Social Network Analysis

Vk project: Friends' graph analysis

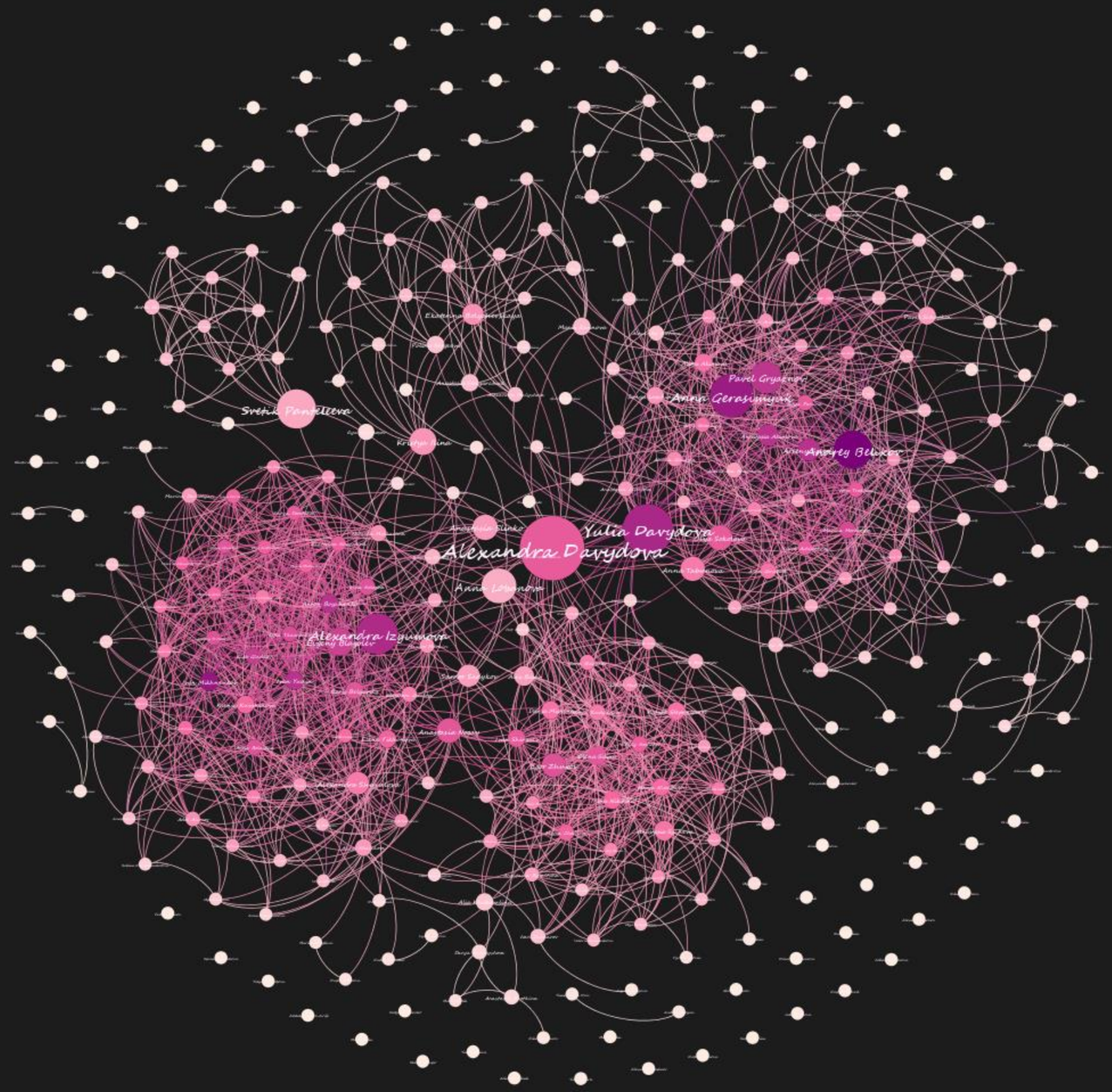
Angelina Parfenova

NETWORK SUMMARY

Nodes: 346
Edges: 1741
Number of connected components: 68

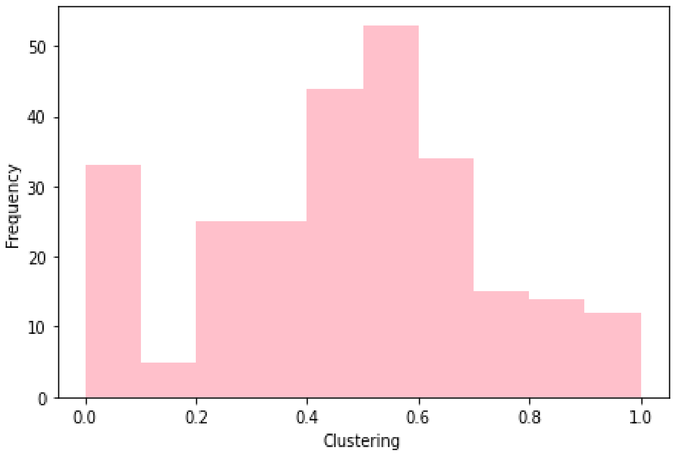
For GCC:

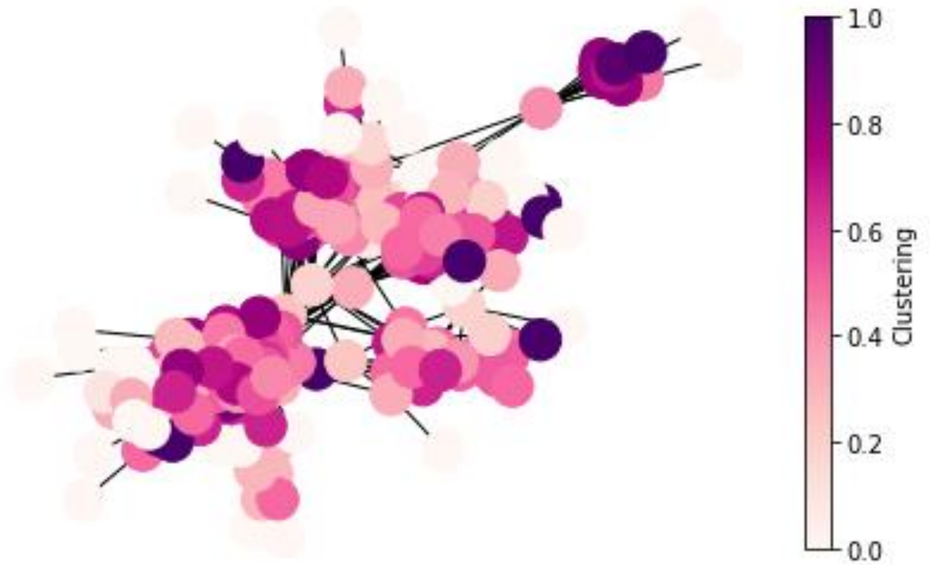
Nodes: 260
Edges: 1715
Radius: 4
Diameter: 7
Average clustering coefficient: 0,47
Average shortest path length: 3,27

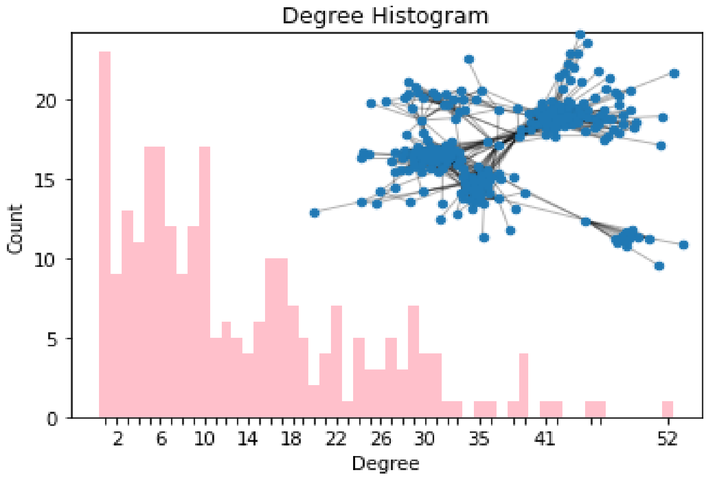


Node size is based on betweenness centrality
Color of nodes is based on their degrees

Clustering coefficients and degree distribution







Average clustering coefficient = 0,47

Fitting model

Model	Av.degree	Av.clustering	Av.path	Radius	Diameter
Real	13.19	0.46	3.27	4	7
Watts-Strogatz	12.0	0.58	3.55	5	7
Barabasi-Albert	13.62	0.11	2.35	3	4
Erdos-Renyi	12.87	0.05	2.44	3	4
Havel-Hakimi	13.19	0.59	4.44	8	16

We can see that the closest model is Watts-Strogatz one

***WS** is based on:*

n (int) – The number of nodes

k (int) – Each node is joined with its k nearest neighbors in a ring topology.

p (float) – The probability of rewiring each edge

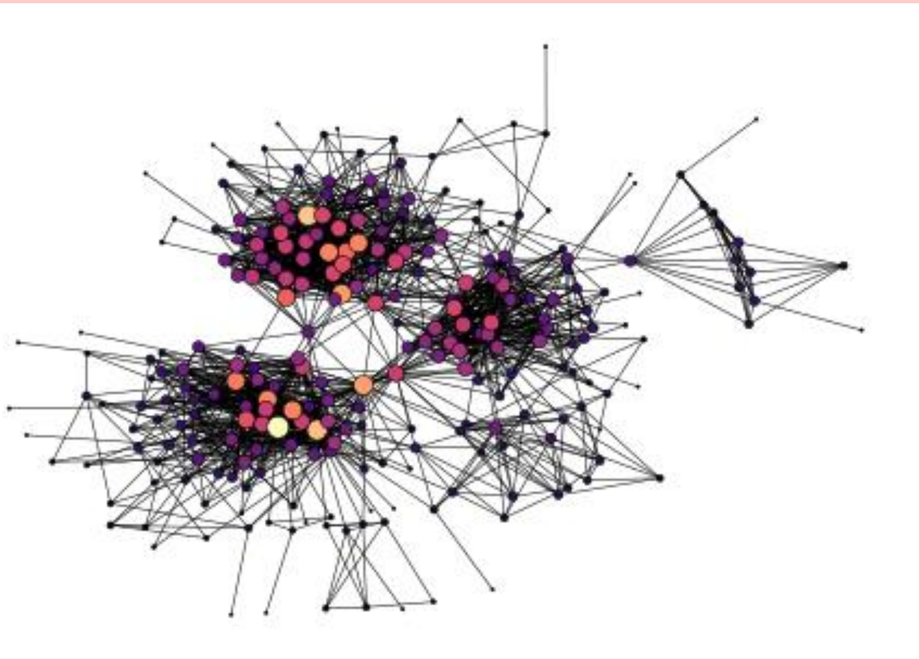
I also tried a Havel-Hakimi random graph model that is based on a degree sequence

DEGREE CENTRALITY

Top 5 nodes:

Andrey Belikov
Ivan Mikhnenkov
Anna Gerasimyuk
Yulia Davydova
Alexandra Izyumova

The degree centrality for a node v is the fraction of nodes it is connected to. So, this top 5 shows people with the highest number of mutual friends.



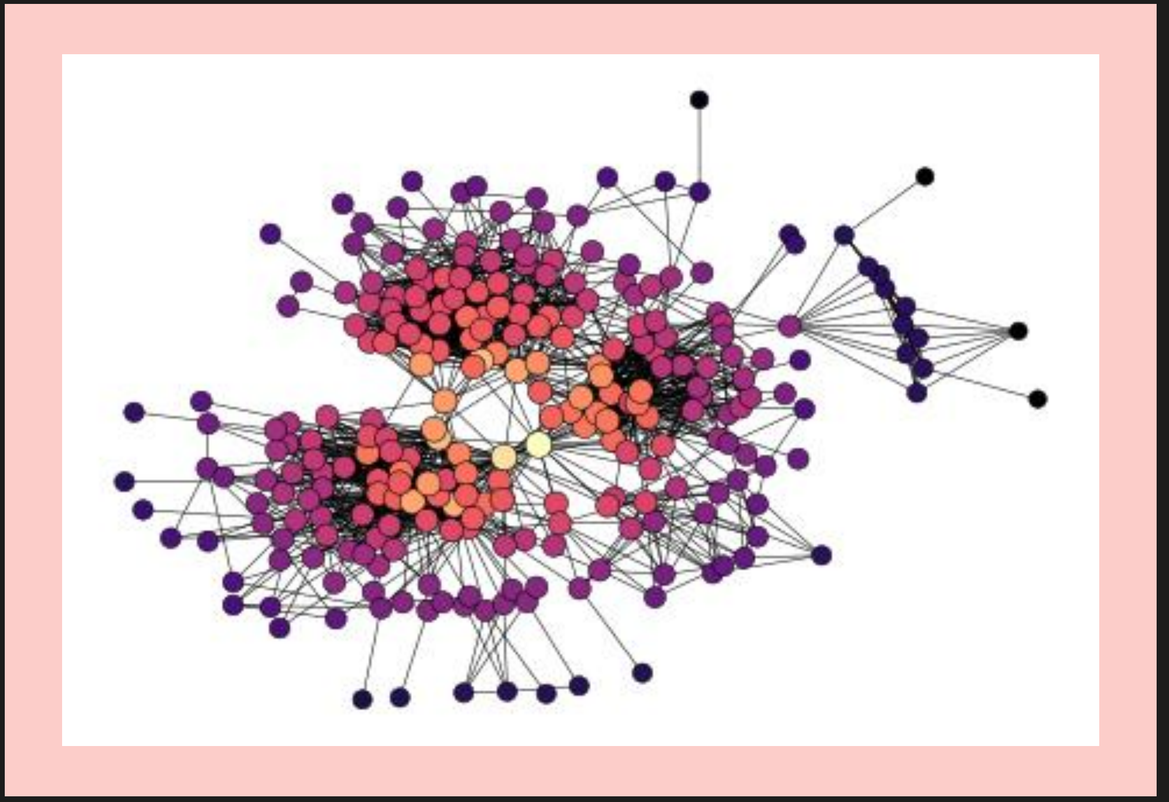
Node size, color(magma) are based on degree centrality

CLOSENESS CENTRALITY

Top 5 nodes:

Alexandra Davydova
Yulia Davydova
Alexandra Izyumova
Olya Sokolova
Anna Gerasimyuk

Closeness centrality of a node u is the reciprocal of the sum of the shortest path distances from u to all n-1 other nodes. So it is easier to find my other friends through these top friends.



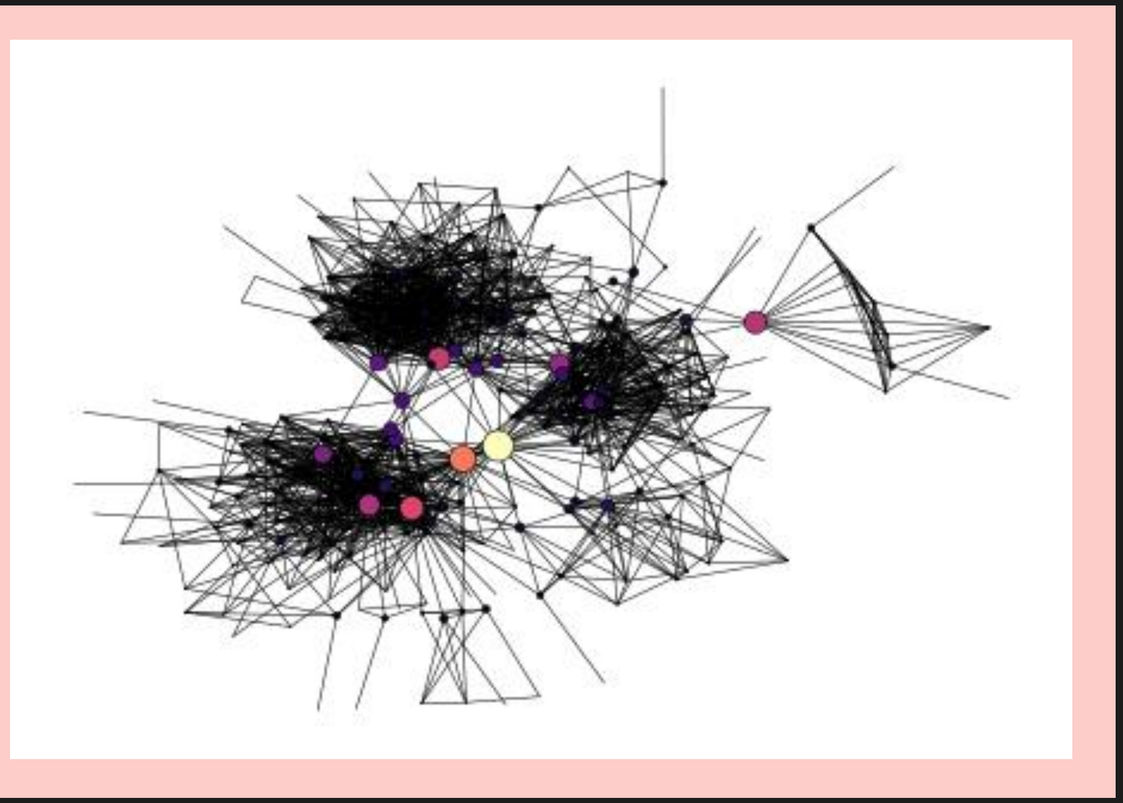
Node size, color(magma) are based on closeness centrality

BETWEENESS CENTRALITY

Top 5 nodes:

Alexandra Davydova
Yulia Davydova
Anna Gerasimyuk
Alexandra Izyumova
Svetik Panteleeva

Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that pass through v. These friends are a bridge between my communities and can lead me towards friends in these communities.



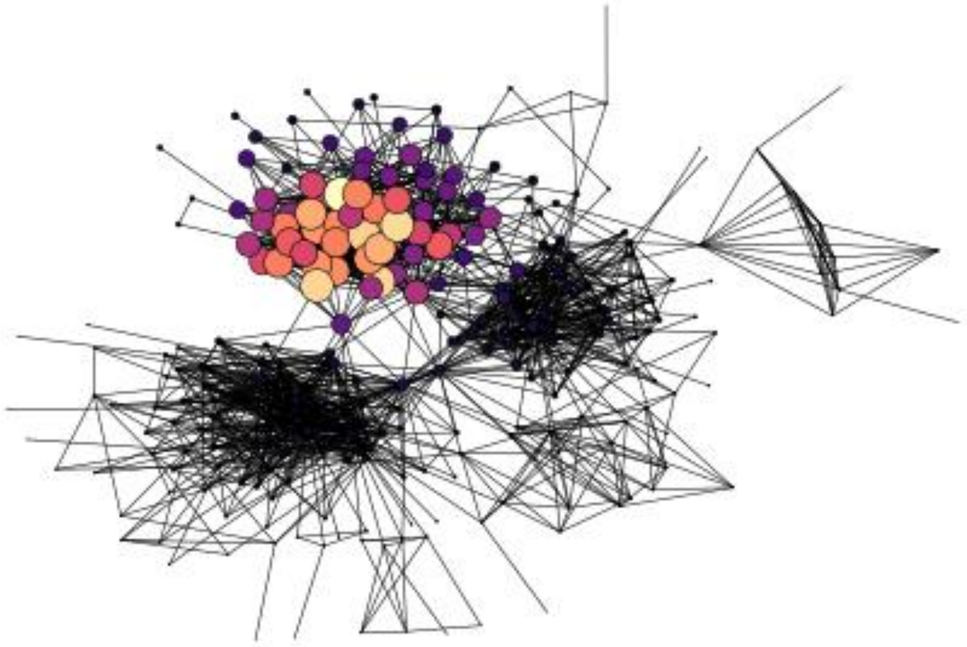
Node size, color(magma) are based on betweenness centrality

EIGENVECTOR CENTRALITY

Top 5 nodes:

Ivan Mikhnenkov
Anton Boychenko
Evgeny Blagolev
Alexandra Izyumova
Gleb Gladkov

Eigenvector centrality computes the centrality for a node based on the centrality of its neighbors. Interestingly, here all my top friends are from bachelor program in HSE.

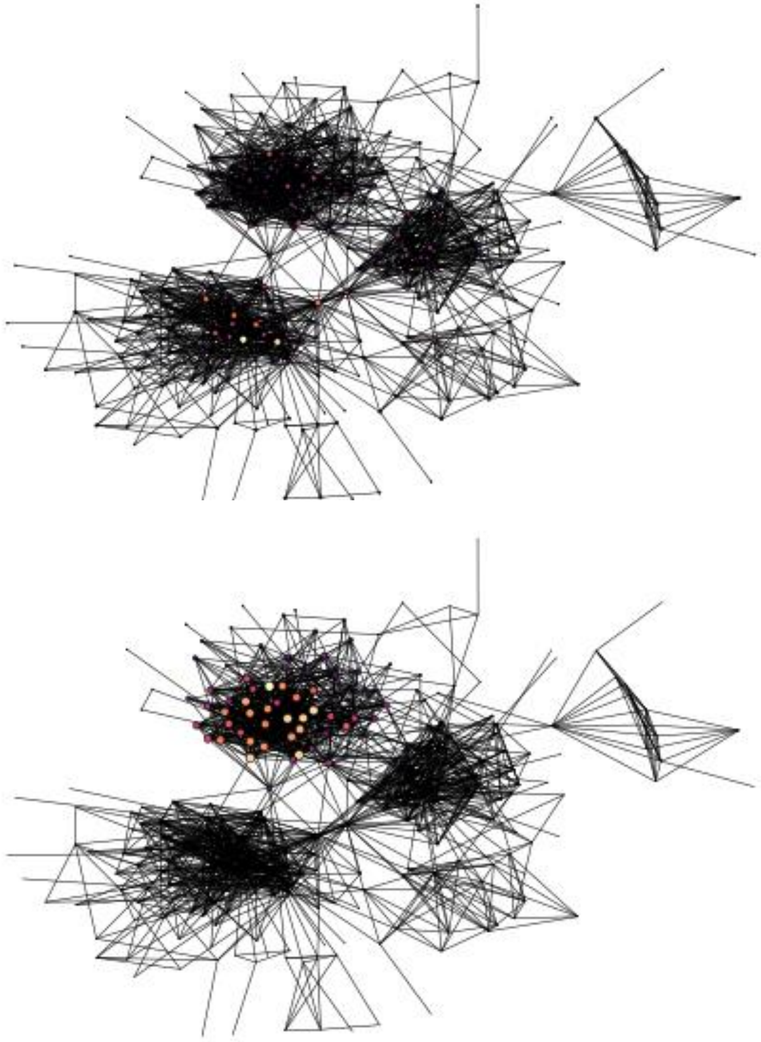


Node size, color(magma) are based on eigenvector centrality

PageRank, HITS

PageRank - computes a ranking of the nodes in the graph G based on the structure of the incoming links. Probability = 0.9.
We can observe that this metrics is very low for all nodes. (max = 0.016)

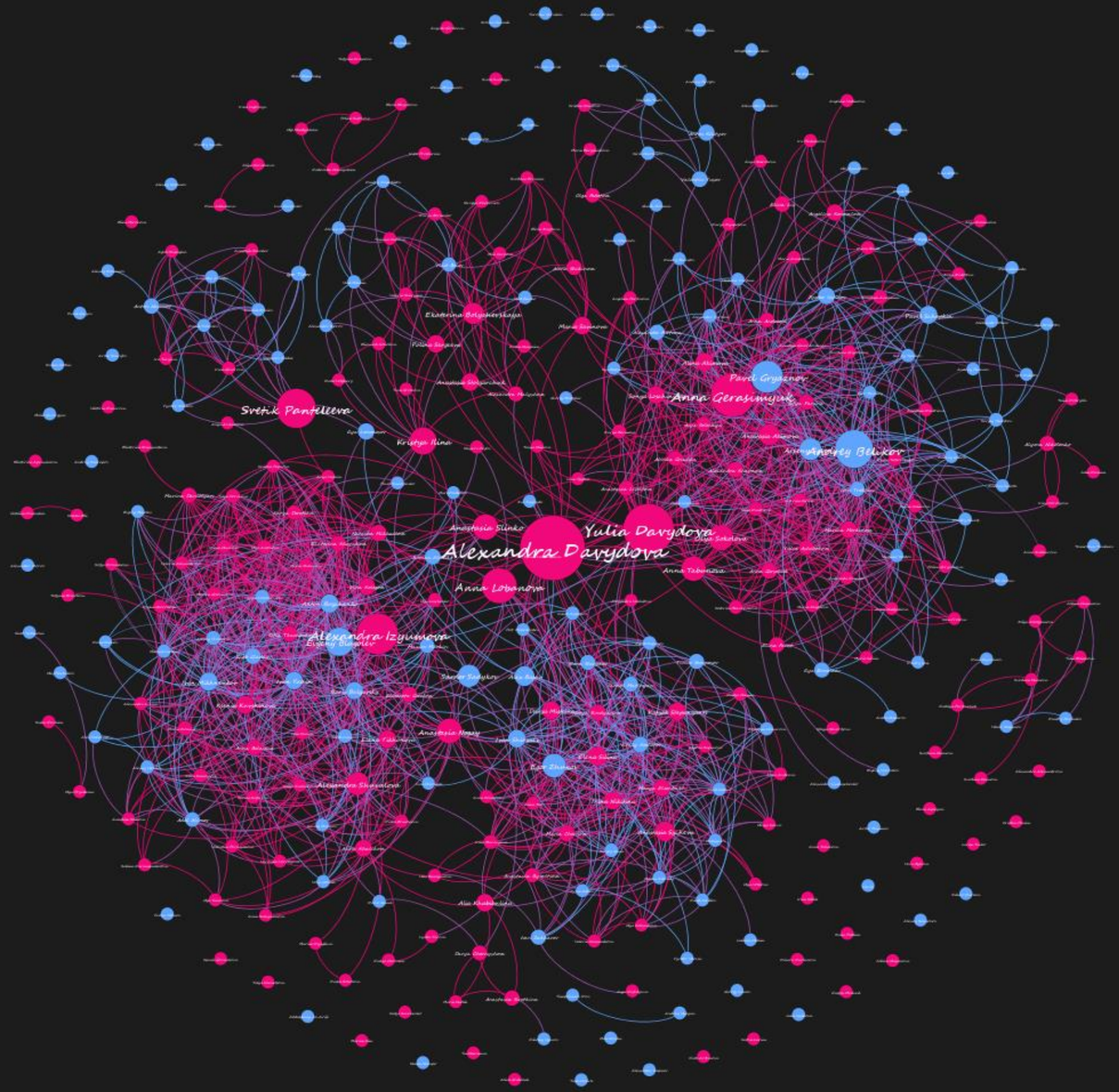
The HITS algorithm computes two numbers for a node. Authorities estimates the node value based on the incoming links. Hubs estimates the node value based on outgoing links. For undirected graph this statistics is the same in both part of tuple.



NETWORK VIZUALISATION BY SEX

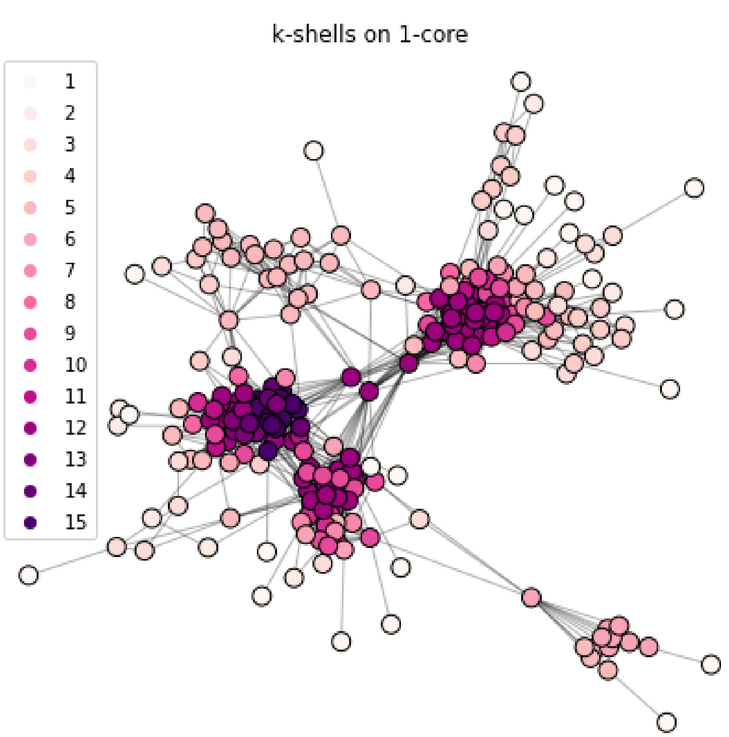
Females: 57,23%

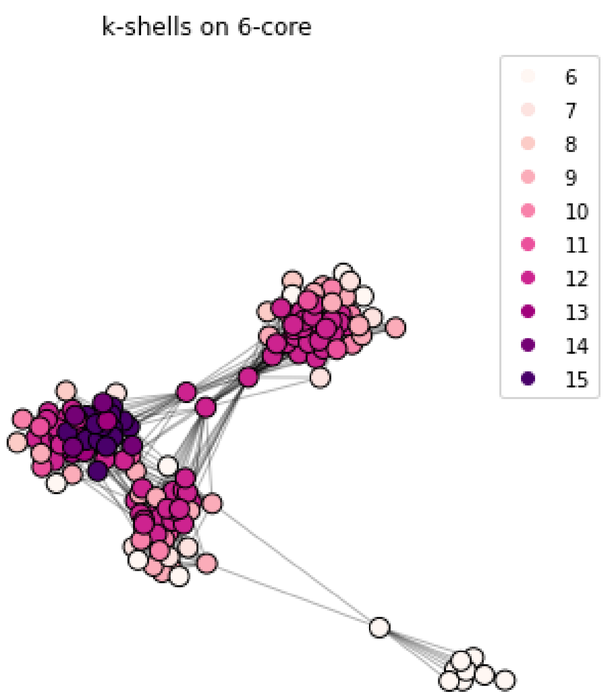
Males: 42,77%

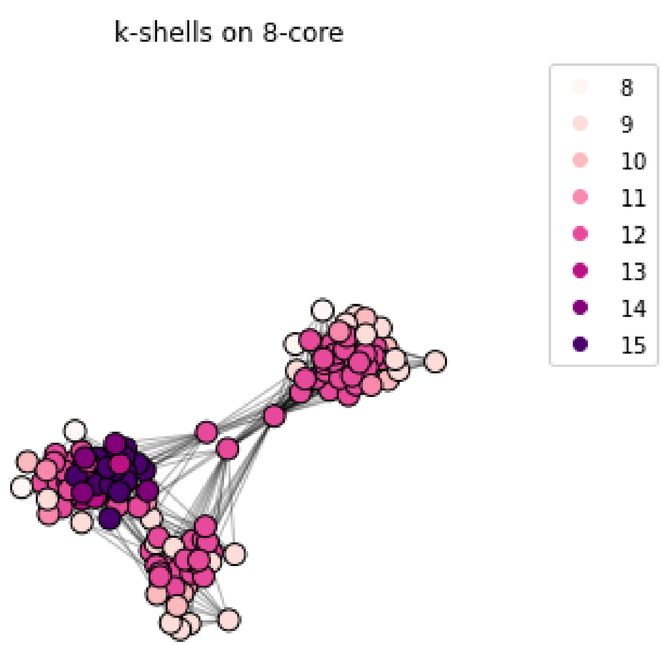


Node size is based on betweenness centrality
Color of nodes is based on the sex

K-core decomposition

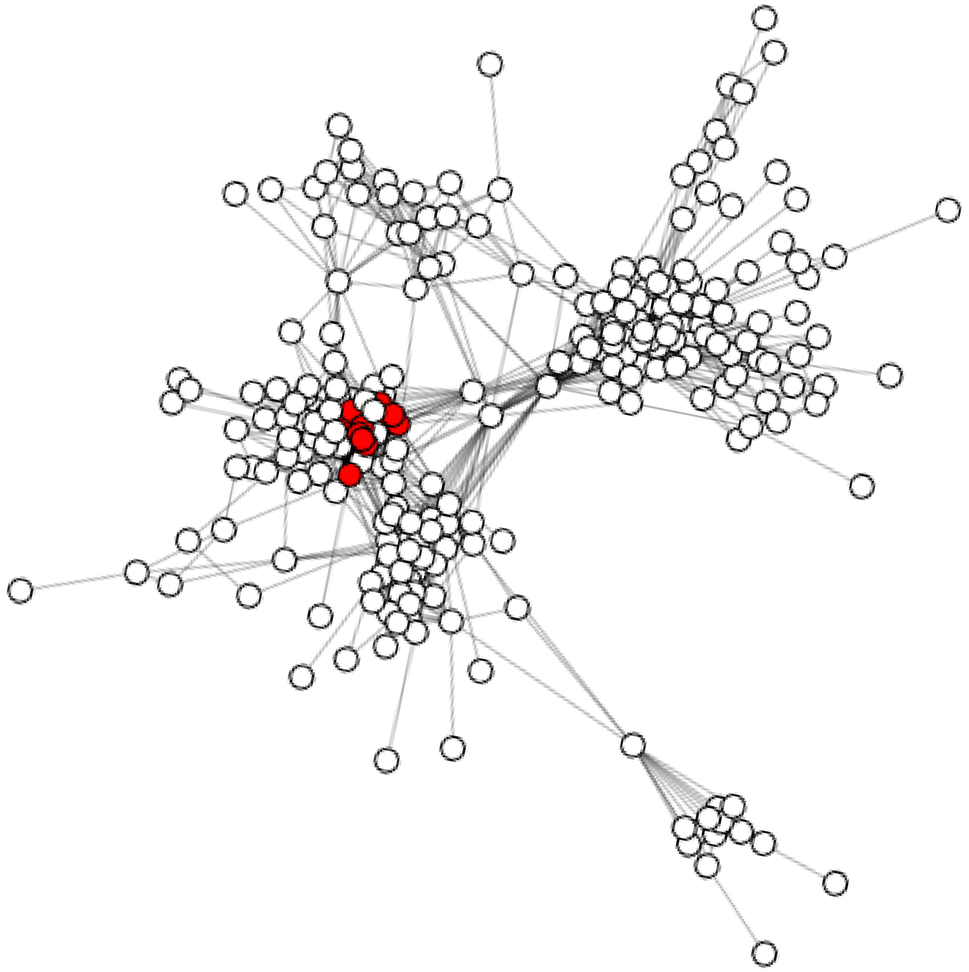






k-core decomposition is a well-established metric which partitions a graph into layers from external to more central vertice

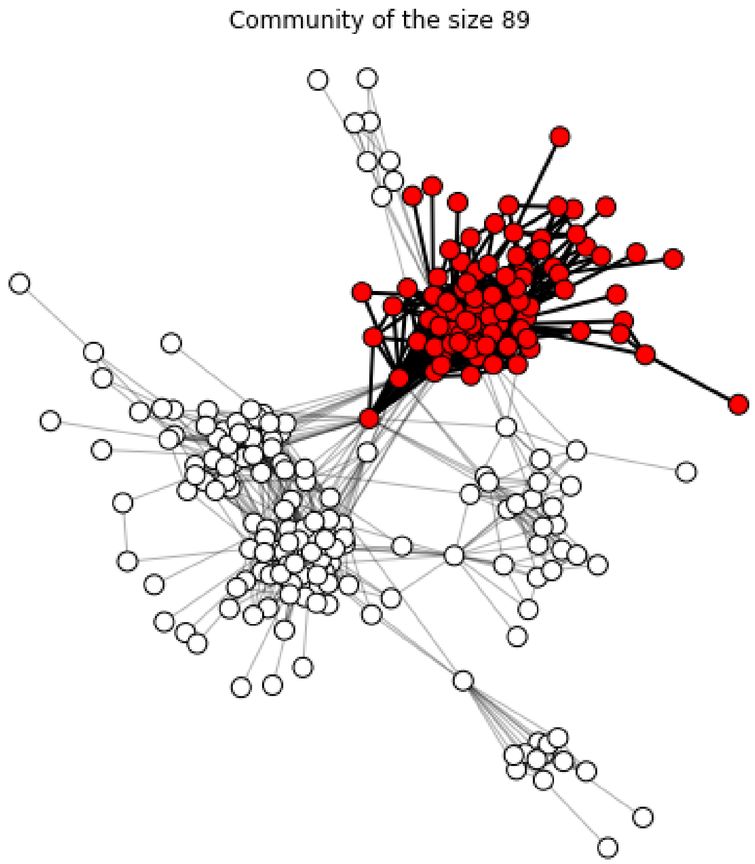
Largest clique (16 degrees)



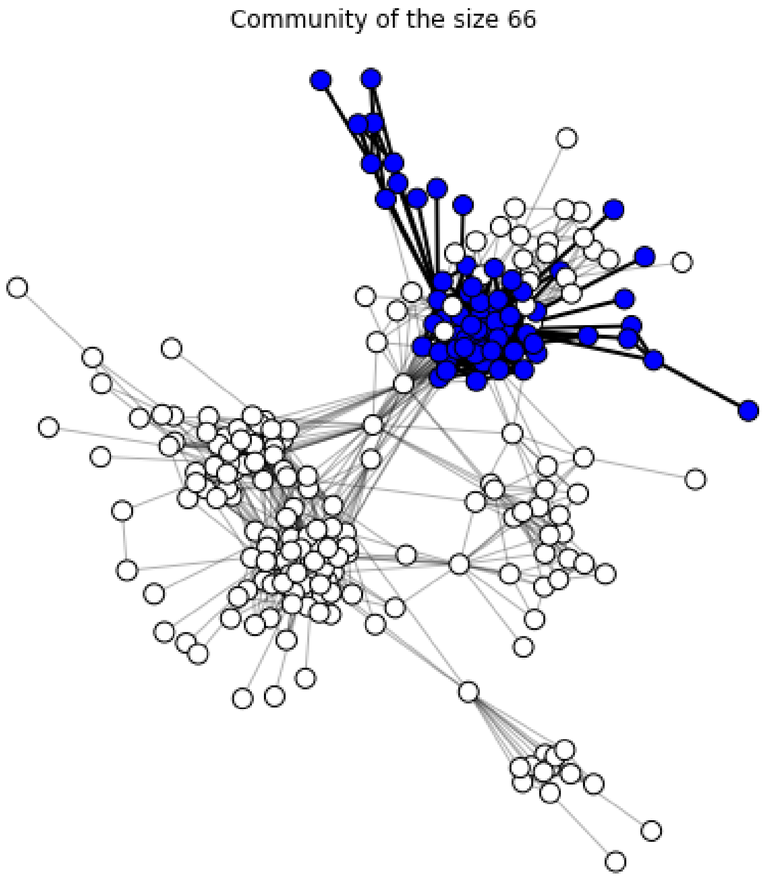
This clique consists of HSE university groupmates from bachelor.

Community detection (largest communities are depicted)

Clauset-Newman-Moore greedy modularity maximization, 7 communities

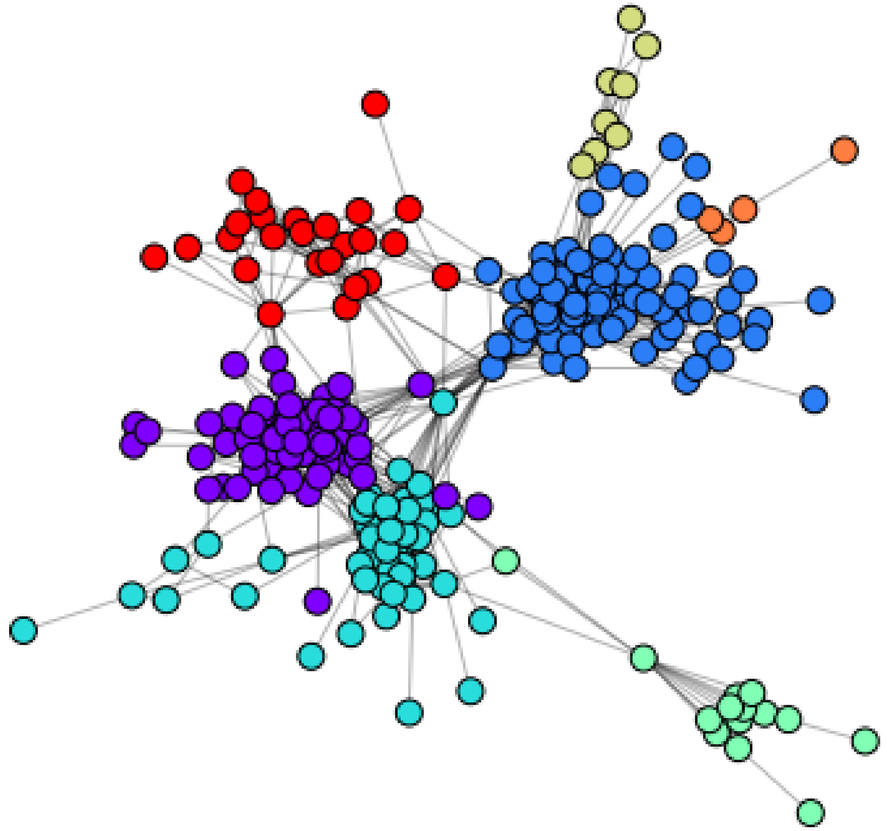


Fluid Communities algorithm., 7 communities (predefined)

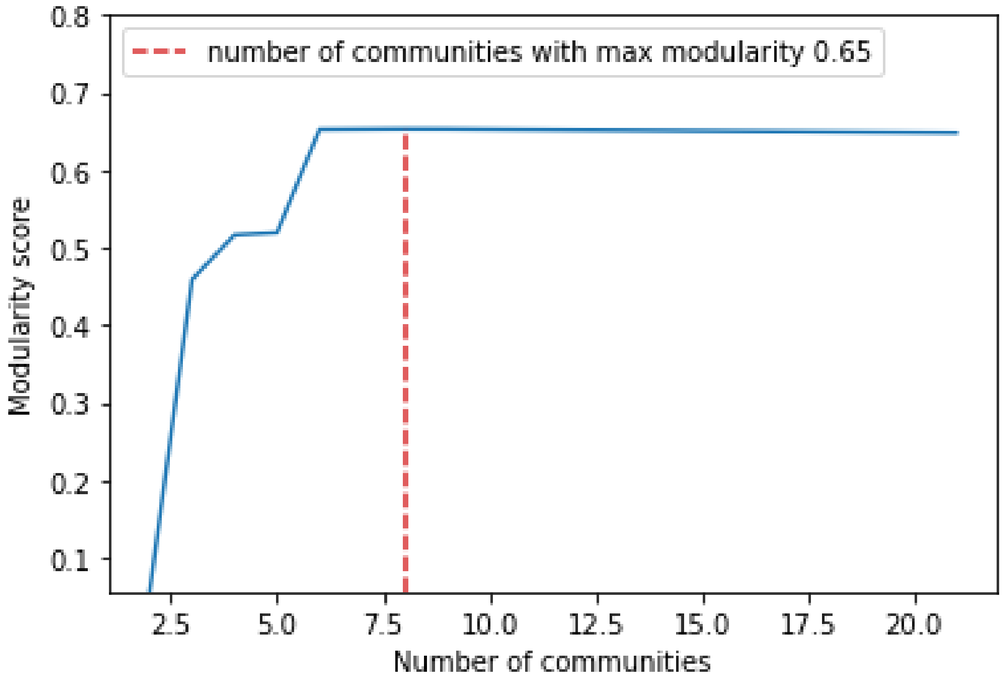


Best result of community detection

Girvan-Newman algorithm
7 communities



Modularity results shows that the best score is achieved with 6-7 communities



Interpretation

Girvan-Newman algorithm
6 communities