

Measures of centrality, PageRank algorithm

You are given iMDB actors collaboration network in Pajek format ([collaboration_imdb.net](http://collaboration.imdb.net)). Your task is to find the most important actors according to different measures of centrality.

1	14													18	VIIIA																					
1	DC													2	EC																					
	Degree Centrality													Eigenvector Centrality																						
2	3	12	4	13													5	9	6	4	7	1	8	1	9	3	10	2								
	BC	2 IIA		CC													SC	C _{COEF}	C _{COEF} ⁻¹	MNC	EC _{COEF}	PR														
	Betweenness Centrality	Closeness Centrality															Subgraph Centrality	Clustering Coefficient	inverse CCOEF	max. neighb. comp.	edge clustering coefficient	PageRank														
3	11	2	12	8													91	14	15	3	16	3	17	2	18	1										
	RL	IC														CC _{xxx}	SC _o	LAC	DMNC	SEC _{COEF}	LR															
	RangeLimited Betweenness	Information Centrality														odd Subgraph Centrality	odd Subgraph Centrality	loc. avg. Connectivity	dens. max. neighb. comp.	sum of ECCOEF	LeaderRank															
4	19	4	20	2	21	1	57	1	58	1	59	1	61	1	62	1	63	1	64	1	65	1	89	1	90	1	32	2	33	1	34	1	35	2	36	2
	BN	RC		IG	DC _{xxx}	BC _{xxx}	CC _{xxx}	ECC _{xxx}	KS _{xxx}	PR _{xxx}	IG _{xxx}	RC _{xxx}	DC _{xxx}	BC _{xxx}	SC _E	KL	COC _{COEF}	PEC _{COEF}	KS																	
	BottleNeck Centrality	Radiality Centrality		Integration										even Subgraph Centrality	Clique Level	coexp. weight CCOEF	PCC×ECCOEF	KatzStatus																		
5	37	1	38	1	39	1	93	1	95	1	94	1	50	1	97	1	96	1	46	1	47	1	48	1	49	1	50	1	51	1	52	1	53	5	54	1
	RWBC	RWCC		CC _{2,3,4}	ECC _{xxx}	PR _{xxx}	KS _{xxx}	CCC _{COEF}	RC _{xxx}	IG _{xxx}	DCBC	BCCC	CCKS	KS _{PR}	DC _{PR}	β	SC ₂	NC	EC ₂																	
	RandomWalk Betweenness	RandomWalk Closeness		2,3,4-localized-CC												Bipartivity	2-localized-SC	Neighborhood Centrality	2-localized-EC																	
6	55	2	56	2	57-71	1	72	1	73	1	74	1	75	1	76	1	77	1	78	1	79	1	80	1	81	1	82	1	50	1	84	1	85	1	86	1
	σ	ECC		WDC	DCECC	CCECC	BCECC	KSECC	PRECC	IGECC	DCCC	BCKS	CCPR	KSIG	DCIG	DCC _{COEF}	SC ₃	LI	EC ₃																	
	stress Centrality	Eccentricity		Weighted Degree													3-localized-SC	Lobby Index	3-localized-EC																	
7	87	1	88	1	89-103	1	104	1	105	1	106	1	107	1	108	1	109	1	110	1	111	1	112	1	113	1	114	1	115	1	50	1	117	1	118	1
	BC _{2,3,4}	ECC ⁻¹		SDC	DCRC	CCRC	BCRC	KSRC	PRRC	IGRC	DCKS	BCPR	CCIG	DCPR	BCIG	ECCRC	BCC _{COEF}	SC ₄	EC ₄																	
	2,3,4-localized-BC	Inverse Eccentricity		Sphere Degree Centrality														4-localized-SC	4-localized-EC																	

z	mass	Hybrid
C	Name	

22	2	23	2	24	2	25	1	26	1	27	1	117	1
FC	FD	US	DIS	ASS	DAM	UC							
Functional Centrality	Functional Diversity	UniScore	Pairwise Dis-connectivity	Assortative Mixing	Damage	United compl. Centrality							

28	1	29	1	30	1	31	1	13	1	116	0	118	1
EI	CM	NαC	MC	HGI	HYP	HC							
Essentiality Index	Complexity Measure	Normalized α Centrality	Moduland Centrality	Harary Graph Information	Hyperbolic Index	Harmonic Centrality							

- Betweenness-based
- Distance-based
- Linear Combinations
- Subgraph-based
- Clustering Coefficient-based
- Edge Clustering Coefficient-based
- Spectral-based
- Miscellaneous

Schoch (University of Konstanz)

I. Degree centrality and clustering coefficients

- (code)** Find the most important actors according to the degree centrality $d_i = \frac{k_i}{n-1}$, where n is the number of network nodes and k_i is the degree of node i . Which actors have the highest d_i (e.g., Hollywood, international, unknown)?
- (code)** Find the most important actors according to the clustering coefficient $C_i = \frac{2t_i}{k_i(k_i-1)}$, where k_i is the degree of node i and t_i is the number of triangles including node i . You should use the link triad counting algorithm from previous labs. Which actors have the highest C_i (e.g., Hollywood, international, unknown)?
- (homework)** Find the most important actors according to the μ -corrected clustering coefficient $C_i^\mu = \frac{2t_i}{k_i\mu}$, where k_i is the degree of node i , t_i is the number of triangles including node i and μ is the maximum number of triangles over a link. You should use the link triad counting algorithm from

previous labs. Which actors have the highest C_i^H (e.g., Hollywood, international, unknown)?

II. Eigenvector centrality and PageRank algorithm

- (code)** Find the most important actors according to the eigenvector centrality $e_i = \lambda_1^{-1} \sum_j A_{ij} e_j$, where A is the network adjacency matrix and λ_1 is a normalizing constant. You should use the power iteration algorithm shown below. Which actors have the highest e_i (e.g., Hollywood, international, unknown)?
- (code)** Find the most important actors according to the PageRank algorithm $p_i = \alpha \sum_j A_{ij} \frac{p_j}{k_j} + \frac{1-\alpha}{n}$, where A is the network adjacency matrix, n is the number of network nodes, k_i is the degree of node i and α is the damping factor set to 0.85. You should use the PageRank algorithm shown below. Which actors have the highest p_i (e.g., Hollywood, international, unknown)?

```

input graph G, precision  $\epsilon$ 
output eigenvector centrality E
1:  $E \leftarrow$  array of ones
2: do
3:    $U \leftarrow$  array of zeros
4:   for nodes  $i \in N$  do
5:     for neighbors  $j \in \Gamma_i$  do
6:        $U[i] \leftarrow U[i] + E[j]$ 
7:    $u \leftarrow \|U\|$ 
8:   for nodes  $i \in N$  do
9:      $U[i] \leftarrow U[i] \cdot n/u$ 
10:   $\Delta \leftarrow \|E - U\|$ 
11:   $E \leftarrow U$ 
12: while  $\Delta > \epsilon$ 
13: return E

```

```

input graph G, damping  $\alpha$ , precision  $\epsilon$ 
output PageRank ranks P
1:  $P \leftarrow$  array of  $n^{-1}$ -s
2: do
3:    $U \leftarrow$  array of zeros
4:   for nodes  $i \in N$  do
5:     for predecessors  $j \in \Gamma_i^{in}$  do
6:        $U[i] \leftarrow U[i] + P[j] \cdot \alpha/k_j^{out}$ 
7:    $u \leftarrow \|U\|$ 
8:   for nodes  $i \in N$  do
9:      $U[i] \leftarrow U[i] + (1 - \alpha)/n$ 
10:   $\Delta \leftarrow \|P - U\|$ 
11:   $P \leftarrow U$ 
12: while  $\Delta > \epsilon$ 
13: return P

```

III. Closeness and betweenness centrality

- (code)** Find the most important actors according to the closeness centrality $\ell_i^{-1} = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d_{ij}}$, where n is the number of network nodes and d_{ij} is the distance between nodes i and j . You should use the breadth-first search algorithm from previous labs. Which actors have the highest ℓ_i^{-1} (e.g., Hollywood, international, unknown)?
- (homework)** Find the most important actors according to the betweenness centrality $\sigma_i = \frac{1}{n^2} \sum_{st} \frac{g_{st}^i}{g_{st}}$, where n is the number of network nodes, g_{st} is the number of shortest paths between nodes s and t , and g_{st}^i is the number of such paths through node i . Which actors have the highest σ_i (e.g., Hollywood, international, unknown)?