

# LECTURE 6: CENTRALITY MEASURES

Prof. Pan Hui

CSIT 6000K: Social Networks and Social Computing: A Data Science Perspective

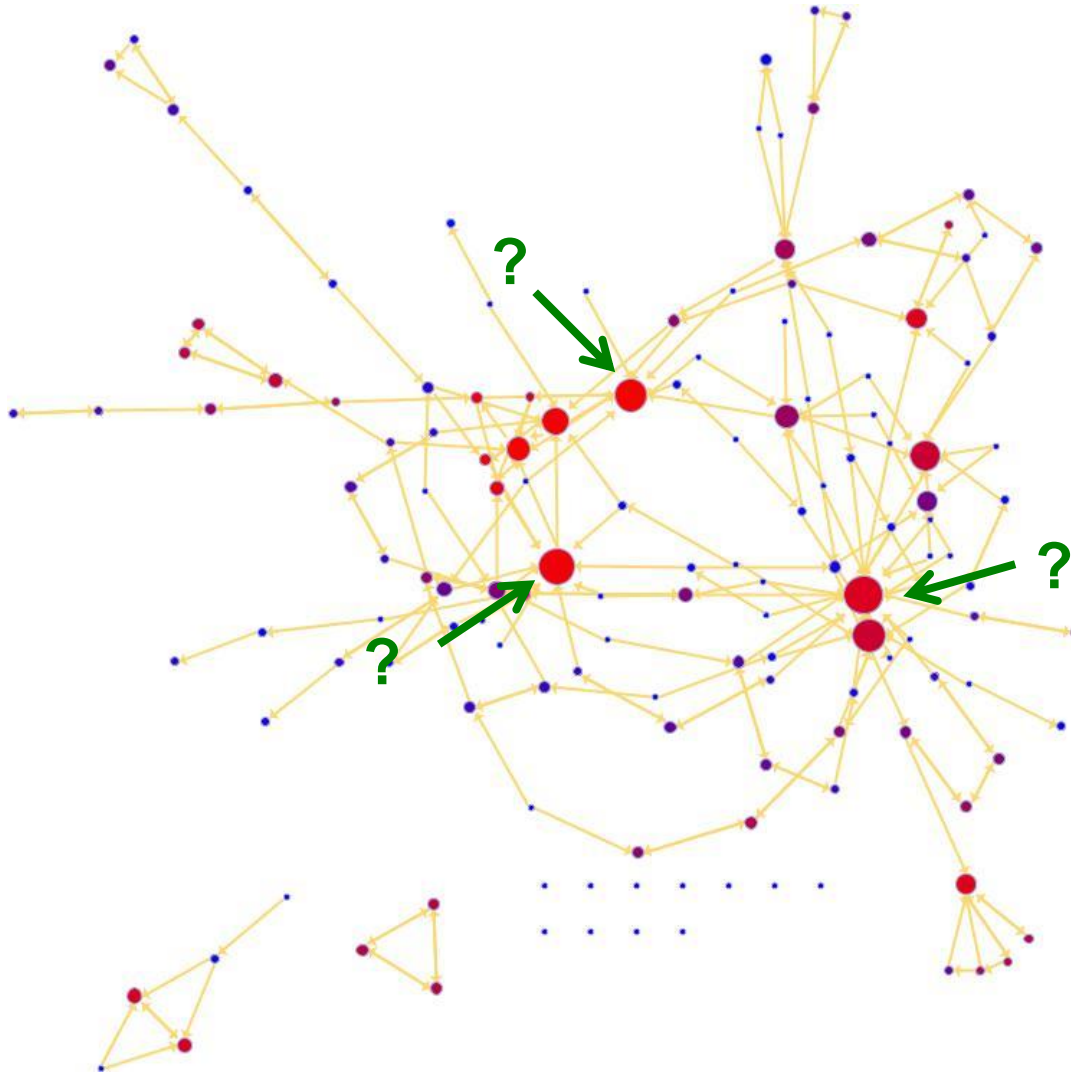
Thursdays 07:30 PM - 10:20 PM

# In This Lecture



- We will introduce the concept of centrality and the various measures which have been associated to this concept.

# Characterizing networks: Who is most central?



# Network Centrality

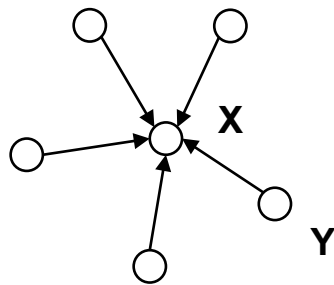
- Which nodes are most ‘central’ ?
- Definition of ‘central’ varies by context/purpose
- Local measure:
  - ▣ degree
- Relative to rest of network:
  - ▣ closeness, betweenness, eigenvector (Bonacich power centrality), Katz, PageRank, ...
- How evenly is centrality distributed among nodes?
  - ▣ Centralization, hubs and authorities, ...

# Centrality

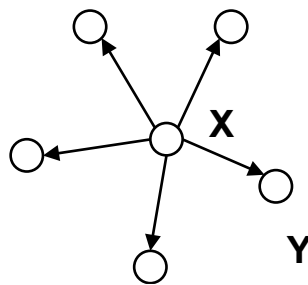
- Finding out which is the most central node is important:
  - ▣ It could help disseminating information in the network faster
  - ▣ It could help stopping epidemics
  - ▣ It could help protecting the network from breaking

# Centrality: visually

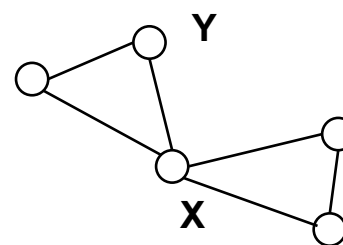
- Centrality can have various meanings:



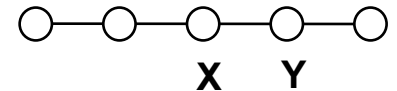
indegree



outdegree

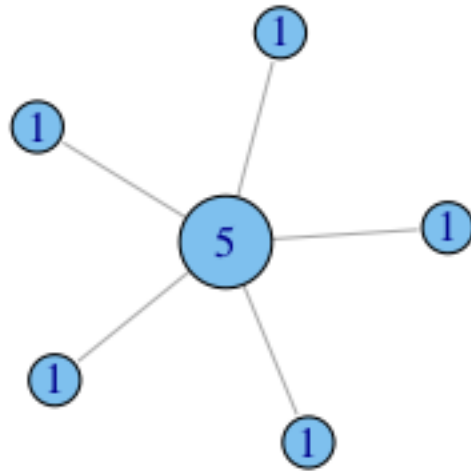


betweenness



closeness

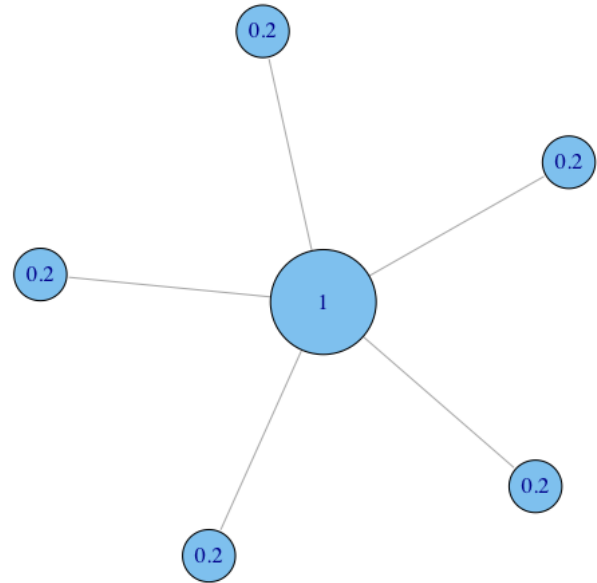
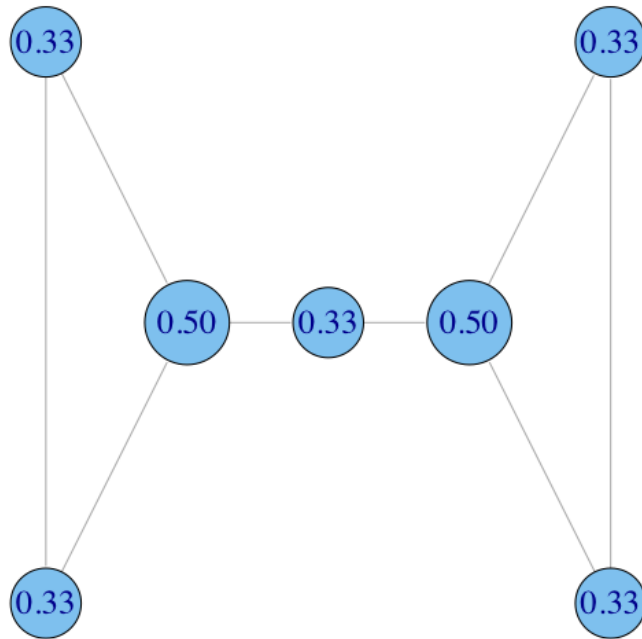
# Degree Centrality



- When is the number of connections the best centrality measure?
- people who will do favors to you
  - people you can talk to / have a beer with

# Normalization

- Divide for the max number of nodes ( $N-1$ )





# Freeman's Network Centrality

- How do we calculate the value of centrality of the network
  - ▣ To check how much variation there is among the nodes (heterogeneity?)

Max value of the above: when network is a star: 1 node has  $C=N-1$  and all others  $(N-1)$  have 1.

Max value of Degree  
Centrality in the Network

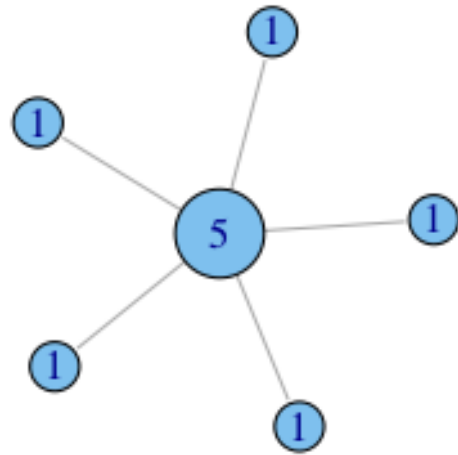
$$C_D = \frac{\sum_{i=1}^N [C_D(n^*) - C_D(i)]}{[(N-1)(N-2)]}$$

# Freeman Network Centrality Explained

- Explanation of the denominator:
- In the star topology one node has degree  $N-1$  and all other nodes have degree of 1:

$$0 + ((n-1) - 1) * (n-1) = (n-2) * (n-1)$$

# Freeman's Network Centrality

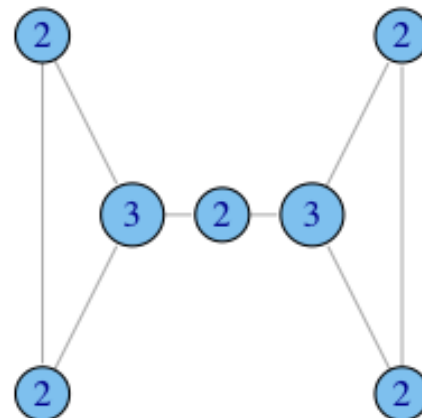


$$C_D = 1.0$$

$$\begin{aligned} &(4+4+4+4+4)/5*4 \\ &(1+0+0+0+1)/4*3=1/6 \\ &(1+1+0+1+0+1+1)/6*5=5/30 \end{aligned}$$

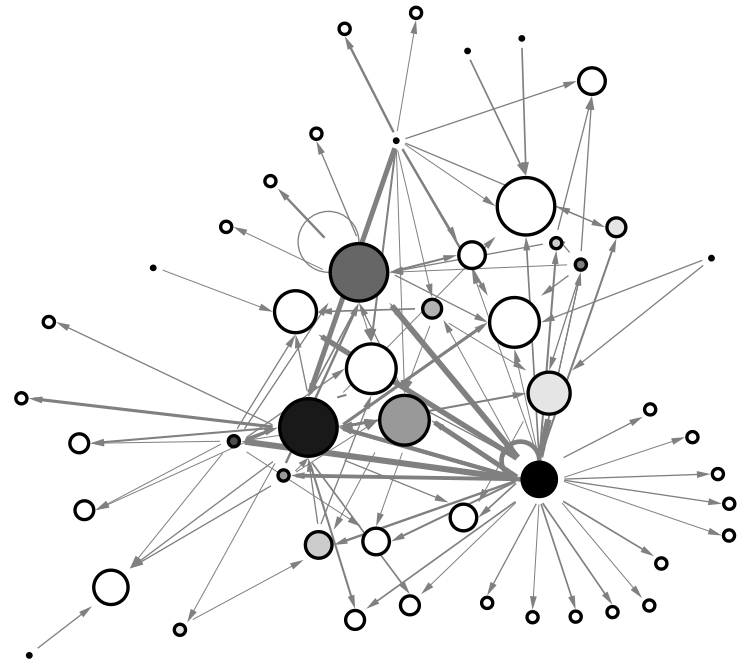
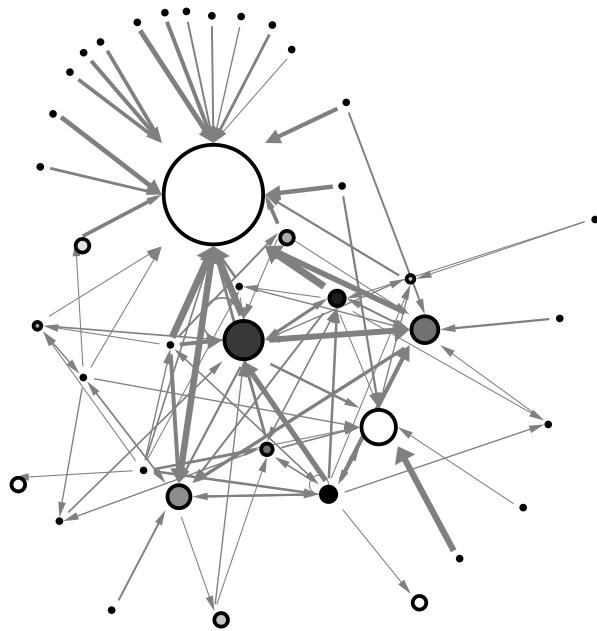


$$C_D = 0.167$$

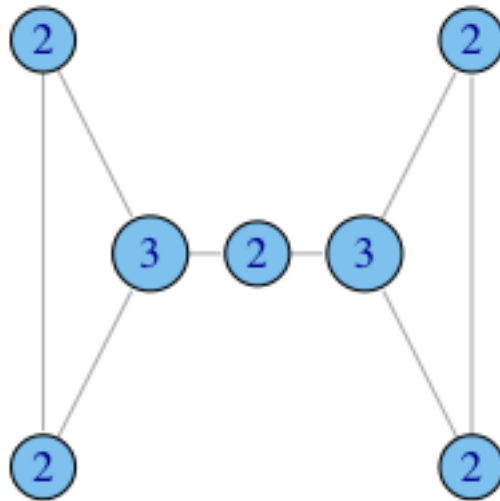


$$C_D = 0.167$$

# Examples: Financial Networks



# When is Degree Centrality not so good?



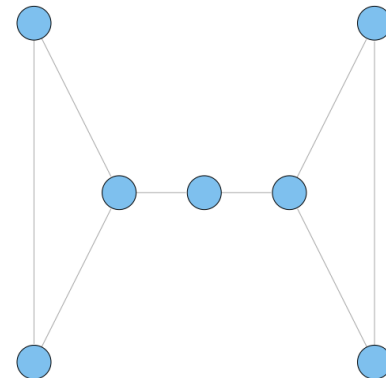
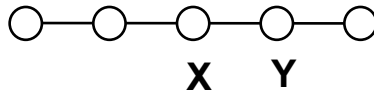
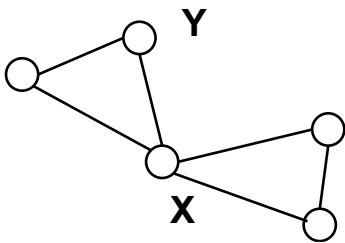
# When is Degree Centrality not so good ?



- 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 (Formally)

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

Where  $g_{jk}(i)$  = the number of shortest paths connecting  $jk$  passing through  $i$

$g_{jk}$  = total number of shortest paths

Usually normalized by:

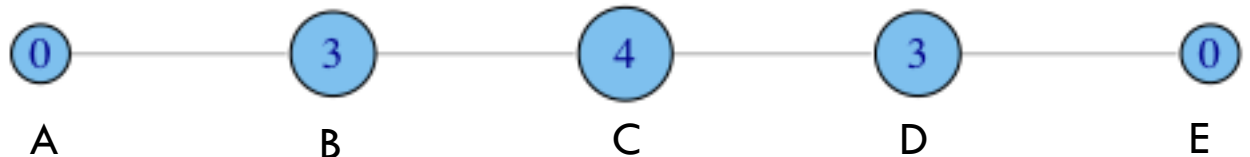
$$C'_B(i) = C_B(i) / [(n-1)(n-2) / 2]$$

number of pairs of vertices  
excluding the vertex itself



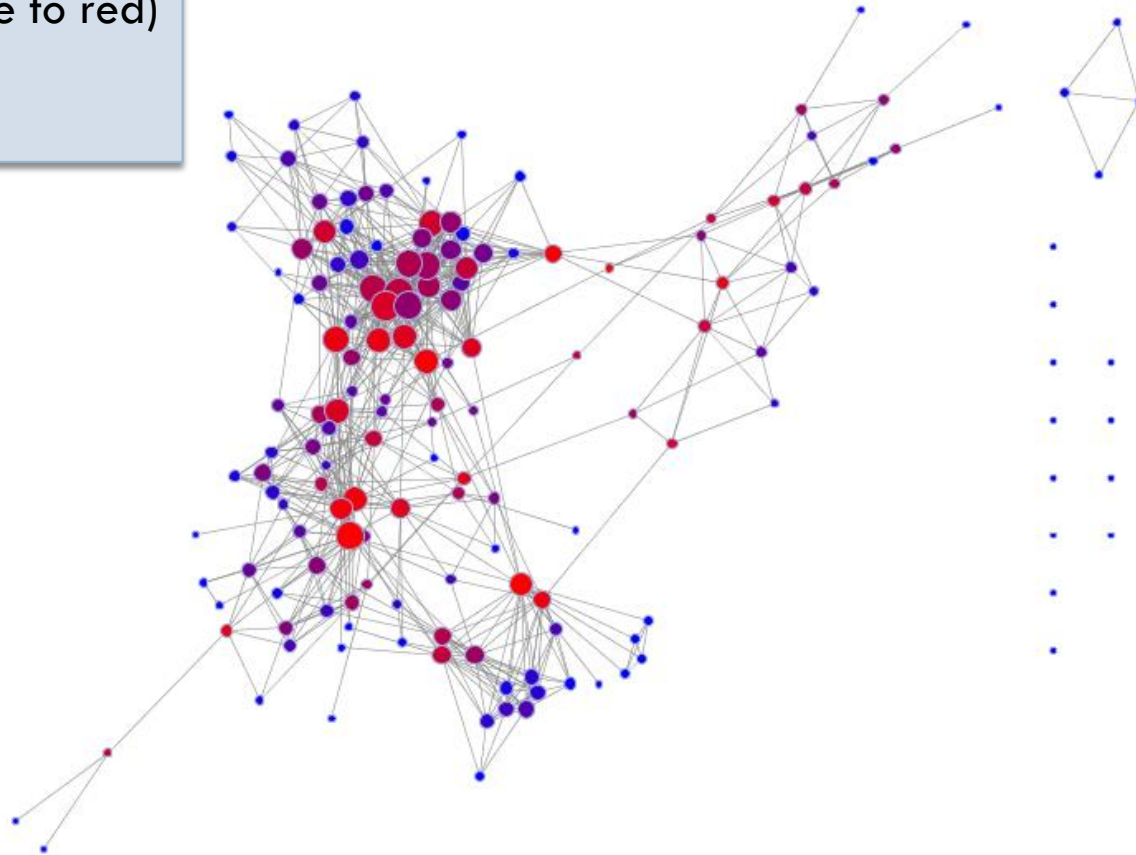
# Betweenness: Example

- 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 alternative paths for these pairs to take, so C gets full credit



# Facebook Example (Adamic)

Color(from blue to red)  
is betweenness  
Size is degree.



# 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

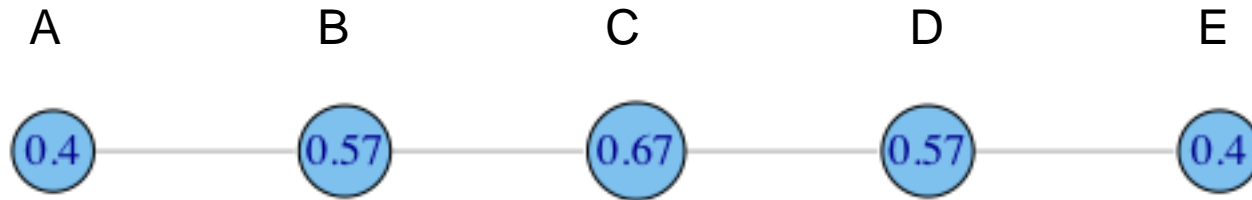
# Closeness Centrality (Formally)

- Closeness is based on the length of the average shortest path between a vertex and all vertices in the graph

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

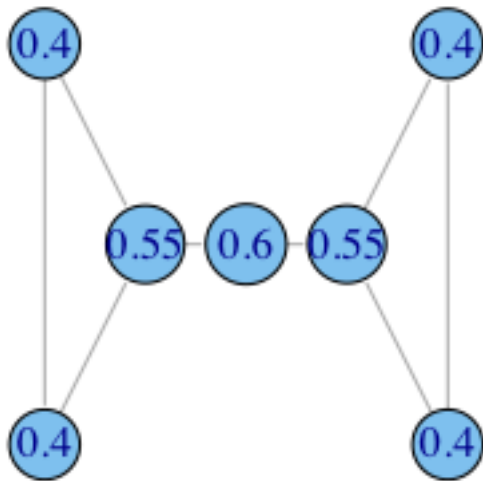
$$C'_c(i) = \left[ \sum_{j=1}^N d(i, j) / (N - 1) \right]^{-1}$$

# Closeness: Example

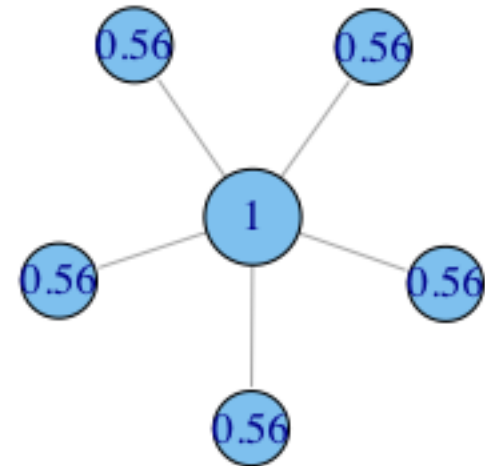


$$C'_c(A) = \left[ \frac{\sum_{j=1}^N d(A, j)}{N-1} \right]^{-1} = \left[ \frac{1+2+3+4}{4} \right]^{-1} = \left[ \frac{10}{4} \right]^{-1} = 0.4$$

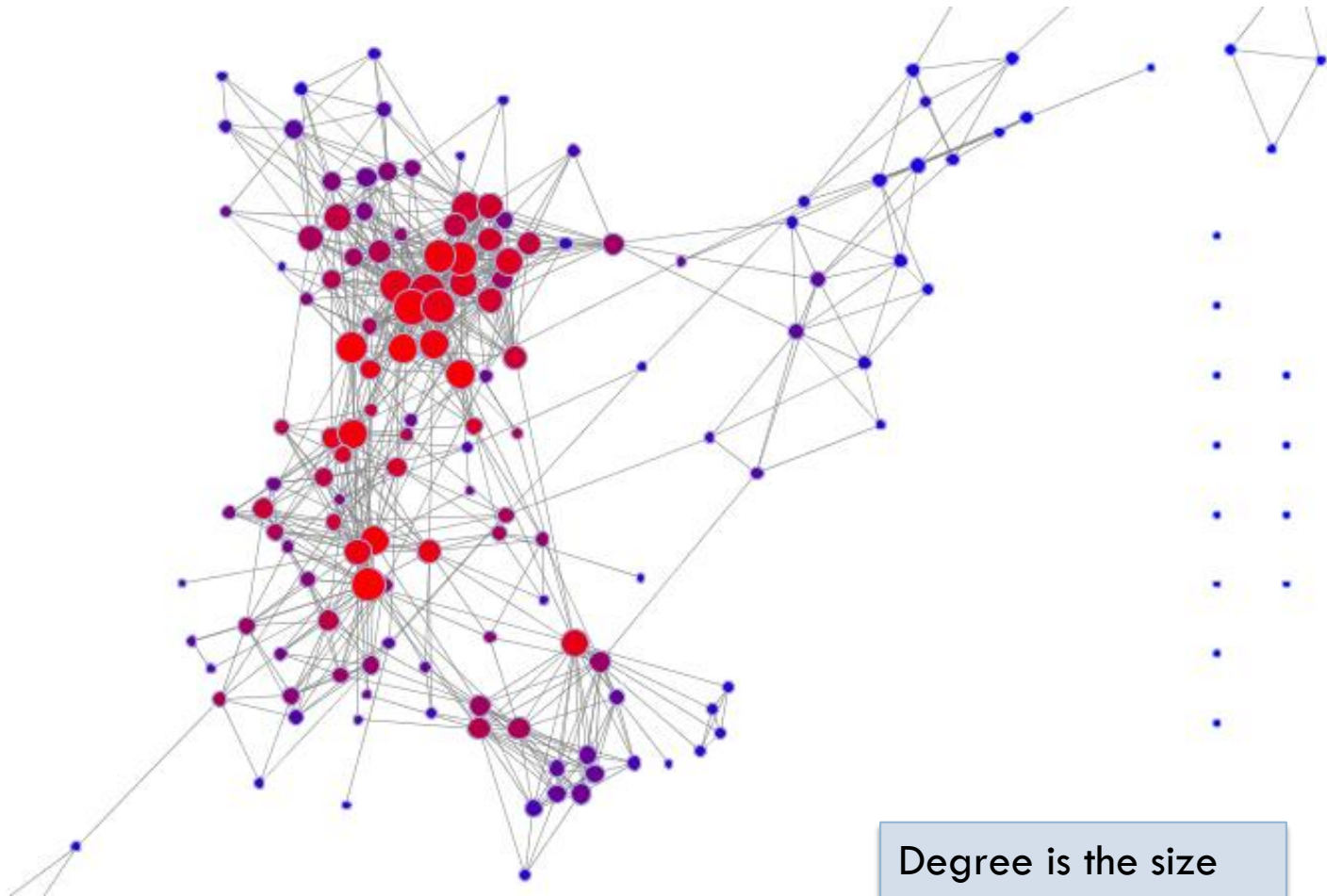
# Examples



$$((1+1+2+3+4+4)/6)^{-1} = 6/15 = 0.4$$



# Example: Facebook (Adamic)



Degree is the size  
Color is closeness

# Centrality Comparison

Comparing across these 3 centrality values

- Generally, the 3 centrality types will be positively correlated
- When they are not (low) correlated, it probably tells you something interesting about the network.

|                  | Low Degree   | Low Closeness  | Low Betweenness   |
|------------------|--|--|---|
| High Degree      |  | Embedded in cluster that is far from the rest of the network   | Ego's connections are redundant - communication bypasses him/her                        |
| High Closeness   | Key player tied to important important/active alters |  | Probably multiple paths in the network, ego is near many people, but so are many others |
| High Betweenness | Ego's few ties are crucial for network flow          | Very rare cell. Would mean that ego monopolizes the ties from a small number of people to many others. |   |



# Eigenvector Centrality

- Degree Centrality depends on having many connections: but what if these connections are pretty isolated?
- A central node should be one connected to powerful nodes

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$$

Neighbourhood of v

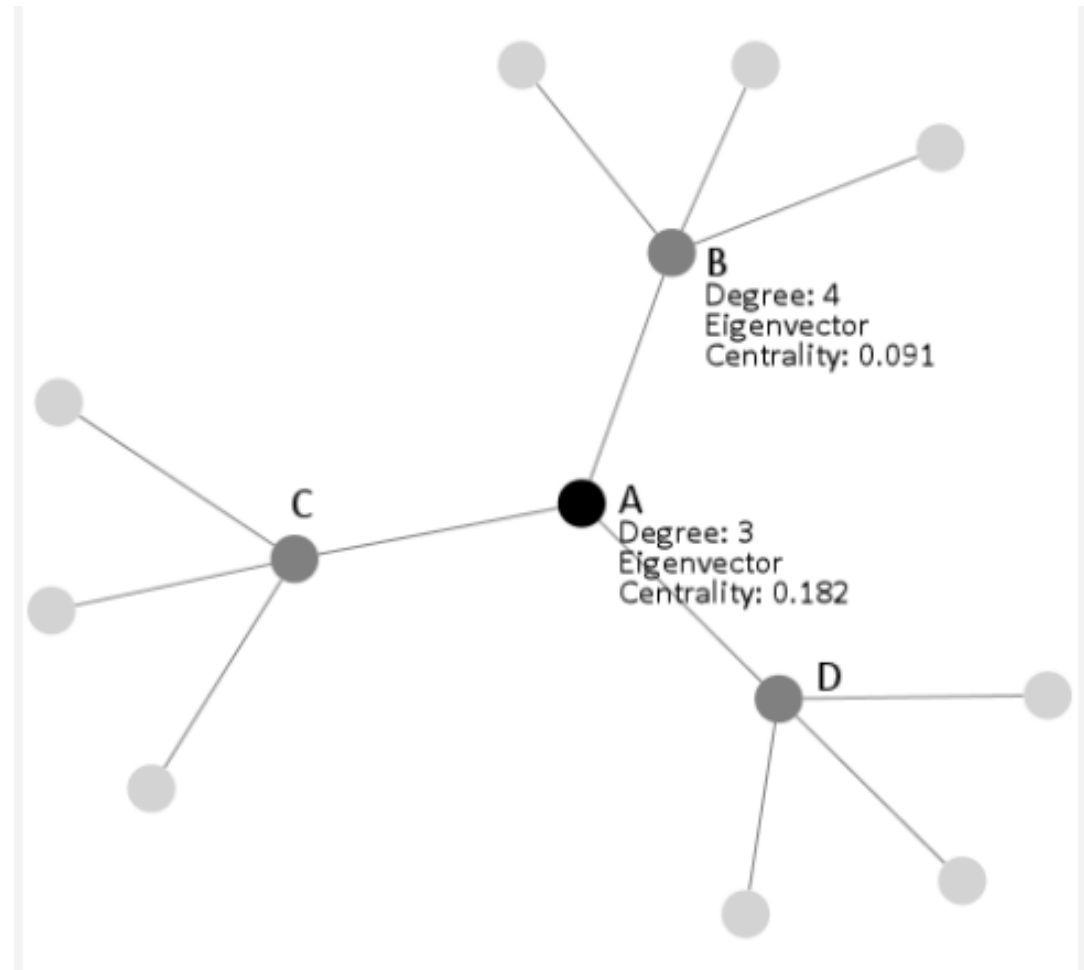
$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

Adjacency Matrix of the graph

# Eigenvector Centrality Algorithm

- 1. Start by assigning centrality score of 1 to all nodes ( $v_i = 1$  for all  $i$  in the network)
- 2. Recompute scores of each node as weighted sum of centralities of all nodes in a node's neighborhood:  $v_i = \sum_{j \in N} a_{ij} * v_j$
- 3. Normalize  $v$  by dividing each value by the largest value
- 4. Repeat steps 2 and 3 until values of  $v$  stop changing.

# Example



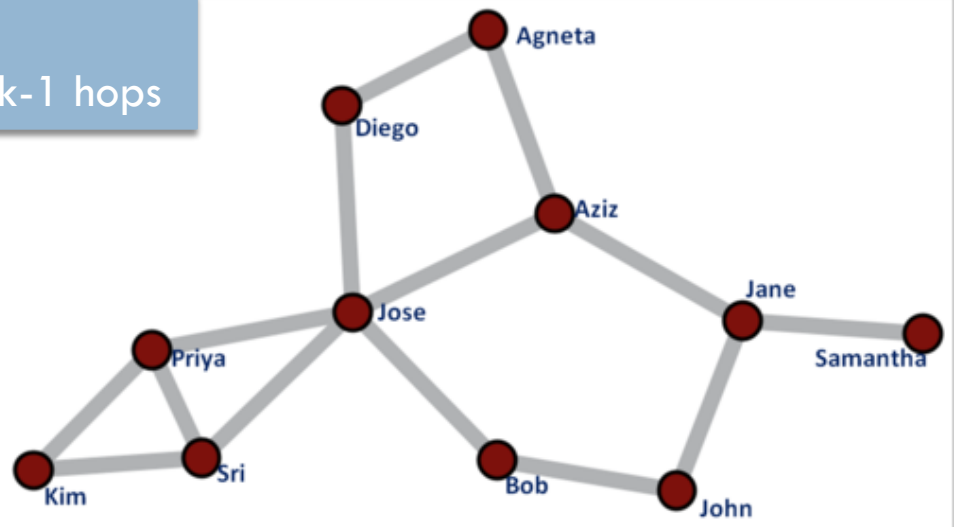
# Katz Centrality

- Closeness counts the number of shortest paths, but one could count the **number of paths**.

$$C_{\text{Katz}}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^n \alpha^k (A^k)_{ji}$$

Alpha is an attenuation factor

$A^k(ij)$  indicates if  $ij$  are connected by  $k-1$  hops



# Prestige in Directed Social Networks

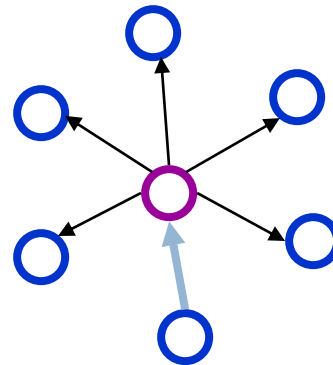
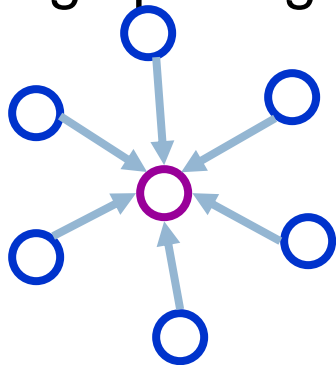
- When 'prestige' may be the right word
  - ▣ admiration
  - ▣ influence
  - ▣ gift-giving
  - ▣ trust
- Directionality especially important in instances where ties may not be reciprocated (e.g. dining partners choice network)
- When 'prestige' may not be the right word
  - ▣ gives advice to (can reverse direction)
  - ▣ gives orders to (- " -)
  - ▣ lends money to (- " -)
  - ▣ dislikes
  - ▣ distrusts

# Extensions of Undirected Degree Centrality - Prestige

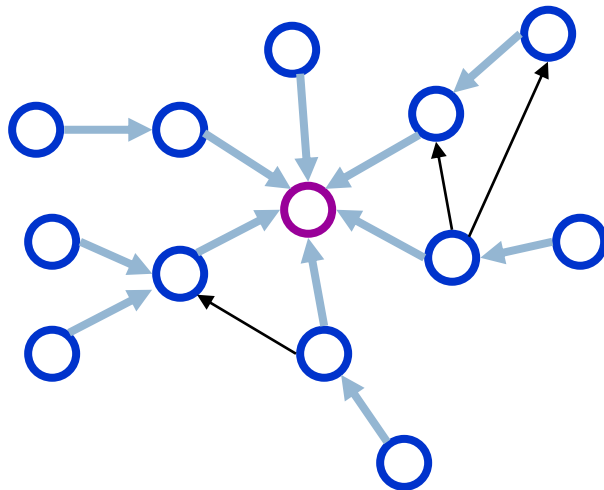
- Degree centrality

- ▣ 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

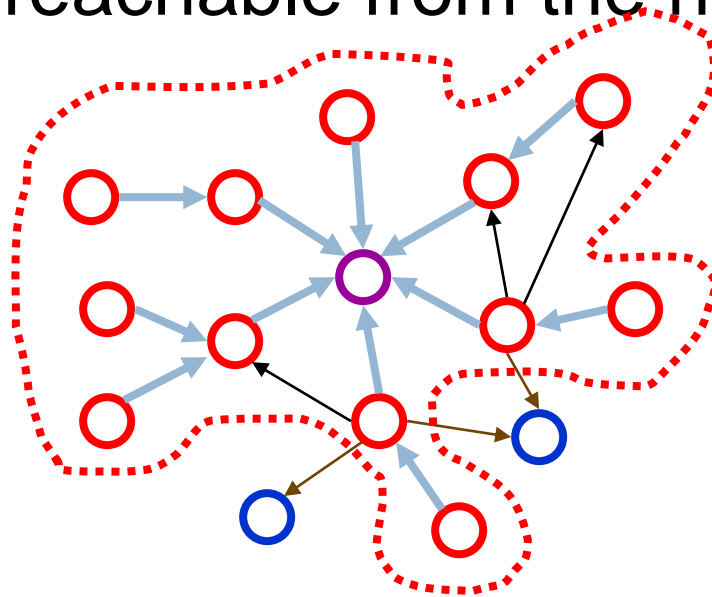


- Closeness centrality usually implies
  - ▣ all paths should lead to you
  - ▣ paths should lead from you to everywhere else
- Usually consider only vertices from which the node  $i$  in question can be reached



# Influence Range

- The influence range of  $i$  is the set of vertices who are reachable from the node  $i$





# Extending Betweenness Centrality to 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  $i$

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

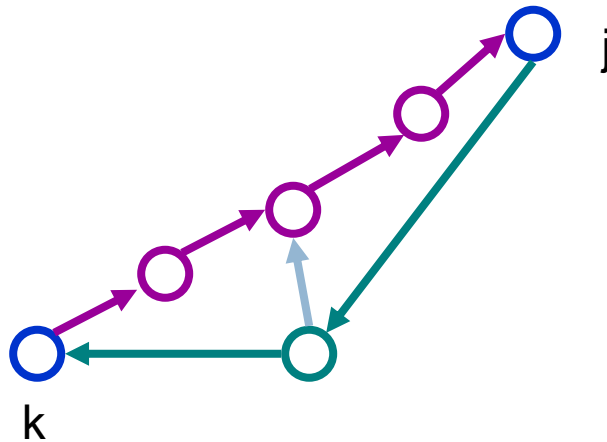
all paths between  $j$  and  $k$

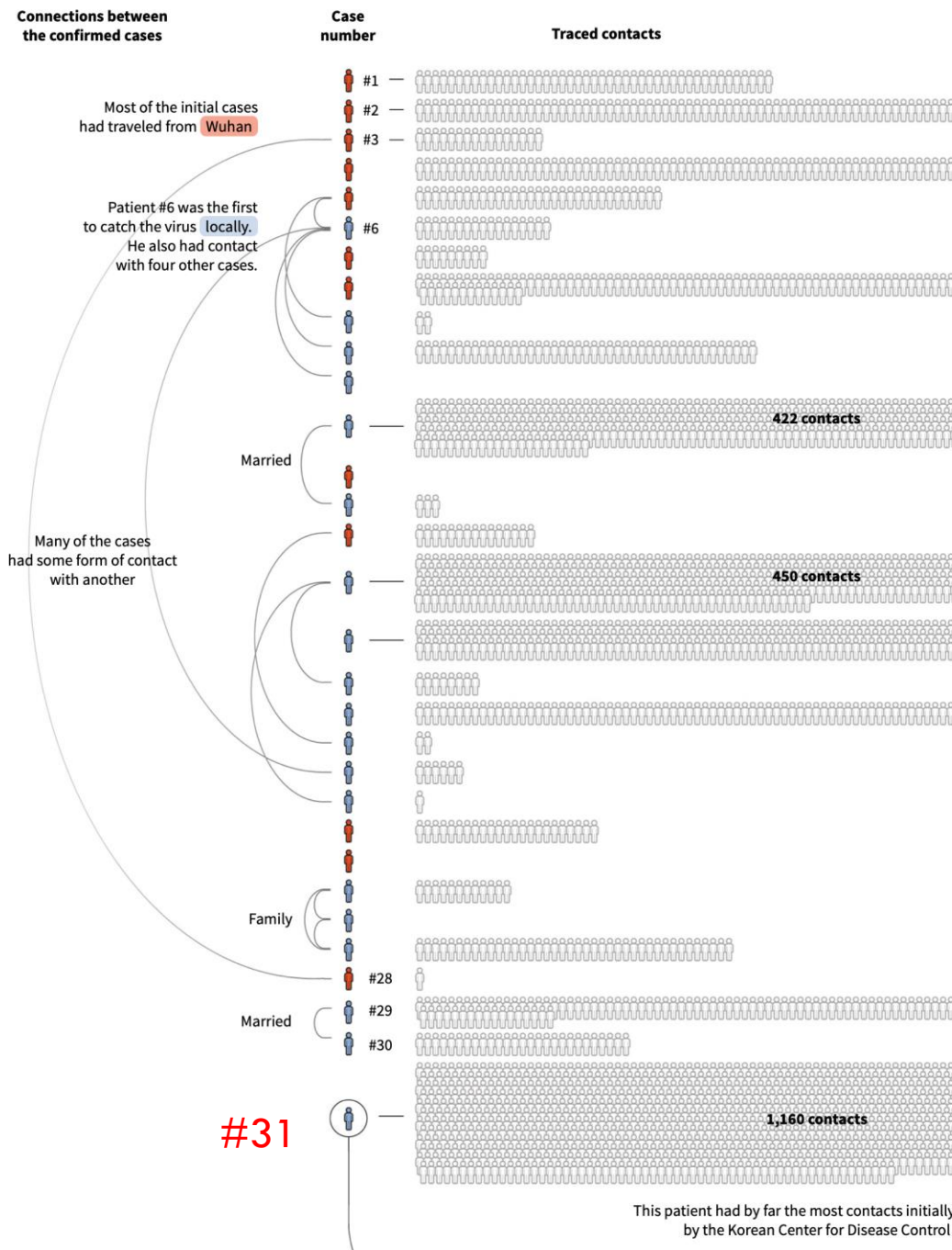
- Only modification: when normalizing, we have  $(N-1)*(N-2)$  instead of  $(N-1)*(N-2)/2$ , because we have twice as many ordered pairs as unordered pairs

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

# Directed geodesics

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





It's not clear where Patient 31 became infected with the virus, but in the days before her diagnosis, she travelled to crowded spots in Daegu, as well as in the capital Seoul. On February 6 she was in a minor traffic accident in Daegu, and checked herself into an Oriental medicine hospital. While at that hospital, she attended services at the Daegu branch of the Shincheonji Church of Jesus, on February 9 and again on February 16.

In between those visits, on February 15, doctors at the hospital said they first suggested she be tested for the coronavirus, as she had a high fever. Instead, the woman went to a buffet lunch with a friend at a hotel. In an interview with local newspaper JoongAng Ilbo, the woman denied that doctors had advised her to be tested. As her symptoms worsened, however, doctors say they once again advised her to be tested. On February 17, she finally went to another hospital for the test. The next day, health authorities announced she was the country's 31st confirmed case. In only a matter of days, those numbers had soared as hundreds of people at the Shincheonji Church and surrounding areas tested positive.

# World City System

TABLE 3  
RANKING OF CITIES ON MEASURES OF POWER AND PRESTIGE

| Rank    | Outdegree     | Value | Closeness     | Value | Betweenness   | Value | Indegree      | Value |
|---------|---------------|-------|---------------|-------|---------------|-------|---------------|-------|
| 1 ..... | Tokyo         | 3,639 | Paris         | 55.51 | Paris         | 25.65 | New York      | 1,425 |
| 2 ..... | New York      | 2,601 | Tokyo         | 53.59 | Tokyo         | 15.04 | London        | 1,086 |
| 3 ..... | Paris         | 2,535 | London        | 53.47 | Düsseldorf    | 13.61 | Paris         | 944   |
| 4 ..... | London        | 1,955 | New York      | 52.87 | London        | 13.31 | Tokyo         | 762   |
| 5 ..... | Düsseldorf    | 1,278 | San Francisco | 51.47 | New York      | 10.01 | Los Angeles   | 538   |
| 6 ..... | Amsterdam     | 897   | Düsseldorf    | 50.90 | San Francisco | 7.29  | Chicago       | 477   |
| 7 ..... | Zurich        | 893   | Amsterdam     | 50.84 | Munich        | 4.89  | Brussels      | 452   |
| 8 ..... | Munich        | 881   | Munich        | 50.05 | Oslo          | 4.60  | Amsterdam     | 435   |
| 9 ..... | Osaka         | 787   | Chicago       | 49.55 | Vevey         | 4.46  | Singapore     | 434   |
| 10 .... | San Francisco | 755   | Stockholm     | 49.43 | Zurich        | 4.32  | Hong Kong     | 424   |
| 11 .... | Frankfurt     | 515   | Toronto       | 49.06 | Beijing       | 4.23  | Toronto       | 412   |
| 12 .... | Vevey         | 491   | Zurich        | 48.97 | Atlanta       | 4.22  | Madrid        | 338   |
| 13 .... | Chicago       | 455   | Los Angeles   | 48.62 | Amsterdam     | 4.09  | Philadelphia  | 334   |
| 14 .... | Stockholm     | 427   | Madrid        | 48.58 | Stockholm     | 3.99  | Milan         | 322   |
| 15 .... | Dallas        | 413   | Dallas        | 48.46 | Osaka         | 3.98  | San Francisco | 321   |
| 16 .... | Detroit       | 359   | Houston       | 48.38 | Saint Louis   | 2.95  | Mexico City   | 280   |
| 17 .... | Utrecht       | 336   | Detroit       | 48.28 | Detroit       | 2.71  | Sydney        | 262   |
| 18 .... | Toronto       | 324   | Singapore     | 48.26 | Melbourne     | 2.61  | São Paulo     | 260   |
| 19 .... | Saint Louis   | 315   | Brussels      | 48.19 | Dallas        | 2.49  | Dallas        | 252   |
| 20 .... | Basel         | 304   | Seoul         | 48.18 | Omaha         | 2.33  | Munich        | 250   |
| 21 .... | Philadelphia  | 299   | Osaka         | 48.15 | Chicago       | 2.32  | Detroit       | 243   |
| 22 .... | Atlanta       | 285   | Atlanta       | 48.08 | Basel         | 2.19  | Houston       | 235   |
| 23 .... | Oslo          | 283   | Saint Louis   | 48.02 | Philadelphia  | 1.98  | Washington    | 227   |
| 24 .... | Beijing       | 273   | Mexico City   | 47.87 | Turin         | 1.72  | Atlanta       | 224   |
| 25 .... | Hamilton      | 250   | Milan         | 47.76 | Houston       | 1.69  | Bangkok       | 212   |
| 26 .... | Omaha         | 245   | Hong Kong     | 47.72 | Ludwigshafen  | 1.65  | Stockholm     | 194   |

# Centrality Other Options

There are other options, usually based on generalizing some aspect of those above:

- *Random Walk Betweenness* (Mark Newman). Looks at the number of times you would expect node  $i$  to be on the path between  $k$  and  $j$  if information traveled a 'random walk' through the network.
- *Peer Influence* based measures (Friedkin and others). Based on the assumed network autocorrelation model of peer influence. In practice it's a variant of the eigenvector centrality measures.
- *Subgraph centrality*. Counts the number of cliques of size 2, 3, 4, ...  $n-1$  that each node belongs to. Reduces to (another) function of the eigenvalues. Very similar to influence & information centrality, but does distinguish some unique positions.
- *Fragmentation centrality* – Part of Borgatti's Key Player idea, where nodes are central if they can easily break up a network.
- Moody & White's *Embeddedness* measure is technically a group-level index, but captures the extent to which a given set of nodes are nested inside a network
- • *Removal Centrality* – effect on the rest of the (graph for any given statistic) with the removal of a given node. Really gets at the system-contribution of a particular actor.