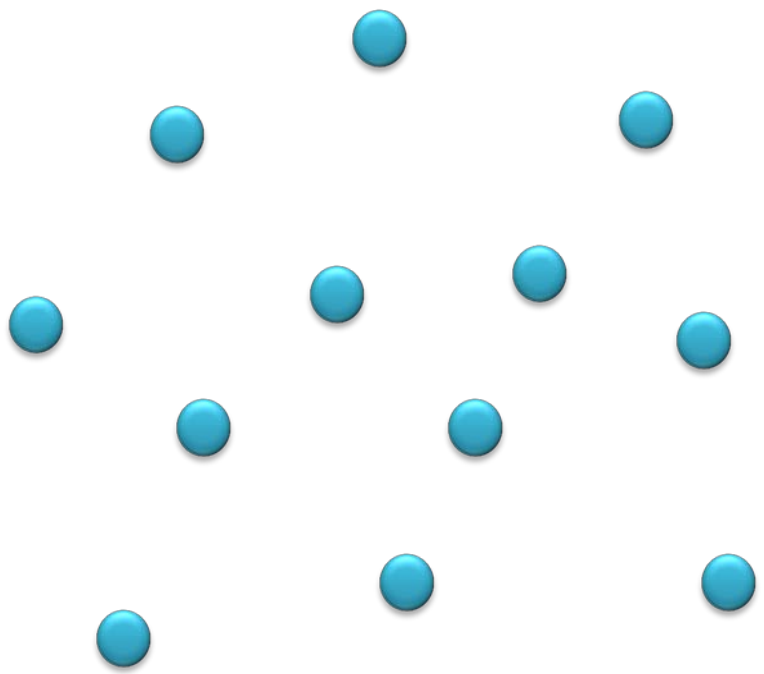
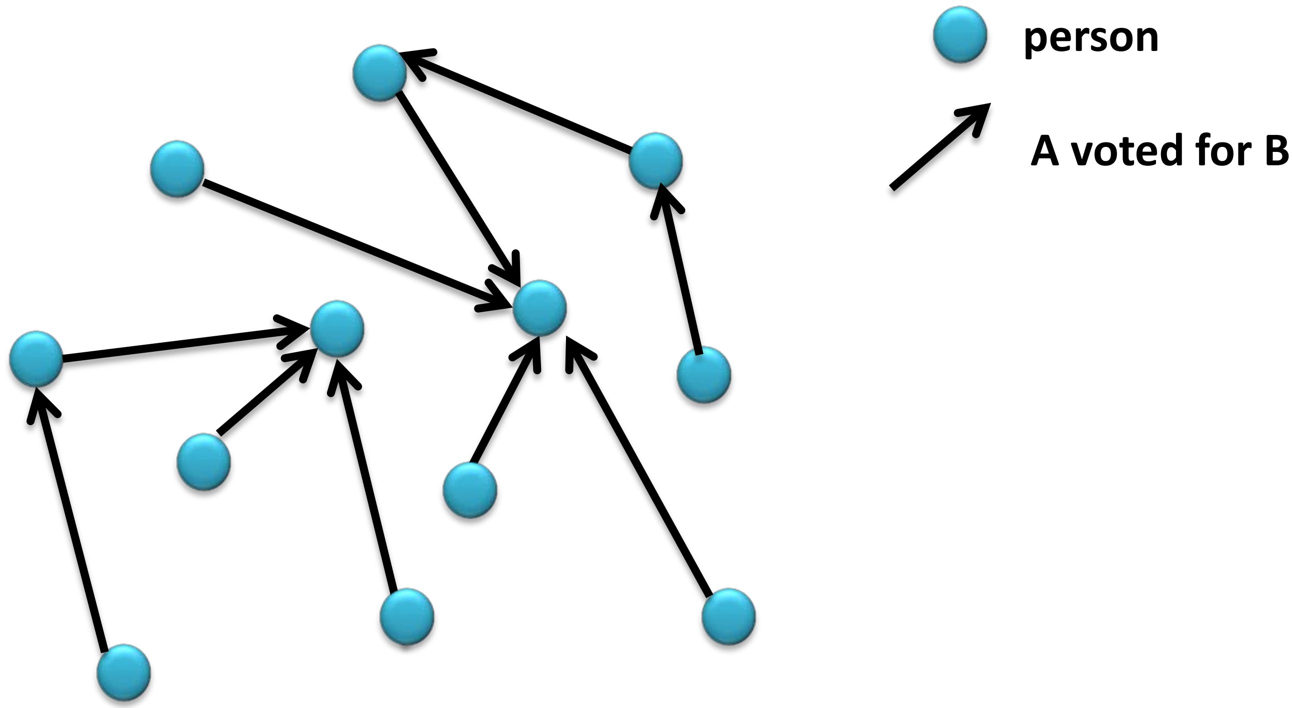
Complex Social System, Elections



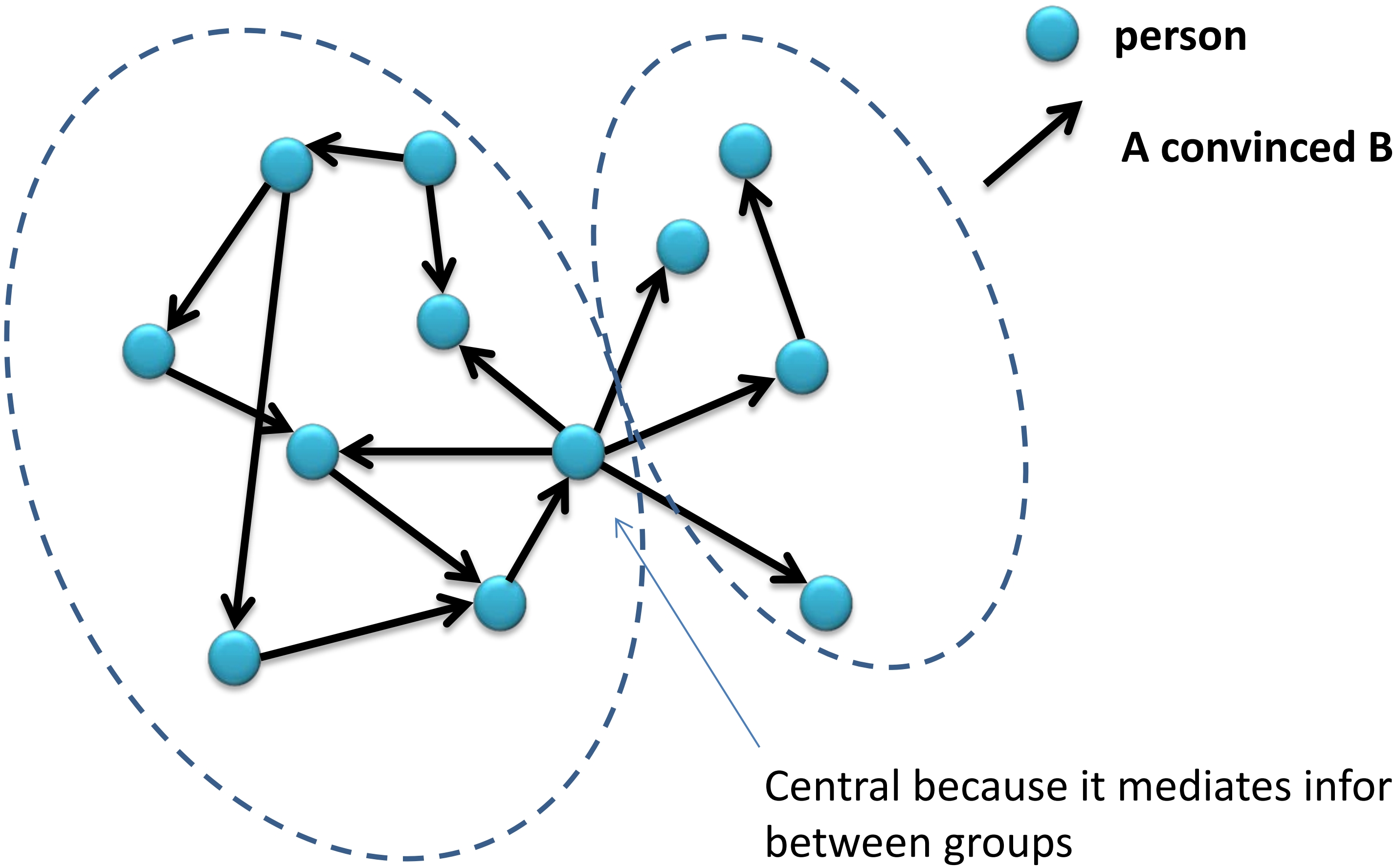
Centrality measures an importance of network's element (nodes, links, edges).

Complex Social System, Network I



A is more "central" than B if more people voted for A
In-degree centrality index = number of in-edges

Complex Social System, Network II

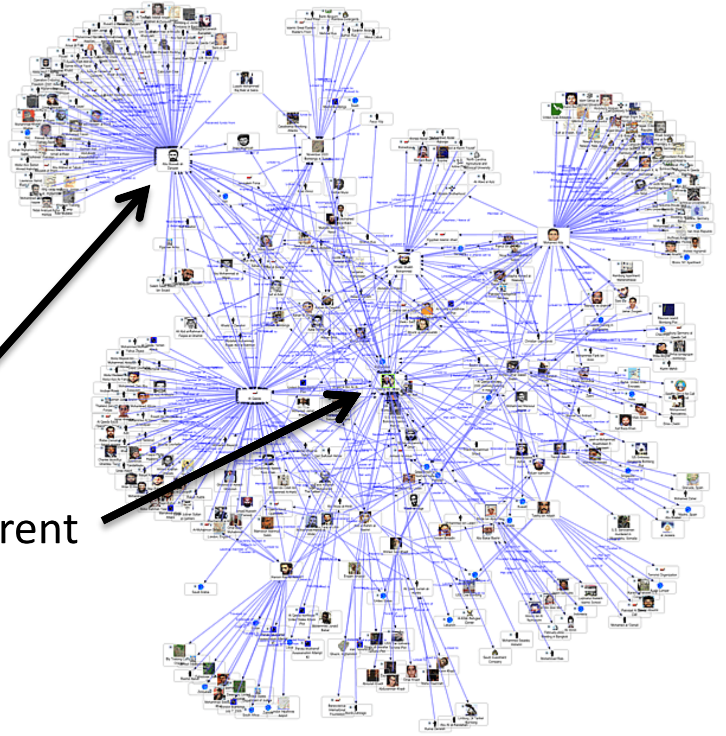


Central because it mediates information between groups
- Betweeness centrality (between groups)

Another example: friendship network, degree-based centrality

Measures and Metrics

Q: What are the most important nodes and edges?



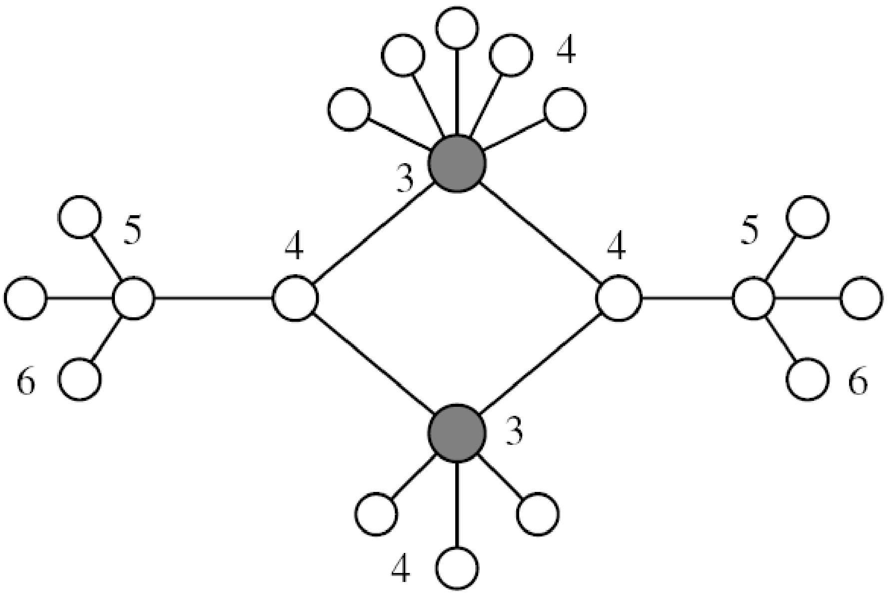
There are many different ways to quantify the importance

- Degree centrality: $\forall i \in V \;\; d(i)$
- Closeness centrality: $\forall i \in V \;\; l_i = \frac{1}{n} \sum_{j \in V} \delta_{ij}, \; \delta_{ij} = \text{distance from } i \text{ to } j$
- Edge removal centrality: $\forall e \in E \;\; r_e = \frac{\sum_{i,j \in V(G)} \delta_{ij}}{\sum_{i,j \in V(G_{|e})} \delta_{ij}}$

Can you come up with your interesting new measure of importance? Is it important for different types of networks? Can you compute it efficiently?

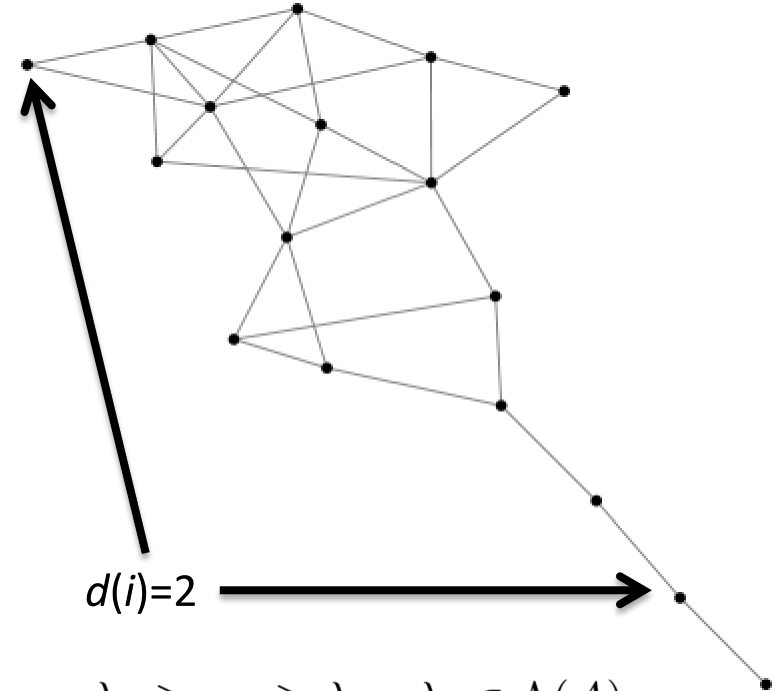
Eccentricity

$\forall i \in V$ we define $e(i) = \max\{\delta_{ij} | j \in V\}$
eccentricity centrality $c_E(i) = 1/e(i)$



U. Brandes "Network Analysis"

Eigenvector Centrality



- $\forall i \in V \;\; x_i^{(0)} = 1$
- $\forall i \in V \;\; x_i^{(1)} = \sum_j A_{ij} x_j$

`- - - - degree centrality - - - -`

Repeating last step *t* times

$$\mathbf{x}^{(t)} = A^t \mathbf{x}^{(0)}$$

Rewrite $x^{(0)}$ with linear combination of eigenvectors of *A*

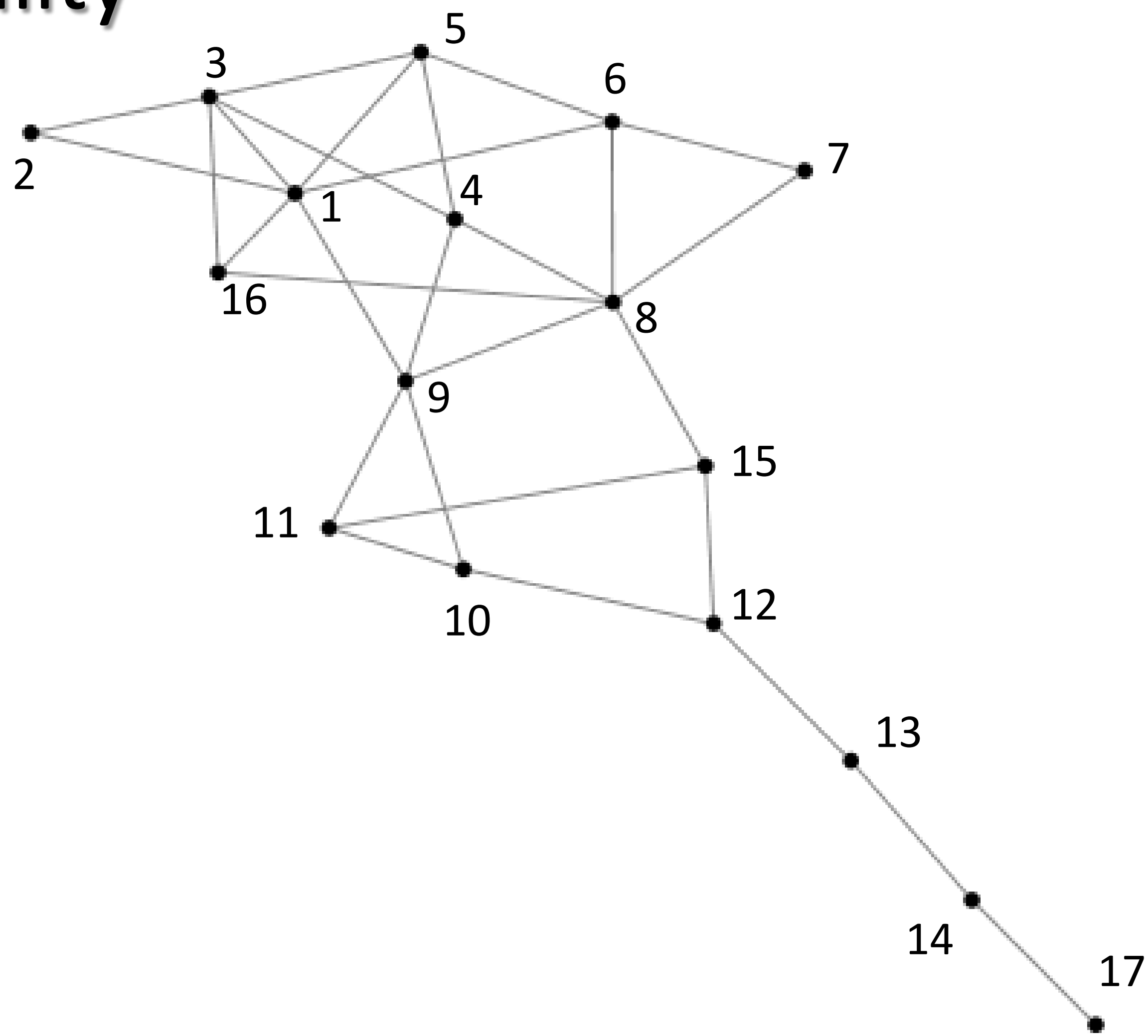
$$\mathbf{x}^{(t)} = A^t \sum_i c_i \mathbf{v}_i = \sum_i c_i \lambda_i^t \mathbf{v}_i = \lambda_1^t \sum_i c_i \left(\frac{\lambda_i}{\lambda_1}\right)^t \mathbf{v}_i, \text{ where } \lambda_1 \geq \cdots \geq \lambda_n, \;\; \lambda_i \in \Lambda(A)$$

$$\lim_{t \to T} \mathbf{x}^{(t)} = c_1 \lambda_1^T \mathbf{v}_1$$

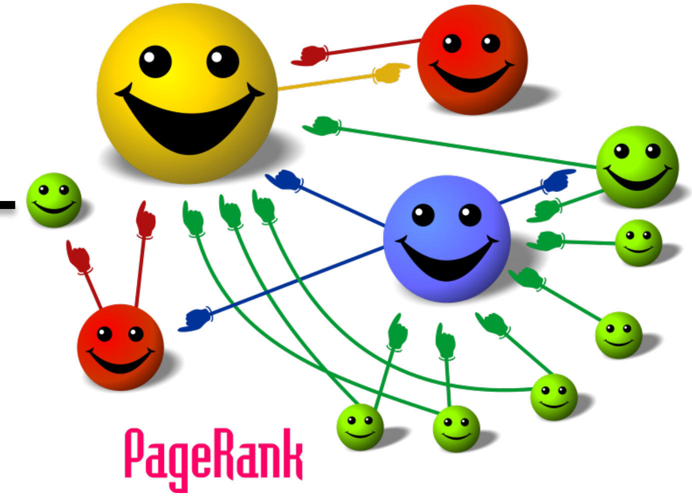
Eigenvector centrality is defined by values $x_i = \lambda_1^{-1} \sum_j A_{ij} x_j$

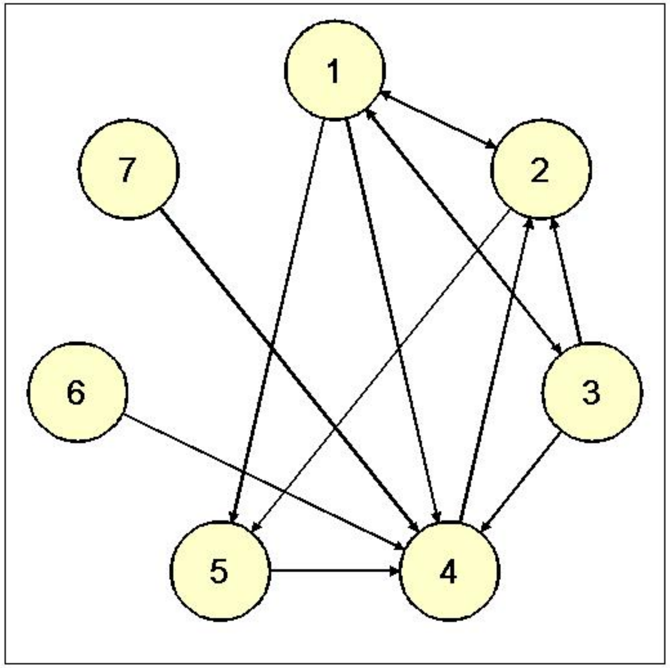
Eigenvector Centrality

1	0.120
3	0.103
8	0.101
5	0.095
4	0.093
9	0.092
6	0.085
16	0.077
2	0.053
7	0.044
11	0.039
15	0.038
10	0.035
12	0.018
13	0.005
14	0.001
17	0.000



PageRank





$$P = \begin{bmatrix} 0 & 0.50 & 0.33 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.25 & 0.00 & 0.33 & 0.50 & 0.00 & 0.00 & 0.00 \\ 0.25 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.25 & 0.00 & 0.33 & 0.50 & 1.00 & 1.00 & 1.00 \\ 0.25 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}$$

transition matrix

$x = (\frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7}, \frac{1}{7})$
$P^{20}x$ - see pagerank-example.m

Markov Chain: probabilities of visiting the pages after *k* steps is $P^k x$.

Problem: dangling nodes ($d^+(i)=0$)
Solution: damping factor α (usually small)

$$P' = (1-\alpha)P + \alpha T, \text{ where } T = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

regular random walk

teleportation

Note: *P′* is still a Markov matrix with no 0s

Eigenproblem-based Centralities: Computational Problems

- These problems are equivalent to solving eigenproblems, namely, *P'v = v*.
- Gaussian elimination is very expensive.

$$P' = (1-\alpha)P + \alpha T, \text{ where } T = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$$

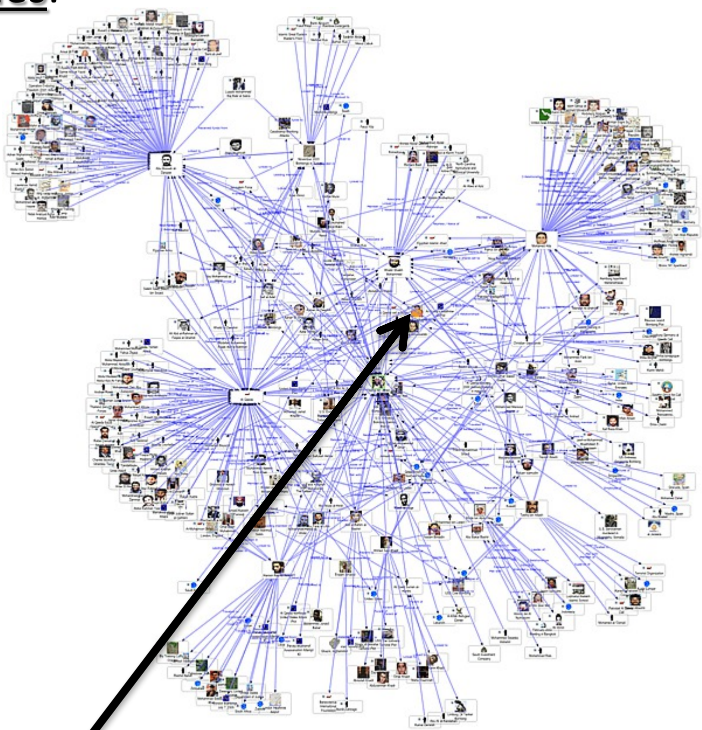
- Iterative computation of *P'v* is less expensive because *P' is* usually sparse.
- Principal eigenvector of *P'* is a PageRank. It can be solved by Power Method (iterative) and by Algebraic Multigrid.
- Parallel computation
- Sublinear computation
- Updating PageRank vector, evolving networks

Further reading: Berkhin ``A Survey on PageRank computing''
Langville, Meyer ``Deeper Inside PageRank"

Hubs and Authorities

Observation: In some networks a vertex may be important if it points to others with high centrality.

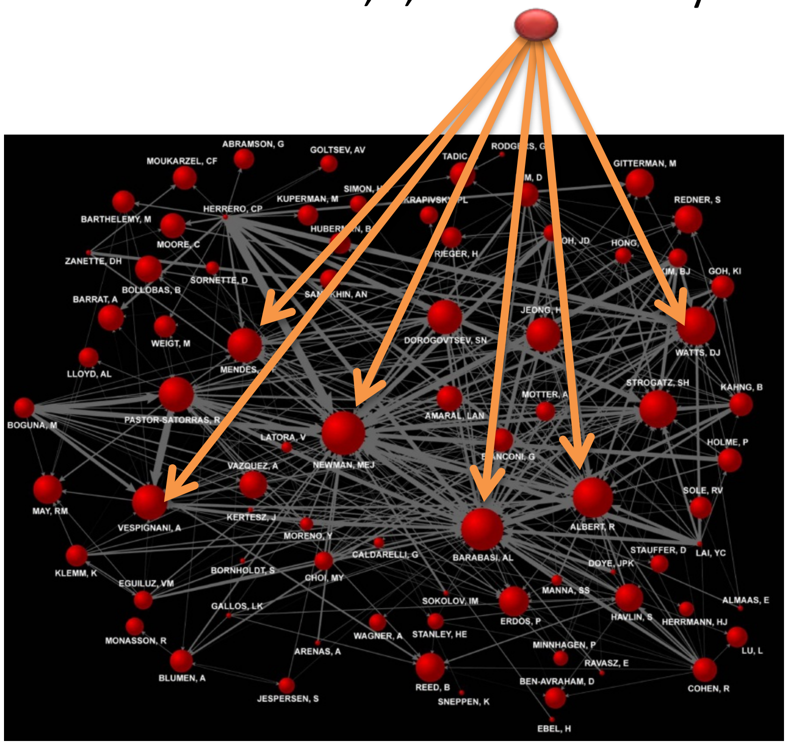
Examples:



9/11 contacts

X sends an information to many leaders but they don't know X. X is influential.

X, Y, Z "Cool Survey"



Scientific paper citations

Raddicchi et al. "Diffusion of scientific credits and the ranking of scientists"

Hubs and Authorities, HITS Algorithm

Authorities are nodes that contain useful information on a topic of interest.
Hubs are nodes that point to best authorities.

HITS Algorithm computes for each node *i* its authority and hub centralities x_i and y_i, respectively

$$x_i = \alpha \sum_j A_{ij} y_j \text{ and } y_i = \beta \sum_j A_{ji} x_j, \text{ where } \alpha,\ \beta \text{ are constants}$$
$$\implies AA^T x = \lambda x \text{ and } A^T A y = \lambda y, \text{ where } \lambda = (\alpha\beta)^{-1}$$

same leading (see eigenvector centrality) eigenvalue in both cases!

All eigenvalues of AA^T and A^TA are the same (check and prove!). Multiplying both sides of the first equation by A^T gives

$$A^T \cdot AA^T x = A^T \cdot \lambda x \implies A^T A(A^T x) = \lambda(A^T x) \implies y = A^T x$$

i.e., once we have a vector of authorities, the hub centrality can be calculated faster

Note: Authority centrality of A = Eigenvector centrality of co-citation matrix AA^T

Current applications: Teoma.com and Ask.com

Further reading: Kleinberg, "Authoritative sources in a hyper-linked environment"

Homework

Paper review

Jon Kleinberg "Authoritative Sources in a Hyperlinked Environment"

Submit your review at EasyChair by 1/27.
Note: Your review must be **detailed**, i.e., include summary, general comments, constructive criticism