

Проектная работа «Анализ социальной сети VK»

Анастасия Валеева,
ФПМИ, М05-116

2021г.

Как же мне
построить граф



ГРАФ ДРУЗЕЙ

- 38 вершин
- 55 рёбер
 - слишком маленький и несвязный

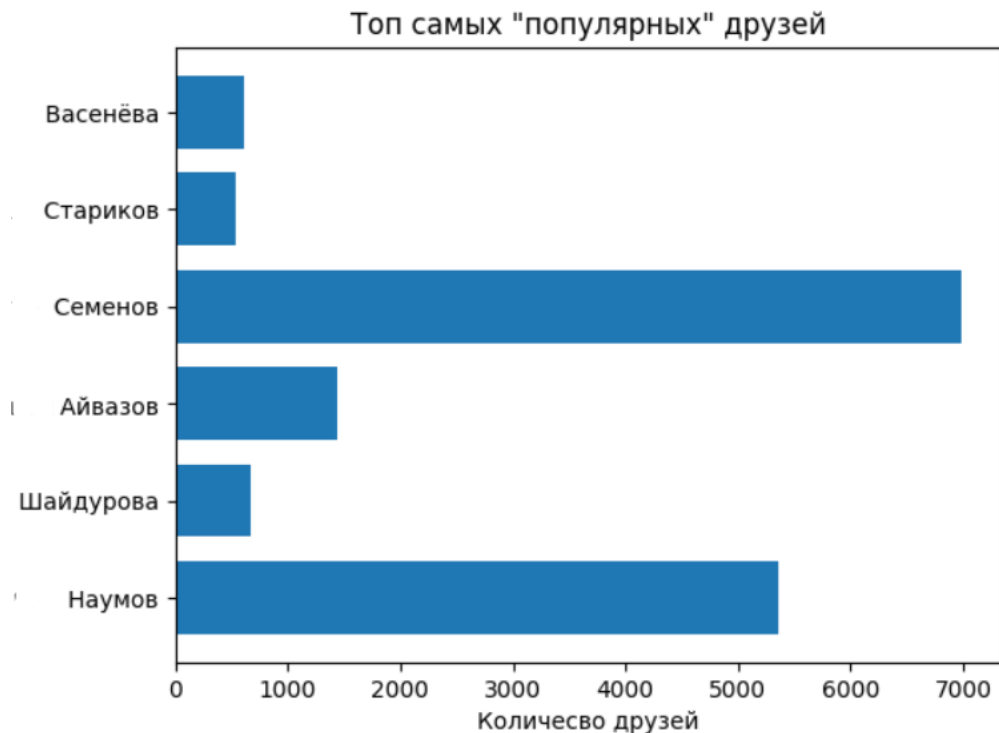
ГРАФ ДРУЗЕЙ ДРУЗЕЙ

- Более 16.000 вершин, степень некоторых >6.000
- На порядок больше рёбер
 - слишком большой и малоинформативный

Гипотеза 70-7

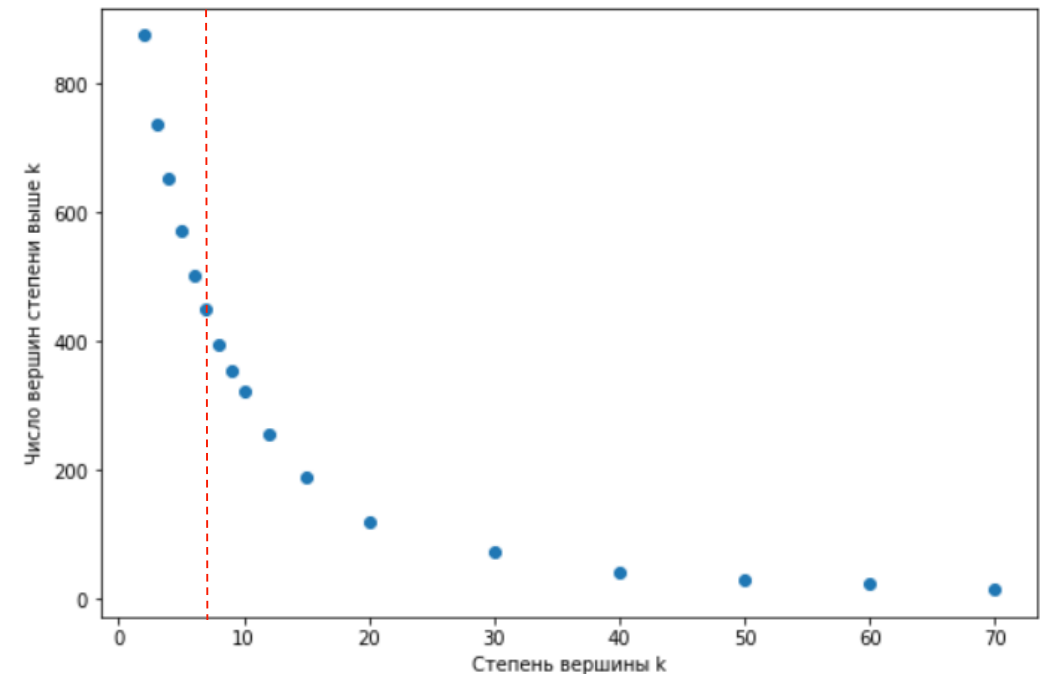
70 друзей

Добавим только 70 случайных друзей каждого моего друга и все связи между ними



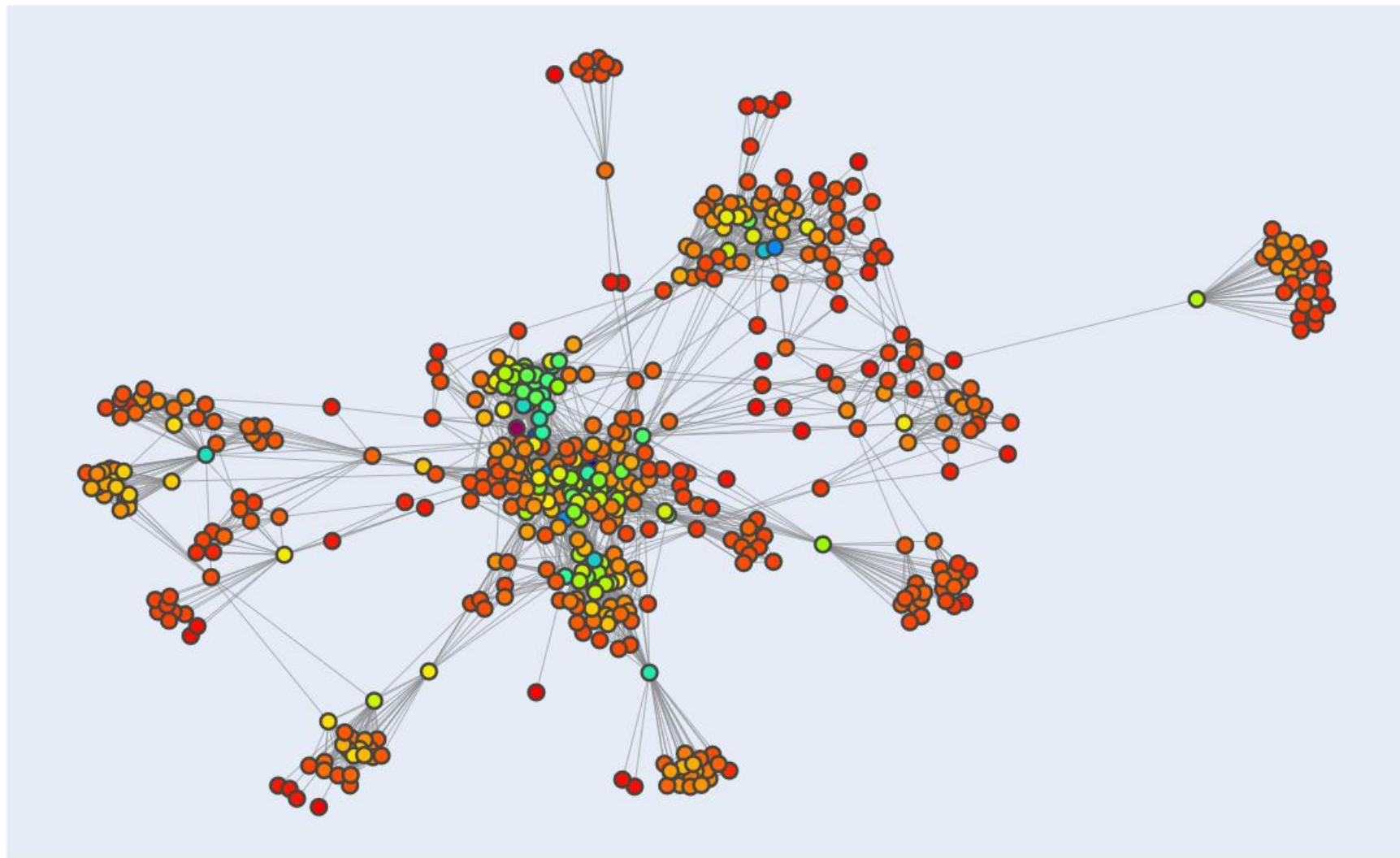
Степень вершины < 7

Удалим вершины степени ниже 7, избавимся от слабосвязанных

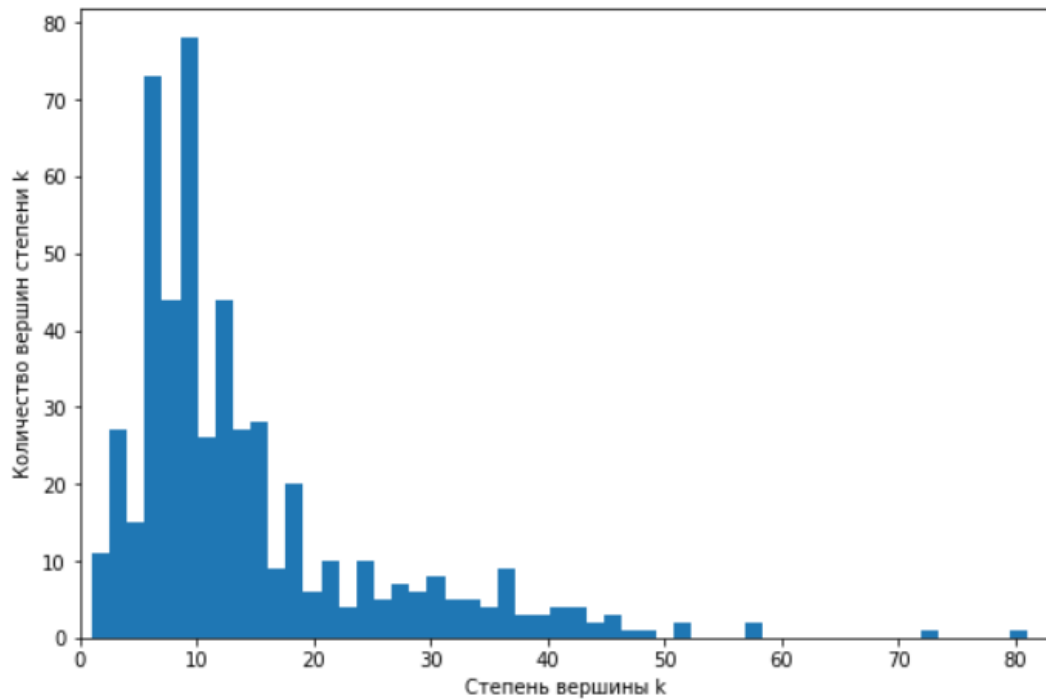


Описание графа

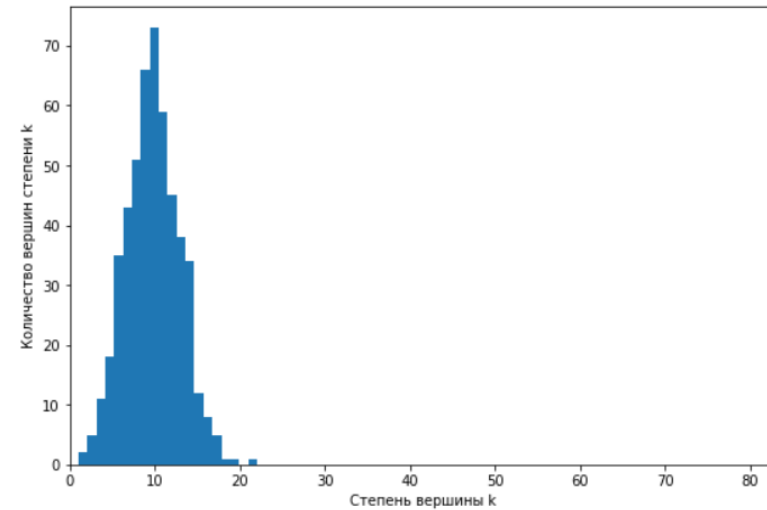
- Вершины: 508
- Рёбра: 3704
- Диаметр: 4
- Радиус: 2
- Коэффициент кластеризации: 0.6
- Средняя длина пути: 3.38
- Женщины: 45%,
мужчины: 55%



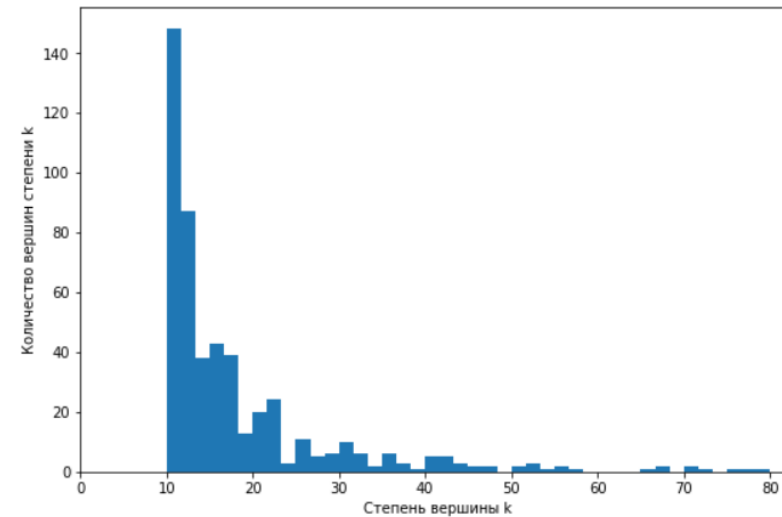
Распределение степеней вершин



Мой граф



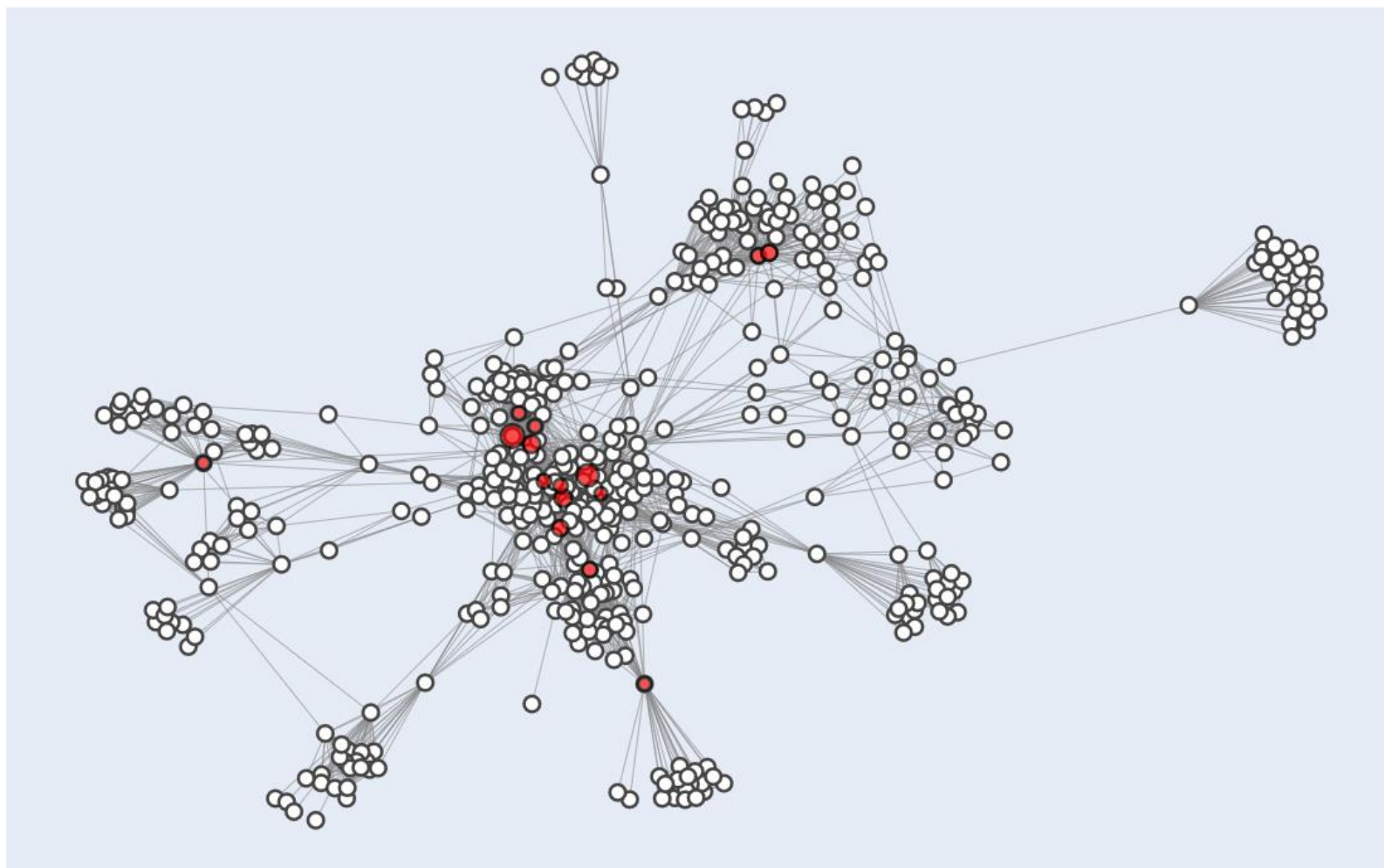
Эрдёш-Реньи
(508, 0.02)



Барбаши-Альберт
(508, 10)

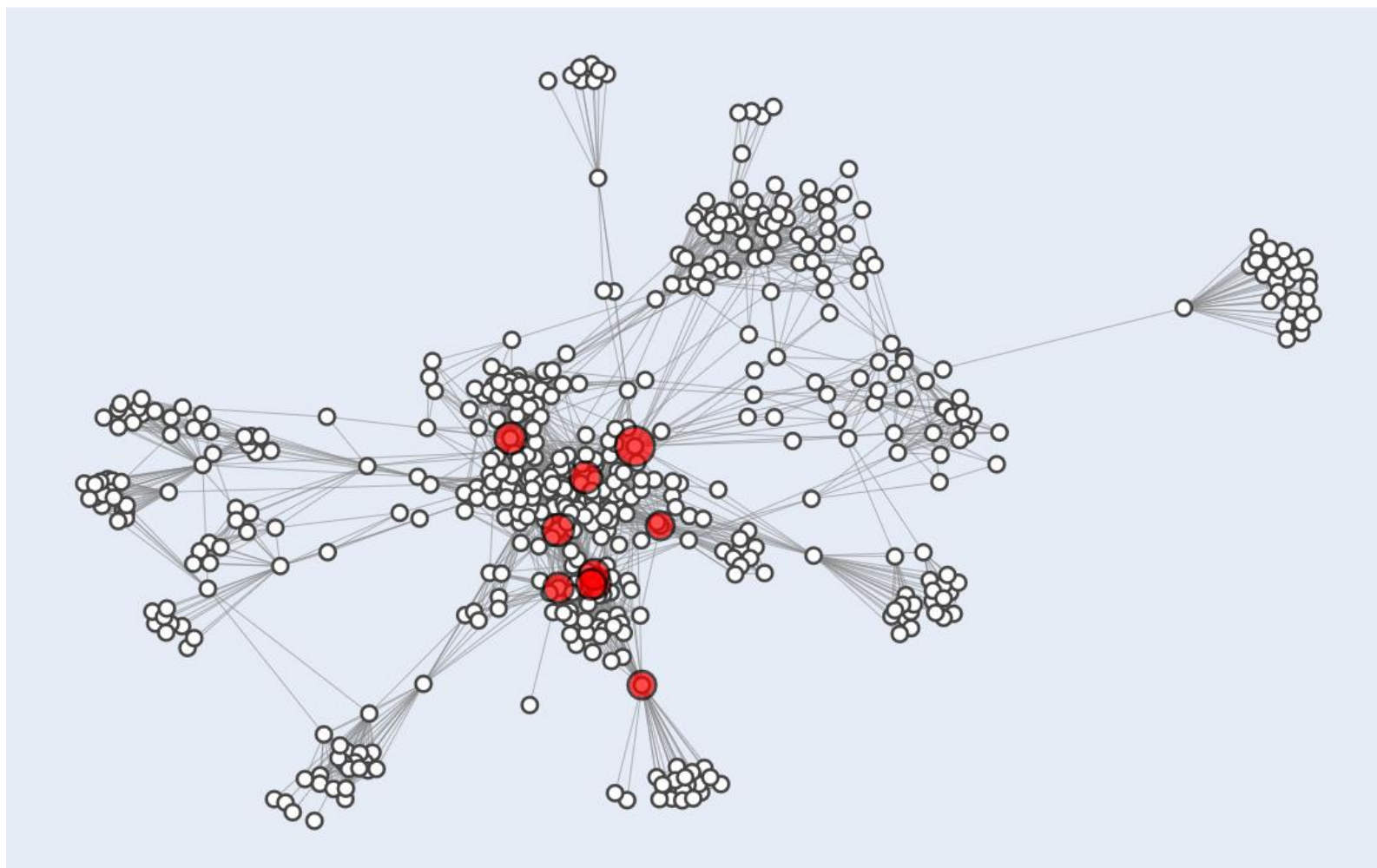
Центральности. Degree centrality

Екатерина Шайдурова 0.1598
Денис Айвазов 0.142
Александр Гришутин 0.1124
Максим Абрамов 0.1124
Руслан Еникеев 0.1026
Наталья Лунева 0.1006
Иван Эрлих 0.0966
Никита Семёнов 0.0927
Александр Ненашев 0.0907
Евгений Молчанов 0.0907
Владимир Гуровиц 0.0888
Маргарита Автандилова 0.0868
Андрей Власов 0.0868
Михаил Голофаев 0.0848
Виталий Павленко 0.0848



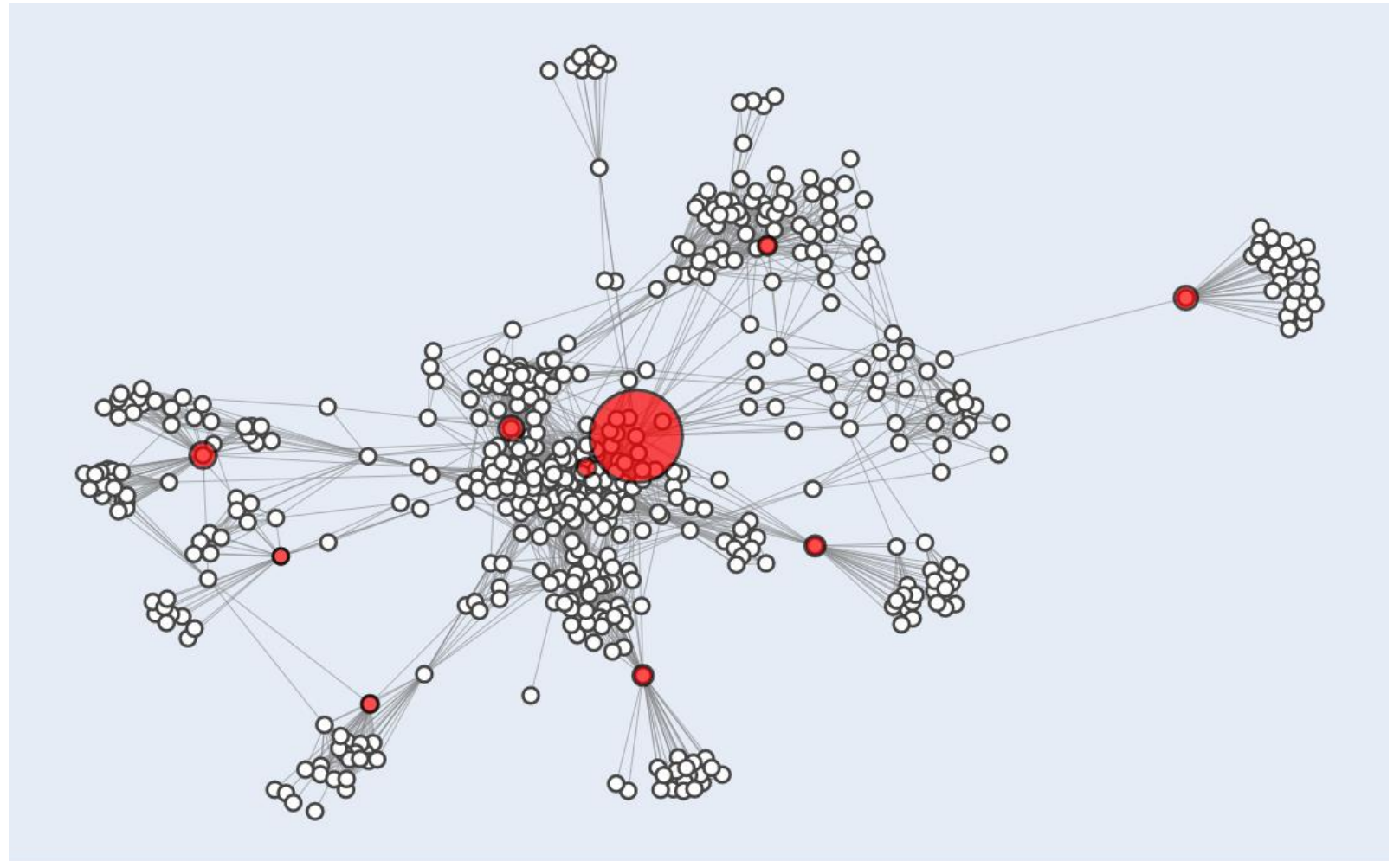
Центральности. Closeness centrality

Анастасия Валеева 0.5195
Денис Айвазов 0.4159
Екатерина Шайдурова 0.4139
Наталья Лунева 0.4005
Максим Регенов 0.3949
Владислав Рыхлов 0.3912
Айрат Валиев 0.3876
Александра Штепен 0.3856
Евгения Бобренкова 0.3798
Михаил Голофаев 0.3798



Центральности. Betweenness centrality

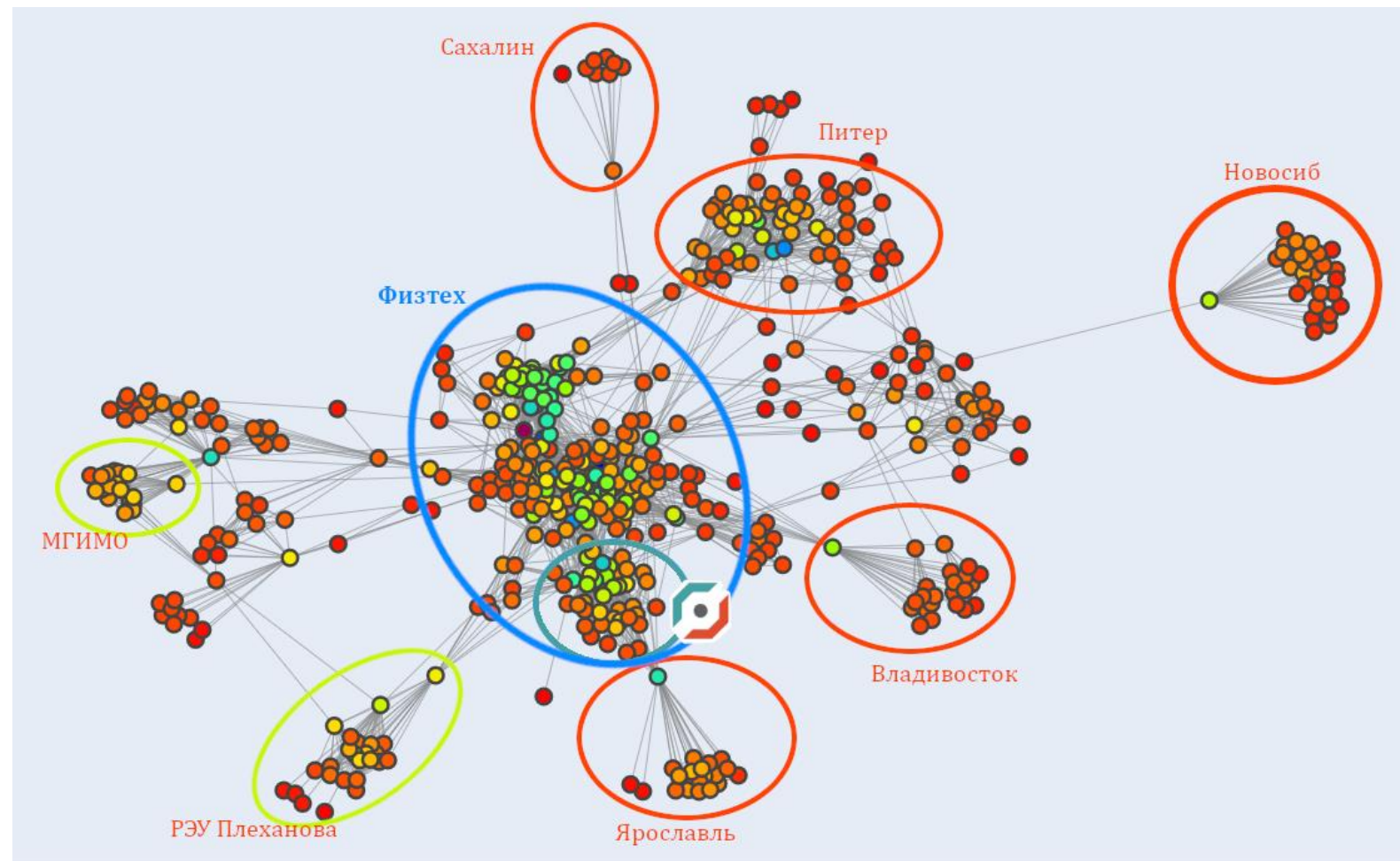
Анастасия Валеева 0.5486
Маргарита Автандилова 0.1249
Екатерина Шайдурова 0.1162
Татьяна Сорокина 0.1022
Каролина Акохова 0.0862
Михаил Голофаев 0.0858
Руслан Еникеев 0.0795
Денис Айвазов 0.0795
Дарья Серегина 0.0681
Дмитрий Черников 0.0592



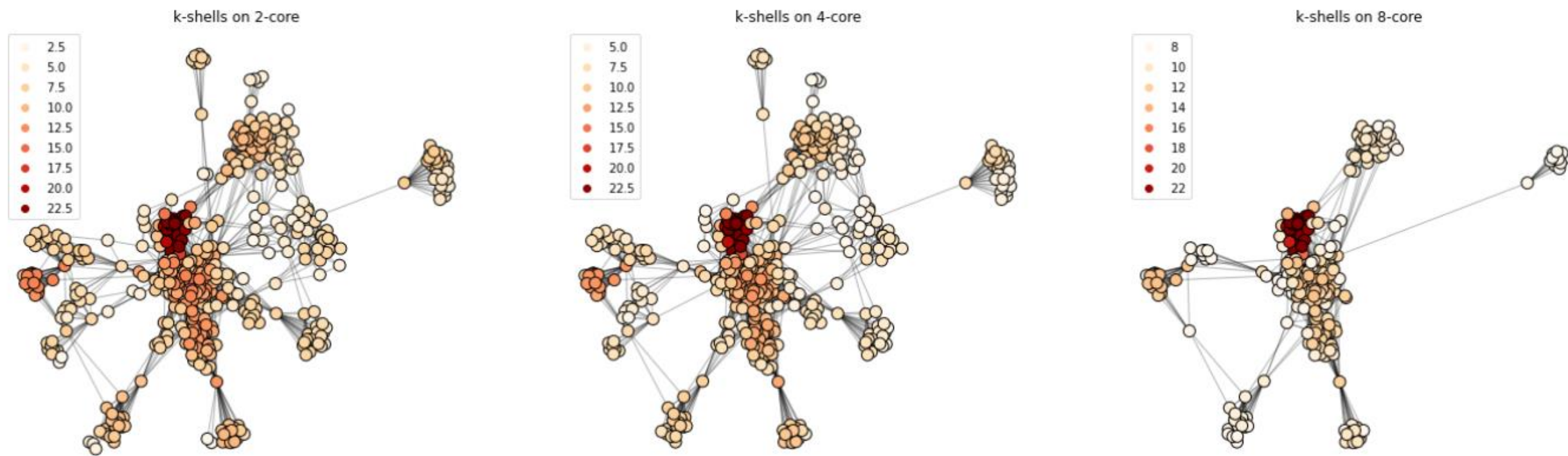
Определение сообществ

Что мы хотим
найти?

Проведем ручной анализ графа:
наблюдается деление по
географическому и
университетскому признакам



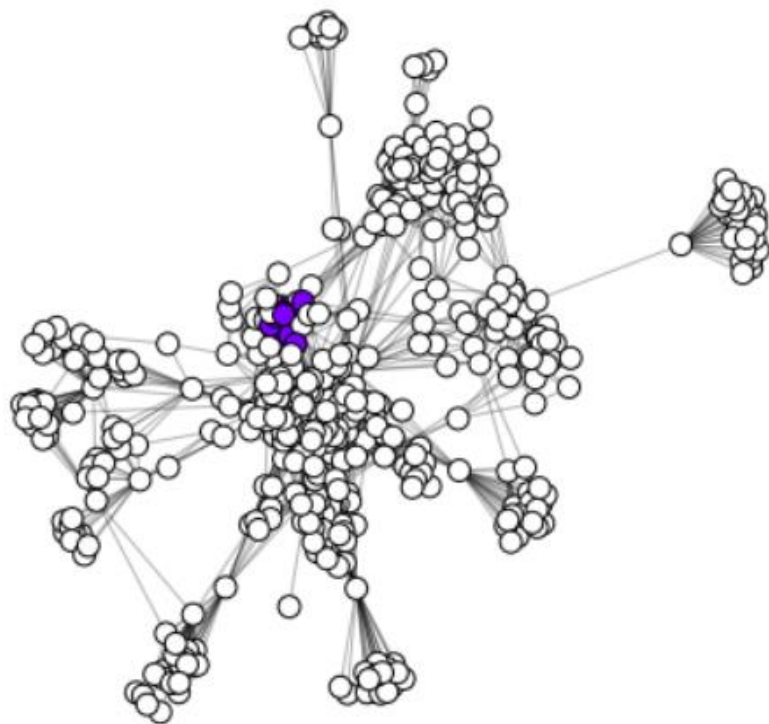
Определение сообществ. k-cores



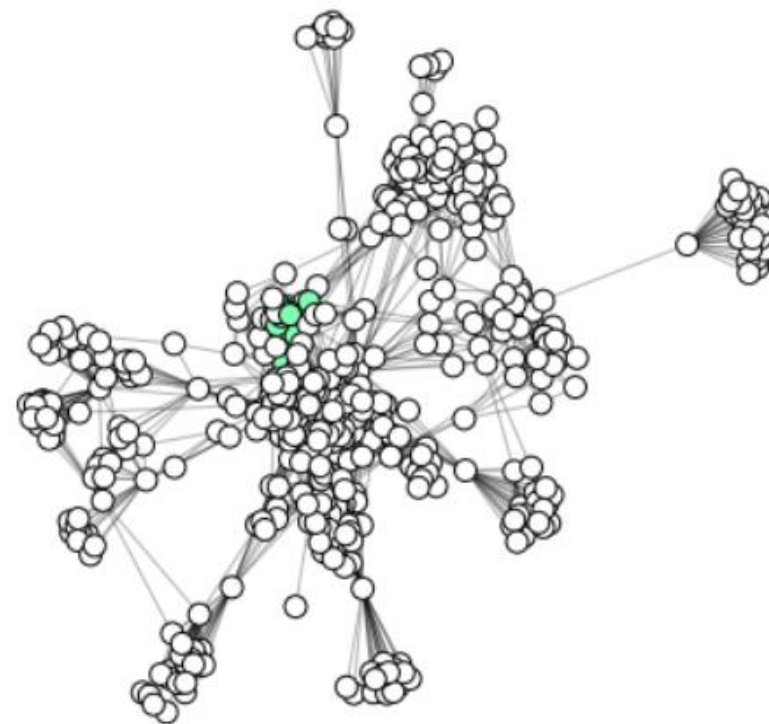
Получилось выделить основное ядро Физтеха. Не особо информативно

Определение сообществ. Cliques

Clique of the size 22

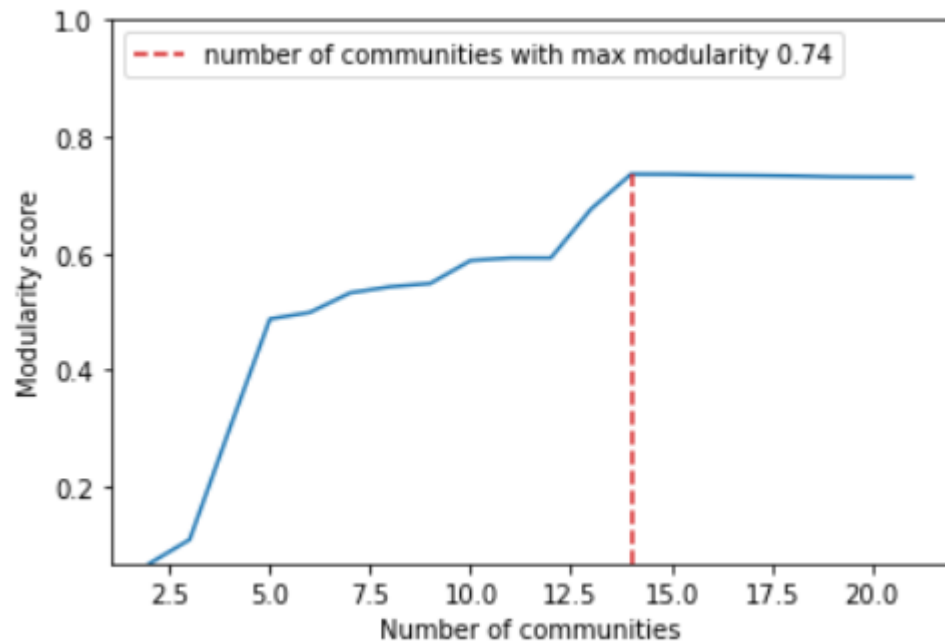


Clique of the size 22

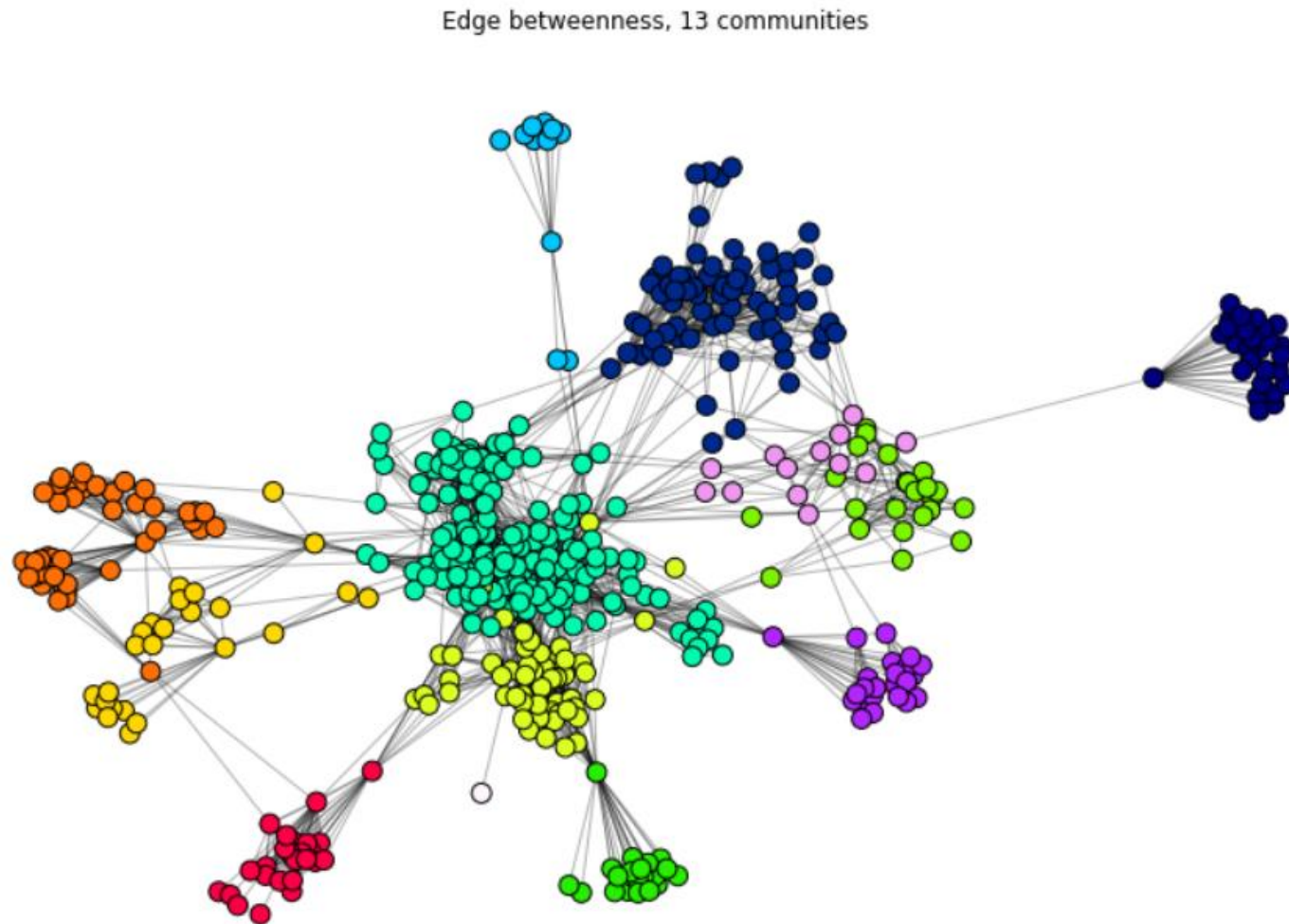


Похожий на **k-cores** результат. Выделил ядро, но чуть меньшего размера.

Определение сообществ. Girvan Newman



Максимизация **modularity** привела к очень точному результату, выделив все осмысленные группы!



Заключение

- Гипотеза 70 – 7 о сохранении структуры графа при понижении размерности подтвердилась, все полученные метрики продемонстрировали корректно интерпретируемые значения
- Распределение степеней вершин соответствует **power law**
- С помощью алгоритма Гирвана-Ньюмана удалось выделить все существующие сообщества в графе