

node *centrality*

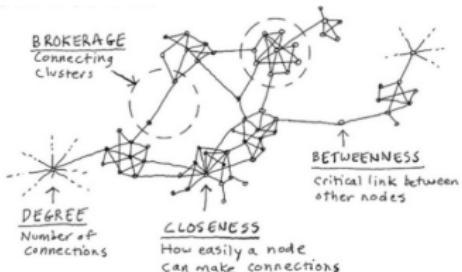
advanced topics in *network science* (*ants*)

Lovro Šubelj & Jure Leskovec
University of Ljubljana
spring 2019/20

centrality *measures*

which *nodes* are most *important*?

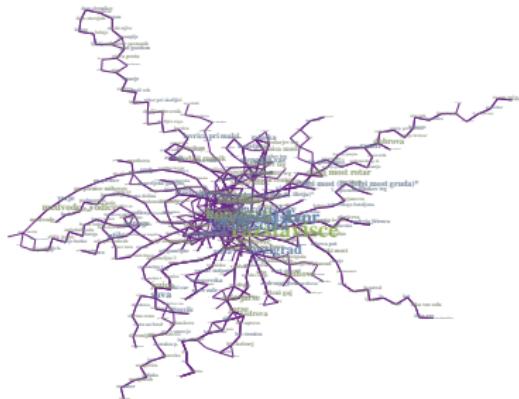
- *node centrality measures* for (*un*)*directed* networks
 - *clustering coefficients* [WS98, SV05, dNMB05]
 - *distance-based* centrality [Fre77, FBW91, New05]
 - *spectral analysis* centrality [Kat53, Bon87, BP98]
 - *fragment-based* centrality [MSOI⁺02, Prž07, EK15]



- *link analysis algorithms* for *directed* networks

networkology *LPP*

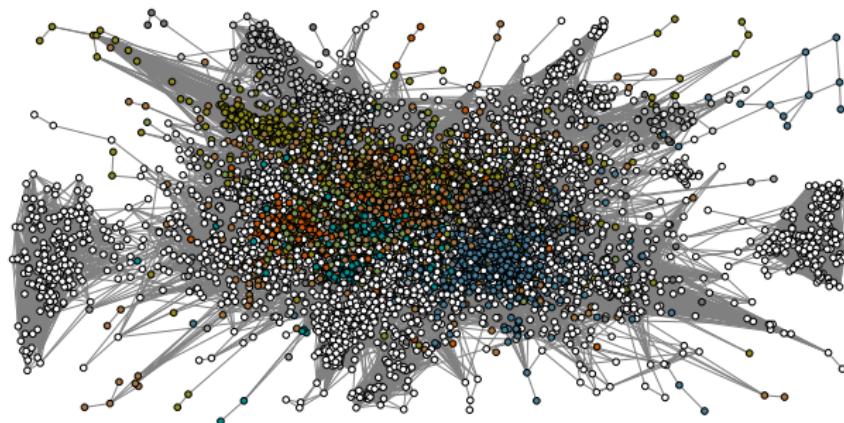
- partial *LPP public bus transport network**
 - $n = 416$ bus stops with $\langle k \rangle = 5.62$ connections
 - giant component 95.4% nodes (6 components)
 - “small-world” with $\langle C \rangle = 0.09$ and $\langle d \rangle = 14.26$
 - “scale-free” with $\gamma = 2.62$ for cutoff $k_{min} = 5$



* reduced to largest connected component

networkology *iMDB*

- *iMDB movie actors collaboration network*
- $n = 17577$ actors with $\langle k \rangle = 32.7$ collaborations
- *giant component* 99.3% nodes (19 components)
- *small-world* with $\langle C \rangle = 0.34$ and $\langle d \rangle = 4.82$
- *scale-free* with $\gamma = 2.21$ for cutoff $k_{min} = 25$



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*

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

- ω -corrected clustering coefficient C^ω [SV05] of *i* is
 - ω_i is *maximum possible* t_i with respect to $\{k\}$

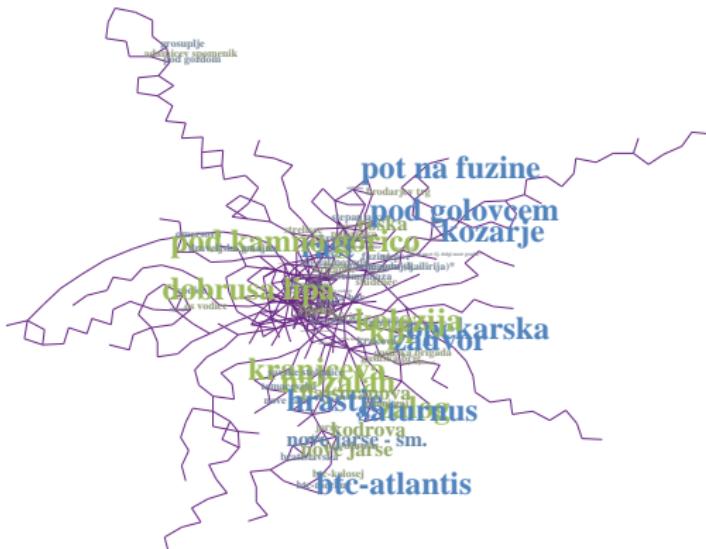
$$C_i^\omega = \frac{t_i}{\omega_i} \quad C_i^\omega = 0 \text{ for } \omega_i = 0$$

- μ -corrected clustering coefficient C^μ [dNMB05] of *i* is
 - μ is *maximum* number of *triangles* over *links*

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

networkkology *LPP*

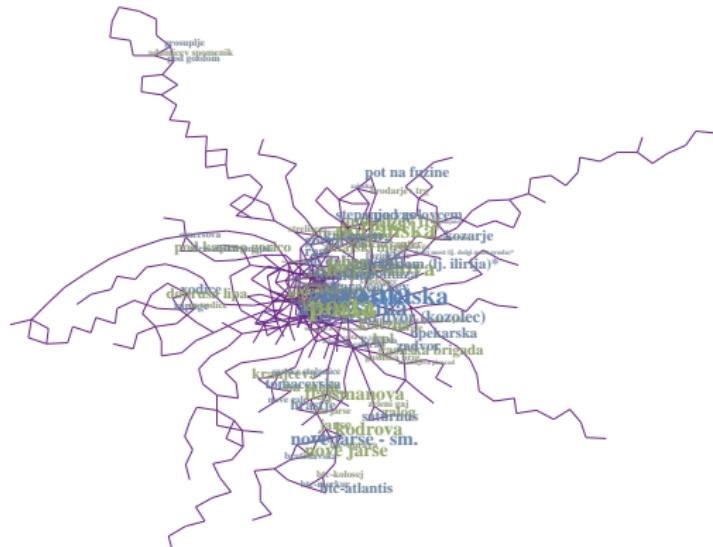
- clustering coefficient C in partial LPP network[†]
- highest $C_i = 1.0$ nodes are *Na Žalah* etc. with $k_i = 2$



[†]reduced to simple undirected graph

networkology *LPP*

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



\dagger reduced to simple undirected graph

networkology *iMDB*

- clustering coefficient C in iMDB network
- highest C nodes are less known actors

#	actor	k	C
1	Rachner, Jonathan	46	1.0000
2	Doucette, Jeff	23	1.0000
3	Willis, Susan	20	1.0000
4	Andersson, Kris	20	1.0000
5	Kurtis, Bill	19	1.0000
6	Cantillana, Nestor	16	1.0000
7	Tolkien, J.R.R.	15	1.0000
8	Faris, Anna	15	1.0000
9	Kurata, Tetsuo	13	1.0000
10	Margera, Jess	12	1.0000
11	Rakeyohn	12	1.0000
12	Raab, Chris	12	1.0000

networkology *iMDB*

- μ -corrected clustering C^μ in iMDB network
- highest C^μ nodes are *wrestling actors/models*

#	actor	k	C^μ
1	Helms, Shane	216	0.3689
2	Wilson, Torrie	217	0.3688
3	Matthews, Darren	212	0.3688
4	Greenwald, Nora	215	0.3687
5	Keibler, Stacy	215	0.3687
6	Jindrak, Mark	214	0.3683
7	Wight, Paul	221	0.3677
8	Guerrero Jr., Chavo	215	0.3676
9	Bischoff, Eric	218	0.3675
10	Hugger, John	212	0.3675
11	Flair, Ric	219	0.3672
12	Layfield, John	192	0.3667

centrality *closeness*

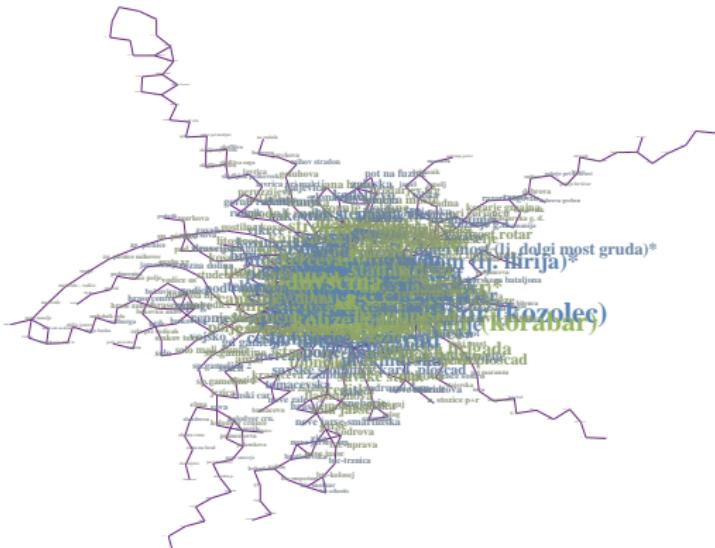
important *nodes* are *close to other* nodes

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

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

networkology *LPP*

- *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

networkology *iMDB*

- *closeness centrality* ℓ^{-1} in iMDB network ¶
- *highest* ℓ^{-1} nodes are *Hollywood actors*

#	actor	k	ℓ^{-1}
1	Goldberg, Whoopi	398	0.3506
2	Hanks, Tom	457	0.3500
3	Jackson, Samuel L.	427	0.3497
4	Berry, Halle	376	0.3468
5	Diaz, Cameron	361	0.3459
6	Stiller, Ben	382	0.3452
7	Lopez, Jennifer	410	0.3427
8	Myers, Mike	345	0.3409
9	Douglas, Michael	263	0.3403
10	Cruise, Tom	336	0.3401
11	Travolta, John	335	0.3388
12	Schwarzenegger, Arnold	333	0.3385

¶ reduced to largest connected component

centrality *betweenness*

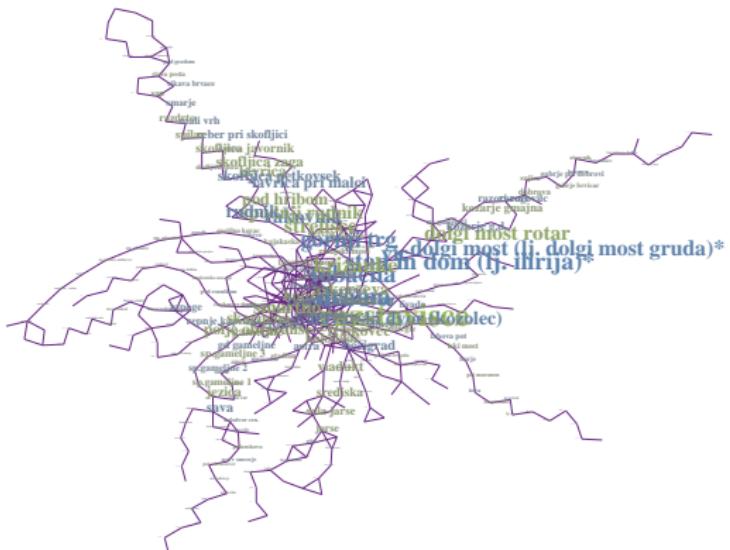
important *nodes* are *bridges for other* nodes

- for (*un*)directed G betweenness centrality σ [Fre77] of i is
 - g_{st} is number of geodesic paths between s and t
 - g_{st}^i is number of such geodesic paths through i
- σ considers *only* geodesic paths [FBW91, New05]
- σ mixes *local centers* with *global bridges* [JMK⁺16]

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

networkology *LPP*

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



|| reduced to simple undirected graph

networkology *iMDB*

- *betweenness centrality* σ in iMDB network**
- *highest* σ nodes are *international actors*

#	actor	k	σ
1	Jeremy, Ron	471	0.0640
2	Chan, Jackie	135	0.0310
3	Cruz, Penelope	182	0.0284
4	Shahlavi, Darren	8	0.0282
5	Del Rosario, Monsour	6	0.0280
6	Depardieu, Gerard	159	0.0265
7	Bachchan, Amitabh	66	0.0169
8	Jackson, Samuel L.	427	0.0167
9	Soualem, Zinedine	121	0.0155
10	Del Rio, Olivia	168	0.0152
11	Jaenicke, Hannes	73	0.0140
12	Hayek, Salma	185	0.0139

** reduced to largest connected component

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 dⁱⁿ* of *i* is

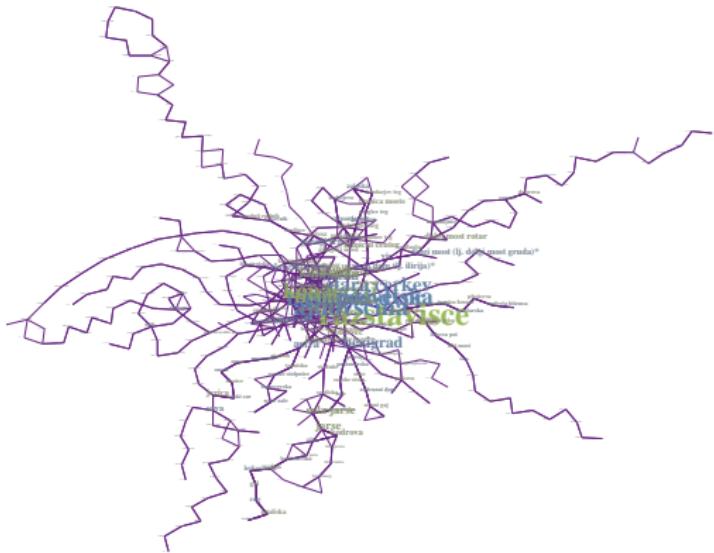
$$d_i^{in} = \frac{1}{n-1} \sum_{j \neq i} A_{ij} = \frac{k_i^{in}}{n-1}$$

- 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 *LPP*

- 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 $k_i^{out} = 21$



networkology *iMDB*

- *degree centrality d* in iMDB network
- *highest d* nodes are *pornographic actors*

#	actor	k	d
1	Davis, Mark	784	0.0446
2	Sanders, Alex	610	0.0347
3	North, Peter	599	0.0341
4	Marcus, Mr.	584	0.0332
5	Tedeschi, Tony	561	0.0319
6	Dough, Jon	555	0.0316
7	Stone, Lee	545	0.0310
8	Voyeur, Vince	533	0.0303
9	Lawrence, Joel	500	0.0284
10	Steele, Lexington	493	0.0280
11	Ashley, Jay	490	0.0279
12	Boy, T.T.	475	0.0270

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 v₁*

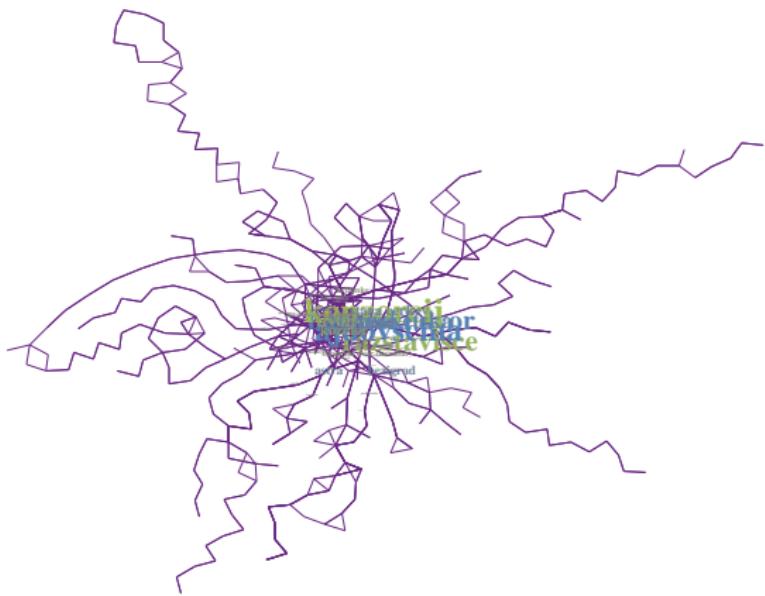
$$e(t) = A^t e(0) = A^t \sum_i C_i v_i = \sum_i C_i \lambda_i^t v_i = \lambda_1^t \sum_i C_i \left[\frac{\lambda_i}{\lambda_1} \right]^t v_i \rightarrow C_1 \lambda_1^t v_1$$

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

- in *directed G e = 0* for *kⁱⁿ = 0 nodes etc.*

networkology *LPP*

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



networkology *iMDB*

- *eigenvector centrality e* in iMDB network
- *highest e* nodes are *wrestling actors*

#	actor	k	e
1	Benoit, Chris	225	0.005261
2	Guerrero, Eddie	225	0.005261
3	Storm, Lance	225	0.005236
4	Wight, Paul	221	0.005231
5	Jericho, Chris	253	0.005230
6	Runnels, Dustin	220	0.005206
7	Flair, Ric	219	0.005203
8	Huffman, Booker	219	0.005199
9	Bischoff, Eric	218	0.005192
10	Wilson, Torrie	217	0.005192
11	Gruner, Peter	218	0.005183
12	Levy, Scott	226	0.005181

centrality *Katz*

nodes get small amount of importance for free

- for (*un*)directed G *Katz centrality* \mathbf{z} [Kat53] of i is

- α and β are some *positive constants*

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

- for *convenience* $\beta = 1$ 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 are some *positive constants*

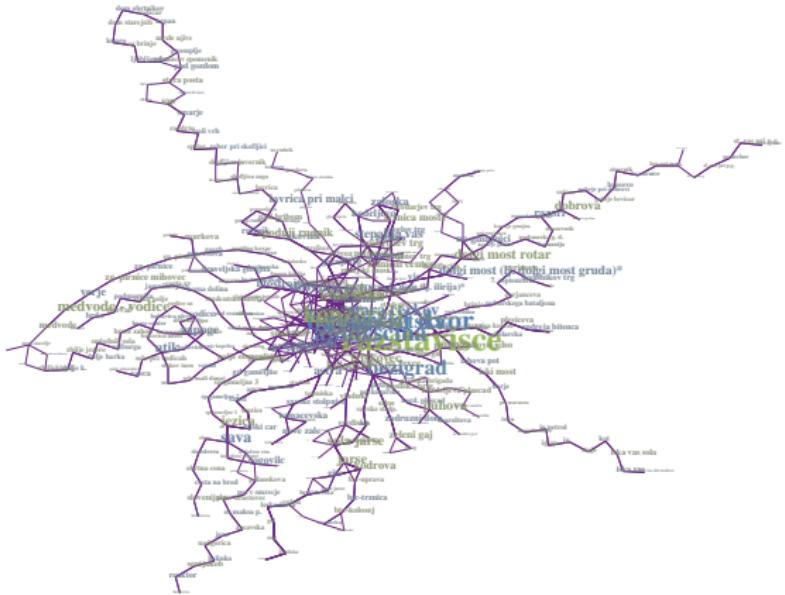
$$p_i = \alpha \sum_j A_{ij} \frac{p_j}{k_j^{out}} + \beta_j$$

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

see PageRank algorithm NetLogo demo

networkology *LPP*

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



networkology *iMDB*

- *PageRank centrality p* in iMDB network
- *highest p* nodes are *Hollywood actors*

#	actor	k	p
1	Hanks, Tom	457	0.000660
2	Jackson, Samuel L.	427	0.000642
3	Goldberg, Whoopi	398	0.000607
4	Stiller, Ben	382	0.000561
5	Davis, Mark	784	0.000547
6	Lopez, Jennifer	410	0.000542
7	Berland, Francois	194	0.000540
8	Berry, Halle	376	0.000536
9	Diaz, Cameron	361	0.000514
10	Travolta, John	335	0.000486
11	Jeremy, Ron	471	0.000476
12	Myers, Mike	345	0.000468

centrality overview

which *nodes* are most *important*?

1 IA		2 IIA		13 IIIA - 17 VIIA												18 VIIIIA																																
1	DC Degree Centrality	3	BC Betweenness Centrality	4	CC Closeness Centrality	5	SC Subgraph Centrality	6	Coeff Clustering Coefficient	7	Coef ⁻¹ Inverse Clustering Coefficient	8	MNC max. neighbor comp.	9	EC _{Coef} edge clustering coefficient	10	PR PageRank																															
2	RL Ring/Local Betweenness	11	IC Information Centrality	12	IG Integration	13	RC Radiality Centrality	14	DCCox DCCox	15	BCCox BCCox	16	ECCox ECCox	17	KGcox KGcox	18	LAC Loc. avg. Connectivity	19	LR LeaderRank																													
3	BN Bavelock Centrality	19	RWBC RandomWalk Betweenness	20	RWCC RandomWalk Closeness	21	CC _{2,4} 2,3,4-localized-CC	22	BCOox BCOox	23	PBcox PBcox	24	KGcox KGcox	25	COCCop COCCop	26	PRcox PRcox	27	DMNC dens. max. neighbor comp.	28	SEC _{Coef} sum. avg. edge Coef																											
4	σ stress Centrality	37	WDC Weighted Degree	38	WDECC Weighted DECC	39	DCECC DCECC	40	COECC COECC	41	BCECC BCECC	42	KBECC KBECC	43	PRECC PRECC	44	IGECC IGECC	45	PECC _{Coef} PCC > ECOP	46	KS KatzIndex																											
5	ω eccentricity	45	EC ₁ Eccentricity	46	WDC Weighted Degree	47	BC _{1,4} 2,3,4-localized-BC	48	ECC ⁻¹ Inverse Eccentricity	49	SDC Sphere Degree Centrality	50	DORC DORC	51	CCRC CCRC	52	PRRC PRRC	53	KGRC KGRC	54	EC ₂ 2-localized-EC																											
6	7	87	BC _{1,4} 2,3,4-localized-BC	88	BC _{1,4} 2,3,4-localized-BC	89	EC ₁ Eccentricity	90	EC ₂ 2-localized-EC	91	EC ₃ 3-localized-EC	92	EC ₄ 4-localized-EC	93	EC ₅ 5-localized-EC	94	EC ₆ 6-localized-EC	95	EC ₇ 7-localized-EC	96	EC ₈ 8-localized-EC																											
z mass		C None		Hybrid		<table border="1"> <tr> <td>22</td><td>FC Functional Centrality</td><td>23</td><td>FD Functional Diversity</td><td>24</td><td>US UniScore</td><td>25</td><td>DIS Pairwise Dis-connectivity</td><td>26</td><td>ASS Assortative Mixing</td><td>27</td><td>DAM Damage</td><td>28</td><td>UC United comp. Centrality</td></tr> <tr> <td>29</td><td>EI Essentiality Index</td><td>30</td><td>CM Complexity Measure</td><td>31</td><td>NoC Normalized or Centrality</td><td>32</td><td>MC Modulated Centrality</td><td>33</td><td>HGI Harary Graph Information</td><td>34</td><td>HYP Hyperbolic Index</td><td>35</td><td>HC Hannan's Centrality</td></tr> </table>													22	FC Functional Centrality	23	FD Functional Diversity	24	US UniScore	25	DIS Pairwise Dis-connectivity	26	ASS Assortative Mixing	27	DAM Damage	28	UC United comp. Centrality	29	EI Essentiality Index	30	CM Complexity Measure	31	NoC Normalized or Centrality	32	MC Modulated Centrality	33	HGI Harary Graph Information	34	HYP Hyperbolic Index	35	HC Hannan's Centrality	<ul style="list-style-type: none"> ■ Betweenness-based ■ Distance-based ■ Linear Combinations ■ Subgraph-based ■ Clustering Coefficient-based ■ Edge Clustering Coefficient-based ■ Spectral-based ■ Miscellaneous 	
22	FC Functional Centrality	23	FD Functional Diversity	24	US UniScore	25	DIS Pairwise Dis-connectivity	26	ASS Assortative Mixing	27	DAM Damage	28	UC United comp. Centrality																																			
29	EI Essentiality Index	30	CM Complexity Measure	31	NoC Normalized or Centrality	32	MC Modulated Centrality	33	HGI Harary Graph Information	34	HYP Hyperbolic Index	35	HC Hannan's Centrality																																			

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, Andrej Mrvar, and Vladimir Batagelj.
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.

centrality *references*

-  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.
-  Nataša Pržulj.
Biological network comparison using graphlet degree distribution.
Bioinformatics, 23(2):e177–e183, 2007.
-  Sara Nadin Soffer and Alexei Vázquez.
Network clustering coefficient without degree-correlation biases.
Phys. Rev. E, 71(5):057101, 2005.

centrality *references*



D. J. Watts and S. H. Strogatz.
Collective dynamics of 'small-world' networks.
Nature, 393(6684):440–442, 1998.