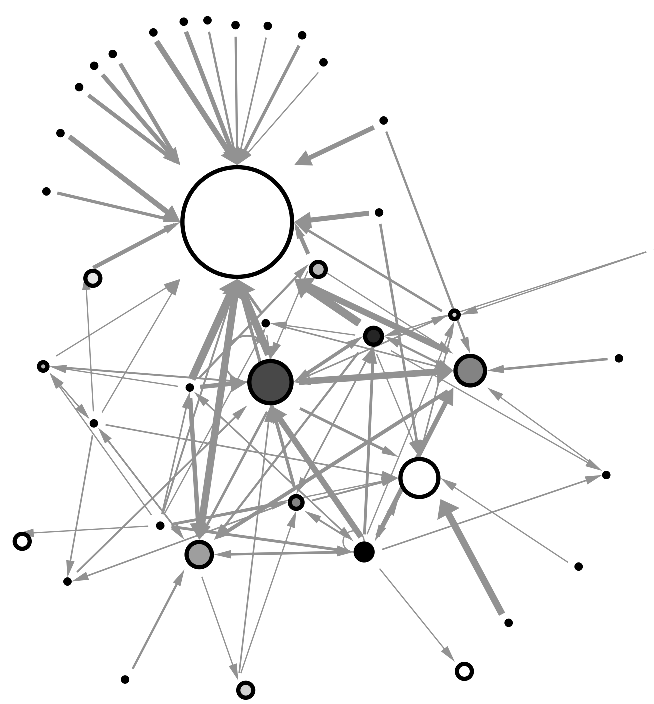


$C_B(G) = 0$

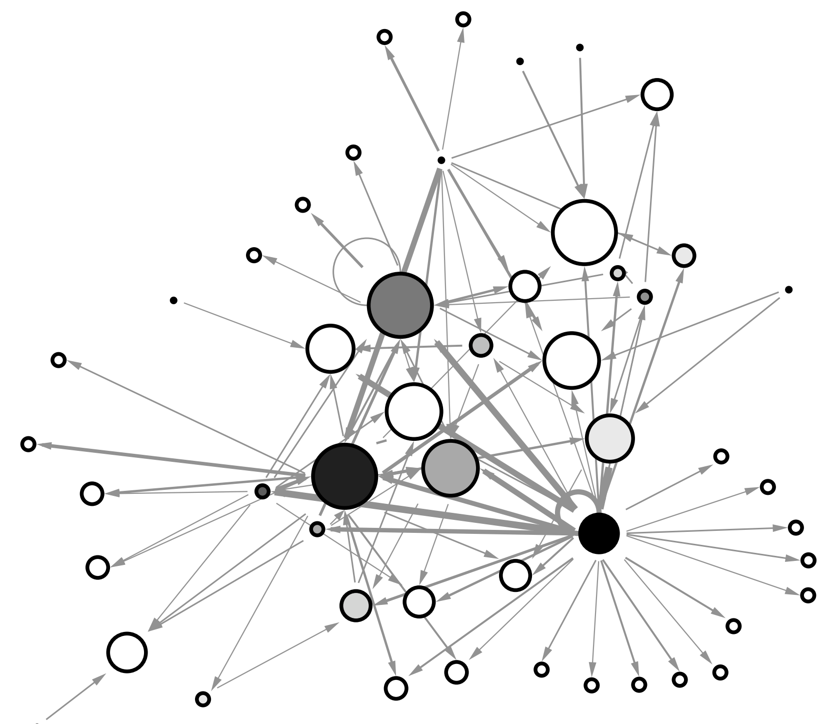
Centralization

Centralization tells us about how influence is spread across the network

Example: Financial Trading Networks



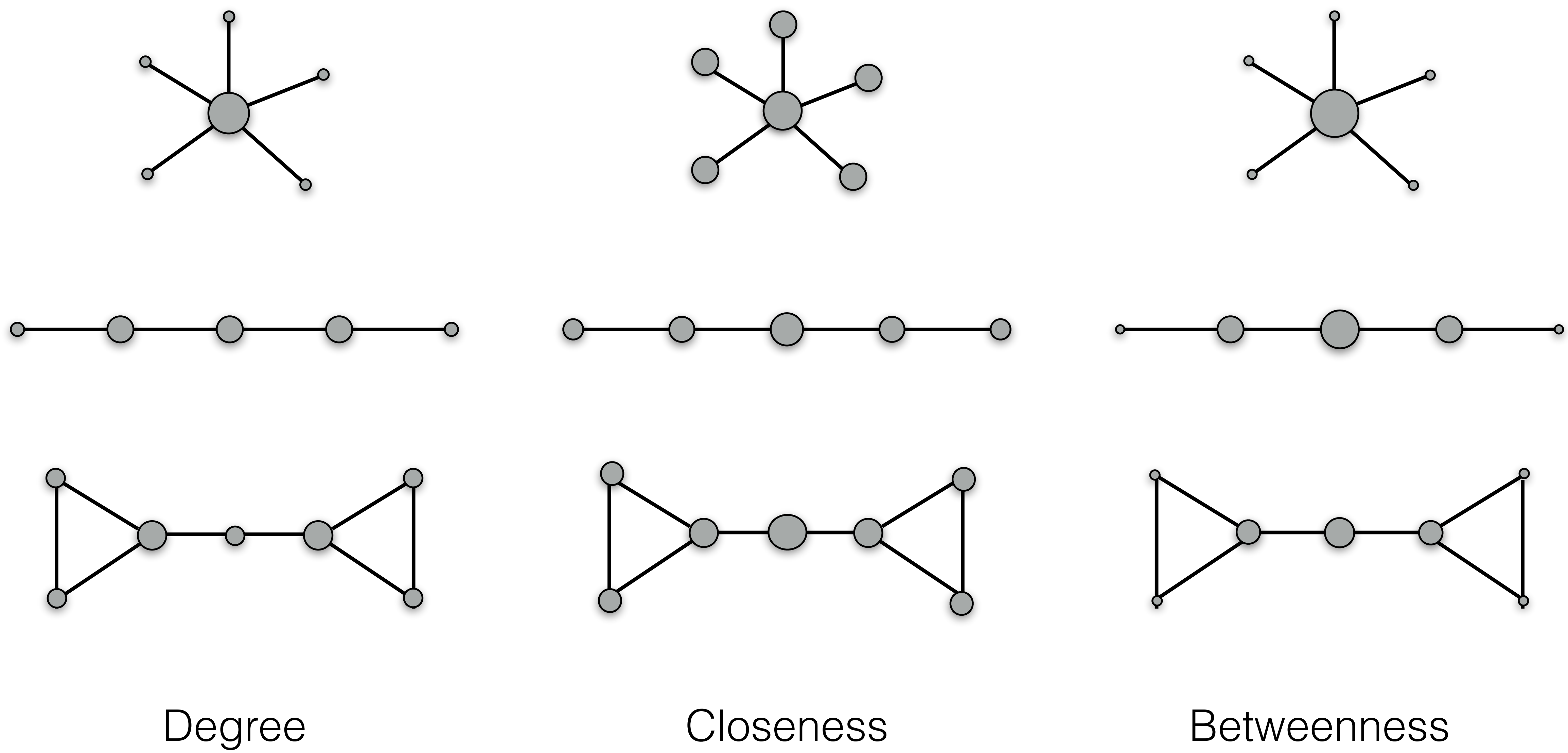
High centralization:
one node dominates
the network



Low centralization:
trades are more evenly
distributed

example & graphics: Adamic lecture

Comparing Centrality Measures



The three are clearly related, but they each get at something slightly different

Directed Networks: “Prestige”

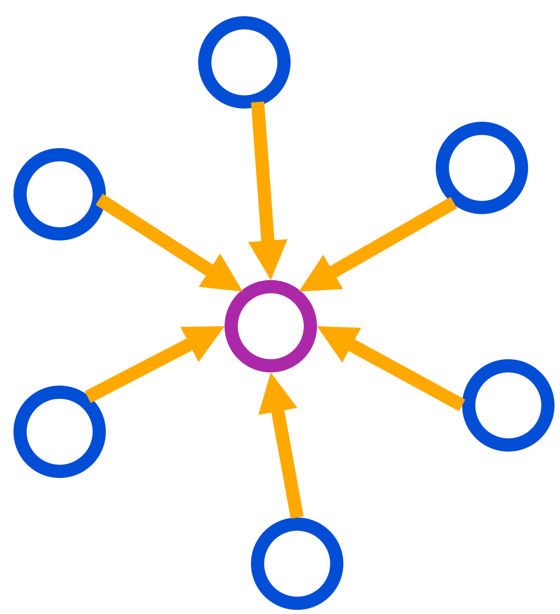
- Centrality in directed networks is called “prestige”
- This is sometimes a fine name:
    - admiration or trust
    - influence
    - friendship
    - trade
- But depending on the type of link, it might be misleading:
    - money lending
    - giving advice
    - hatred or distrust

→Lesson: Context matters! Always consider the interpretation of a measure *in a particular context*

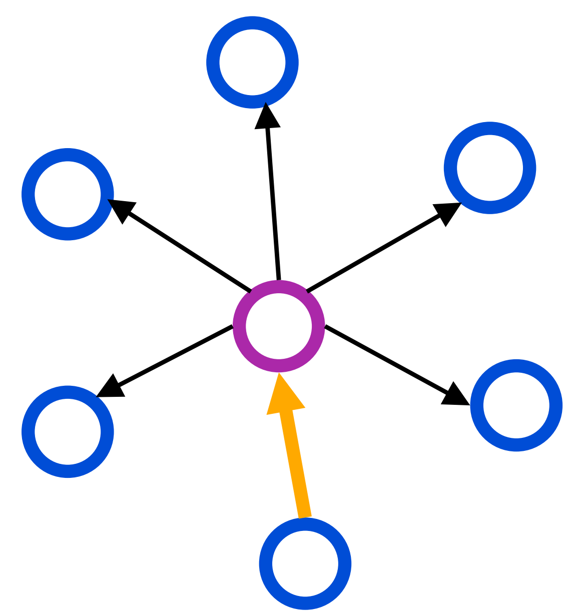
Directed Networks: "Prestige"

- In-degree
  - A website that is linked to often has high prestige
  - A person who is frequently nominated for a reward has high prestige

$$C_D(v_i) = \frac{d_{in}(v_i)}{(N-1)}$$



High



Low

example & graphics: Adamic lecture

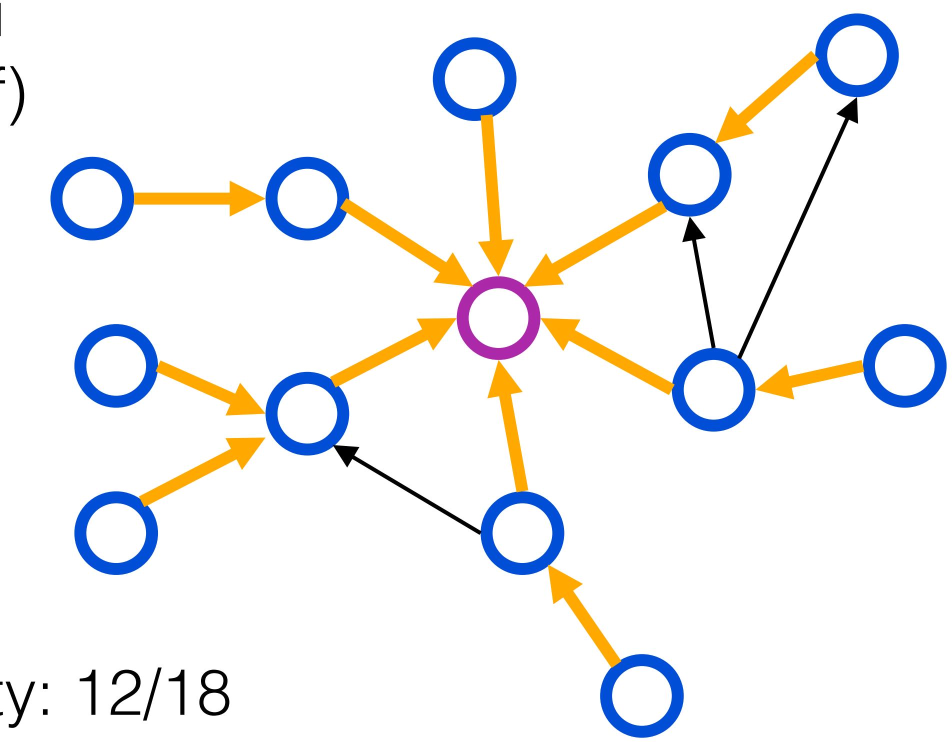
Directed Networks: "Prestige"

- Closeness Analogue: Proximity
  - Uses shortest *directed* path length: directed geodesic
  - Considers only nodes that can reach the selected node

Number of nodes that can reach i (including i itself)

$$C_C(v_i) = \frac{N_i - 1}{\sum_{v_j \in G_i} d(v_i, v_j)}$$

Nodes that can reach i



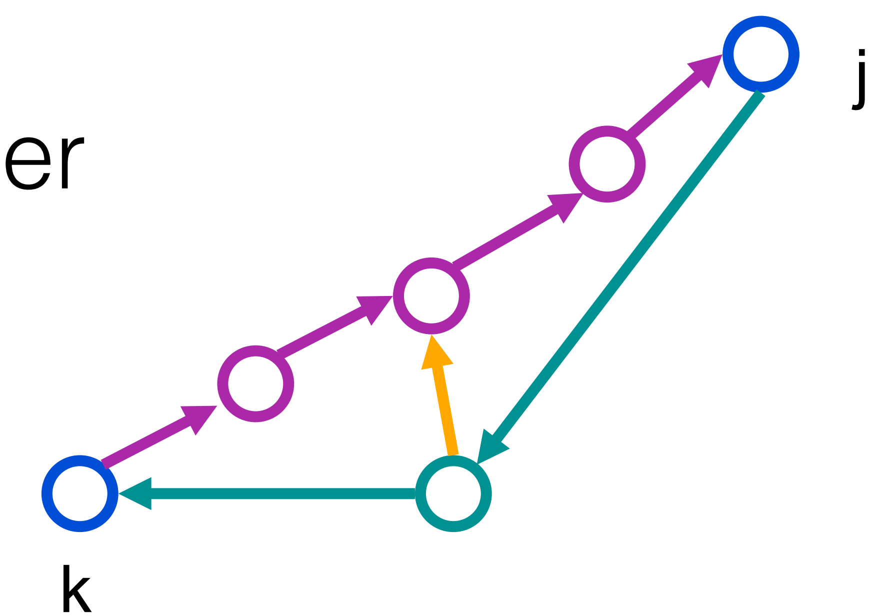
Directed closeness centrality: 12/18

example & graphics: Adamic lecture

Directed Networks: “Prestige”

A note on directed geodesics:

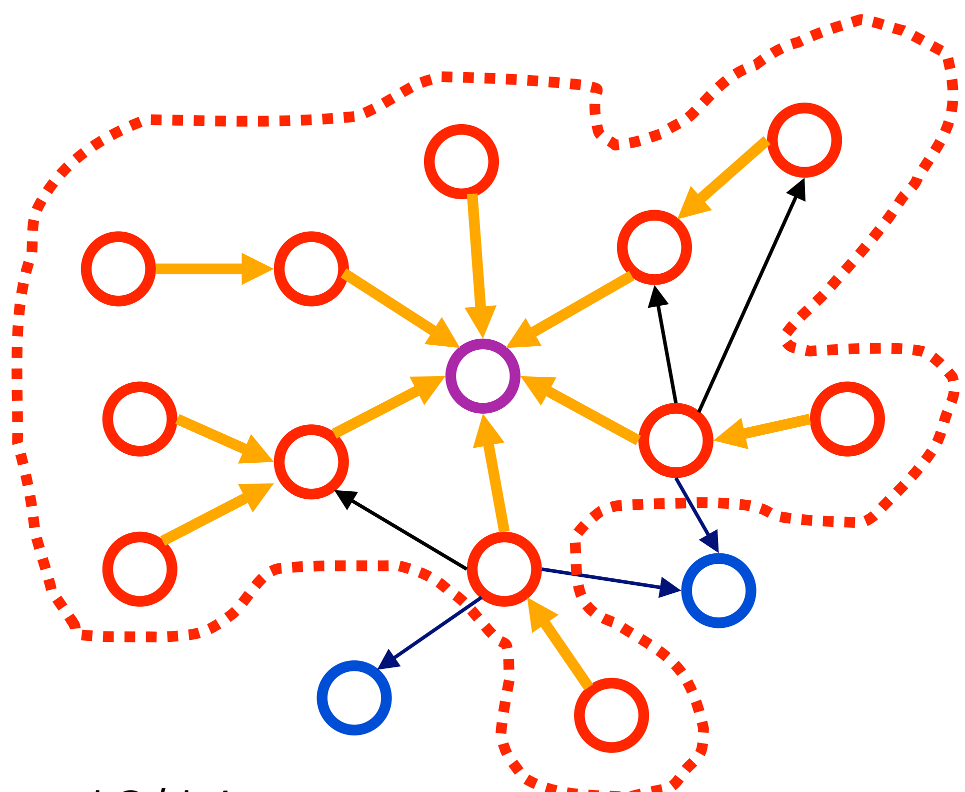
- You need to follow the arrows when tracing a path through the network
- The shortest directed path may not be the geodesic on the related undirected network
- The directed geodesic from j to k may be shorter than the directed geodesic from k to j



example & graphics: Adamic lecture

Directed Networks: “Prestige”

- Influence range
    - The influence range is what fraction of the nodes in the network can reach you via directed paths



Influence Range: 12/14

example & graphics: Adamic lecture

Directed Networks: “Prestige”

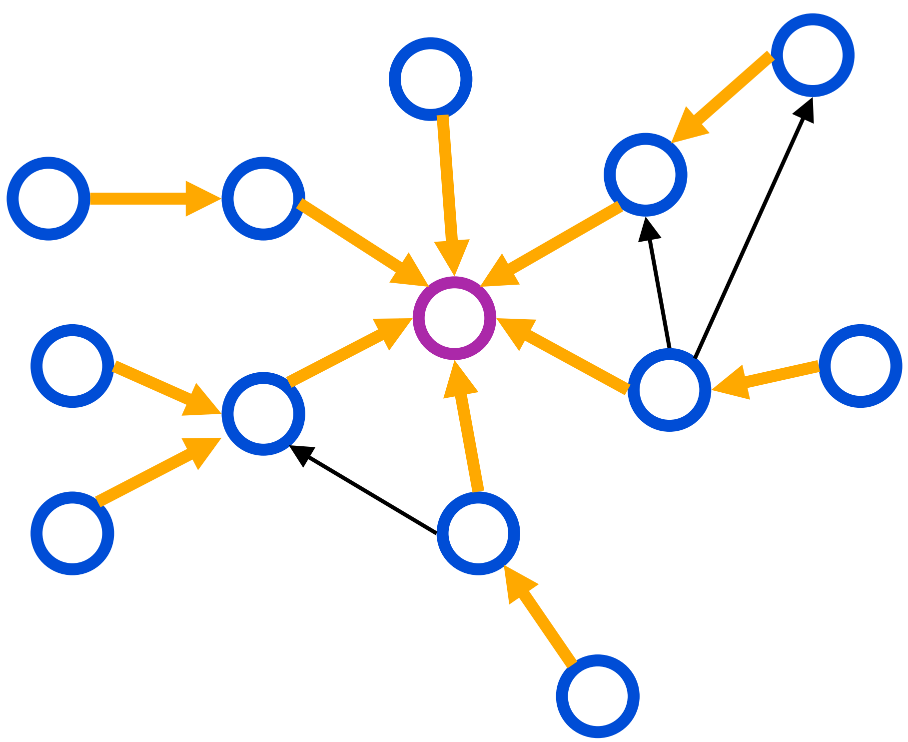
- Directed Betweenness: Almost exactly the same as betweenness, but with directed geodesics and normalized in a directed way

Number of directed geodesics between j and k containing i

$$C_B\left(v_i\right) = \frac{1}{\left(N-1\right)\left(N-2\right)} \sum_{j,k} \frac{g_{jk}\left(v_i\right)}{g_{jk}}$$

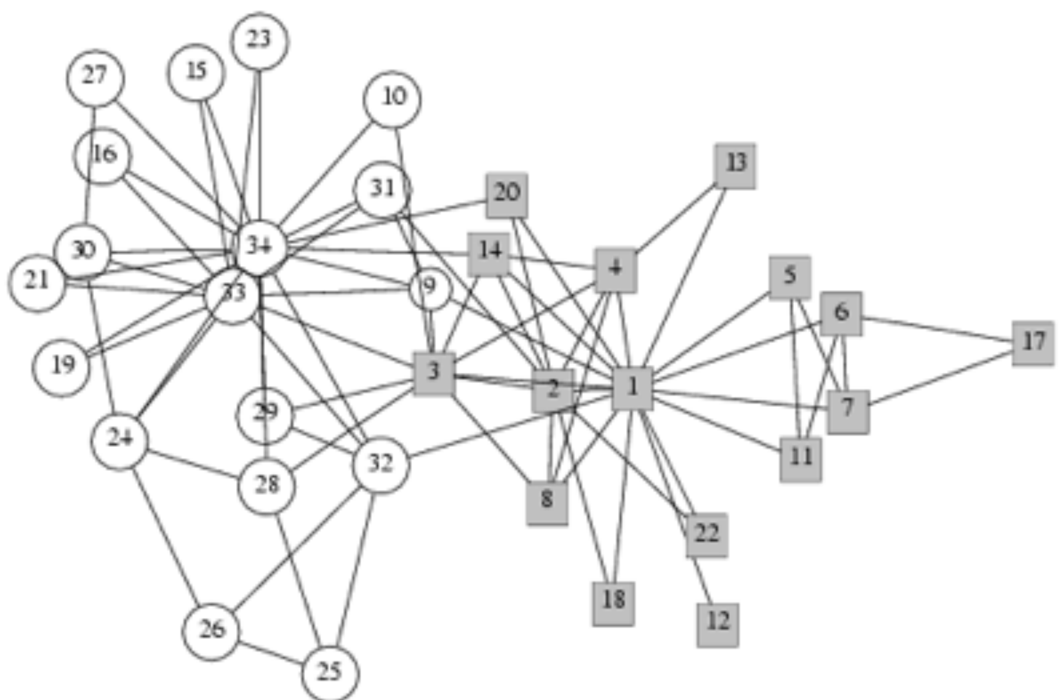
Note: *both* directions

Total number of directed geodesics between j and k



Directed Betweenness Centrality: 0

Summing up…



graphic: Newman Karate Club

There are lots of ways for a node to be “central” to a network

- Degree
- Closeness
- Betweenness
- etc!

- Different types of centrality are relevant in different contexts.
- Which is most interesting is a judgment call!