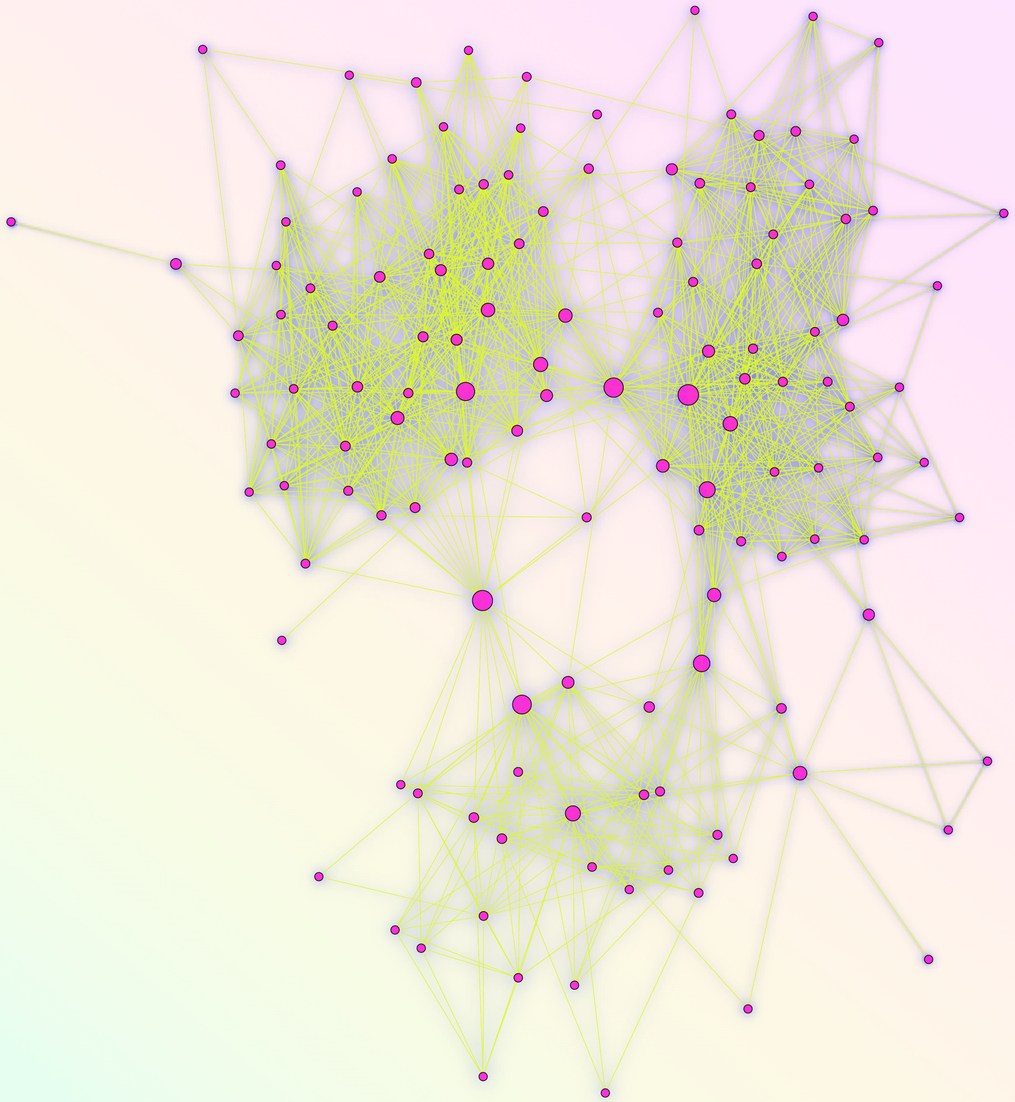
VK FRIENDS: GRAPH ANALYSIS

IVAN SKVORTSOV

GRAPH SOURCE

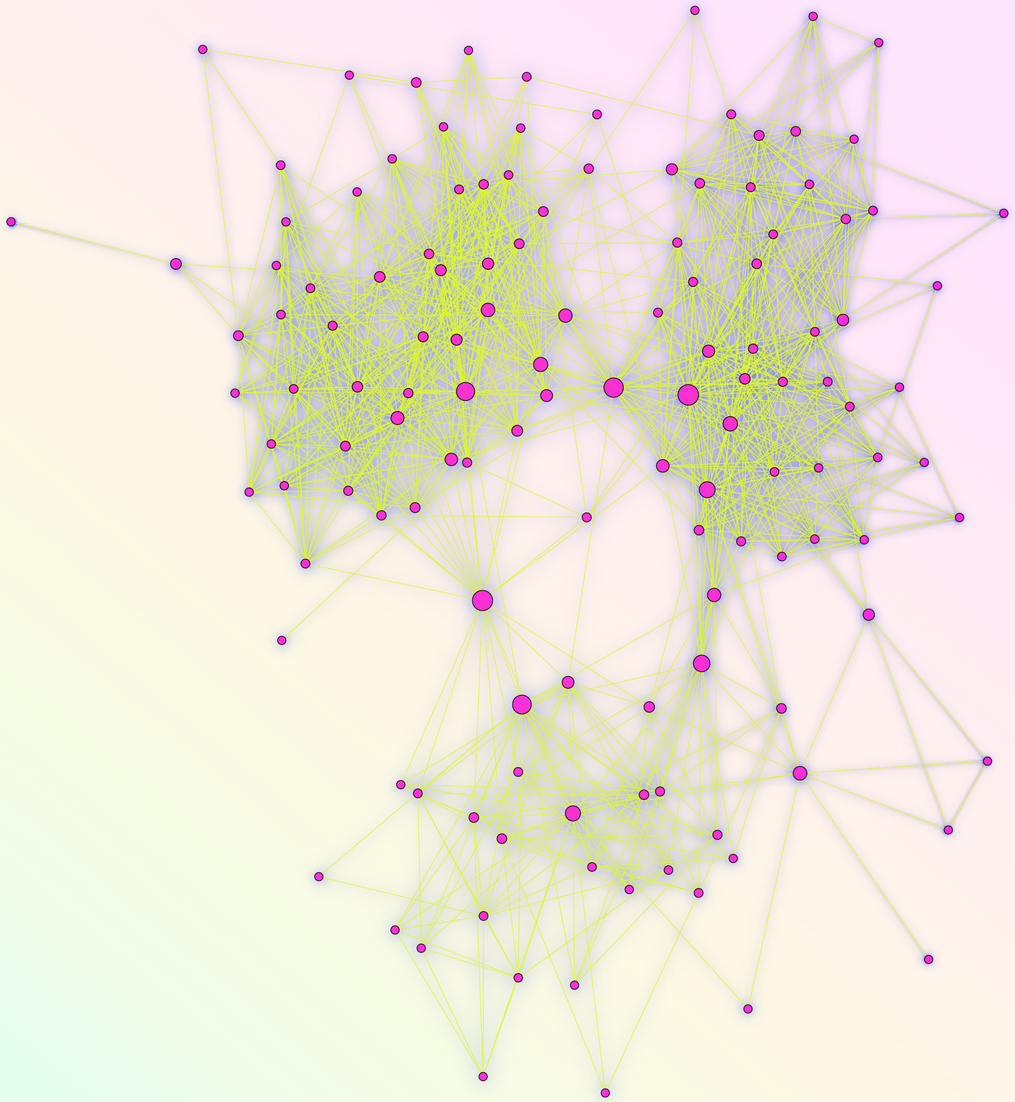
The graph was parsed using VK API and represents my ego network with 133 nodes and 1054 edges.

- The graph is undirected, since edges represent friendship connections between VK users.
- Additional node attributes, such as gender, city and university, were also collected.
- Nodes without any connections were removed, so we are looking at a connected graph.



GRAPH SUMMARY

- The «small world» property holds in this graph, with its diameter equal to 6.
- The graph is quite clustered, with average clustering coefficient equal to 0.56

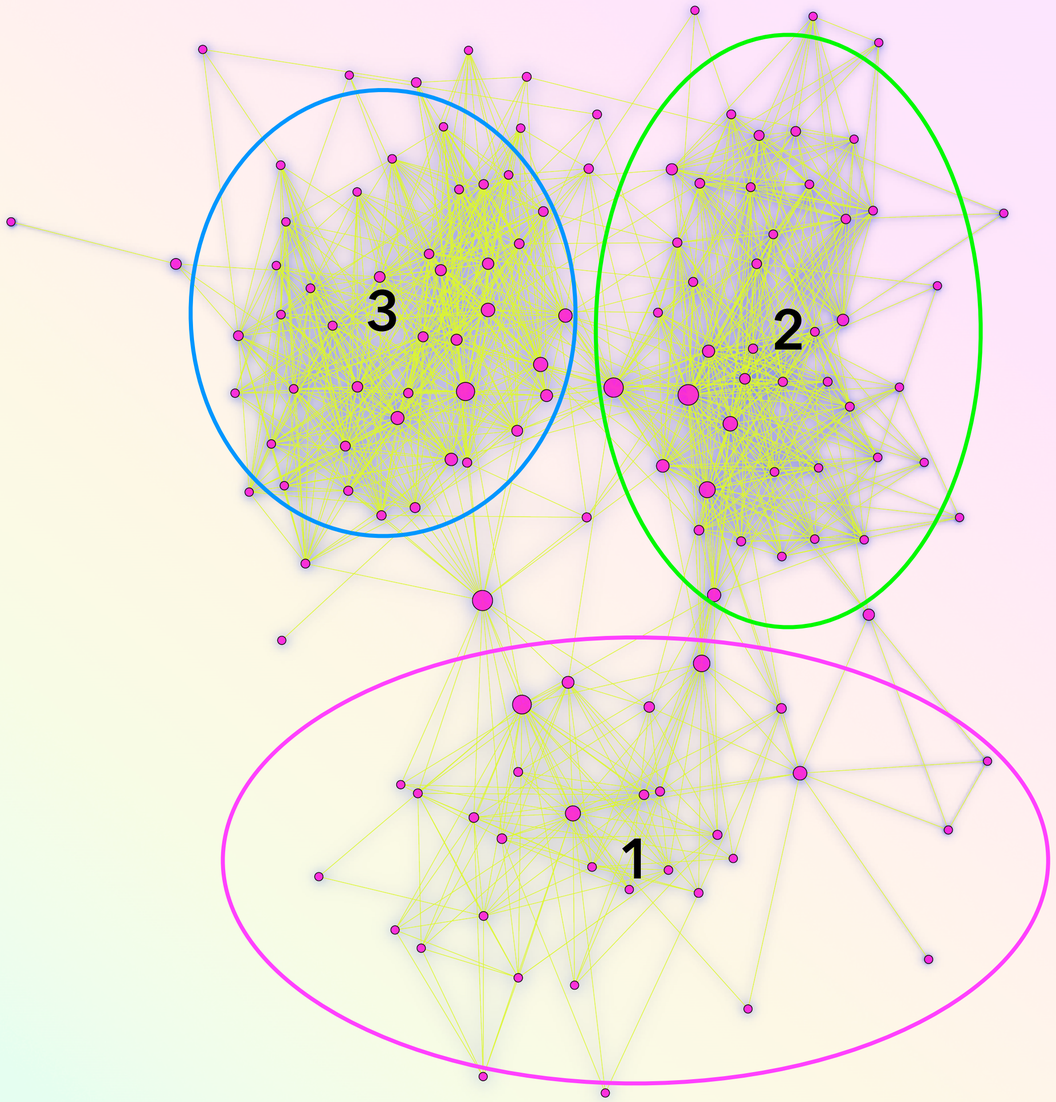


GRAPH SUMMARY

The graph has three distinct segments (communities), which represent three primary sources of my friends:

1. School pals
2. Guys from social sciences olympiad
3. HSE Faculty of Economics classmates

These communities will be defined later.

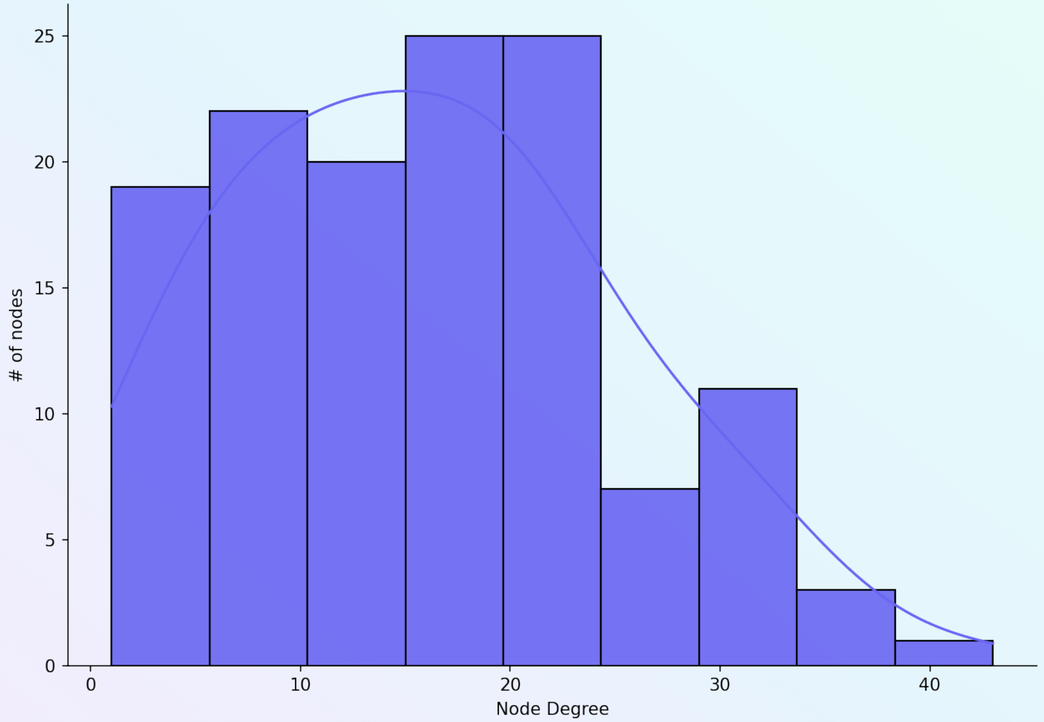


DEGREE DISTRIBUTION

Two models were fitted:

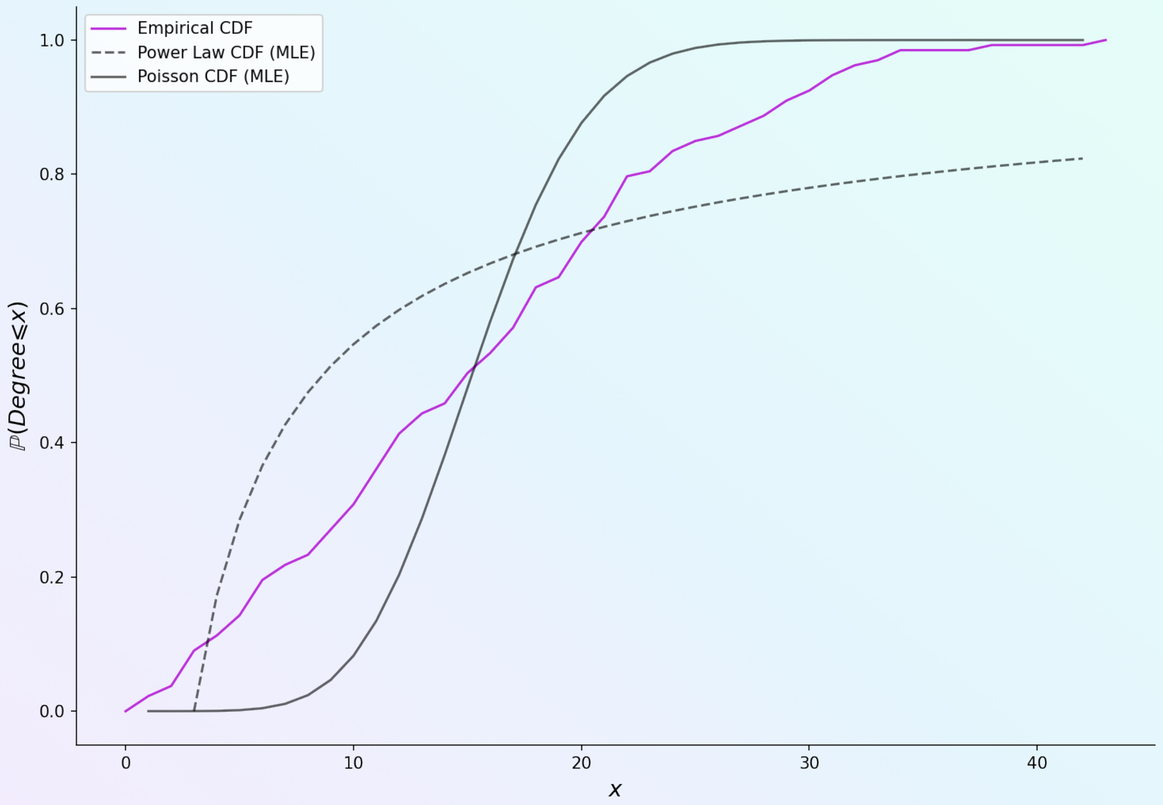
- Power law MLE + Kolmogorov-Smirnov test;
- Poisson MLE.

Both models failed to fit the real degree distribution of the network 🥲



DEGREE DISTRIBUTION

Model	Parameters	Estimated Value
Power Law	α	1.66
	xmin	3
Poisson	λ	15.8



RANDOM GRAPH MODELS

To compare our real-world graph to random ones, I used 4 models:

- Erdős–Rényi, with parameters (N, p) estimated directly from the empirical graph;
- Barabási–Albert, with N = 133 and M = 2;
- Watts–Strogatz, with parameters N, k estimated from the empirical graph and with p = 0.5

Model	Diameter	Radius	Average clustering	Average shortest path length
Empirical	6	3	0.555	2.531
Erdős-Rényi	3	3	0.126	1.995
Barabási-Albert	6	3	0.067	3.275
Watts-Strogatz	3	3	0.168	2.126

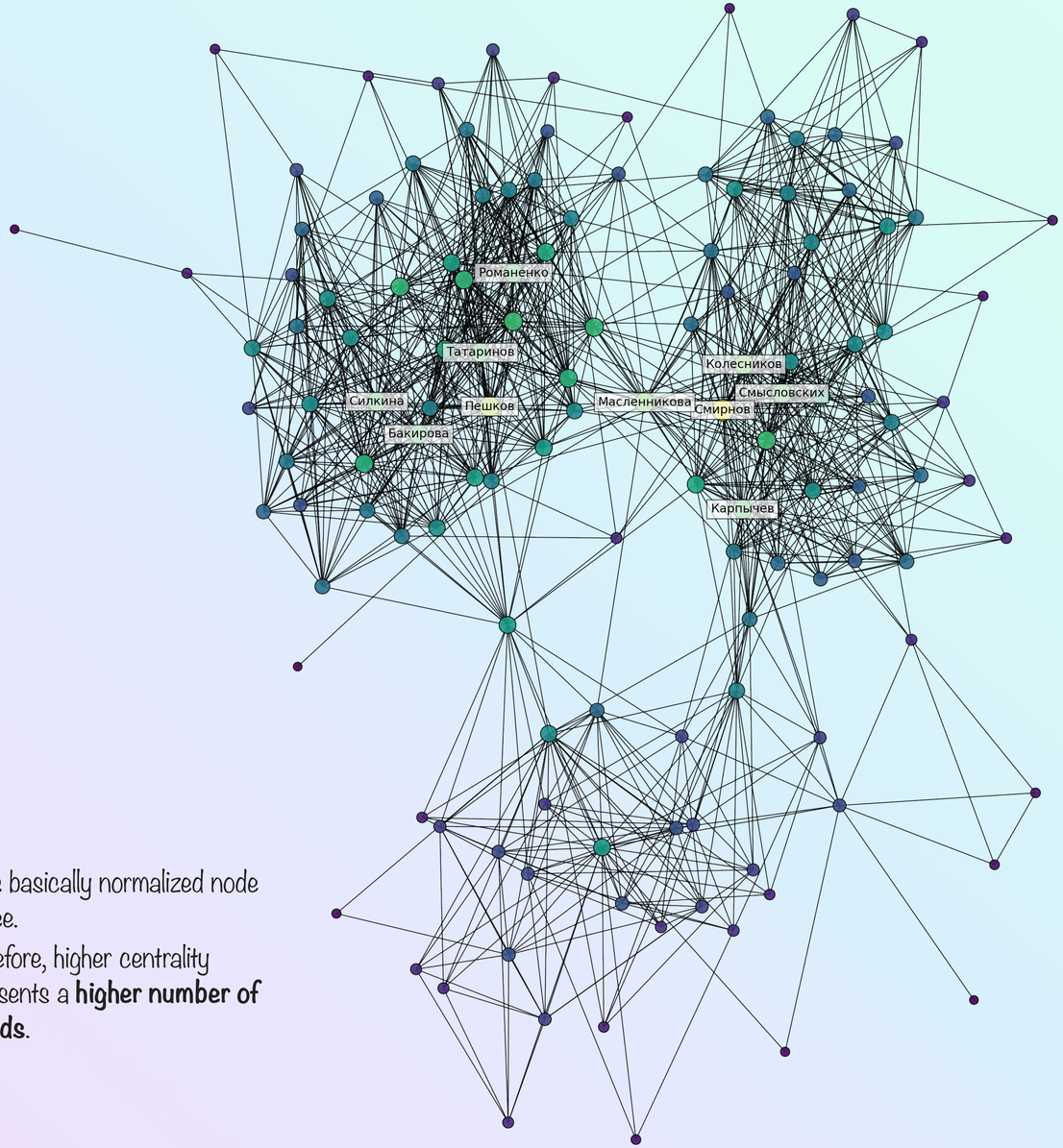
BEST RESULTS ARE HIGHLIGHTED

CENTRALITIES

DEGREE CENTRALITY

First Name	Last Name	DC
Фил	Смирнов	0.326
Максим	Пешков	0.288
Игорь	Колесников	0.258
Света	Масленникова	0.258
Елена	Смысловских	0.250
Борис	Карпычев	0.242
Артем	Татаринов	0.242
Таня	Бакирова	0.235
Саша	Романенко	0.235
Мария	Силкина	0.235

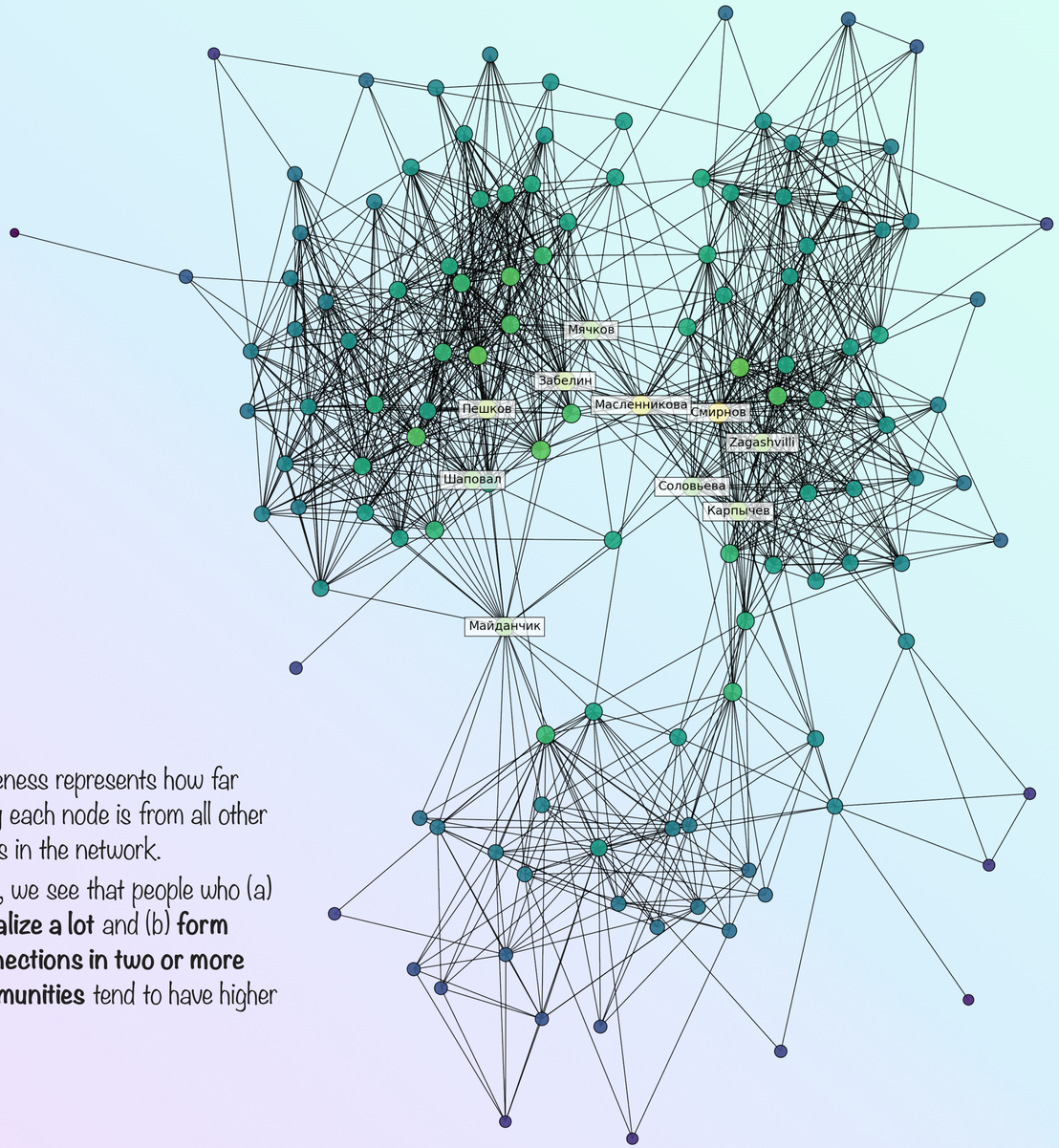
- DC is basically normalized node degree.
- Therefore, higher centrality represents a **higher number of friends**.



CENTRALITIES

CLOSENESS CENTRALITY

First Name	Last Name	CC
Фил	Смирнов	0.548
Света	Масленникова	0.532
Максим	Пешков	0.510
Алексей	Забелин	0.506
Борис	Карпычев	0.500
Angelika	Zagashvilli	0.493
Дмитрий	Мячков	0.489
Миша	Майданчик	0.485
Света	Шаповал	0.484
Арина	Соловьева	0.482

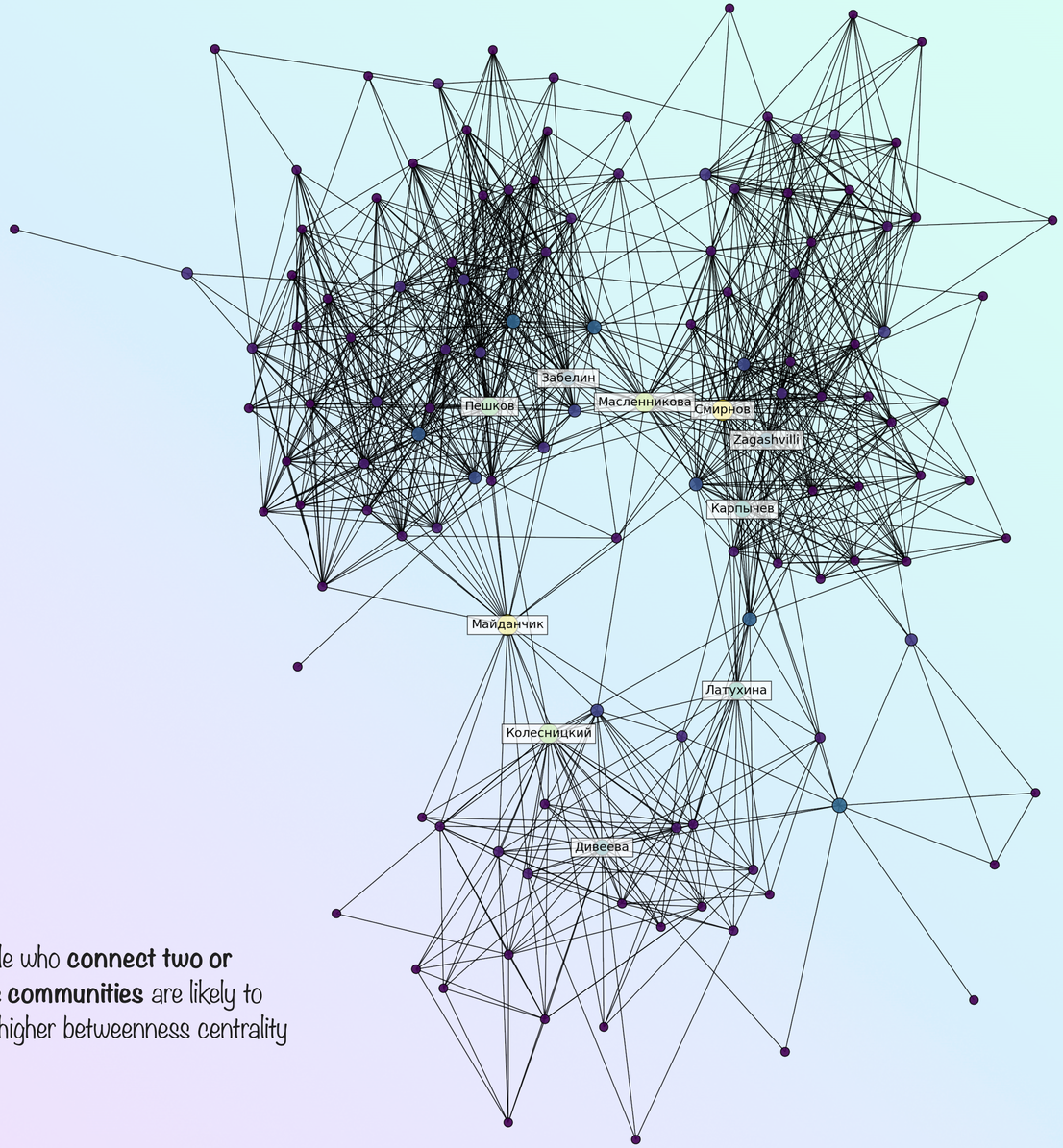


- Closeness represents how far away each node is from all other nodes in the network.
- Thus, we see that people who (a) **socialize a lot** and (b) **form connections in two or more communities** tend to have higher CC.

CENTRALITIES

BETWEENNESS CENTRALITY

First Name	Last Name	BC
Фил	Смирнов	0.112
Миша	Майданчик	0.108
Света	Масленникова	0.097
Андрей	Колесницкий	0.090
Максим	Пешков	0.083
Екатерина	Латухина	0.065
Борис	Карпычев	0.059
Варвара	Дивеева	0.051
Angelika	Zagashvilli	0.044
Алексей	Забелин	0.042



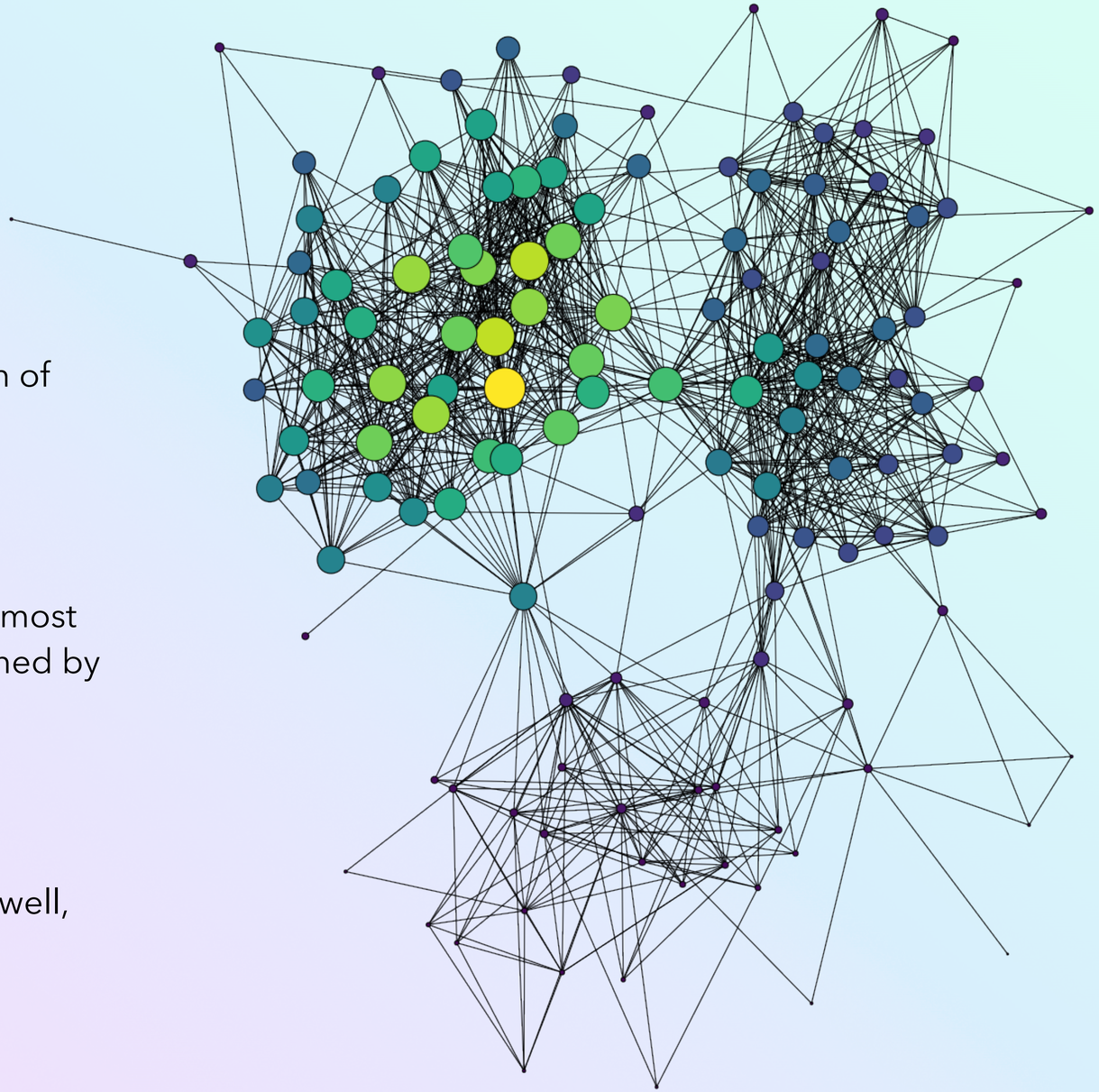
- People who **connect two or more communities** are likely to have higher betweenness centrality

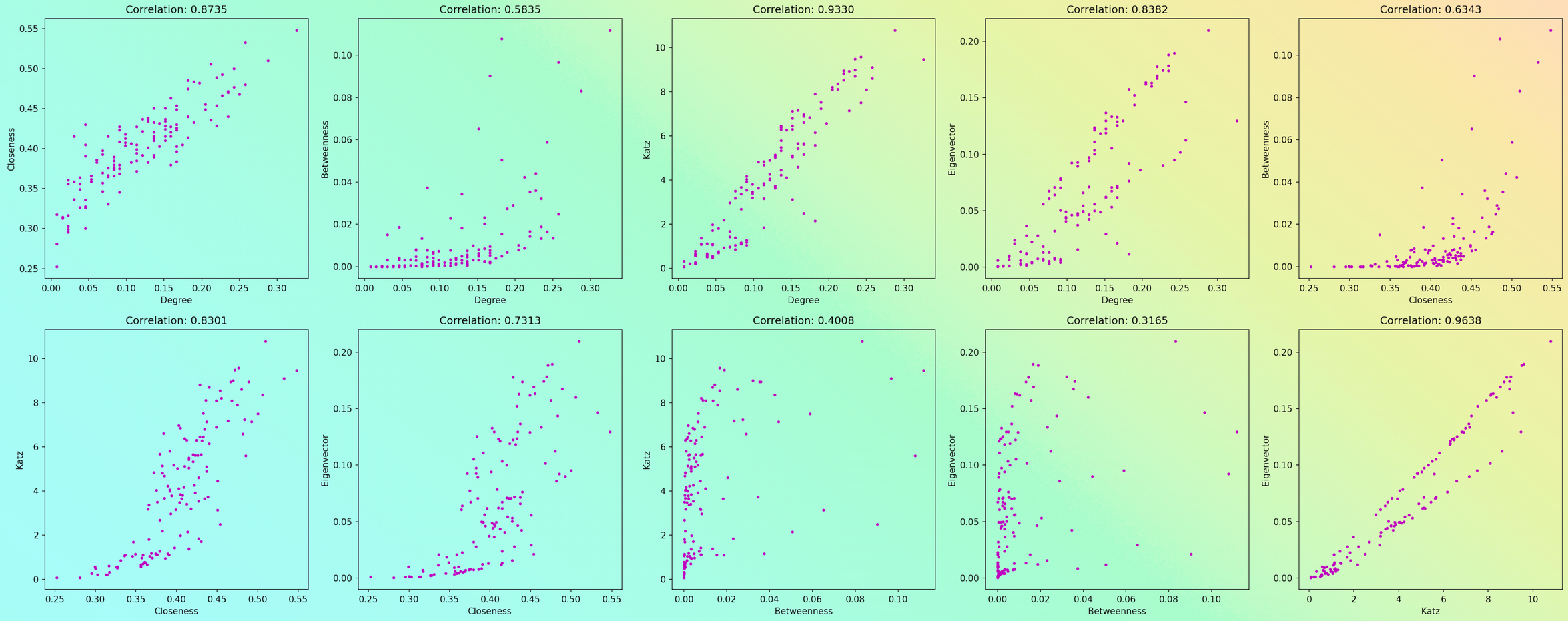
CENTRALITIES

EIGENVECTOR CENTRALITY

Note how eigenvector centrality differs between communities!

- In this case, centrality of a node is defined recursively as sum of centralities of its neighbors.
- So this centrality highlights people with most «influential» connections.
- It seems that Faculty of Economics community is overall the most important in terms of eigenvector centrality. Which is explained by two factors:
    - Economists tend to have more connections within the community.
    - Some of them communicate with other communities as well, which makes them influential neighbors and raises eigenvector centrality.



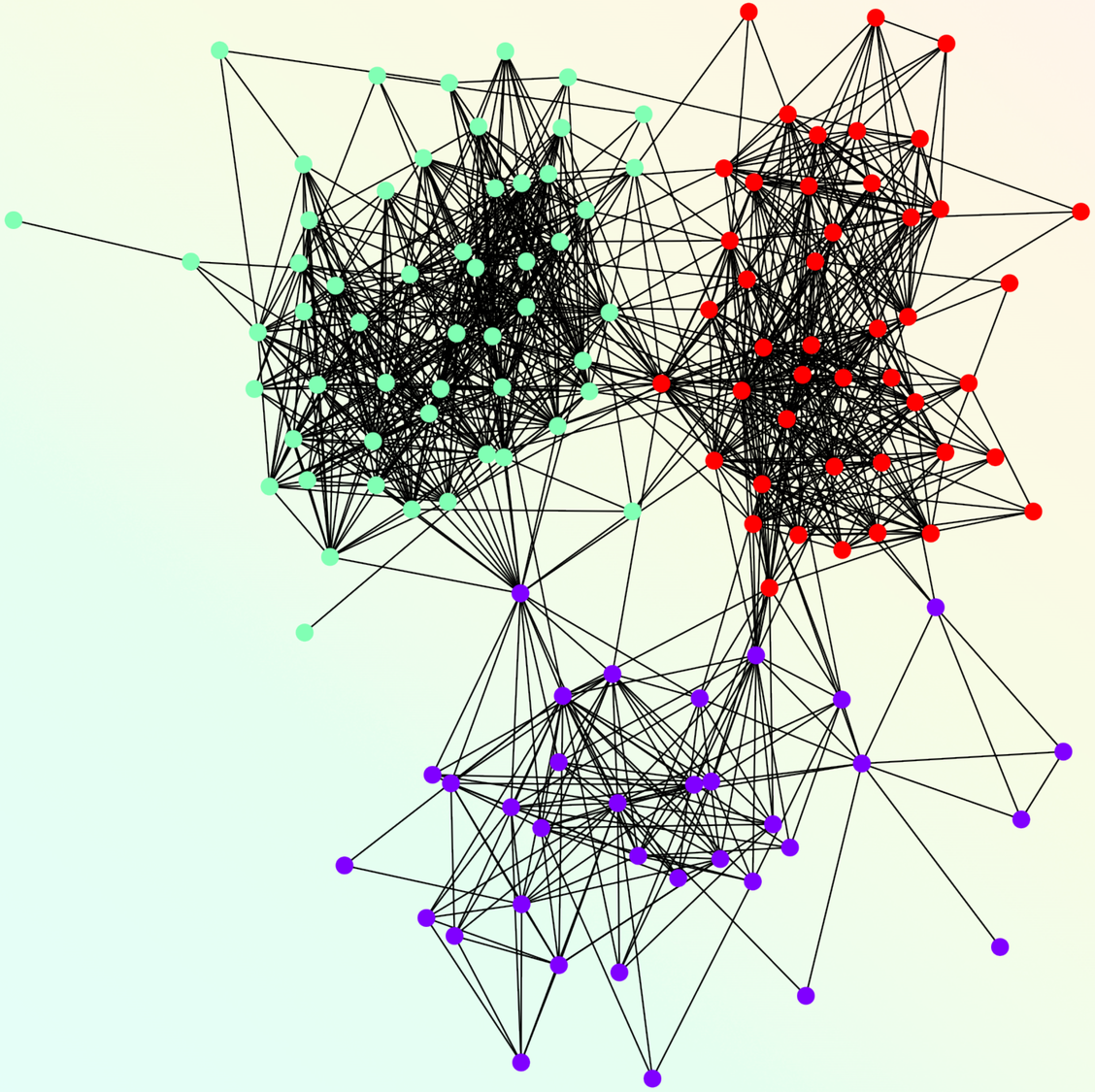

Correlation: 0.8735
Closeness
Degree
Correlation: 0.5835
Betweenness
Degree
Correlation: 0.9330
Katz
Degree
Correlation: 0.8382
Eigenvector
Degree
Correlation: 0.6343
Betweenness
Closeness
Correlation: 0.8301
Katz
Closeness
Correlation: 0.7313
Eigenvector
Closeness
Correlation: 0.4008
Katz
Betweenness
Correlation: 0.3165
Eigenvector
Betweenness
Correlation: 0.9638
Eigenvector
Katz

COMMUNITIES

LAPLACIAN & K-MEANS

I found that for this network, using spectral embedding along with K-Means is the best option.

This method works perfectly and yields exactly what we expect from community detection 👌



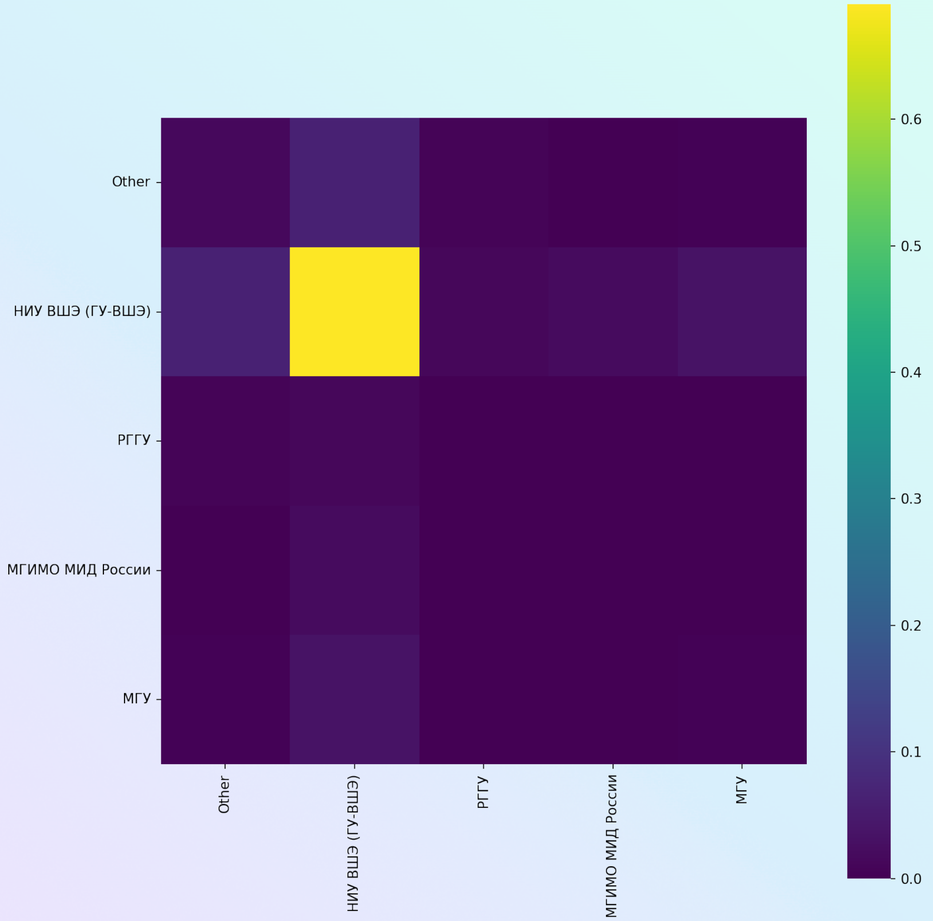
ASSORTATIVE MIXING

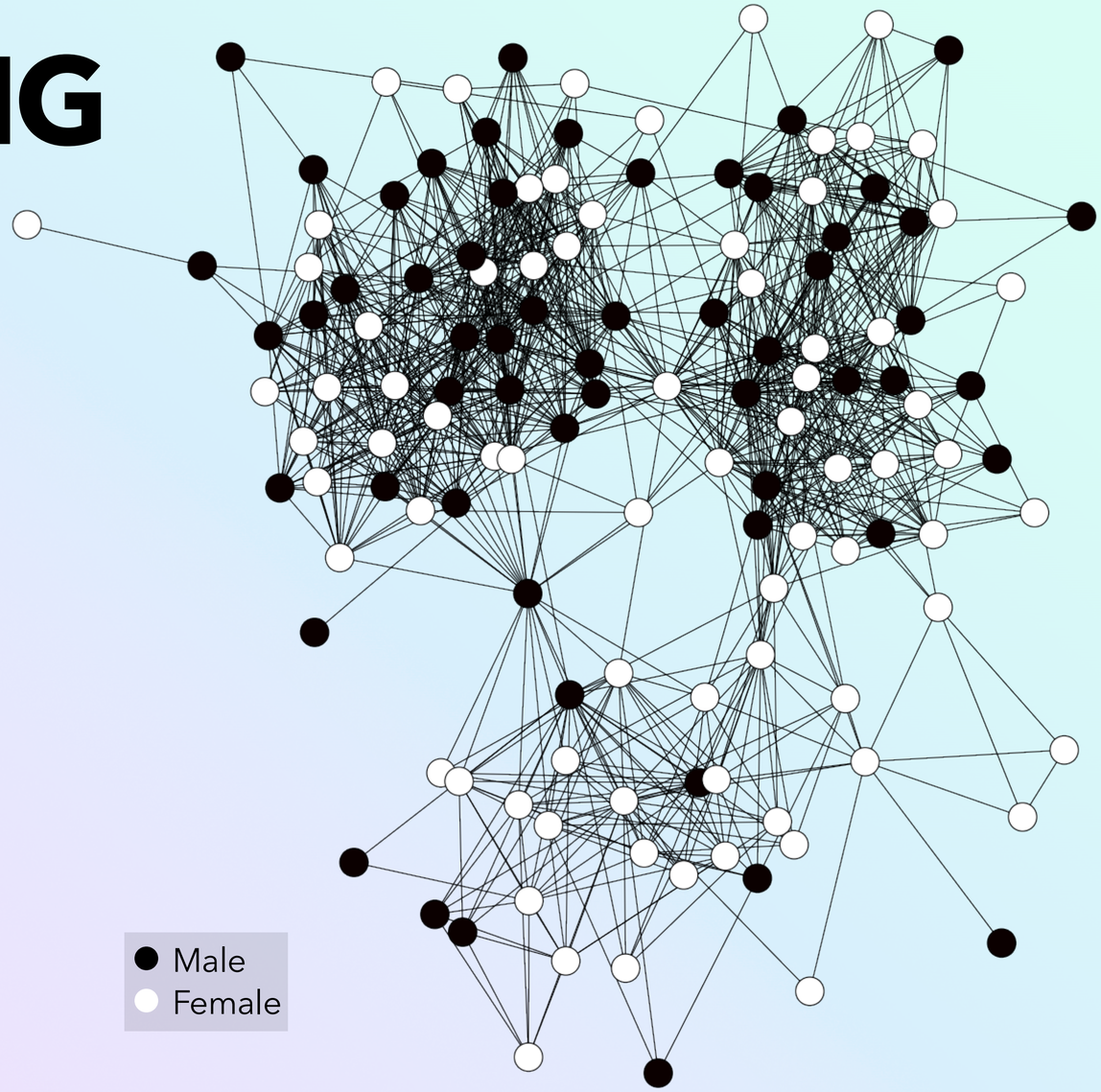
UNIVERSITY

Most of my friends are from HSE. Let's see if HSE students love other HSE students, i. e. if assortative mixing for university = «HSE» is high

(spoiler: of course it is 🤪)

⚠️ It's worth noting that the graph is highly biased towards HSE students. It is safe to assume that other universities in population also have high assortative mixing, they're just not well represented in this particular graph.



ASSORTATIVE MIXING

SEX