node *centrality*

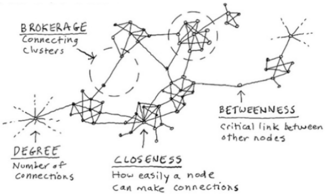
introduction to *network analysis* (*ina*)

Lovro Šubelj
University of Ljubljana
spring 2022/23

centrality *measures*

which *nodes* are most *important*?

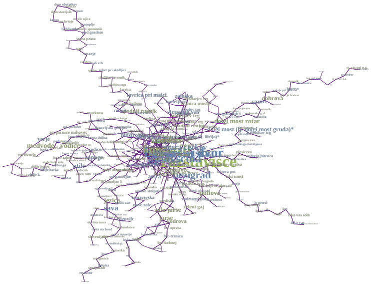
- *node centrality measures* for (*un*)*directed* networks
  - *clustering coefficients* [WS98, SV05, dNMB05]
  - *distance-based* measures [Fre77, FBW91, New05]
  - *spectral analysis* measures [Kat53, Bon87, BP98]
  - *fragment-based* measures [MSOI$^+$02, Prž07, EK15]



- *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

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

— ω-*corrected clustering coefficient* C^ω [SV05] of i is

  – ω_i is *maximum possible* t_i with *respect to* $\{k\}$

$$C_i^\omega = \frac{t_i}{\omega_i} \qquad C_i^\omega = 0 \text{ for } \omega_i = 0$$

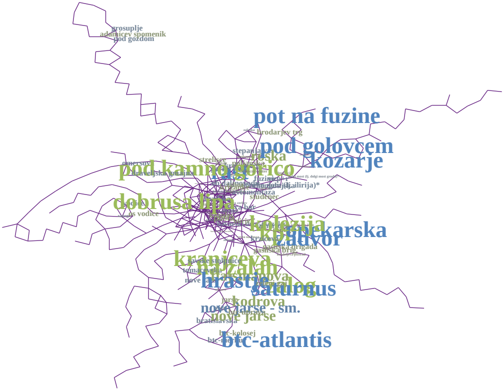
— μ-*corrected clustering coefficient* C^μ [dNMB05] of i is

  – μ is *maximum* number of *triangles* over *links*

$$C_i^\mu = \frac{2t_i}{k_i\mu} \qquad C_i^\mu = 0 \text{ for } k_i = 0$$

networkology *clustering*

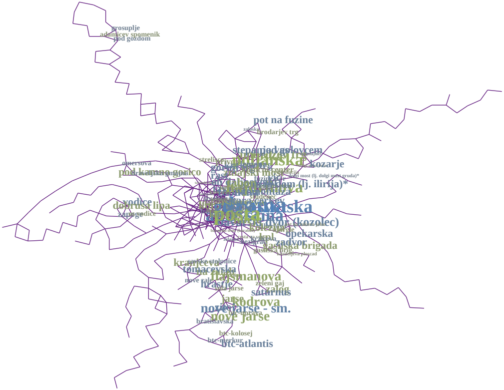
- *clustering coefficient* C in partial LPP network[†]
- *highest* $C_i = 1.0$ nodes are *Na Žalah etc.* with $k_i = 2$



[†] reduced to simple undirected graph

networkology *μ-clustering*

- *μ-corrected clustering C^{μ}* in partial LPP network‡
- *highest* $C_i^{\mu} = 0.44$ node is *Drama* with $k_i = 10$



‡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

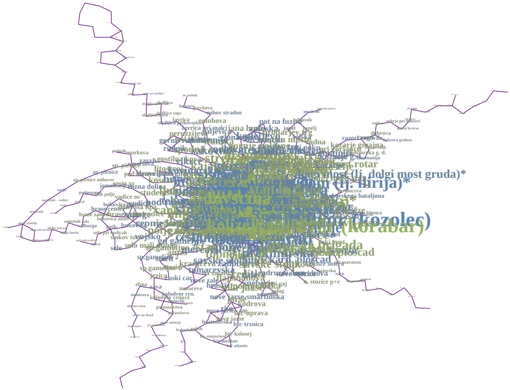
- d_{ij} is (*un*)*directed distance* between i and j
- $d_{ij} = \infty$ for nodes in *different components*

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

— 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

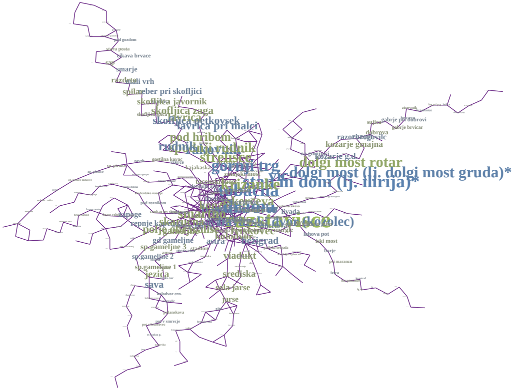
$$\sigma_i = \frac{1}{n^2} \sum_{st} \frac{g_{st}^i}{g_{st}}$$

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

— σ mixes *local centers* with *global bridges* [JMK+16]

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

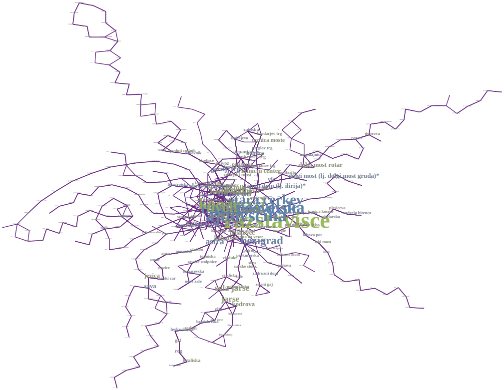
$$d_i^{in} = \frac{1}{n-1} \sum_{j \neq i} A_{ij} = \frac{k_i^{in}}{n-1}$$

— in *directed* G *out-degree centrality* d^{out} 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^{in} 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
  - v and λ are *eigenvectors* and *eigenvalues* of A
  - e is *proportional* to *leading eigenvector* v_1

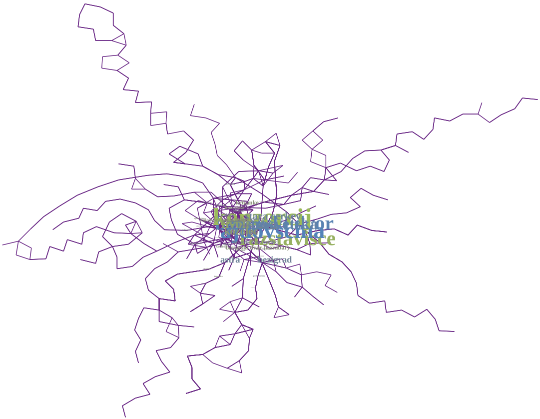
$$e(t) = A^t e(0) = A^t \sum_i C_i v_i = \sum_i C_i \lambda_i^t v_i = \lambda_1^t \sum_i C_i \left[\frac{\lambda_i}{\lambda_1}\right]^t v_i \to C_1 \lambda_1^t v_1$$

$$e_i = \lambda_1^{-1} \sum_j A_{ij} e_j$$

- in *directed* G $e = 0$ 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_i are some *positive constants*

$$z_i = \alpha \sum_j A_{ij} z_j + \beta_i$$

- for *convenience* $\beta_i = 1$ whereas $\alpha < \lambda_1^{-1}$
  - λ_1 is *leading eigenvalue* of A

centrality *PageRank*

nodes distribute equal amount of *importance*

— for (*un*)*directed* G *PageRank centrality* p [BP98] of i is

- α and β_i are some *positive constants*

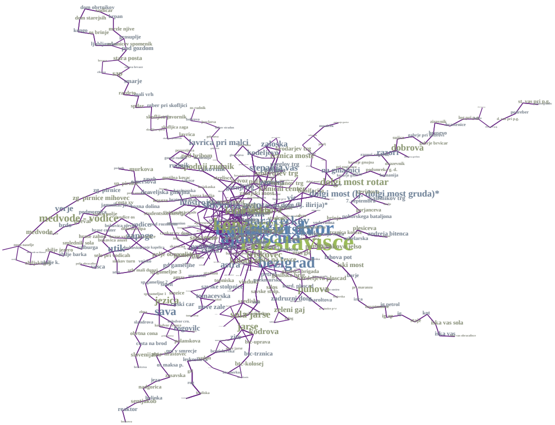
$$p_i = \alpha \sum_j A_{ij} \frac{p_j}{k_j^{out}} + \beta_i$$

— for *convenience* $\beta_i = \frac{1-\alpha}{n}$ whereas $\alpha = 0.85$

see PageRank algorithm NetLogo demo

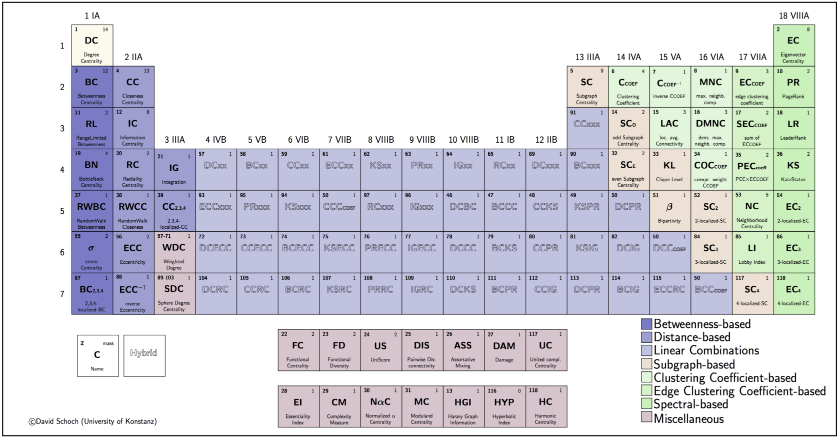
networkology *PageRank*

- *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*

A.-L. Barabási.
Network Science.
Cambridge University Press, Cambridge, 2016.

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, Andrej Mrvar, and Vladimir Batagelj.
Exploratory Social Network Analysis with Pajek.
Cambridge University Press, Cambridge, 2005.

Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj.
Exploratory Social Network Analysis with Pajek: Expanded and Revised Second Edition.
Cambridge University Press, Cambridge, 2011.

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.

centrality *references*

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

Pablo Jensen, Matteo Morini, Marton Karsai, Tommaso Venturini, Alessandro Vespignani, Mathieu Jacomy, Jean-Philippe Cointet, Pierre Merckle, and Eric Fleury.
Detecting global bridges in networks.
J. Complex Netw., 4(3):319–329, 2016.

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

R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon.
Network motifs: Simple building blocks of complex networks.
Science, 298(5594):824–827, 2002.

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.

Mark E. J. Newman.
Networks.
Oxford University Press, Oxford, 2nd edition, 2018.

centrality *references*

Nataša Pržulj.
Biological network comparison using graphlet degree distribution.
Bioinformatics, 23(2):e177–e183, 2007.

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.