My VK Friends Network Analysis

Rim Shayakhmetov

Higher School of Economics
Department of Computer Science

2016

Outline

My Network Summary

- Network Source
- My Network Layout
- Friends

Structural Analysis

- Degree Distribution
- Main Properties
- Node Similarity
- Node Centrality and Network Layout

Community Detection

- Clique
- Community detection

Outline

My Network Summary
- Network Source
- My Network Layout
- Friends

Structural Analysis
- Degree Distribution
- Main Properties
- Node Similarity
- Node Centrality and Network Layout

Community Detection
- Clique
- Community detection

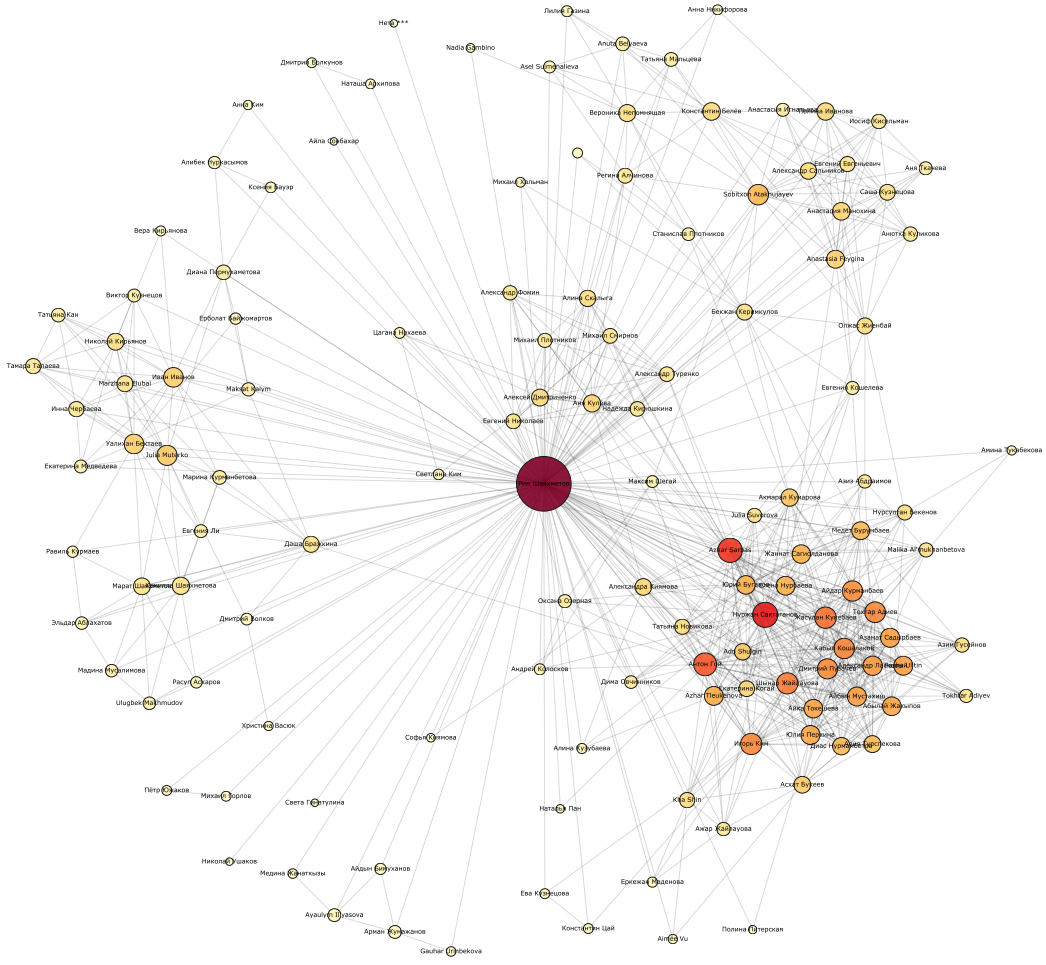
Network Source

Data

- VK API
  - friends.get
  - users.get
- Users' attributes
  - last name
  - first name
  - city
  - country
  - gender
  - birthday
- Size
  - 131 nodes
  - 760 edges

My Network

PageRank (size) and Degree Centrality (color)



Friends

My view

- MSU friends
  - friends from Astana
  - friends from Moscow
- HSE friends
  - data science programme
  - dormitory
- Home Town
  - school
  - relatives
- Others

Outline

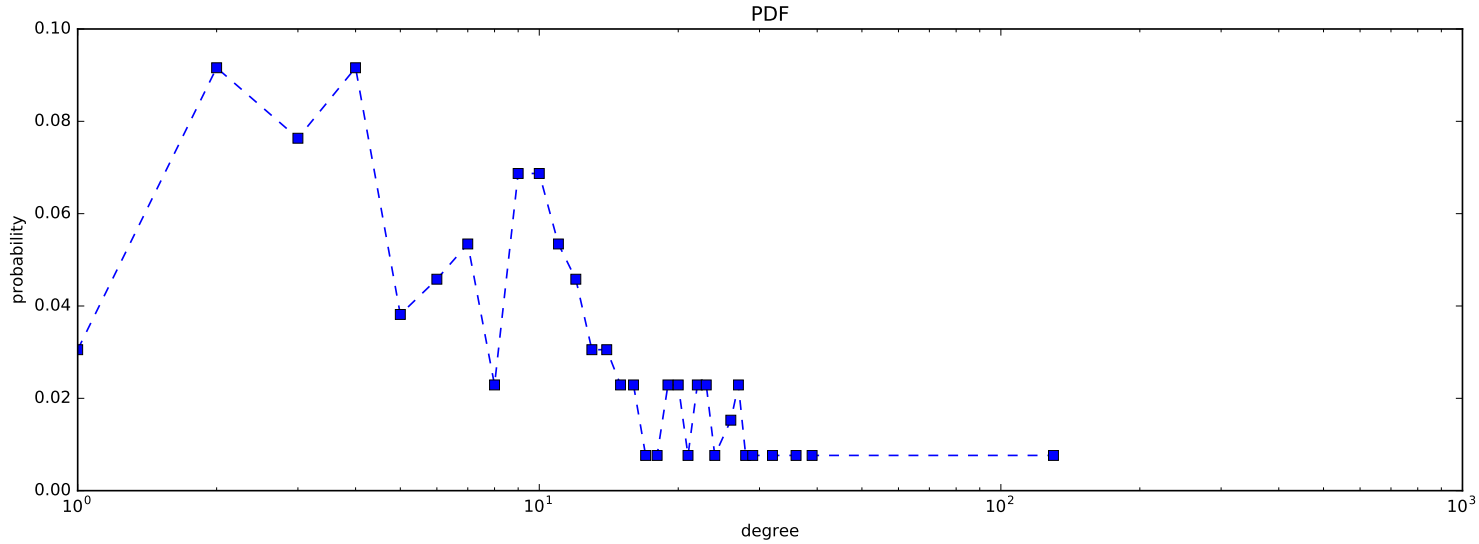
My Network Summary
- Network Source
- My Network Layout
- Friends

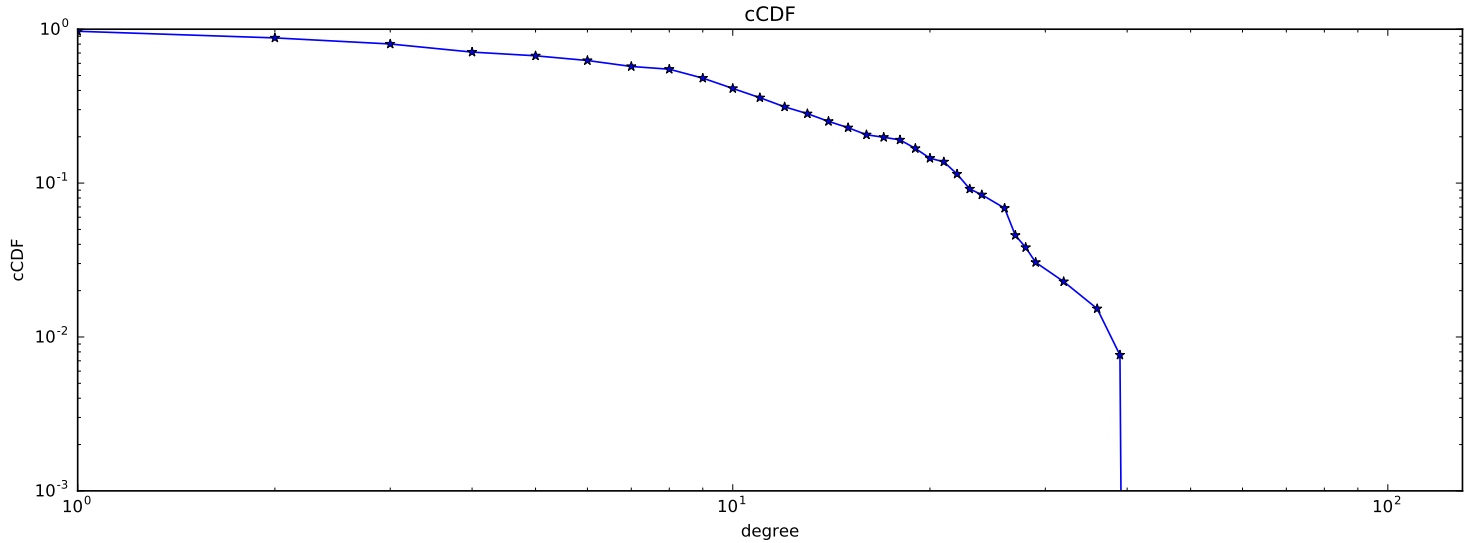
Structural Analysis
- Degree Distribution
- Main Properties
- Node Similarity
- Node Centrality and Network Layout

Community Detection
- Clique
- Community detection

Degree Distribution

PDF and cCDF of the degree distribution of the nodes



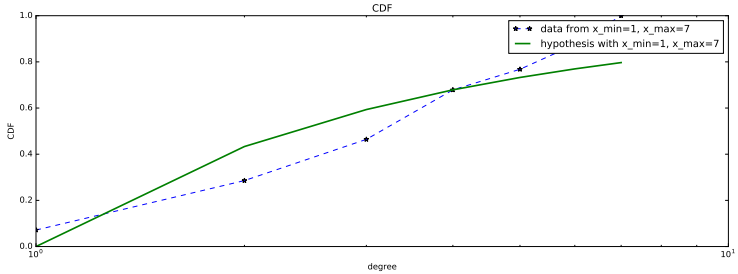


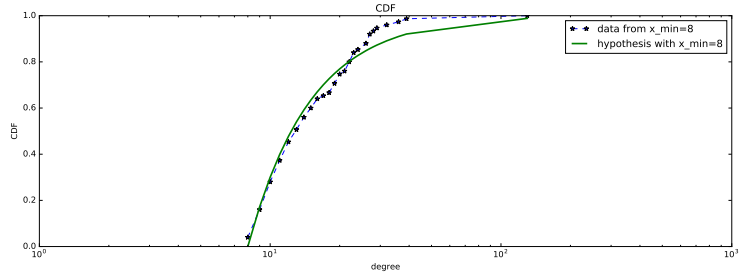
Degree Distribution

Hypothesis

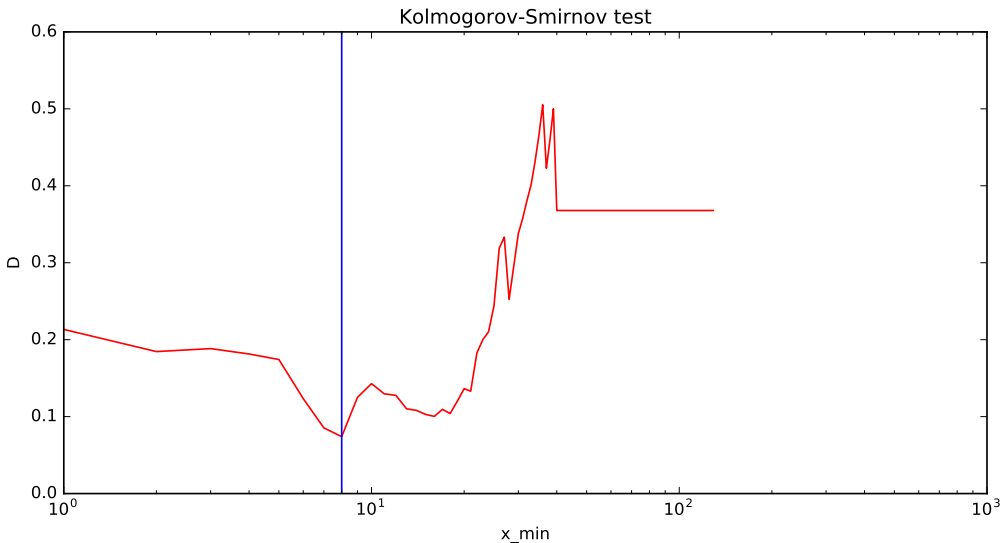
Divide degrees by some threshold s.t.:

Power Law holds on each region





Finding an optimal split by Kolmogorov-Smirnov test

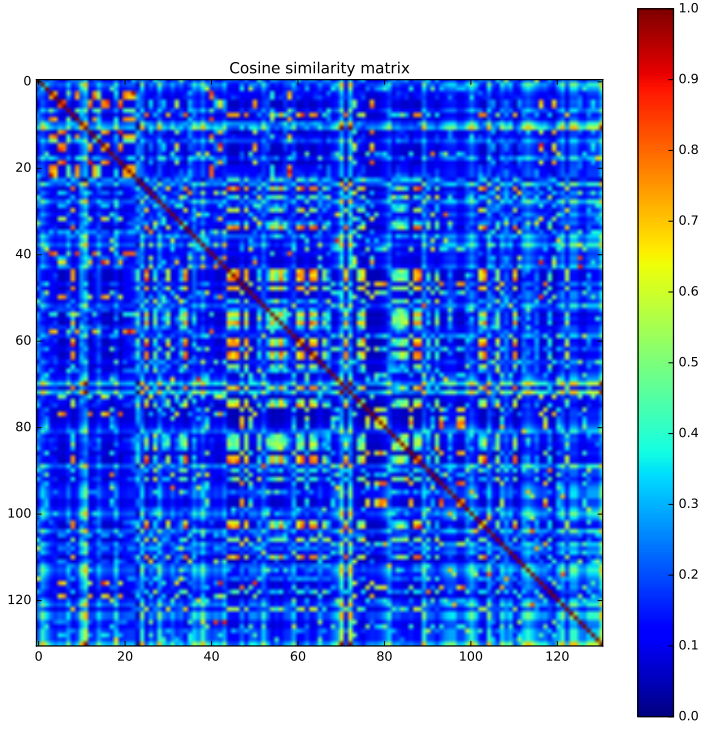


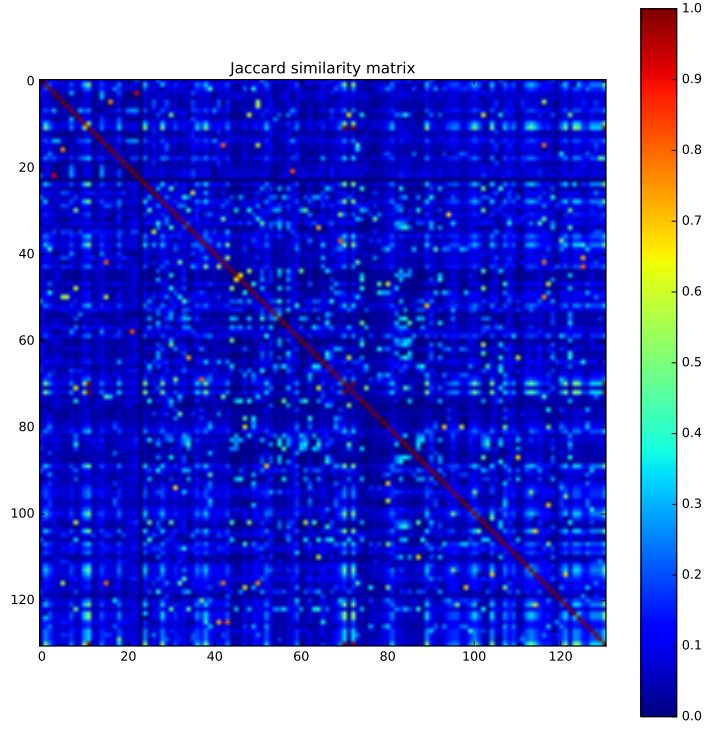
Main Properties

- Average Clustering Coefficient = 0.71
- Diameter = 2
- Average Shortest Path Length = 1.91
- Transitivity = 0.38
- Assortativity coefficient
  - by gender = 0.01
  - by city = 0.04
  - by country = 0.05
  - by degree = -0.17

Node Similarity

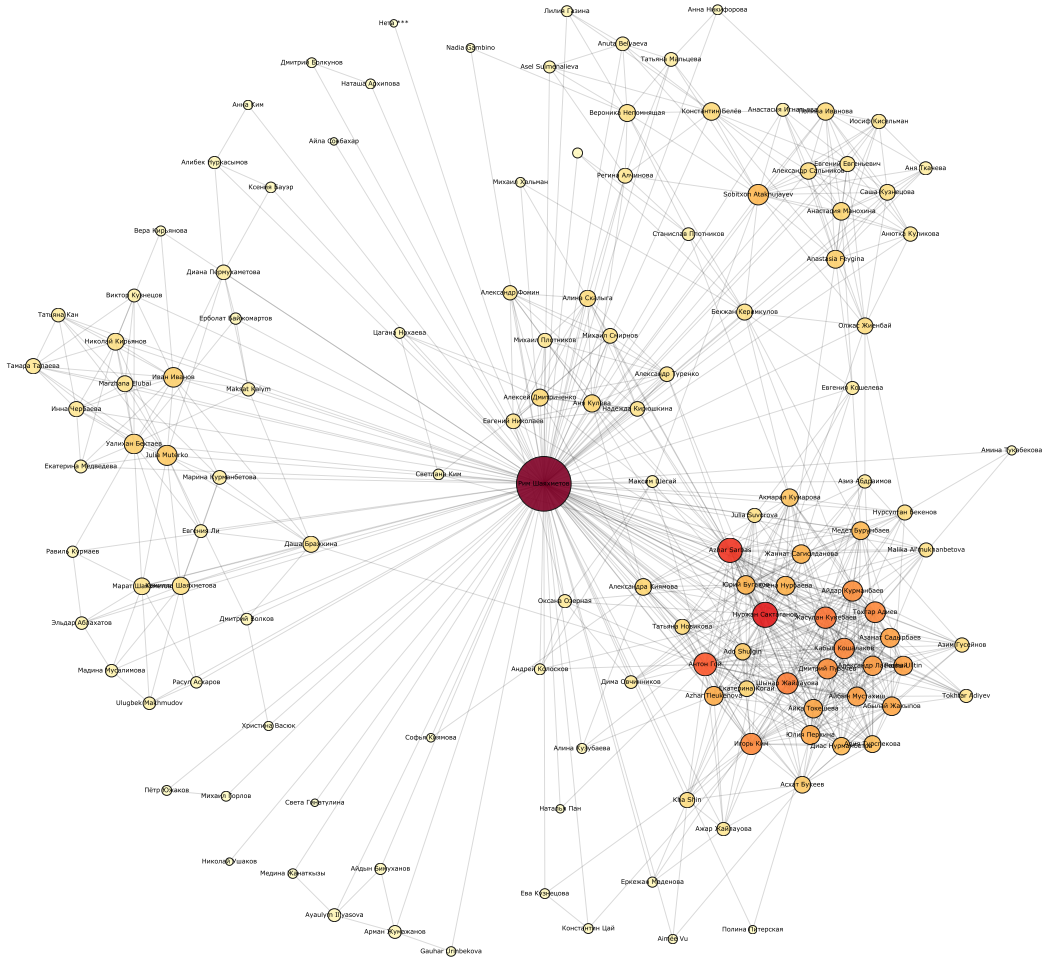
Cosine and Jaccard Similarity





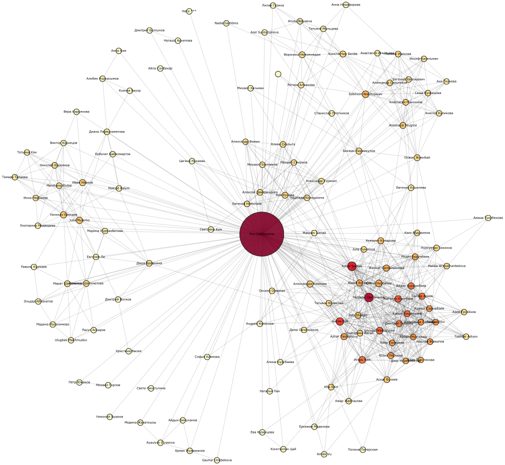
Node Centrality

PageRank (size) and Degree Centrality (color)



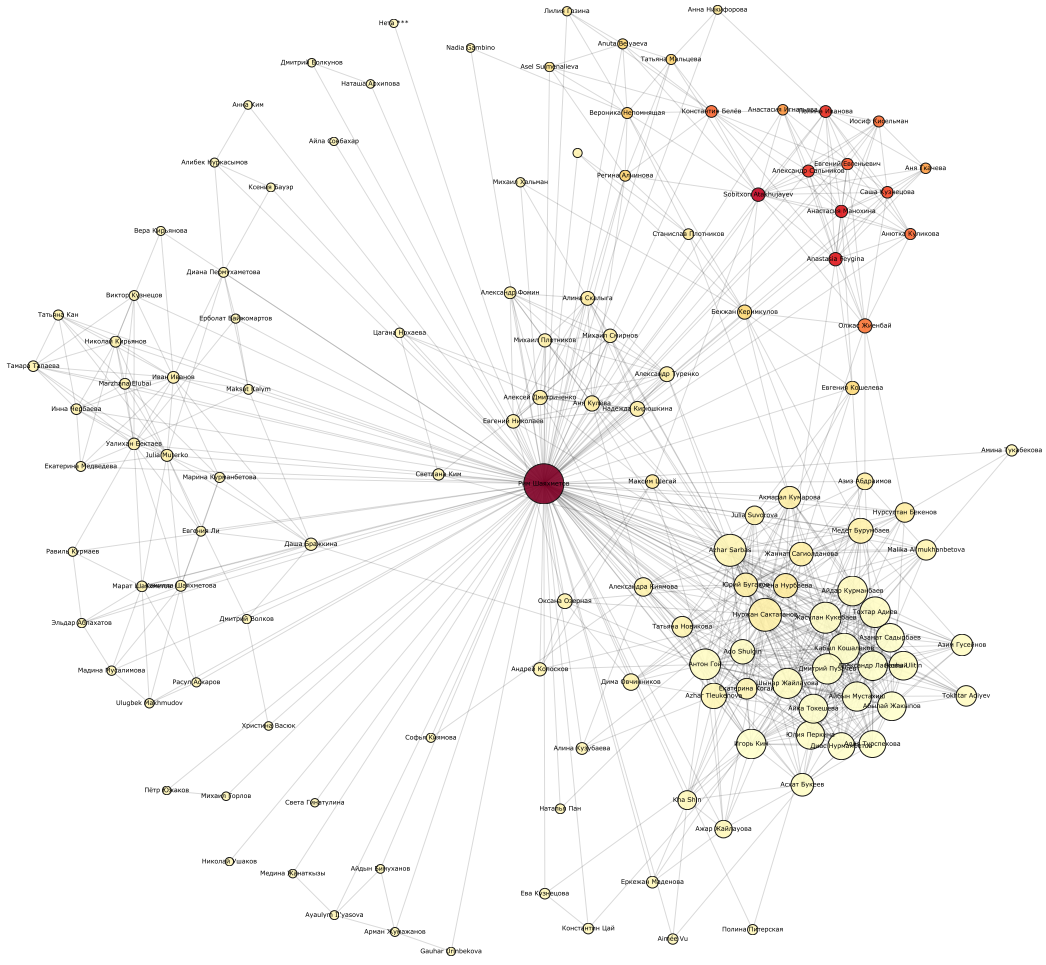
Node Centrality

Betweeness (size) and Closeness Centrality (color)



Node Centrality

Eigenvector (size) and Katz Centrality (color)



Outline

My Network Summary
- Network Source
- My Network Layout
- Friends

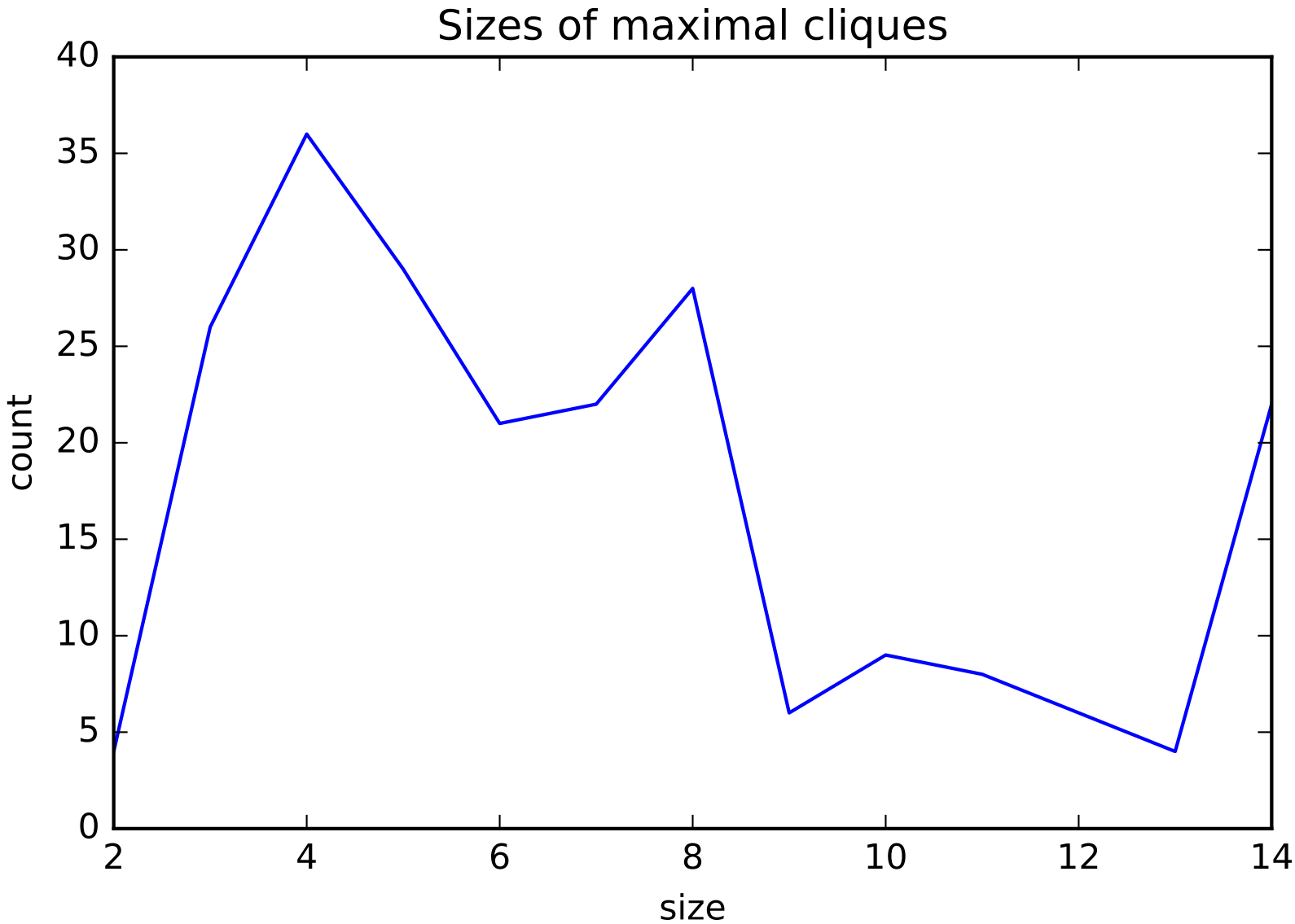
Structural Analysis
- Degree Distribution
- Main Properties
- Node Similarity
- Node Centrality and Network Layout

Community Detection
- Clique
- Community detection

Clique

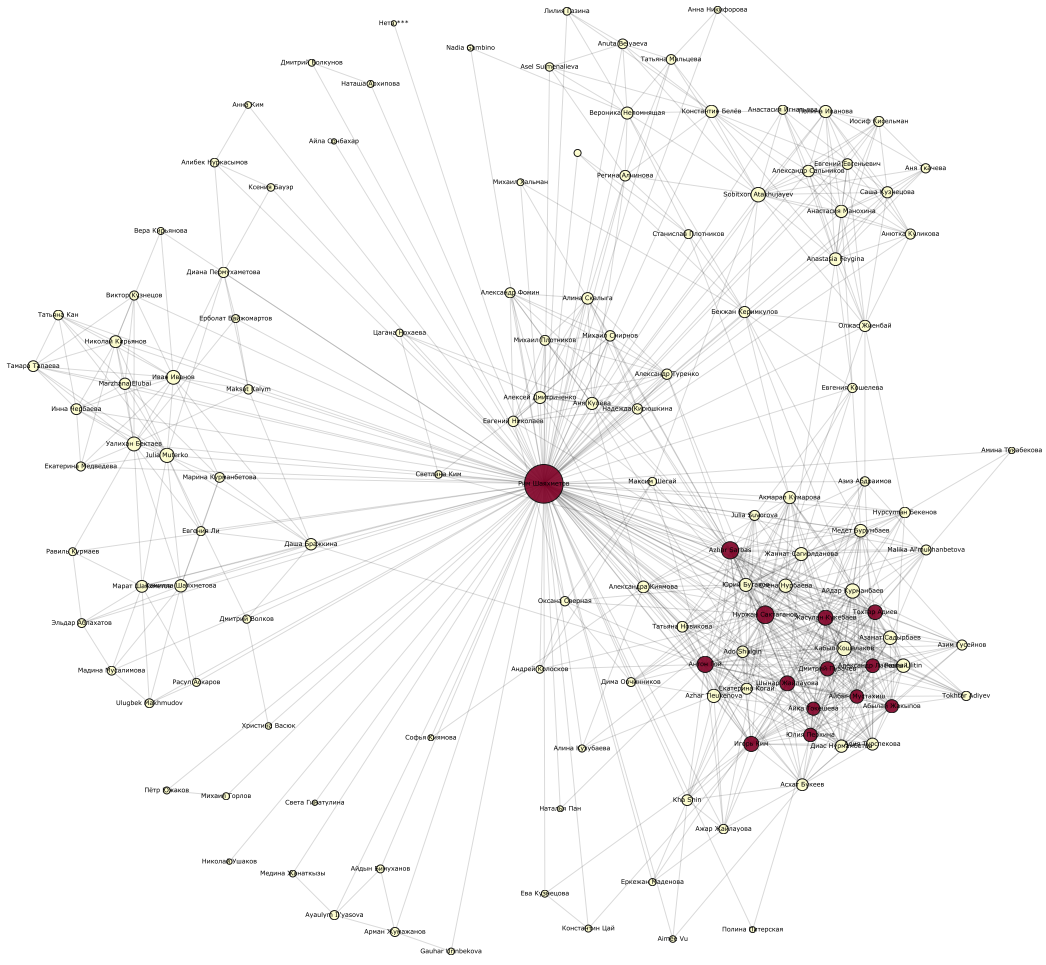
Maximal Cliques

The distribution of sizes of maximal cliques



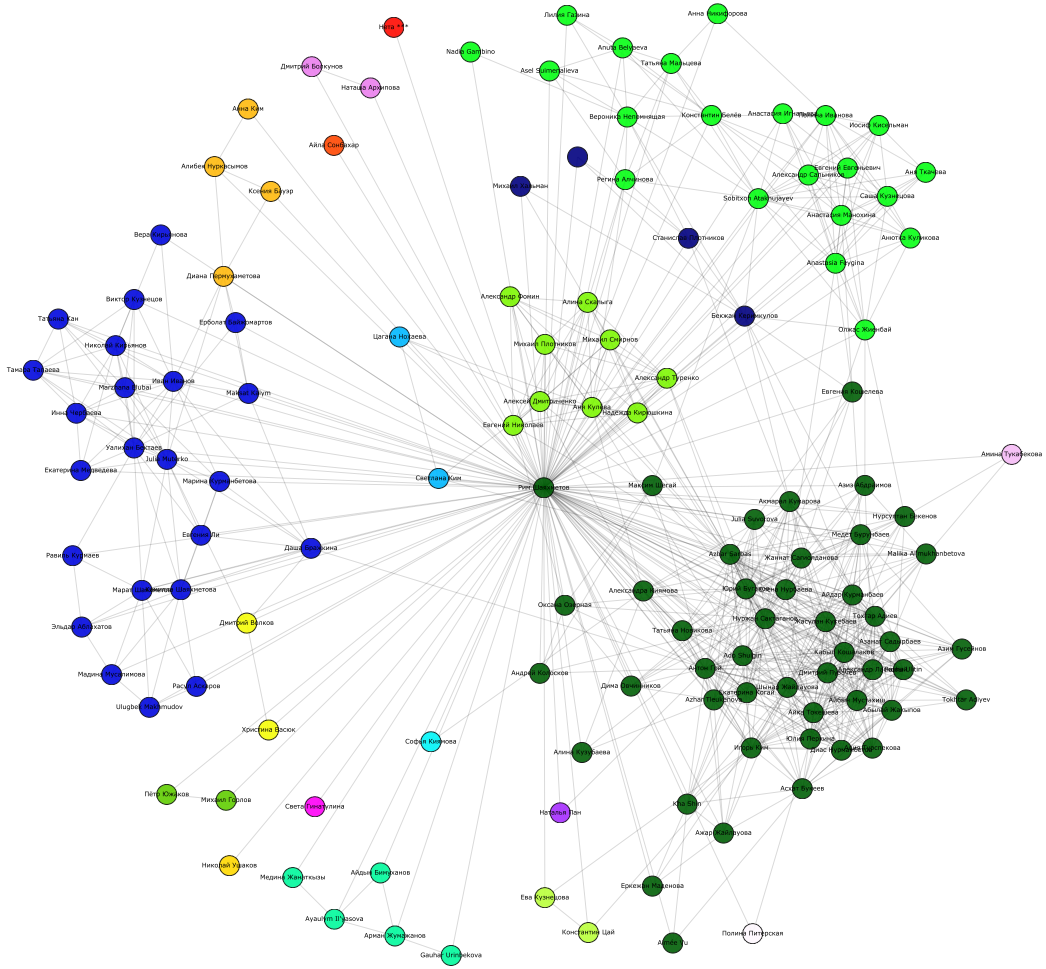
Clique

Some Maximal Maximum Clique (color)



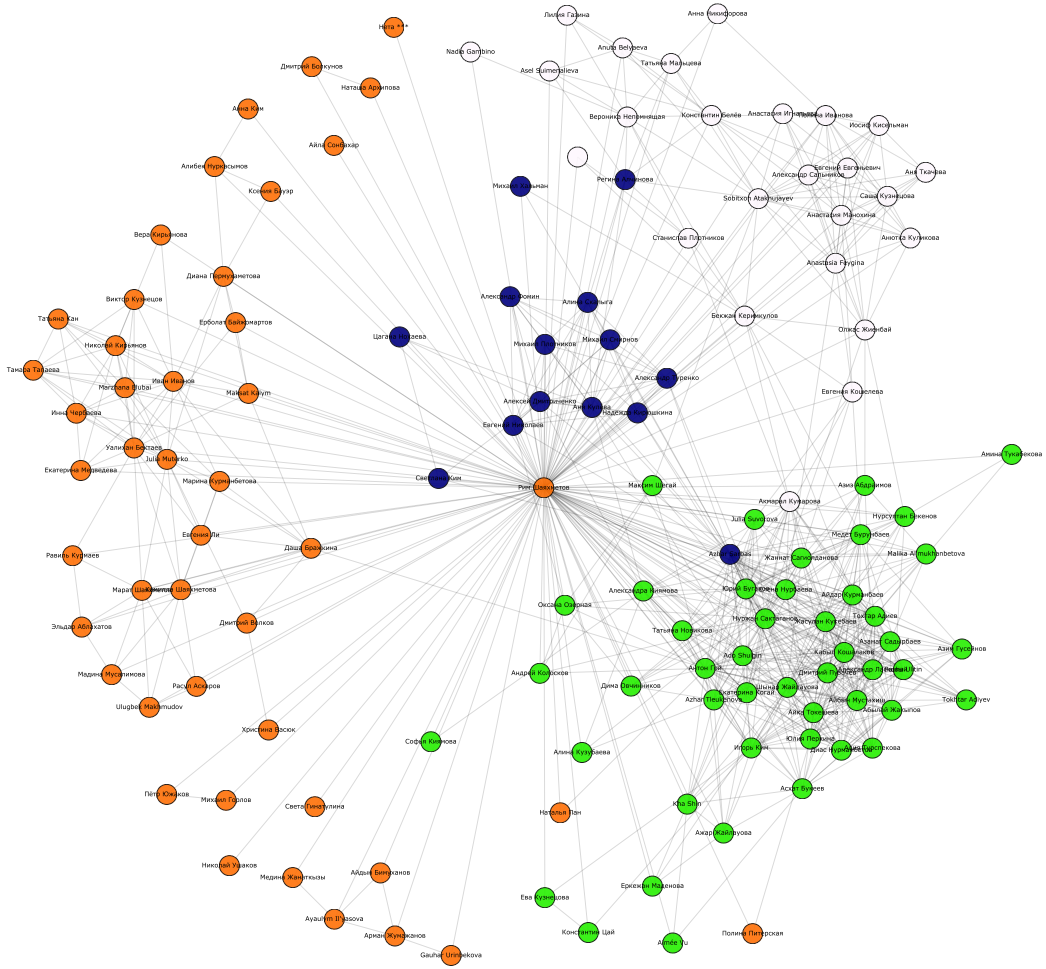
Community detection

Edge Betweenness (color)



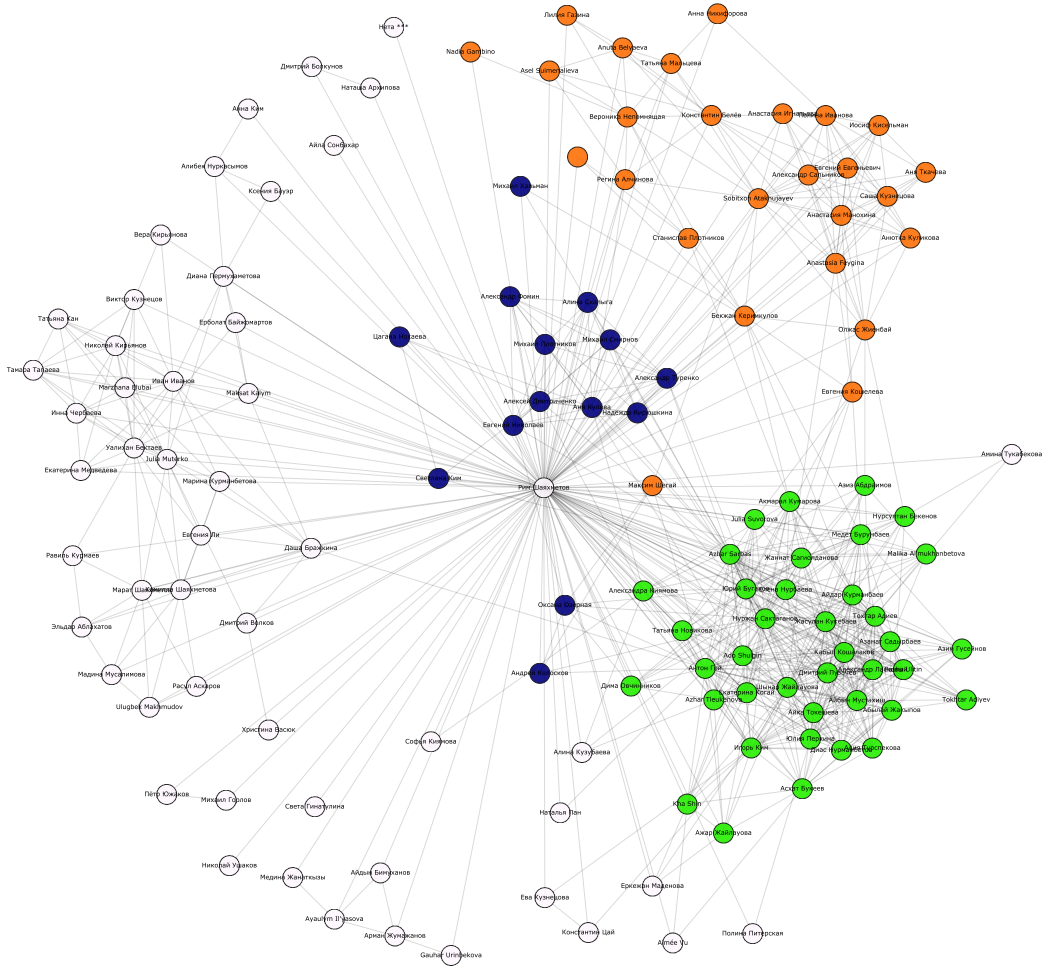
Community detection

Fast Greedy (color)



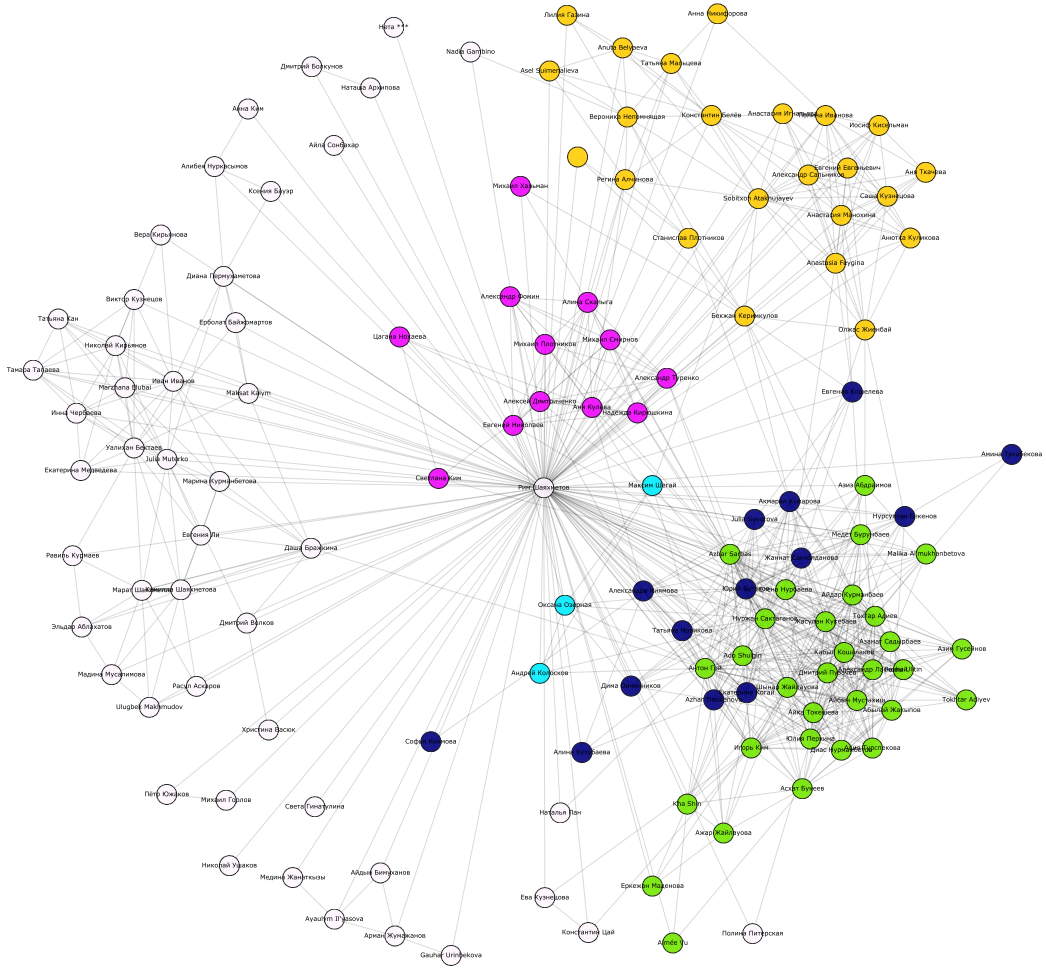
Community detection

Leading Eigenvector (color)



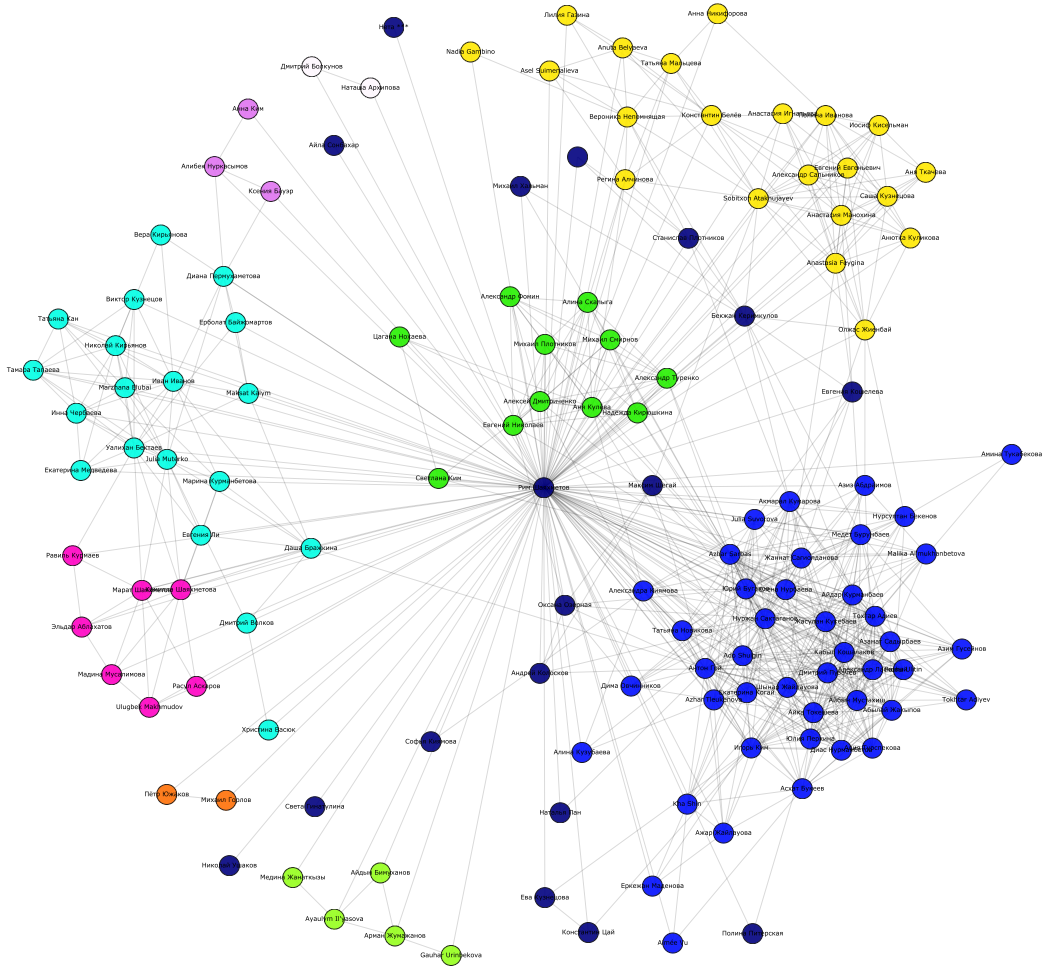
Community detection

Springlass (color)



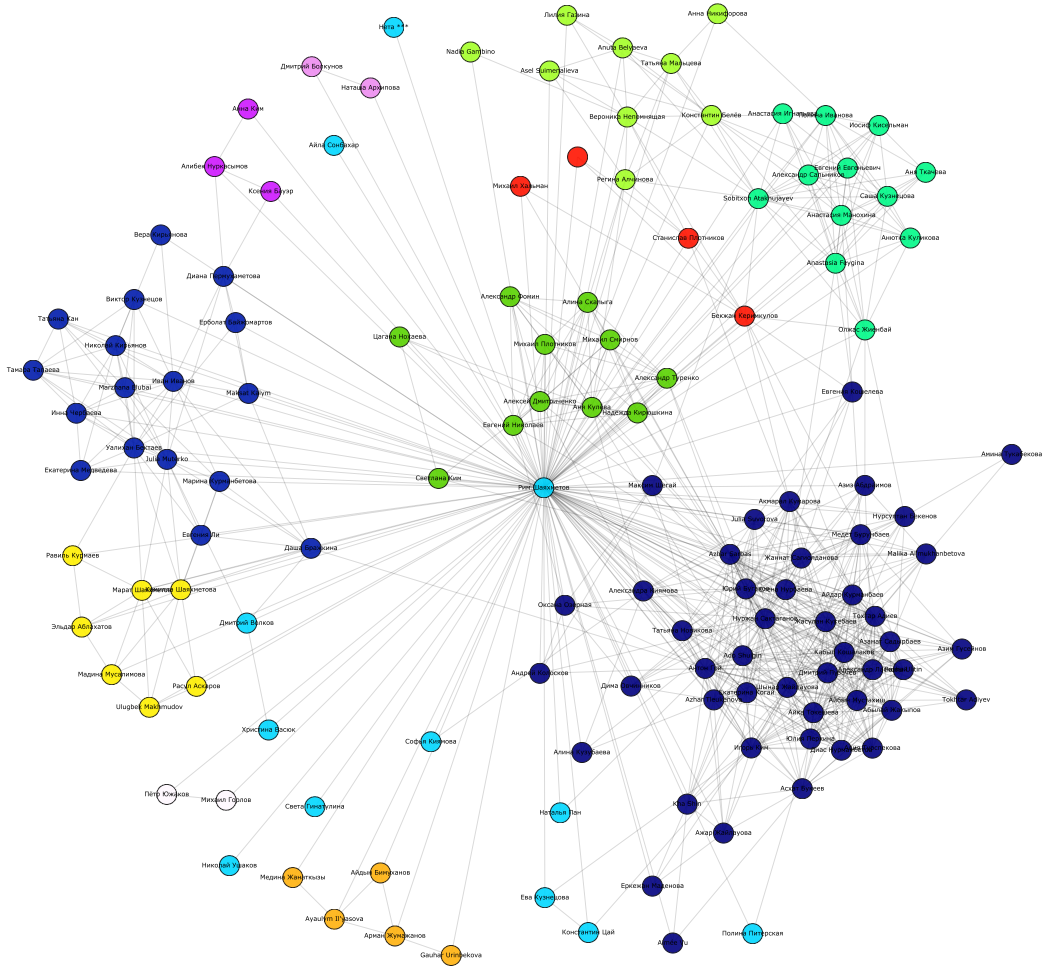
Community detection

Walktrap (color)



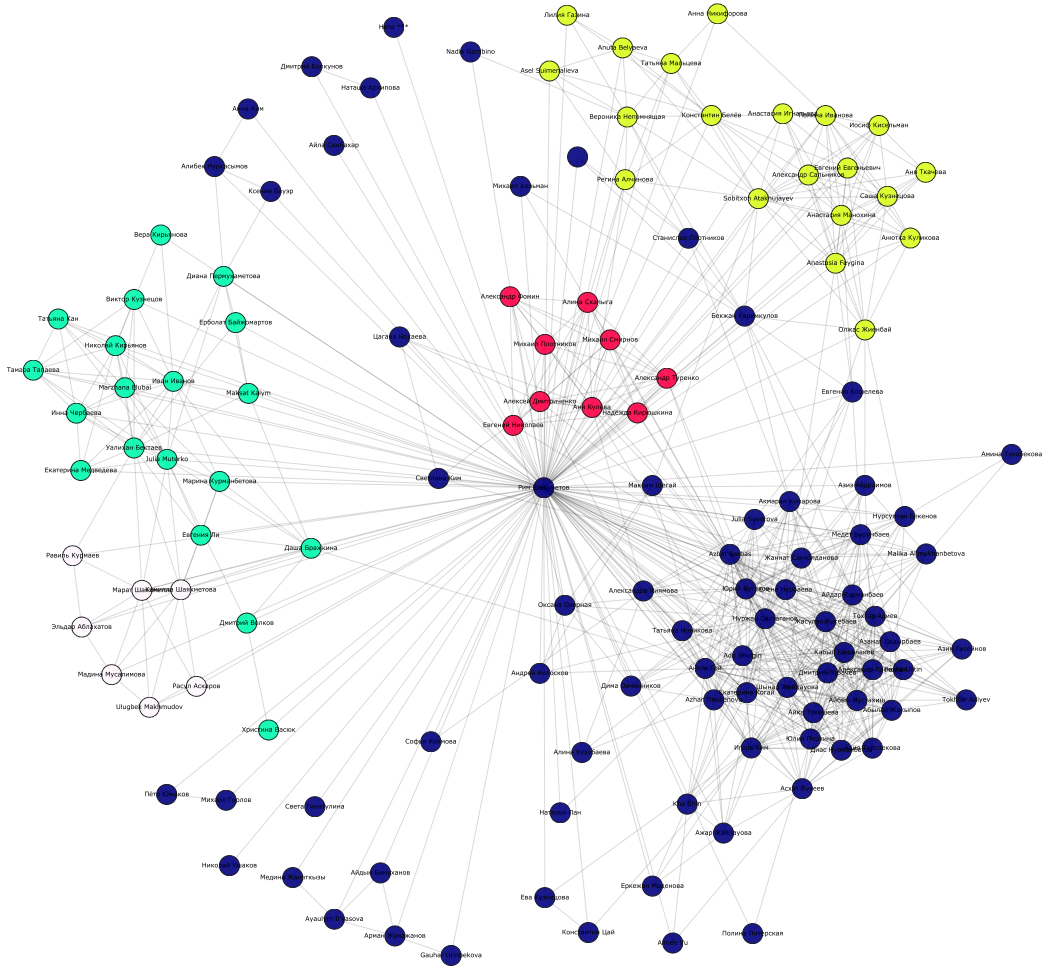
Community detection

Infomap (color)



Community detection

Label Propagation (color)



Summary

- Maximal Maximum Cliques correspond to my previous classmates.
- The most reasonable community detection has been made with Edge Betweenness and Infomap.
- Springlass was good at identifying small communities inside the big MSU community.
- Centrality measures correlate with how the person is sociable (except for me, which is obvious as the network is centered on me)

Used tools I

- Python (numpy/pandas)
- NetworkX
- igraph