Node Centrality and Ranking on Networks

Social Network Analysis. MAGoLEGO course.

Lecture 4

Leonid Zhukov

lzhukov@hse.ru
www.leonidzhukov.net/hse/2019/sna

National Research University Higher School of Economics
School of Data Analysis and Artificial Intelligence, Department of Computer Science

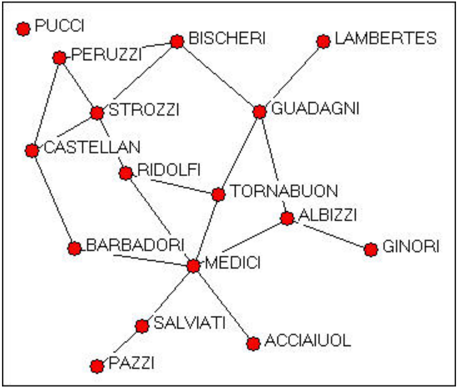
Centrality

Which vertices are important?

image from M.Grandjean, 2014

Centrality Measures

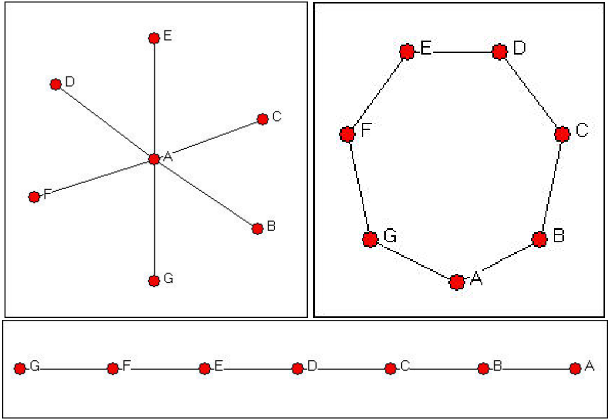
Determine the most ”important” or ”prominent” actors in the network based on actor location.



Marriage alliances among leading Florentine families 15th century.

Padgett, 1993

Three graphs



Star graph　　Circle graph　　Line Graph

Degree centrality

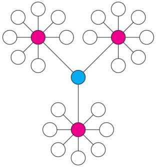
Degree centrality: number of nearest neighbors

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

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

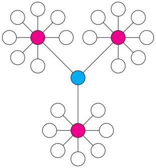
Closeness centrality: how close an actor to all the other actors in network

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

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 others



Betweenness centrality

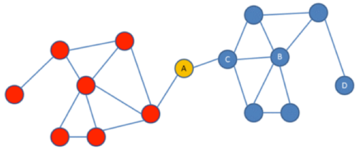
Betweenness centrality: number of shortest paths going through the actor $\sigma_{st}(i)$

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

Normalized betweenness centrality

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

Hight betweenness centrality - vertex lies on many shortest paths
Probability that a communication from *s* to *t* will go through *i*



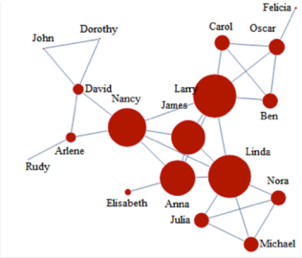
Eigenvector centrality

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

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

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

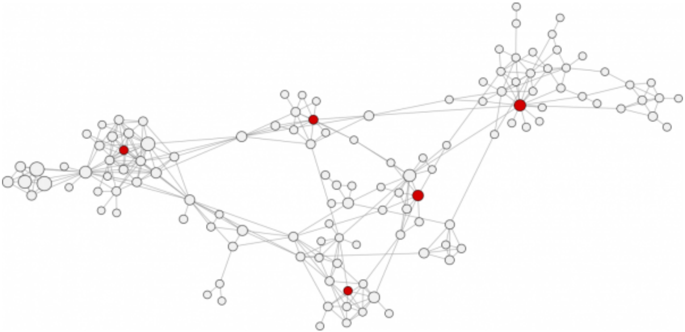
$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$



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

Centrality examples

Closeness centrality

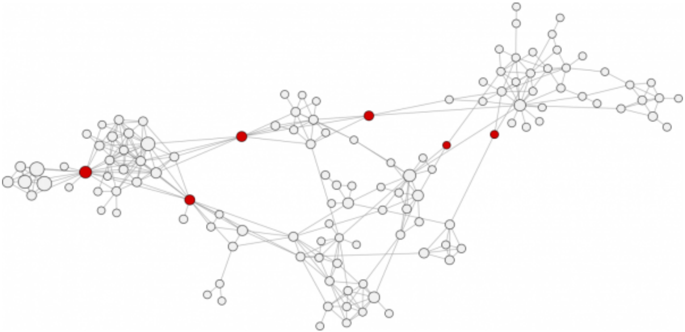


igraph:closeness()

from www.activenetworks.net

Centrality examples

Betweenness centrality



igraph:betweenness()

from www.activenetworks.net

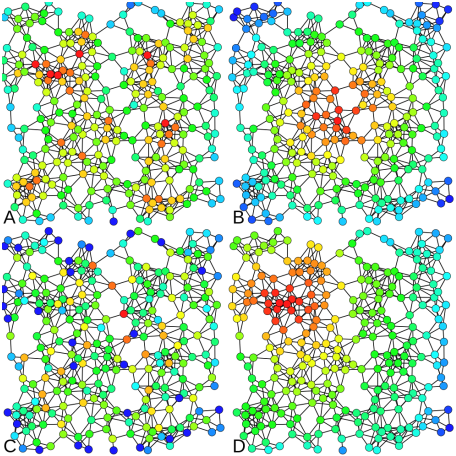
Centrality examples

Eigenvector centrality

igraph:evcent()

from www.activenetworks.net

Centrality examples



from Claudio Rocchini

- A) Degree centrality
- B) Closeness centrality
- C) Betweenness centrality
- D) Eigenvector centrality

Centralization

Centralization (network measure) - how central the most central node in the network in relation to all other nodes.

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

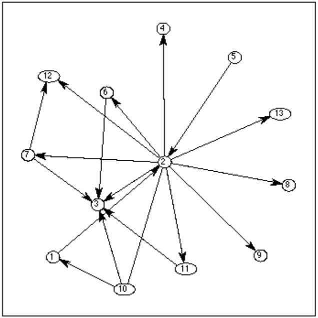
C_x - one of the centrality measures
p_* - node with the largest centrality value
max - is taken over all graphs with the same number of nodes (for degree, closeness and betweenness the most centralized structure is the star graph)

igraph: centralization.degree(), centralization.closeness(), centralization.betweenness(), centralization.evcent()

Linton Freeman, 1979

Directional relations

Directed graph: distinguish between choices made (outgoing edges) and choices received (incoming edges)



sending - receiving
export - import
cite papers - being cited

Centrality measures

All based on outgoing edges

- Degree centrality (normalized):

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

- Closeness centrality (normalized):

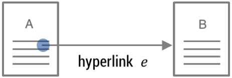
$$C_C^*(i) = \frac{n-1}{\sum_j d(i,j)}$$

- **Betweenness centrality (normalized):

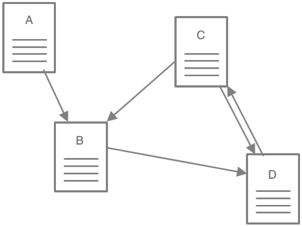
$$C_B^*(i) = \frac{1}{(n-1)(n-2)} \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Web as a graph

- Hyperlinks - implicit endorsements

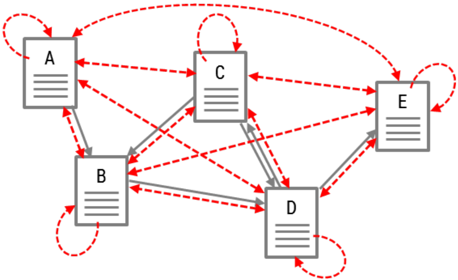


- Web graph - graph of endorsements (sometimes reciprocal)



PageRank

”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

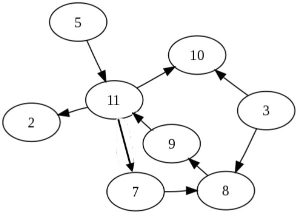
Random walk

- Random walk on graph

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

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

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}, \;\; \mathbf{D}_{ii} = diag\{d_i^{out}\}$$



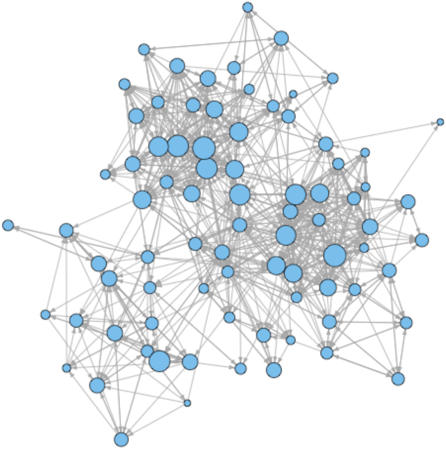
- with teleportation

$$\mathbf{p}^{t+1} = \alpha \mathbf{P}^T \mathbf{p}^t + (1-\alpha)\mathbf{v}$$

Perron-Frobenius Theorem guarantees existence and uniqueness of the solution $\lim_{t\to\infty} \mathbf{p} = \pi$

$$\pi = \alpha \mathbf{P}^T \pi + (1-\alpha)\mathbf{v}$$

PageRank



igraph: page.rank()

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

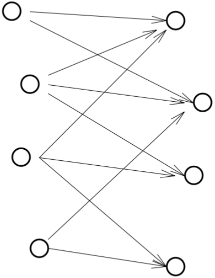
Mutual recursion

- Good authorities referred by good hubs

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

- Good hubs point to good authorities

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



HITS

System of linear equations

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

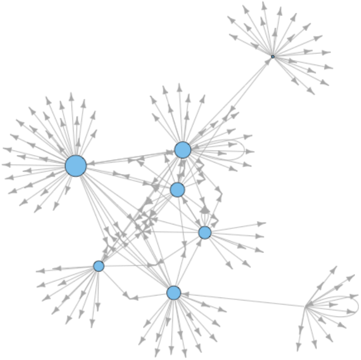
Symmetric eigenvalue problem

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

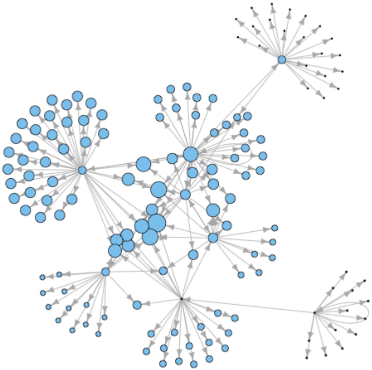
where eigenvalue $\lambda = (\alpha\beta)^{-1}$

Hubs and Authorities

Hubs

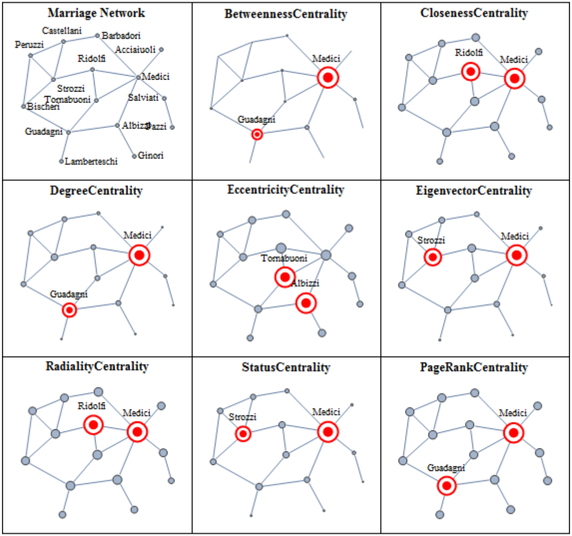


Authorities



igraph: hub.score(), authority.score()

Florentines families



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
- Authoritative Sources in a Hyperlinked Environment. Jon M. Kleinberg, Proc. 9th ACM-SIAM Symposium on Discrete Algorithms
- 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