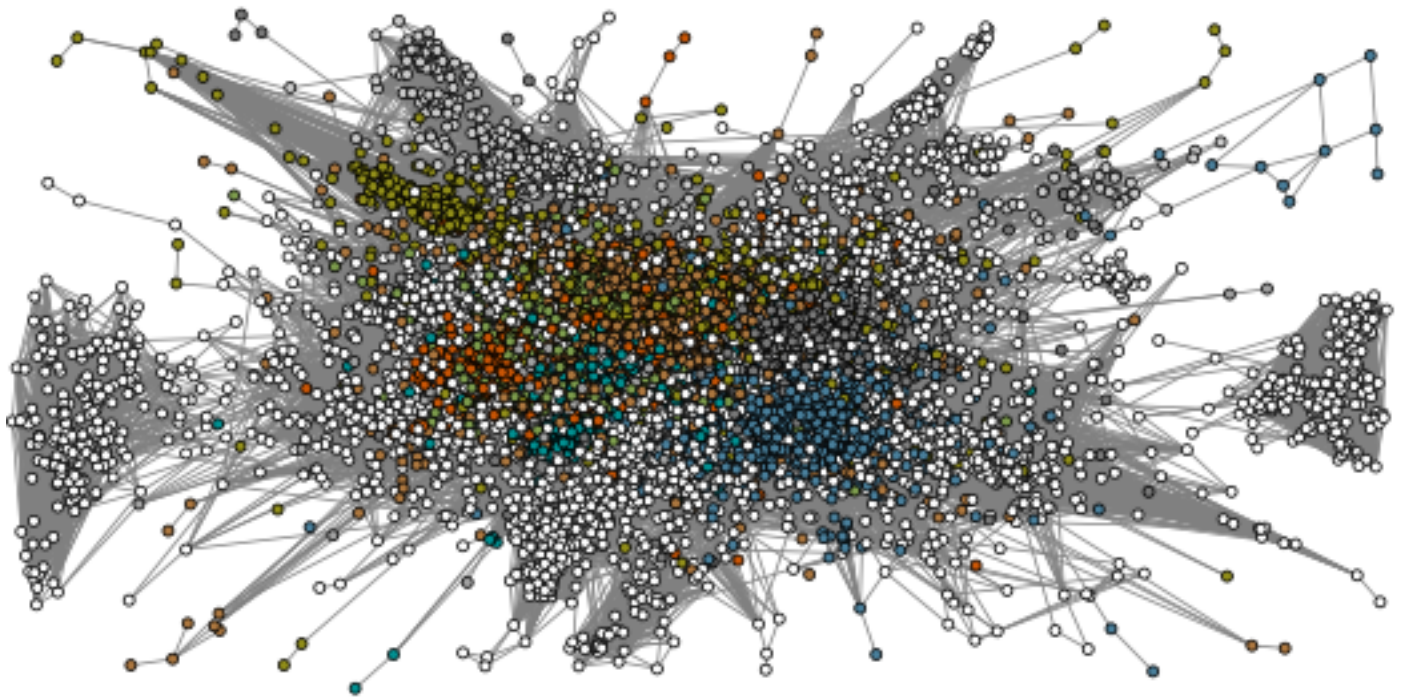


## Node position in IMDb actors collaboration network

You are given [IMDb actors collaboration network](#) in Pajek format. Your task is to find the **most important actors** according to different measures of node centrality. You can either use the methods provided by your network analysis library or implement the algorithms by yourself.



### I. Degree centrality and clustering coefficients

1. Find the **most important actors according to degree centrality**  $d_i = \frac{k_i}{n-1}$ , where  $n$  is the number of nodes and  $k_i$  is the degree of node  $i$ . Which actors have the highest  $d$  (e.g. Hollywood, international, unknown)?

*Computational complexity is linear  $\mathcal{O}(n)$  and applicable to any network that fits in your memory.*

2. Find the **most important actors according to 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  $i$ . Which actors have the highest  $C$  (e.g. Hollywood, international, unknown)?

*Computational complexity is superlinear  $\mathcal{O}(m\langle k \rangle)$  and applicable to all but the largest networks.*

3. (homework) Find the **most important actors according to  $\mu$ -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  $\mu$  is the maximum number of triangles over a single link. Which actors have the highest  $C^\mu$  (e.g. Hollywood, international, unknown)?

## II. Eigenvector centrality and PageRank algorithm

1. (homework) Find the **most important actors according to eigenvector centrality**

$e_i = \lambda_1^{-1} \sum_j A_{ij} e_j$ , where  $A$  is the adjacency matrix and  $\lambda_1$  is a normalizing constant. Which actors have the highest  $e$  (e.g. Hollywood, international, unknown)?

2. Find the **most important actors according to PageRank score**  $p_i = \alpha \sum_j A_{ij} \frac{p_j}{k_j} + \frac{1-\alpha}{n}$ , where  $A$  is the adjacency matrix,  $n$  is the number of nodes,  $k_i$  is the degree of node  $i$  and  $\alpha$  is the damping factor set to 0.85. Which actors have the highest  $p$  (e.g. Hollywood, international, unknown)?

Computational complexity is  $\approx$  linear  $\mathcal{O}(m)$  and applicable to any network that fits in your memory.

```

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

1. (homework) Find the **most important actors according to closeness centrality**

$\ell_i^{-1} = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d_{ij}}$ , where  $n$  is the number of nodes and  $d_{ij}$  is the distance between nodes  $i$  and  $j$ . Which actors have the highest  $\ell^{-1}$  (e.g. Hollywood, international, unknown)?

2. Find the **most important actors according to 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  $\sigma$  (e.g. Hollywood, international, unknown)?

Computational complexity is inevitably quadratic  $\mathcal{O}(nm)$  and applicable only to smaller networks.