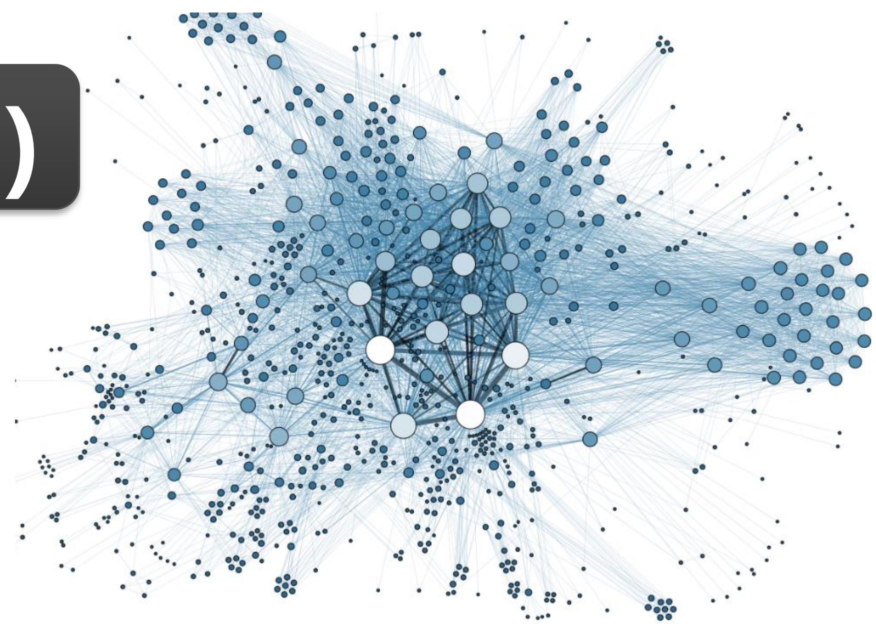
Social Media Analytics

INFS7450

Node Measures (I)



Prof. Hongzhi Yin

School of EECS

The University of Queensland

Mandatory training: Respect at UQ

– UQ has introduced a mandatory **Respect at UQ** training module for all students to complete by **5pm on Sunday 10 March 2024.**

– Your access to Blackboard will be restricted if you do not complete the module by the deadline.

– A link to the module and more information is available at respect.uq.edu.au/respect-at-uq.

We will focus on how to compute various node centrality measures efficiently.

But we will not delve into mathematical proofs about these computation methods or algorithms.

Why Do We Need Centrality Measures?

- Who are the important/central figures (influential individuals) in the network?
  - **Centrality**
- To answer the question, one first needs to define measures for quantifying **centrality of nodes**

Klout Influence Score

Citation Score

Hongzhi Yin

FOLLOW

Professor and ARC Future Fellow, University of Queensland
Verified email at uq.edu.au - Homepage

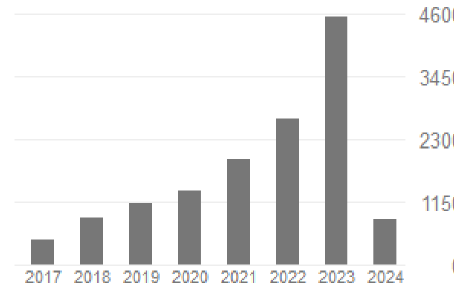
recommender system graph learning trustworthy intelligence decentralized intelligence

TITLE	CITED BY	YEAR
Lcars: a location-content-aware recommender system H Yin, Y Sun, B Cui, Z Hu, L Chen Proceedings of the 19th ACM SIGKDD international conference on Knowledge ...	447	2013
Learning Graph-based POI Embedding for Location-based Recommendation M Xie, H Yin*, F Xu, H Wang, W Chen, S Wang The 25th ACM International Conference on Information and Knowledge ...	433	2016
Self-supervised hypergraph convolutional networks for session-based recommendation X Xia, H Yin, J Yu, Q Wang, L Cui, X Zhang Proceedings of the AAAI conference on artificial intelligence 35 (5), 4503-4511	375	2021
Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection T Chen, X Li, H Yin, J Zhang Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2018 ...	354	2018
Challenging the long tail recommendation H Yin, B Cui, J Li, J Yao, C Chen arXiv preprint arXiv:1205.6700	335	2012
Rethinking the item order in session-based recommendation with graph neural networks R Qiu, J Li, Z Huang, H Yin Proceedings of the 28th ACM international conference on information and ...	327	2019

GET MY OWN PROFILE

Cited by VIEW ALL

	All	Since 2019
Citations	14537	12606
h-index	70	63
i10-index	191	188



Public access VIEW ALL

8 articles — 188 articles

not available — available

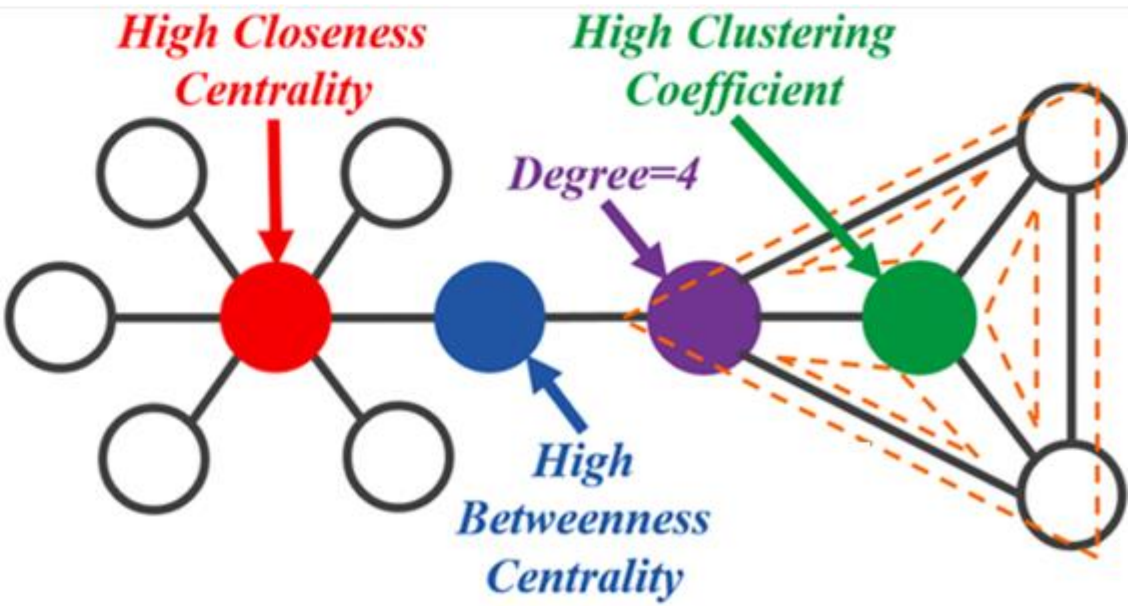
Based on funding mandates

Node Centrality

Centrality defines how important a node is within a network

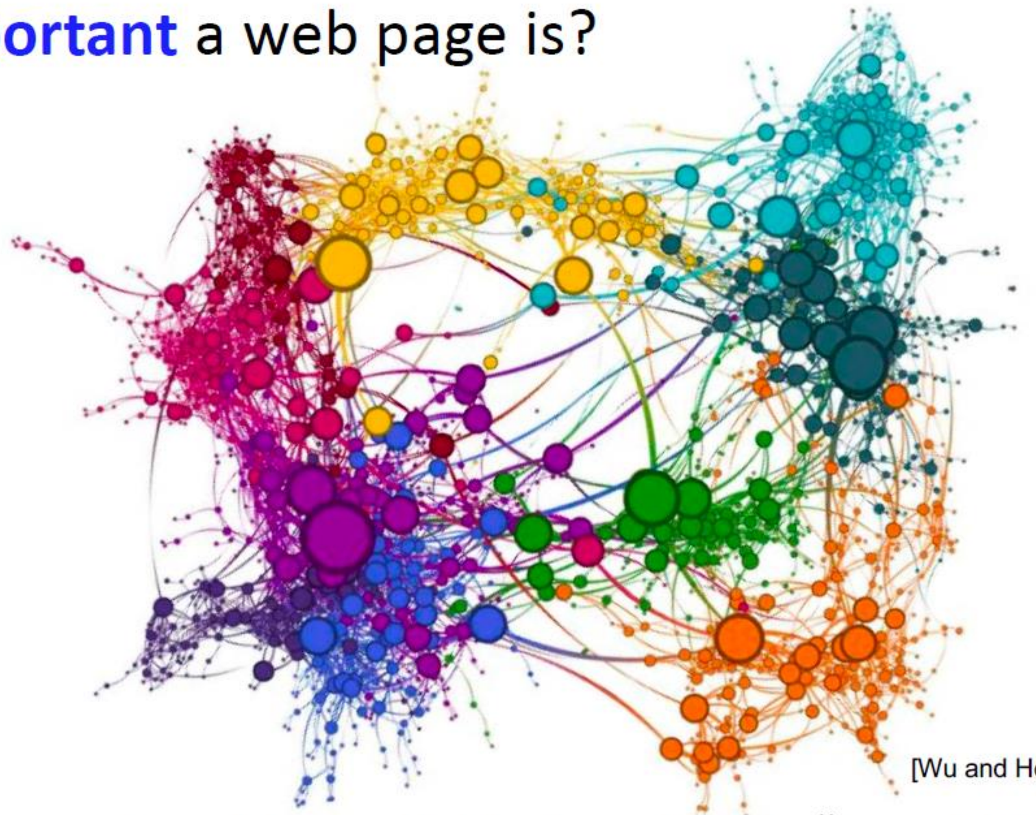
Network Centrality

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



Network Centrality

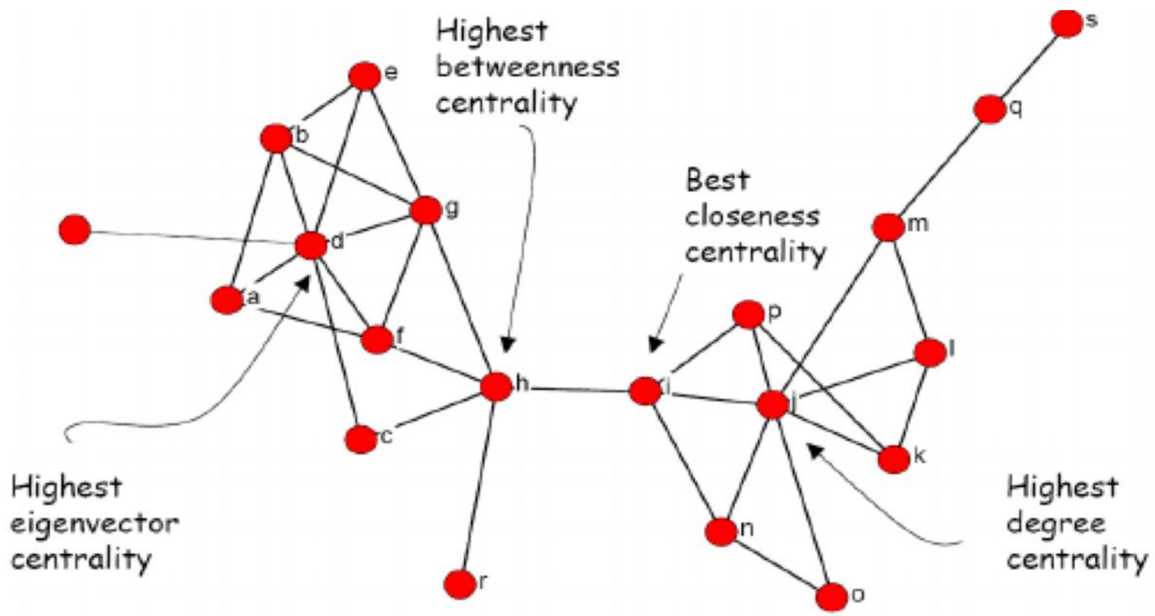
- **Centrality** is used often for detecting:
  - How **influential** a person is in a social network?
  - How **well used** a road is in a transportation network?
  - How **important** a web page is?



[Wu and He'15]

Centrality Measures

- **Geometric Measures**:
  - Importance of a node is a **function of distances** to others.
- **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



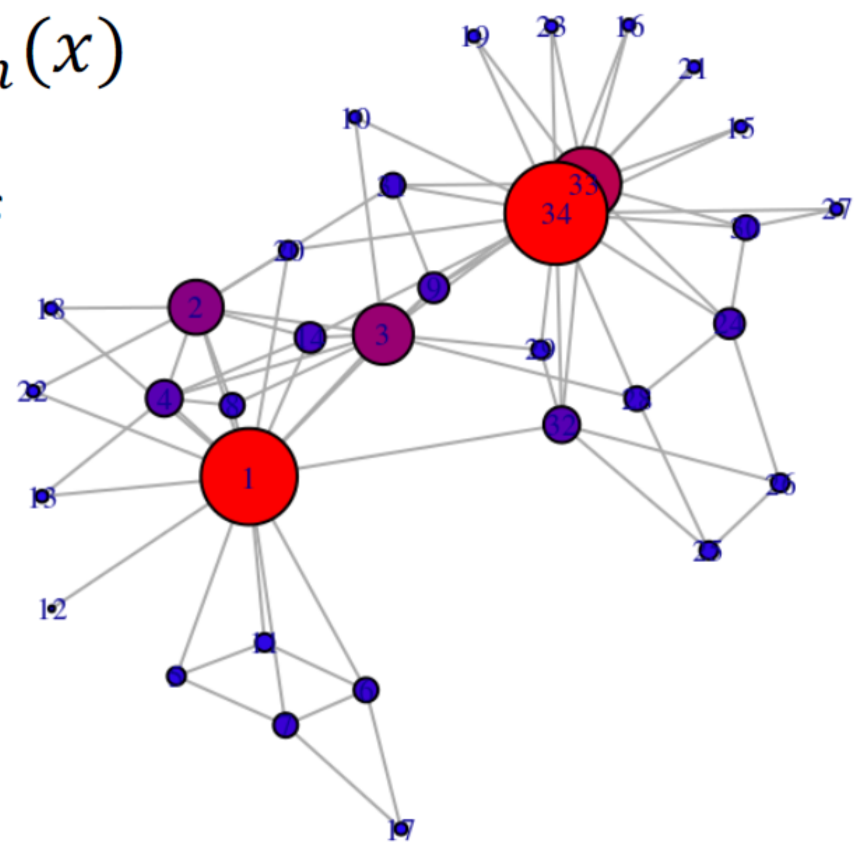
Geometric Centrality Measures

- (In)Degree Centrality
- Closeness Centrality
- Harmonic Centrality

In-degree Centrality

- **Geometric measures**
  - **(In-)Degree Centrality**: The number of incoming links

$$c_{\deg}(x) = d_{in}(x)$$



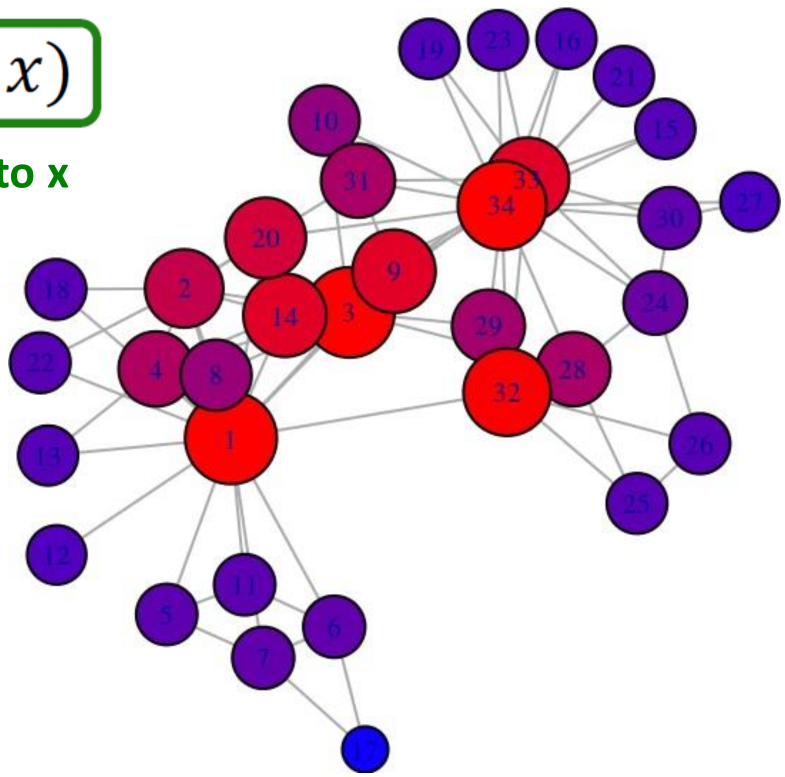
  - Or equivalently, number of nodes at distance one
  - Equivalent to majority voting

Closeness Centrality

- **Geometric measures**
  - **Closeness Centrality**
  - Nodes that are more central have smaller distances

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

length of the shortest path from y to x

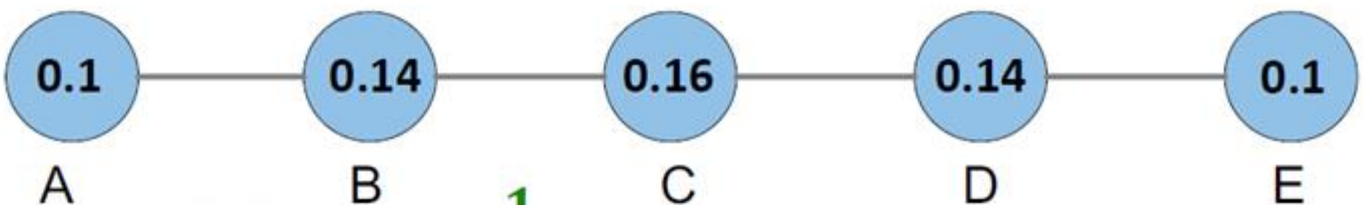


- Nodes that are more central have smaller distances to other nodes, and higher centrality

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

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

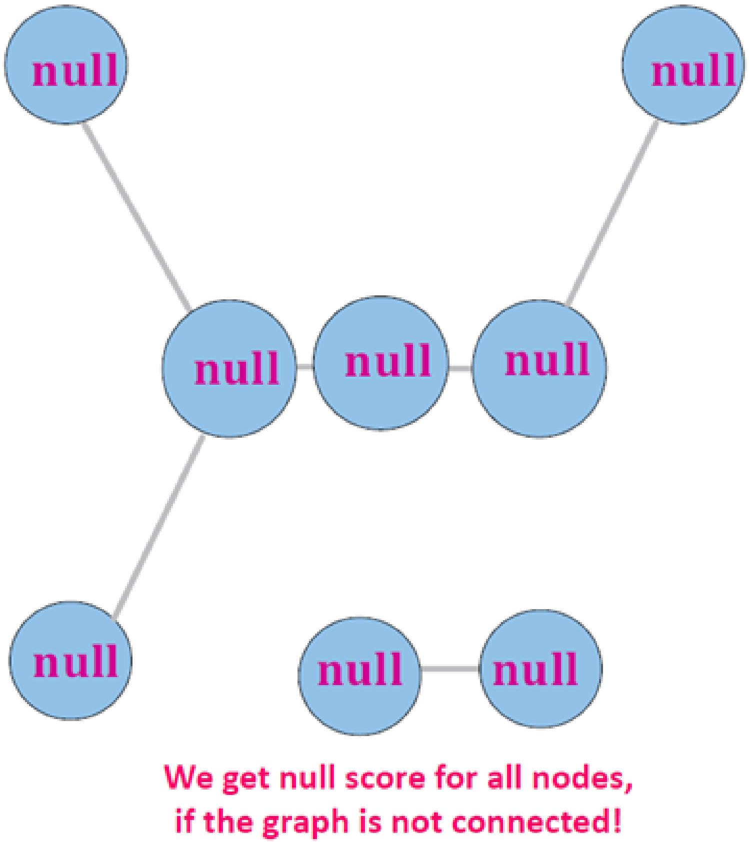
length of the shortest path from x to y

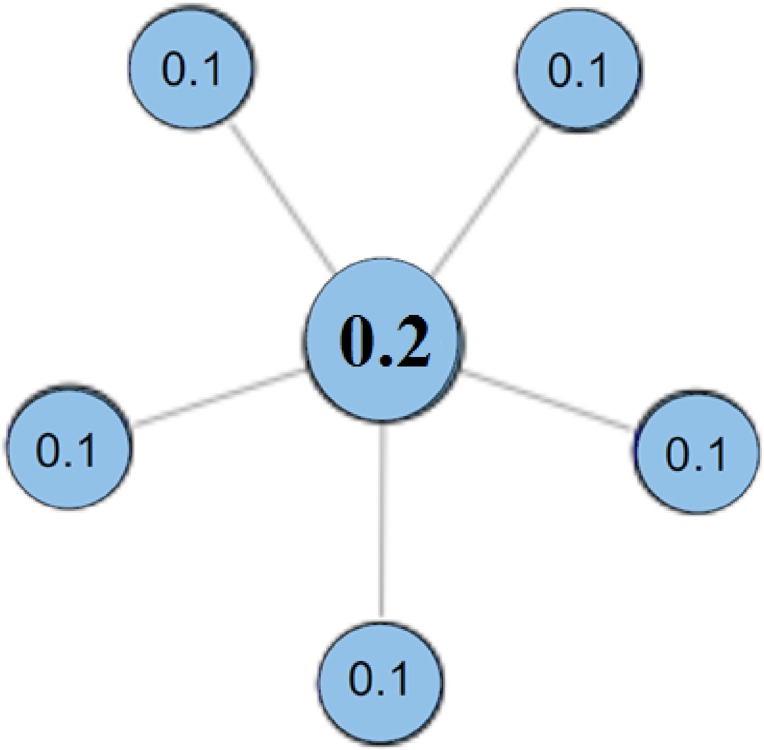


$$c_{clos}(A) = \frac{1}{1+2+3+4} = 0.1$$

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

Closeness Centrality: Example





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

For any pair of nodes from two components, their distance is **infinite**.

Harmonic Centrality

Rather than summing the distances of a node to all other nodes, the harmonic centrality algorithm sums the inverse of those distances. This enables it to deal with infinite values.

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

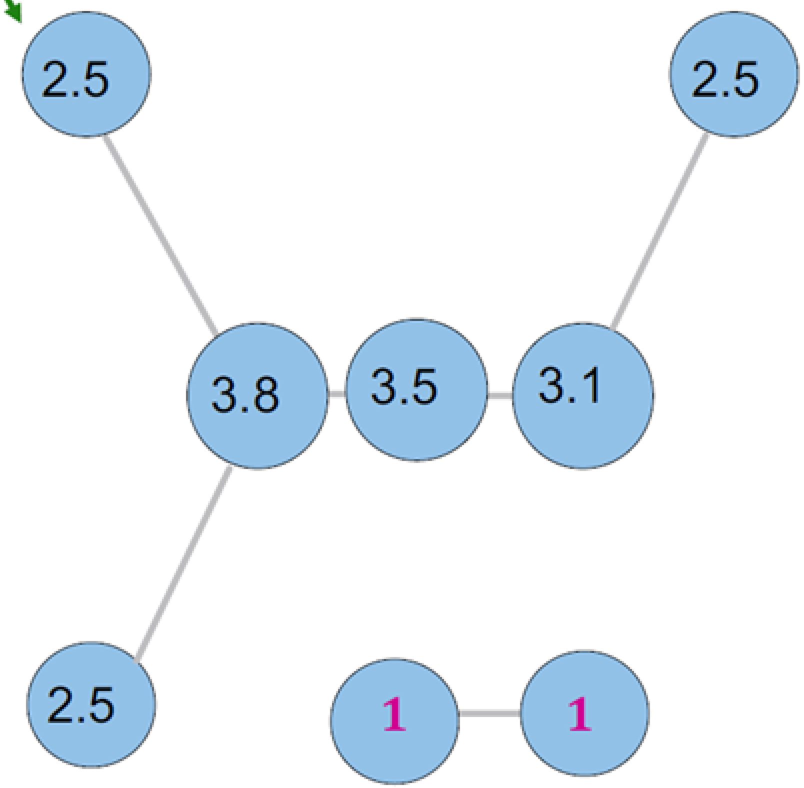
$$c_{\text{har}}(x) = \frac{1}{\text{n-1}} \boxed{\sum_{y \neq x} \frac{1}{d(y,x)}} = \frac{1}{\text{n-1}} \sum_{d(y,x)<\infty, y \neq x} \frac{1}{d(y,x)}$$

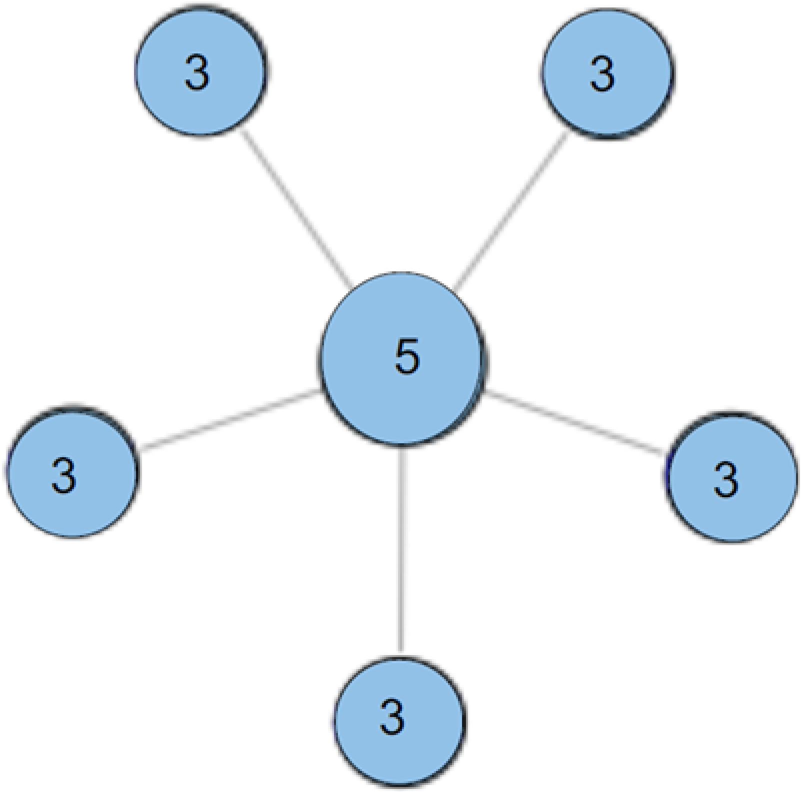
normalized version

- 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

$$c_{harm} = \frac{1}{1} + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \mathbf{2.58}$$



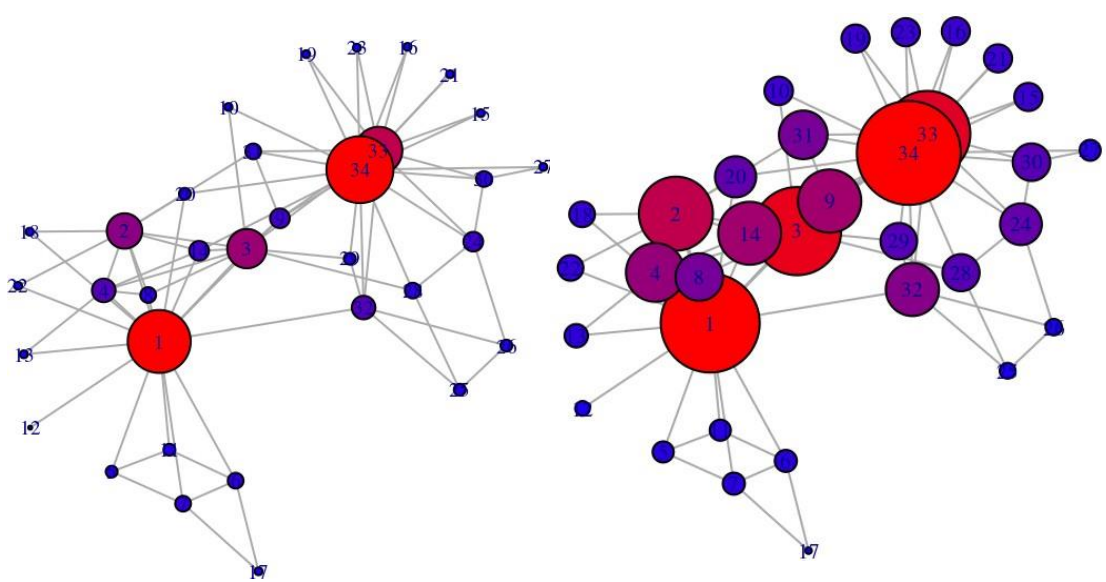


Spectral Centrality Measures

- Eigenvector Centrality
- Kat’s Index
- PageRank
- Hits

Spectral Centrality

- **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 (RECURSIVE DEFINITION)**
    - 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**



Eigenvector Centrality

- Having more friends does not by itself guarantee that someone is more important
  - Having more **important friends** provides a stronger signal

- Eigenvector centrality generalizes degree centrality by incorporating the importance of the neighbors (undirected)

- For directed graphs, we can use incoming edges

Formulation

- Let's assume the eigenvector centrality of a node is $c_e(v_i)$ (unknown) $\mathbf{C}_e = (C_e(v_1), C_e(v_2), \ldots, C_e(v_n))^T$
- We would like $c_e(v_i)$ to be higher when important neighbors (node v_j with higher $c_e(v_j)$) point to it
  - Incoming neighbors
  - For incoming neighbors $A_{j,i} = 1$
- Idea: each node starts with the same score, and then each node gives away its score to its successors

$$c_e(v_i) = \sum_{j=1}^{n} A_{j,i} c_e(v_j)$$

- Is this summation bounded?
  - We have to normalize! $c_e(v_i) = \frac{1}{\lambda} \sum_{j=1}^{n} A_{j,i} c_e(v_j)$

λ is the norm of the centrality vector of all nodes

Eigenvector Centrality (Matrix Formulation)

- Let $\mathbf{C}_e = (C_e(v_1), C_e(v_2), \ldots, C_e(v_n))^T$

$$\rightarrow \lambda \mathbf{C}_e = A^T \mathbf{C}_e$$ Characteristic equation

- This means that $\boldsymbol{C_e}$ is an eigenvector of adjacency matrix A^T (or A when undirected) and λ is the corresponding eigenvalue
- Which eigenvalue-eigenvector pair should we choose?
  - Eigenvector centrality is a recursive definition.
  - $\boldsymbol{C_e}$ converges to the dominant eigenvector of adj. matrix A^T
  - λ converges to the dominant eigenvalue of adj. matrix A^T

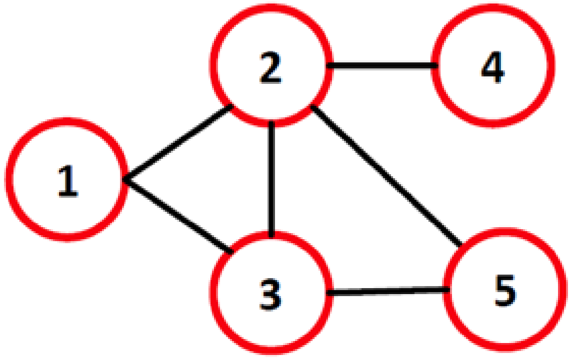
Eigen-Centrality: How to Compute

- How to compute eigenvector centrality?
  1. We compute the eigenvalues of A
  2. Select the largest eigenvalue λ
  3. And the corresponding eigenvector of λ is $\mathbf{C_e}$.

It is time-consuming to directly compute eigenvalues and eigenvectors and this method cannot apply to large-scale networks.

Eigen-Centrality: How to Compute

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



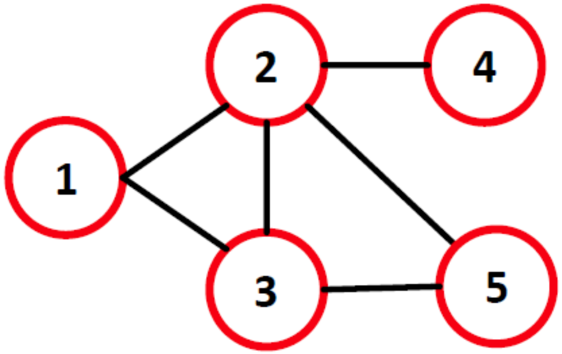
$$||x||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

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

$$c = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Eigen-Centrality: How to Compute

- **Power Iteration:**



Iteration 1

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

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

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

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

Katz Centrality

- A major problem with eigenvector centrality arises when it deals with directed graphs
- For nodes without incoming edges in a directed graph, their centrality values are zero
- Eigenvector centrality can only consider the effect of network topology structure and cannot capture the external knowledge.

- To resolve this problem we add bias term β to the centrality values for all nodes

Similar to Eigenvector Centrality

$$C_{\mathrm{Katz}}(v_i) = \boxed{\alpha \sum_{j=1}^{n} A_{j,i} C_{\mathrm{Katz}}(v_j)} + \beta$$

Katz Centrality, cont.

$$C_{\text{Katz}}(v_i) = \alpha \sum_{j=1}^{\text{n}} A_{j,i} C_{\text{Katz}}(v_j) + \beta$$

Controlling term

Bias term

Rewriting equation in a vector form

$$\mathbf{C}_{\text{Katz}} = \alpha A^T \mathbf{C}_{\text{Katz}} + \beta \mathbf{1}$$

vector of all 1's

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Katz centrality: $\mathbf{C}_{\text{Katz}} = \beta(\mathbf{I} - \alpha A^T)^{-1} \cdot \mathbf{1}$

The time complexity of matrix inversion operation is $O(n^3)$

Identity Matrix

- An identity matrix is a **square matrix** in which all the elements of **the principal diagonal** are ones and all other elements are zeros.
- The **effect of multiplying a given matrix by an identity matrix** is to leave the given matrix unchanged.

$$
\begin{aligned}
M \times I &= \begin{bmatrix} -4 & -3 \\ -6 & 5 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} -4\times1+-3\times0 & -4\times0+-3\times1 \\ -6\times1+5\times0 & -6\times0+5\times1 \end{bmatrix} \\
&= \begin{bmatrix} -4 & -3 \\ -6 & 5 \end{bmatrix}
\end{aligned}
$$

How to select controlling term

Since we are inverting a matrix here, not all α values are acceptable.

When $\alpha = 0$, the eigenvector centrality part is removed, and all nodes get the same centrality value β.

when $\det(\mathbf{I} - \alpha A^T) = 0$, the matrix $\mathbf{I} - \alpha A^T$ becomes non-invertible and the centrality values diverge.

The $\det(\mathbf{I} - \alpha A^T)$ becomes 0 when $\alpha = 1/\lambda$, where λ is an eigenvalue of A^T.

In this case, αA^T is an identity matrix.

$$\lambda \mathbf{C} = A^T \mathbf{C} \Rightarrow \mathbf{C} = \boxed{1/\lambda\, A^T} \mathbf{C}$$

How to select controlling term

The $\det(\mathbf{I} - \alpha A^T)$ becomes 0 when $\alpha = 1/\lambda$, where λ is an eigenvalue of A^T.

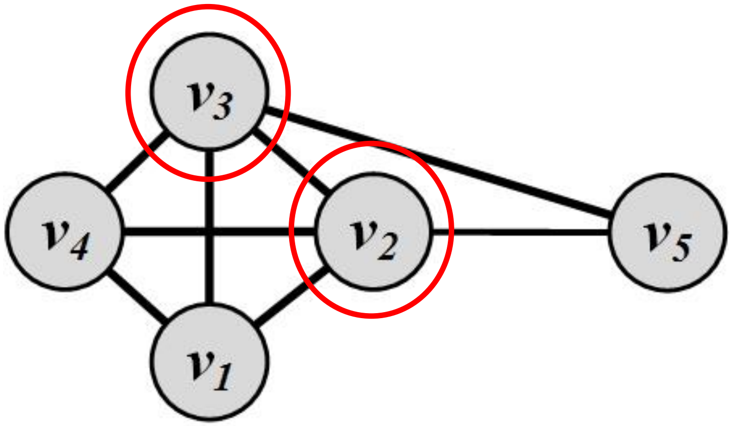
In this case, αA^T is an identity matrix.

$$\lambda \mathbf{C} = A^T \mathbf{C} \Rightarrow \mathbf{C} = 1/\lambda \, A^T \mathbf{C}$$

The $\det(\mathbf{I} - \alpha A^T)$ first becomes 0 when $\alpha = 1/\lambda$, where λ is the largest eigenvalue of A^T.

$\boxed{\alpha < 1/\lambda \text{ is selected so that the matrix is invertible}}$

Katz Centrality Example



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

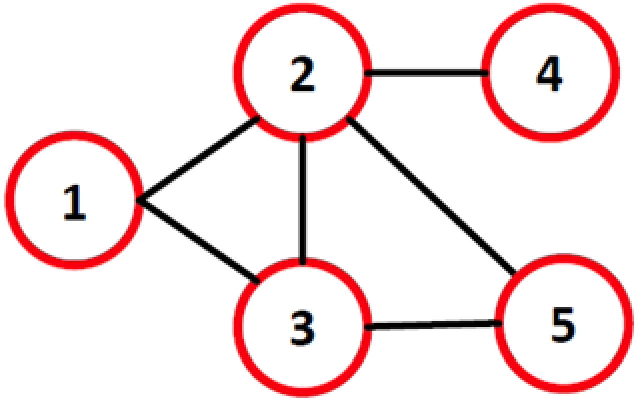
- The Eigenvalues are -1.68, -1.0, -1.0, 0.35, $\boxed{3.32}$
- We assume α=0.25 < $\boxed{1/3.32}$ and $\beta = 0.2$

$$\mathbf{C}_{Katz} = \beta(\mathbf{I} - \alpha A^T)^{-1} \cdot \mathbf{1} = \begin{bmatrix} 1.14 \\ 1.31 \\ 1.31 \\ 1.14 \\ 0.85 \end{bmatrix}$$

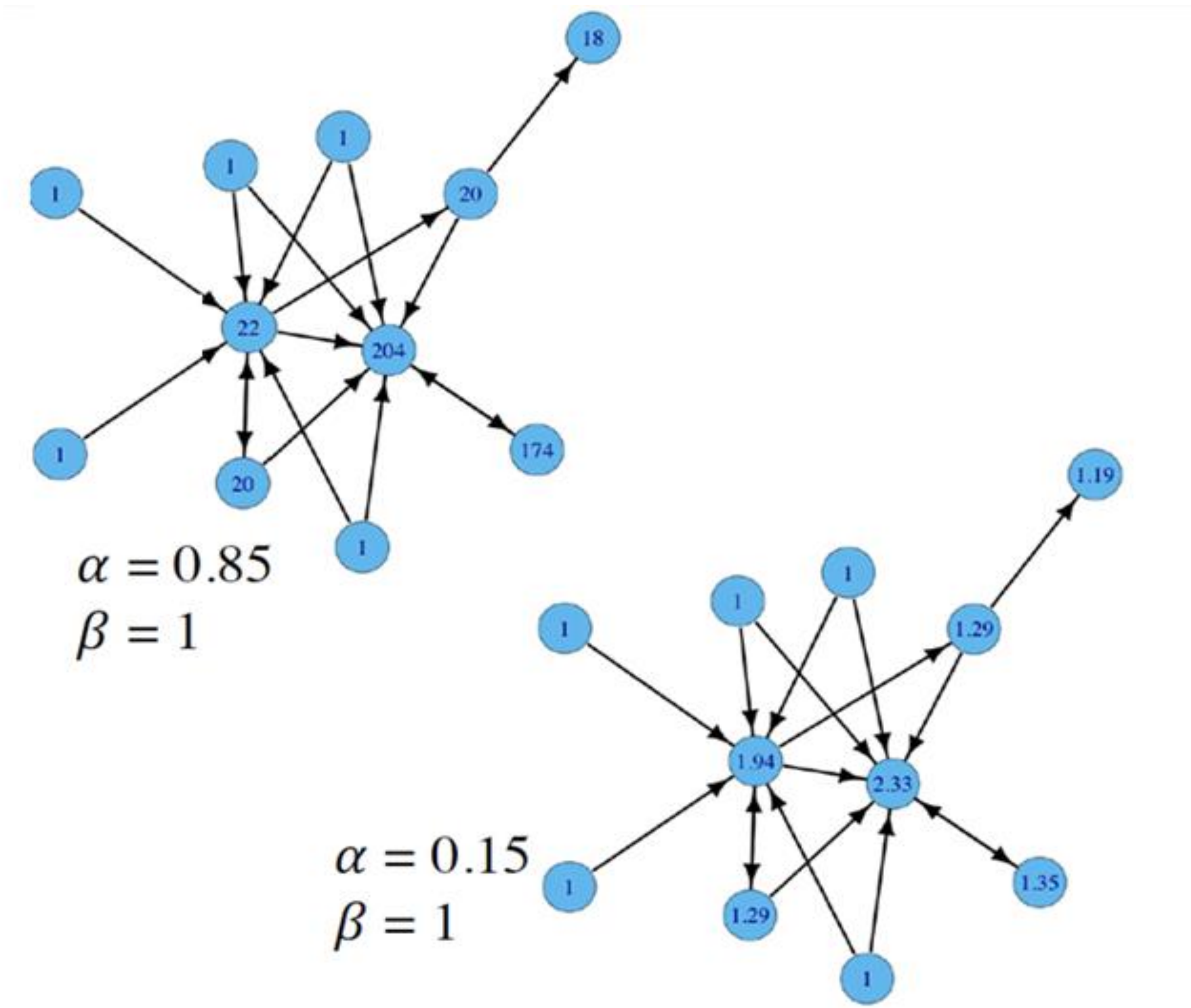
Most important nodes!

How to efficiently compute Katz Centrality

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



Katz Centrality Example



References

- R. Zafarani, M. A. Abbasi, and H. Liu, Social Media Mining: An Introduction, Cambridge University Press, 2014.
- http://socialmediamining.info/
- Stanford CS224W Analysis of Networks