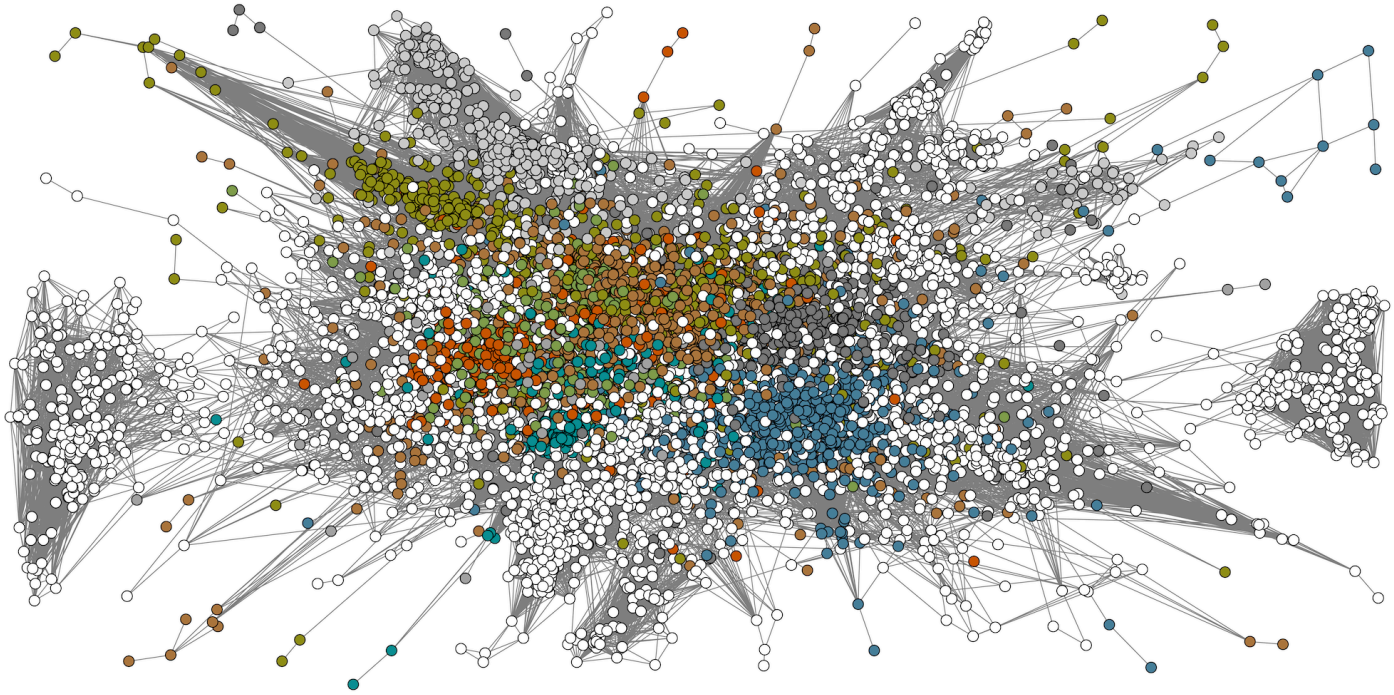
# Node position in IMDb actors collaboration network

You are given **IMDb actors collaboration network** in Pajek format (imdb.net). 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 library or implement the algorithms yourself.



## I. Degree centrality and clustering coefficients

1. *(starter)* 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. 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 an appropriate constant (i.e. 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. *(tentative)* 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 ϵ
output eigenvector centrality E
E ← array of ones
do
    U ← array of zeros
    for nodes i ∈ N do
        for neighbors j ∈ Γ_i do
            U[i] ← U[i] + E[j]
    u ← ||U||
    for nodes i ∈ N do
        U[i] ← U[i] · n/u
    Δ ← ||E − U||
    E ← U
while Δ > ϵ
return E
```

```
input  graph G, damping α, precision ϵ
output PageRank ranks P
P ← array of n^−1-s
do
    U ← array of zeros
    for nodes i ∈ N do
        for predecessors j ∈ Γ_i^in do
            U[i] ← U[i] + P[j] · α/k_j^out
    u ← ||U||
    for nodes i ∈ N do
        U[i] ← U[i] + (1 − u)/n
    Δ ← ||P − U||
    P ← U
while Δ > ϵ
return P
```

## III. Closeness and betweenness centrality

1. *(tentative)* 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 quadratic $\mathcal{O}(nm)$ and applicable only to medium sized networks.*