

Node Centrality Metrics and Link Analysis

Leonid E. Zhukov

School of Applied Mathematics and Information Science
National Research University Higher School of Economics

11.09.2013



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Sociology. Linton Freeman, 1979.

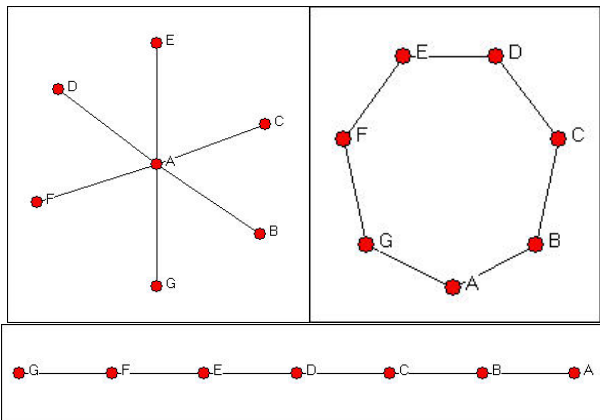
Most "important" actors: actor location in the social network

- Actor Centrality - involvement with other actors, many ties, source or recipient
- Actor Prestige - recipient (object) of many ties, ties directed to an actor

Three graphs:

- Star graph
- Circle graph
- Line graph

Three graphs



Degree Centrality

$$C_D(i) = k(i) = \sum_j A_{ij} = \sum_j A_{ji}$$

Normalized degree centrality

$$C_D^*(i) = \frac{1}{n-1} C_D(i)$$

High centrality degree - direct contact with many other actors

Low degree - not active, peripheral

Closeness Centrality

How close an actor to all the other actors in network

$$C_C(i) = \left[\sum_j d(i,j) \right]^{-1}$$

Normalized closeness centrality

$$C_C^*(i) = (n - 1)C_C(i)$$

Actor in the center can quickly interact with all others, short communication path to others, minimal number of steps to reach others

Betweenness Centrality

Betweenness Centrality

Number of shortest paths going through the actor $\sigma_{st}(i)$

$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Normalized betweenness centrality

$$C_B^*(i) = \frac{2}{(n-1)(n-2)} C_B(i)$$

Probability that a communication from s to t will go through i (geodesics)

Edge betweenness

Degree Prestige

$$P_D(i) = k_{in}(i) = \sum_j A_{ji}$$

Normalized degree prestige

$$P_D^*(i) = \frac{1}{n-1} P_D(i)$$

Prestigious actors receive many nominations

Proximity Prestige

Influence domain - set of actors that can reach i directly and indirectly.

l_i - size of influence domain. Average distance $\sum_j d(j, i)/l_i$

$$P_p(i) = \frac{l_i/(n-1)}{\sum_j d(j, i)/l_i}$$

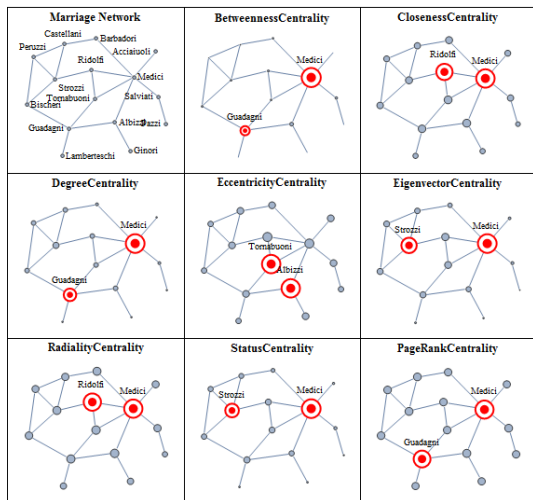
Centralization - how central the most central node in relation to all other nodes.

$$C_x = \frac{\sum_i^N [C_x(p_*) - C_x(p_i)]}{\max \sum_i^N [C_x(p_*) - C_x(p_i)]}$$

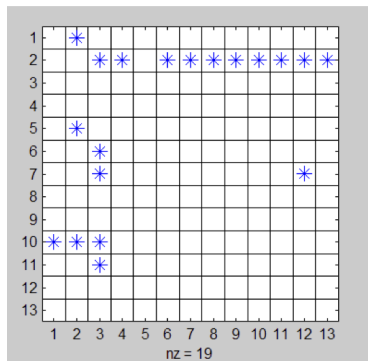
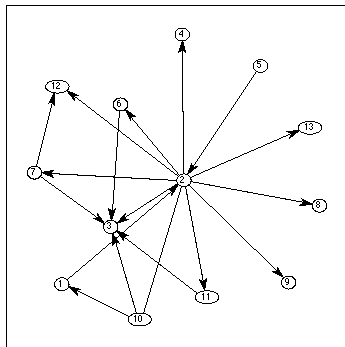
C_x - one of the centrality measures

Centrality and Prestige of Florentine Families

The Medici family marriage network



Graph $G(n, m)$ and adjacency matrix A_{ij} , edge $i \rightarrow j$



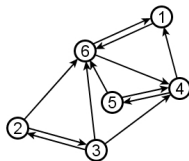
Status or Rank Prestige

Leo Katz, 1953.

Take into account status (prestige) of directly connected actors

$$p_i \leftarrow \sum_{j \in N(i)} p_j = \sum_j A_{ji} p_j$$

$$p_i = \sum_j A_{ji} p_j$$



$$\mathbf{p} = \mathbf{A}^T \mathbf{p}$$

$$(\mathbf{I} - \mathbf{A}^T) \mathbf{p} = \mathbf{0}$$

Nontrivial solution only if $\det(\mathbf{I} - \mathbf{A}^T) = 0$. Need to constraint matrix

Leo Katz, 1953

An actor gives equal parts of its prestige to all nearest neighbours

$$p_i \leftarrow \sum_j A_{ji} \frac{p_j}{k_{out}(j)} = \sum_j \frac{A_{ji}}{k_{out}(j)} p_j$$

$$\mathbf{p} = (\mathbf{D}^{-1}\mathbf{A})^T \mathbf{p}$$

where $\mathbf{D}_{ii} = \max(k_i, 1)$

$\mathbf{D}^{-1}\mathbf{A}$ - stochastic matrix, $\sum_j (\mathbf{D}^{-1}\mathbf{A})_{ij} = 1$, guaranteed $\lambda_{max} = 1$

$$(\mathbf{D}^{-1}\mathbf{A})^T \mathbf{p} = \lambda \mathbf{p}$$

Phillip Bonacich, 1972.

$$c_i \leftarrow \sum_j A_{ji} c_j$$

$$c_i = \frac{1}{\kappa} \sum_j A_{ji} c_j$$

$$\kappa \mathbf{c} = \mathbf{A}^T \mathbf{c}$$

$$(\kappa \mathbf{I} - \mathbf{A}^T) \mathbf{p} = 0$$

Nontrivial solution only when $\det(\kappa \mathbf{I} - \mathbf{A}^T) = 0$. Eigenvalue problem.

Choose eigenvector corresponding do maximum eigenvalue:

$$\kappa_{max} = \kappa_1, \mathbf{c} = \mathbf{c}_1$$

Status or Rank Prestige. Eigenvector Centrality

Phillip Bonacich, 1987.

Parametrized centrality measure $c(\alpha, \beta)$

$$c_i = \sum_j (\alpha + \beta c_j) A_{ji}$$

$$\mathbf{c} = \alpha \mathbf{A}^T \mathbf{e} + \beta \mathbf{A}^T \mathbf{c}$$

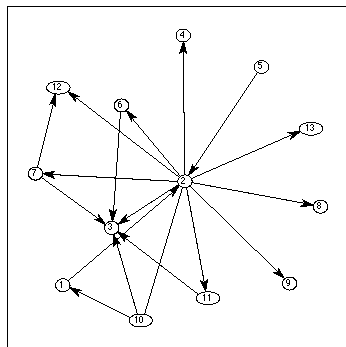
$$\mathbf{c} = \alpha (\mathbf{I} - \beta \mathbf{A}^T)^{-1} \mathbf{A}^T \mathbf{e}$$

α - found from normalization $\|\mathbf{c}\|_2 = \sum c_i^2 = 1$

β - parameter, degree and direction of dependence on others

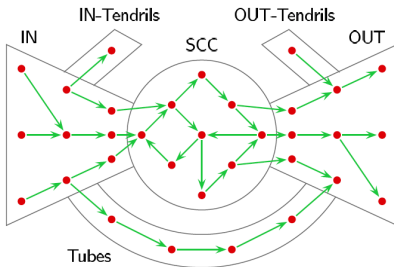
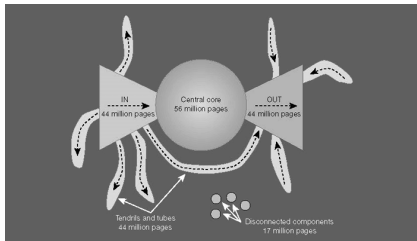
Link Analysis

Graph $G(n, m)$



- zero out degree nodes, $k_{out}(i) = 0$
- zero in degree nodes, $k_{in}(i) = 0$

Bow tie structure of the web



Perron-Frobenius Theorem

Oscar Perron, 1907, Georg Frobenius, 1912.

Eigenvalue problem:

$$\mathbf{P}\mathbf{p} = \lambda\mathbf{p}$$

Perron-Frobenius theorem: Real square matrix with positive entries

- stochastic (non-negative and rows sum up to one)
- irreducible (strongly connected graph)
- aperiodic

then unique largest eigenvalue $\lambda_{max} = 1$, with positive left eigenvector and power iterations converges to it. Solution satisfies $|\mathbf{p}|_1 = 1$

Stationary distribution of Markov chain

PageRank

Sergey Brin and Larry Page, 1998

Transition matrix:

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$$

Stochastic matrix:

$$\mathbf{P}' = \mathbf{P} + \frac{\mathbf{d}\mathbf{e}^T}{n}$$

PageRank matrix:

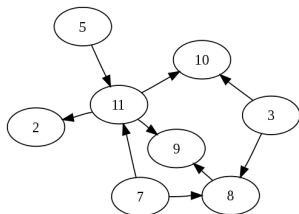
$$\mathbf{P}'' = \alpha\mathbf{P}' + (1 - \alpha)\frac{\mathbf{e}\mathbf{e}^T}{n}$$

Eigenvalue problem (choose solution with $\lambda = 1$):

$$\mathbf{P}''^T \mathbf{p} = \lambda \mathbf{p}$$

Notations:

\mathbf{e} - unit column vector, \mathbf{d} - absorbing nodes indicator vector (column)



- Eigenvalue problem

$$\left[\alpha \left((\mathbf{D}^{-1}\mathbf{A})^T + \frac{\mathbf{e}\mathbf{d}^T}{n} \right) + (1 - \alpha) \frac{\mathbf{e}\mathbf{e}^T}{n} \right] \mathbf{p} = \lambda \mathbf{p}$$

- Power iterations

$$\mathbf{p} \leftarrow \alpha (\mathbf{D}^{-1}\mathbf{A})^T \mathbf{p} + \alpha \frac{\mathbf{e}}{n} (\mathbf{d}^T \mathbf{p}) + (1 - \alpha) \frac{\mathbf{e}}{n} (\mathbf{e}^T \mathbf{p})$$

$$\mathbf{p} \leftarrow \frac{\mathbf{p}}{\|\mathbf{p}\|}$$

- Sparse linear system ($\lambda = 1$, $\|\mathbf{p}\|_1 = 1$)

$$\left[\mathbf{I} - \alpha \left((\mathbf{D}^{-1}\mathbf{A})^T + \frac{\mathbf{e}\mathbf{d}^T}{n} \right) \right] \mathbf{p} = (1 - \alpha) \frac{\mathbf{e}}{n}$$

Hubs and Authorities

HITS, Jon Kleinberg, 1999

Citation networks. Reviews vs original research (authoritative) papers

- authorities, contain useful information, a_i
- hubs, contains links to authorities, h_i

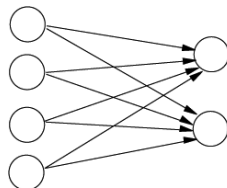
Mutual recursion

- Good authorities referred by good hubs

$$a_i \leftarrow \sum_j A_{ji} h_j$$

- Good hubs point to good authorities

$$h_i \leftarrow \sum_j A_{ij} a_j$$



System of linear equations

$$\mathbf{a} = \alpha \mathbf{A}^T \mathbf{h}$$

$$\mathbf{h} = \beta \mathbf{A} \mathbf{a}$$

Symmetric eigenvalue problem

$$(\mathbf{A}^T \mathbf{A}) \mathbf{a} = \lambda \mathbf{a}$$

$$(\mathbf{A} \mathbf{A}^T) \mathbf{h} = \lambda \mathbf{h}$$

where eigenvalue $\lambda = (\alpha\beta)^{-1}$

- A new status index derived from sociometric analysis, L. Katz, Psychometrika, 19, 39-43, 1953.
- Power and centrality: A family of measures, P. Bonacich, American Journal of Sociology, 92,1170-1182,1987.
- Graph structure in the Web, Andrei Broder et al, Procs of the 9th international World Wide Web conference on Computer networks, 2000.
- The PageRank Citation Ranknig: Bringing Order to the Web, S. Brin, L. Page, R. Motwany, T. Winograd, Stanford Digital Library Technologies Project, 1998
- Authoritative Sources in a Hyperlinked Environment, Jon M. Kleinberg, Proc. 9th ACM-SIAM Symposium on Discrete Algorithms,

- Centrality in Social Networks. Conceptual Clarification, Linton C. Freeman, *Social Networks*, 1, 215-239, 1979
- Power and Centrality: A Family of Measures, Phillip Bonacich, *The American Journal of Sociology*, Vol. 92, No. 5 (Mar., 1987), 1170-1182.