Centrality

CE642: Social and Economic Networks
Maryam Ramezani
Sharif University of Technology
maryam.ramezani@sharif.edu

01

Centrality

Why a centrality measure?

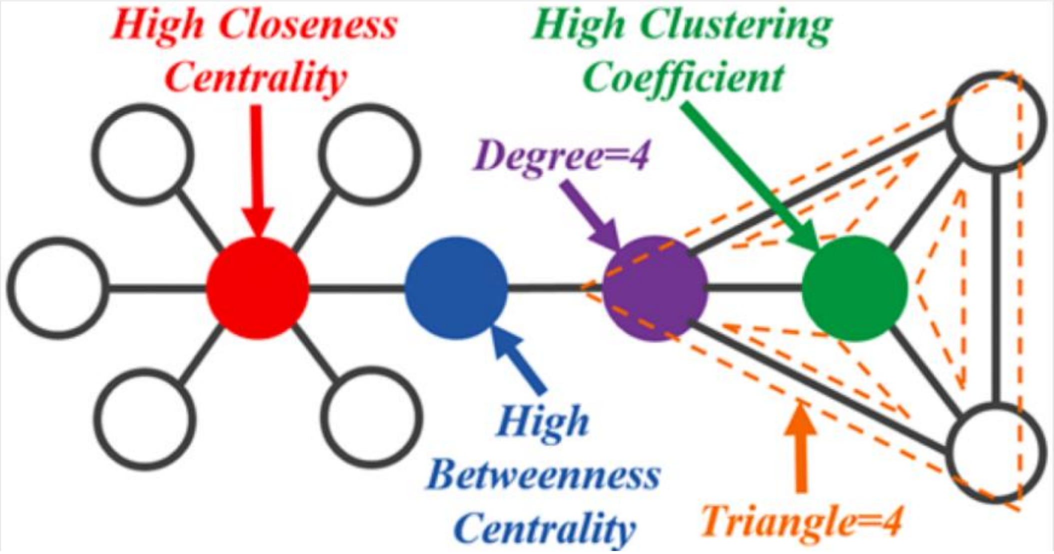
- Often, we are interested in identifying IMPORTANT network components
  - Nodes
  - Edges
- Central components may play critical role in network functions
  - Robustness
  - Collective behavior
  - Synchronization
  - Information spreading
  - Social dynamics
  - …

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
  - …

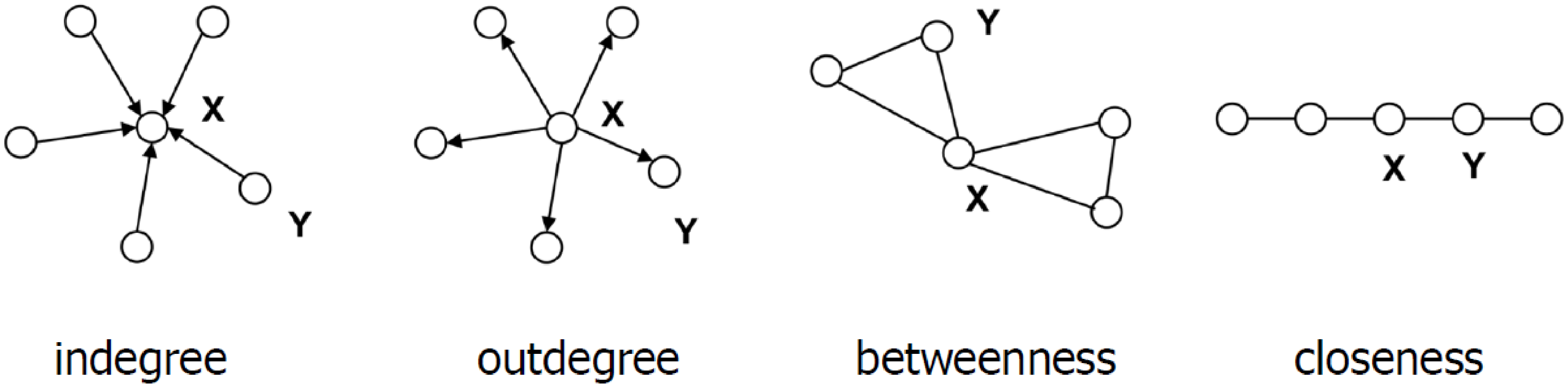
Network Centrality

- Given a social network, which nodes are more **important** or **influential**?
- **Centrality measures** were proposed to account for the importance of the nodes of a network



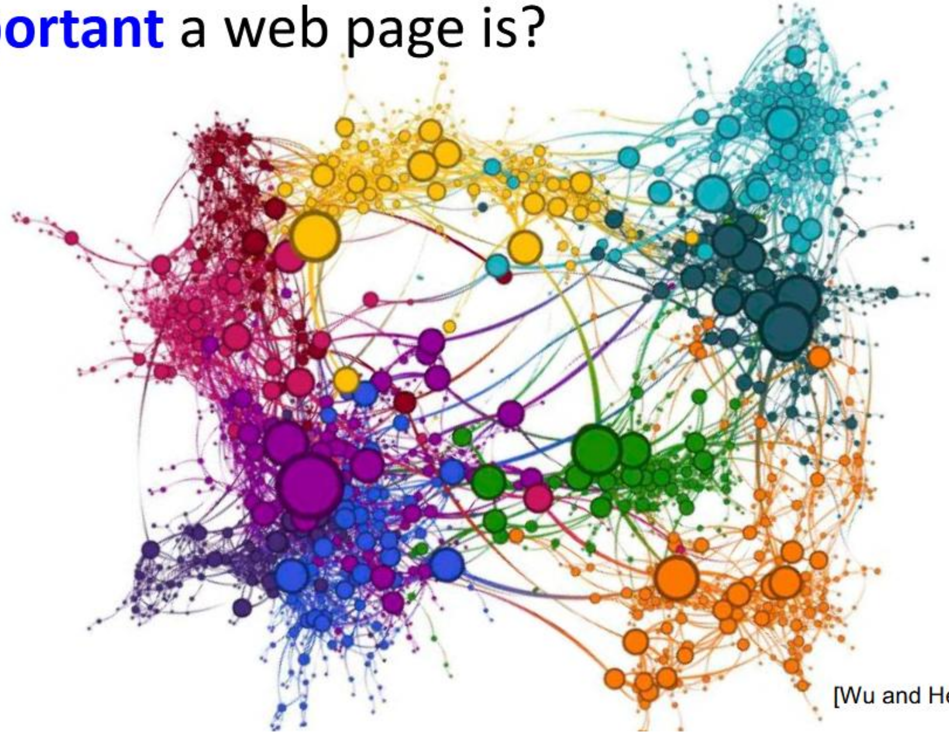
Network Centrality

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



Network Centrality

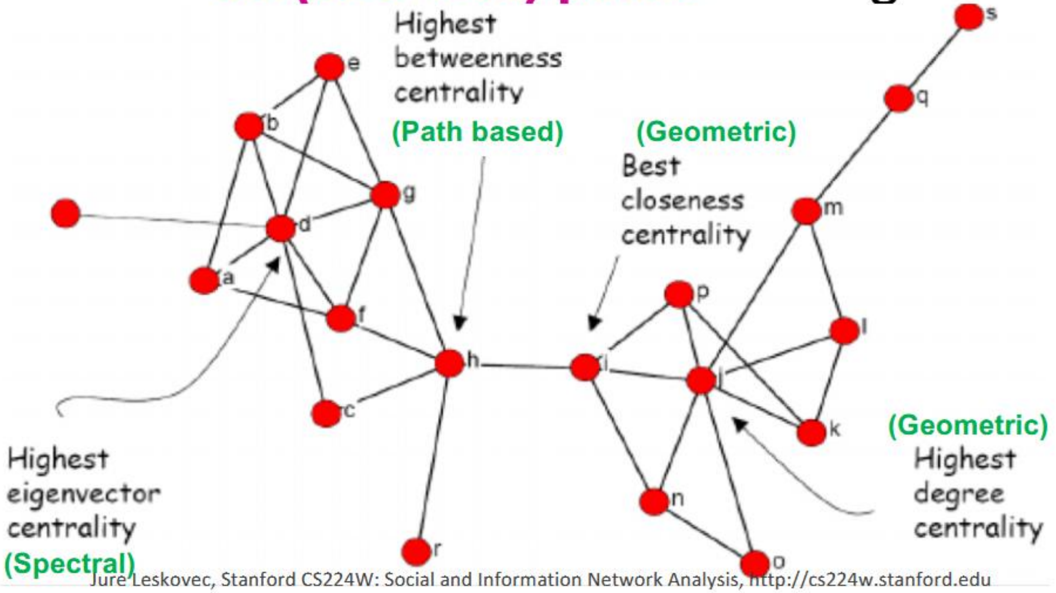
- **Centrality** is used often for detecting:
  - How **influential** a person is in a social network?
  - How **well used** a road is in a transportation network?
  - How **important** a web page is?



[Wu and He'15]

Centrality Measures

- **Geometric Measures**:
  - Importance is a **function of distances** to other nodes.
- **Spectral Measures**:
  - Based on the **eigen-structure** of some graph-related matrix
- **Path-based Measures**:
  - Take into account **all (shortest) paths** coming into a node



Jure Leskovec, Stanford CS224W: Social and Information Network Analysis, http://cs224w.stanford.edu 4

02

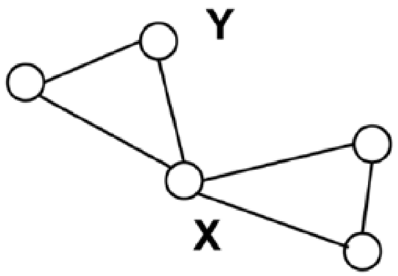
Path-based Measures for Centrality

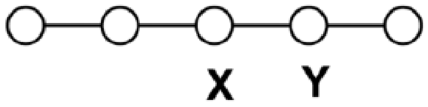
Betweenness Centrality

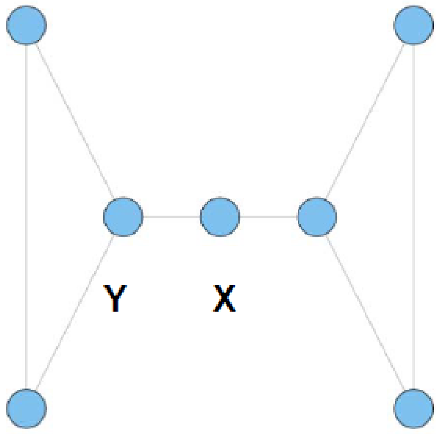
- Measure of centrality in a graph based on shortest paths:
  - Edge Betweenness Centrality
  - Node Betweenness Centrality
- In a telecommunications network, a node with higher betweenness centrality would have more control over the network, because more information will pass through that node.

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?







Edge Betweenness Centrality

$$\rho_{ij} = \sum_{p \neq q} \left(\Gamma_{pq}\left(e_{ij}\right) / \Gamma_{pq} \right)$$

Γ_{pq} is the number of shortest paths from the *p*-th to the *q*-th node

$\Gamma_{pq}(e_{ij})$ is the number of these paths making use of e_{ij}.

- **Usually the betweenness is normalized by** $\boxed{/\,[(n-1)(n-2)/2]}$

Number of possible edges

Node Betweenness Centrality

$$C_i = \sum_{p \neq i \neq q} \left(\Gamma_{pq}(i) / \Gamma_{pq} \right)$$

Γ_{pq} is the number of shortest paths from the *p*-th to the *q*-th node

$\Gamma_{pq}(i)$ is the number of these shortest paths making use of the *i*-th node (except those that are start or end nodes is *i*).

$$/[(n-1)(n-2)/2]$$

- **Usually the betweenness is normalized by**

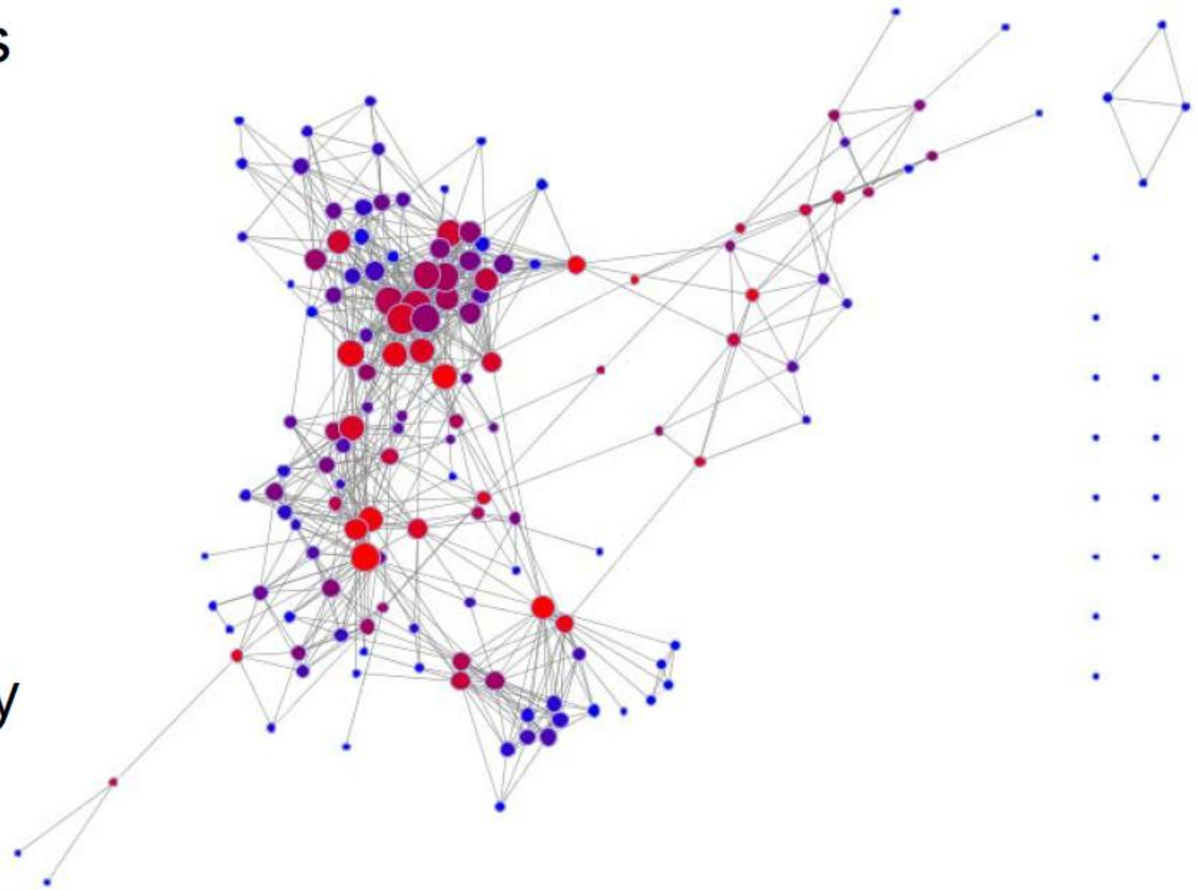
number of pairs of vertices excluding the vertex itself

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?

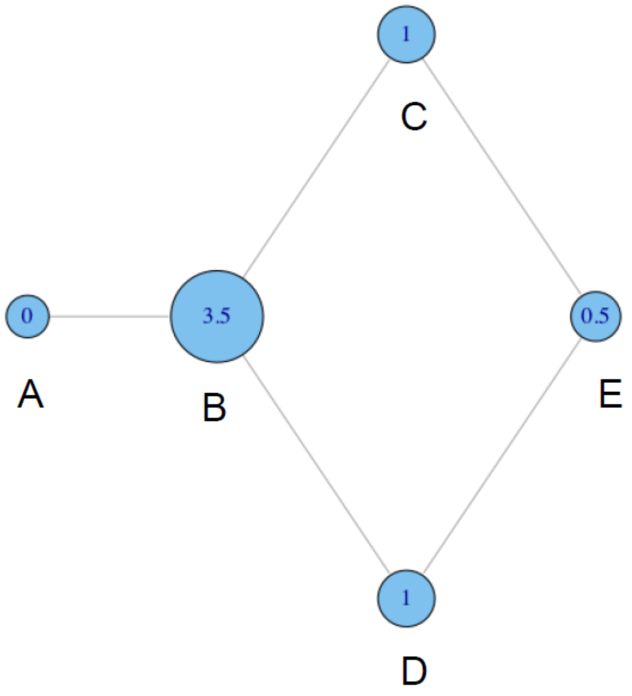


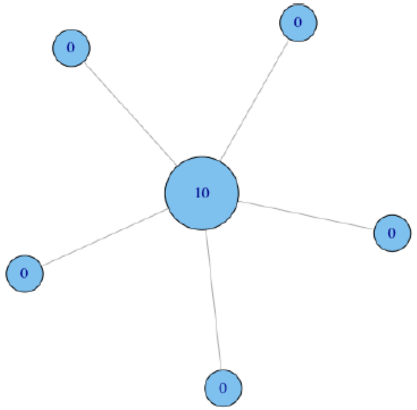
Betweenness Centrality

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: $\frac{1}{2}+\frac{1}{2} = 1$

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





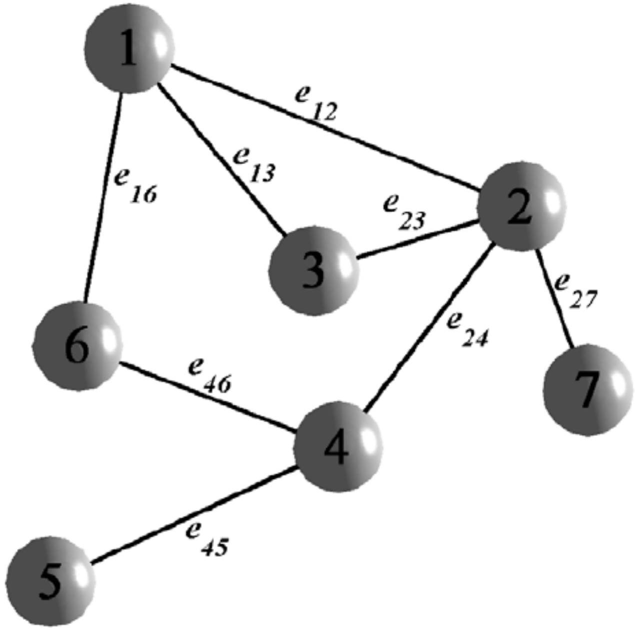
Connection Graph Stability Scores

- 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



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

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

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

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

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

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

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

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

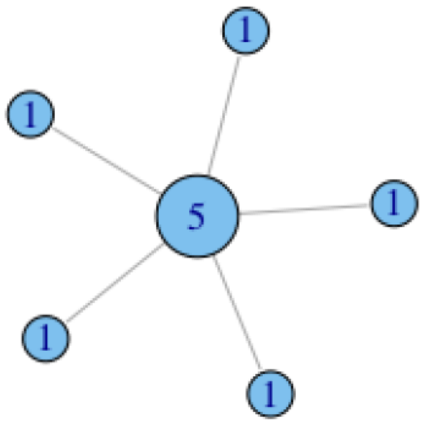
$$b_{46} = |P_{15}| + |P_{26}| + |P_{46}| + |P_{56}| = 3 + 2 + 1 + 2 = 8.$$

03

Geometric Measures for Centrality

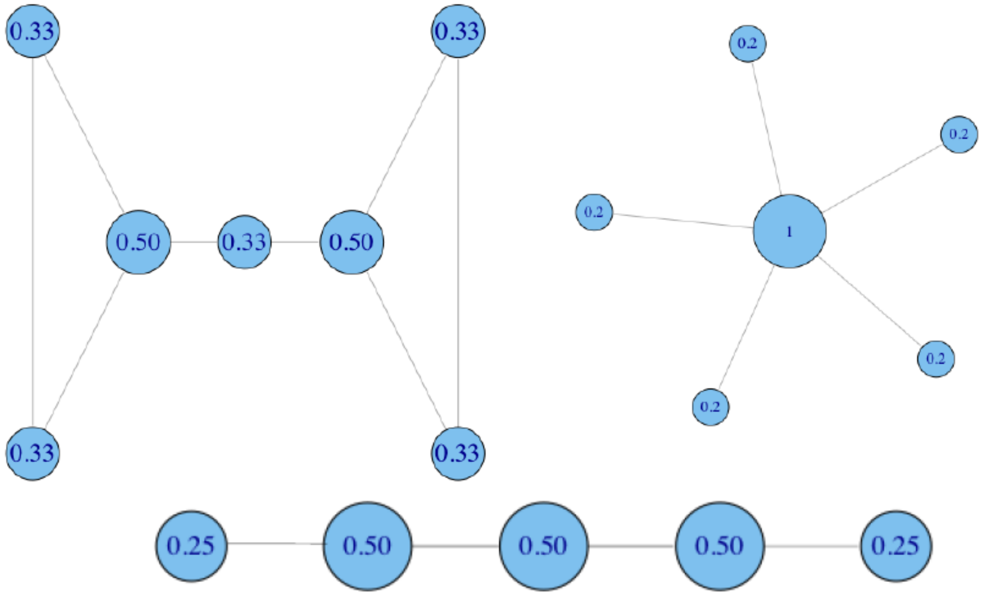
Degree centrality (undirected)

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



When is the number of connections the best centrality measure?

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



Normalized degree centrality:

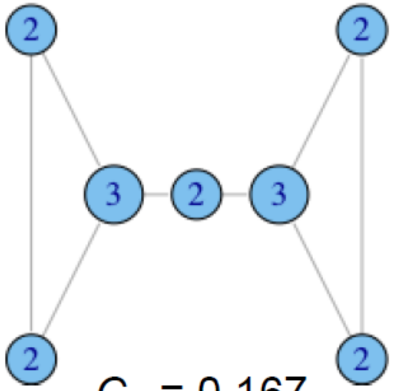
Degree is divided by the max. possible, i.e. (N-1)

How equal are the nodes?

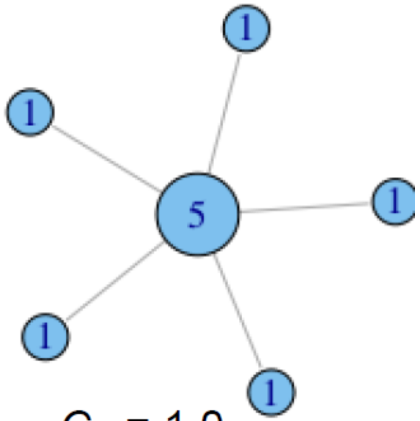
- How much variation is there in the centrality scores among the nodes?
- Freeman’s general formula for centralization (can use other metrics, e.g. Gini coefficient or standard deviation):

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 = 0.167



C_D = 1.0

C_D = 0.167

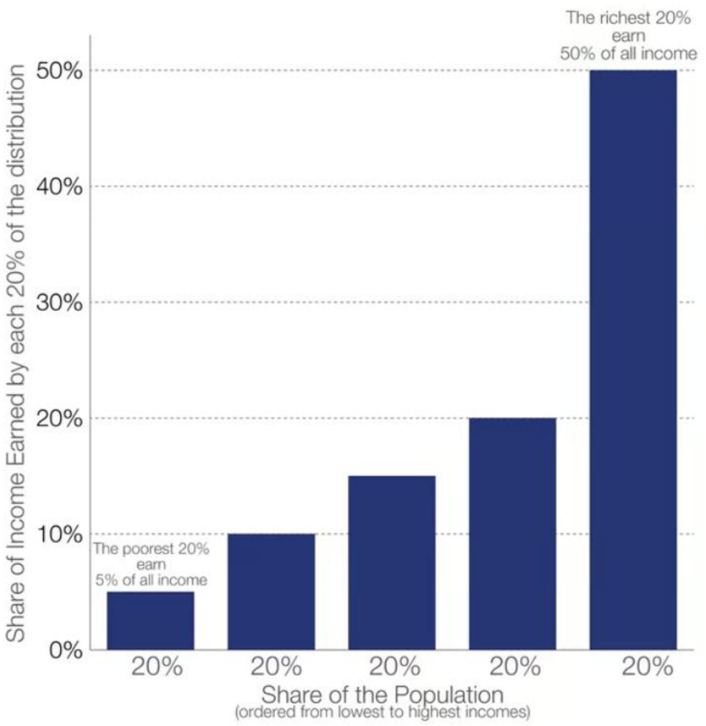
Gini Coefficient (Index)

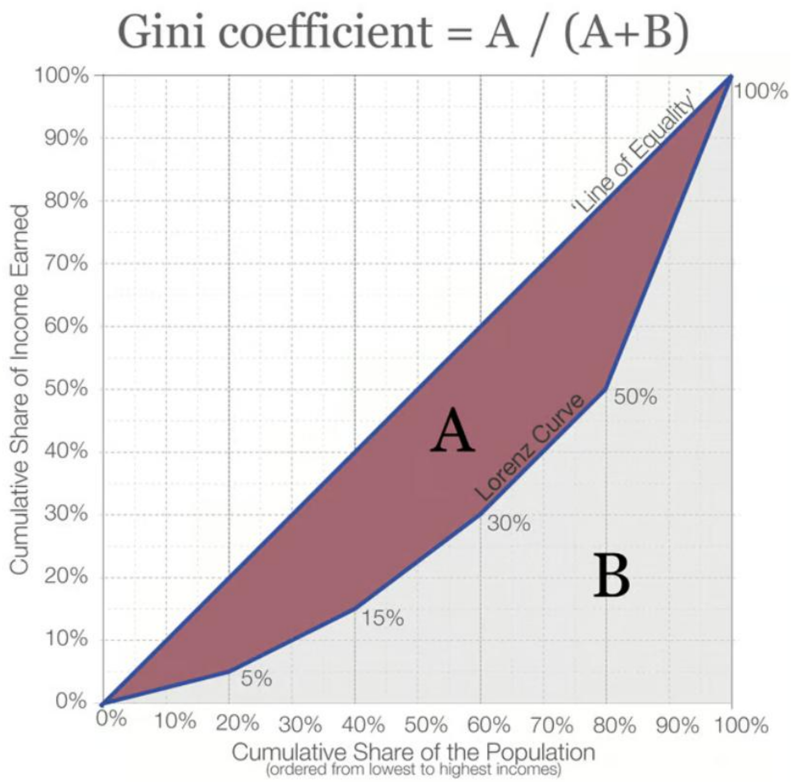
The bar chart on the left shows a simple distribution of incomes. The total population is split up in 5 parts and ordered from the poorest to the richest 20%.
The bar chart shows how much income each 20% part of of the income distribution earns.

The chart on the right shows the same information in a different way, both axis show the cumulative shares:
The poorest 20% of the population earn 5% of the total income, the next 20% earn 10% – so that the poorest 40% of the population earn 15% etc.
The curve resulting from this way of displaying the data is called the Lorenz Curve.

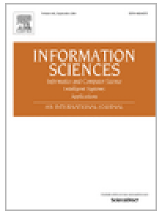
If there was no income inequality the resulting Lorenz Curve would be a straight line – the 'Line of Equality'.
A larger area (A) between the Lorenz Curve and the Line of Equality means a higher level of inequality.

The ratio of A/(A+B) is therefore a measure of inequality and is referred to as the Gini coefficient, Gini index, or simply the Gini.





Gini Coefficient (Index)



Information Sciences
Volume 462, September 2018, Pages 16-39

Sparsity measure of a network graph: Gini index

Swati Goswami [a] [b] [1], C.A. Murthy [a], Asit K. Das [b]

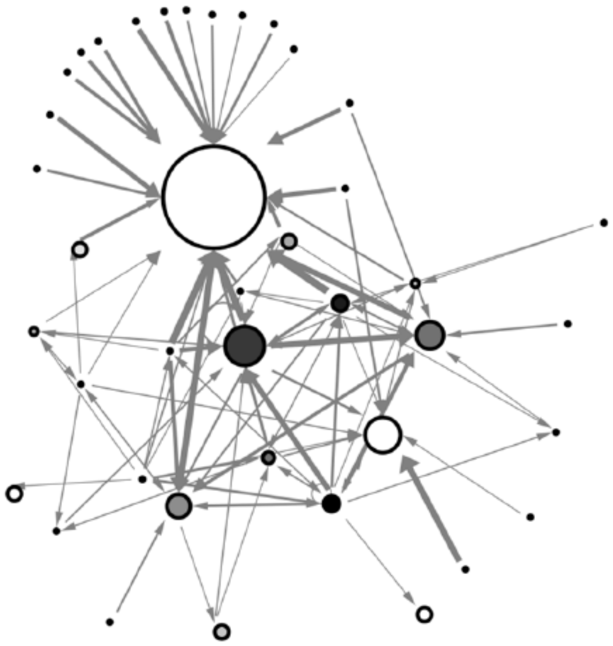
Show more

Add to Mendeley Share Cite

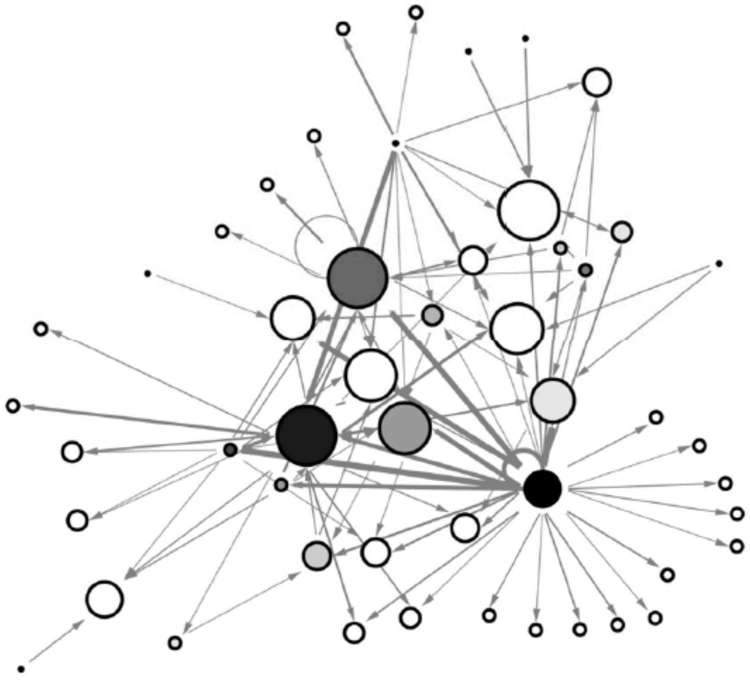
https://doi.org/10.1016/j.ins.2018.05.044

Get rights and content

Example: financial trading networks



high centralization: one node trading with many others



low centralization: trades are more evenly distributed

Characteristic path length

- A network with N nodes
  - Compute the shortest path (distance) between any two nodes d_{ij}
  - The length of the path is the number of edges (unweighted networks) or the weighted sum of the edges (weighted networks)
  - If the nodes are not connected, the path length between them is set to infinity
  - It is also called average geodesic distance
  - If dij is infinity, it diverges
  - Many times we compute the average only over the connected pairs of nodes (that is, we ignore “infinite” length paths)
  - $l = \frac{1}{N(N-1)} \sum_{i,j,i \neq j} d_{ij}$

Efficiency

- In this way the divergence is avoided
- The inverse of efficiency E is called harmonic mean
- Efficiency is an indicator of the traffic capacity of the
- network
- The couple of disconnected nodes have a contribution of zero in computing E
- The more the values of E are the more the communication-efficient the network is
- It is also called global efficiency of the network.

$$E = \frac{1}{N(N-1)} \sum_{i,j,i \neq j} \frac{1}{d_{ij}}$$

Efficiency

- Higher Efficiency → Faster communication, better connectivity.
- Why is Efficiency Important?
  - 🚀 High Efficiency → Network is Well-Connected
    - Information spreads quickly.
    - Fewer intermediate steps needed.
    - Helps in optimizing transportation, communication, and social interactions.
  - 🐢 Low Efficiency → Poor Connectivity
    - Long paths between nodes.
    - Slower communication and bottlenecks.
    - Less effective in handling information flow.

$$E = \frac{1}{N(N-1)} \sum_{i,j,i \neq j} \frac{1}{d_{ij}}$$

The Role of Connectivity in Efficiency

Network Type	Effect on E
Fully Connected (Strongly Connected Component - SCC)	🔺 High E, every node can reach any other node efficiently.
Weakly Connected Component (WCC)	🔻 Lower E, some nodes may only be accessible in one direction.
Disconnected Components	⛔ Very Low E, as some distances are infinite (ignored in practical calculations).

The Power of Shortcuts

- A shortcut in a graph is an additional edge that significantly reduces the shortest path distance between two nodes without being essential for the overall connectivity of the network.
- Extra direct connections between distant nodes.
- Reduce the shortest path distance d_{ij}.
- Improve efficiency without requiring full connectivity.
- Real-World Examples:
  - Social Networks → Influencers create bridges between distant groups.
  - Transport Networks → Highways and express routes reduce travel time.
  - Computer Networks → Fast routing through backbone connections (CDNs).
- Fewer hops → Shorter paths → Higher Efficiency.

Vulnerability

- It is important to know which component (nodes or edges) are crucial to the best performance.
- The more the drop in the efficiency by removing a component the more crucial that component.
- Degree (hub node) might be a criterion
  - Only degree is not enough, e.g. all vertices of a binary tree network have equal degree, i.e. no hub, but disconnection of vertices closer to the root and the root itself have a greater impact than of those near the leaves.
- The amount of change in the efficiency (or other network properties) as a component is removed can be an indicator of the vulnerability

Vulnerability

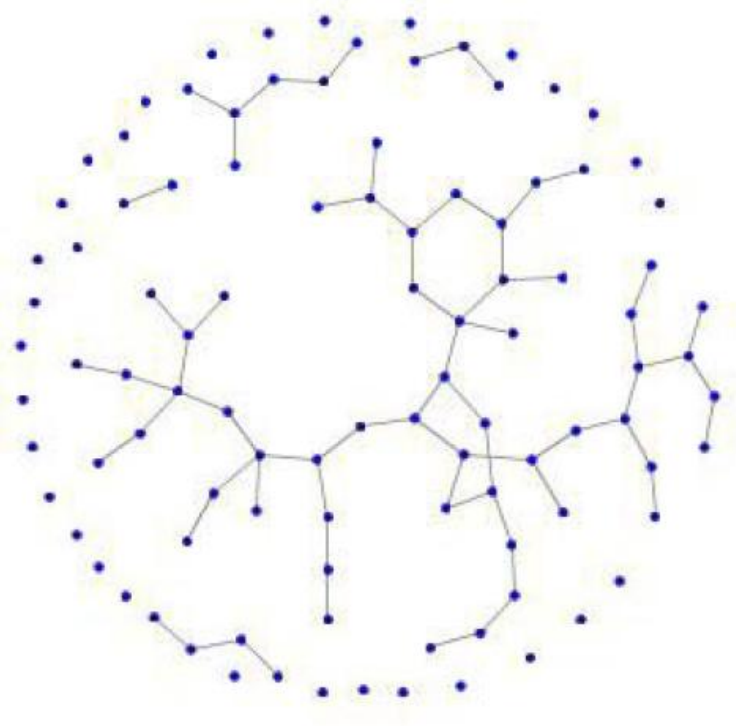
$$V_i = \frac{E - E_i}{E} \qquad V = \max_i V_i$$

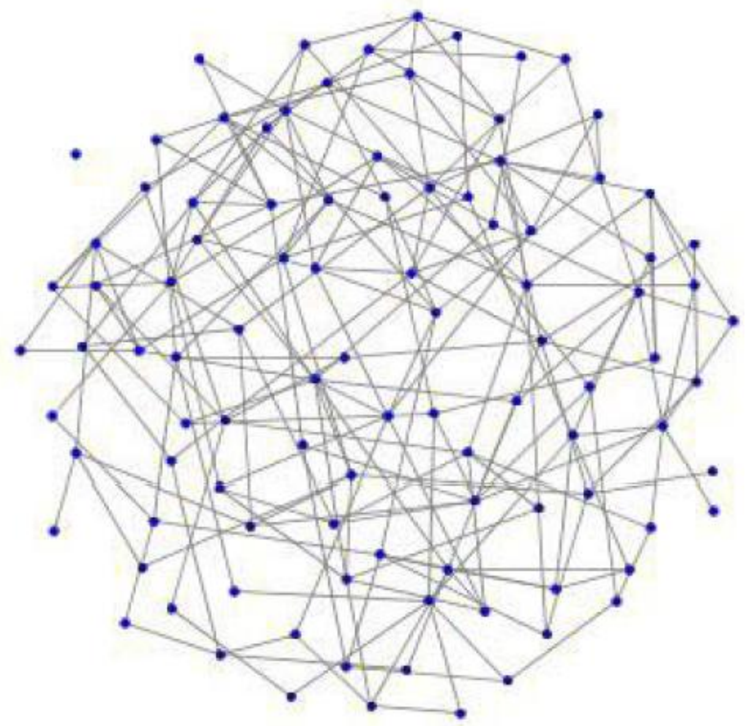
- where V_i is the vulnerability of component i and E_i is the of i E efficiency the networks by removed that component.

- V can be regarded as the vulnerability of the network the ordered distribution of nodes with respect to their vulnerability V_i is related to the network hierarchy.
- The most vulnerable (critical) node occupies the highest position in the network hierarchy.
- The same is also true for the edges.

Density

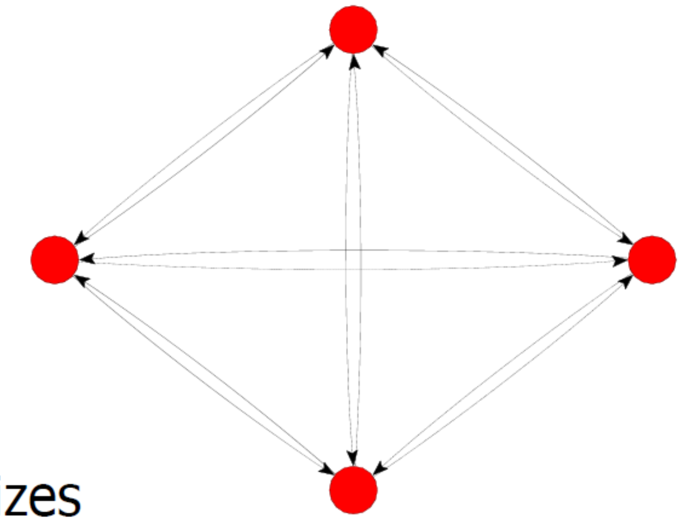
- How dense the networks are?





Density

- Number of the connections that may exist between n nodes
  - directed graph
    $e_{max} = n*(n-1)$
    each of the n nodes can connect to (n-1) other nodes
  - undirected graph
    $e_{max} = n*(n-1)/2$
    since edges are undirected, count each one only once
- What fraction are present?
  - density = e / e_{max}
  - For example, out of 12 possible connections, this graph has 7, giving it a density of 7/12 = 0.583
- Would this measure be useful for comparing networks of different sizes (different numbers of nodes)?



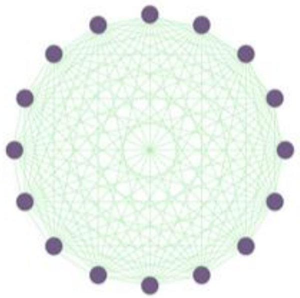
Density

- for *undirected* G *density* ρ is defined as

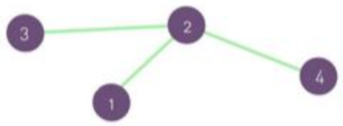
$$\rho = \frac{2m}{n(n-1)} = \frac{\langle k \rangle}{n-1}$$

- for *directed* G *density* $\rho^{\cdot}$ is defined as

$$\rho^{\cdot} = \frac{m}{n(n-1)} = \frac{\langle k^{\cdot} \rangle}{n-1}$$



complete $m = \binom{n}{2}$



tree $m = n - 1$

- G is *dense* if $\rho \to \text{const.}$ as $n \to \infty$ thus $\langle k \rangle = \mathcal{O}(n)$
- G is *sparse* if $\rho \to 0$ as $n \to \infty$ thus $\langle k \rangle \neq \mathcal{O}(n)$

Closeness

- What if it’s not so important to have many direct friends?
  - Degree Centrality is not important
- Or be “between” others
  - Betweeness Centrality is not important
- But one still wants to be in the “middle” of things, not too far from the center.
- Closeness is based on the length of the average shortest path between a node and all other nodes in the network

Example

- **High Degree Centrality**
  - A person who has many direct friends in a social network.
- **High Betweenness Centrality**
  - A person who acts as a bridge between two separate groups.
- **High Closeness Centrality**
  - A person who can reach anyone in the network with the fewest intermediaries, even if they don’t have many direct friends.
- Closeness Centrality focuses on being "near" other nodes in terms of short paths across the entire network, rather than just having many direct connections!

Closeness

Closeness Centrality:

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

Normalized Closeness Centrality

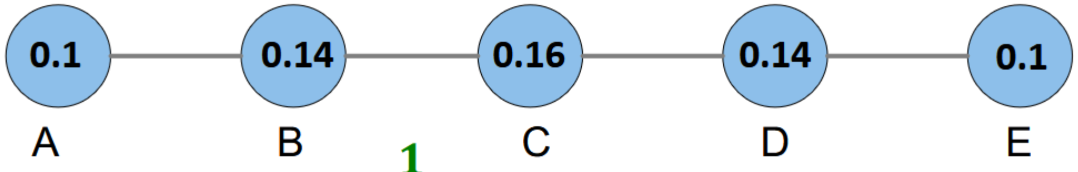
$$C'_C(i) = \left[\sum_{j=1}^{N} d(i,j)/(N-1)\right]^{-1}$$

- Formula:
  - **Closeness Centrality**:

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

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

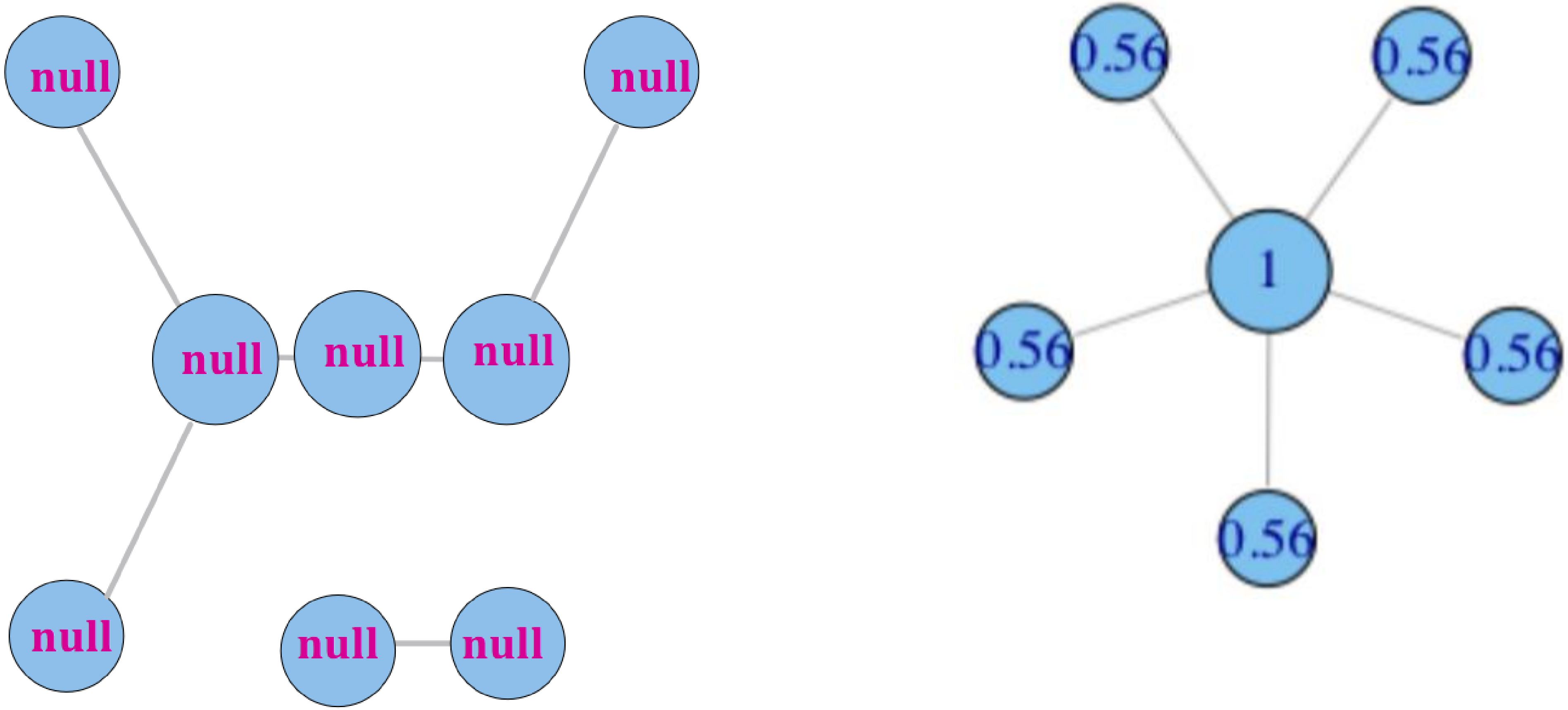


$$c_{clos}(A) = \frac{1}{1+2+3+4} = 0.1$$

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

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

Closeness Example



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

Harmonic Centrality

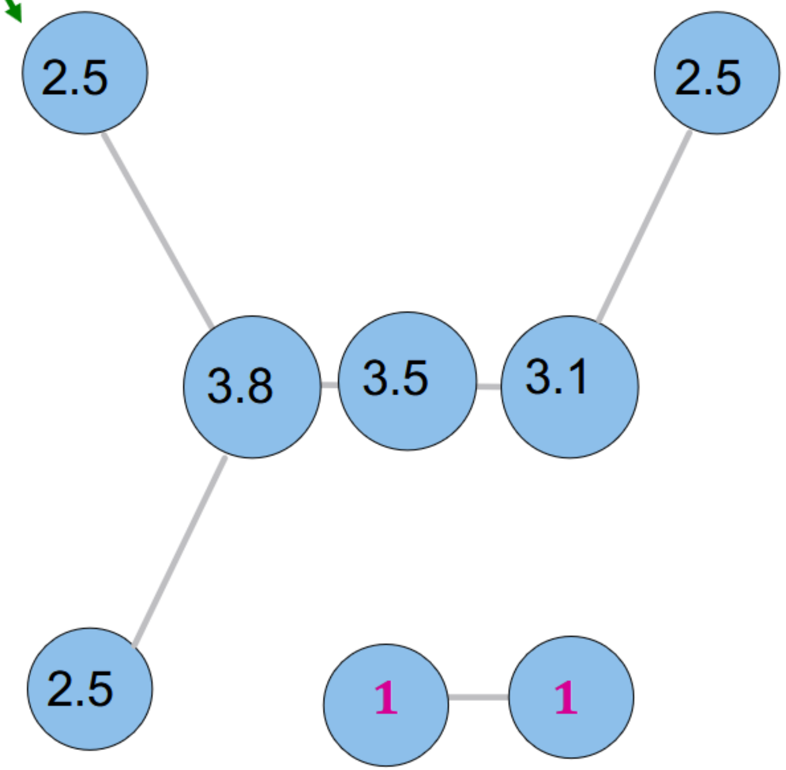
- **Geometric measures**
  - **Harmonic Centrality**:
    - Replace the average distance with the harmonic mean of all distances.
    - The $n(n-1)$ distances between every pair of distinct nodes:

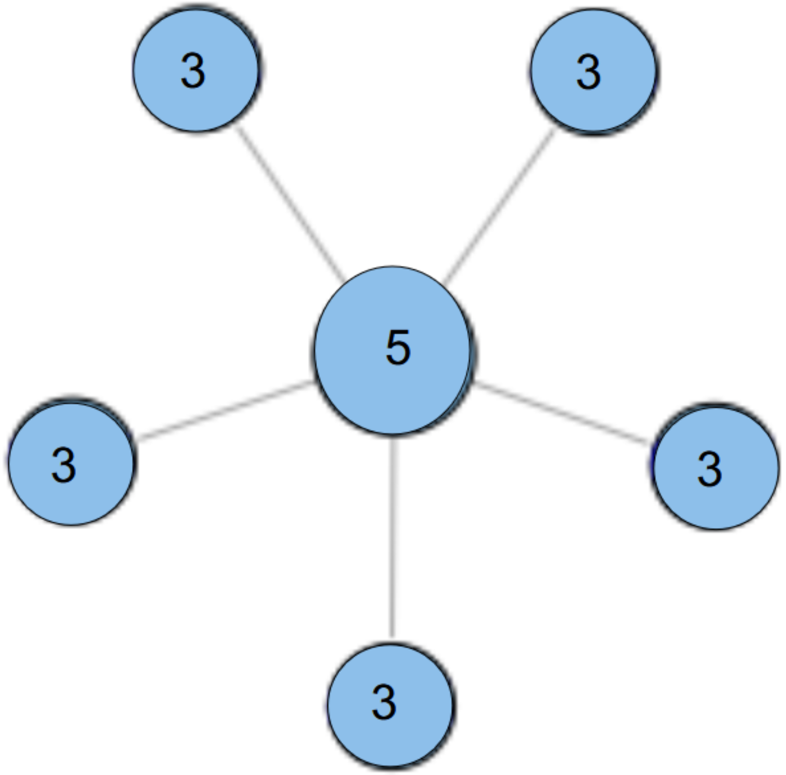
$$c_{\text{har}}(x) = \underbrace{\boxed{\sum_{y \neq x} \frac{1}{d(y,x)}}}_{\text{Harmonic mean}} = \sum_{d(y,x)<\infty, y \neq x} \frac{1}{d(y,x)}$$

    - 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

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

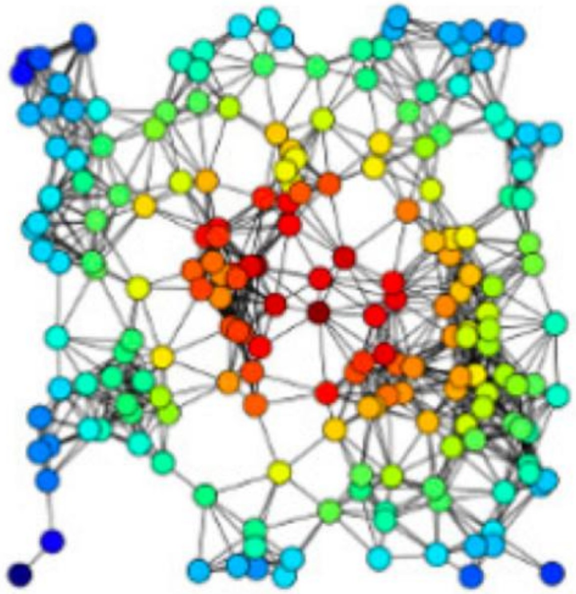




Comparison

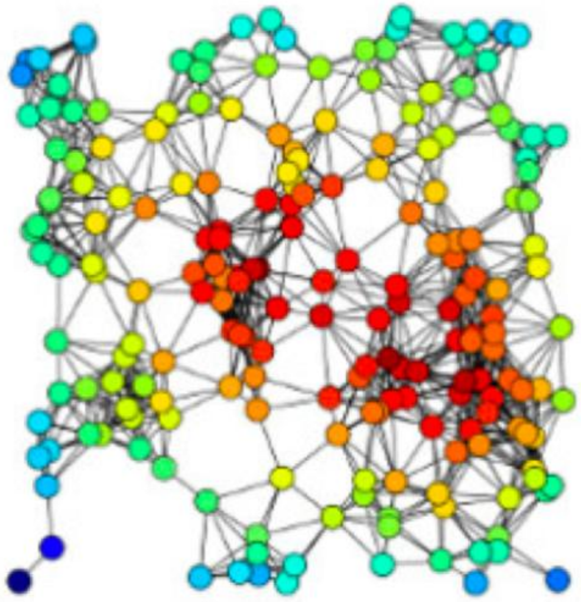
- Closeness Centrality is affected by average distances, while Harmonic Centrality is influenced by nearby nodes.
- A node can be well-positioned (high Closeness) but still have many distant nodes that lower its Harmonic score.
- Harmonic Centrality gives more weight to close neighbors, whereas Closeness considers all distances equally.

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.

Let's Think

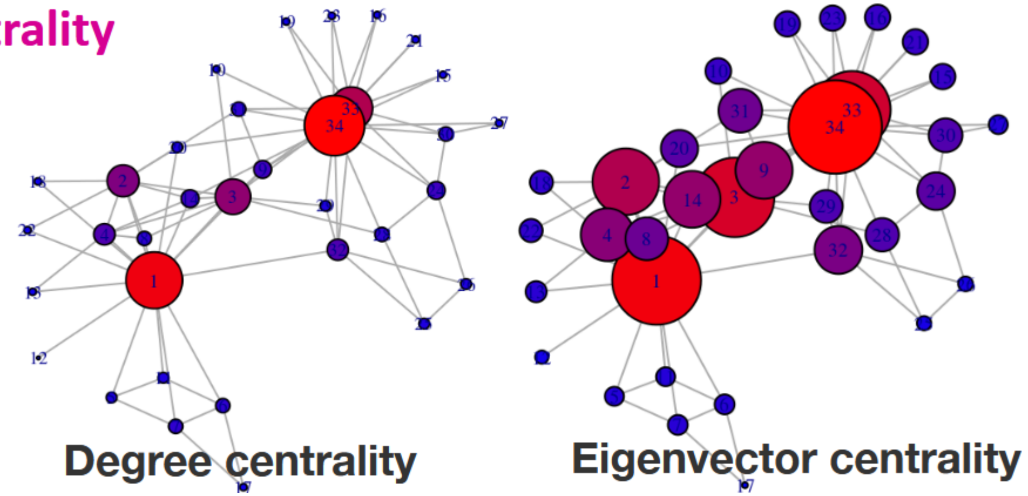
- Can a Node Have High Harmonic Centrality but Low Degree?
  - Imagine you have only two friends, but they are well-connected influencers in the network. Your degree is low (only 2 connections). However, because your friends have strong connections, you can quickly reach many people.
- Can a Node Have High Degree but Low Harmonic Centrality?
  - Imagine a node has 10 direct connections, but all these connections are to each other and not to the rest of the network. Degree is high (10 connections), but reaching other parts of the network requires multiple hops.
- Can a Node Have High Closeness but Low Harmonic Centrality?
  - Consider the tree or ring networks.
- Can a Node Have Low Closeness but High Harmonic Centrality?
  - In a tree structure, a node close to a highly connected hub can still have low Closeness (because reaching deep parts of the tree takes many steps). However, its Harmonic Centrality is high because it can reach nearby nodes very efficiently.

04

Spectral Measures for Centrality

Spectral Measures

- **Spectral measures**
  - Compute the left dominant eigenvector of some matrix derived from the graph
  - **Idea:** A node's centrality is a function of the **centrality of its neighbors**
    - Nodes connected to central nodes has a larger centrality score than those connected to non-central nodes.
    - Eigenvector Centrality
    - Katz's Index
    - Page Rank
    - Hits

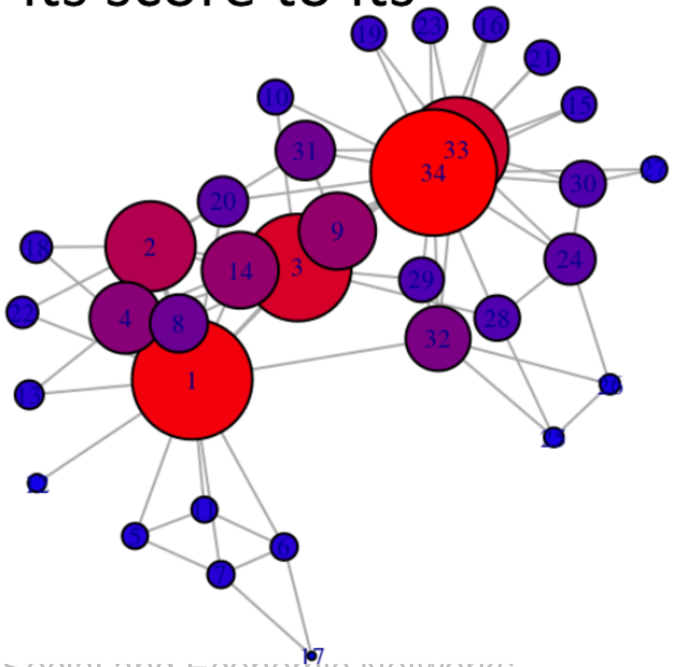


Eigencentrality

- **Spectral measures**
  - **Eigenvector Centrality**: Measure of the **influence** of a node in a network
  - **Idea:** Every node starts with the same score, and then each node gives away its score to its successors

$$c_{\text{eig}}(x) = \frac{1}{\lambda} \sum\nolimits_{y \to x} c_{\text{eig}}(y)$$

Normalization constant = $\|c_{eig}\|_2$



- **Intuitively**: Degree counts walks of length one, the eigenvalue centrality counts walks of length infinity

Eigencentrality

- **Spectral measures**
  - **Eigenvector Centrality**: Measure of the **influence** of a node in a network:

$$c_{\mathrm{eig}}(x) = \frac{1}{\lambda} \sum_{y \to x} c_{\mathrm{eig}}(y)$$

    - c_{eig} converges to the dominant eigenvector of adj. matrix $\boldsymbol{A}$
    - λ converges to the dominant eigenvalue of adj. matrix $\boldsymbol{A}$

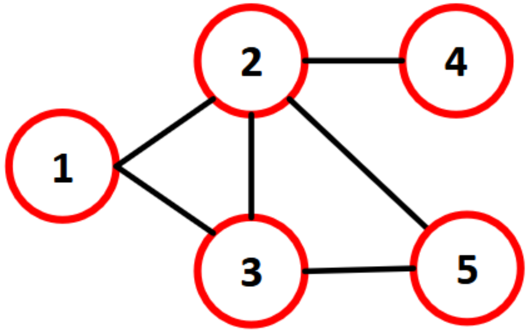
  - Equivalently, eigencentrality is the eigenvector correspond to the dominant eigenvalue (λ) of $\boldsymbol{A}$

$$\boldsymbol{AX} = \lambda \boldsymbol{X}$$

  - **Problem**: Graph should be **strongly connected**!

How to compute Eigencentrality?

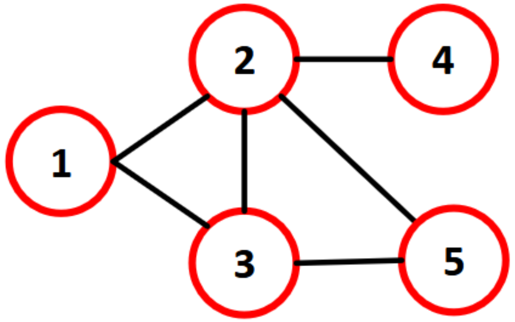
- **Power Iteration:**
  - Set $\boldsymbol{c}^{(0)} \leftarrow 1, k \leftarrow 1$
  - **1:** $\boldsymbol{c}^{(k)} \leftarrow \boldsymbol{A}\boldsymbol{c}^{(k-1)}$
  - **2:** $\boldsymbol{c}^{(k)} = \boldsymbol{c}^{(k)} / \left\|\boldsymbol{c}^{(k)}\right\|_2$
  - **3:** If $\left\|\boldsymbol{c}^{(k)} - \boldsymbol{c}^{(k-1)}\right\| > \varepsilon$:
  - **4:** $\quad k \leftarrow k + 1$, goto **1**



$$A = \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} & \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \end{array} \qquad c = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

How to compute Eigencentrality?

- **Power Iteration:**



Iteration 1

$$\begin{array}{c c} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \end{array} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 3 \\ 1 \\ 2 \end{bmatrix} \equiv \begin{bmatrix} 0.34 \\ 0.68 \\ 0.51 \\ 0.17 \\ 0.34 \end{bmatrix}$$

$$\boldsymbol{A} \qquad \boldsymbol{c}^{(0)} \quad \boldsymbol{c}^{(1)} = \boldsymbol{A}\boldsymbol{c}^{(0)} \quad \boldsymbol{c}^{(1)} = \boldsymbol{c}^{(1)} / \|\boldsymbol{c}^{(1)}\|_2$$

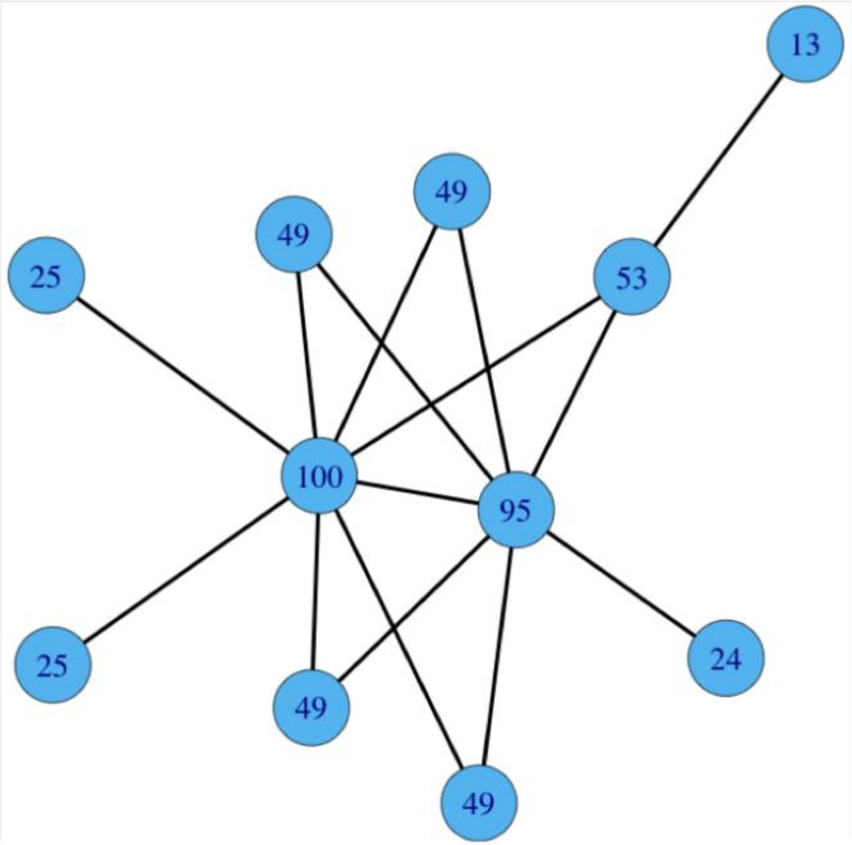
Iteration 2

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.34 \\ 0.68 \\ 0.51 \\ 0.17 \\ 0.34 \end{bmatrix} = \begin{bmatrix} 1.19 \\ 1.36 \\ 1.36 \\ 0.68 \\ 1.19 \end{bmatrix} \equiv \begin{bmatrix} 0.45 \\ 0.51 \\ 0.51 \\ 0.25 \\ 0.45 \end{bmatrix}$$

Iteration 3

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.45 \\ 0.51 \\ 0.51 \\ 0.25 \\ 0.45 \end{bmatrix} = \begin{bmatrix} 1.02 \\ 1.66 \\ 1.41 \\ 0.51 \\ 1.02 \end{bmatrix} \equiv \begin{bmatrix} 0.38 \\ 0.62 \\ 0.53 \\ 0.19 \\ 0.38 \end{bmatrix} \quad \dots \quad c = \begin{bmatrix} 1 \\ 1.41 \\ 1.27 \\ 0.52 \\ 1 \end{bmatrix}$$

Example



Eigenvalue centrality counts walks of length infinity

Katz's Index

- **Spectral measures**
  - **Katz's Index**: Measures **influence** by taking into account the **total number of walks** between a pair of nodes

$$c_{\text{katz}}(x) = \boldsymbol{\beta} \sum_{k=0}^{\infty} \sum_{x \to y} \alpha^k \boxed{\left(A^k\right)_{xy}}$$

Total number of walks of length k between x, y

  - α is an attenuation factor in range $\left(0, \frac{1}{\lambda}\right)$, where λ is the largest eigenvalue of A
  - β is to give some nodes more privilege
  - **Long paths are weighted less than short ones**

Katz's Index

- **Spectral measures**
  - **Katz's Index**: Measures **influence** by taking into account the **total number of walks** between a pair of nodes

$$c_{\text{katz}}(x) = \beta \sum_{k=0}^{\infty} \sum_{x \to y} \boxed{\alpha^k} \boxed{\left(A^k\right)_{xy}}$$

$\alpha < 1$: Long paths are weighted less

Total number of walks of length k between x, y

$$A^0 = \mathbf{I}$$

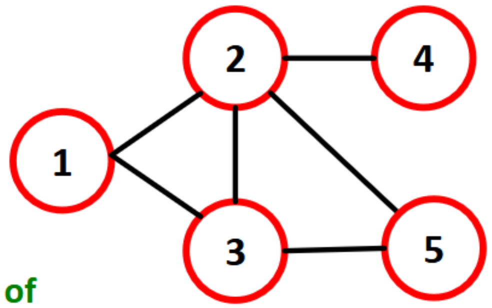
$$A^1 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$A^2 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 2 & 1 & 1 & 1 & 2 \\ 1 & 4 & 2 & 0 & 1 \\ 1 & 2 & 3 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 & 2 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}^3 = \begin{bmatrix} 2 & 6 & 5 & 1 & 2 \\ 6 & 4 & 6 & 4 & \boxed{6} \\ 5 & 6 & 4 & 2 & 5 \\ 1 & 4 & 2 & 0 & 1 \\ 2 & 6 & 5 & 1 & 2 \end{bmatrix}$$

Number of walks of length 3 between 2, 5

(2,1,3,5), (2,4,2,5), (2,3,2,5),
(2,1,2,5), (2,5,3,5), (2,5,2,5)



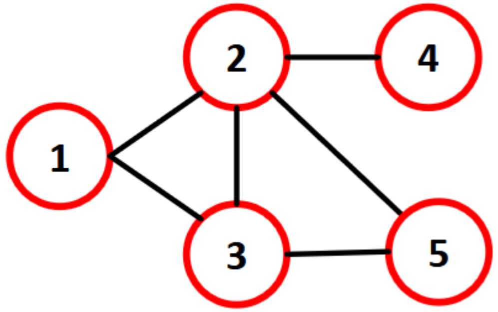
How to compute Katz's Index?

- **Spectral measures**
  - **Katz's Index**: Give each node a small amount of centrality for free

$$c_{\text{katz}}(x) = \alpha \sum_{y \to x} (c_{\text{Katz}}(y) + \beta)$$

Normalization constant

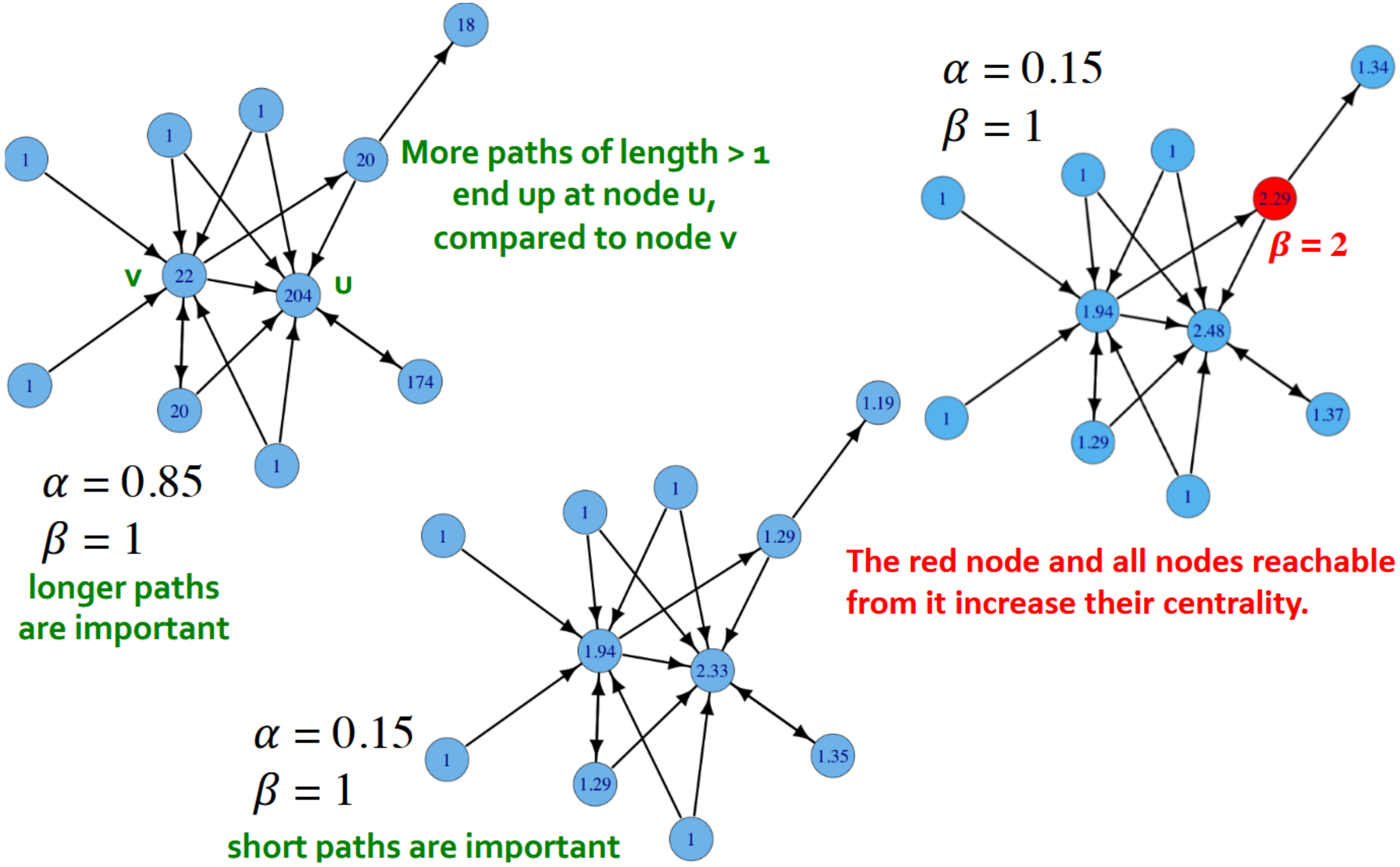
- **Power Iteration:**
  - Set $\boldsymbol{c}^{(0)} \leftarrow 1, k \leftarrow 1$
  - **1:** $\boldsymbol{c}^{(k)} \leftarrow \alpha \boldsymbol{A} \boldsymbol{c}^{(k-1)} + \beta \boldsymbol{1}$
  - **2:** If $\left\| \boldsymbol{c}^{(k)} - \boldsymbol{c}^{(k-1)} \right\| > \varepsilon$:
  - **3:** $k \leftarrow k + 1$, goto **1**



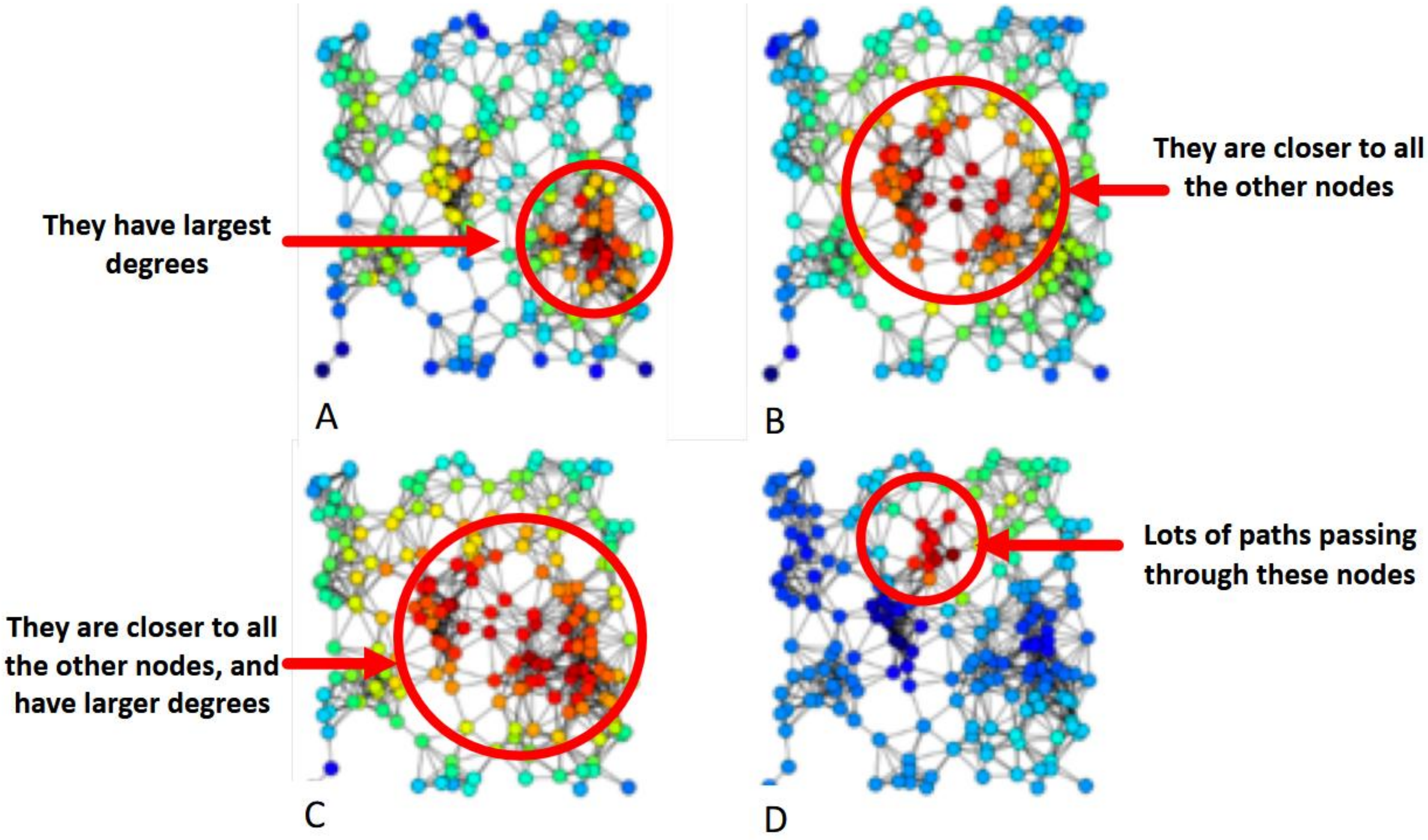
Katz's Index

- **Spectral measures**
  - **Katz's Index**: Suitable for **directed acyclic graphs**
  - **How to choose α?**
    - For **α close to 0**, the contribution given by paths longer than one rapidly declines, and thus
      - Katz scores are mainly **influenced by short paths** (mostly in-degrees)
    - When the **α is large**, long paths are devalued smoothly, and
      - Katz scores are more **influenced the by topology** of the network
    - The measure diverges at $\alpha > \frac{1}{\lambda}$
    - The dominant eigenvector of A is the limit of Katz centrality as α approaches $1/\lambda$ from below

Example


18
1
1
1
20
22
204
174
20
1
1
v
u
More paths of length > 1
end up at node u,
compared to node v
$\alpha = 0.85$
$\beta = 1$
longer paths
are important
$\alpha = 0.15$
$\beta = 1$
1.34
1
1
1
2.29
$\beta = 2$
1.94
2.48
1
1.29
1.37
1
1.19
1
1
1
1.29
1.94
2.33
1
1.29
1.35
1
$\alpha = 0.15$
$\beta = 1$
short paths are important
The red node and all nodes reachable
from it increase their centrality.

Example



Examples of A) Degree centrality, B) Closeness centrality, C) Harmonic Centrality and D) Katz centrality of the same graph.

An Interesting Comparison!

- Comparing three centrality values
  - Generally, the 3 centrality types will be positively correlated
  - When they are not (or low correlation), it usually reveals interesting information

	Low Degree	**Low Closeness**	**Low Betweenness**
High Degree		*Node is embedded in a community that is far from the rest of the network*	*Ego's connections are redundant - communication bypasses the node*
High Closeness	*Key node connected to 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! Ego monopolizes the ties from a small number of people to many others.*	

This slide is modified from a slide developed by James Moody

Any Question?