Node centrality and ranking on networks

Lecture 4

Natalia Semenova

Higher School of Economics

nasemenova_2@edu.hse.ru

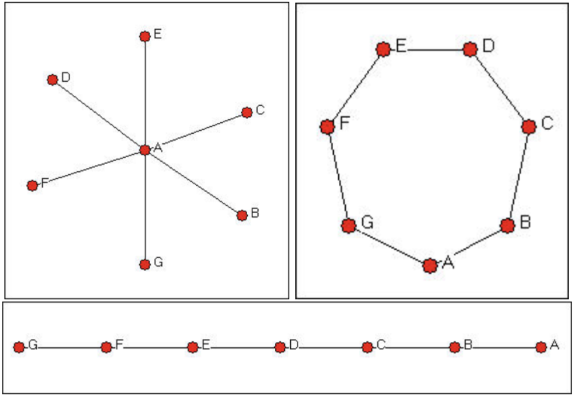
Structural Analysis and Visualization of Networks
23.01.2024

Lecture Overview

Node centrality

Which vertices are important?

Which one is important?



Star graph Circle graph Line Graph

Degree centrality

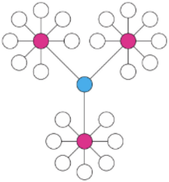
Degree centrality: number of nearest neighbours

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

Normalized degree centrality

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

High centrality degree -direct contact with many other actors



Closeness centrality

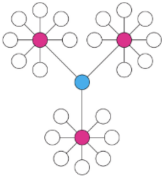
Closeness centrality: how close an actor to all the other actors in network

$$C_C(i) = \frac{1}{\sum_j d(i,j)}$$

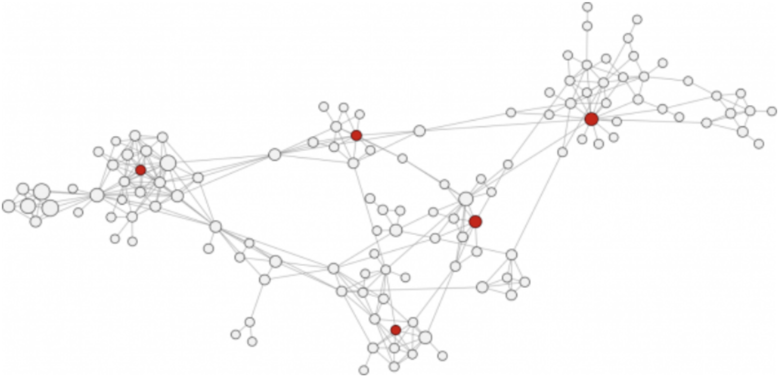
Normalized closeness centrality

$$C_C^*(i) = (n-1)C_C(i) = \frac{n-1}{\sum_j d(i,j)}$$

High closeness centrality - short communication path to others, minimal number of steps to reach all others



Closeness centrality: example



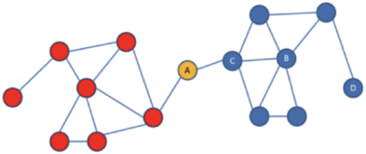
Betweenness centrality

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

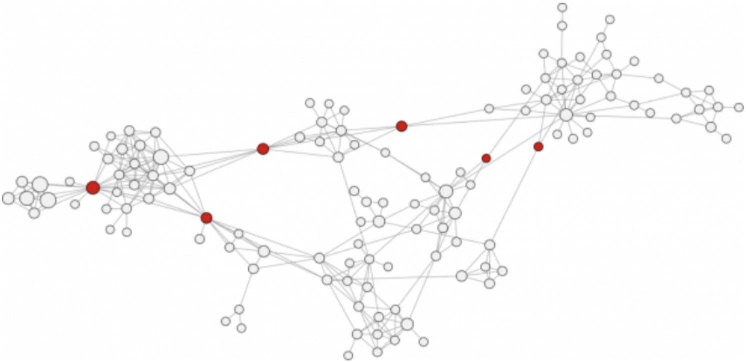
$$C_B^*(i) = \frac{2}{(n-1)(n-2)} C_B(i) = \frac{2}{(n-1)(n-2)} \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$



High betweenness centrality - vertex lies on many shortest paths

Linton Freeman, 1977

Betweenness centrality: example



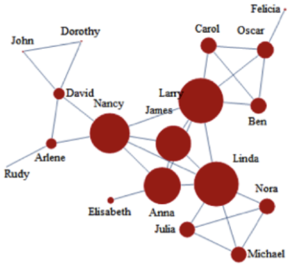
Eigenvector centrality

Importance of a node depends on the importance of its neighbors (recursive definition)

$$v_i \leftarrow \sum_j A_{ij} v_j$$

$$v_i = \frac{1}{\lambda} \sum_j A_{ij} v_j$$

$$\mathbf{Av} = \lambda \mathbf{v}$$



Select an eigenvector associated with largest eigenvalue $\lambda = \lambda_1$, $\mathbf{v} = \mathbf{v}_1$

Eigenvector centrality: example

Katz status index

Weighted count of all paths coming to the node: the weight of path of length n is counted with attenuation factor β^n, $\beta < \frac{1}{\lambda_1}$

$$k_i = \beta \sum_j A_{ij} + \beta^2 \sum_j A_{ij}^2 + \beta^3 \sum_j A_{ij}^3 + \ldots$$

$$\mathbf{k} = (\beta \mathbf{A} + \beta^2 \mathbf{A}^2 + \beta^3 \mathbf{A}^3 + \ldots)\mathbf{e} = \sum_{n=1}^{\infty} (\beta^n \mathbf{A}^n)\mathbf{e} = (\sum_{n=0}^{\infty} (\beta \mathbf{A})^n - \mathbf{I})\mathbf{e}$$

$$\sum_{n=0}^{\infty} (\beta \mathbf{A})^n = (\mathbf{I} - \beta \mathbf{A})^{-1}$$

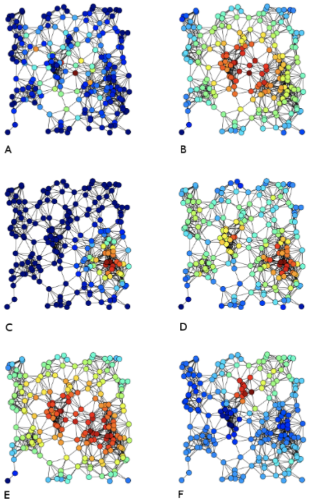
$$\mathbf{k} = ((\mathbf{I} - \beta \mathbf{A})^{-1} - \mathbf{I})\mathbf{e}$$

$$(\mathbf{I} - \beta \mathbf{A})\mathbf{k} = \beta \mathbf{A}\mathbf{e}$$

$$\mathbf{k} = \beta \mathbf{A}\mathbf{k} + \beta \mathbf{A}\mathbf{e}$$

$$\mathbf{k} = \beta (\mathbf{I} - \beta \mathbf{A})^{-1} \mathbf{A}\mathbf{e}$$

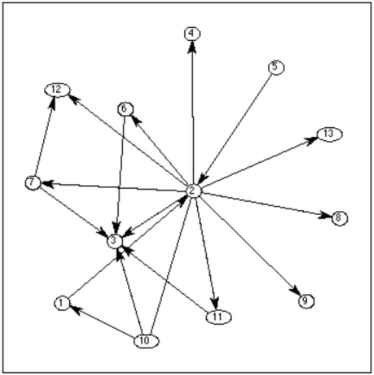
Centrality examples



from Wikipedia

- A) Betweenness centrality
- B) Closeness centrality
- C) Eigenvector centrality
- D) Degree centrality
- F) Harmonic centrality
- E) Katz centrality

Directed graphs



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

Centrality measures for directed graphs

All based on outgoing edges

- Degree centrality (normalized):

$$C_D^{in}(i) = \frac{k^{in}(i)}{n-1}; \quad C_D^{out}(i) = \frac{k^{out}(i)}{n-1}$$

- **Closeness centrality (normalized):

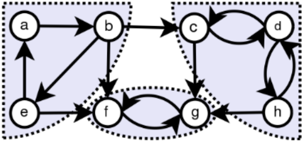
$$C_C(i) = \frac{n-1}{\sum_j d(i,j)}$$

- **Betweenness centrality (normalized):

$$C_B(i) = \frac{1}{(n-1)(n-2)} \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Connectivity

- Graph is **strongly connected** if every vertex is reachable form every other vertex.
- **Strongly connected components** are partitions of the graph into subgraphs that are strongly connected



- In strongly connected graphs there is a path is each direction between any two pairs of vertices

image from Wikipedia

Aperiodicity

- A directed graph is **aperiodic** if the greatest common divisor of the lengths of its cycles is one (there is no integer $k > 1$ that divides the length of every cycle of the graph)

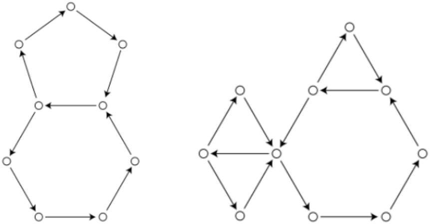
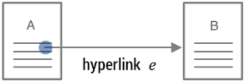


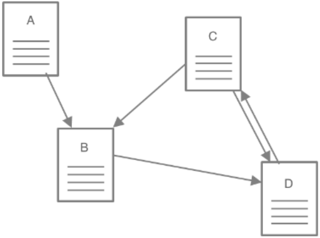
image from Wikipedia

Web as a graph

- Hyperlinks - implicit endorsements

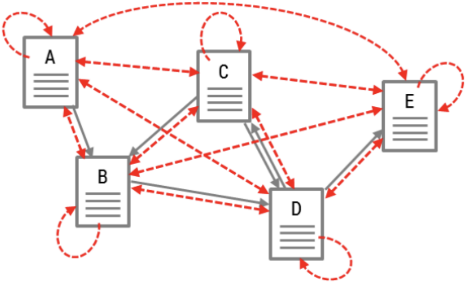


- Web graph - graph of endorsements (sometimes reciprocal)



Page Rank

"PageRank can be thought of as a model of user behavior. We assume there is a "random surfer" who is given a web page at random and keeps clicking on links, never hitting "back" but eventually gets bored and starts on another random page. The **probability** that the random surfer visits a page is its **PageRank**."



Sergey Brin and Larry Page, 1998

Random walk

- Random walk on a directed graph:

$$p_i^{t+1} = \sum_{j \in N(i)} \frac{p_j^t}{d_j^{out}} = \sum_j \frac{A_{ji}}{d_j^{out}} p_j$$

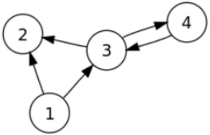
$$\mathbf{D}_{ii} = diag\{d_i^{out}\}$$

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

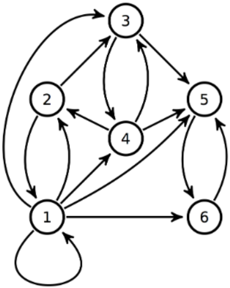
$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$$

- Power iterations

$$\mathbf{p}^{t+1} \leftarrow \mathbf{P}^T \mathbf{p}^t$$



a)



b)

PageRank formulation

- Power iterations:

$$\mathbf{p} \leftarrow \alpha \mathbf{P}^T \mathbf{p} + (1-\alpha)\frac{\mathbf{e}}{n}, \quad \alpha \text{ - teleportation coefficient}$$

- Sparse linear system:

$$(\mathbf{I} - \alpha \mathbf{P}^T)\mathbf{p} = (1-\alpha)\frac{\mathbf{e}}{n}$$

- Eigenvalue problem ($\lambda = 1$):

$$\left(\alpha \mathbf{P}^T + (1-\alpha)\mathbf{E}\right)\mathbf{p} = \lambda \mathbf{p}$$

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

Perron-Frobenius Theorem

Perron-Frobenius theorem (Fundamental Theorem of Markov Chains)

If matrix is

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

then

$$\exists \lim_{t\to\infty} \bar{\mathbf{p}}^t = \bar{\pi}$$

and can be found as a left eigenvector

$$\bar{\pi}P = \lambda\bar{\pi}, \text{ where } ||\bar{\pi}||_1 = 1, \lambda = 1$$

$\bar{\pi}$ - stationary distribution of Markov chain, row vector

Oscar Perron, 1907, Georg Frobenius,1912.

PageRank variations

- Power iterations

$$\mathbf{p} \leftarrow \alpha \mathbf{P}^T \mathbf{p} + (1-\alpha)\mathbf{v}, \quad \mathbf{v} \text{ - teleportation vector}$$

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

$$\mathbf{p} \leftarrow \mathbf{P}'^T \mathbf{p}, \ ||\mathbf{p}|| = 1$$

- Topic specific PageRank

v - set of pages on specific topics

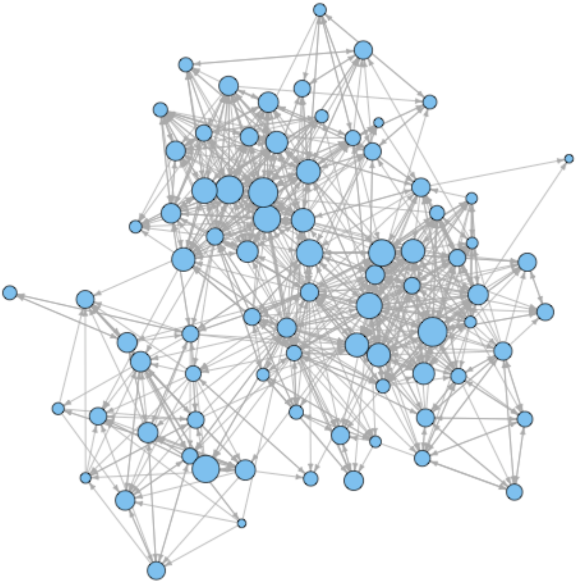
- TrustRank

v - set of trusted pages

- Personalized PageRank

v - set of personal preference pages

PageRank



PageRank beyond the Web

1. GeneRank
2. ProteinRank
3. FoodRank
4. SportsRank
5. HostRank
6. TrustRank
7. BadRank
8. ObjectRank
9. ItemRank
10. ArticleRank
11. BookRank
12. FutureRank
13. TimedPageRank
14. SocialPageRank
15. DiffusionRank
16. ImpressionRank
17. TweetRank
18. TwitterRank
19. ReversePageRank
20. PageTrust
21. PopRank
22. CiteRank
23. FactRank
24. InvestorRank
25. ImageRank
26. VisualRank
27. QueryRank
28. BookmarkRank
29. StoryRank
30. PerturbationRank
31. ChemicalRank
32. RoadRank
33. PaperRank
34. Etc...

Hubs and Authorities (HITS)

Citation networks. Reviews vs original research (authoritative) papers

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

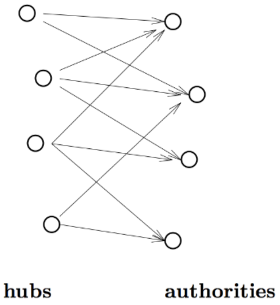
Mutual recursion

- Good authorities reffered 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

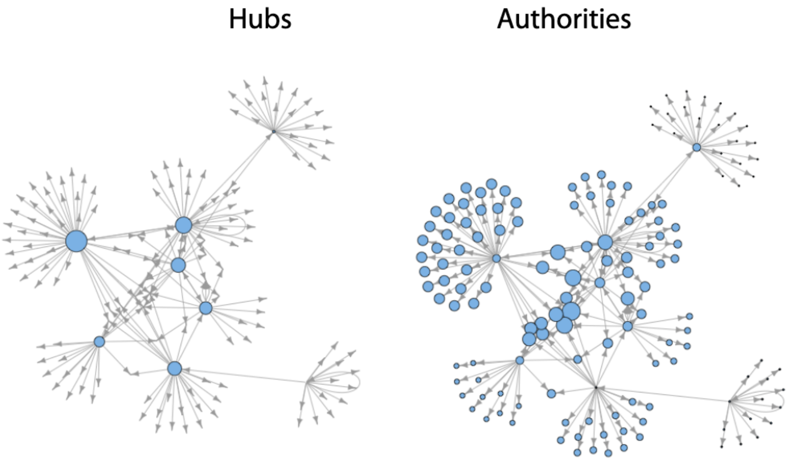
$$
\begin{aligned}
\mathbf{a} &= \alpha \mathbf{A}^T \mathbf{h} \\
\mathbf{h} &= \beta \mathbf{A} \mathbf{a}
\end{aligned}
$$

Symmetric eigenvalue problem

$$
\begin{aligned}
(\mathbf{A}^T \mathbf{A})\mathbf{a} &= \lambda \mathbf{a} \\
(\mathbf{A} \mathbf{A}^T)\mathbf{h} &= \lambda \mathbf{h}
\end{aligned}
$$

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

Hubs and Authorities



References

- 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, 1170-1182, 1987
- A new status index derived from sociometric analysis, L. Katz, Psychometrika, 19, 39-43, 1953.
- Eigenvector-like measures of centrality for asymmetric relations, Phillip Bonacich, Paulette Lloyd, Social Networks 23, 191-201, 2001
- The PageRank Citation Ranknig: Bringing Order to the Web. S. Brin, L. Page, R. Motwany, T. Winograd, Stanford Digital Library Technologies Project, 1998

References

- Authoritative Sources in a Hyperlinked Environment. Jon M. Kleinberg, Proc. 9th ACM-SIAM Symposium on Discrete Algorithms
- Graph structure in the Web, Andrei Broder et all. Procs of the 9th international World Wide Web conference on Computer networks, 2000
- A Survey of Eigenvector Methods of Web Information Retrieval. Amy N. Langville and Carl D. Meyer, 2004
- PageRank beyond the Web. David F. Gleich, arXiv:1407.5107, 2014