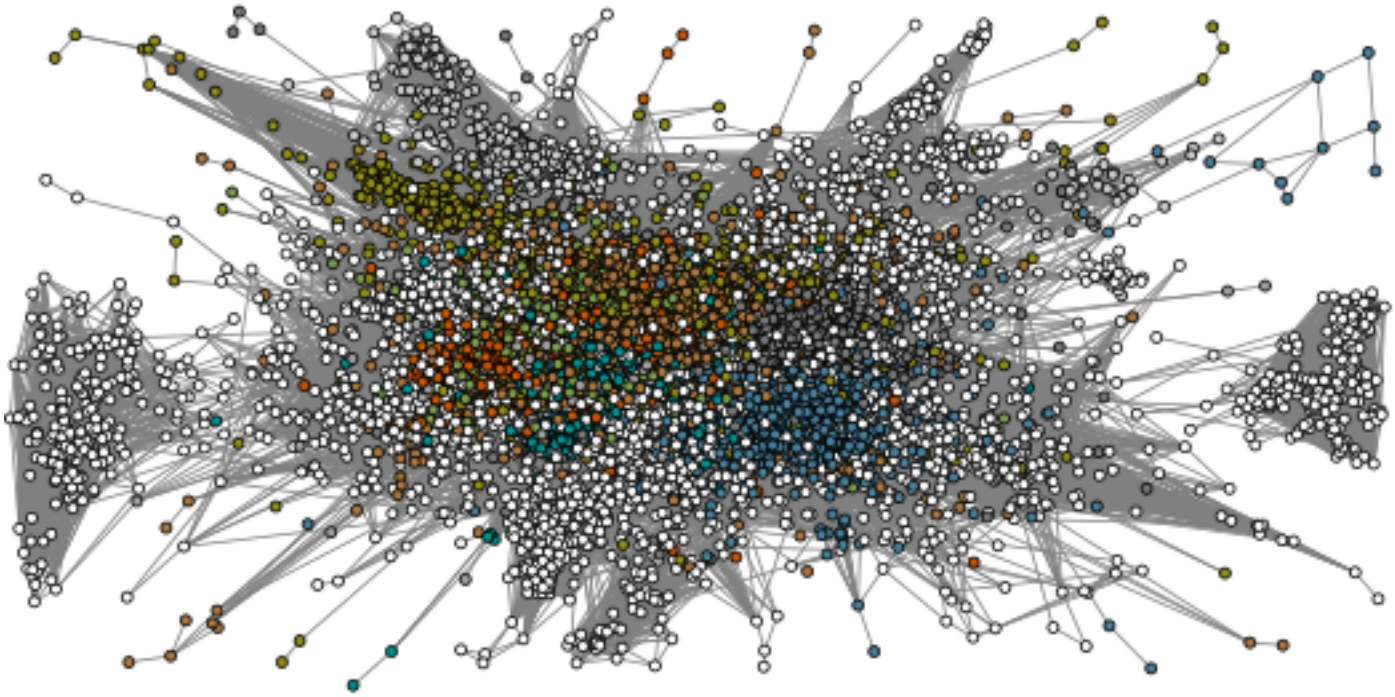
Node position in IMDb actors collaboration network

You are given **IMDb actors collaboration network** in Pajek, edge list and LNA formats. 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 network nodes and k_i is the degree of node i. Which actors have the highest d_i (e.g. Hollywood, Bollywood, international, unknown, other)? *(Computational complexity of these computations is linear $\mathcal{O}(n)$ and should be applicable to any network you are able to fit 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 triads including node i. Which actors have the highest C_i (e.g. Hollywood, Bollywood, international, unknown, other)? *(Computational complexity of these computations is superlinear $\mathcal{O}(m\langle k\rangle)$ and should be applicable to all but the largest networks.)*

3. *(tentative)* Find the **most important actors according to 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 triads including node i and μ is the maximum number of triads over a single link. Which actors have the highest C_i^{μ} (e.g. Hollywood,

Bollywood, international, unknown, other)? *(Computational complexity of these computations is superlinear $\mathcal{O}(m\langle k\rangle)$ and should be applicable to all but the largest networks.)*

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 network adjacency matrix and λ_1 is a normalizing constant. Which actors have the highest e_i (e.g. Hollywood, Bollywood, international, unknown, other)? *(Computational complexity of these computations is close to linear $\mathcal{O}(m)$ and should be applicable to any network you are able to fit in your memory.)*

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 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. Which actors have the highest p_i (e.g. Hollywood, Bollywood, international, unknown, other)? *(Computational complexity of these computations is close to linear $\mathcal{O}(m)$ and should be applicable to any network you are able to fit 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 network nodes and d_{ij} is the distance between nodes i and j. Which actors have the highest ℓ_i^{-1} (e.g. Hollywood, Bollywood, international, unknown, other)? *(Computational complexity of these computations is inevitably quadratic $\mathcal{O}(nm)$ and will be performed by the workshop instructor beforehand.)*

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 σ_i (e.g. Hollywood,

Bollywood, international, unknown, other)? *(Computational complexity of these computations is inevitably quadratic* $\mathcal{O}(nm)$ *and will be performed by the workshop instructor beforehand.)*