Ανάλυση Κοινωνικών Δικτύων Και Εφαρμογές

Μετρικές Ανάλυσης ΙΙ

Συμεών Παπαβασιλείου (papavass@mail.ntua.gr)
Βασίλειος Καρυώτης (vassilis@netmode.ntua.gr)

Topics

- Centrality metrics
  - General definition & Examples of Use
  - Degree Centrality
  - Closeness Centrality
  - Betweenness Centrality
    - Shortest path
    - Routing
  - Eigenvector Centrality
  - Network Centralization
- Comparative Examples

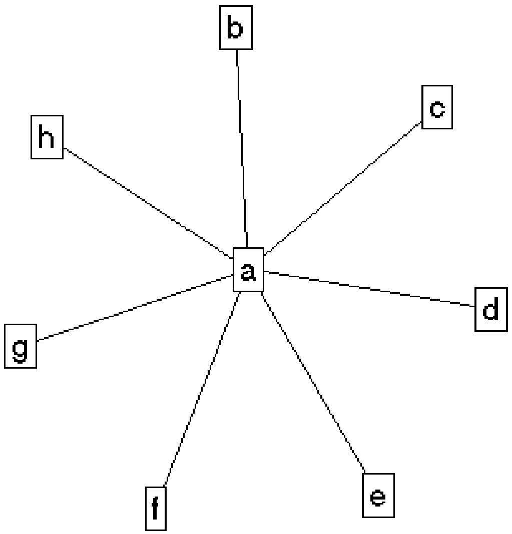
Network Centrality Issues

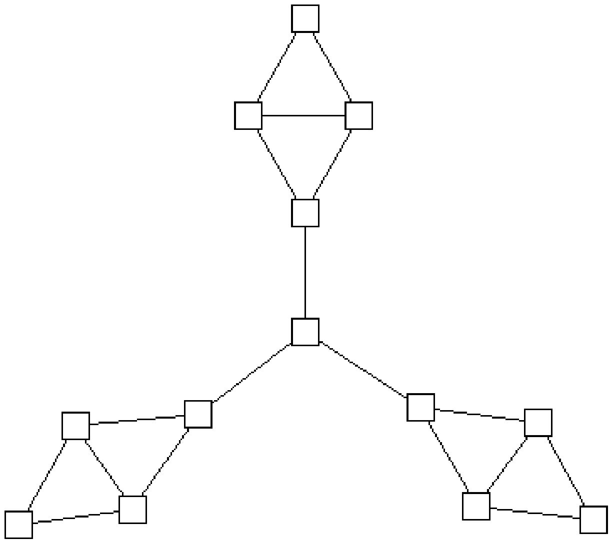
- ✓ Which nodes are most ‘central’?
- ✓ Definition of ‘central’ varies by context/purpose.
- ✓ **Local measure:**
  degree
- ✓ **Relative to rest of network:**
  closeness, betweenness,
  eigenvector (Bonacich power centrality)
- ✓ How evenly is centrality distributed among nodes?
  centralization…

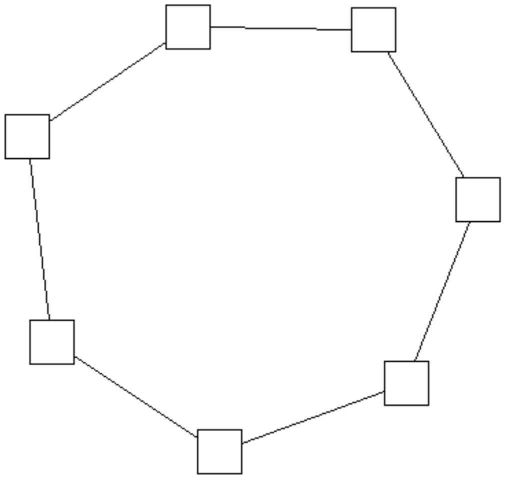
Centrality in Social Networks

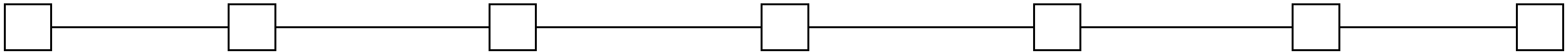
Intuitively, we want a method that allows us to distinguish "important" actors

Consider the following graphs:









Centrality: Definition

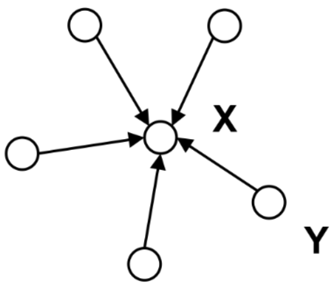
- Structural centrality $\rightarrow$ importance of a node
- Absolute/relative measures of centrality

Centrality Types

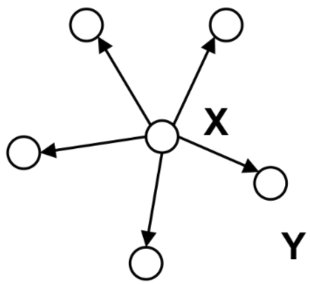
- **Node centralities:** importance of a node in a network
  - Strategic location of an actor within the network:
    1. Measures based on degree
    2. Measures based on geodesics
    3. Measures based on spectral properties of the graph
- **Network centralities:** importance of groups

Centrality: who's important based on their network position

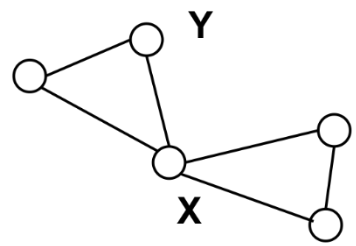
In each of the following networks, *X* has higher centrality than *Y* according to a particular measure



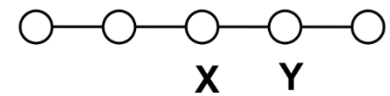
in-degree



out-degree



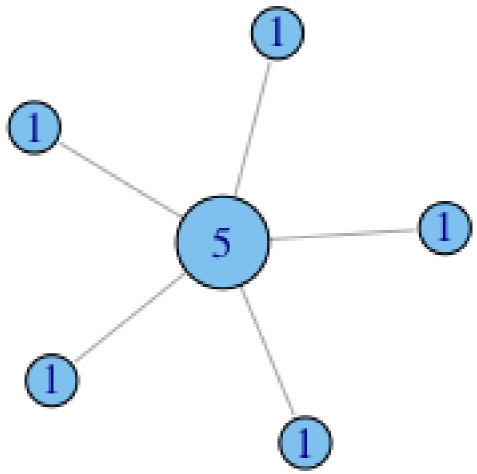
betweenness



closeness

Degree Centrality (1)

He who has many friends is most important.



When is the number of connections the best centrality measure?

- people who will do favors for you
- people you can talk to

Degree Centrality (2)

- Number of one-hop neighbors of an actor (similar to node degree)
- Ties of an actor make him visible

$$C_D(k) = \sum_{i=1}^{n} a_{ik}$$

Adjacency Matrix

$$C'_D = \frac{\sum_{i=1}^{n} a_{ik}}{n-1}$$

Number of nodes in the network

- Expresses:
  - ✓ Frequency of visits by something taking an infinitely long random walk in a network
  - ✓ Measure of immediate influence-ability to infect others directly or in one time period due to popularity. Similarly for information control

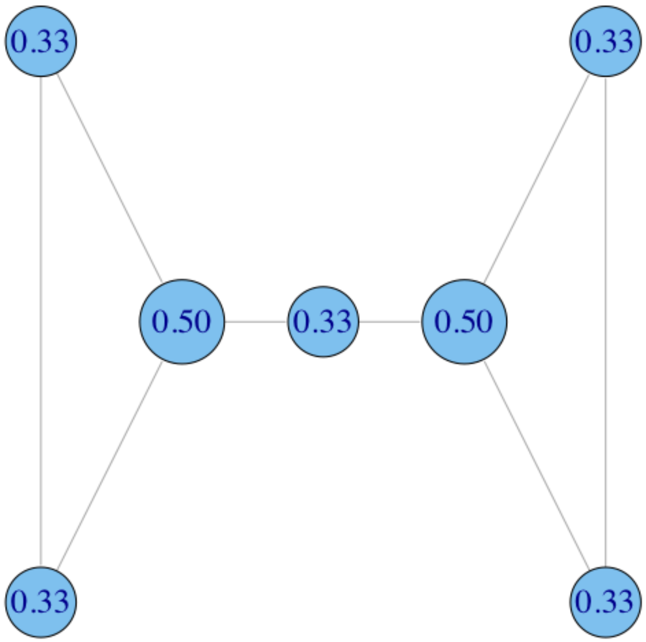
Degree Centrality (3)

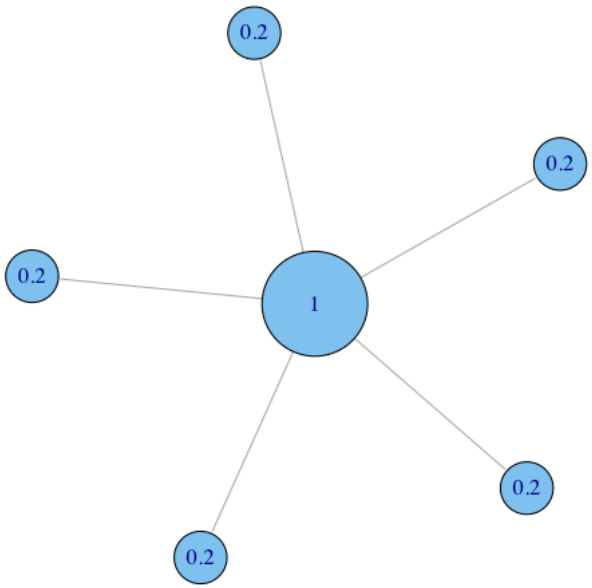
- Motivation comes from the star topology (the most central node has the highest degree)
- Linear function of node degree most often used
- Easy computation

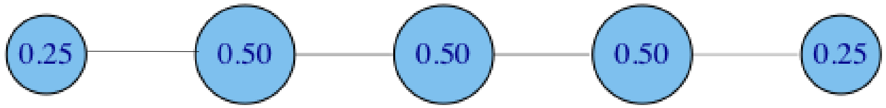
- **Does not capture** dynamics of information flow $\rightarrow$ the obtained centrality values may not be representative
  - i.e., suitable for malware propagation, not suitable in routing or delay and throughput management (bottlenecks may not be nodes with high degree)

Normalized Degree Centrality (4)

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







Centralization (Degree) of a graph (1)

If we want to measure the degree to which the graph as a whole is centralized, we look at the *dispersion* of centrality:

Simple: variance of the individual centrality scores.

$$S_D^2 = \left[\sum_{i=1}^{g} (C_D(n_i) - \bar{C}_d)^2 \right] / \text{n}$$

Mean degree centrality

Or, using **Freeman's general** formula for centralization (which ranges from 0 to 1):

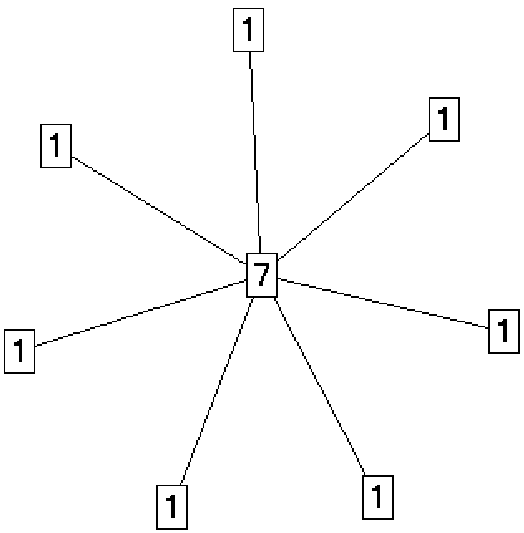
Maximum degree centrality

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

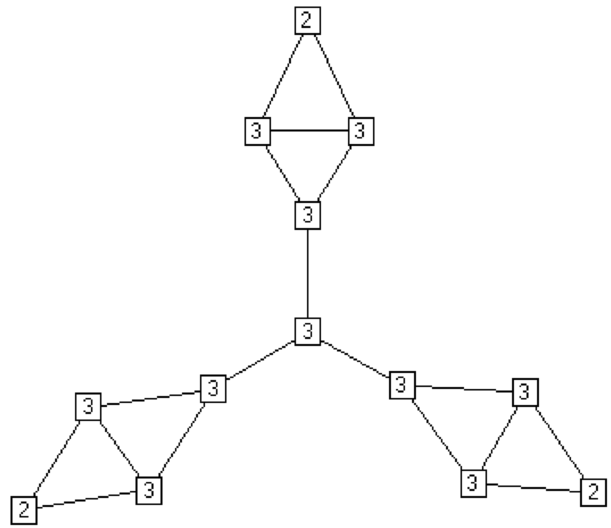
UCINET, SPAN, PAJEK and most other network software will calculate these measures.

Centralization (Degree) of a graph (2)

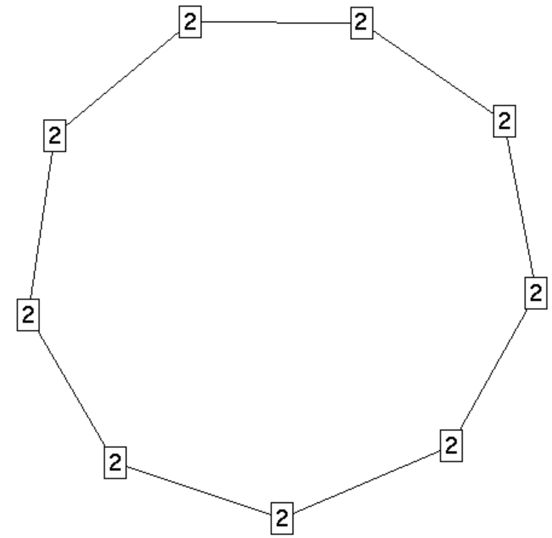
Degree Centralization Scores



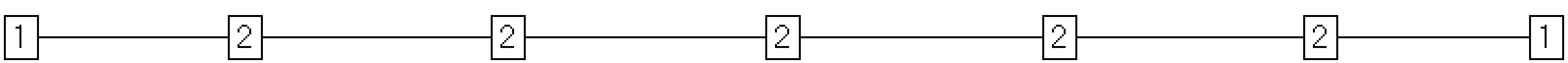
Freeman: 1.0
Variance: 3.9



Freeman: .02
Variance: .17



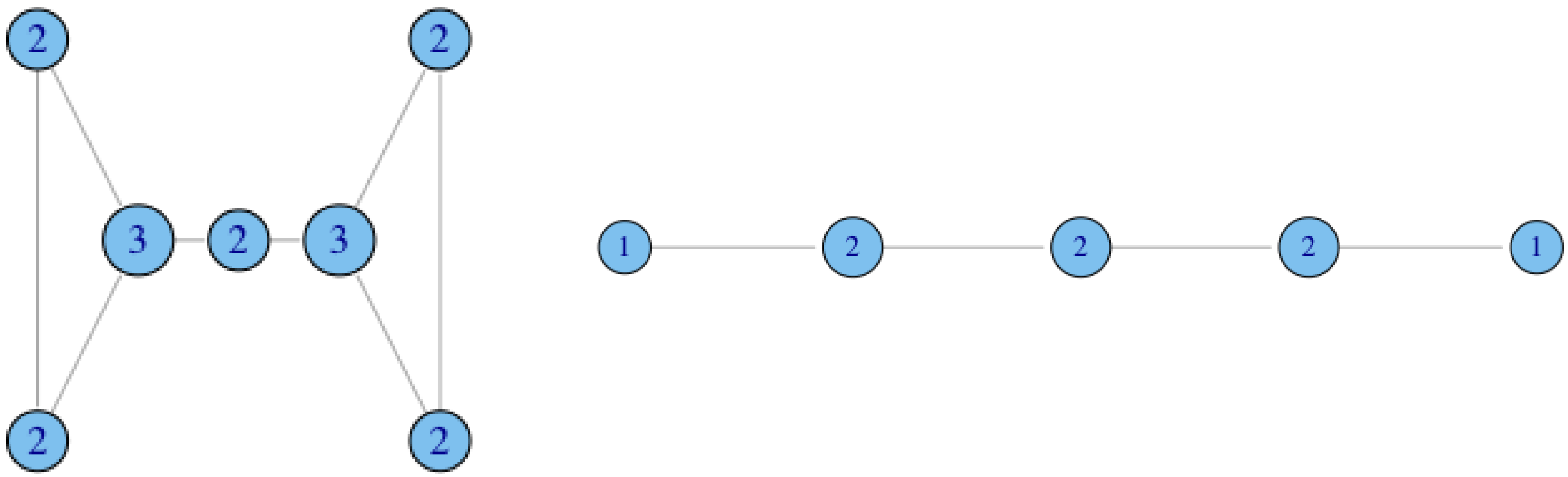
Freeman: 0.0
Variance: 0.0



Freeman: .07
Variance: .20

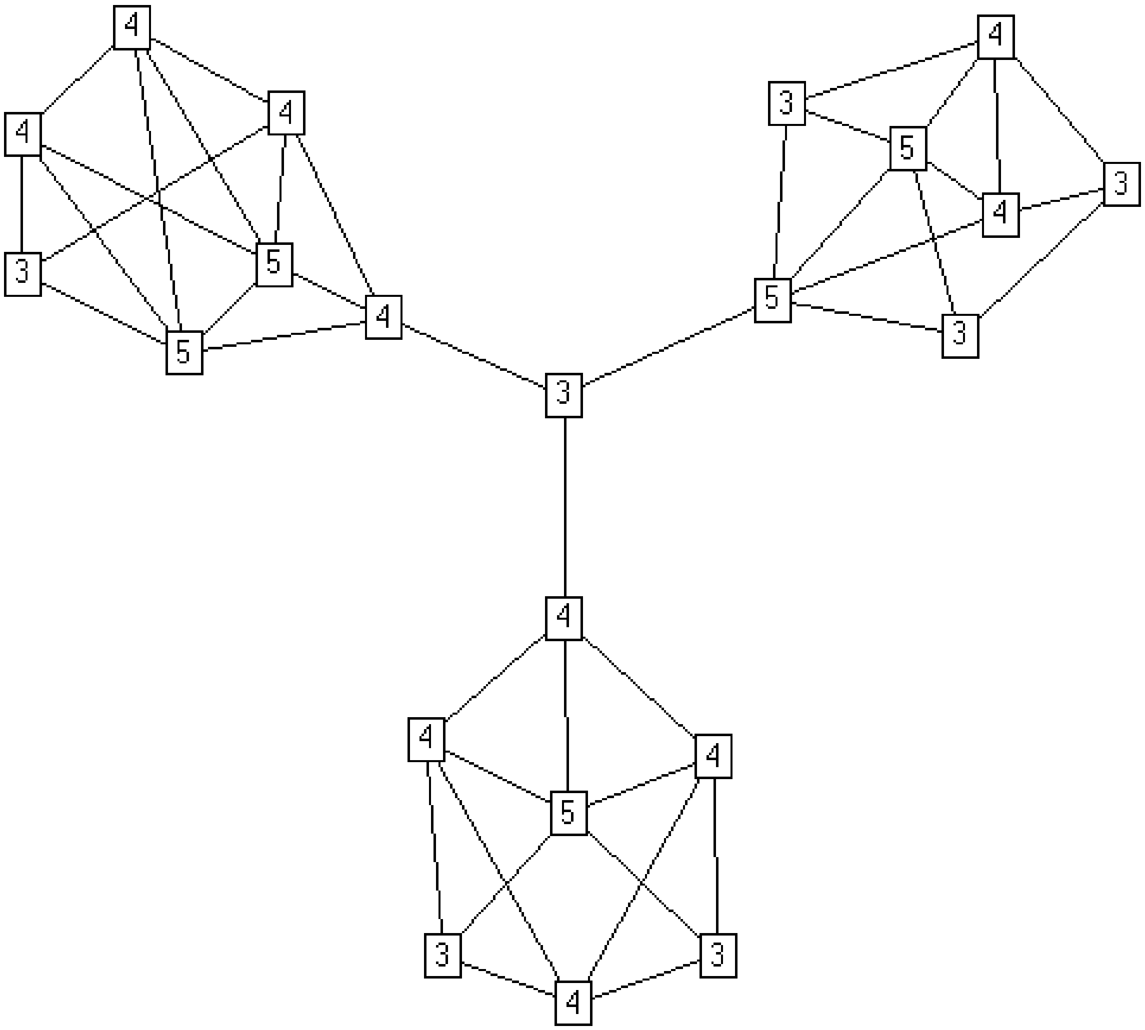
When degree isn't everything (1)

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…

When degree isn't everything (2)



Degree centrality, however, can be deceiving, *because it is purely local measure*.

Closeness Centrality (1)

- Defined via the distance of a node from all other nodes

$$C_P(k) = \left(\sum_{i=1}^{n} d(i,k) \right)^{-1}$$

$$C'_P = \left(\frac{\sum_{i=1}^{n} d(i,k)}{n-1} \right)^{-1} = \frac{n-1}{\sum_{i=1}^{n} d(i,k)}$$

Shortest distance (e.g. in hops)

- Expresses:
  - ✓ Efficiency of information propagation from an actor to all other actors
  - ✓ Identifies the most spatially central nodes:
    - Nodes are central with respect to the edges of the topology and the employed distance metric

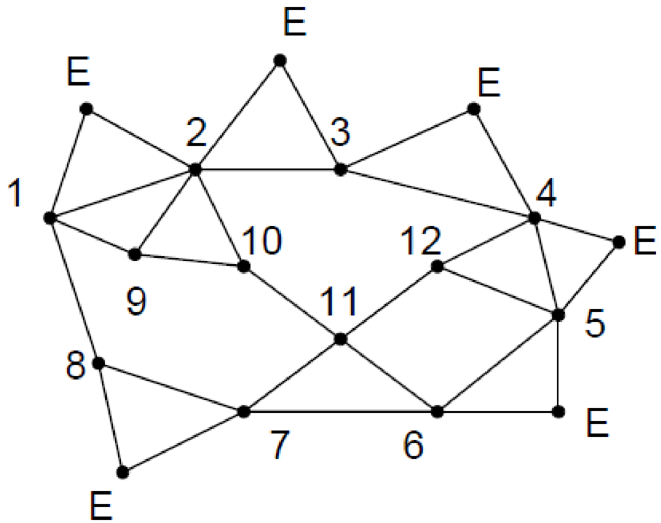
Closeness Centrality (2)

✓ Suitable when information travels through shortest paths:

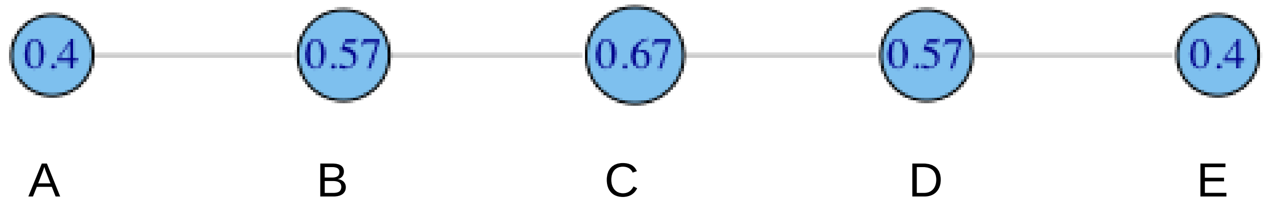
Low closeness score ➡ low distance from others ➡ receive information sooner (well positioned to receive novel information earlier)

✓ Expected time until arrival

- Problem: The network should be fully connected



Closeness Centrality (3): Toy Example

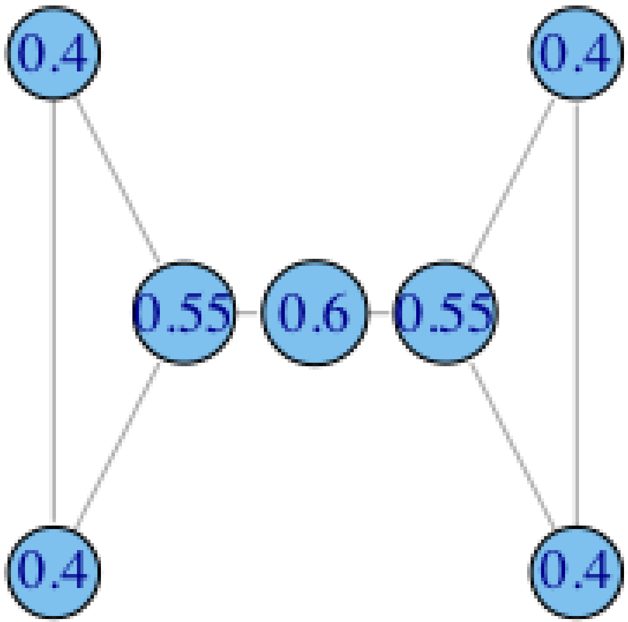


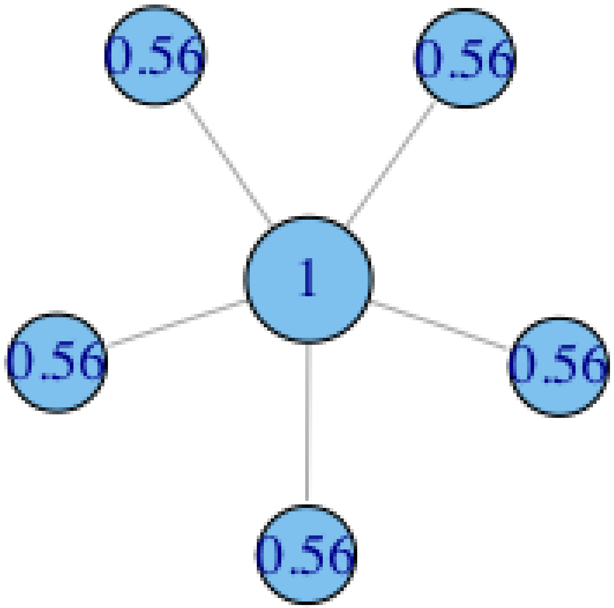
$$C_P(A) = \left[\frac{\sum_{j=1}^{n} d(A,j)}{n-1}\right]^{-1} = \left[\frac{1+2+3+4}{4}\right]^{-1} = \left[\frac{10}{4}\right]^{-1} = 0.4$$

$$C_P(B) = \left[\frac{\sum_{j=1}^{n} d(B,j)}{n-1}\right]^{-1} = \left[\frac{1+1+2+3}{4}\right]^{-1} = \left[\frac{7}{4}\right]^{-1} = 0.57$$

$$C_P(C) = \left[\frac{\sum_{j=1}^{n} d(C,j)}{n-1}\right]^{-1} = \left[\frac{2+1+1+2}{4}\right]^{-1} = \left[\frac{6}{4}\right]^{-1} = 0.67$$

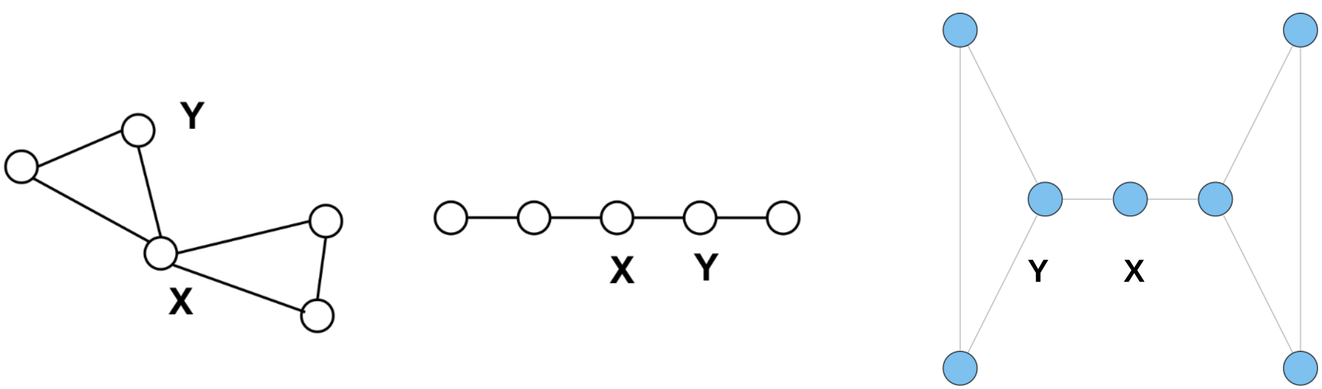
Closeness Centrality (4): Toy Example





Betweenness: another centrality measure

- **intuition**: how many 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

- Based on the notion of shortest paths, on the frequency that a node falls in the shortest paths connecting other pairs of nodes.
- Potential of control of central nodes: Central nodes control higher part of the information flow if routing takes place among shortest paths.
- Measures the dominance of a single node.

Betweenness Centrality: Shortest Paths

- Partial betweenness $b_{ij}(p_k)=0$ if disconnected
  - Assume g_{ij} the # of geodesics linking p_i, p_j and $g_{ij}(p_k)$ those containing p_k
- Partial betweenness of node p_k

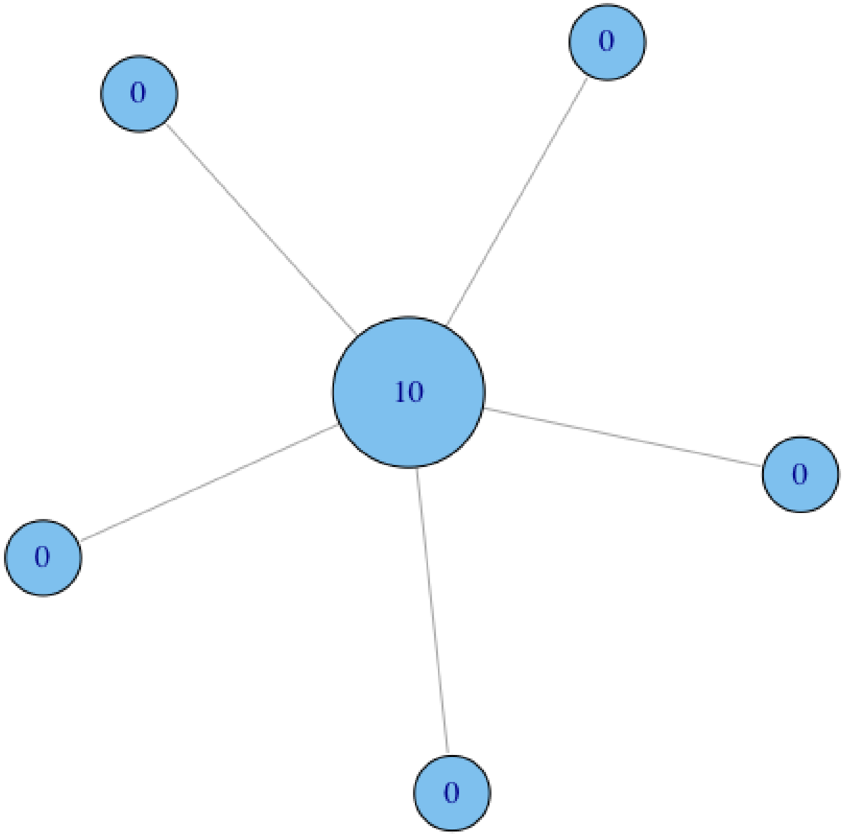
$$b_{ij}(p_k) = \left(\frac{1}{g_{ij}}\right)\left(g_{ij}(p_k)\right) = \frac{g_{ij}(p_k)}{g_{ij}}$$

- Prob. node p_k falls on a random geodesic linking p_i, p_j
- Centrality of node p_k is the sum of all unordered pairs:

$$C_B(p_k) = \sum_{i<j}^{n} \sum^{n} b_{ij}(p_k)$$

Betweenness Centrality Star Graph

- non-normalized version:



Betweenness Centrality: Shortest Paths

- Maximum possible Betweenness centrality: Star or Wheel graph (Freeman)

$$C_B^{\max}(k) = \frac{n(n-1)}{2} - (n-1) = \frac{n^2 - 3n + 2}{2}.$$

- Betweenness centrality independent of network size:

$$C'_B(k) = \frac{2C_B(k)}{n^2 - 3n + 2}$$

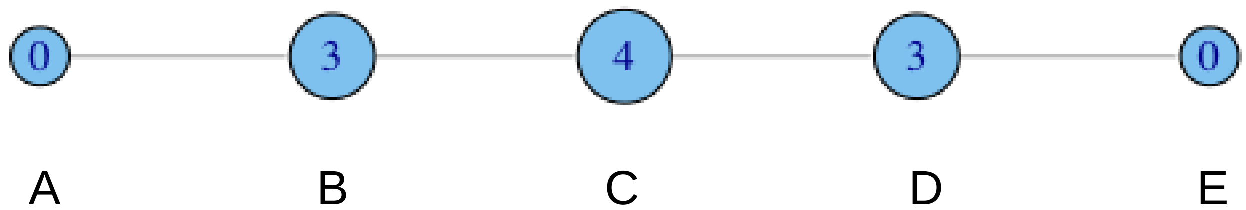
- Demanding computation especially for large scale networks:

  Approximations, Ego centralities

- Betweenness centrality works both for connected & disconnected networks

Betweenness on toy networks (1)

- non-normalized version:

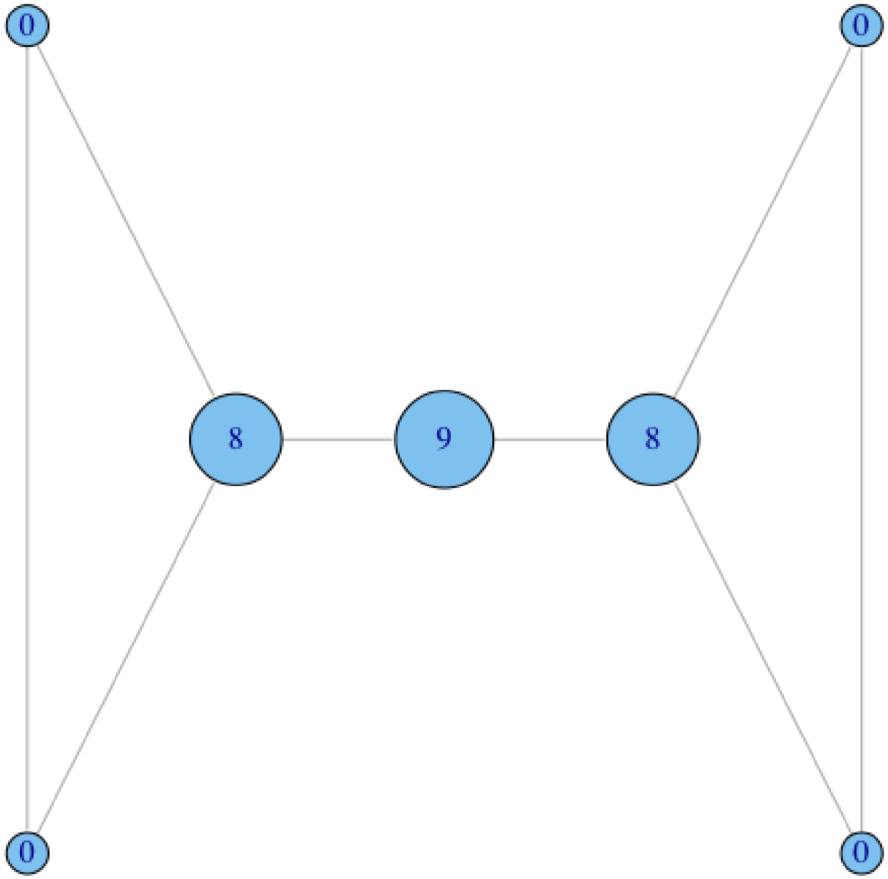


- A lies between no two other vertices
- B lies between A and 3 other vertices: C, D, and E
- C lies between 4 pairs of vertices (A,D),(A,E),(B,D),(B,E)

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

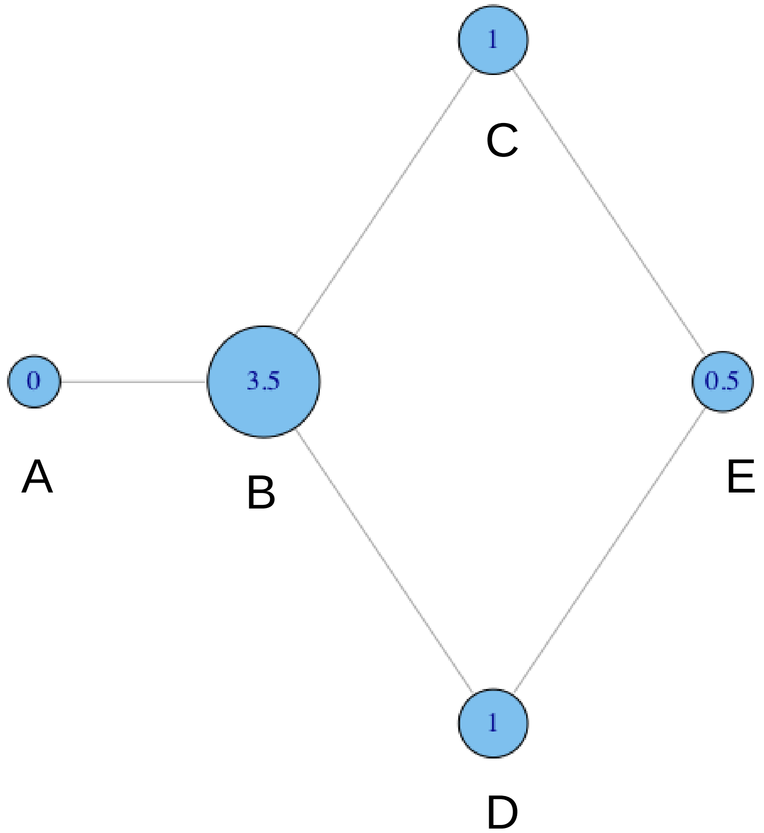
Betweenness on toy networks (2)

- non-normalized version:



Betweenness on toy networks (3)

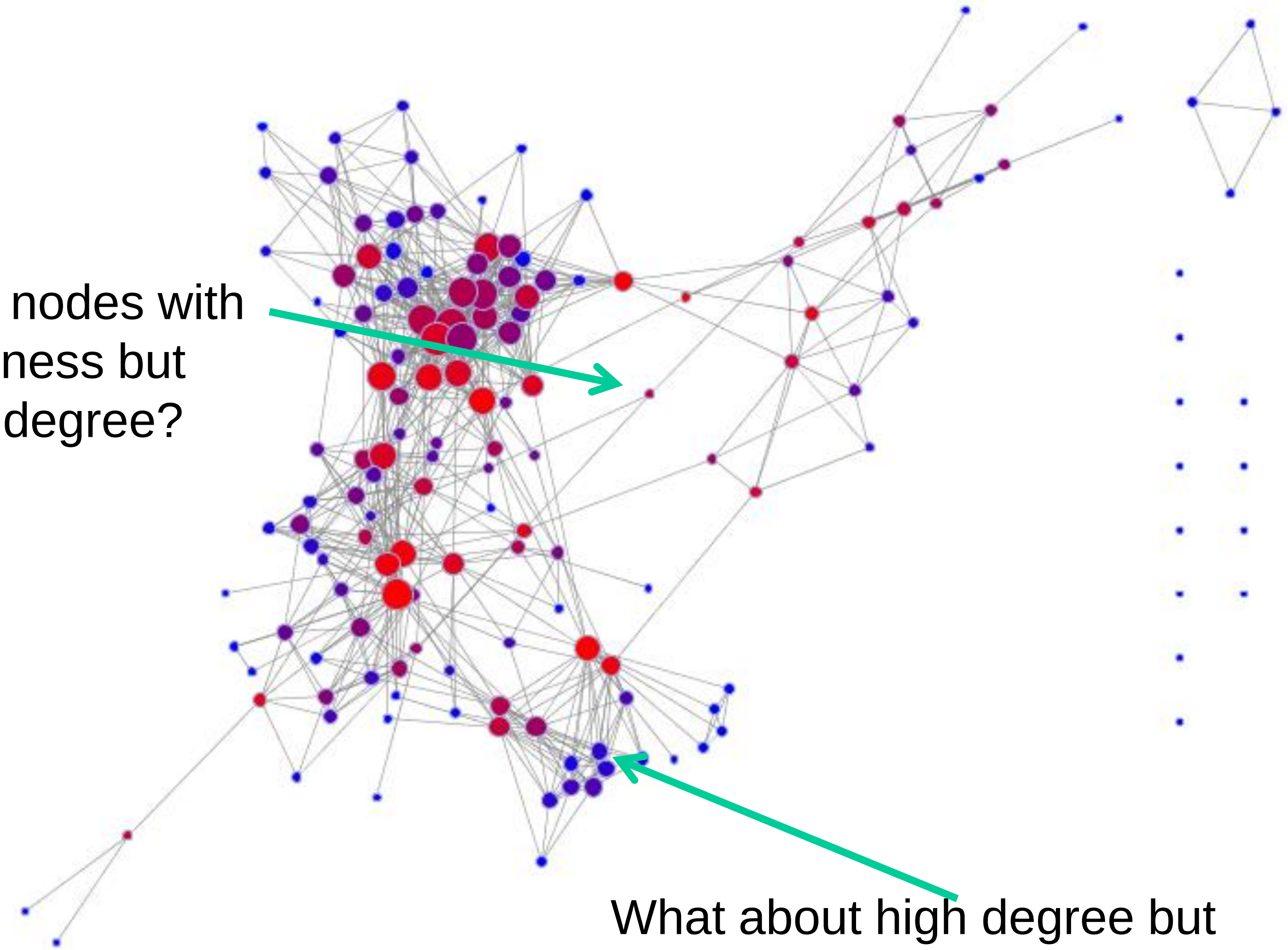
- non-normalized version:



- 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
- Can you figure out why B has betweenness 3.5 while E has betweenness 0.5?

Betweenness vs. Degree Centrality

Nodes are sized by degree, and colored by betweenness.



Can you spot nodes with high betweenness but relatively low degree?

What about high degree but relatively low betweenness?

Comparing Centrality Measures (Properties)

Comparing across these 3 centrality values:

- Generally, the 3 centrality types will be positively correlated.

	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.	

More Types of Centrality (1)

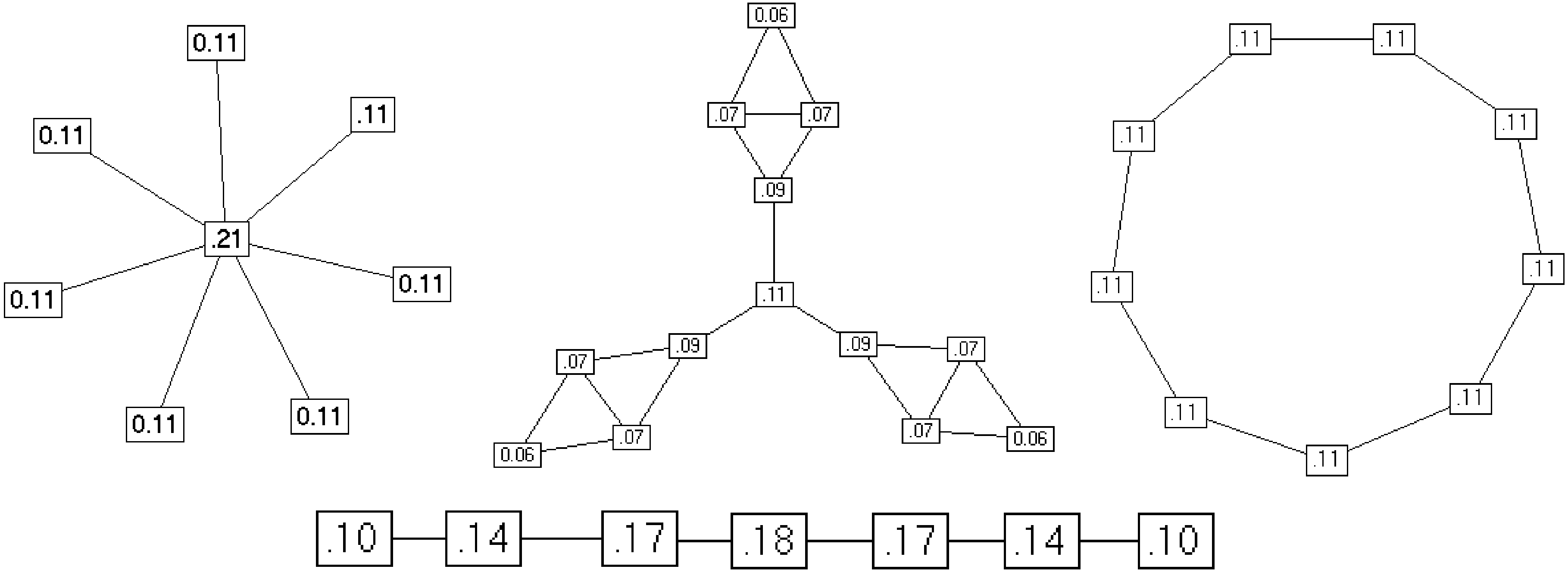
It is quite likely that information can flow through paths other than the geodesic.

Information Centrality:

- Scores all paths in the network, and weights them based on their length.

Routing Centrality:

- More general definition than shortest path betweenness centrality.
- Loop free routing scheme.



More Types of Centrality (2)

- Bridging Centrality: multiplying the shortest- path betweenness centrality by a bridging coefficient
  - ✓ The bridging coefficient is the ratio of the inverse of a node degree to the sum of the inverses of all its neighbor degrees
  - ✓ The bridging coefficient of an actor describes how well the actor is located between high-degree actors

- Spectral centrality (Eigenvector, Bonacich): takes into account both the number of ties of an actor and the quality of neighboring actors

…..and many more….

More Types of Centrality (3)

…..and many more….

e.g.

- Bounded-distance Betweenness Centrality
- Length-scaled Betweenness (like information) Centrality
- Edge Betweenness Centrality
- Group Betweenness Centrality
- Maximal-Flow Betweenness Centrality
- Stress Centrality
- Traffic Load Centrality

Eigenvector (Spectral) Centrality

- Principal eigenvector of the adjacency matrix A (largest eigenvalue)
- Takes into consideration the importance or centrality of the direct neighbors of a node (similar to Google's Pagerank)

$$\lambda v = Av \Rightarrow$$
$$v = \frac{1}{\lambda} Av$$

$$v_i = \frac{1}{\lambda} \sum_{j=1}^{N} a_{ij} v_j, \ \forall\ i$$

- Connected graph $\rightarrow$ A irreducible $\rightarrow$ (Perron Frobenius)

$$v_i > 0, \ \forall\ i.$$