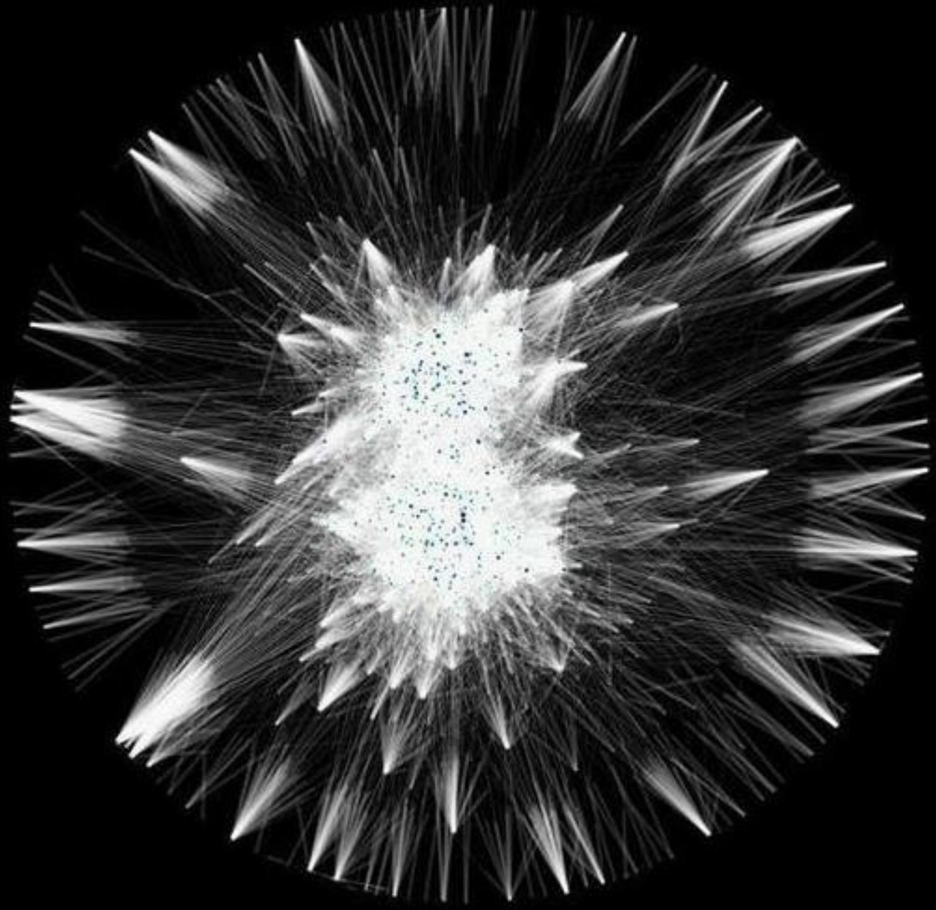
HIGHER SCHOOL OF ECONOMICS
NATIONAL RESEARCH UNIVERSITY

Project 1
Social network analysis.
Friends network in vk.com

Yulia Gurova
мНОД18



# of vertices: 1050 # of edges: 22764 radius: 5 diameter: 9

Outline

- Overview
- Subjects and methods
- Results
- Discussion and future dtudies

Network summary

- **Source**: vk.com, personal page

Data extracted using api.vk.com with access token, functions friends.get, users.get

- **Preprocessing**:
    - the maximal connected component is used for further analysis
    - nodes for closed and deleted pages are deleted
    - the people who are friends to each other are connected
- **Size:**

Network summary

- **Features** for nodes:

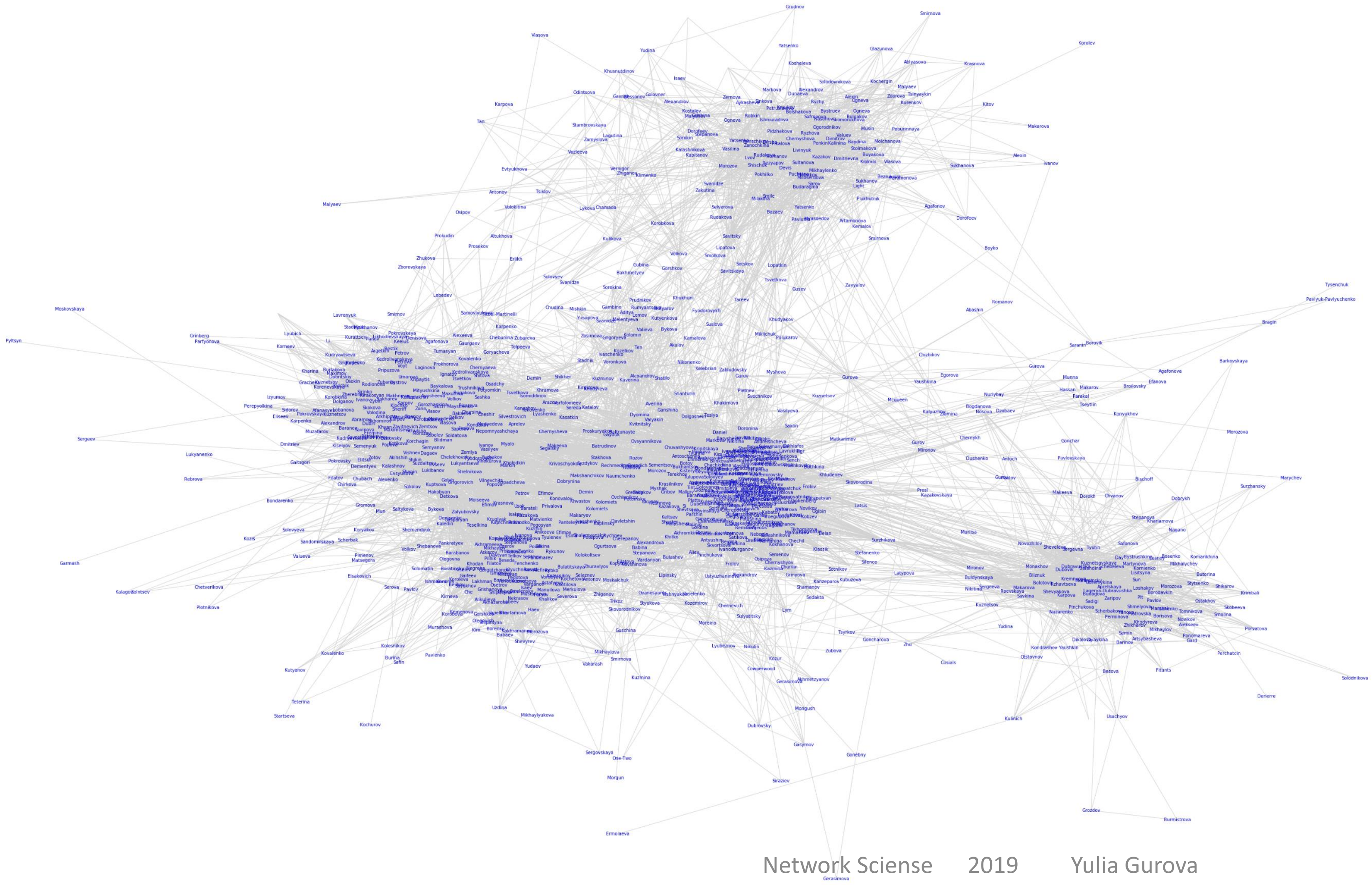
```
{'id': 22656638,
'first_name': 'Yulia’,
'last_name': 'Gurova’,
'is_closed': False,
'can_access_closed': True,
'sex': 1,
'bdate': '11.12’,
'city': {'id':1,'title':'Moscow'}}
```

- **Size** and other parameters:

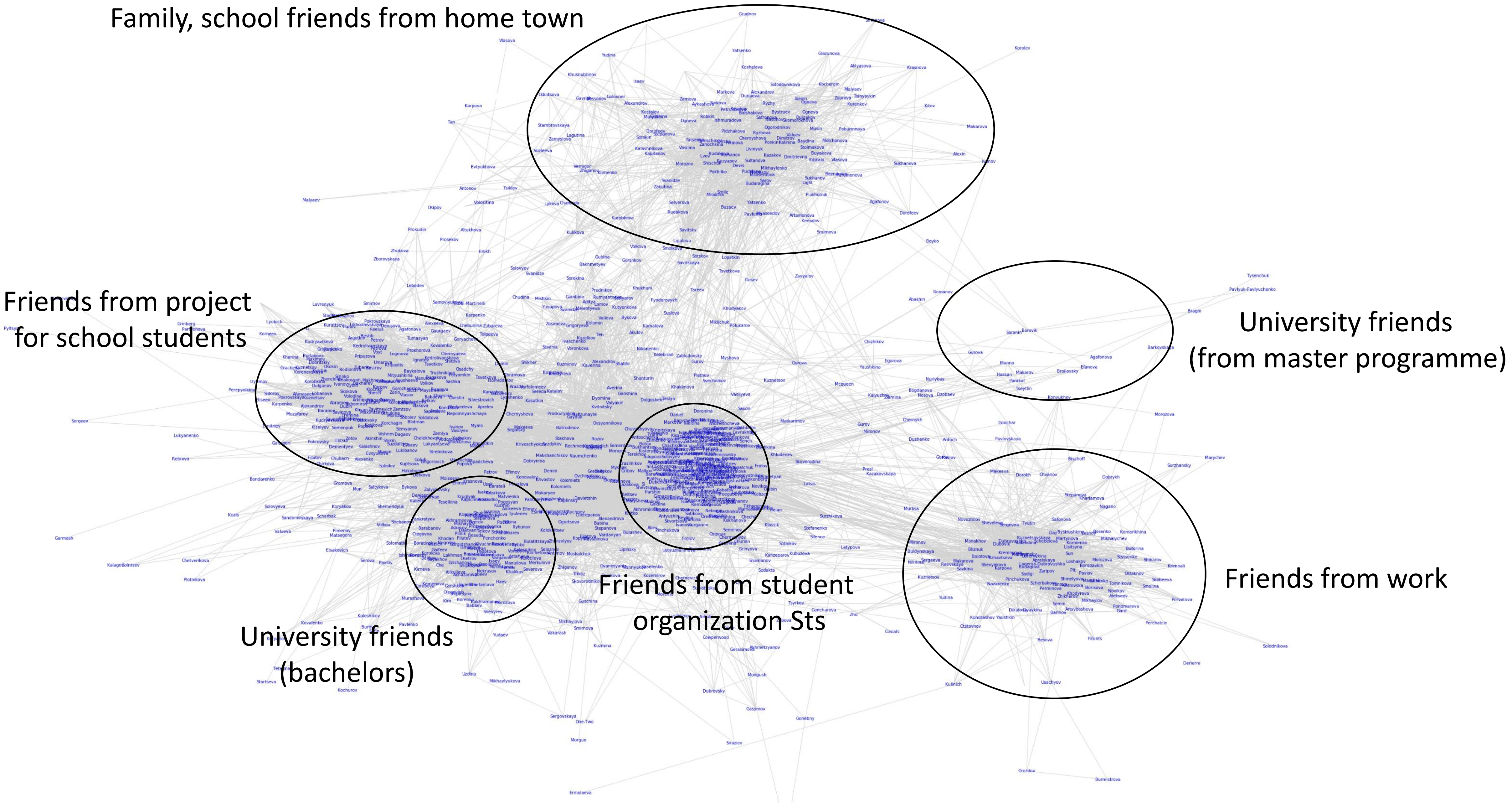
```
# of vertices: 1050
# of edges: 22764
radius: 5
diameter: 9
mean degree: 43.19
average shortest path length: 2.92
clustering coefficient
(transitivity): 0.5091
```

So, the network satisfies some of the small-world properties, having small diameter and average shortest path, and rather high clustering coefficient.

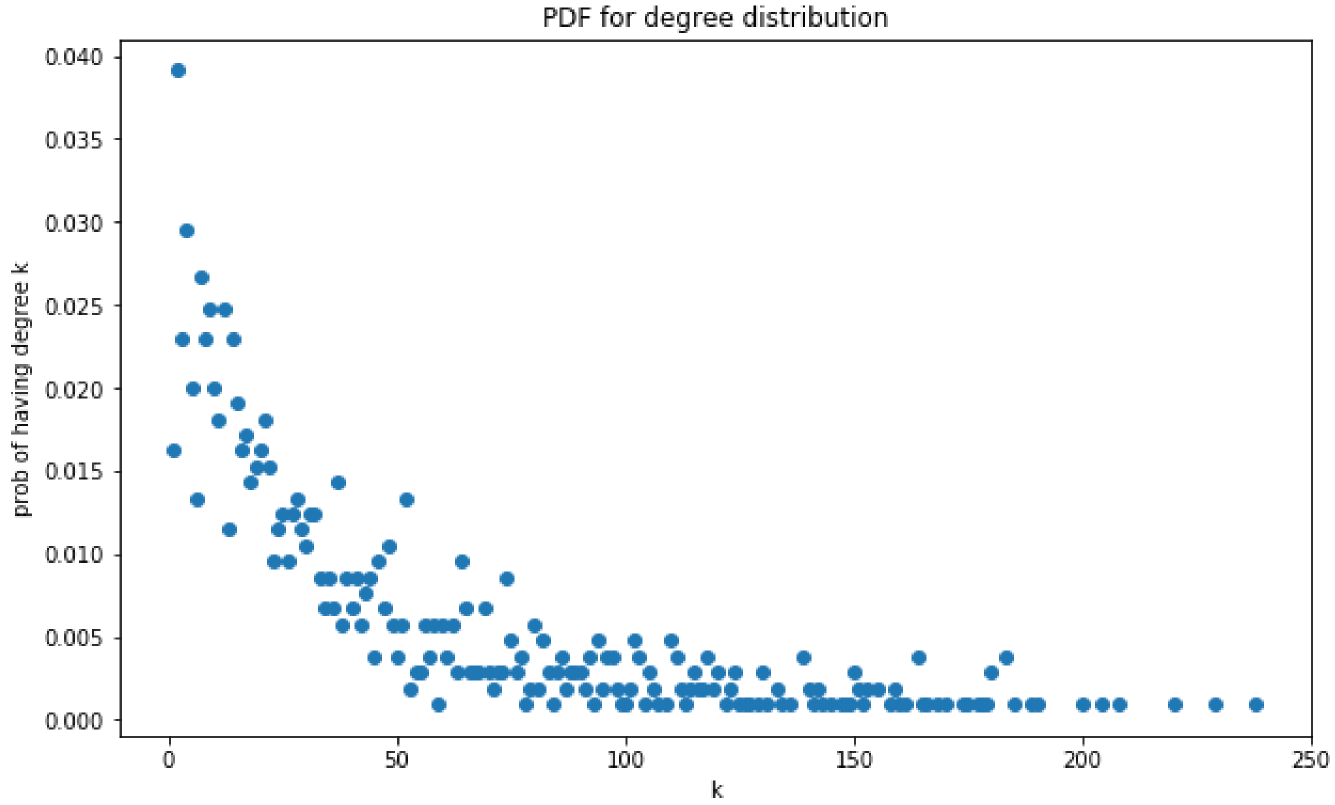
Network Layout

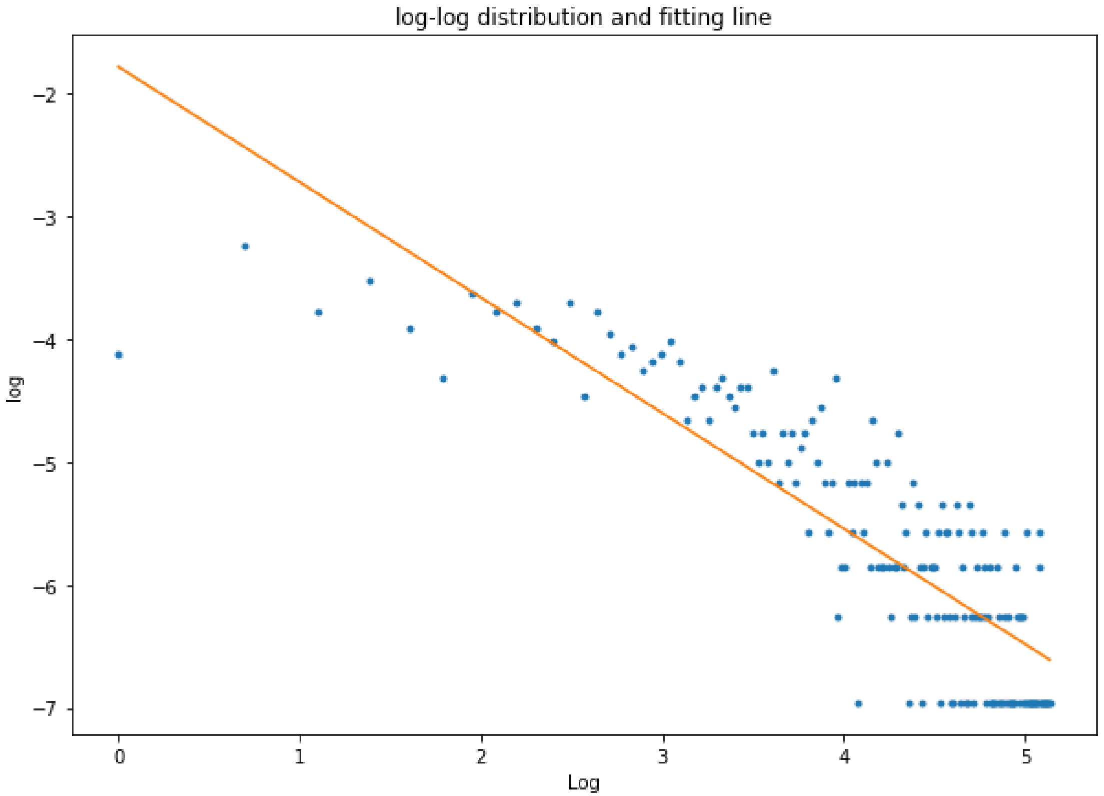


The names are unreadable, but it makes a lot of sence if looking closer (see the communities next slide).
I tried other layouts: circular_layout, shell, spectral and even random layouts, they all look much worser.
I consider this variant is the best for the whole 1000 nodes graph.


Family, school friends from home town
Friends from project for school students
University friends (from master programme)
Friends from work
Friends from student organization Sts
University friends (bachelors)

Structural analysis: Degree Distribution





This is probability density function (dots as it is jumping) for degree distr and the log-log scale to check if it satisfies the power law. Looks close to it. Estimated by linear regression gamma is 0.9407

Structural analysis: Degree Centrality

The nodes with highest degree centrality:

Sts 0.2273
Fyodorovykh 0.2187
Ozoglu 0.2101
Sadykov 0.1987
Tint 0.1948
Glozman 0.1910
Prokofyev 0.1815
Gordeeva 0.1805
Miller 0.1767
Frankenberg 0.1748

These are the people, with whom I have the largest number of common friends.

The highest one – STS -- is the page of a student organisation, in which I was a member for a long time

My BF.

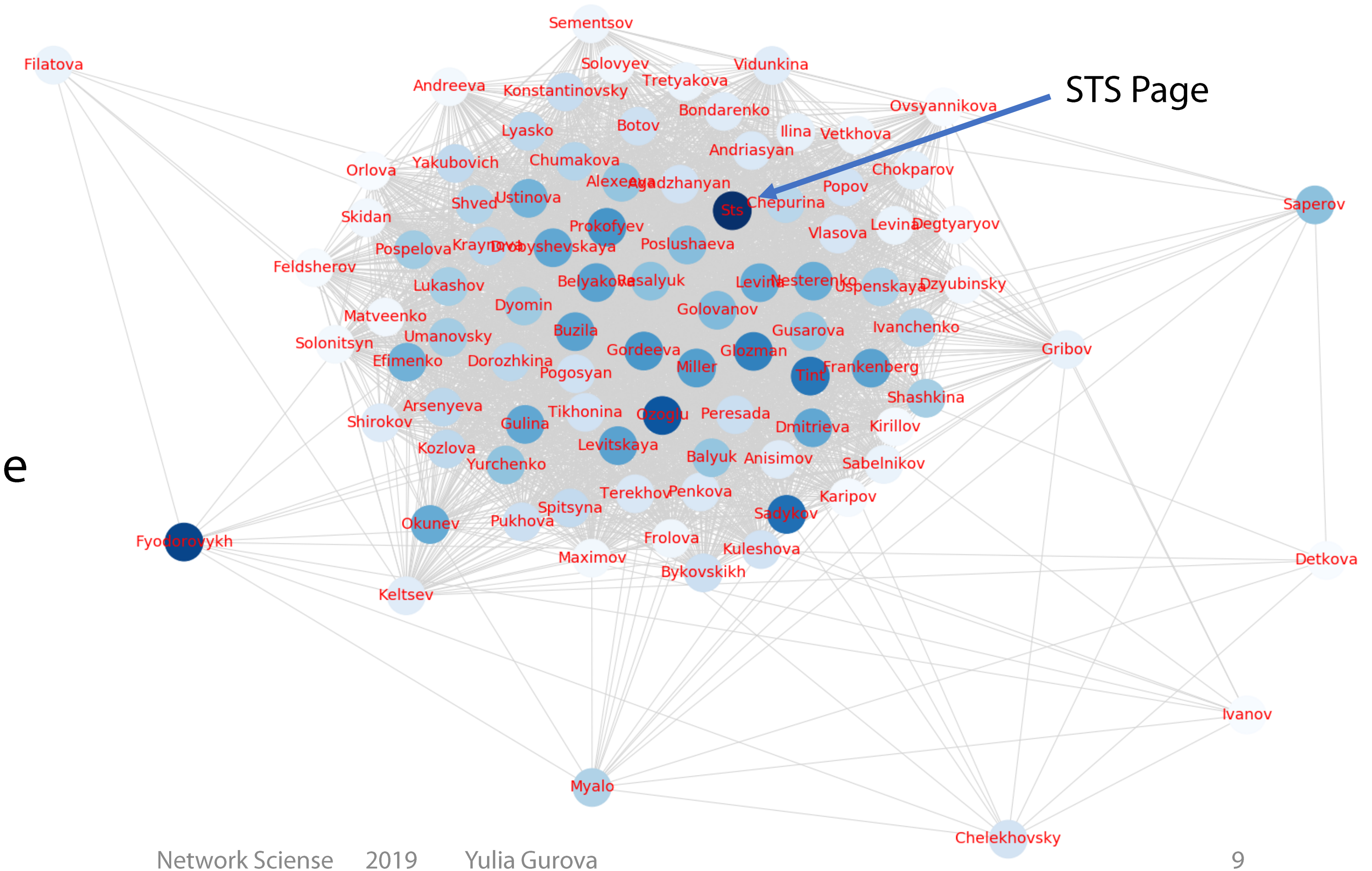
And these people are the members of that organisation and friends to each other. It is a large community, and most of these people are friends to each other as well

Structural analysis: Degree Centrality

Subgraph of nodes with degree centrality > 0.11

The darker is the color – the higher is centrality

The dense subgraph in the middle are STS members, the others are other university friends.



Structural analysis: Closeness Centrality

The nodes with highest closeness centrality

Sts 0.4906
Fyodorovykh 0.4858
Sadykov 0.4723
Okunev 0.4712
Tint 0.4672
Ozoglu 0.4623
Glozman 0.4620
Efimenko 0.4586
Frankenberg 0.4566
Myalo 0.4546
Keltsev 0.4538
Filatova 0.4532
Saperov 0.4436
Chelekhovsky 0.4431
Sementsov 0.4431

← Popular guys from studlife

← University friends (bachelors)

First pages in this list are the pages with the highest amount of friends.

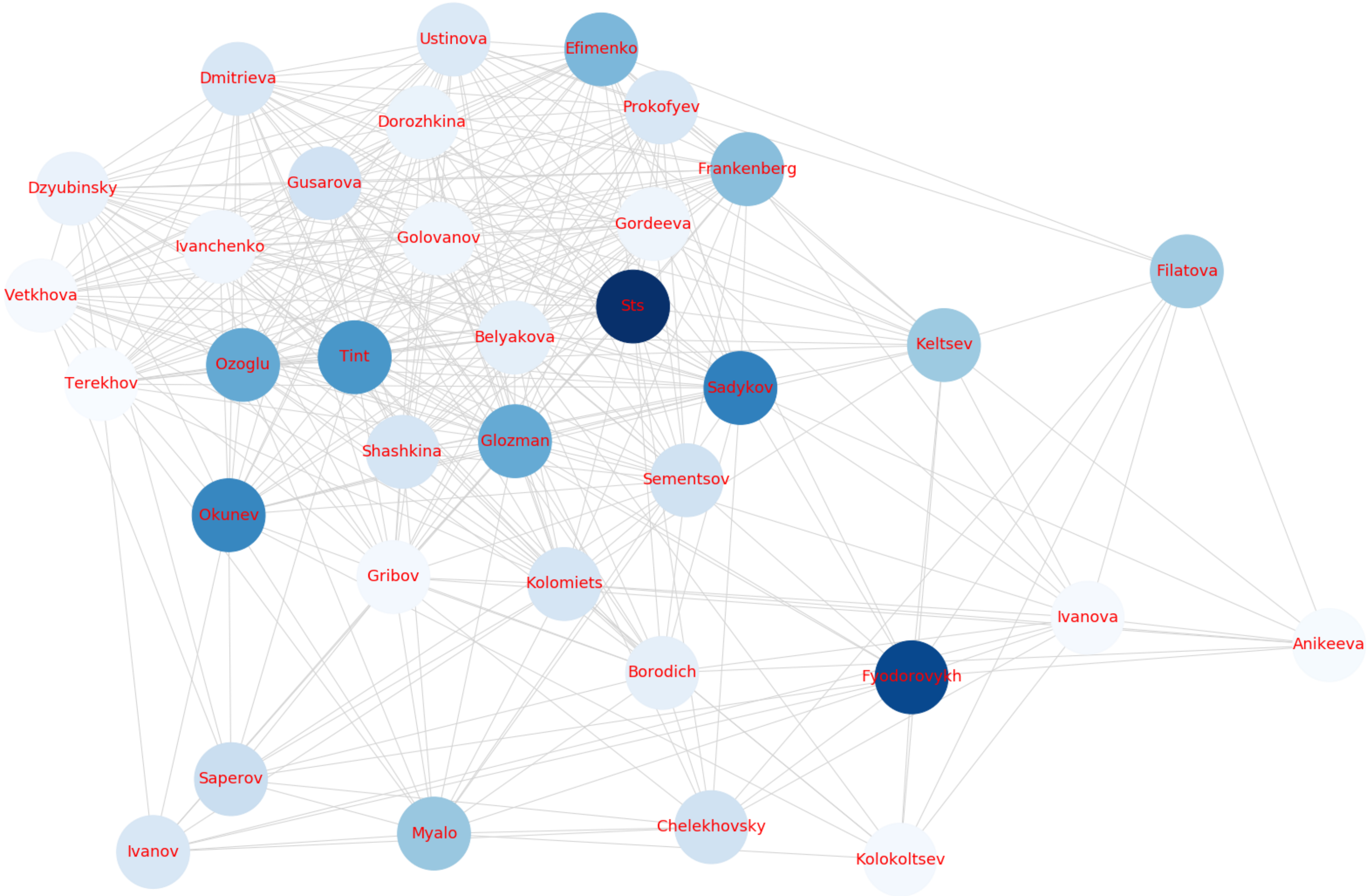
There are couple of very popular at student's life guys, some central people of the friends companies in which I'm included.

And, surprisingly, my best friend Liza Filatova, who has only less then 500 friends, which is smaller then all the other people in the list (they have more than 1000 mostly). That is because she knows at least somebody in all my groups of friends, and so she becomes central here according to closeness.

Structural analysis: Closeness Centrality

Subgraph of nodes with closeness centrality > 0.43

The darker is the color – the higher is centrality



Structural analysis: Betweenness Centrality

The nodes with highest Betweenness centrality

Fyodorovykh 0.0864 ← Boyfriend

Filatova 0.0574 ← Bestfriend

Frankenberg 0.0530

Sts 0.0405

Sadykov 0.0365

$\left.\right\}$ STS

Dubravushka 0.0332 ← Page of place in which I worked long ago (many friends from there are connected to it

Kantemirovsky 0.0328

Okunev 0.0287

Gurova 0.0287 ← Mom O.o

Surzhikova 0.0264

Volkova 0.0250

First there are the closest people who have connections to different groups of my friends (not whole group, but just someone).

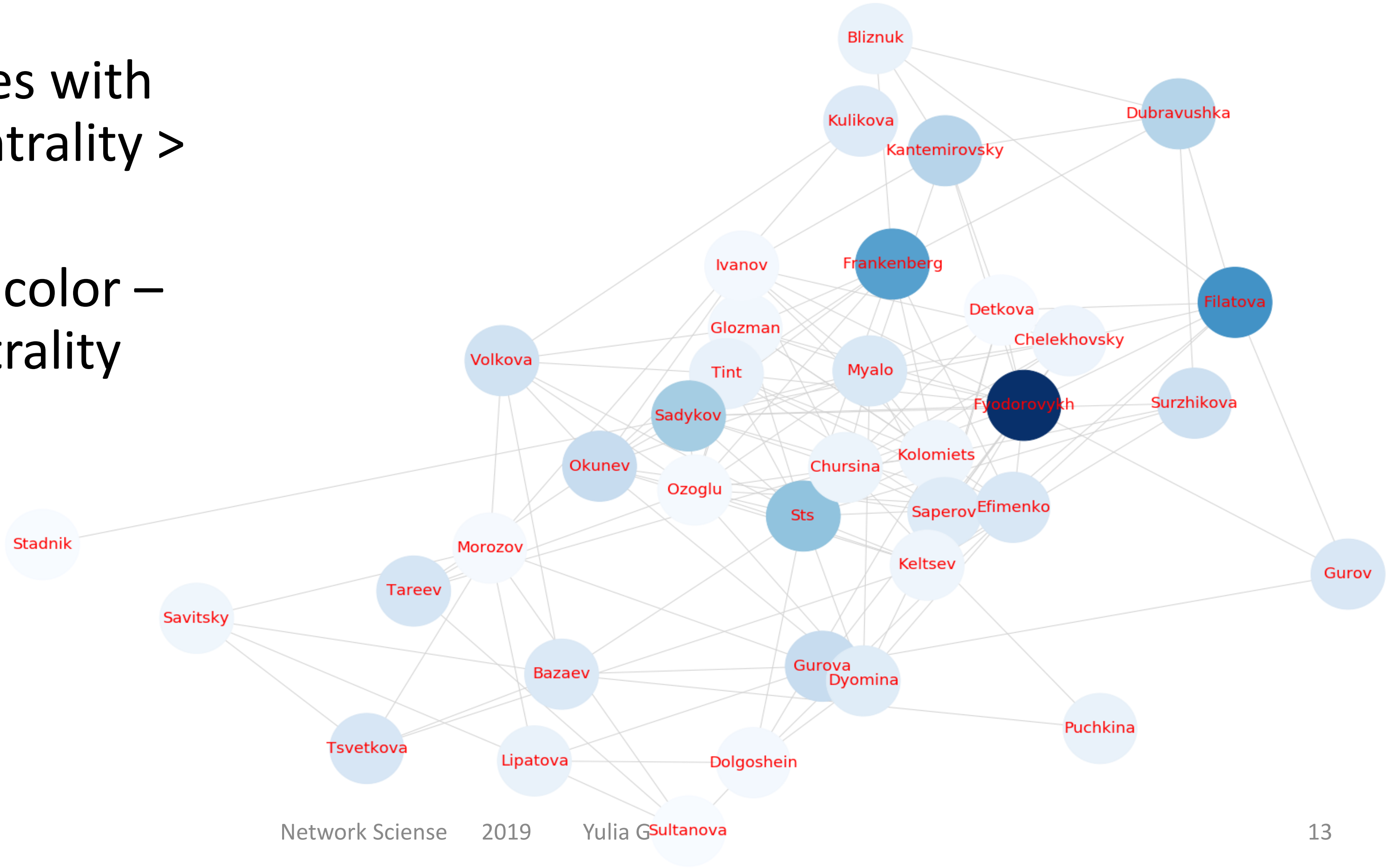
And (wow) my mom is the ningth in the list, despite she has less than 200 friends and only 25 common with me. But of course, these 25 are mostly my best friends from different periods of life, so she is among the most central here.

So, I find the degree centrality on my connections graph the most informative for searching my closest people in the network.

Structural analysis: Betweenness Centrality

Subgraph of nodes with Betweenness centrality > 0.01

The darker is the color – the higher is centrality



Structural analysis: Eigenvector Centrality

The nodes with highest eigenvector centrality

Ozoglu 0.1136
Sts 0.1109
Gordeeva 0.1077
Levitskaya 0.1073
Miller 0.1065
Drobyshevskaya 0.1060
Prokofyev 0.1055
Levina 0.1049
Glozman 0.1047
Belyakova 0.1047
Gulina 0.1044
Buzila 0.1021
Nesterenko 0.1020
Alexeeva 0.1015

STS

All these people are from the largest and highly connected to each other group of my friends, with whom I was a member of the student organisation Ingroup StS.

And the second page on the list is the page of this organisation. So, not surprisingly, the highest people here are from the same group as the connection to the highest centrality person increases your centrality in the case of eigenvector centrality

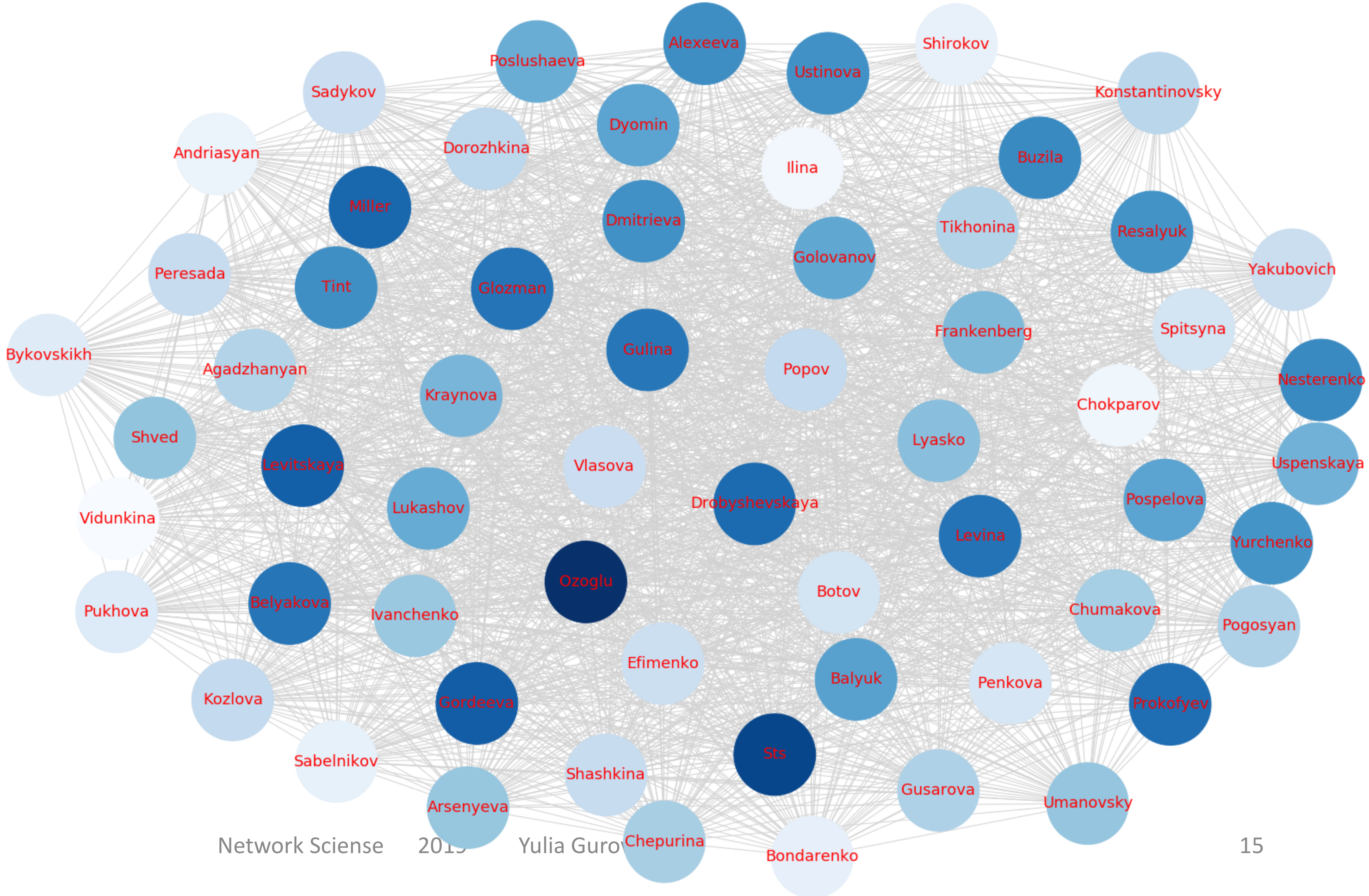
Graph on the next slide demonstrates how dence this subgraph is.

Structural analysis: Eigenvector Centrality

Subgraph of nodes with eigenvector centrality > 0.08

The darker is the color – the higher is centrality.

It is very dense subgraph, almost all these people are connected to each other



Structural analysis: Page-Rank Centrality

The nodes with highest Page-Rank centrality (a=0.7)

Fyodorovykh 0.0050

Dubravushka 0.0036

Filatova 0.0034

Saperov 0.0031

Sadykov 0.0031

Sts 0.0030

Myalo 0.0028

Morozov 0.0028

Chelekhovsky 0.0027

Fominsky 0.0027

Okunev 0.0026

Efimenko 0.0026

Gurova 0.0025

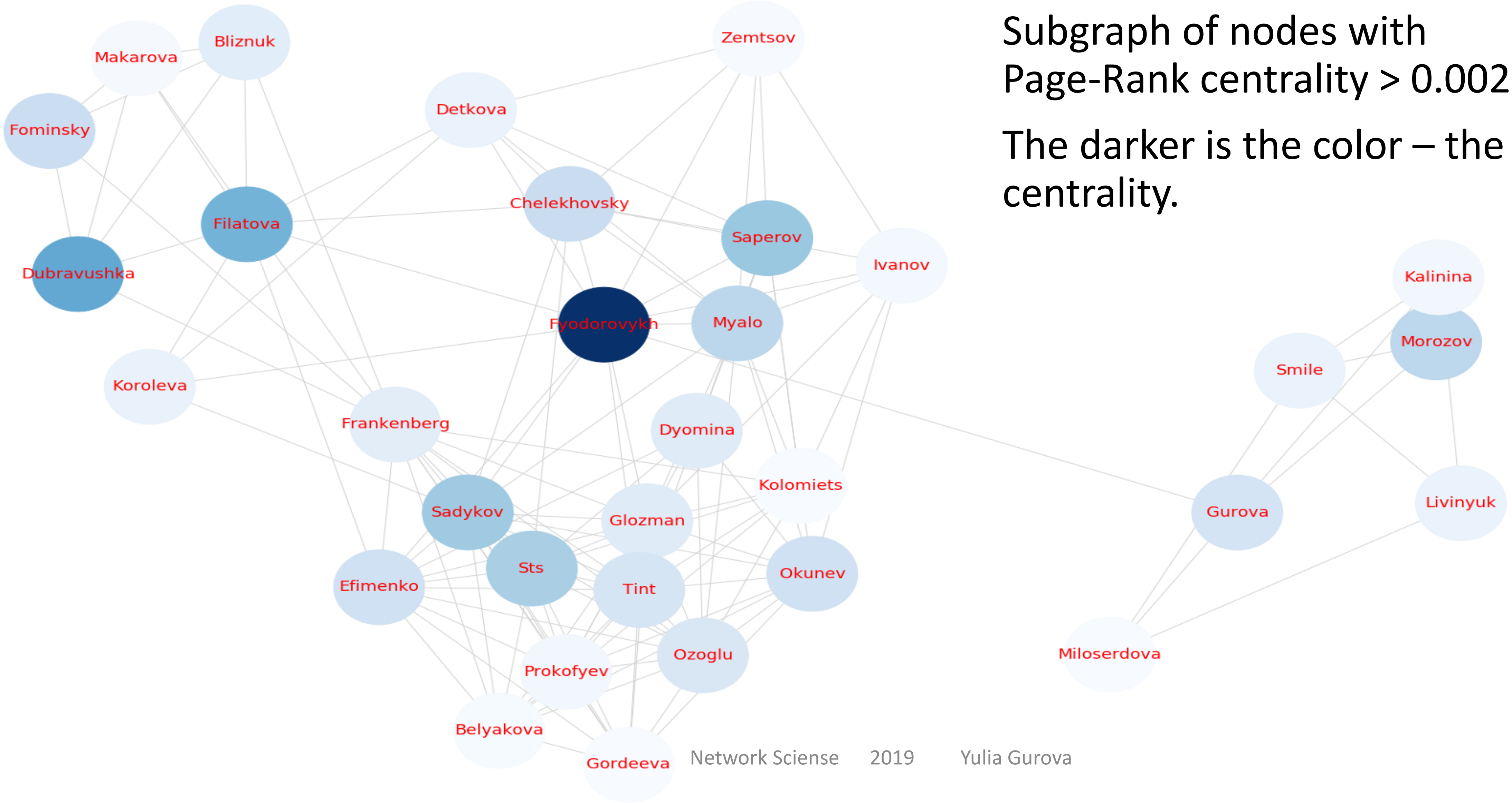
Tint 0.0025

Ozoglu 0.0025

These are people, central in different groups of friends (they cross a lot with the ones with highest betweennes centrality).

Because I took a=0.7, the random walks was not stuck completely in between the densiest STS group, but went through the central people at each group of friends.

Structural analysis: Page-Rank Centrality

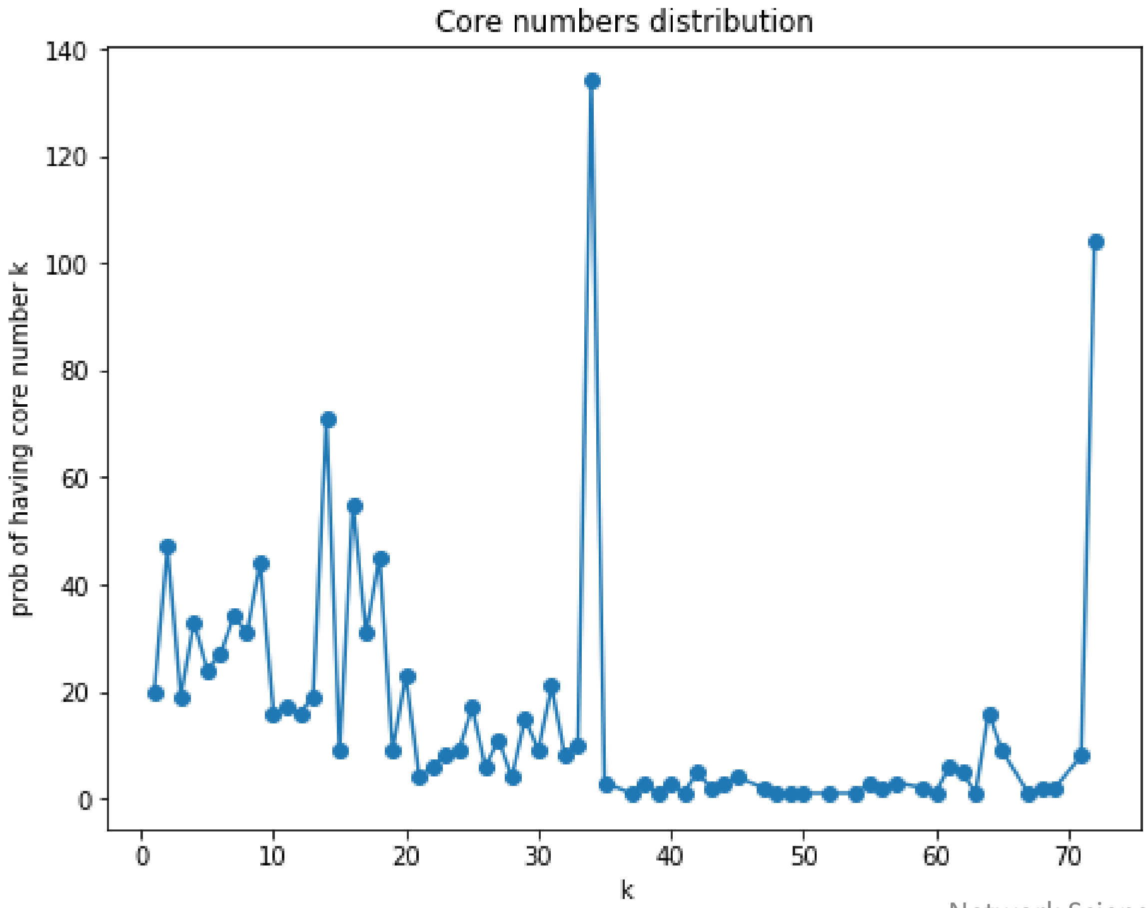


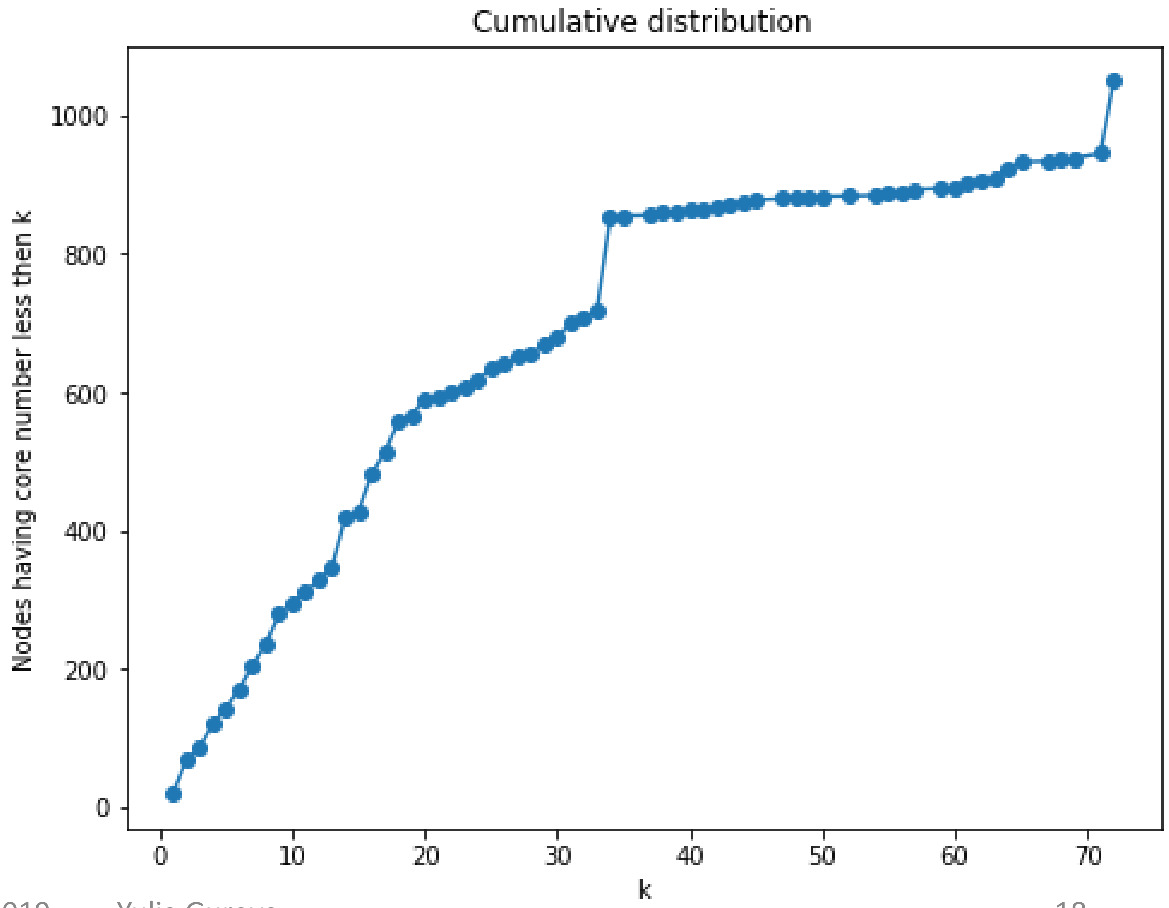
Subgraph of nodes with Page-Rank centrality > 0.002

The darker is the color – the higher is centrality.

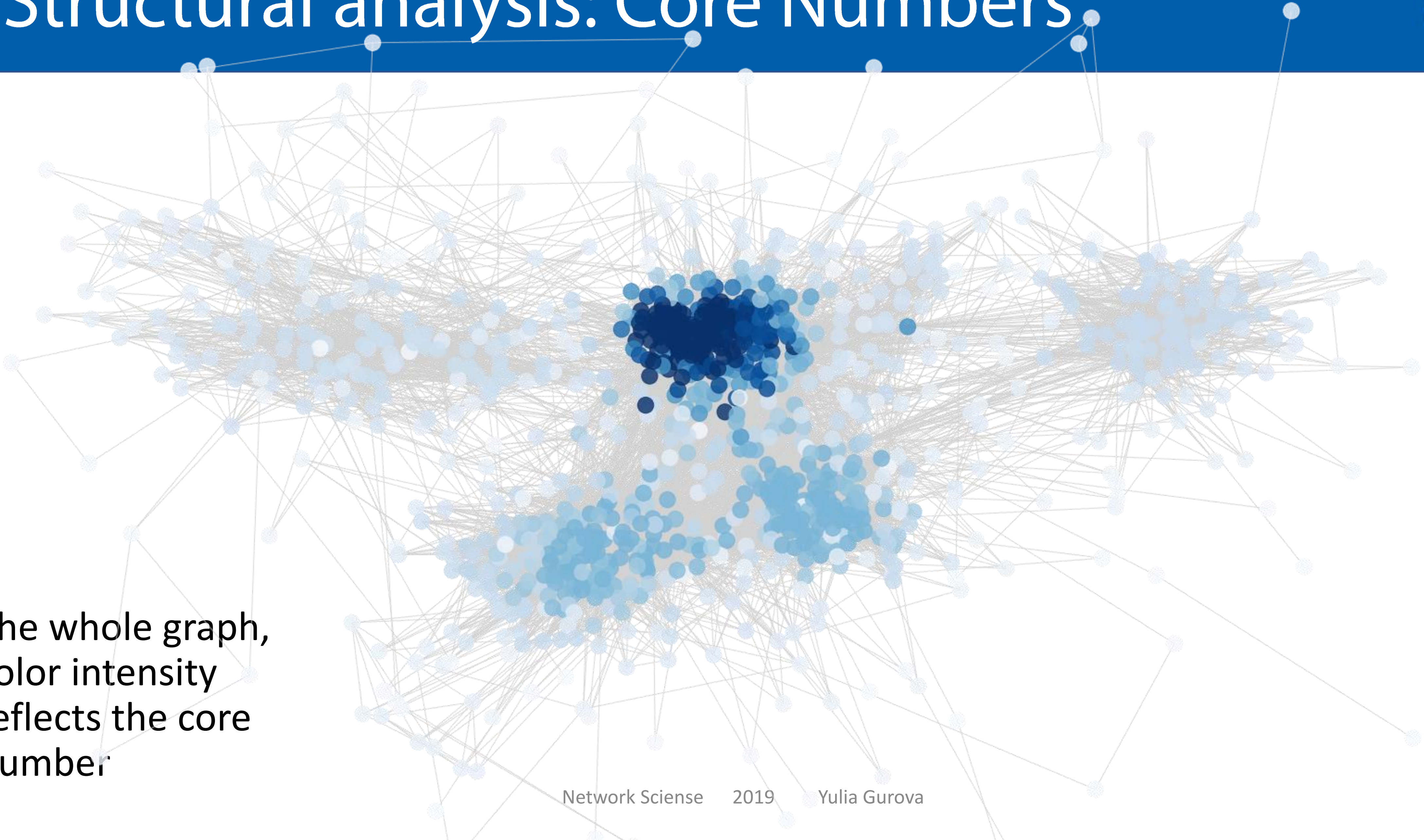
Structural analysis: Core Numbers

Now I calculate core numbers for all the vertices. Max core for my graph is 72. The distribution of the core numbers :





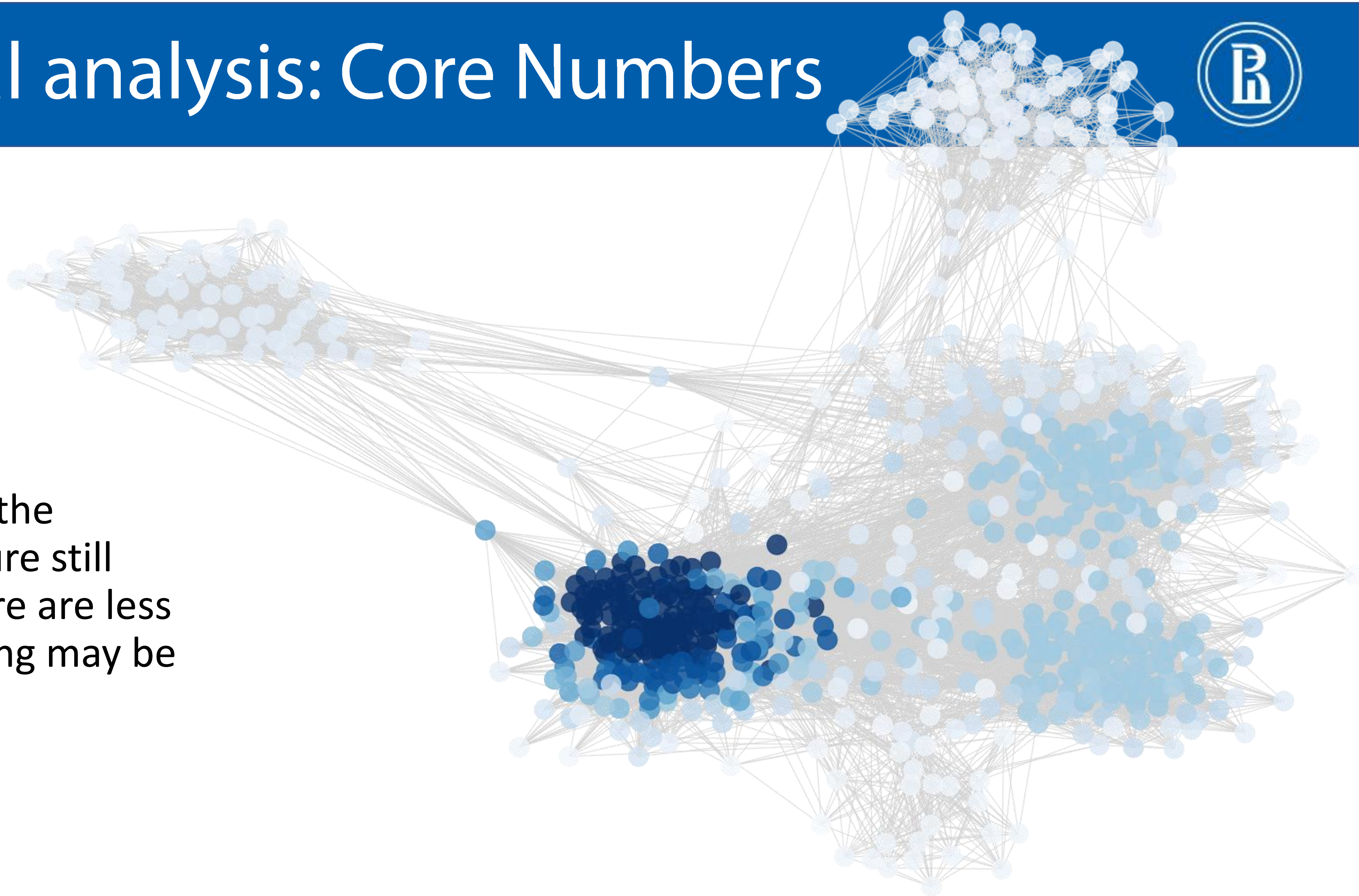
Structural analysis: Core Numbers



The whole graph, color intensity reflects the core number

Structural analysis: Core Numbers

The 12-core. Here the community structure still preserves, but there are less nodes, so everything may be seen more clearly



Structural analysis: Assortativity Coefficient

Degree assortativity coefficient for graph: 0.35095

It is rather high, that's expected for such kind of social network

Attribute assortativity coefficient for graph:

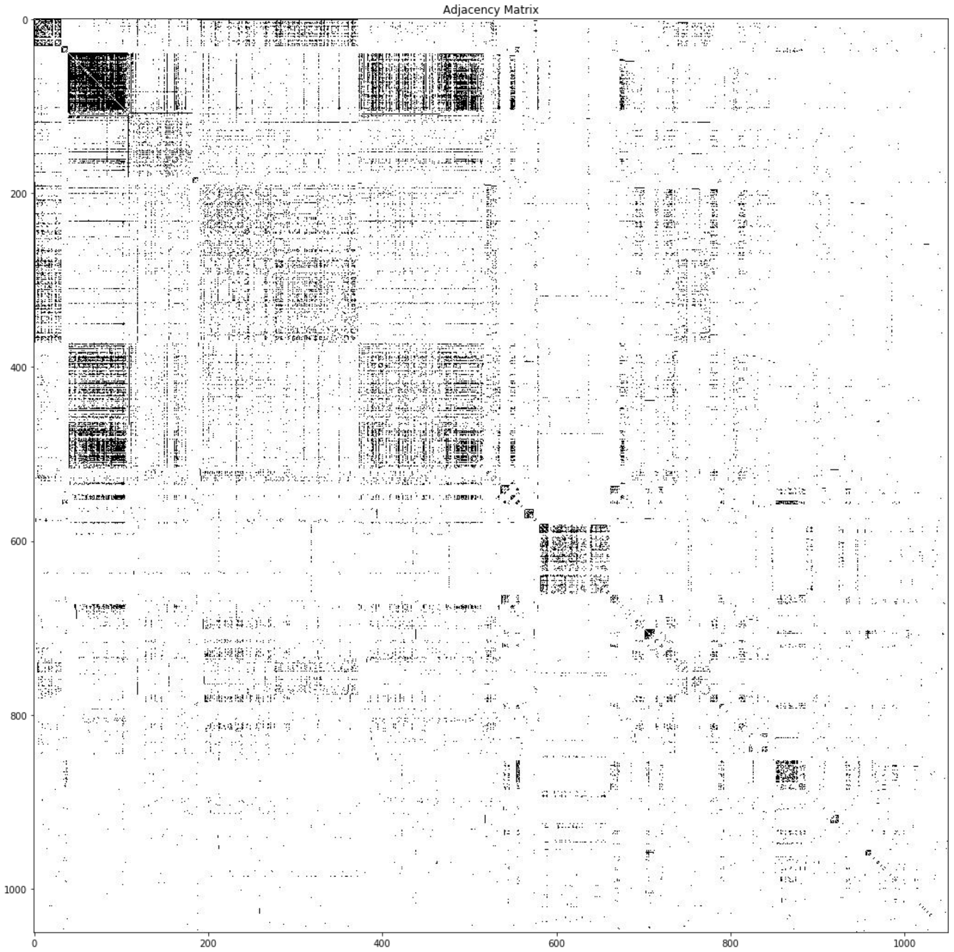
sex: 0.03878

Bdate: -0.0019

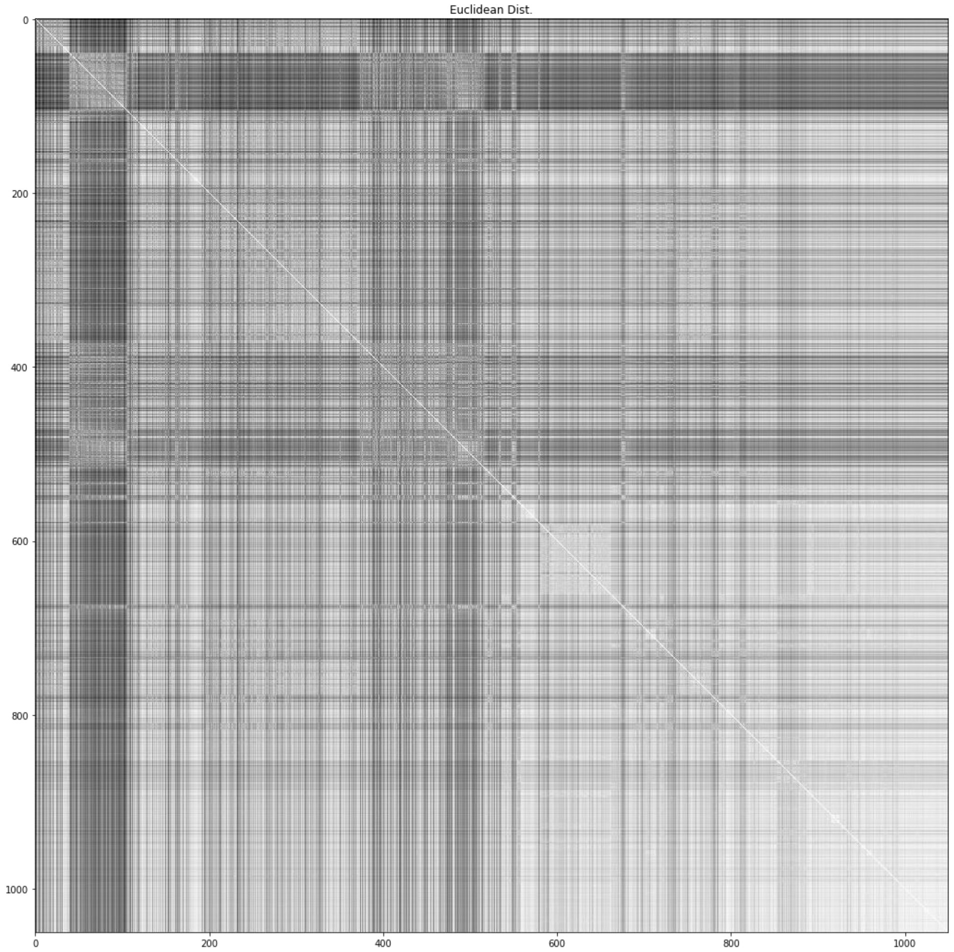
That can be explained by the fact that not every one shares the birthday date and the nearest days are counted as different as far away days

Structural analysis: Ajacency Matrix

Ajacency matrix

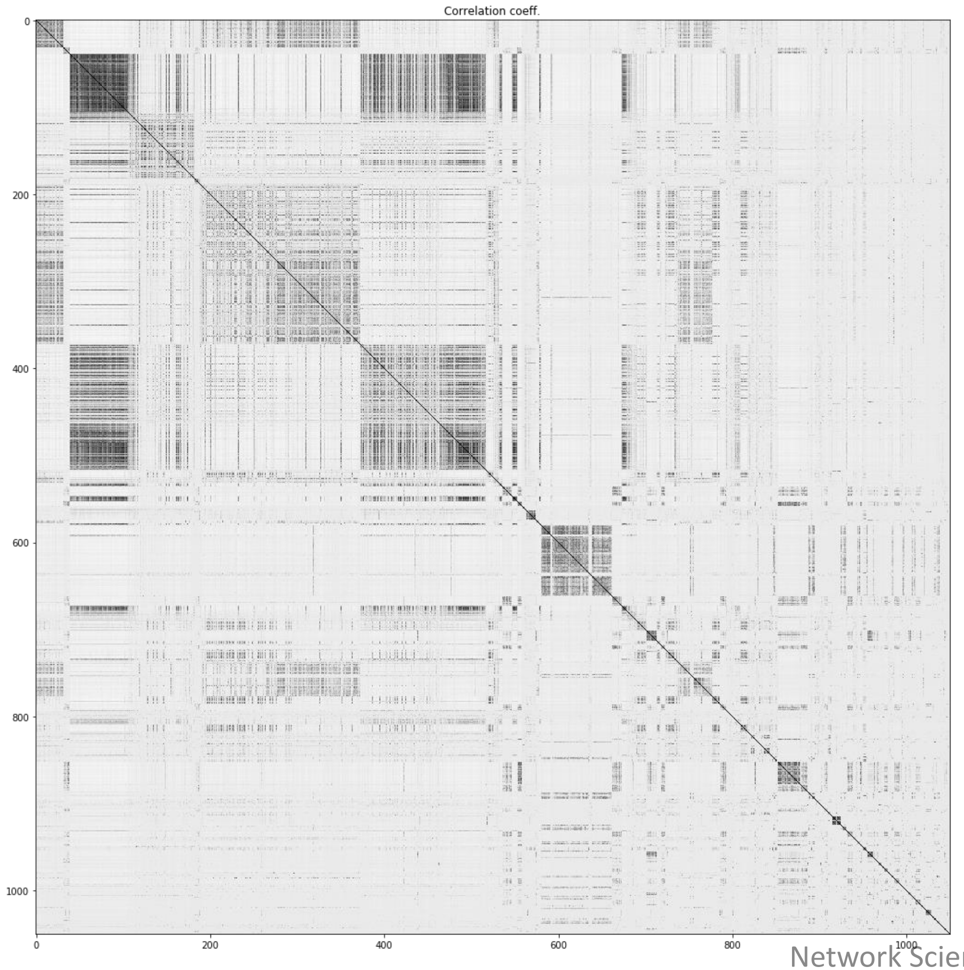


Euclidean distance

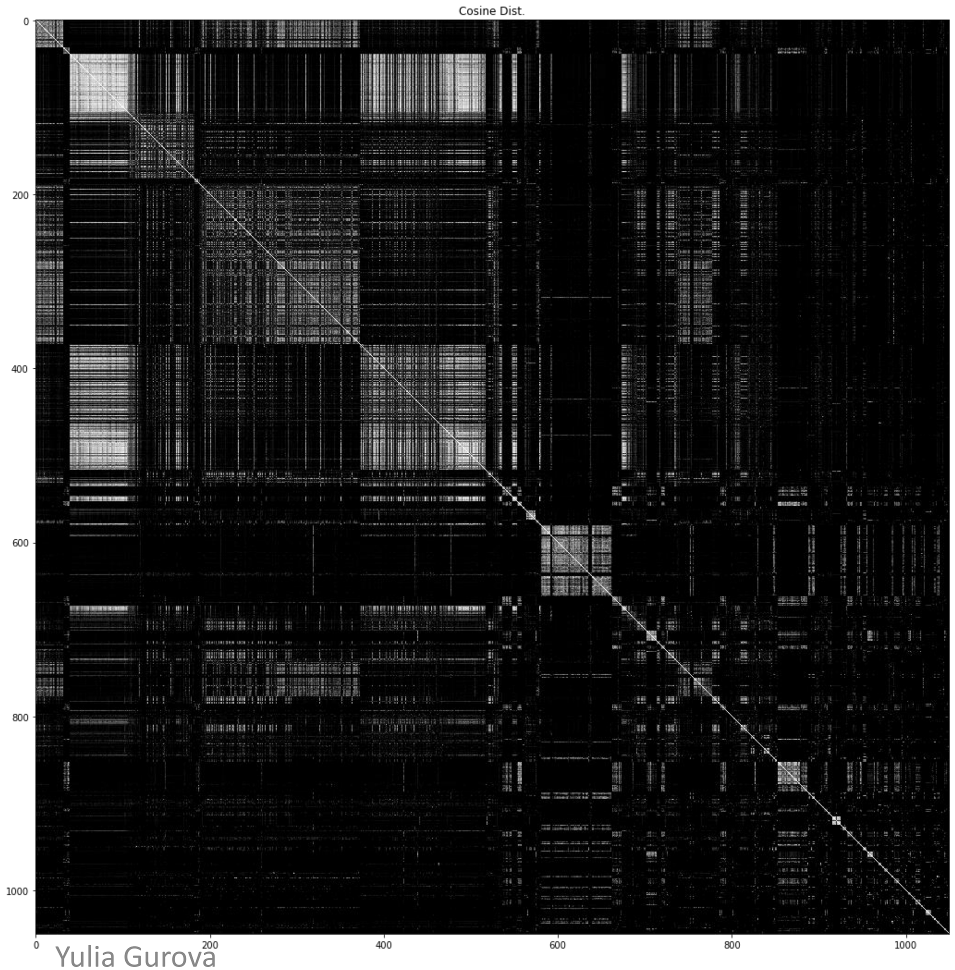


Structural analysis: Ajacency Matrix

Correlation Coefficient



Cosine distance

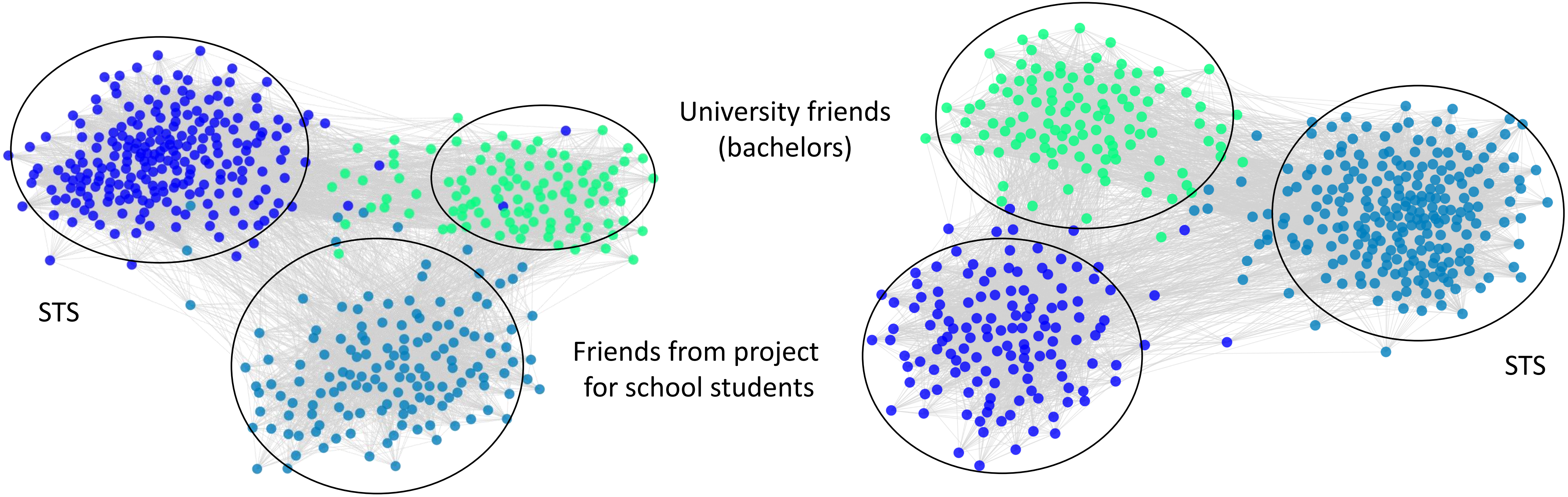


Community Detection

The best results of community detection algorithms. As graph is really large, the clique algorithms do not work in reaslnable time. Even for smaller part (core) of the graph it doesn’t work as it is very dense (20-core contains 483 nodes). So, I used other algorithms.

Community Detection: 20-core

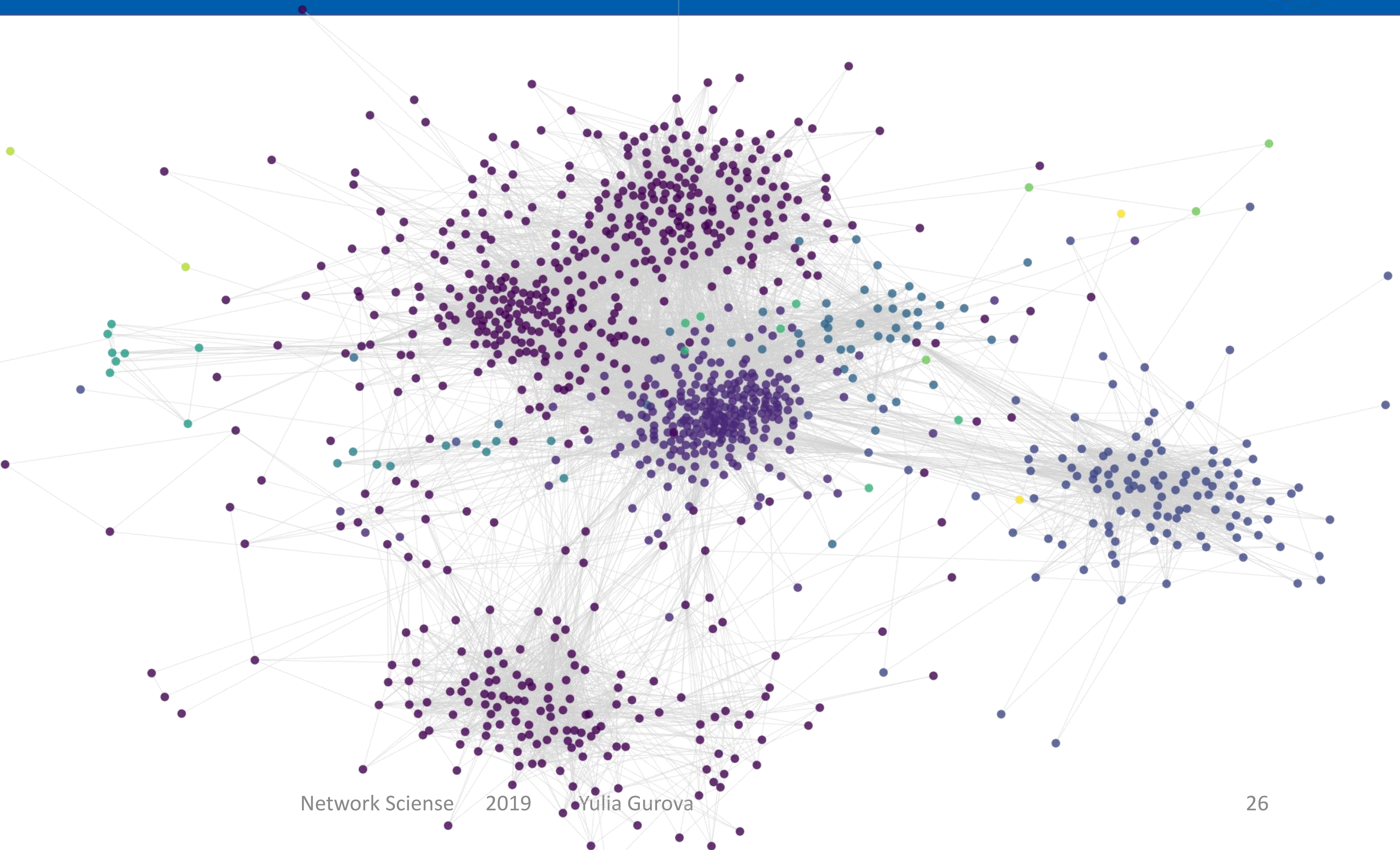
Clauset-Newman-Moore greedy modularity maximization. Fluid Communities algorithm (using Nearest Neighbours).



Both algorithms found interpretable communities. Looks like Fluid Communities algorithm did better, without layouts that CNM has. But it needs the determined number of communities, while CNM finds optimal number itself.

Community Detection: the whole graph

On the whole graph modularity maximization algorithm didn’t work well. It detects 10 communities, but all the 20-core from the previous page was detected as one community. It separated some small groups of nodes, only one of them is really a community (master’s friends). All the others are smatched between each other and are not interpretable



Summary

- The VK-friends graph was analysed
- There vere some problems due to specificity: many nodes, high dense for one group of friends
- Looks like the graph satisfies the small-world properties
- Community-detection algorithms were checked, they work well on the core, but not for the whole graph