node *centrality*

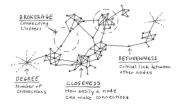
introduction to *network science in Python* (NetPy)

Lovro Šubelj University of Ljubljana 3rd October 2024

centrality *measures*

which *nodes* are most *important*?

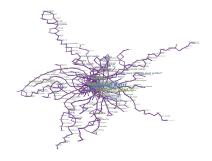
- node centrality measures for (un)directed networks
 - clustering coefficients [WS98, SV05, dNMB05]
 - geodesic-based measures [Fre77, FBW91, New05]
 - spectral analysis measures [Kat53, Bon87, BP98]



— link analysis algorithms primarily for directed networks

networkology LPP

- partial LPP public bus transport network*
- n = 416 bus stops with $\langle k \rangle = 5.62$ connections
- giant component 95.4% nodes (6 components)
- "small-world" with $\langle C \rangle = 0.09$ and $\langle d \rangle = 14.26$
- "scale-free" with $\gamma = 2.62$ for cutoff $k_{min} = 5$



^{*} reduced to largest connected component

centrality clustering

important nodes are strongly embedded

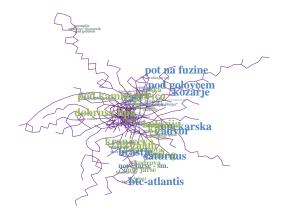
- for undirected G clustering coefficient C [WS98] of i is
 - t_i is number of *linked neighbors* or *triangles* of i

$$C_i = \frac{2t_i}{k_i(k_i-1)}$$
 $C_i = 0$ for $k_i \le 1$

— C fails for hub nodes in scale-free networks [dNMB05]

networkology *clustering*

- clustering coefficient C in partial LPP network[†]
- highest $C_i = 1.0$ nodes are Na Žalah etc. with $k_i = 2$



[†]reduced to simple undirected graph

centrality *closeness*

important *nodes* are *close to other* nodes

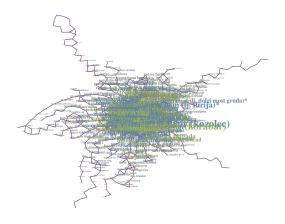
- for (un) directed G closeness centrality ℓ^{-1} [New10] of i is
 - d_{ij} is (un) directed distance between i and j
 - $d_{ij} = \infty$ for nodes in different components

$$\ell_i^{-1} = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d_{ij}}$$

— ℓ^{-1} spans *small range* in *small-world* networks

networkology *closeness*

- closeness centrality ℓ^{-1} in partial LPP network[‡]
- highest $\ell_i^{-1} = 0.208$ node is Gosposvetska with $k_i = 14$



[‡]reduced to simple undirected graph

centrality betweenness

important *nodes* are *bridges between other* nodes

- for (un)directed G betweenness centrality σ [Fre77] of i is
 - g_{st} is number of shortest paths between s and t
 - $-g_{st}^{i}$ is number of such shortest paths through i

$$\sigma_i = \frac{1}{n^2} \sum_{st} \frac{g_{st}^i}{g_{st}}$$

— σ considers *only shortest paths* [FBW91, New05]

networkology betweenness

- betweenness centrality σ in partial LPP network§
- highest $\sigma_i = 0.235$ node is Razstavišče with $k_i = 11$



[§] reduced to simple undirected graph

centrality degrees

important nodes are linked by many nodes

— for undirected G degree centrality d of i is $d_i = \frac{1}{n-1} \sum_{j \neq i} A_{ij} = \frac{k_i}{n-1}$

— in directed G in-degree centrality d^{in} of i is

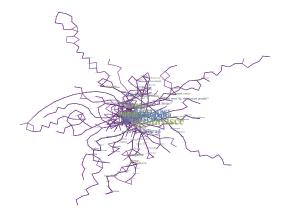
$$d_i^{in} = \frac{1}{n-1} \sum_{j \neq i} A_{ij} = \frac{k_i^{in}}{n-1}$$

— in directed G out-degree centrality dout of i is

$$d_i^{out} = \frac{1}{n-1} \sum_{j \neq i} A_{ji} = \frac{k_i^{out}}{n-1}$$

networkology *degrees*

- degree centrality d in partial LPP network
- highest $d_i = 0.099$ node is Razstavišče with $k_i = 41$
- highest d_i node is Razstavišče with $k_i^{in} = 20$ and $k_i^{out} = 21$



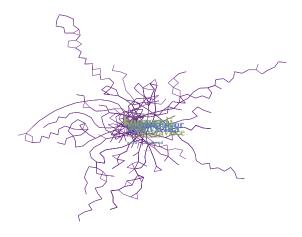
centrality eigenvector

important nodes are linked by important nodes

- for (un)directed G eigenvector centrality e [Bon87] of i is e is leading eigenvector v_1 of A with eigenvalue λ_1^{-1} $e_i = \lambda_1^{-1} \sum_j A_{ij} e_j$
- in directed G = 0 for $k^{in} = 0$ nodes etc.

networkology eigenvector

- eigenvector centrality e in partial LPP network
- highest $e_i = 0.082$ node is Konzorcij with $k_i = 30$



centrality Katz

nodes get small amount of importance for free

- for (un) directed G Katz centrality z [Kat53] of i is
 - α and β are appropriate positive constants

$$\mathbf{z}_i = \alpha \sum_j A_{ij} \mathbf{z}_j + \beta$$

- for *convenience* $\beta = 1$ whereas $\alpha < \lambda_1^{-1}$
 - $-\lambda_1$ is leading eigenvalue of A for eigenvector v_1

centrality PageRank

nodes distribute equal amount of importance

— for (un)directed G PageRank centrality p [BP98] of i is $-\alpha$ and β are appropriate positive constants

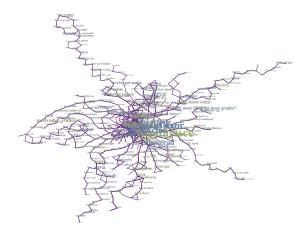
$$p_i = \alpha \sum_j A_{ij} \frac{p_j}{k_j} + \beta$$

- for *convenience* $\beta = \frac{1-\alpha}{n}$ whereas $\alpha = 0.85$
- p probability of random surfer with teleports

see PageRank algorithm NetLogo demo

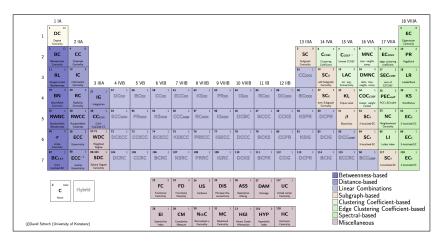
networkology PageRank

- PageRank centrality p in partial LPP network
- highest $p_i = 0.011$ node is Razstavišče with $k_i = 41$



centrality overview

which *nodes* are most *important*?



centrality references



Phillip Bonacich.

Power and centrality: A family of measures.

American Journal of Sociology, 92(5):1170–1182, 1987.



S. Brin and L. Page.

The anatomy of a large-scale hypertextual Web search engine. Comput. Networks ISDN. 30(1-7):107-117, 1998.



Wouter de Nooy, Andrei Mrvar, and Vladimir Batageli.

Exploratory Social Network Analysis with Pajek. Cambridge University Press, Cambridge, 2005.



David Easley and Jon Kleinberg.

Networks, Crowds, and Markets: Reasoning About a Highly Connected World. Cambridge University Press, Cambridge, 2010.



Ernesto Estrada and Philip A. Knight.

A First Course in Network Theory.
Oxford University Press, 2015.



Linton C. Freeman, Stephen P. Borgatti, and Douglas R. White.

Centrality in valued graphs: A measure of betweenness based on network flow. Soc. Networks, 13(2):141–154, 1991.



L. Freeman.

A set of measures of centrality based on betweenness. Sociometry, 40(1):35–41, 1977.



Leo Katz.

A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

centrality references



M. E. J. Newman.

A measure of betweenness centrality based on random walks. Soc. Networks, 27(1):39–54, 2005.



Mark E. J. Newman.

Networks: An Introduction.
Oxford University Press, Oxford, 2010.



Sara Nadiv Soffer and Alexei Vázquez.

Network clustering coefficient without degree-correlation biases. *Phys. Rev. E*, 71(5):057101, 2005.



D. J. Watts and S. H. Strogatz.

Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.