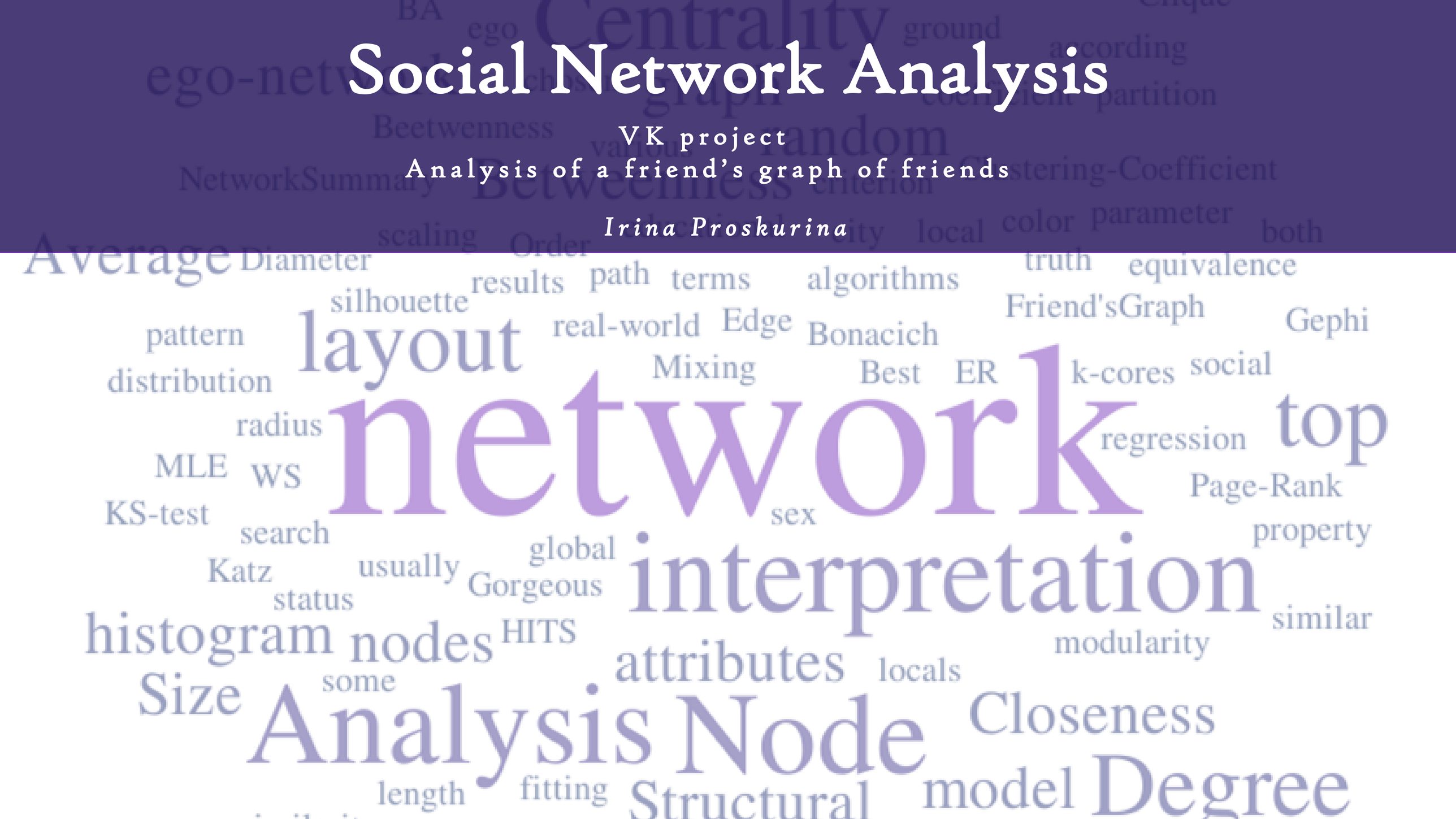
Social Network Analysis

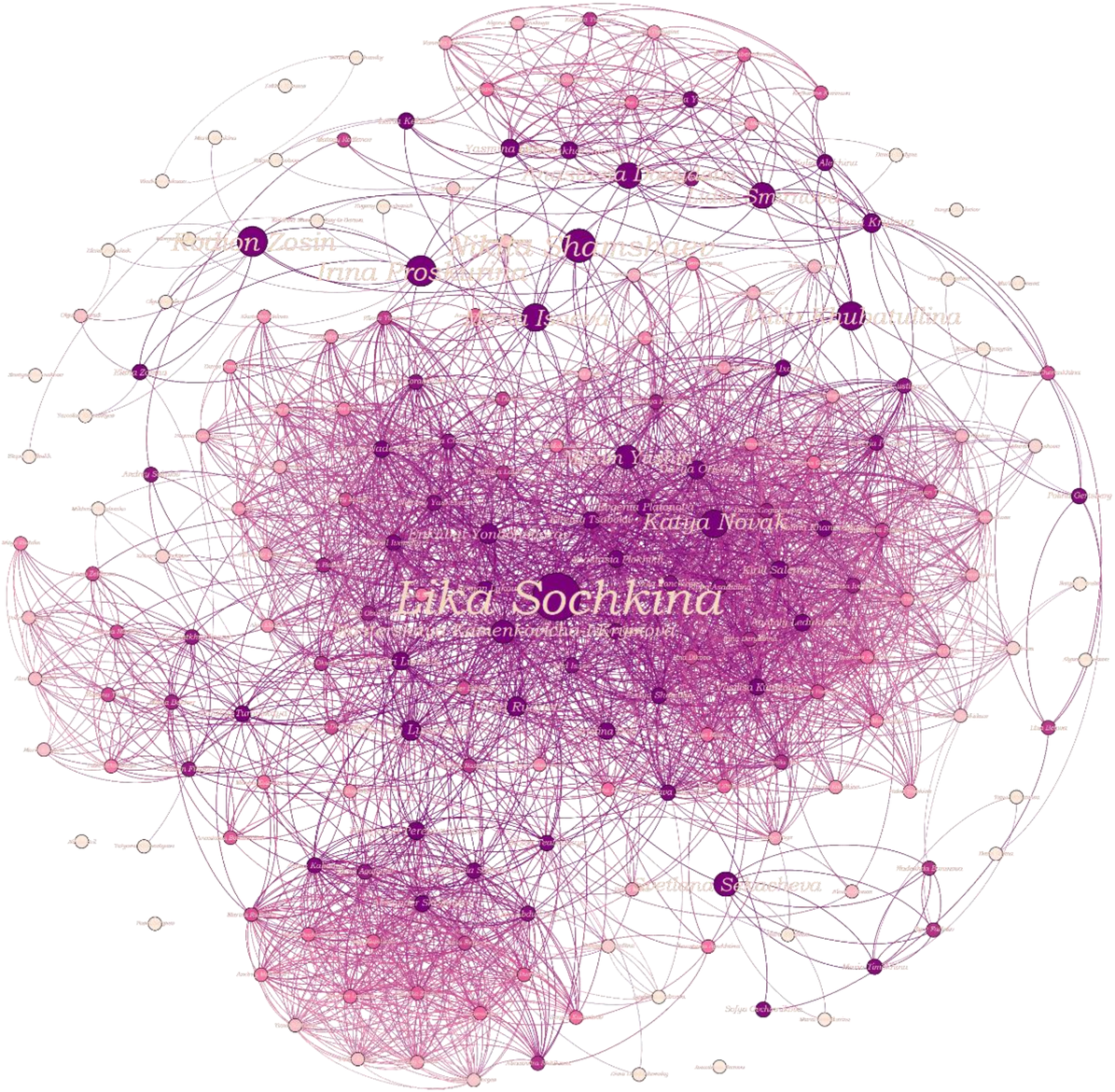
VK project

Analysis of a friend's graph of friends

Irina Proskurina



NETWORK SUMMARY



Node size is based on betweenness centrality
Color of nodes → degrees

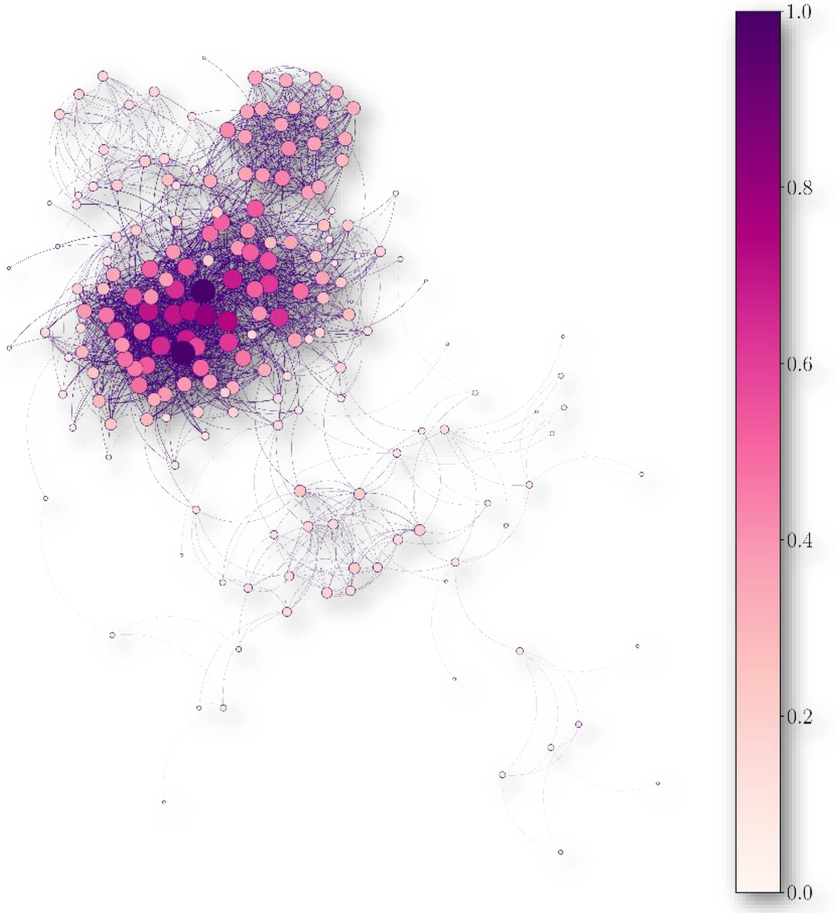
Number of nodes: 201
Number of edges: 1553
Nodes' attributes: sex, city of location, university, educational status
Average node degree: 15,38
Average clustering coefficient: 0,521
Average path: 1,98
Density: 0.076

GCC:

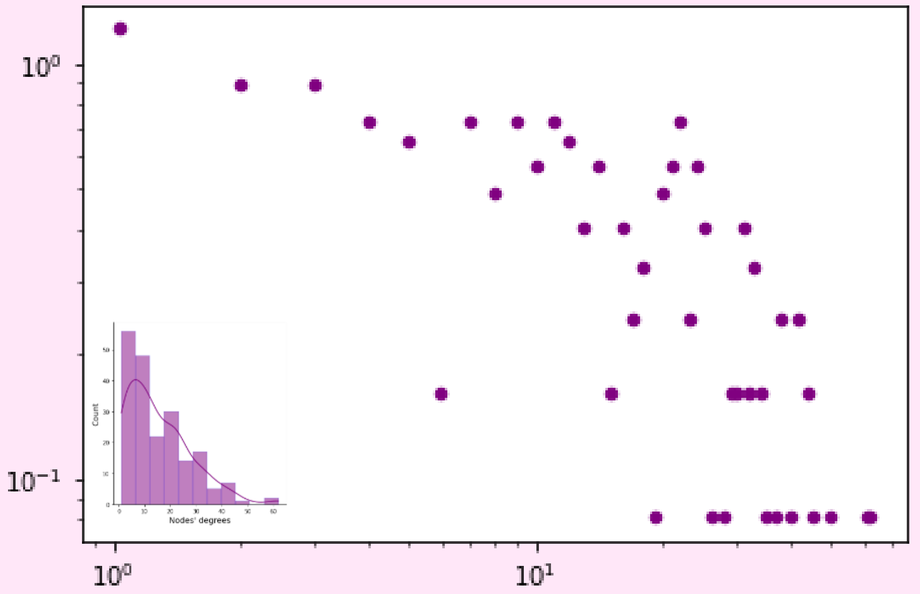
Number of nodes: 190
Number of edges: 1537
Radius: 4
Diameter: 8
Average clustering coefficient: 0, 521
Average shortest path: 3,055

Degree distribution

Average degree = 15,38

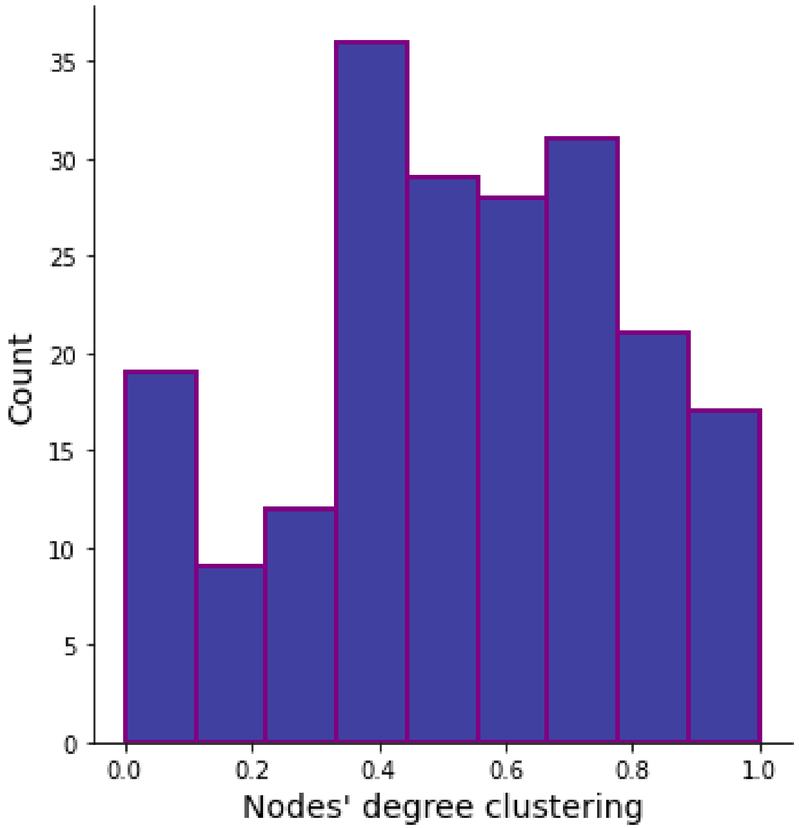


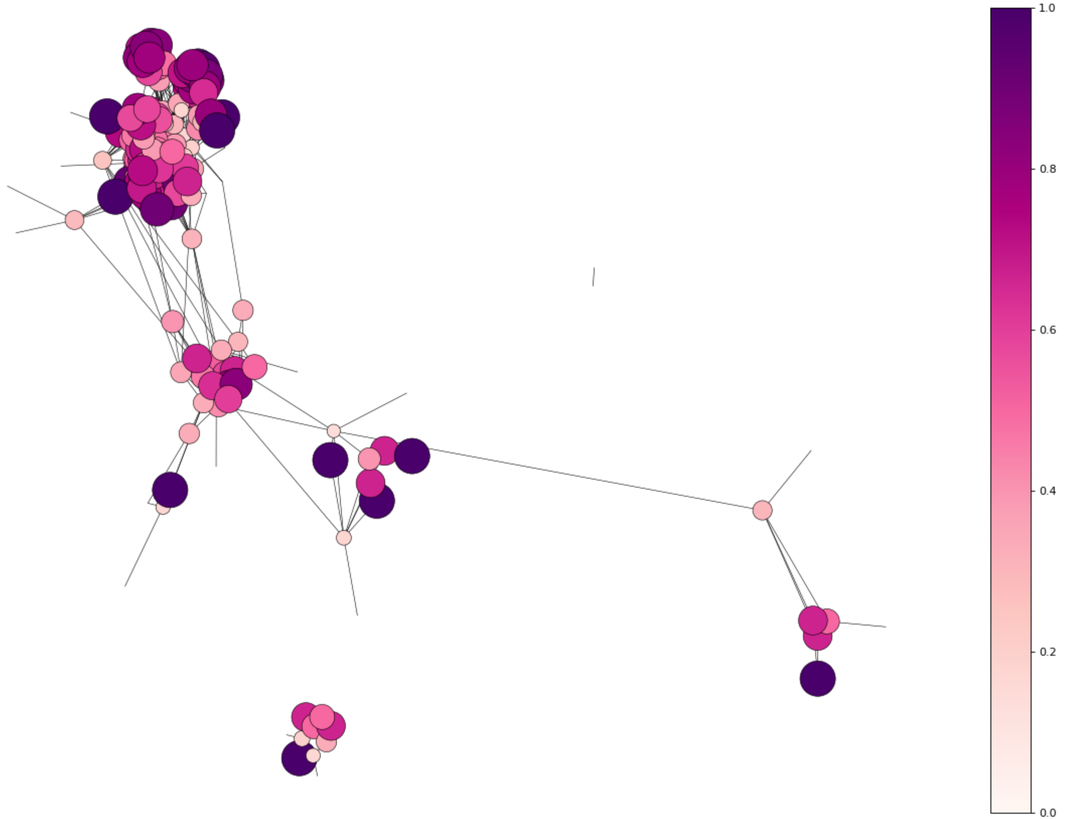
PDF



Clustering coefficients

Average coefficient = 0,521





Fitting models

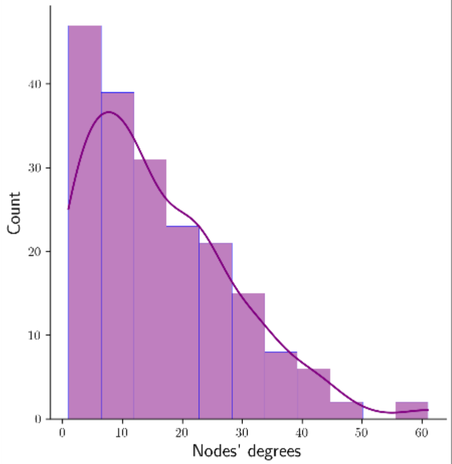
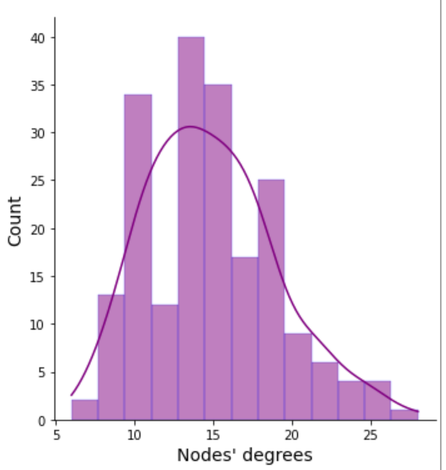
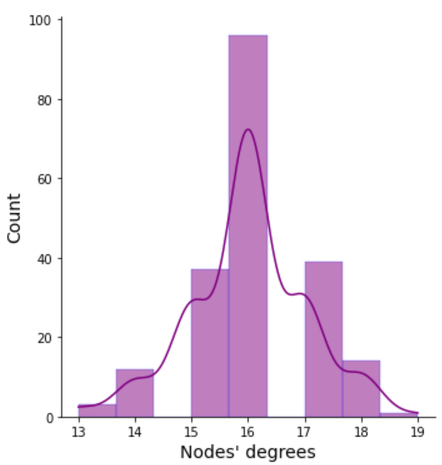
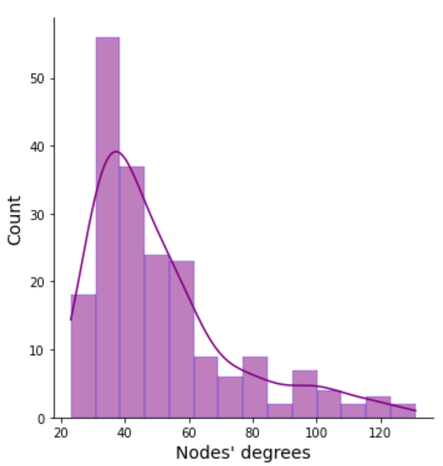
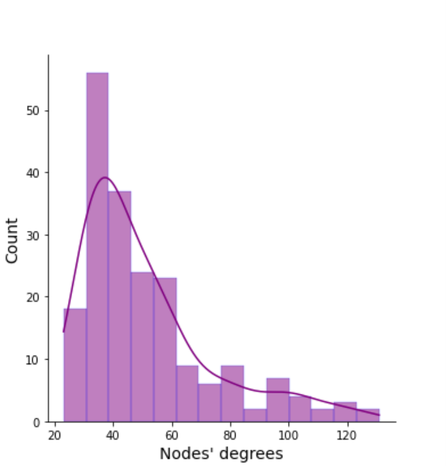
The best models are Watts-Strogatz and Havel-Hakimi

	Original	Erdos-Renyi model	Watts-Strogatz	Barabasi-Albert	Havel-Hakimi
Number of nodes	202	202	202	202	202
Number of edges	1553	1500	1616	5160	1553
Diameter	8	4	4	3	23
Radius	4	3	3	2	12
Average node degree	15,376	14,851	16	51,089	15,376
Average clustering coefficient	0,522	0,075	0,559	0,357	0,535
Average shortest path	1,982	2,24	2,702	1,746	1,364
Density	0,076	0,074	0,08	0,254	0,076

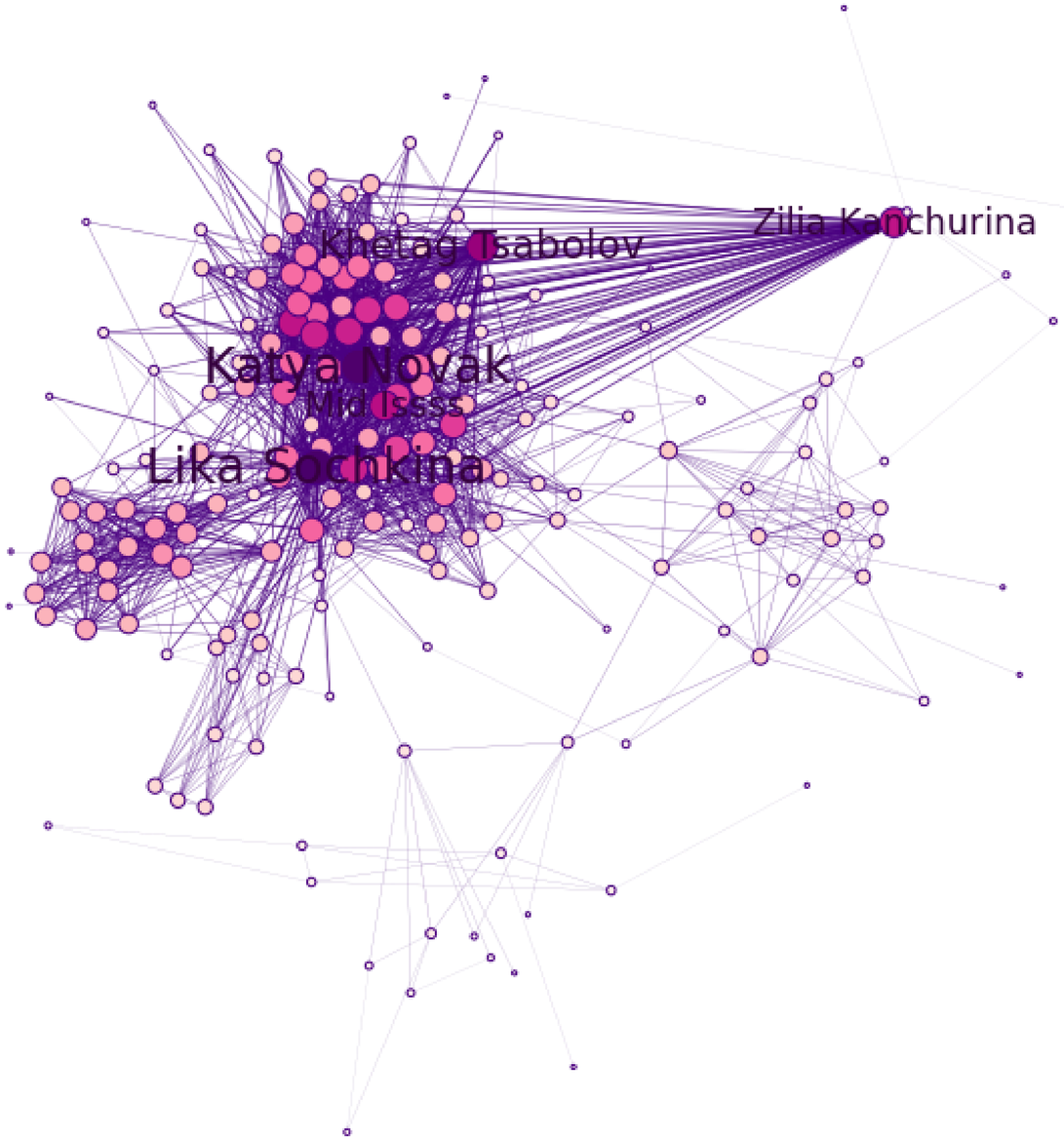
1. *Watts-Strogatz parameters: , k nearest neighbors in ring topology = 16, p = 0.1*
2. *Havel-Hakimi model's parameter is a degree_sequence*

Degree distributions of fitting models

The closest models are: Watts-Strogatz and Havel-Hakimi

Original	Erdos-Renyi model	Watts-Strogatz	Barabasi-Albert	Havel-Hakimi
				

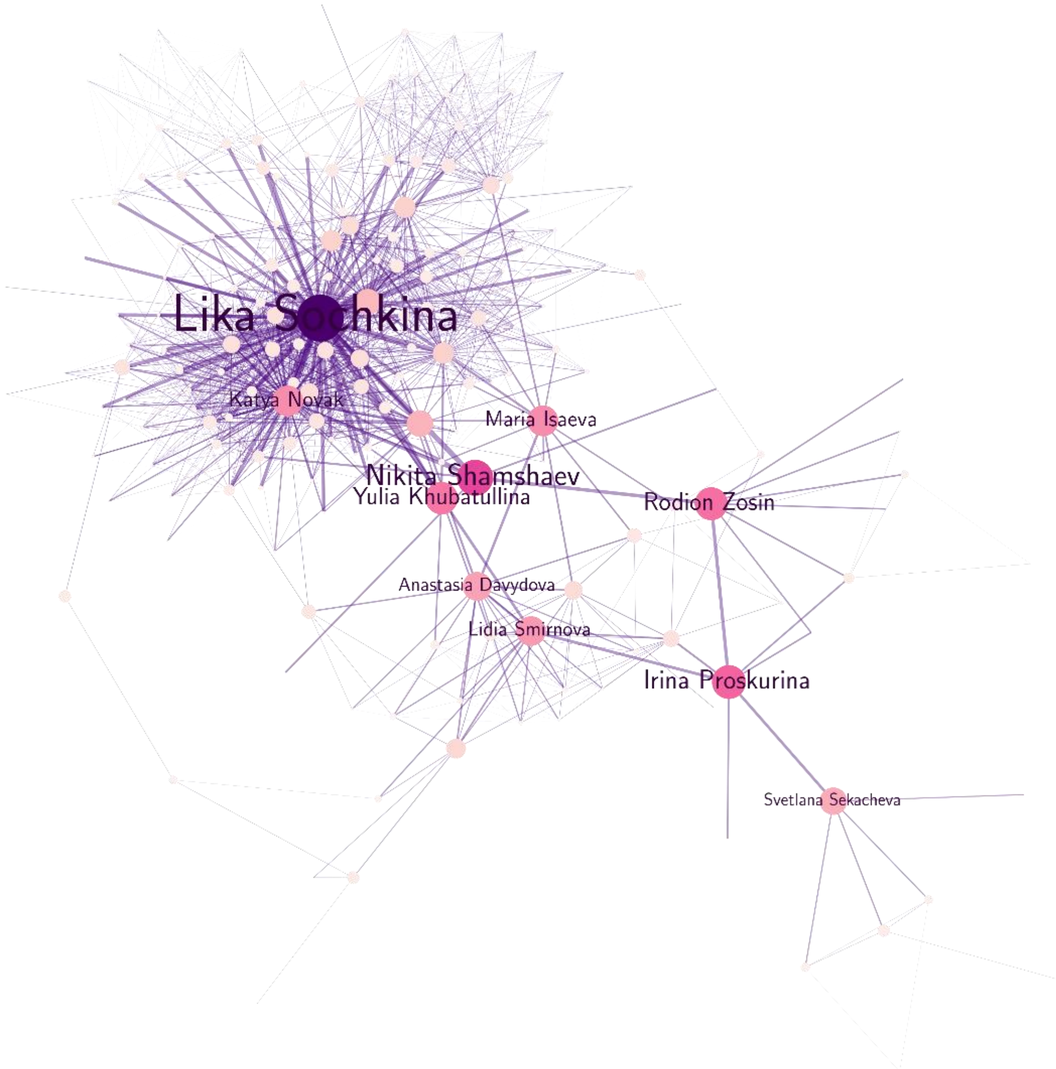
Centralities: degree



Top 5 nodes include a colleague and groupmates:

1. Lika Sochkina(0,308)
2. Katya Novak (0,303)
3. Khetag Tsabolov (0,250)
4. Mid Issss (0,224)
5. Zilia Kanchurina (0,219)

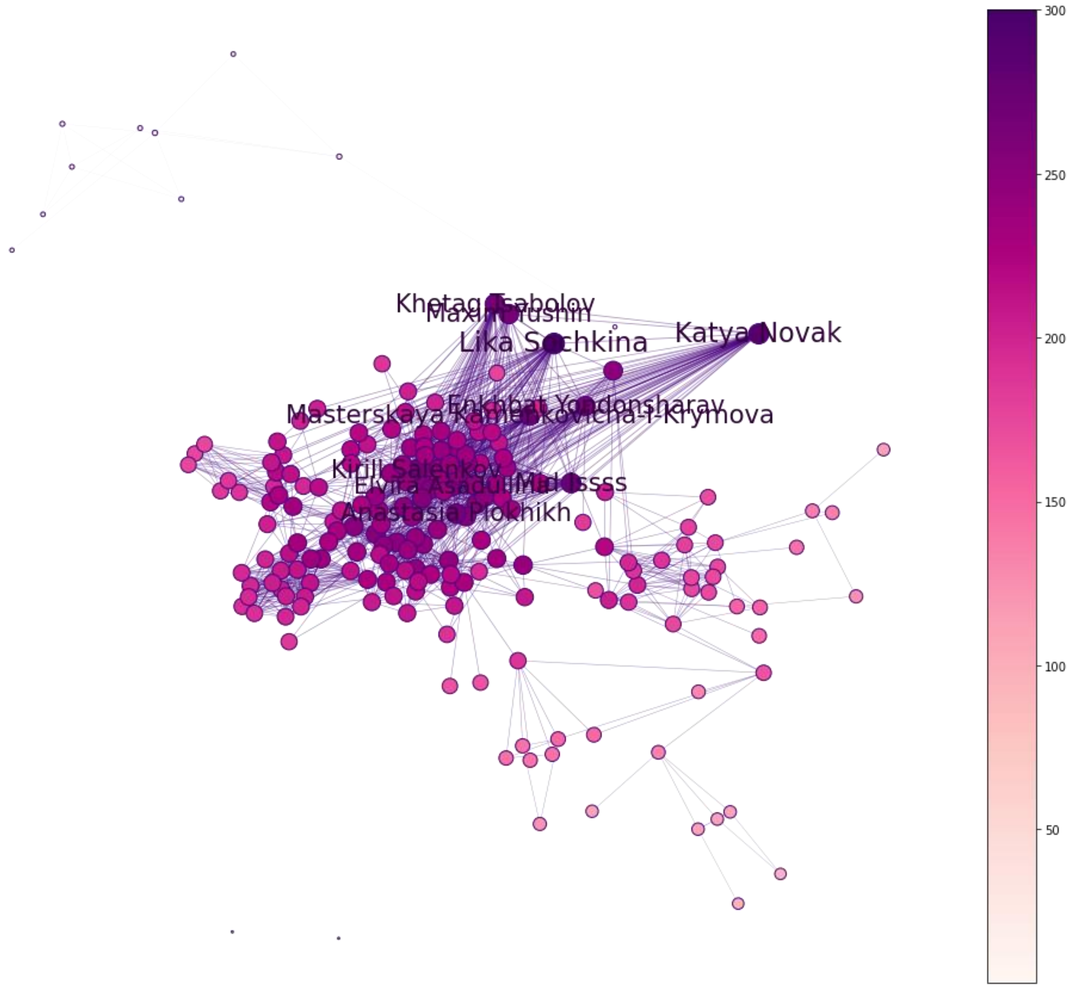
Centralities: betweenness



Top 5 nodes include best friends and a relative:

1. Lika Sochkina (0,171)
2. Nikita Shamshaev (0,097)
3. Yulia Khubatullina (0,087)
4. Irina Proskurina (0,086)
5. Rodion Zosin (0,081)

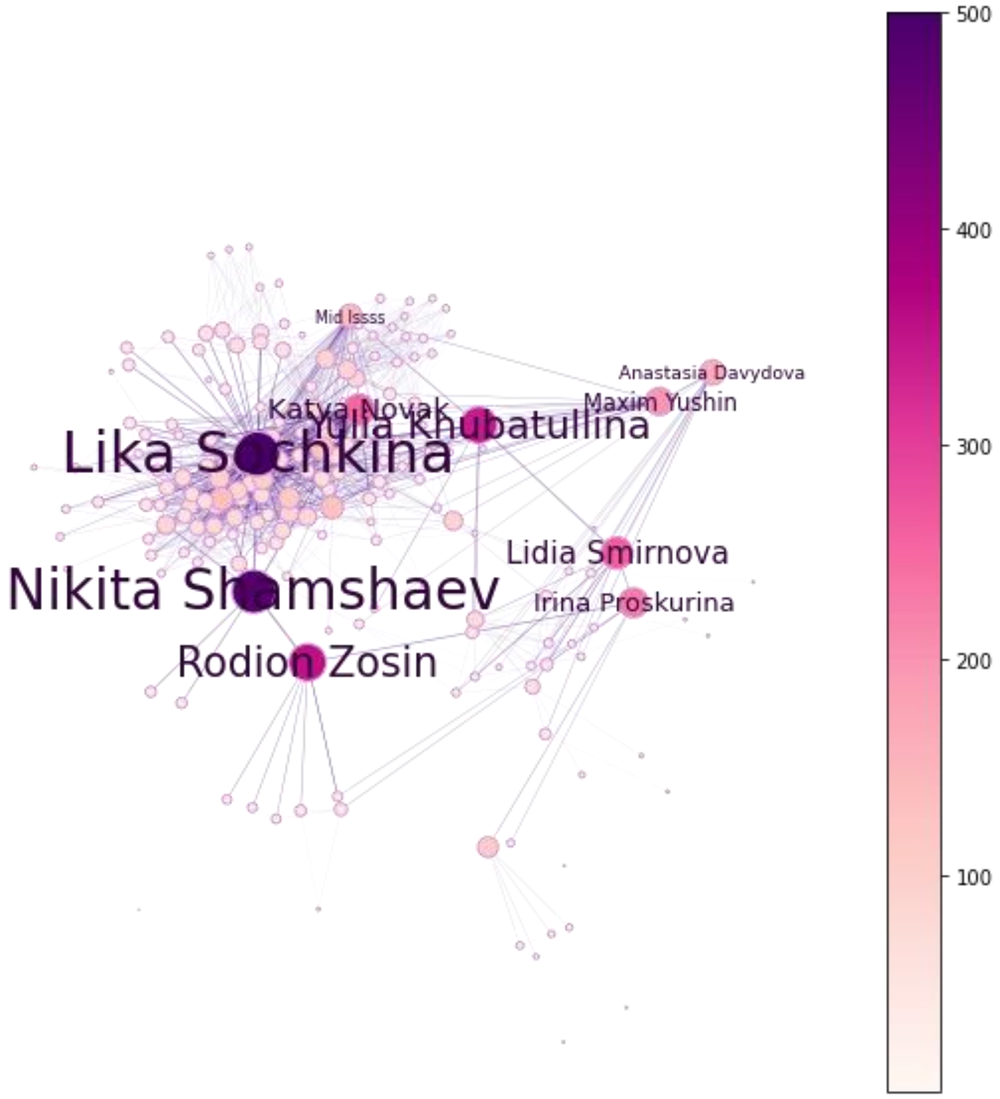
Centralities: closeness



Top 5 nodes include best friends and a relative:

1. Lika Sochkina (0,465)
2. Katya Novak (0,445)
3. Mid Issss (0,424)
4. Masterskaya Kamenkovicha-I-Krymova (0,424)
5. Anastasia Plokhikh (0,413)

Centralities: Katz

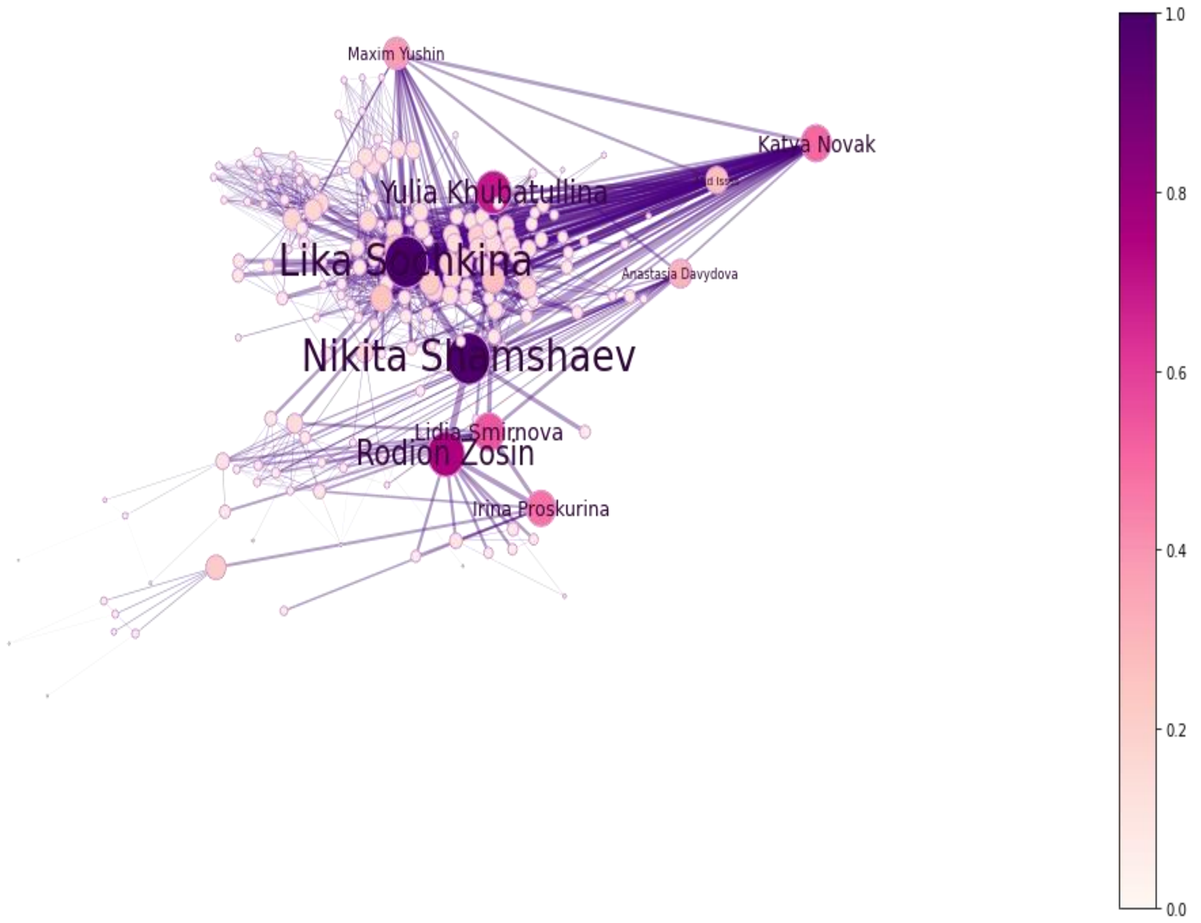


Top 5 nodes include same friends and previous colleagues:

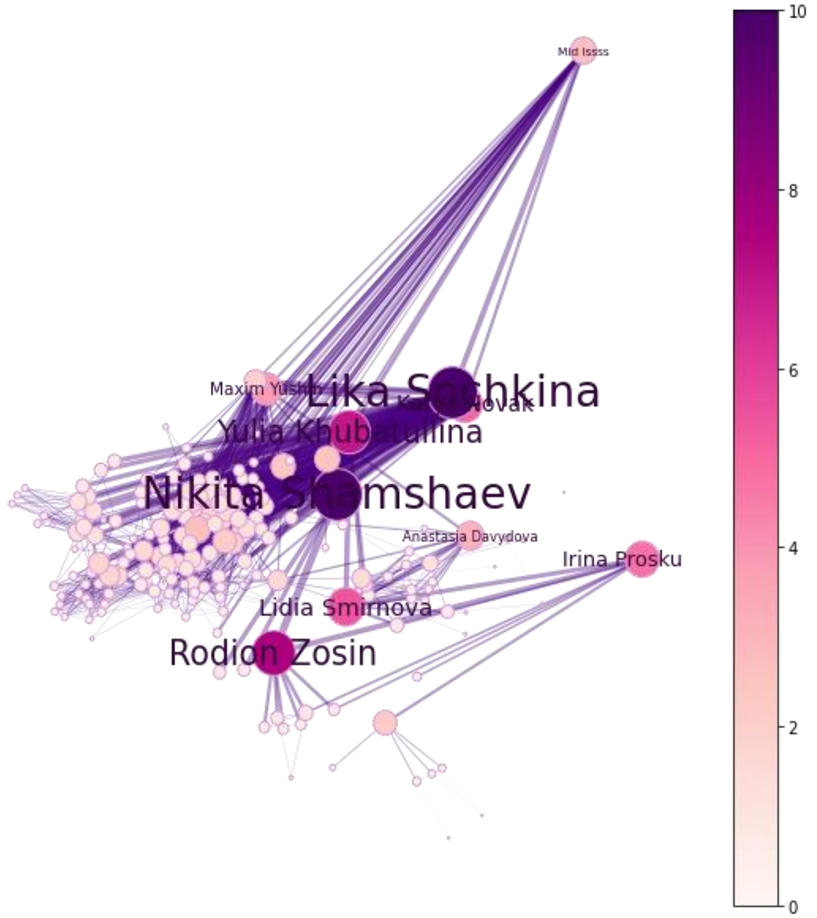
1. Lika Sochkina
2. Nikita Shamshaev
3. Rodion Zosin
4. Yulia Khubatullina
5. Lidia Smirnova

PageRank and HITS

PageRank (max 0,03)



HITS (max 0,055)

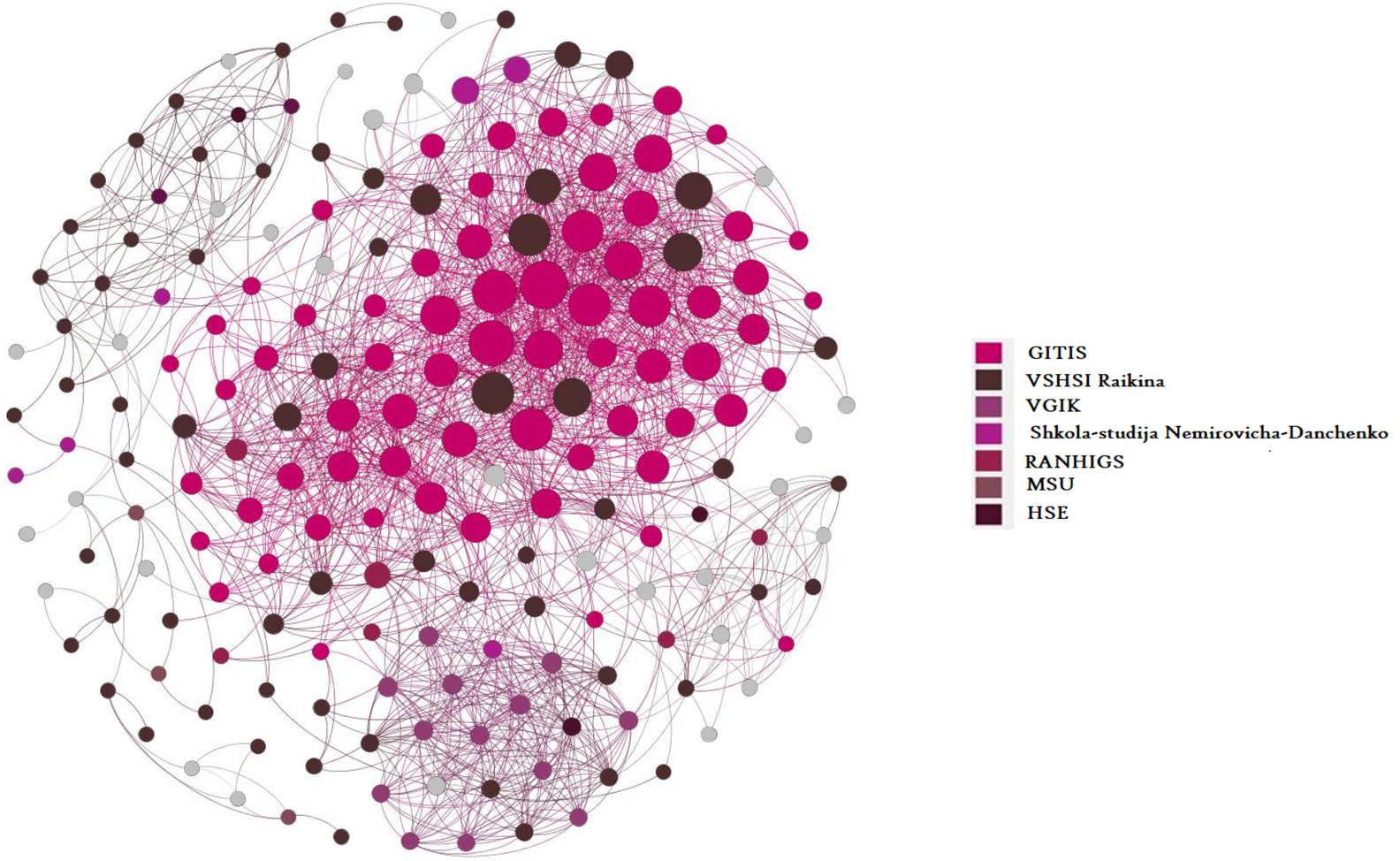


Assortative mixing

Node degree assortativity: *0.217*

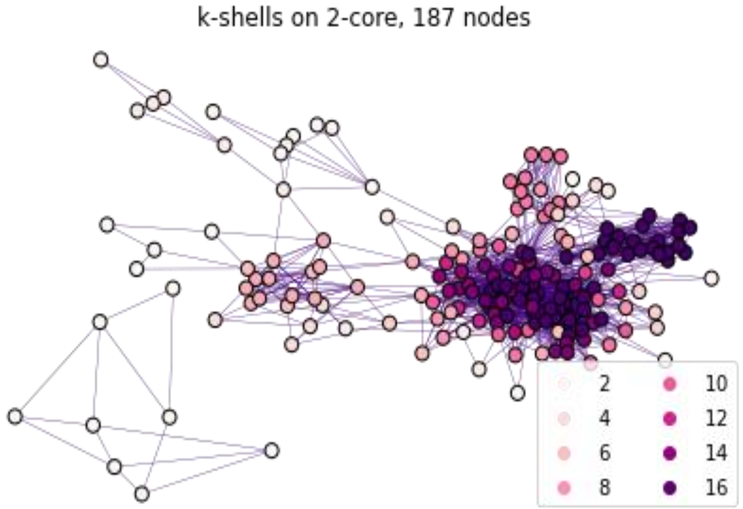
Assortativity for node attributes:

1. City – 0,023
2. Sex – 0,089
3. University – 0,216
4. Education status – 0,020

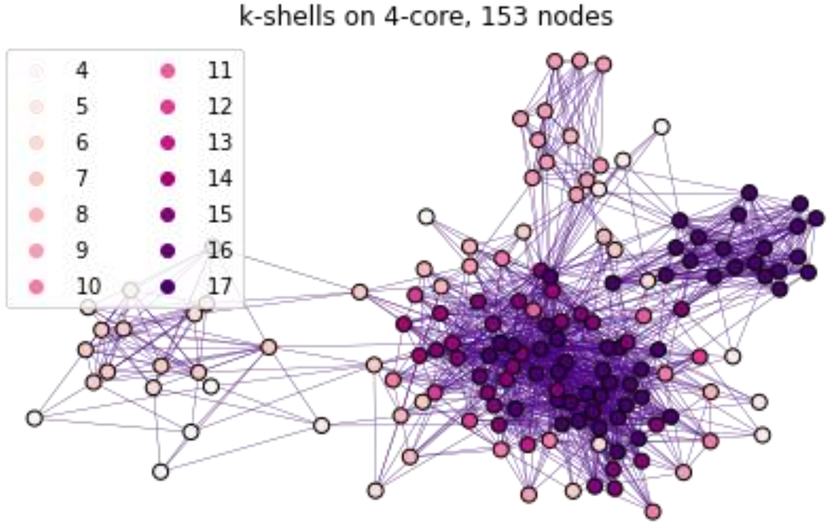


Network visualization by the university

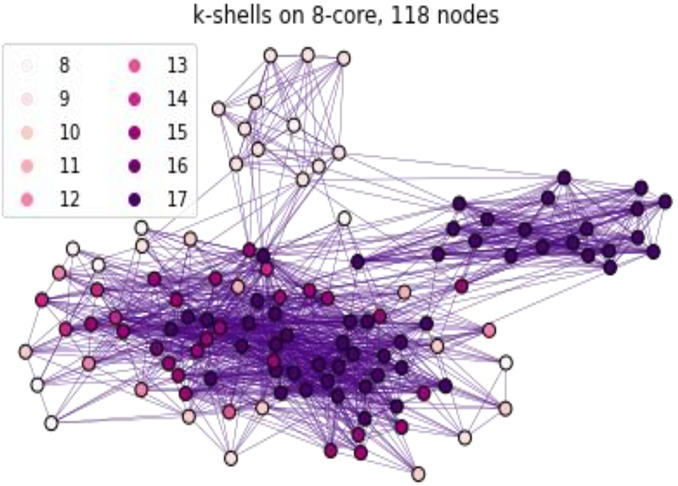
K-core decomposition



K-shells in 2-core

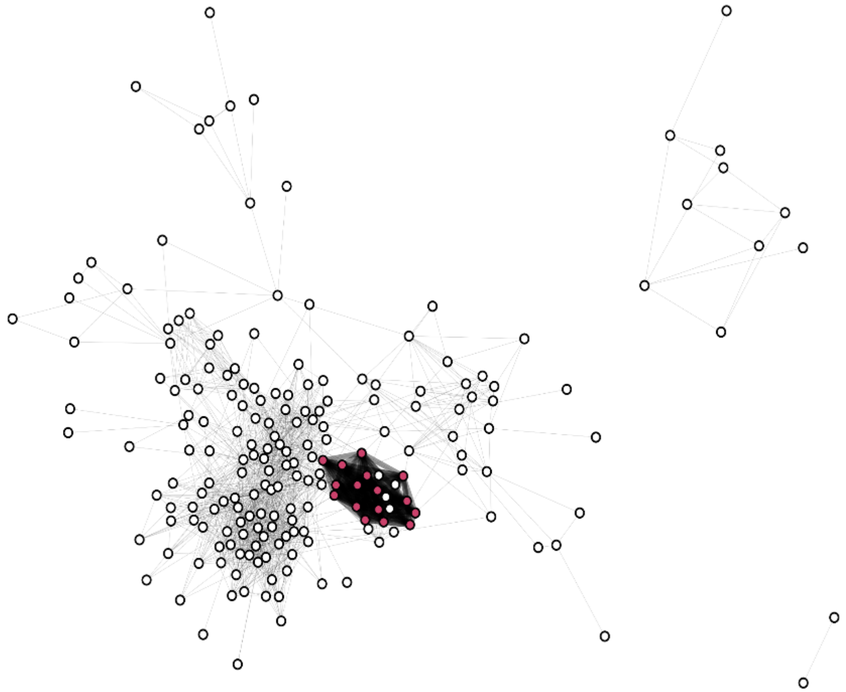


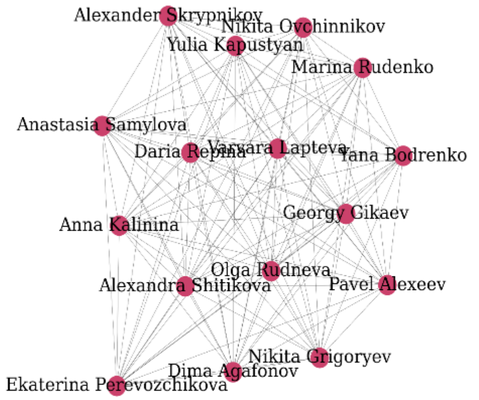
K-shells in 4-core



K-shells in 8-core

Cliques





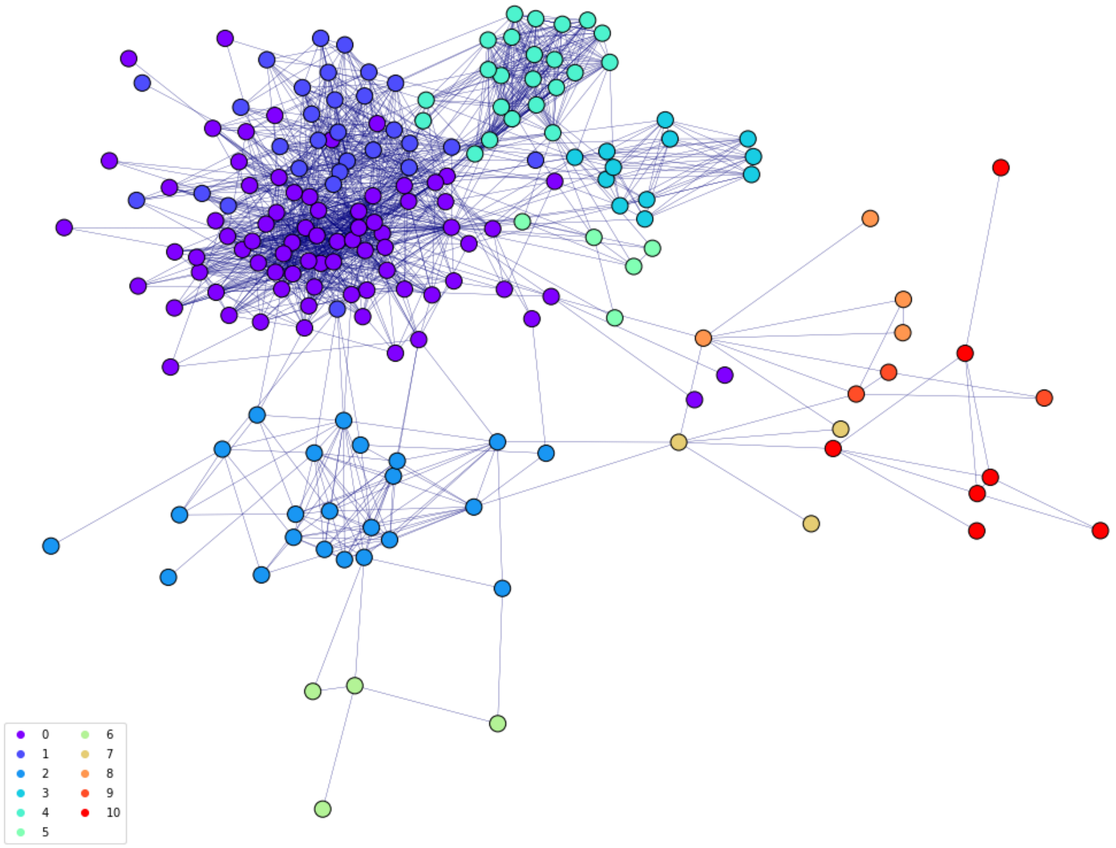
- Number of cliques: 846
- Largest clique is of size: 16
- Average size of clique: 7,132

Max cliques consist of colleagues involved in different projects

Community detection

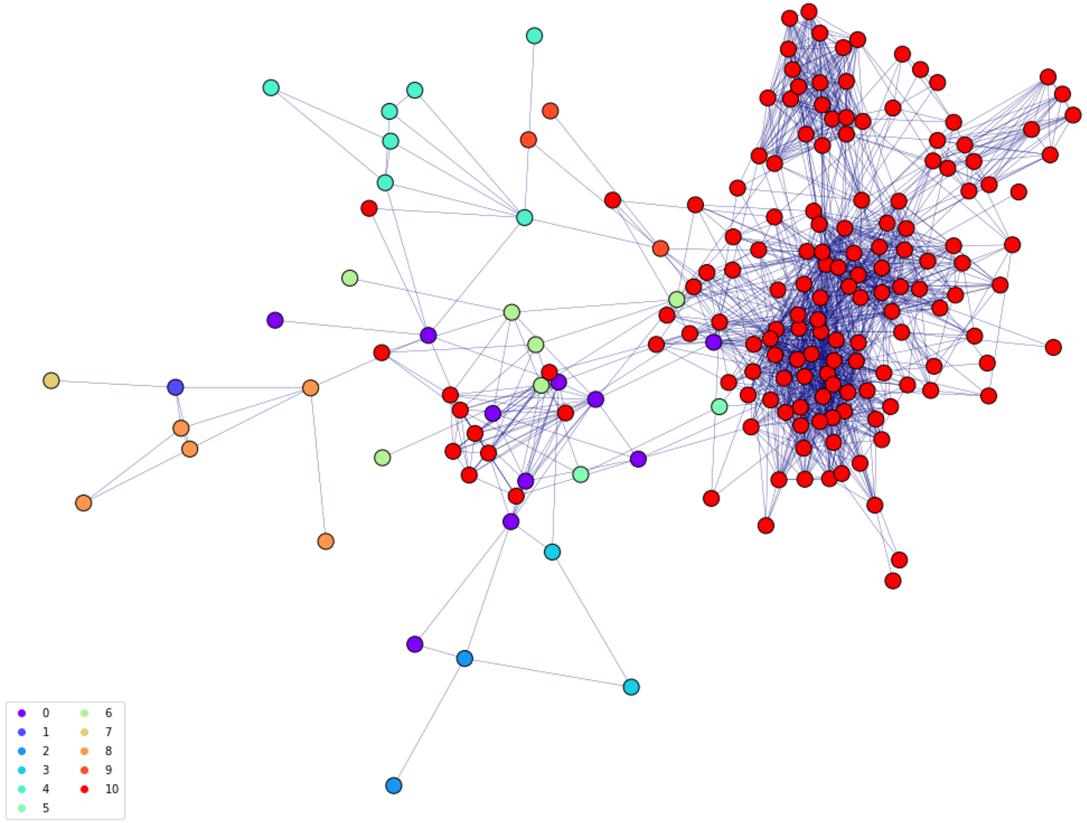
Asynchronous Label Propagation

Modularity = 0.37424174325192544, Coverage = 0.8296488946684005,
Performance = 0.8363130047340573, Silhouette = 0.23922307803817944



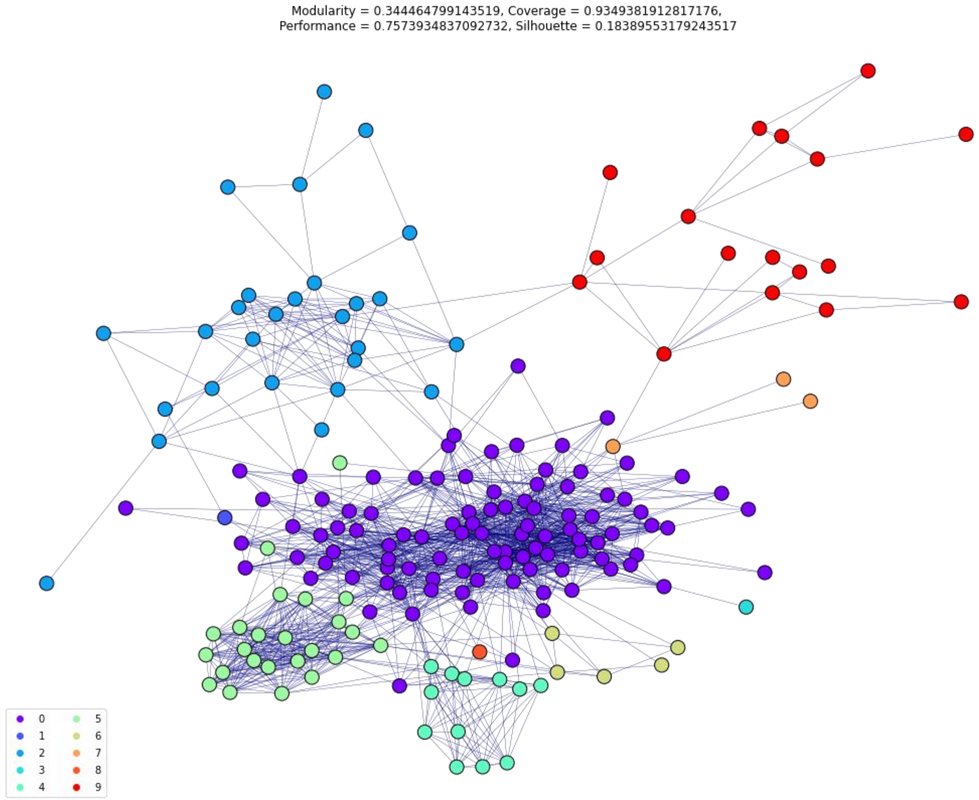
Spectral clustering

Modularity = 0.16295533619734032, Coverage = 0.9414443721535458,
Performance = 0.4394319131161236, Silhouette = 0.11008914852558008

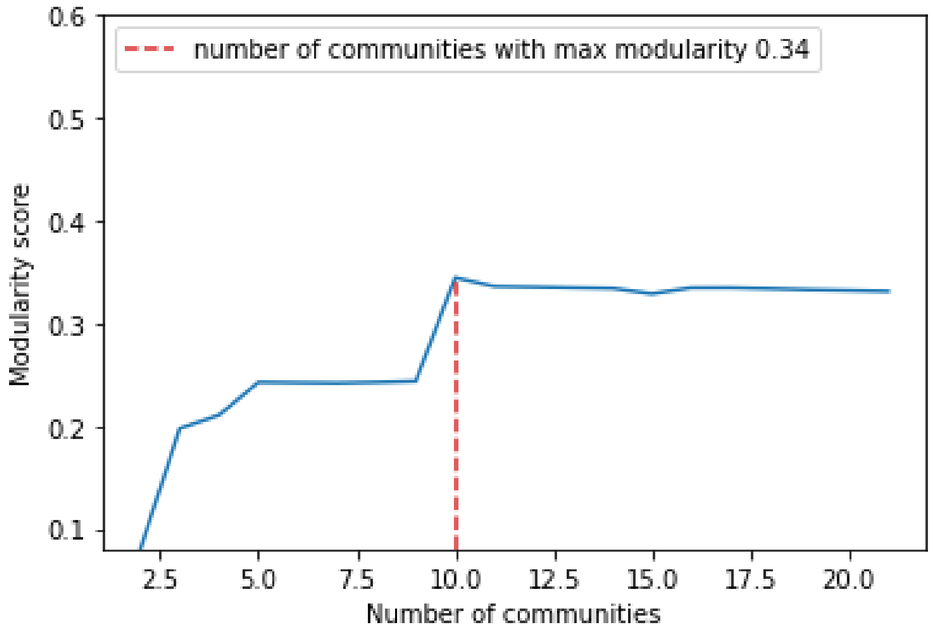


Best result of community detection

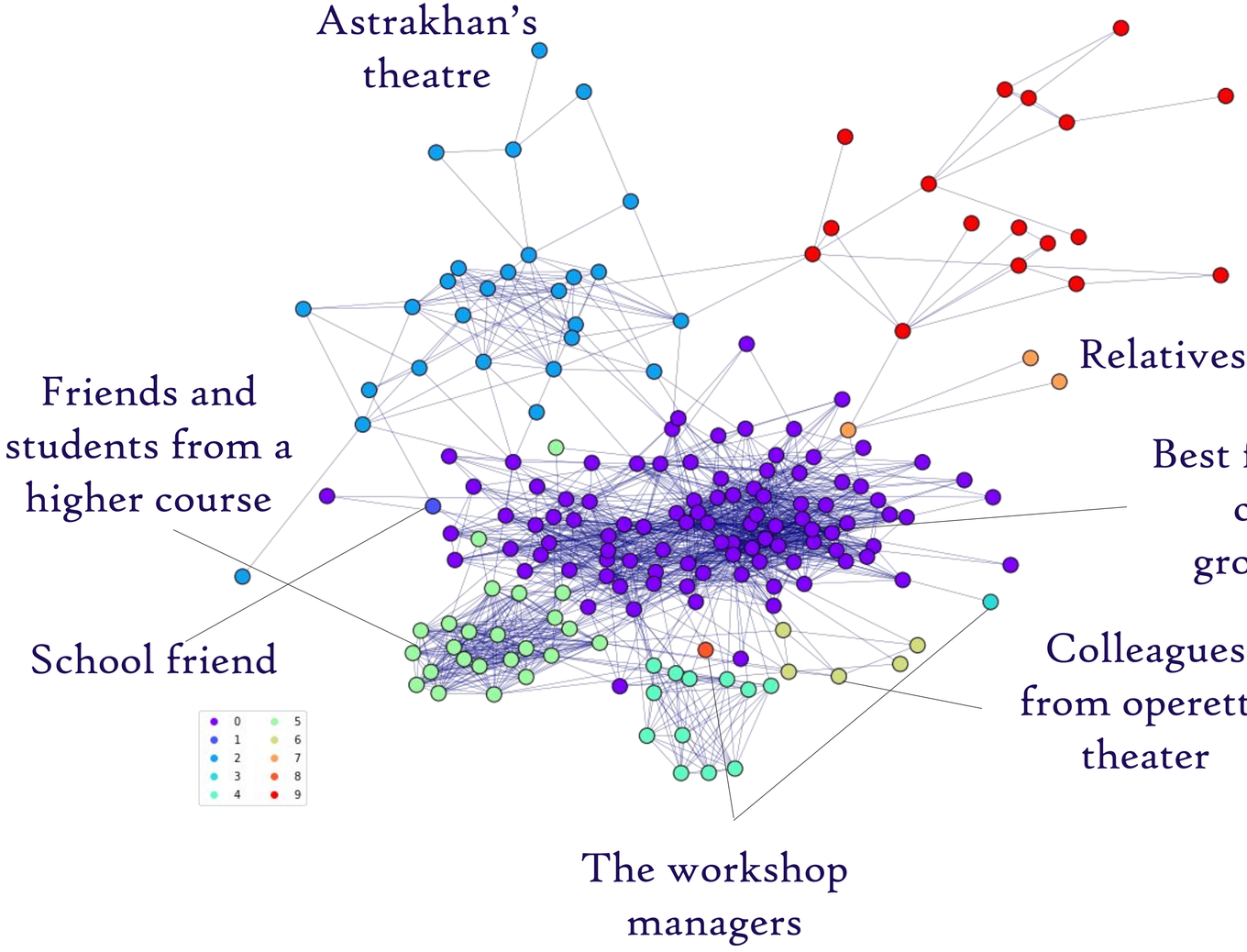
Girvan-Newman algorithm 10 communities



Best score with respect to modularity: 10-11 communities



Interpretation



Girvan-Newman algorithm
10 communities

Results of Community Detection

	Edge-betweenness	Clauset-Newman-Moore	Girvan-Newman	Spectral-clustering	Markov clustering
Modularity	0,336	0,108	0,344	0,163	0,327
Coverage	0,934	0,114	0,935	0,941	0,447
Performance	0,761	0,919	0,757	0,439	0,897
Silhouette	0,202	0,051	0,184	0,11	0,067