

Complex Network Theory

Lecture 3

Network centrality measures

Instructor: S. Mehdi Vahidipour
(Vahidipour@kashanu.ac.ir)

Thanks

A. Rezvani

A. Barabasi, L. Adamic and J. Leskovec

Outline

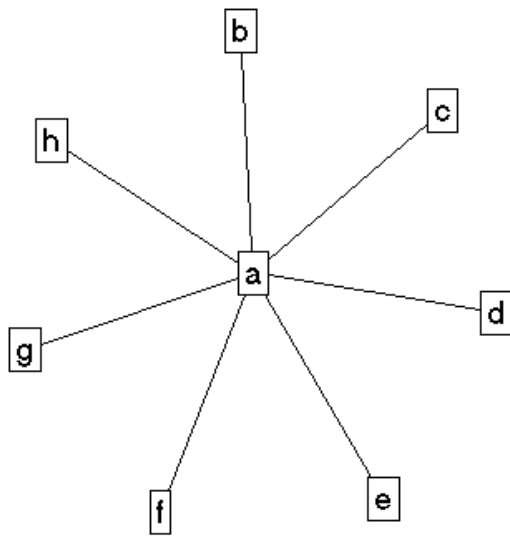
- Overview of class topics
 - Importance of nodes and links
 - Network centrality measures
 - Degree centrality
 - Closeness centrality
 - Betweenness centrality
 - Reach centrality
 - PageRank centrality
 - Vulnerability
 - Network entropy
- Next class:
 - Network analysis

Centrality in Social Networks

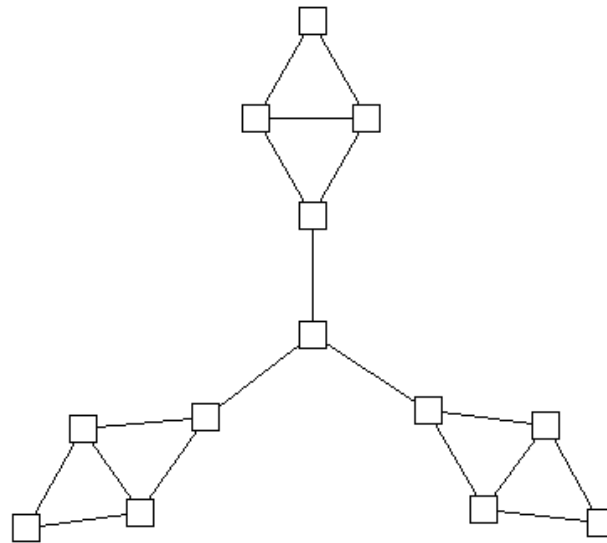
- Background: At the individual level, one dimension of position in the network can be captured through centrality.
- Conceptually, centrality is fairly straight forward: we want to identify which nodes are in the 'center' (**important**) of the network. In practice, identifying exactly what we mean by 'center' is somewhat complicated.
- Approaches:
 - Degree
 - Closeness
 - Betweenness
- Graph level measures: Centralization
- Central components may play critical role in network functions
 - Robustness
 - Collective behavior
 - Information spreading
 - Synchronization
 - Social dynamics

Central nodes

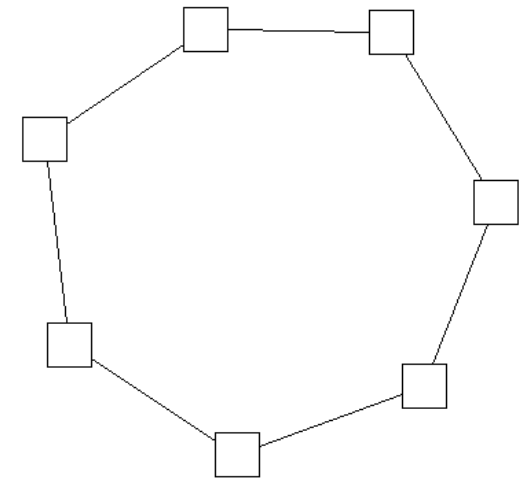
Intuitively, we want a method that allows us to distinguish “important” nodes (users, actors). Consider the following graphs:



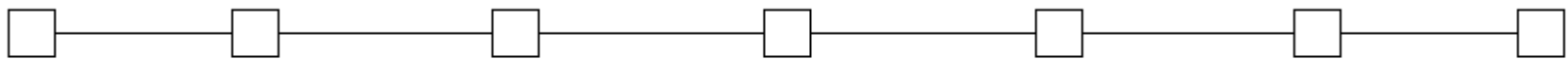
Star



Modular



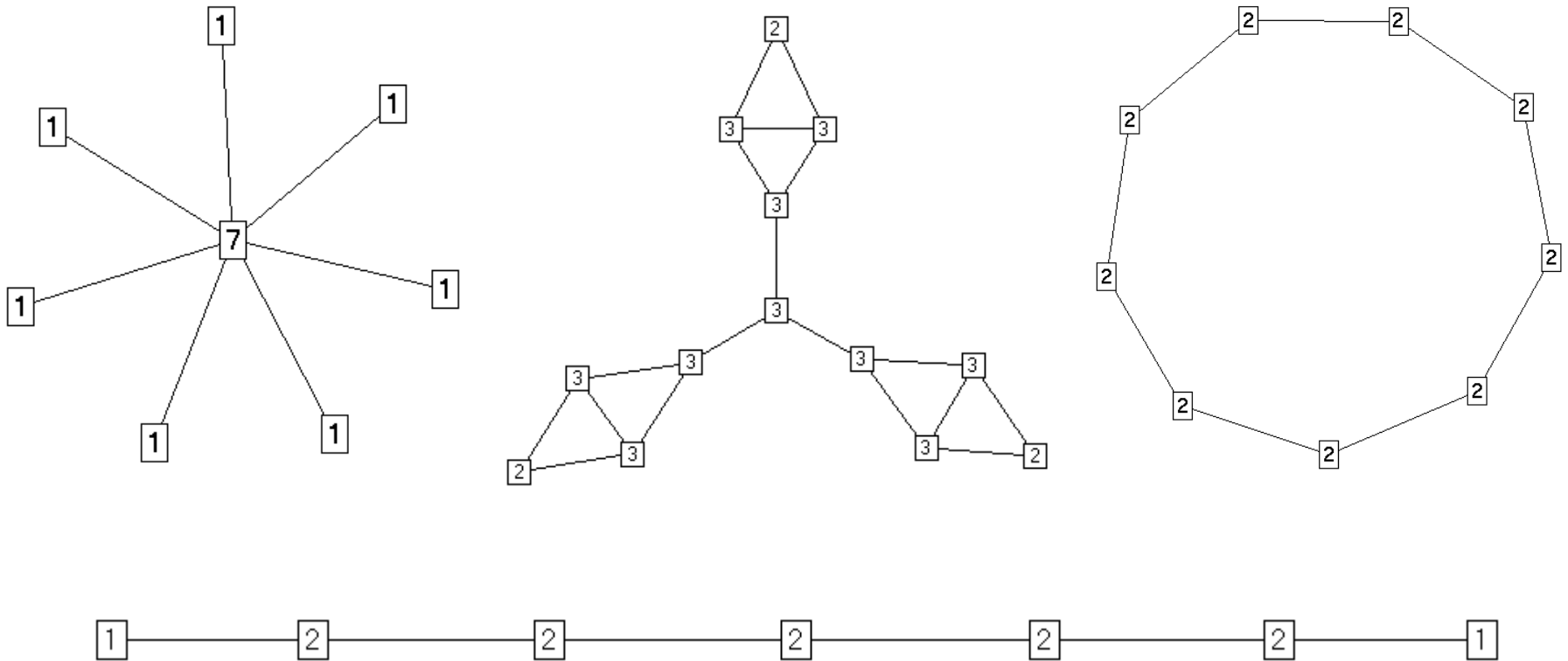
Circle



Line

Degree based centrality

The most intuitive notion of centrality focuses on degree: The node with the most ties is the most important:

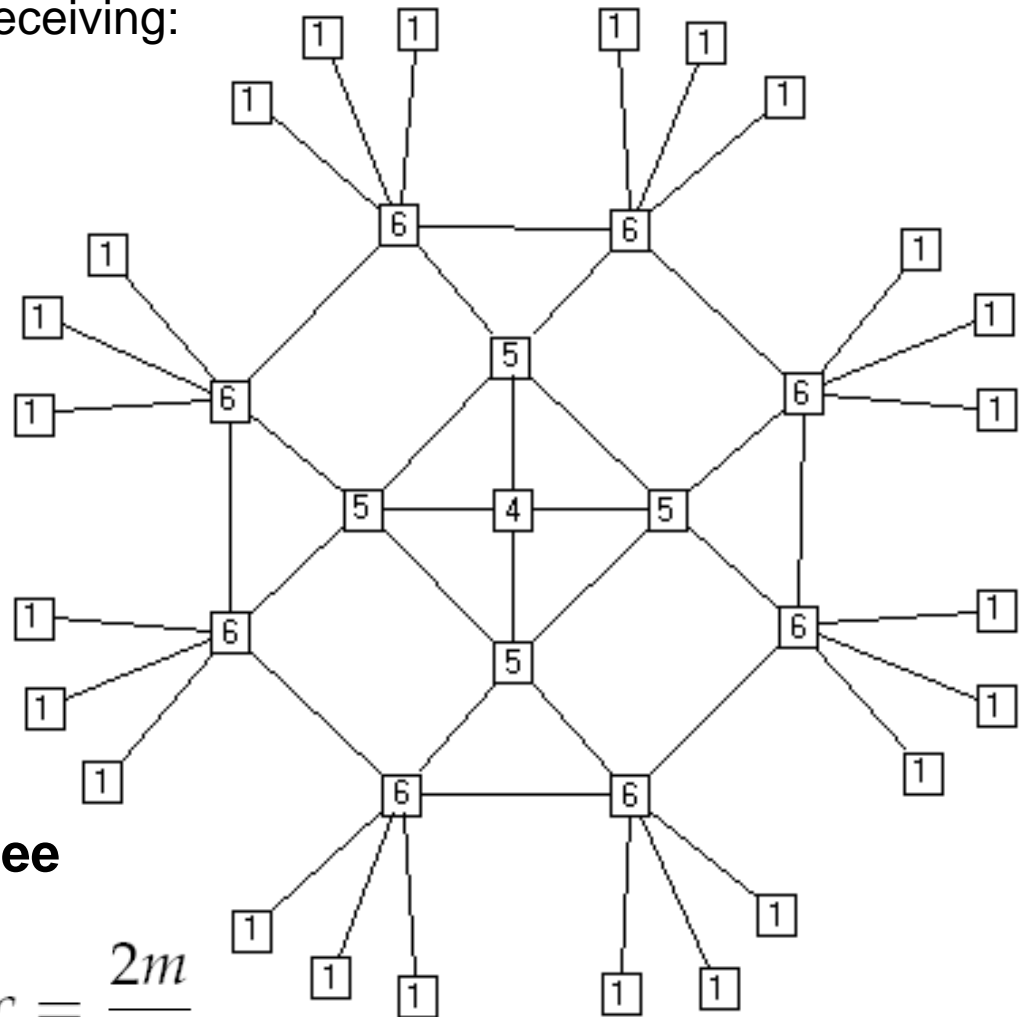


$$C_D(v_i) = d(v_i) = k_i = \sum_j a_{ij}$$

Degree based centrality

Degree centrality, however, can be deceiving:

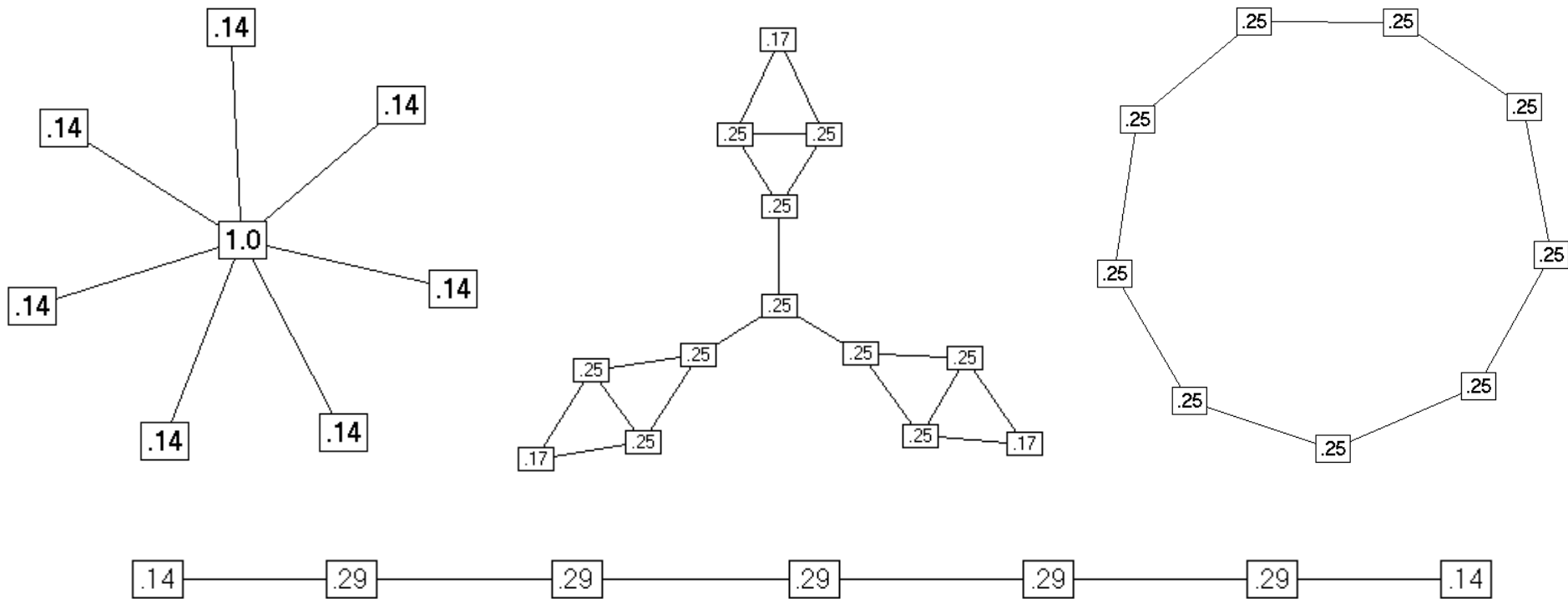
- In what ways does degree fail to capture centrality in the following graphs
 - ability to broker between groups
 - likelihood that information originating anywhere in the network reaches you



Other measures based on degree

- Max degree: k_{\max}
- Mean (average) degree: $\langle k \rangle$, $c = \frac{2m}{n}$
- Degree Distribution: $P(k)$
 - $P_{\text{out}}(k)$, $P_{\text{in}}(k)$

One often standardizes the degree distribution, by the maximum possible (n-1):



$$\overline{C_D}(i) = \overline{d}(v_i) = \overline{k}_i = \frac{\sum_j a_{ij}}{n-1} = \frac{k_i}{n-1}$$

Centralization (skew in distribution)

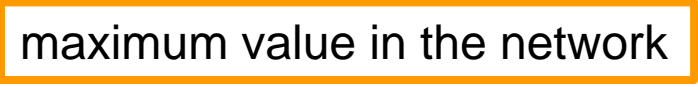
How much variation is there in the centrality scores among the nodes?

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

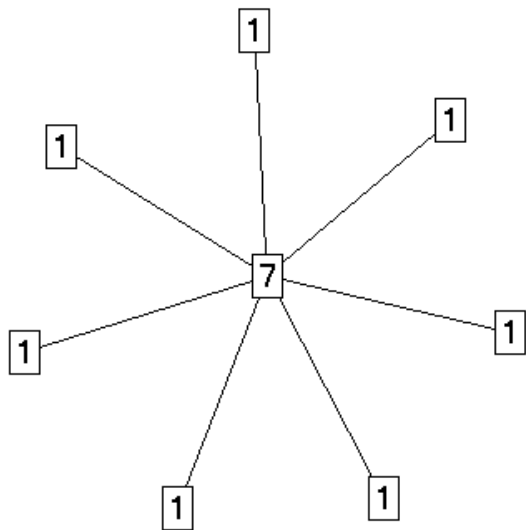
Simple: **variance** of the individual centrality scores.

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

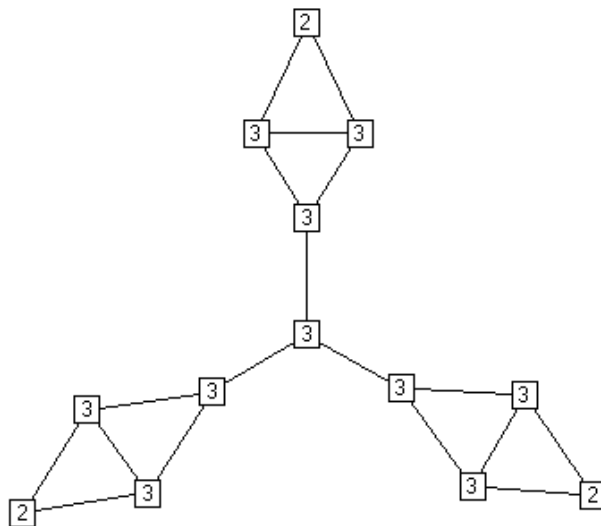
Or, using Freeman's general formula for **centralization**:

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


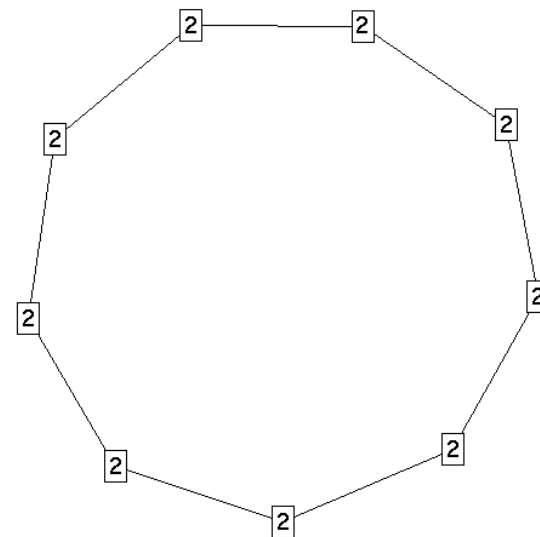
Degree Centralization Scores



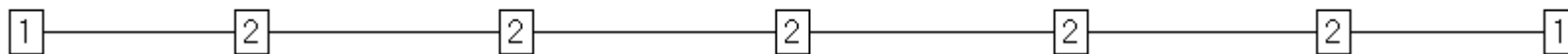
Freeman: 1.0
Variance: 3.9



Freeman: .02
Variance: .17

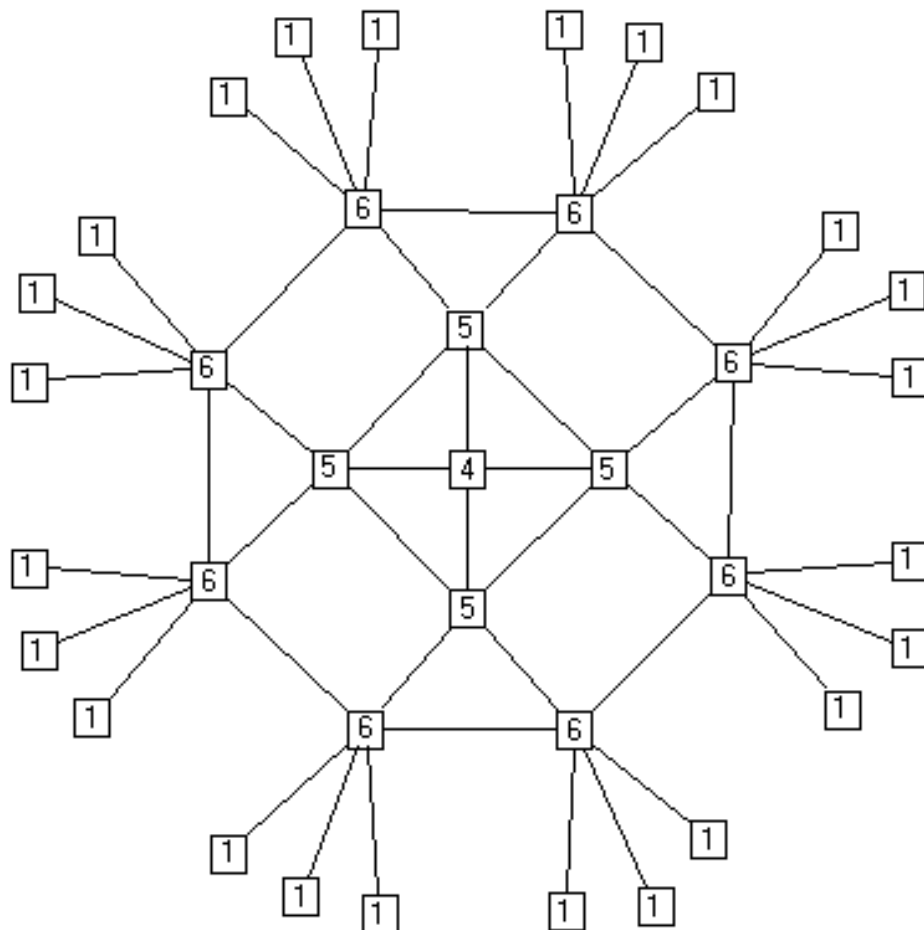


Freeman: 0.0
Variance: 0.0



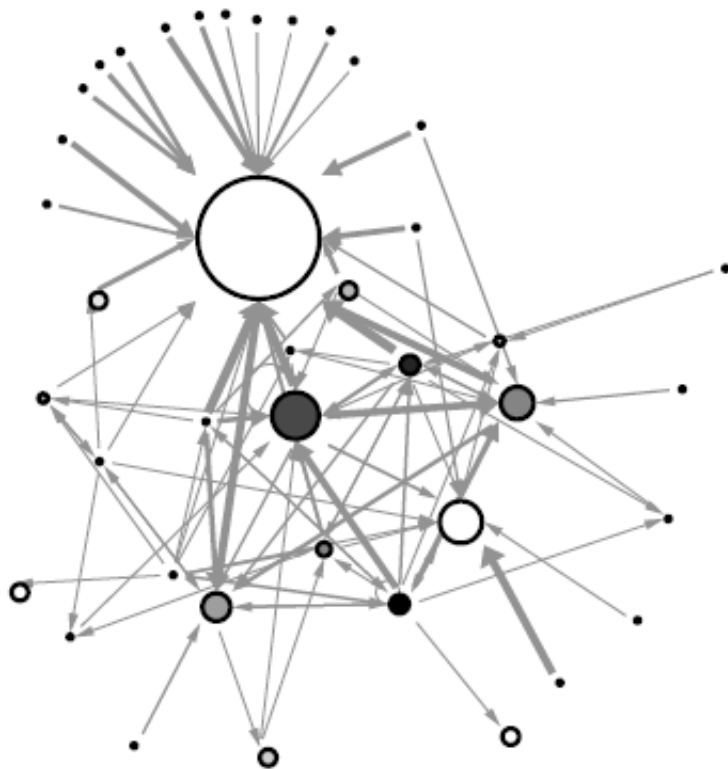
Freeman: .07
Variance: .20

Degree Centralization Scores

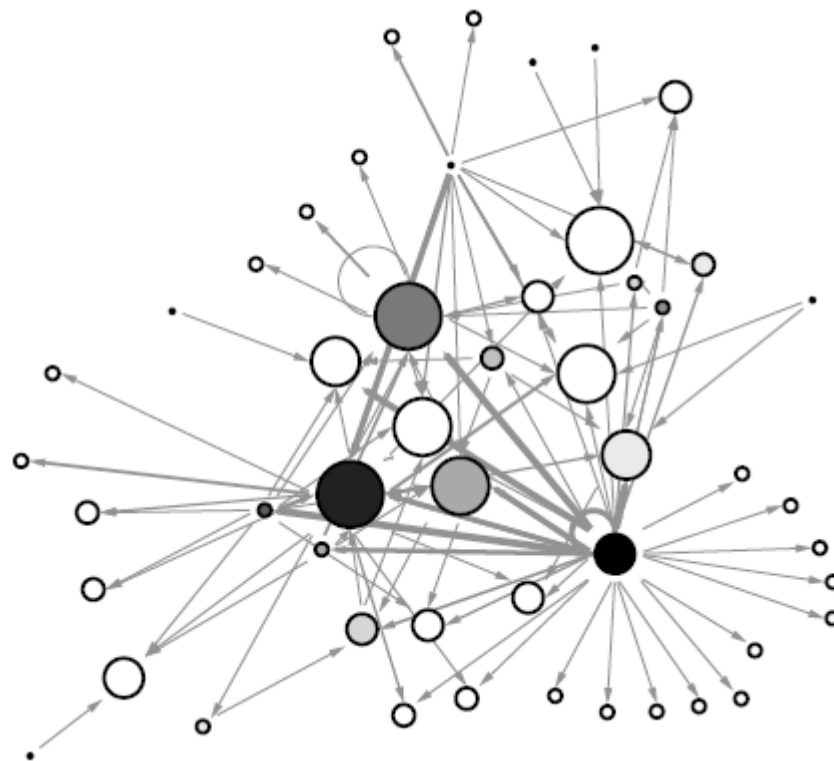


Freeman: 0.1
Variance: 4.84

Degree Centralization (example financial trading networks)



- high in-centralization: one node buying from many others



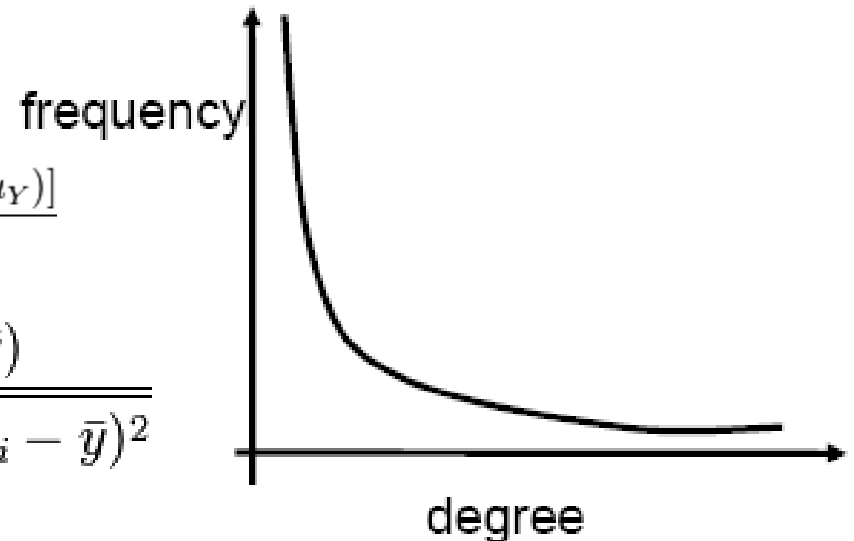
- low in-centralization: buying is more evenly distributed

Degree and degree distribution

- Degree k_i of node i is a measure of its centrality
- Nodes with high degrees are called **hubs**
- Maximum degree $k_{\max} = \max_i(k_i)$ is also an important measure
- The variance of node-degrees can be an indicator of network heterogeneity, i.e. the more the variance the more the heterogeneity
- Degree distribution

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad \rho_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

$$r = r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$



Degree-degree correlation

- It is important to know if the nodes with degree k are connected to nodes with degree k' .
- one way is to use the method proposed by Newman and compute the correlation coefficient
- Degree-degree correlation is computed as

$$r = \frac{\frac{1}{E} \sum_{j>i} k_i k_j a_{ij} - \left[\frac{1}{E} \sum_{j>i} \frac{1}{2} (k_i + k_j) a_{ij} \right]^2}{\frac{1}{E} \sum_{j>i} \frac{1}{2} (k_i^2 + k_j^2) a_{ij} - \left[\frac{1}{E} \sum_{j>i} \frac{1}{2} (k_i + k_j) a_{ij} \right]^2}$$

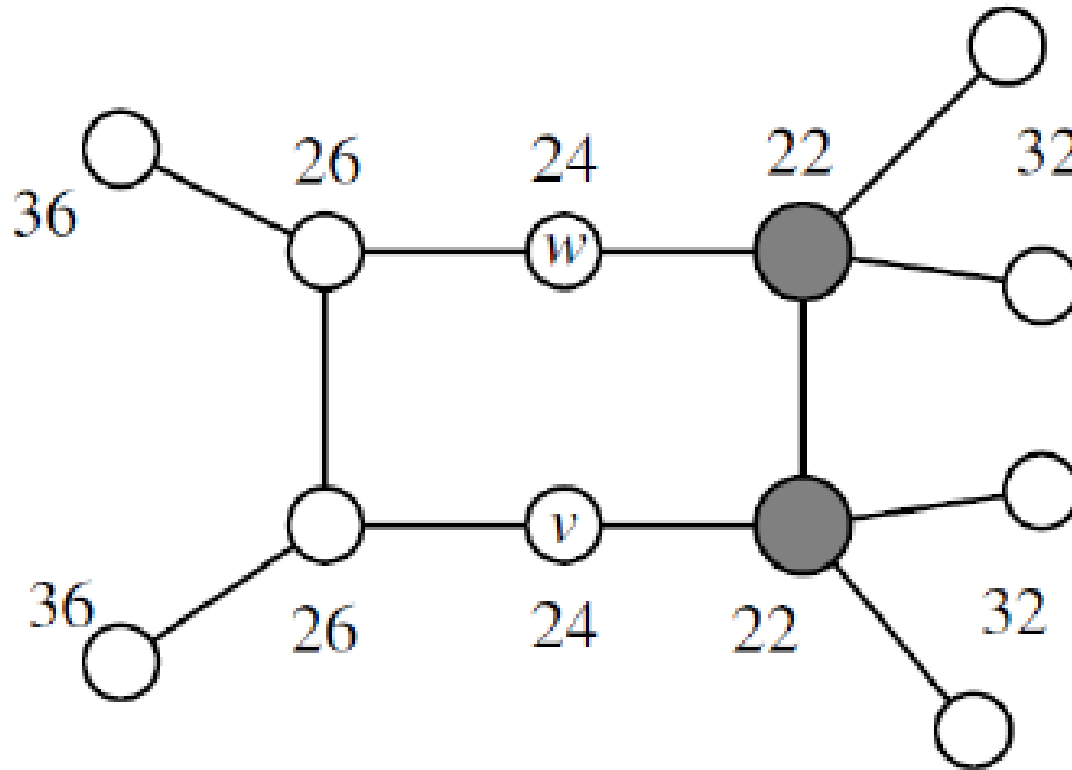
- E is the total number of edges
- a_{ij} is the entry (i,j) of the adjacency matrix
- k_i is the degree of node v_i

Degree-degree correlation

- $r > 0$: the network is called **assortative**
 - Node with **large** degree intent to connect to those with **large** degrees and nodes with **low** degrees intend to connect to those with **low** degrees (rich with rich and poor with poor)
- $r < 0$: the network is called **disassortative**
 - Node with **large** degree intent to connect to those with **low** degrees and nodes with **low** degrees intend to connect to those with **high** degrees (rich with poor)
- $r = 0$: no correlations
 - There is no specific intention in the connection between the nodes in the sense of their degrees

An example: shopping mall location

- The idea is that a node is central if it can quickly interact with all others (these nodes are called median of the graph)



Closeness-based centrality

A second measure of centrality is closeness centrality. A node is considered important if he/she is relatively close to all other actors.

Closeness is based on the inverse of the distance of each actor to every other actor in the network.

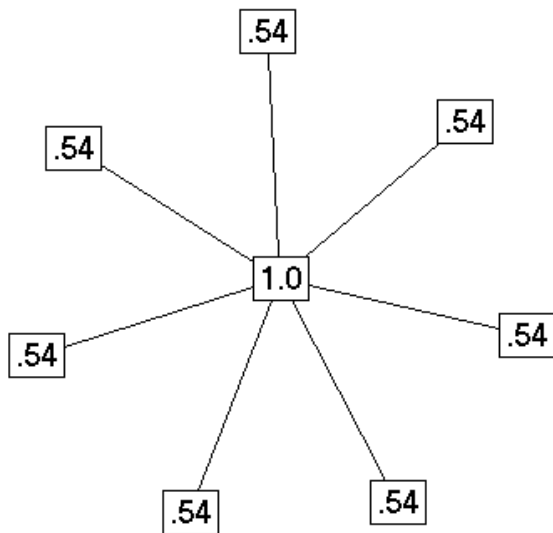
Closeness Centrality:

$$C_c(v_i) = \left[\sum_{j=1}^n d(v_i, v_j) \right]^{-1}$$

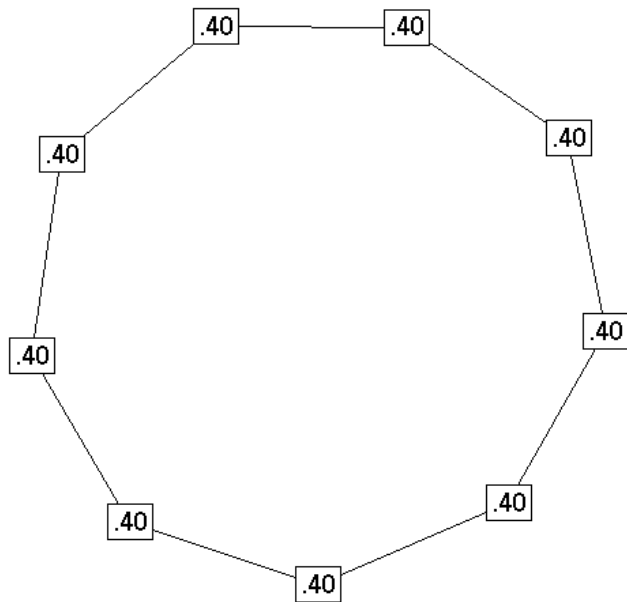
Normalized Closeness Centrality

$$\tilde{C}_c(v_i) = (C_c(v_i))(n-1) = \frac{(n-1)}{\sum_{j=1}^n d(v_i, v_j)}$$

Closeness Centrality in the examples



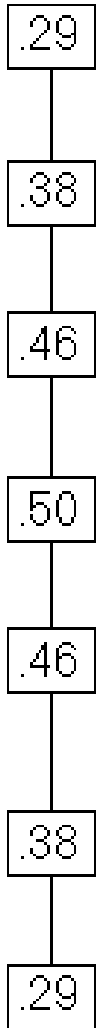
Distance								Closeness	normalized
0	1	1	1	1	1	1	1	.143	1.00
1	0	2	2	2	2	2	2	.077	.538
1	2	0	2	2	2	2	2	.077	.538
1	2	2	0	2	2	2	2	.077	.538
1	2	2	2	0	2	2	2	.077	.538
1	2	2	2	2	0	2	2	.077	.538
1	2	2	2	2	2	0	2	.077	.538
1	2	2	2	2	2	2	0	.077	.538



Distance								Closeness	normalized	
0	1	2	3	4	4	3	2	1	.050	.400
1	0	1	2	3	4	4	3	2	.050	.400
2	1	0	1	2	3	4	4	3	.050	.400
3	2	1	0	1	2	3	4	4	.050	.400
4	3	2	1	0	1	2	3	4	.050	.400
4	4	3	2	1	0	1	2	3	.050	.400
3	4	4	3	2	1	0	1	2	.050	.400
2	3	4	4	3	2	1	0	1	.050	.400
1	2	3	4	4	3	2	1	0	.050	.400

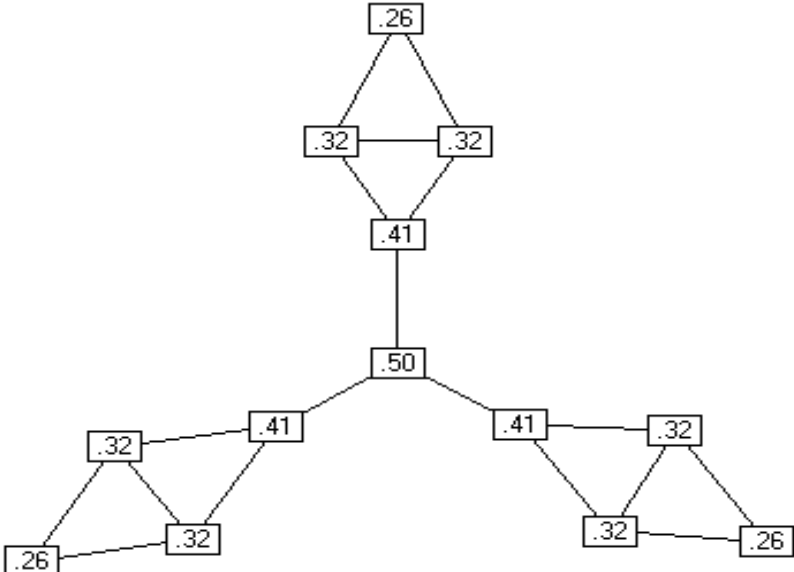
Closeness centrality

Closeness Centrality in the examples



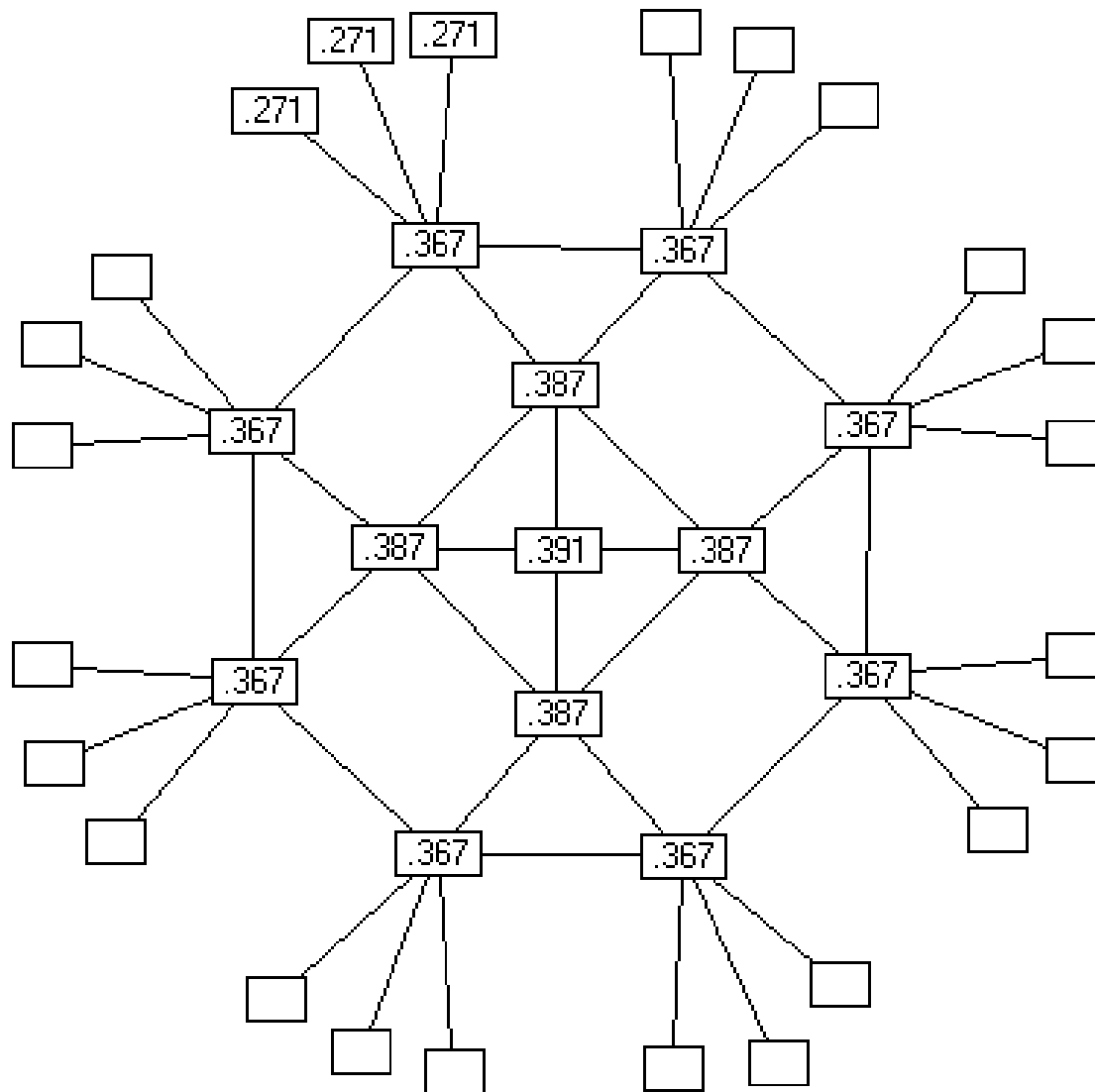
Distance							Closeness	normalized
0	1	2	3	4	5	6	.048	.286
1	0	1	2	3	4	5	.063	.375
2	1	0	1	2	3	4	.077	.462
3	2	1	0	1	2	3	.083	.500
4	3	2	1	0	1	2	.077	.462
5	4	3	2	1	0	1	.063	.375
6	5	4	3	2	1	0	.048	.286

Closeness Centrality in the example



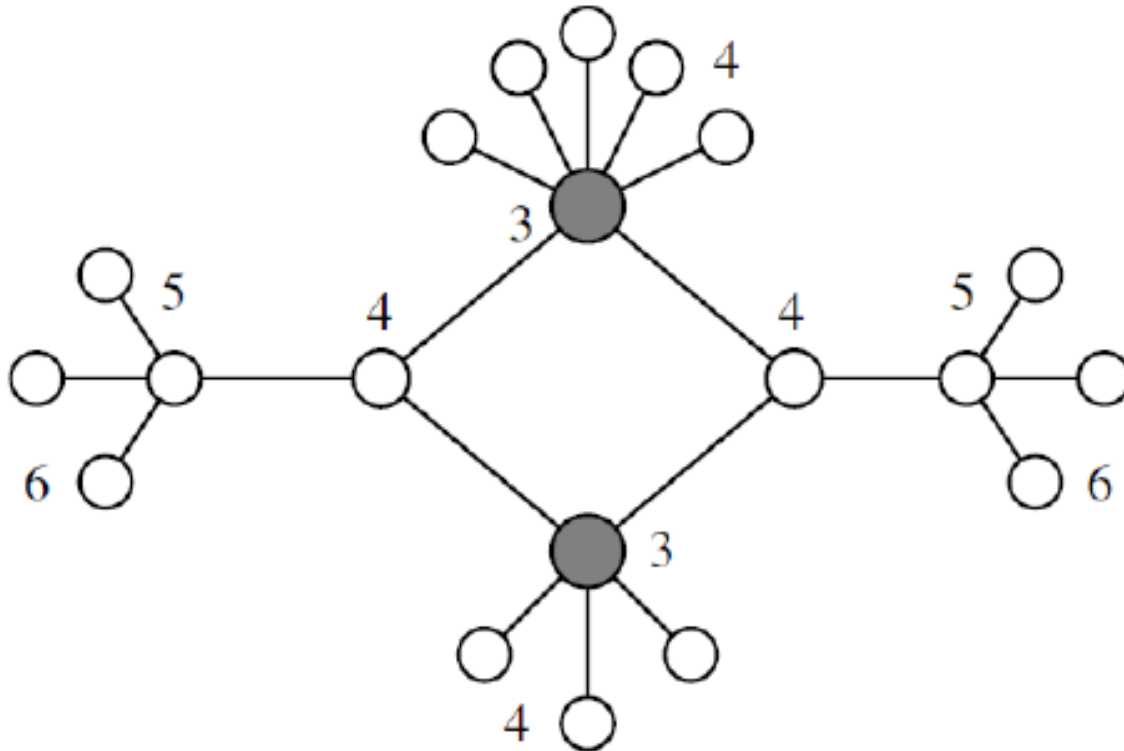
	Distance												Closeness	normalized
0	1	1	2	3	4	4	5	5	6	5	5	6	.021	.255
1	0	1	1	2	3	3	4	4	5	4	4	5	.027	.324
1	1	0	1	2	3	3	4	4	5	4	4	5	.027	.324
2	1	1	0	1	2	2	3	3	4	3	3	4	.034	.414
3	2	2	1	0	1	1	2	2	3	2	2	3	.042	.500
4	3	3	2	1	0	2	3	3	4	1	1	2	.034	.414
4	3	3	2	1	2	0	1	1	2	3	3	4	.034	.414
5	4	4	3	2	3	1	0	1	1	4	4	5	.027	.324
5	4	4	3	2	3	1	1	0	1	4	4	5	.027	.324
6	5	5	4	3	4	2	1	1	0	5	5	6	.021	.255
5	4	4	3	2	1	3	4	4	5	0	1	1	.027	.324
5	4	4	3	2	1	3	4	4	5	1	0	1	.027	.324
6	5	5	4	3	2	4	5	5	6	1	1	0	.021	.255

Closeness Centrality in the example



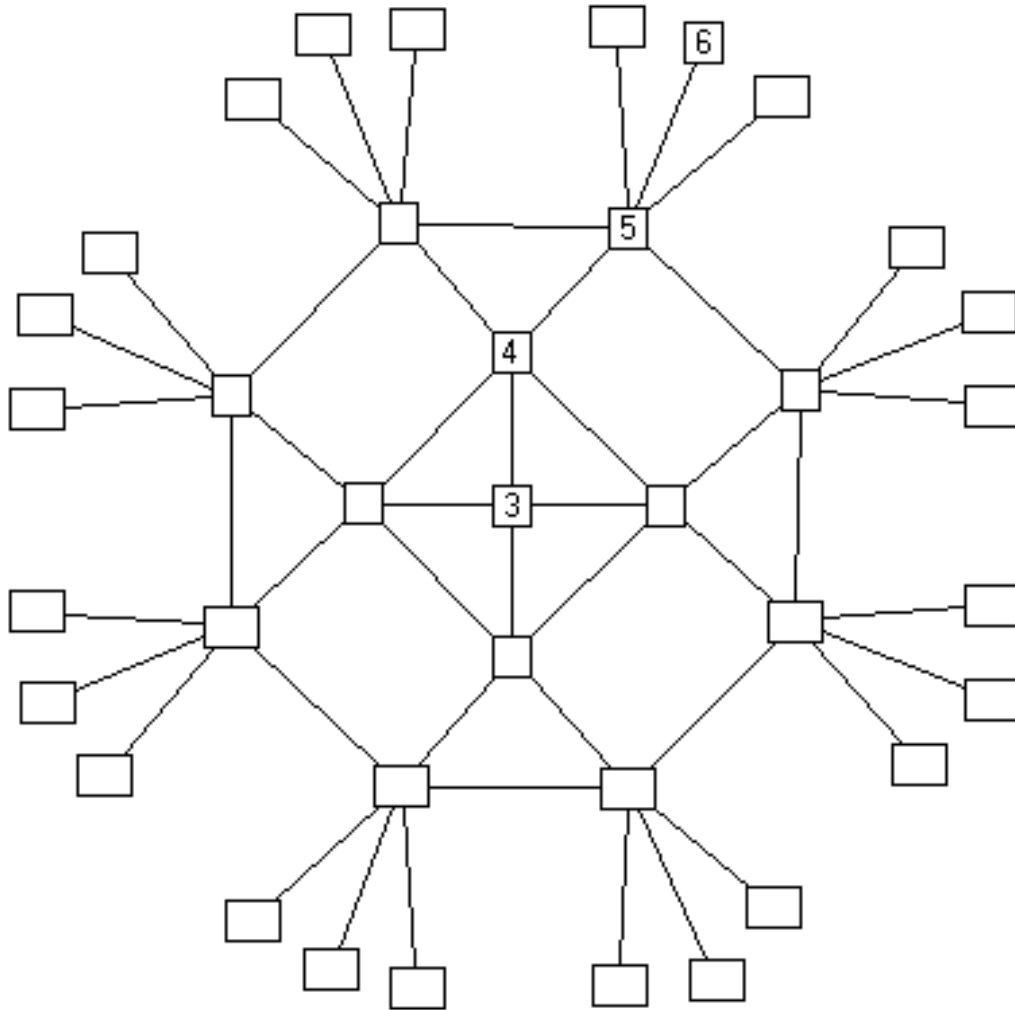
Example: Hospital location problem

- **Eccentricity:** Maximum distance of a node to all other nodes: $e(u) = \max\{d_{uv} : v \in V\}$
- **Radius of graph:** Minimum eccentricity



Graph Theoretic Center (Barry or Jordan Center).

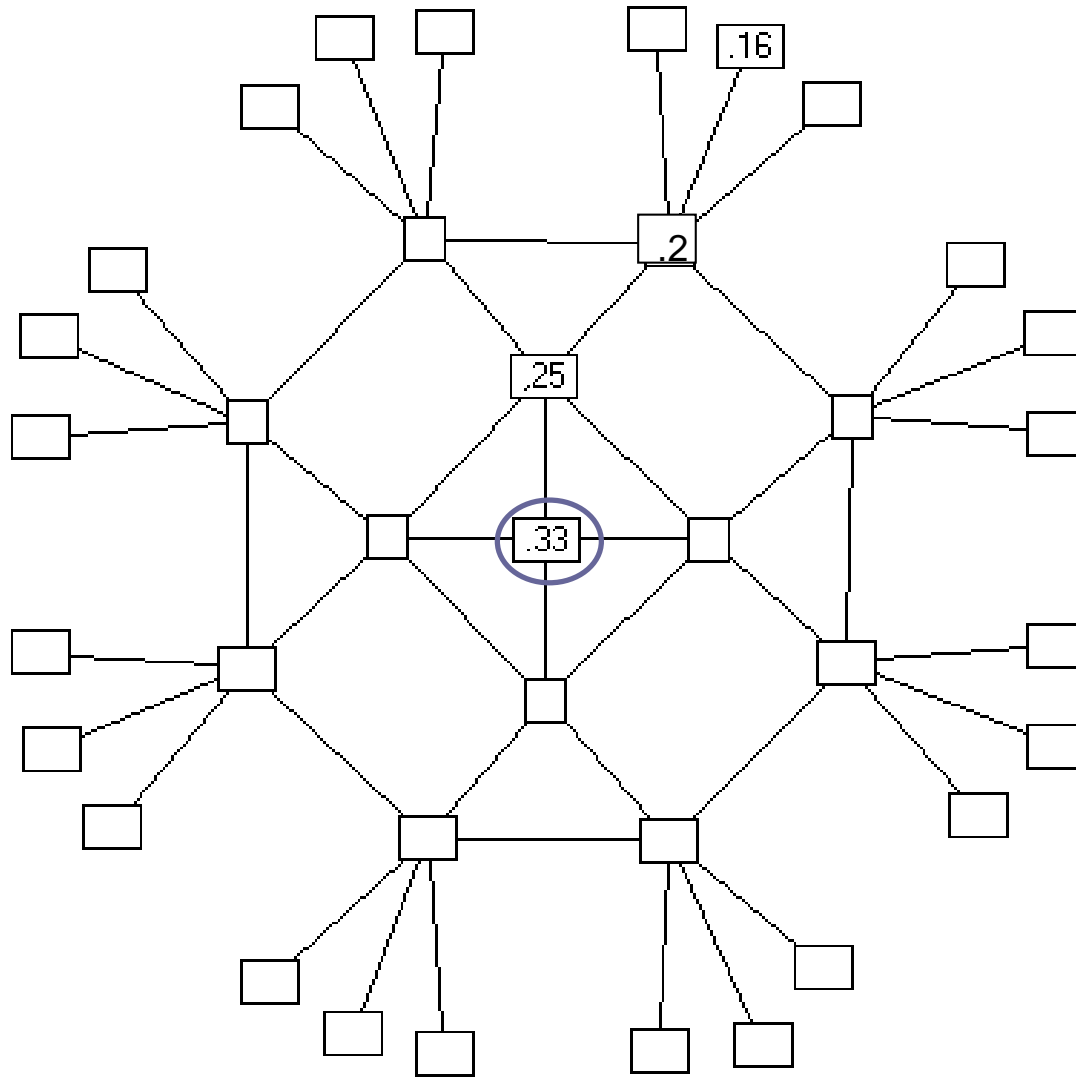
Identify the points with the smallest, maximum distance to all other points.



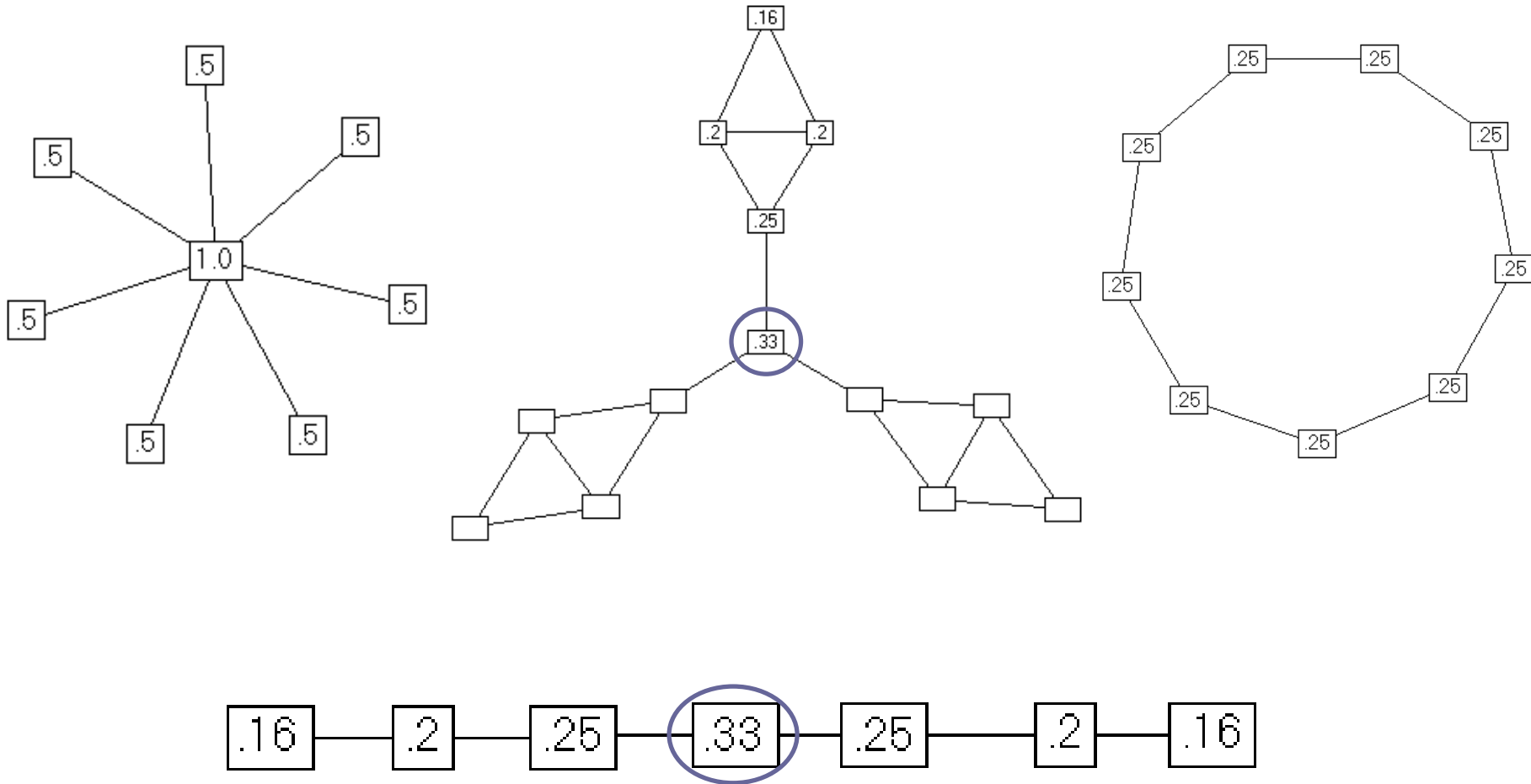
Value = longest distance to any other node.

The graph theoretic center is '3', but you might also consider a continuous measure as the inverse of the maximum geodesic

Graph Theoretic Center (Barry or Jordan Center).

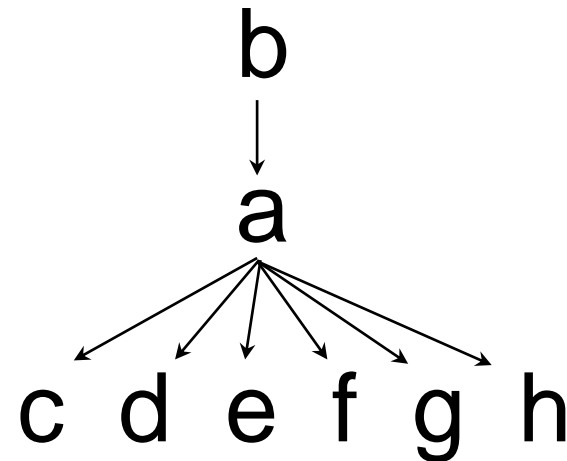
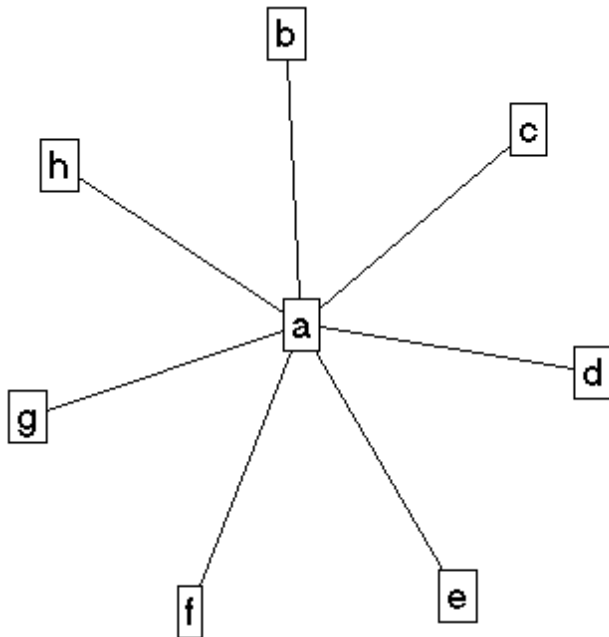


Graph Theoretic Center (Barry or Jordan Center).

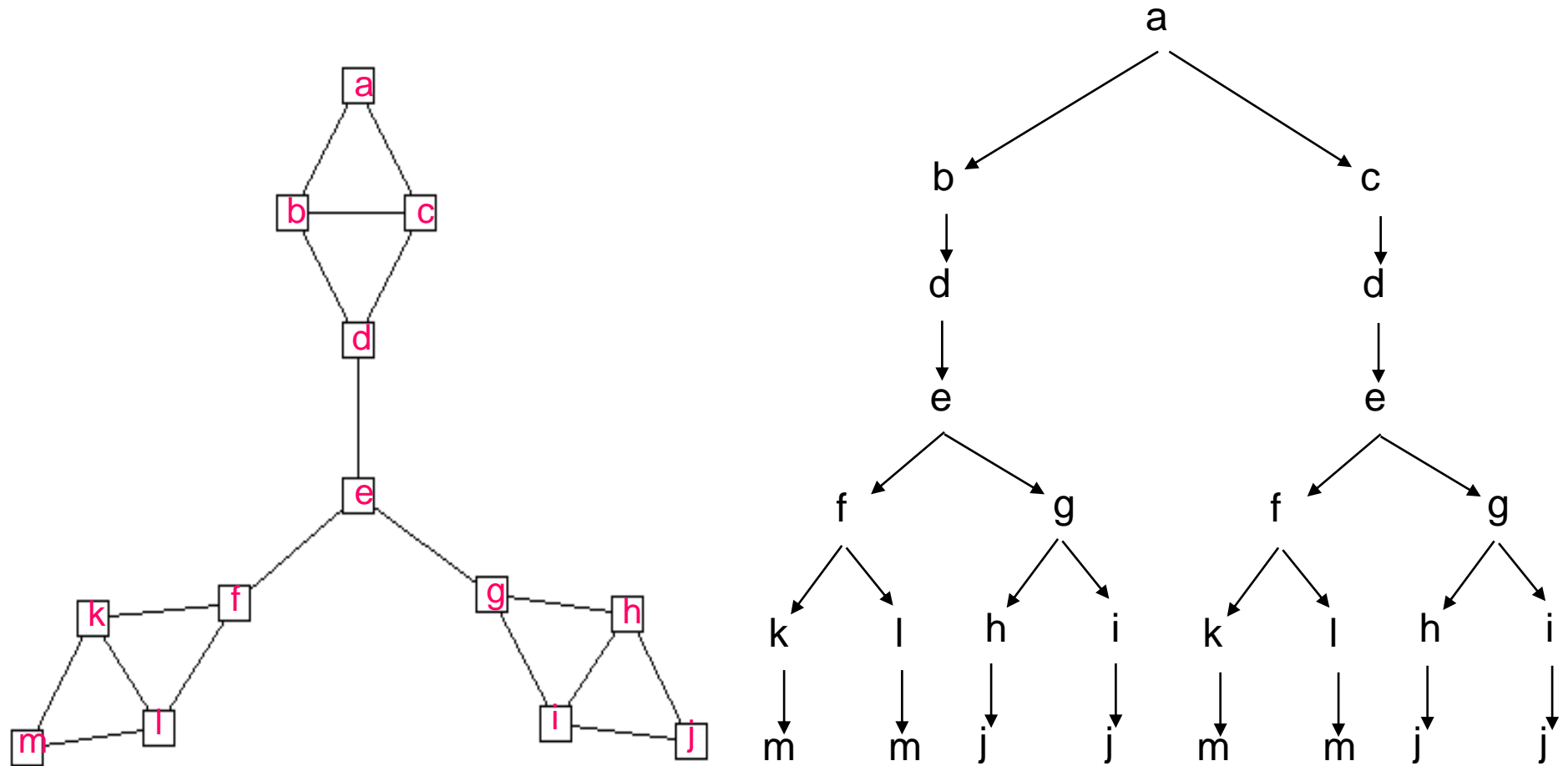


Betweenness Centrality

- Model based on communication flow: A person who lies on communication paths can control communication flow, and is thus important.
- Betweenness centrality counts the number of geodesic (shortest) paths between i and k that actor j resides on.



Betweenness Centrality



Betweenness Centrality

- **Node Betweenness**

$$C_B(v_i) = \frac{\sum_{j,k} g_{jk}(v_i)}{\sum_{j,k} g_{jk}}$$

g_{jk} = the number of shortest paths between v_j and v_k

$g_{jk}(v_i)$ = the number shortest paths between v_j and v_k that pass through node v_i

- **Edge Betweenness:** Its definition is similar to node betweenness

$$C_B(e_i) = \frac{\sum_{j,k} g_{jk}(e_i)}{\sum_{j,k} g_{jk}}$$

g_{jk} = the number of shortest paths between v_j and v_k

$g_{jk}(e_i)$ = the number shortest paths between v_j and v_k that pass through edge e_i

Usually normalized by:

$$\tilde{C}_B(n_i) = \frac{C_B(v_i)}{E_{\max}}$$

Eigenvector centrality

- Bonacich eigenvector, 1972
- Idea: Connections to people who are themselves influential will lend a person more influence than connections to less influential people.
- x_i : centrality of node v_i is proportional to the average of the centralities of j 's neighbors

$$C_{eig}(v_i) = \frac{1}{\lambda} \sum_j a_{ij} x_j$$

- Assuming centralities to be non-negative:
 - λ must be the **largest** eigenvalue of A and \mathbf{x} the corresponding eigenvector (Perron–Frobenius theorem)

Bonacich power centrality

- Bonacich Power Centrality: Actor's centrality (prestige) is equal to a function of the prestige of those they are connected to. Thus, actors who are tied to very central actors should have higher prestige/centrality than those who are not.

$$C(\alpha, \beta) = \alpha(I - \beta R)^{-1} R \mathbf{1}$$

- α is a scaling vector, which is set to normalize the score.
- β reflects the extent to which you *weight* the centrality of people ego is tied to.
- \mathbf{R} is the adjacency matrix (can be valued)
- \mathbf{I} is the identity matrix (1s down the diagonal)
- $\mathbf{1}$ is a matrix of all ones.

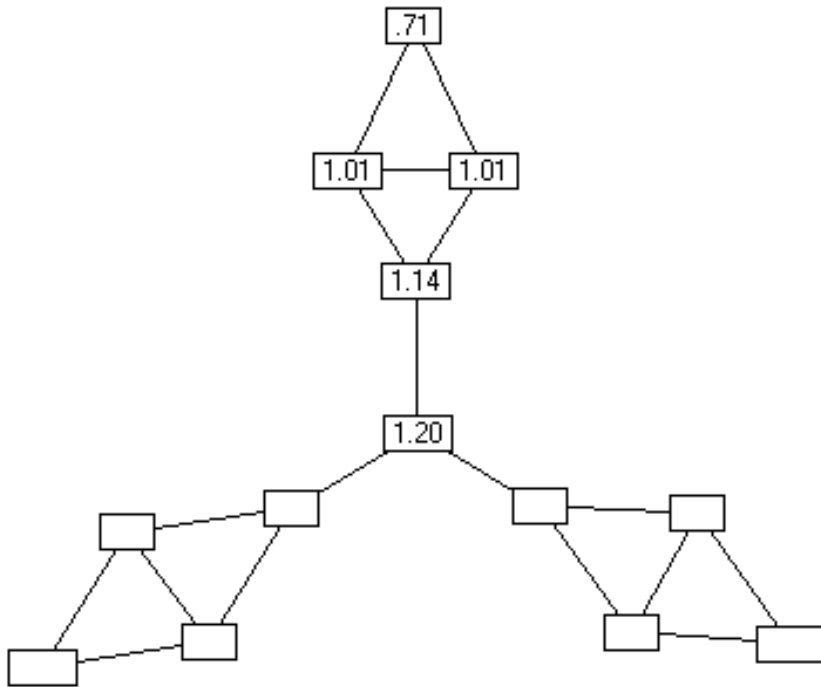
Bonacich power centrality

- The magnitude of β reflects the radius of power.
- Small values of β weight local structure.
- Large values of β weight global structure.

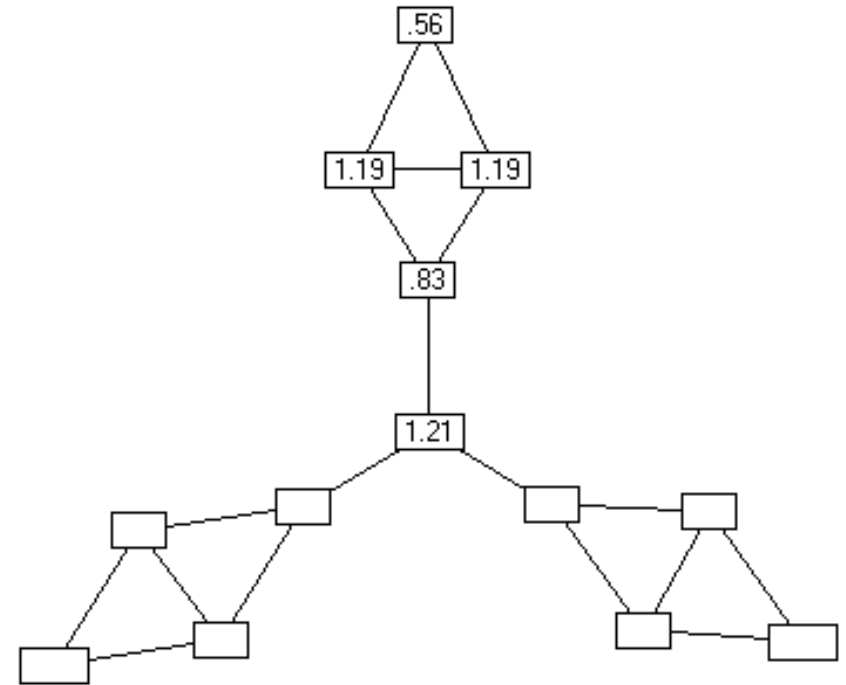
- If $\beta > 0$, the node has higher centrality when tied to people who are central.
- If $\beta < 0$, then node has higher centrality when tied to people who are not central
- $\beta = 0$, It get degree centrality.

Bonacich power centrality

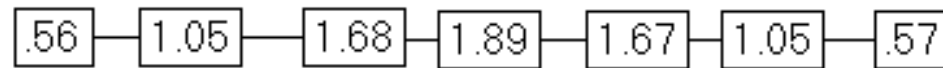
$\beta=0.35$



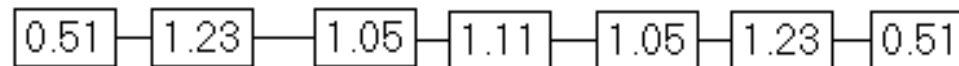
$\beta= -0.35$



$\beta=0.23$

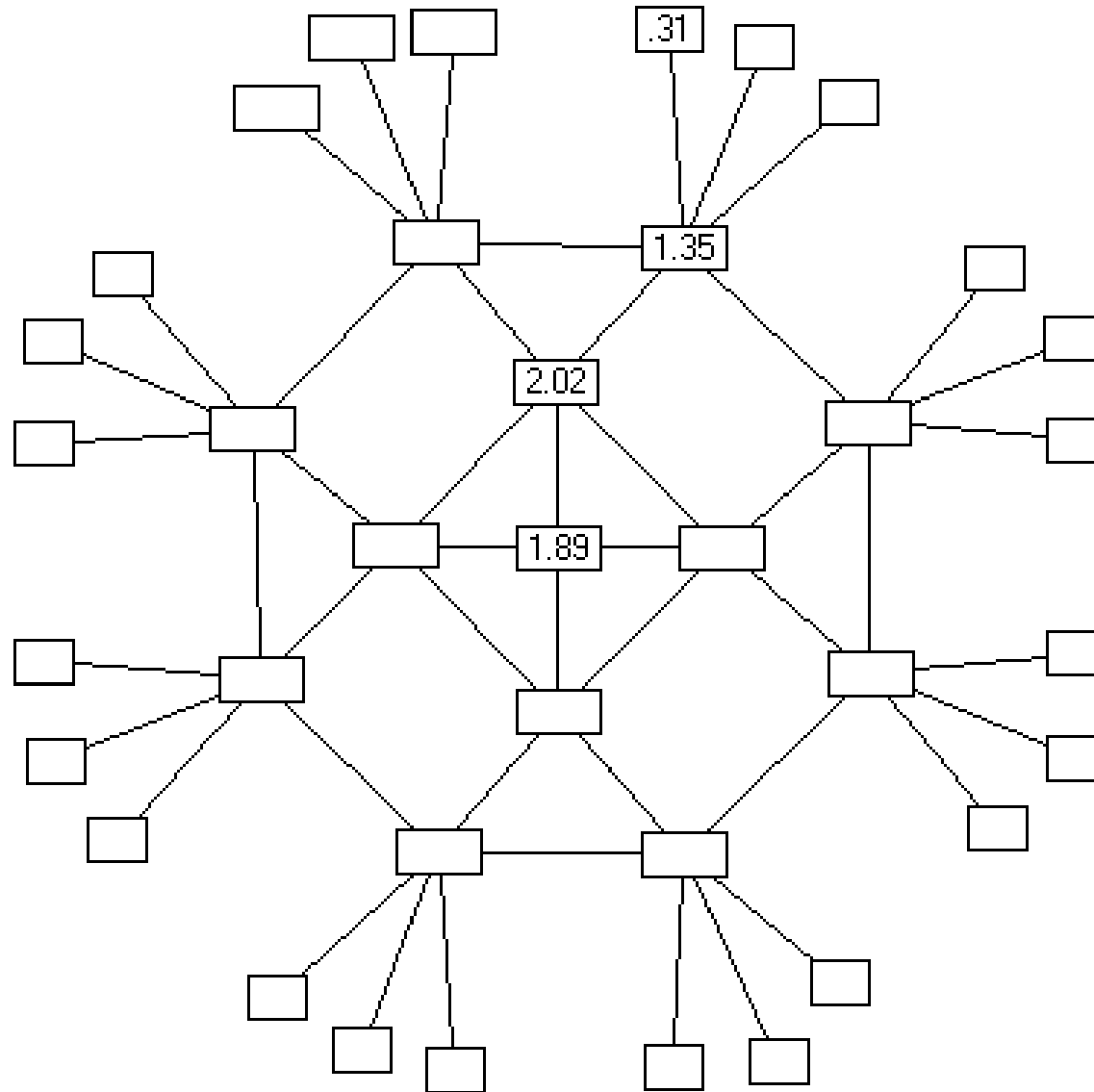


$\beta= -0.23$



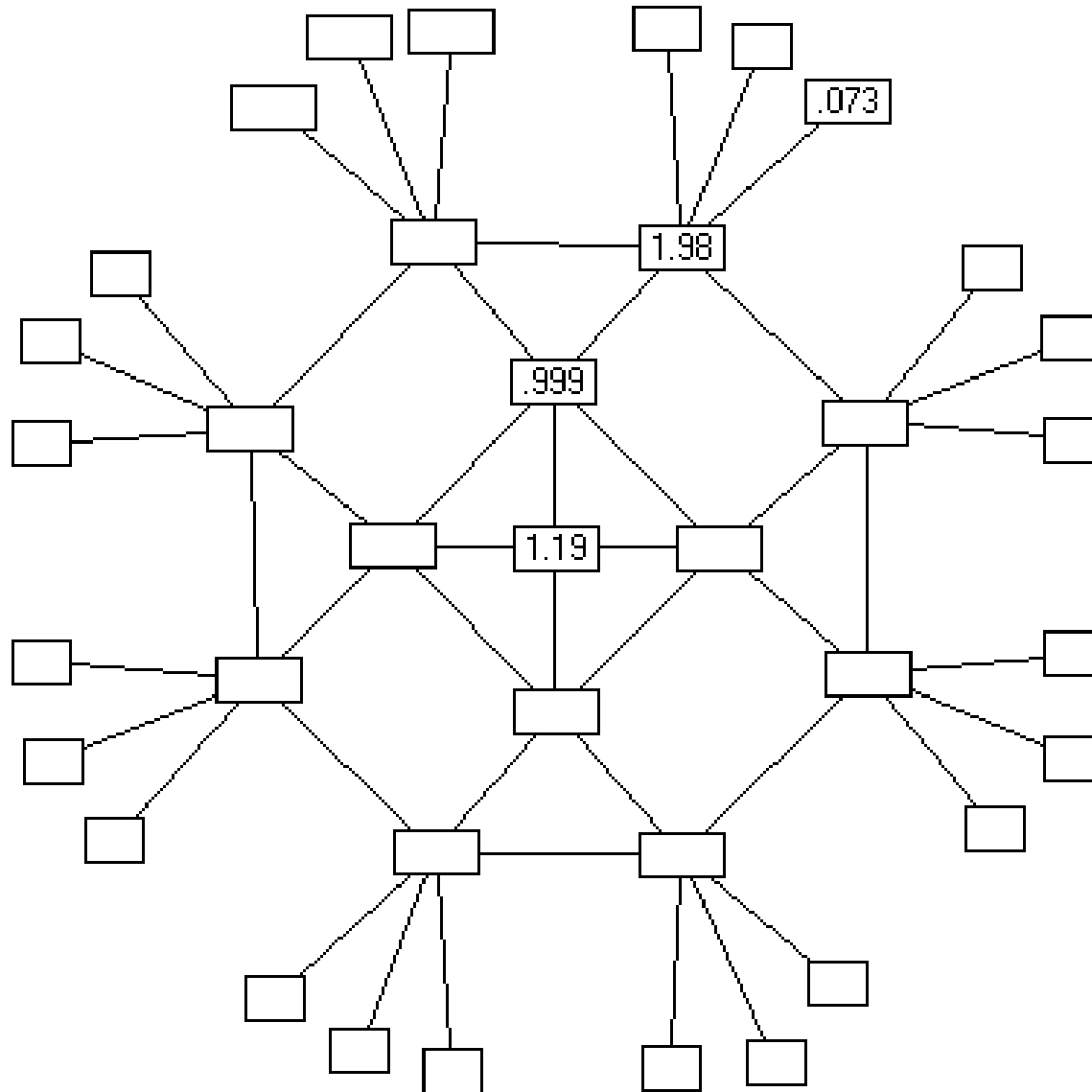
Bonacich power centrality

$\beta = .23$



Bonacich power centrality

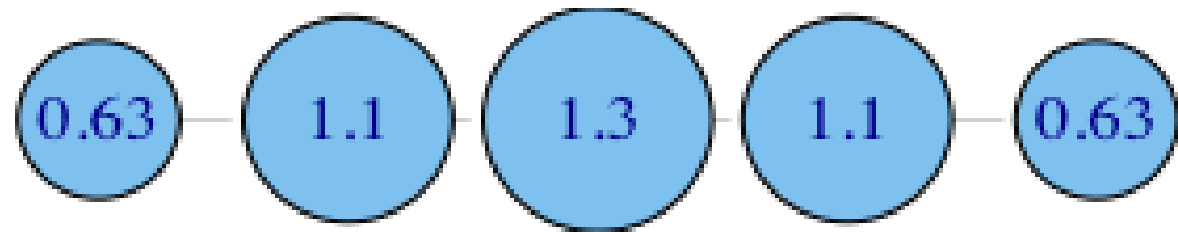
$$\beta = -.23$$



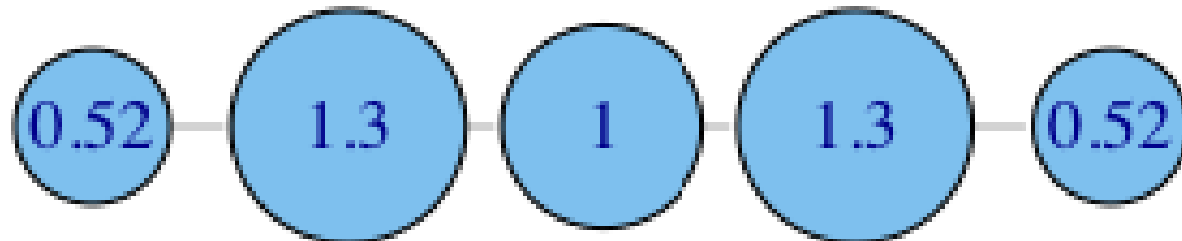
Bonacich power centrality

■ Example

■ $\beta=0.25$



■ $\beta=-0.25$



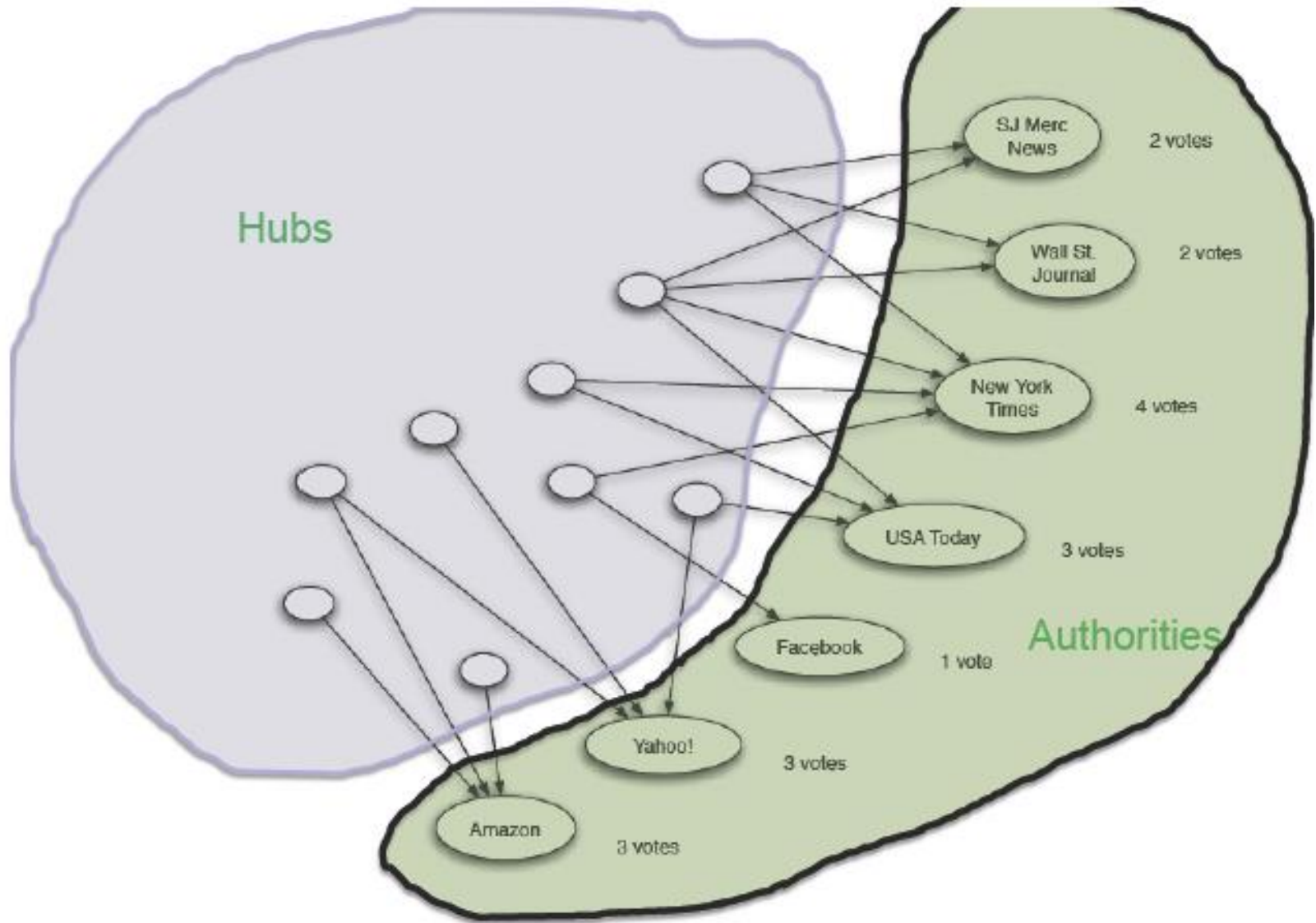
HITS centrality

- Hyperlink-Induced Topic Search
- also known as Hubs and authorities
- Developed by Jon Kleinberg
- Precursor to Page Rank
- Certain web pages, known as hubs, serve as large directories

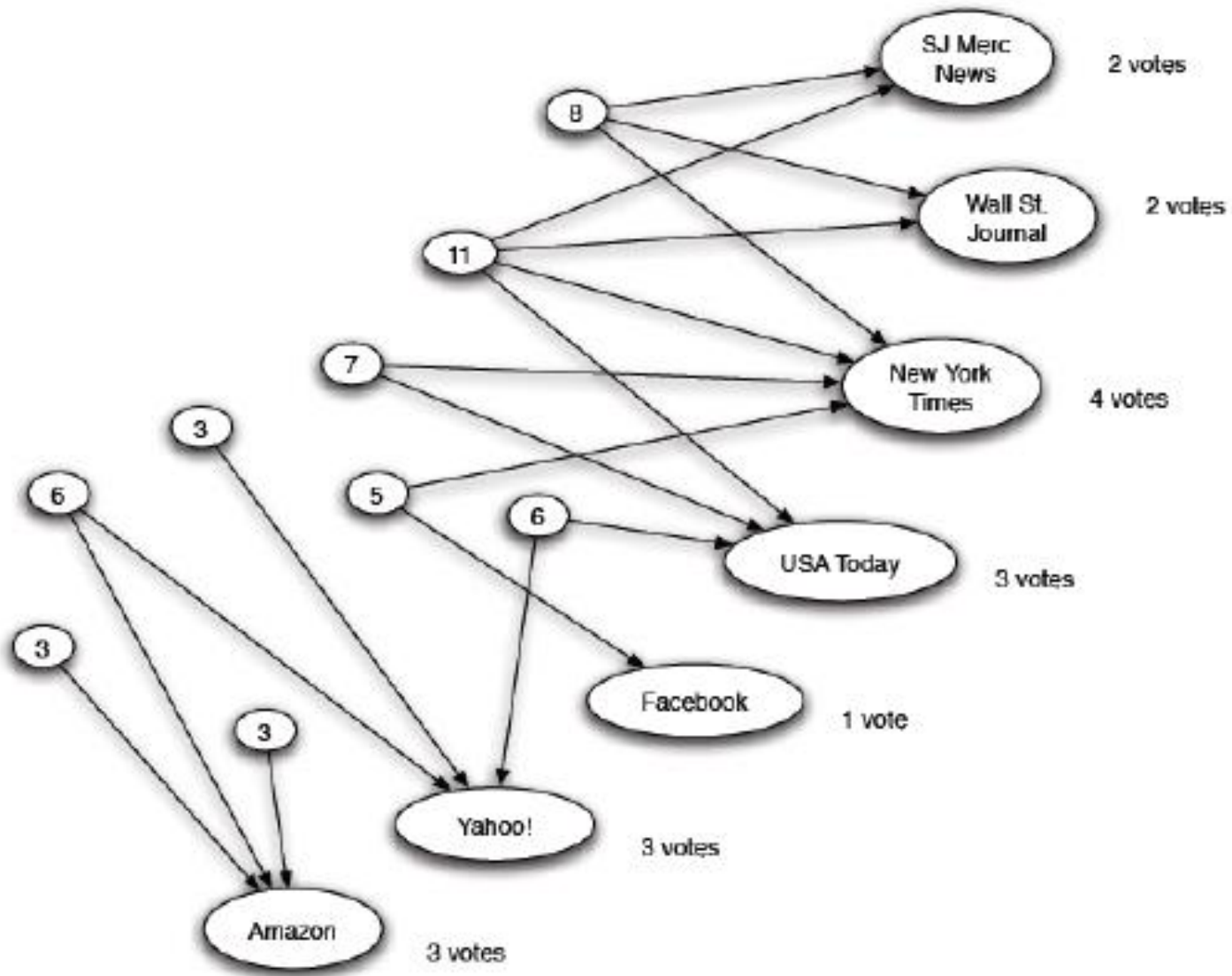
Hubs vs. Authorities

- Hubs are not actually authoritative in the information that it hold, but are used as compilations of a broad catalog of information that lead users directly to other authoritative pages.
- a good hub represents a page that points to many other pages, and
- a good authority represents a page that is linked by many different hubs.

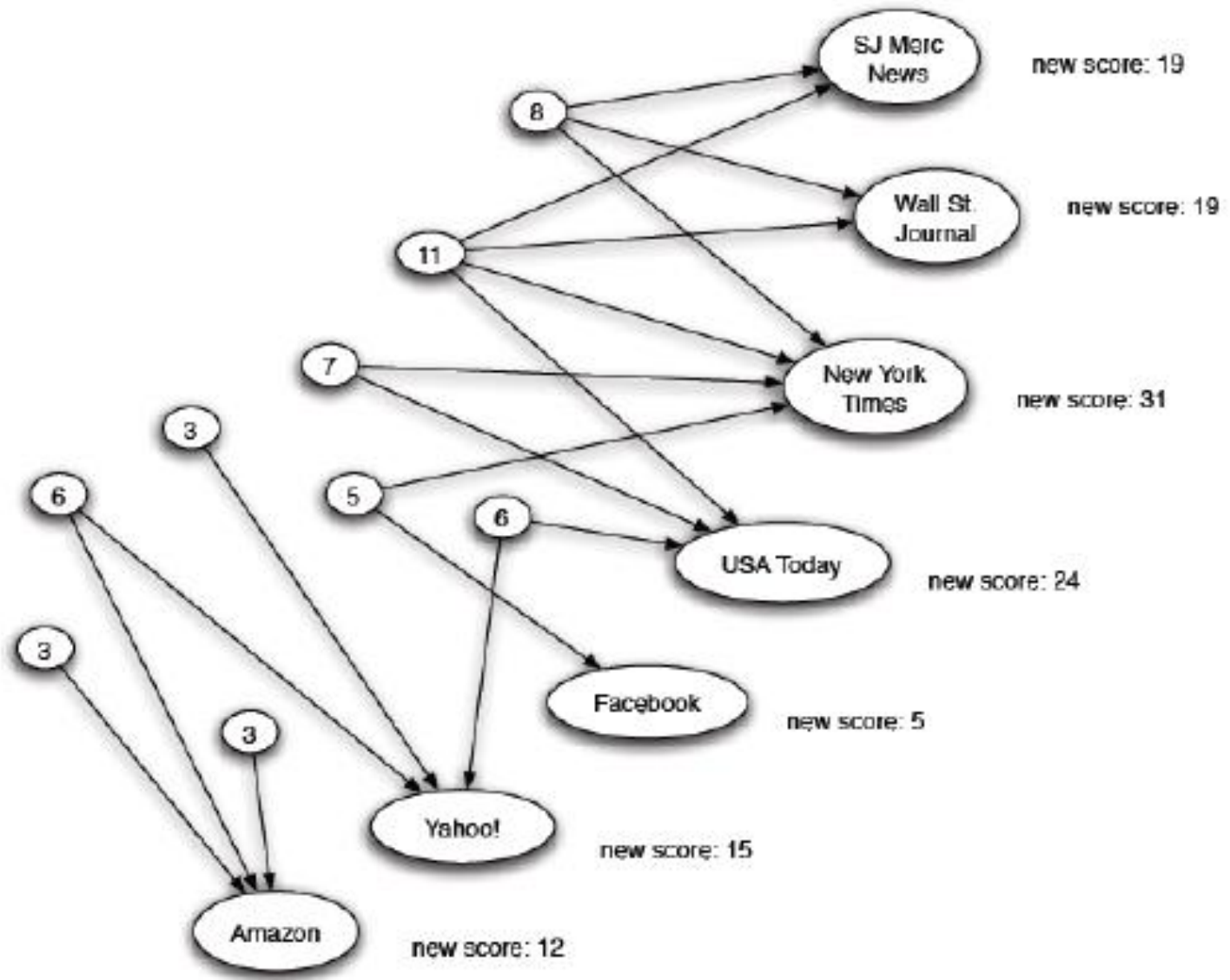
HITS



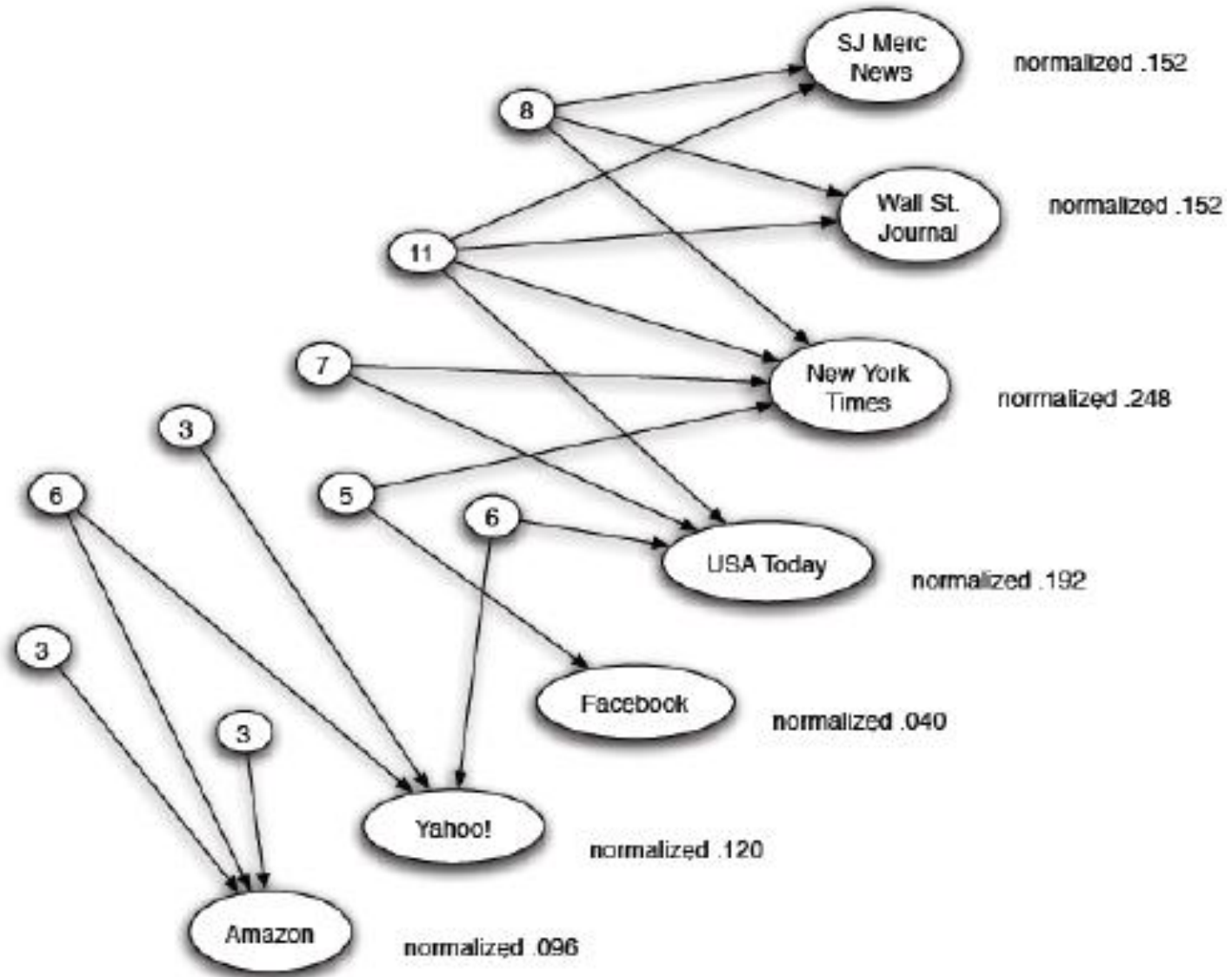
HITS



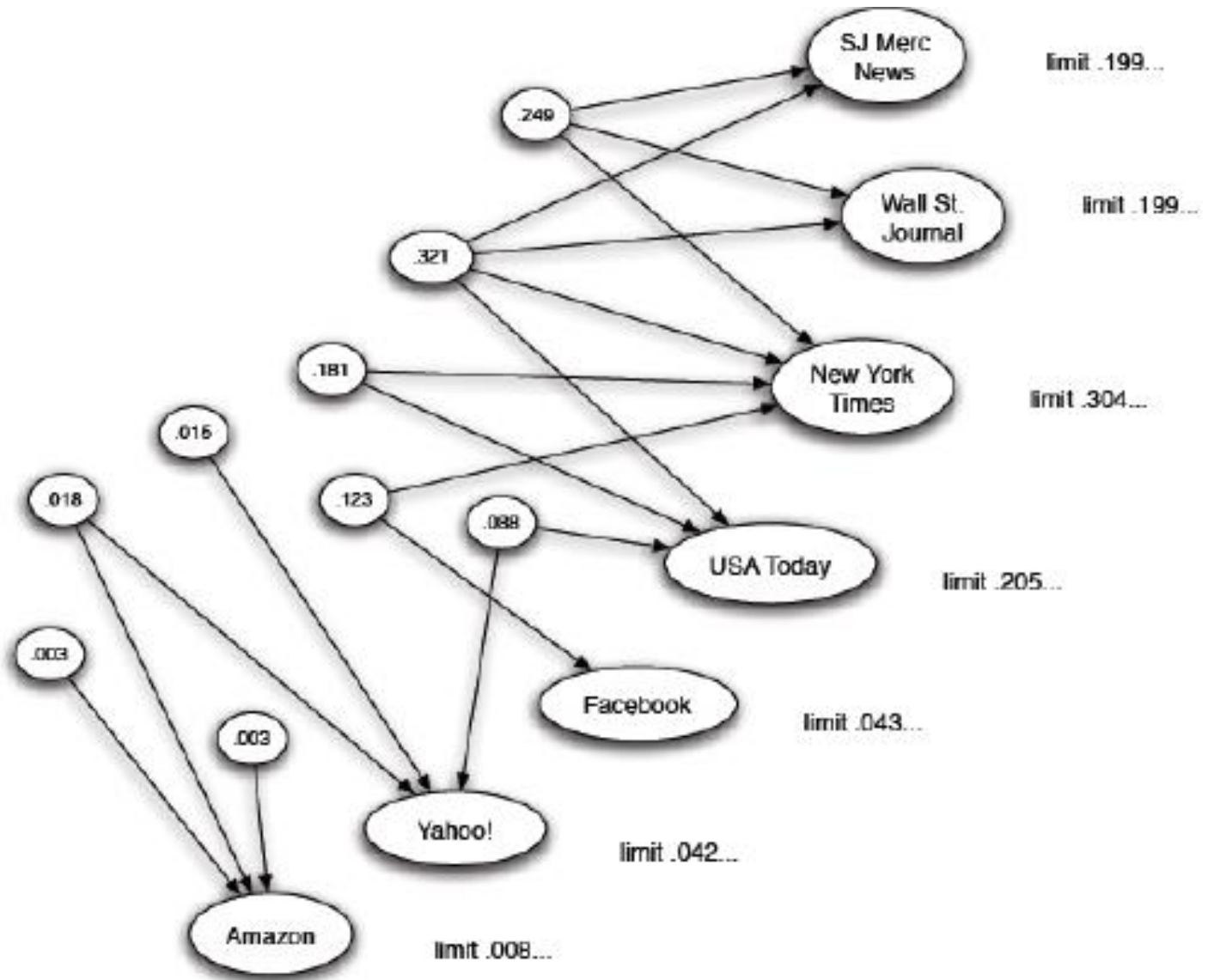
HITS



HITS



HITS



PageRank

■ History !

■ How to organize the Web?

■ First try: Human curated Web directories

- Yahoo, DMOZ, LookSmart

■ Second try: Web Search

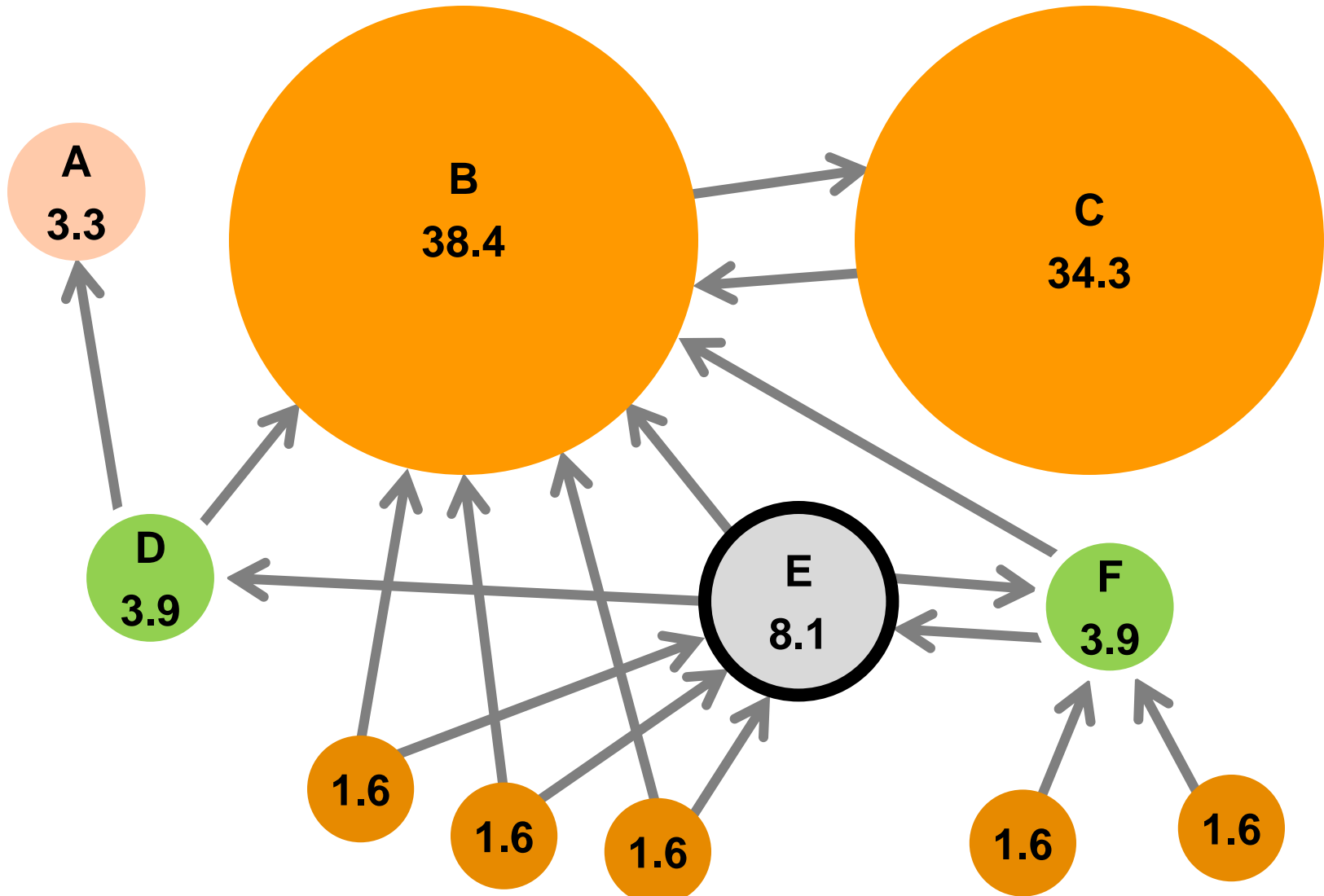
- **Information Retrieval** investigates:
Find relevant docs in a small
and trusted set

- Newspaper articles, Patents, etc.

- **But:** Web is **huge**, full of untrusted documents, random things,
web spam, etc.



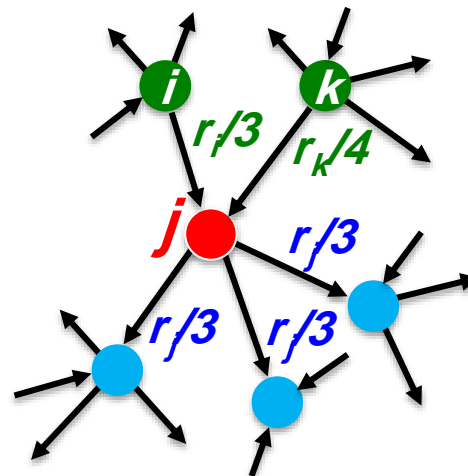
Example: PageRank Scores



Simple Recursive Formulation

- Each link's vote is proportional to the **importance** of its source page
- If page j with importance r_j has n out-links, each link gets r_j/n votes
- Page j 's own importance is the sum of the votes on its in-links

$$r_j = r_i/3 + r_k/4$$



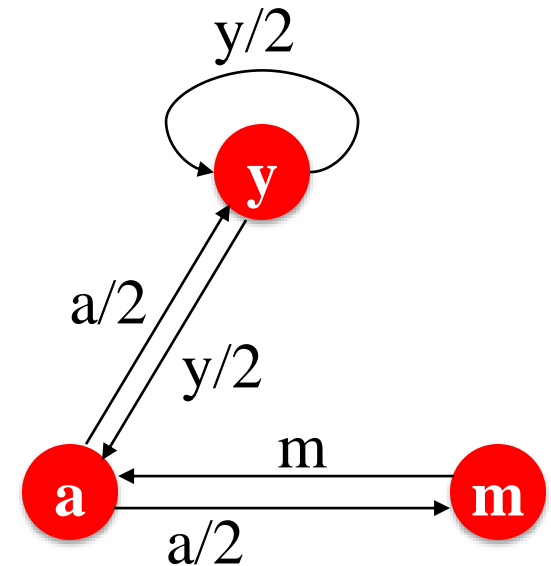
PageRank: The “Flow” Model

- A “vote” from an important page is worth more
- A page is important if it is pointed to by other important pages
- Define a “rank” r_j for page j

$$r_j = \sum_{i \rightarrow j} \frac{r_i}{d_i}$$

d_i ... out-degree of node i

The web in 1839



“Flow” equations:

$$r_y = r_y/2 + r_a/2$$

$$r_a = r_y/2 + r_m$$

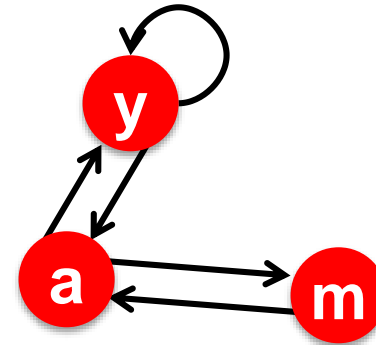
$$r_m = r_a/2$$

PageRank: How to solve?

Power Iteration:

- Set $r_j = 1/N$
- 1: $r'_j = \sum_{i \rightarrow j} \frac{r_i}{d_i}$
- 2: $r = r'$
- Goto 1

Example:



	y	a	m
y	1/2	1/2	0
a	1/2	0	1
m	0	1/2	0

$$\mathbf{r}_y = \mathbf{r}_y/2 + \mathbf{r}_a/2$$

$$\mathbf{r}_a = \mathbf{r}_y/2 + \mathbf{r}_m$$

$$\mathbf{r}_m = \mathbf{r}_a/2$$

$$\begin{pmatrix} \mathbf{r}_y \\ \mathbf{r}_a \\ \mathbf{r}_m \end{pmatrix} = \begin{matrix} 1/3 & 1/3 & 5/12 & 9/24 & & 6/15 \\ 1/3 & 3/6 & 1/3 & 11/24 & \dots & 6/15 \\ 1/3 & 1/6 & 3/12 & 1/6 & & 3/15 \end{matrix}$$

Iteration 0, 1, 2, ...

The Google Matrix

- **PageRank equation** [Brin-Page, '98]

$$r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}$$

- **What is β ?**

- In practice $\beta = 0.85$

$[1/N]_{1 \times N}$... 1 by N matrix
where all entries are 1/N

- **The Google Matrix R :**

$$R = \beta \ell R + (1 - \beta) \left[\frac{1}{N} \right]_{1 \times N}$$

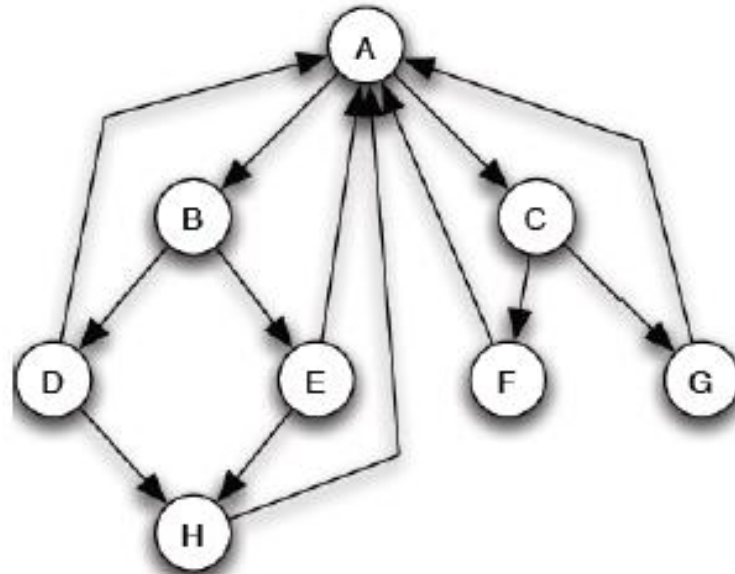
- **We have a recursive problem**

$$\mathbf{R} = \begin{bmatrix} PR(p_1) \\ PR(p_2) \\ \vdots \\ PR(p_N) \end{bmatrix} \quad \begin{bmatrix} \ell(p_1, p_1) & \ell(p_1, p_2) & \cdots & \ell(p_1, p_N) \\ \ell(p_2, p_1) & \ddots & & \vdots \\ \vdots & & \ell(p_i, p_j) & \\ \ell(p_N, p_1) & \cdots & & \ell(p_N, p_N) \end{bmatrix} \quad \sum_{j=1}^N \ell(p_i, p_j) = 1$$

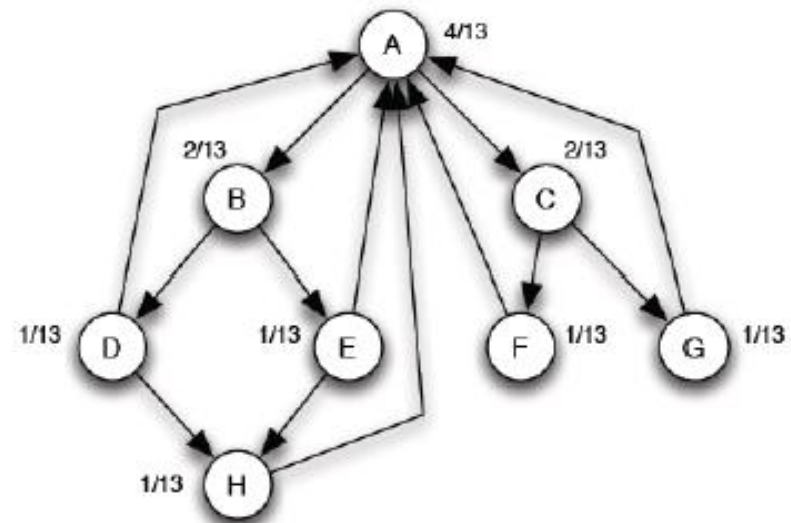
If $\ell(p_i, p_j) > 0$,
then p_i links to p_j

Page Rank

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j},$$

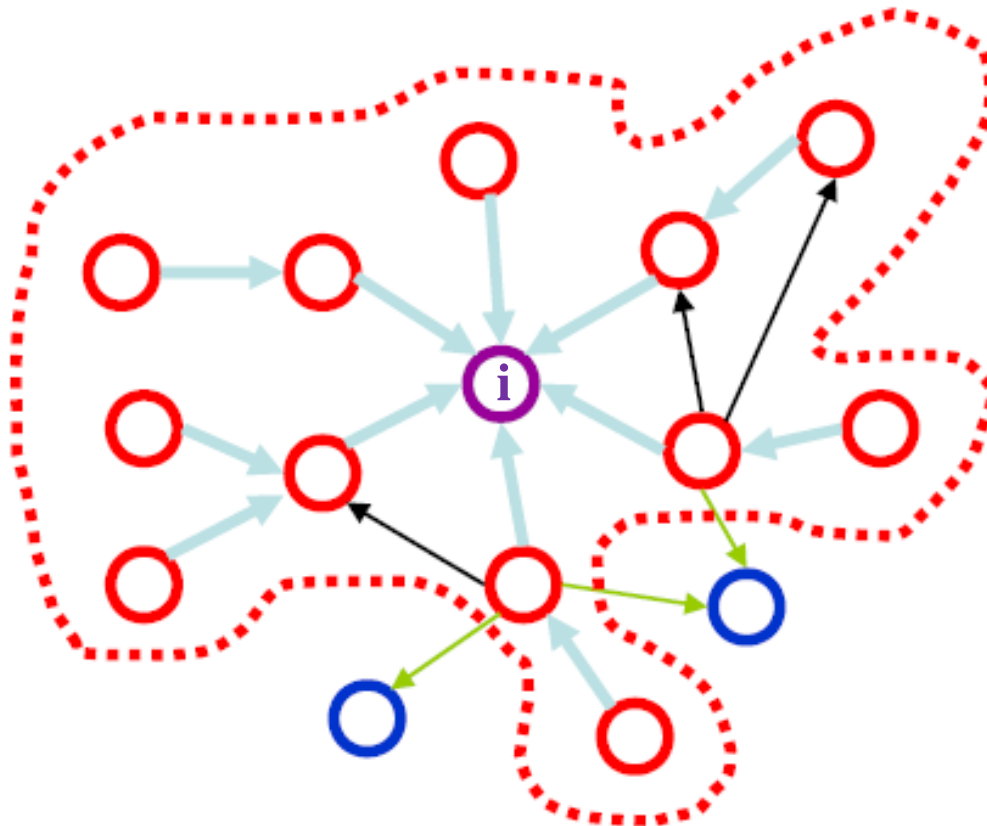


■ Equilibrium Values for Page Rank



Influence range in directed networks

- The influence range of i is the set of vertices who are reachable from the node i
- Alternatively, we can also consider the influential range of node i as a set of the nodes with a path to i



Network entropy (of degree distribution)

- Entropy is an indicator of disorder
- The more the disorder, the more the entropy
- The entropy of the degree distribution provides an average measurement of the heterogeneity of the network

$$H = -\sum_k P(k) \log P(k)$$

- $P(k)$ the probability network
- The maximum value of entropy is obtained for a uniform degree distribution
- The minimum value $H_{\min} = 0$ is achieved whenever all nodes have the same degree

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

$$V_i = \frac{E - E_i}{E}$$

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

$$V = \max_j V_j$$

- 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

Disconnecting and cut sets

- How many edges or nodes must be removed in order to disconnect an originally connected graph?
- If a node is removed then all edges joining it will also be removed
- But, the converse may not be true, i.e. an edge may removed without necessarily removing the nodes tipping to that
- **Disconnecting set:** a set of edges $E_o(G)$, after it is being removed, the graph G will be become disconnected
- **Cut set:** the smallest disconnecting set, i.e. No proper subset of which is a disconnecting set

Readings

- Newman, Mark. ***Networks: an introduction***. Oxford University Press, 2010. (Ch. 7)
- Easley, David, and Jon Kleinberg. **Networks, crowds, and markets: Reasoning about a highly connected world**. Cambridge University Press, 2010. (Ch. 3)
- L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. **Characterization of complex networks: A survey of measurements**. *Advances in Physics*, 56(1):167 – 242, 2007.