

Санкт–Петербургский Государственный Университет

Поляков Иван Михайлович

Отчёт по домашнему заданию №1

«Анализ структуры сети»

«Распределение степеней вершин, кластерный коэффициент»

«Метрики узлов графа-друзей vk»

Направление 01.04.02: «Прикладная математика и информатика»

Образовательная программа ВМ.5505.2021: «Математическое и информационное
обеспечение экономической деятельности»

Руководитель:
кандидат физ.-мат. наук,
доцент Воронкова Ева Боруховна

Санкт-Петербург
2023 г.

1 Анализ структуры сети

В данной части работы были даны 4 графа:

- CA-AstroPh – граф соавторства в области астрофизики (неориентированный);
- vk – граф пользователей ВКонтакте (неориентированный);
- vk_friends – личный граф друзей ВКонтакте (неориентированный);
- web-Google (ориентированный).

В соответствии с заданием были вычислены следующие значения для каждого графа и построены гистограммы:

	CA-AstroPh	vk	vk_friends	web-Google
Вершины	18772	2736186	53	875713
Рёбра	198110	9140970	169	5105039
Плотность	0.0011	0.0000	0.1226	0.0000
Число компонент слабой связности	290	45695	4	2746
Доля вершин в максимальной по мощности компоненте связной связности	95.37%	96.32%	60.38%	97.73%
Число компонент сильной связности	-	-	-	371764
Доля вершин графа в наибольшей компоненте сильной связности	-	-	-	49.65%
Число вершин степени, превышающей среднюю степень	5594	624037	26	265436
Радиус	8	14	3	29
Диаметр	14	16	4	40
90 перцентиль расстояния	5	8	3	14
Число треугольников	1351441	15635422	241	13391903
Средний кластерный коэффициент	0.6306	0.0237	0.5231	0.5143
Глобальный кластерный коэффициент	0.3180	0.0576	0.5583	0.0552
Минимальная степень	1	1	1	1
Средняя степень	21.11	6.68	6.38	11.66
Максимальная степень	504	3331	17	6353

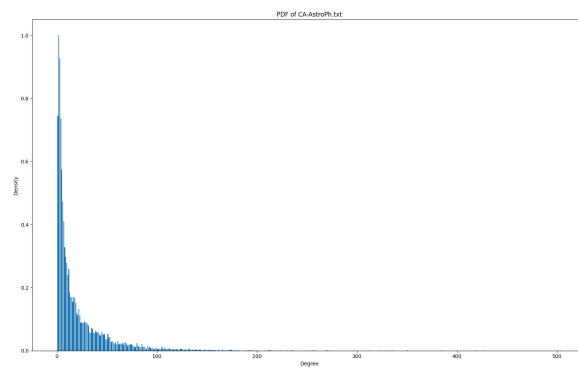


Рис. 1: Функция плотности распределения

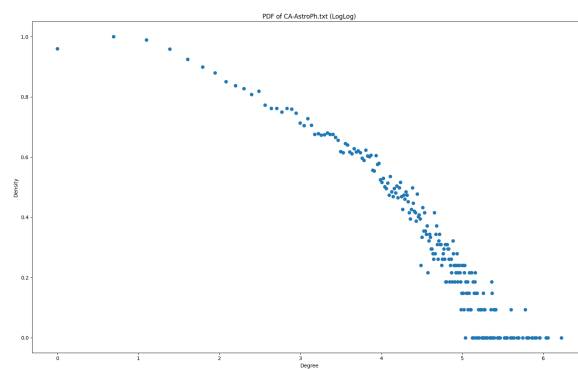


Рис. 2: Функция плотности распределения (LogLog)

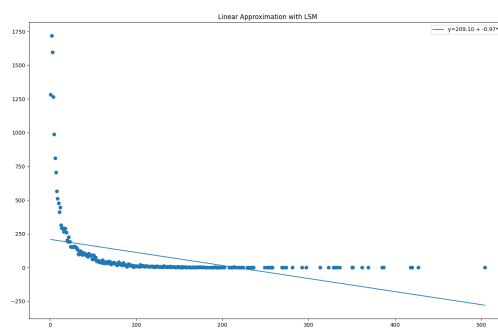


Рис. 3: Аппроксимация линейной функцией

vk

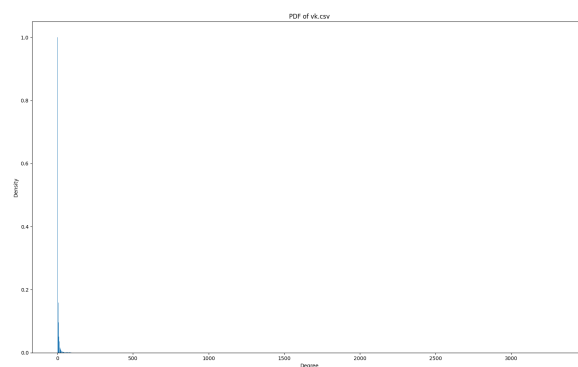


Рис. 4: Функция плотности распределения

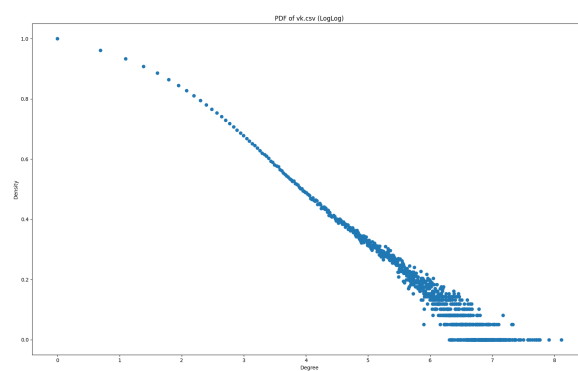


Рис. 5: Функция плотности распределения (LogLog)

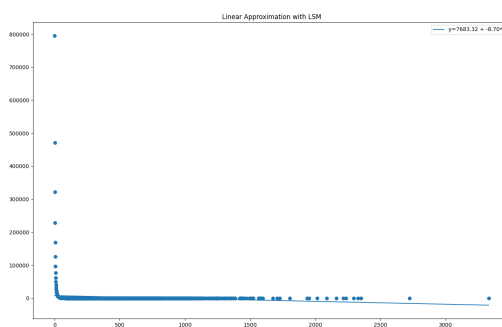


Рис. 6: Аппроксимация линейной функцией

vk_friends

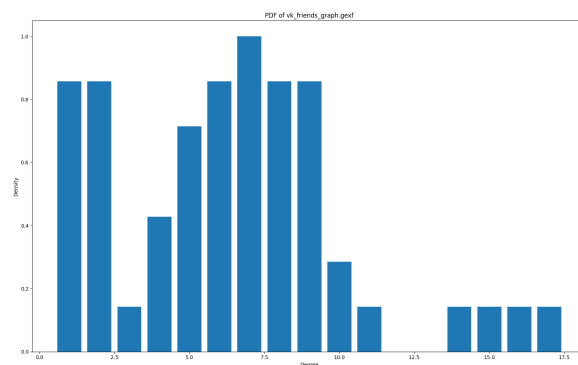


Рис. 7: Функция плотности распределения

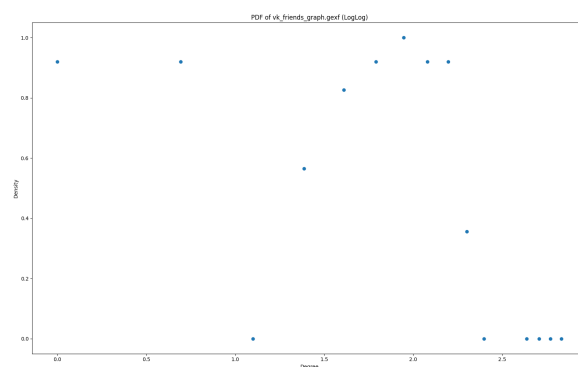


Рис. 8: Функция плотности распределения (LogLog)

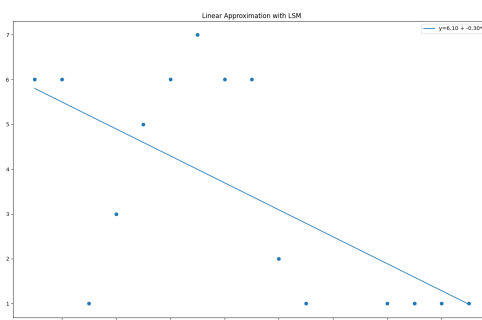


Рис. 9: Аппроксимация линейной функцией

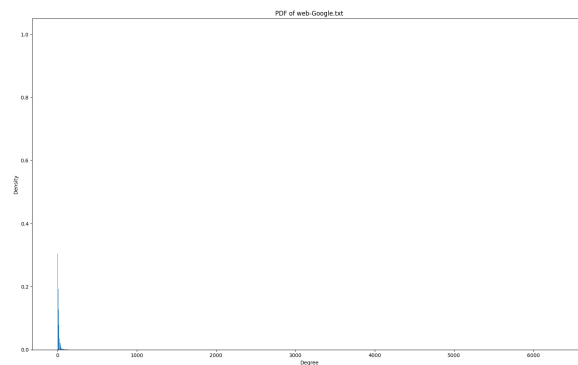


Рис. 10: Функция плотности распределения

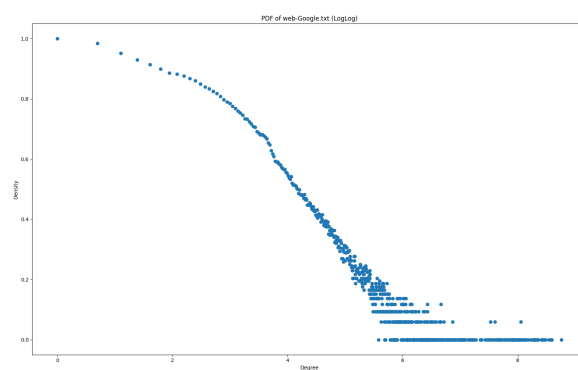


Рис. 11: Функция плотности распределения (LogLog)

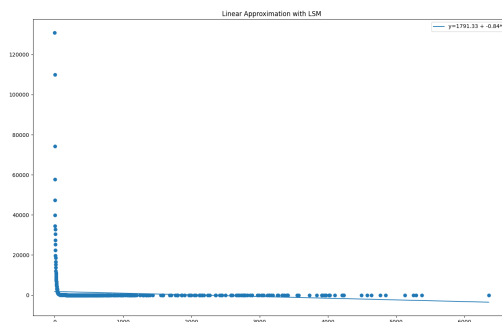


Рис. 12: Аппроксимация линейной функцией

2 Распределение степеней вершин, кластерный коэффициент

В данной части работы предлагалось использовать 3 следующих графа:

1. Граф General Relativity and Quantum Cosmology (GR-QC);
2. Случайный граф $G(n, p)$, где n – число вершин в GR-QC, p – вероятность, которая выбрана так, чтобы математическое ожидание числа рёбер было равно числу рёбер в GR-QC;
3. Случайный граф Модели Малого Мира (WS) с числом вершин и рёбер, совпадающим с числом вершин и рёбер в GR-QC.

Для данных графов были найдены следующие величины:

	GR-QC	$G(n, p)$	WS
Число треугольников	48260	26	5251
Средний кластерный коэффициент	0.5296	0.0009	0.2819
Глобальный кластерный коэффициент	0.6298	0.0010	0.2268
Минимальная степень	1	0	4
Средняя степень	5.5307	5.4136	5.5307
Максимальная степень	81	15	12

Построены функции плотности распределения и кумулятивной функции распределения степеней вершин каждой из сети, а также была проведена аппроксимация функции плотности распределения графа GR-QC с помощью метода наименьших квадратов:

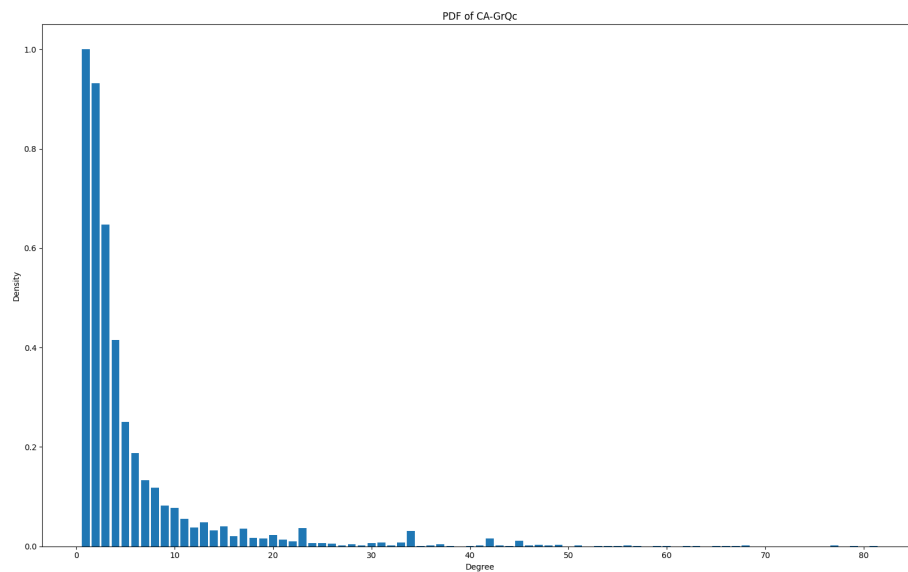


Рис. 13: Функция плотности распределения

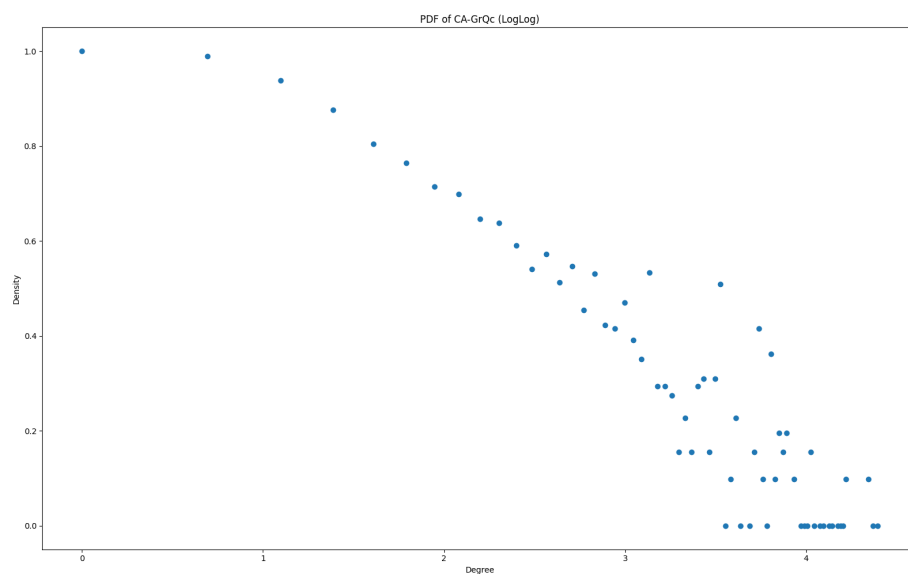


Рис. 14: Функция плотности распределения (LogLog)

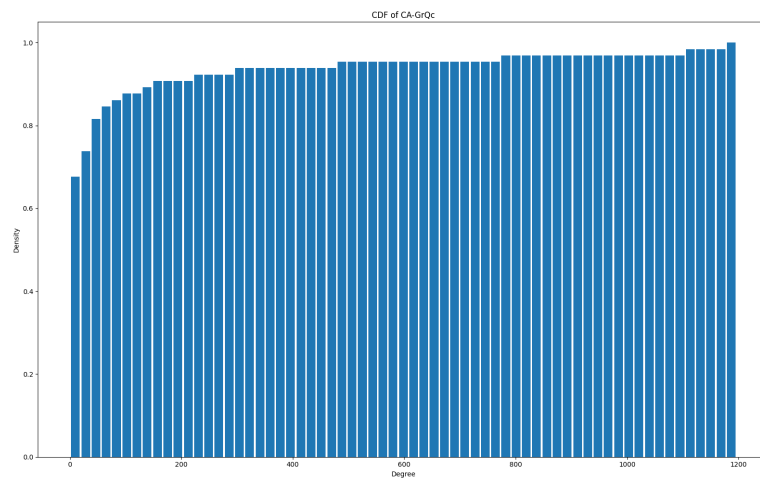


Рис. 15: Кумулятивная функция распределения степеней

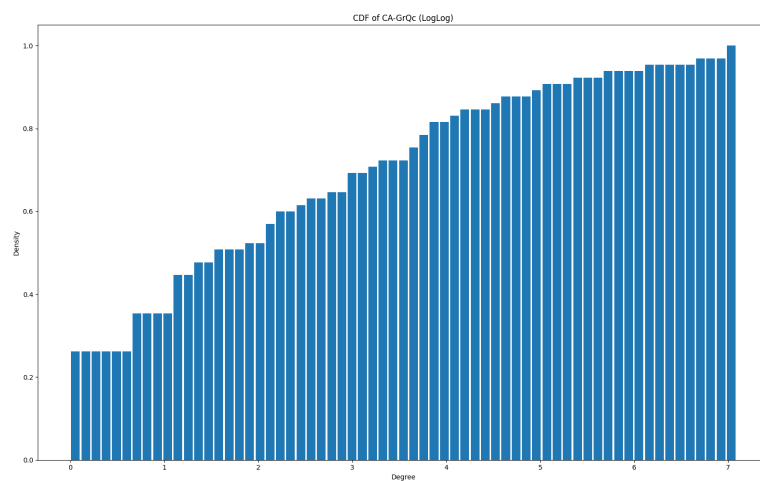


Рис. 16: Кумулятивная функция распределения степеней (LogLog)

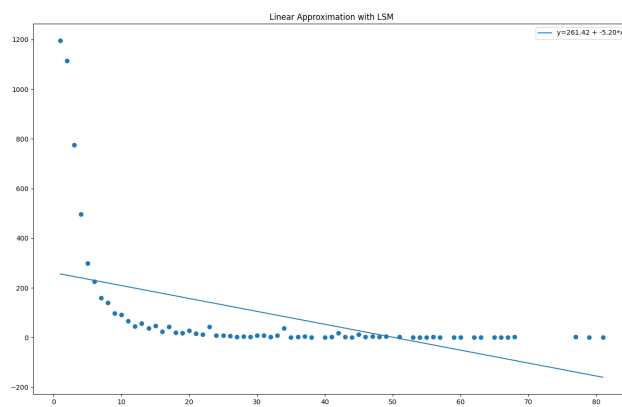


Рис. 17: Аппроксимация линейной функцией

$G(n, p)$

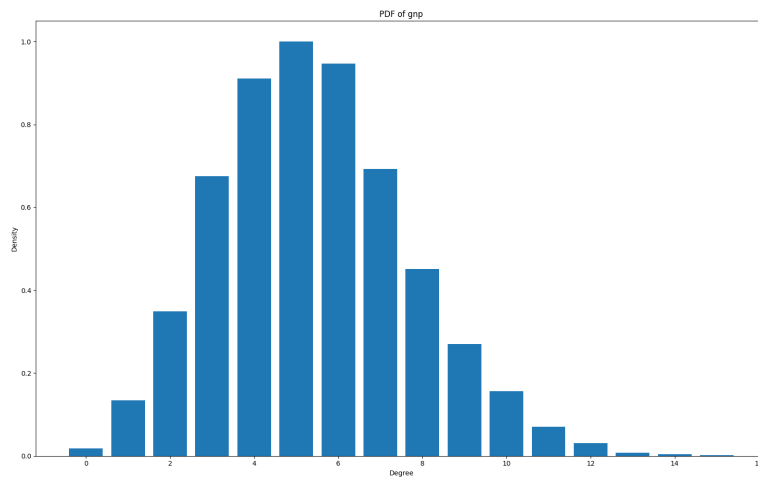


Рис. 18: Функция плотности распределения

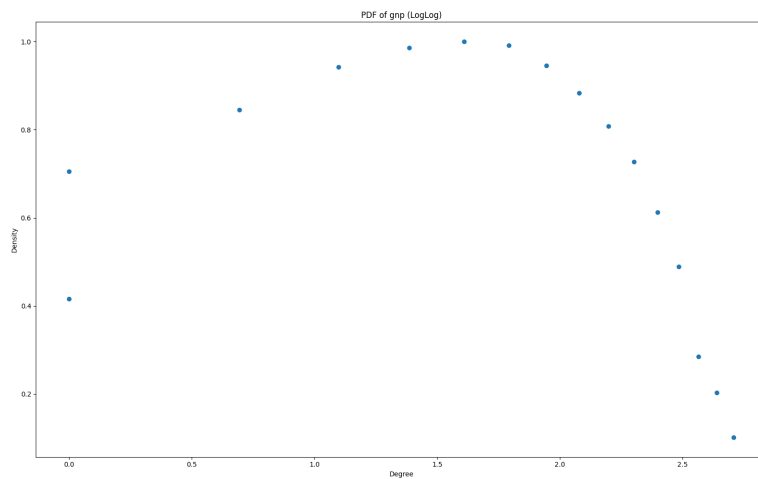


Рис. 19: Функция плотности распределения (LogLog)

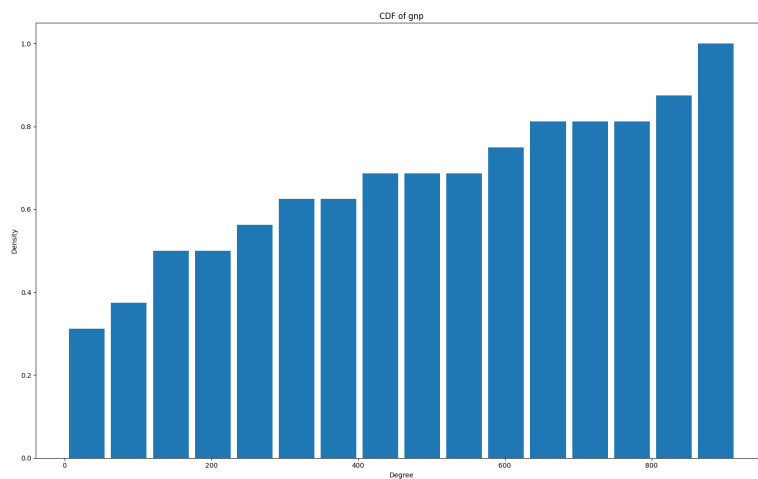


Рис. 20: Кумулятивная функция распределения степеней

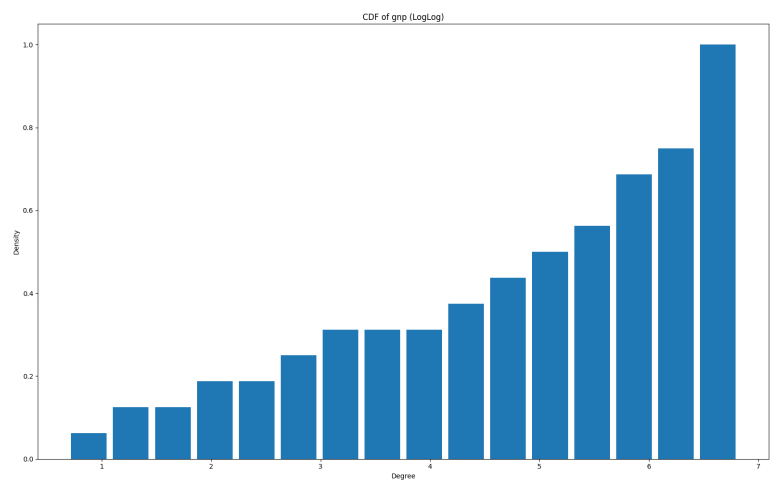


Рис. 21: Кумулятивная функция распределения степеней (LogLog)

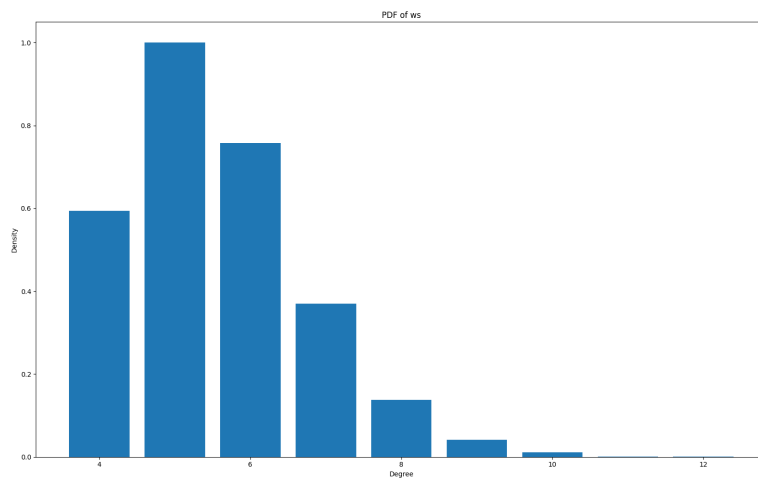


Рис. 22: Функция плотности распределения

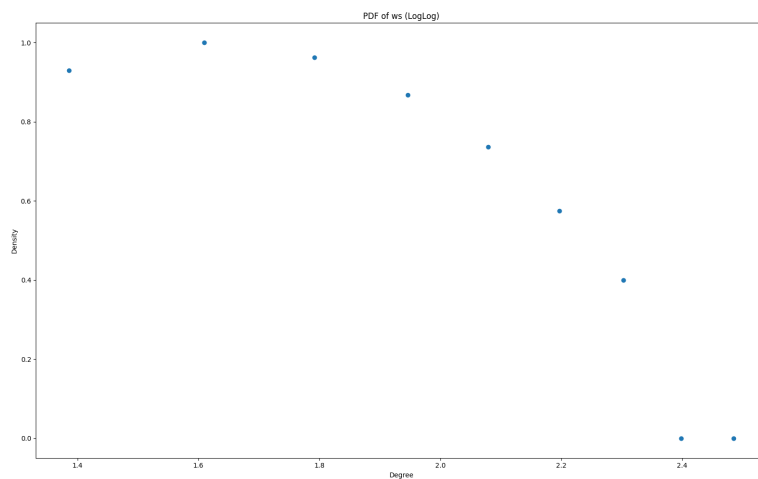


Рис. 23: Функция плотности распределения (LogLog)

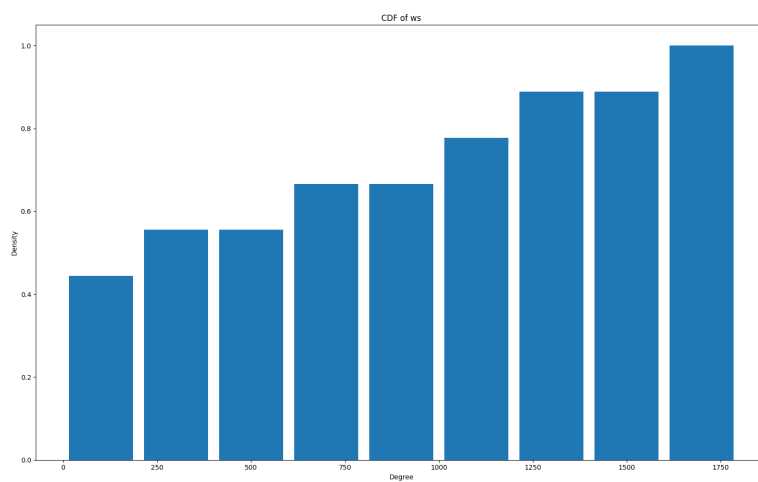


Рис. 24: Кумулятивная функция распределения степеней

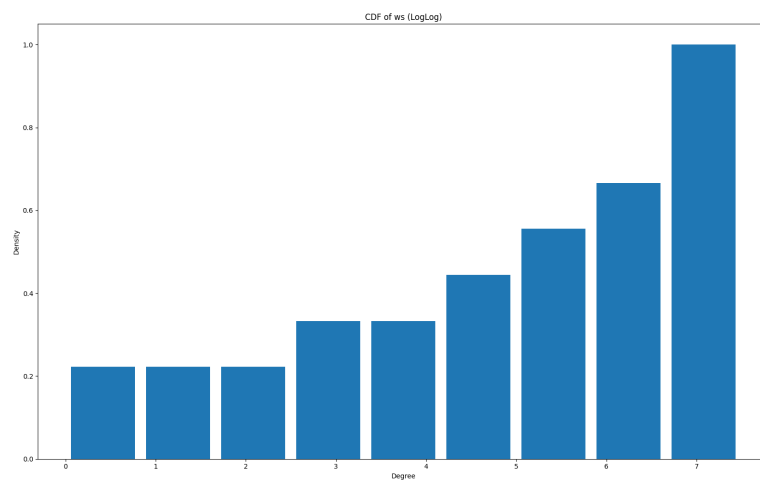


Рис. 25: Кумулятивная функция распределения степеней (LogLog)

3 Метрики узлов графа-друзей *vk*

В данной части работы использовалась наибольшая компонента связности графа друзей *vk*.
Были рассчитаны следующие метрики, а также приведены их результаты:

1. Метрика центральности по степени

- Александр Савченко : 0.5484
- Александр Папернюк : 0.5161
- Marika Oja : 0.4839
- Анастасия Жадан : 0.4516
- Инга Прохорова : 0.3548
- Дмитрий Гирдюк : 0.3226
- Соня Писарева : 0.3226
- Анастасия Раевская : 0.2903
- Артур Саакян : 0.2903
- Александр Кулаков : 0.2903

2. Метрика центральности по близости

- Александр Савченко : 0.6327
- Александр Папернюк : 0.6200
- Анастасия Жадан : 0.5849
- Анастасия Раевская : 0.5741
- Marika Oja : 0.5636
- Антон Гавриков : 0.5536
- Саша Крылатов : 0.5345
- Ефим Жолобов : 0.5345
- Дмитрий Гирдюк : 0.5254
- Александр Кулаков : 0.5167

3. Метрика центральности по посредничеству

- Антон Гавриков : 0.2146
- Александр Савченко : 0.1482
- Анастасия Раевская : 0.1393
- Александр Папернюк : 0.1240
- Дмитрий Гирдюк : 0.1173
- Анастасия Жадан : 0.0668
- Саша Крылатов : 0.0619
- Артур Саакян : 0.0567
- Marika Oja : 0.0517
- Ефим Жолобов : 0.0468

4. Метрика центральности по собственному вектору

- Александр Савченко : 0.3564
- Александр Папернюк : 0.3411
- Marika Oja : 0.3346
- Анастасия Жадан : 0.3207
- Инга Прохорова : 0.2765
- Александр Кулаков : 0.2414
- Соня Писарева : 0.2395
- Диана Евтина : 0.2239
- Святослав Ковалев : 0.2099
- Няшка Веселенькая : 0.2093

5. Метрика *Decay Centrality* ($\delta=0.2$)

- Александр Савченко : 3.8320
- Александр Папернюк : 3.6720
- Marika Oja : 3.3840
- Анастасия Жадан : 3.3200
- Анастасия Раевская : 2.6480
- Дмитрий Гирдюк : 2.6160
- Инга Прохорова : 2.5904
- Антон Гавриков : 2.5840
- Соня Писарева : 2.5840
- Александр Кулаков : 2.4560

6. Метрика *Decay Centrality* ($\delta=0.5$)

- Александр Савченко : 11.5000
- Александр Папернюк : 11.2500
- Анастасия Жадан : 10.6250
- Marika Oja : 10.5000
- Анастасия Раевская : 9.8750
- Антон Гавриков : 9.6250
- Дмитрий Гирдюк : 9.3750
- Соня Писарева : 9.2500
- Александр Кулаков : 9.1250
- Ефим Жолобов : 9.1250

7. Метрика *Decay Centrality* ($\delta=0.8$)

- Александр Савченко : 22.0480
- Александр Папернюк : 21.8880
- Анастасия Жадан : 21.4400
- Marika Oja : 21.2160
- Анастасия Раевская : 21.1520

- Антон Гавриков : 20.8960
- Ефим Жолобов : 20.5760
- Саша Крылатов : 20.5440
- Дмитрий Гирдюк : 20.5440
- Соня Писарева : 20.4160

8. Метрика PageRank

- Александр Савченко : 0.0614
- Александр Папернюк : 0.0580
- Marika Oja : 0.0535
- Анастасия Жадан : 0.0506
- Дмитрий Гирдюк : 0.0430
- Антон Гавриков : 0.0412
- Артур Саакян : 0.0398
- Инга Прохорова : 0.0398
- Анастасия Раевская : 0.0374
- Соня Писарева : 0.0372

9. Метрика *hub* алгоритма *HITS*

- Александр Савченко : 0.0797
- Александр Папернюк : 0.0762
- Marika Oja : 0.0748
- Анастасия Жадан : 0.0717
- Инга Прохорова : 0.0618
- Александр Кулаков : 0.0539
- Соня Писарева : 0.0535
- Диана Евтина : 0.0501
- Святослав Ковалев : 0.0469
- Няшка Веселенькая : 0.0468

10. Метрика *authorities* алгоритма *HITS*

- Александр Савченко : 0.0797
- Александр Папернюк : 0.0762
- Marika Oja : 0.0748
- Анастасия Жадан : 0.0717
- Инга Прохорова : 0.0618
- Александр Кулаков : 0.0539
- Соня Писарева : 0.0535
- Диана Евтина : 0.0501
- Святослав Ковалев : 0.0469
- Няшка Веселенькая : 0.0468

Визуализация данных метрик представлена ниже:

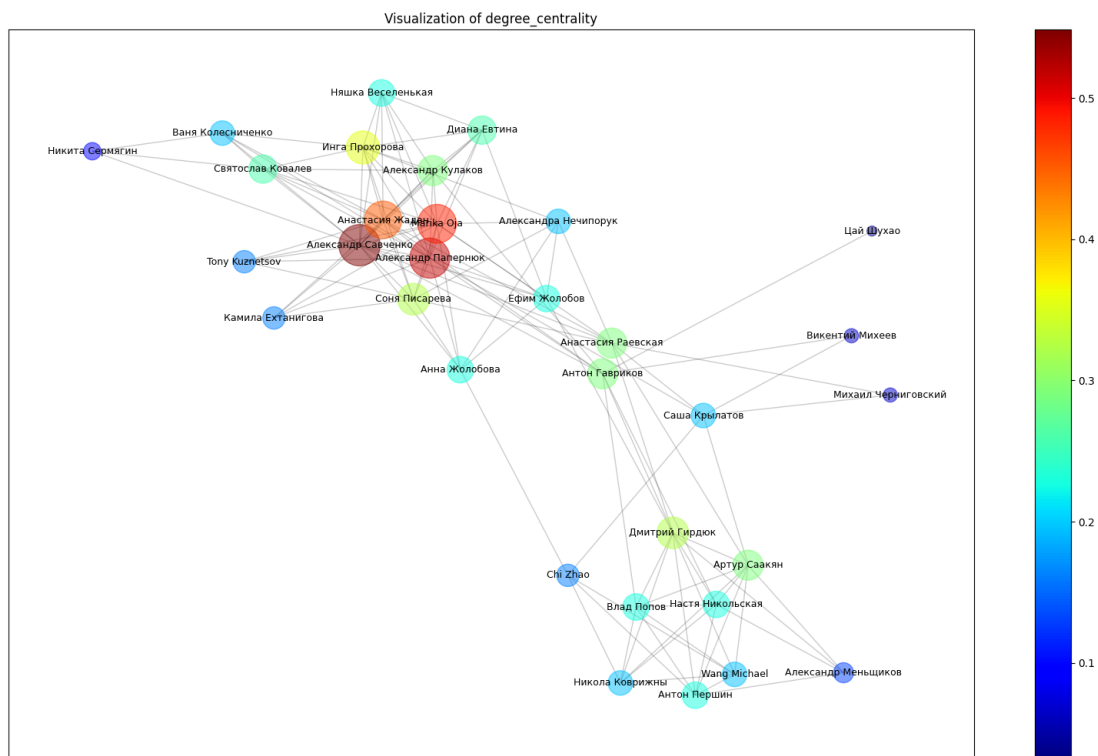


Рис. 26: Центральность по степени

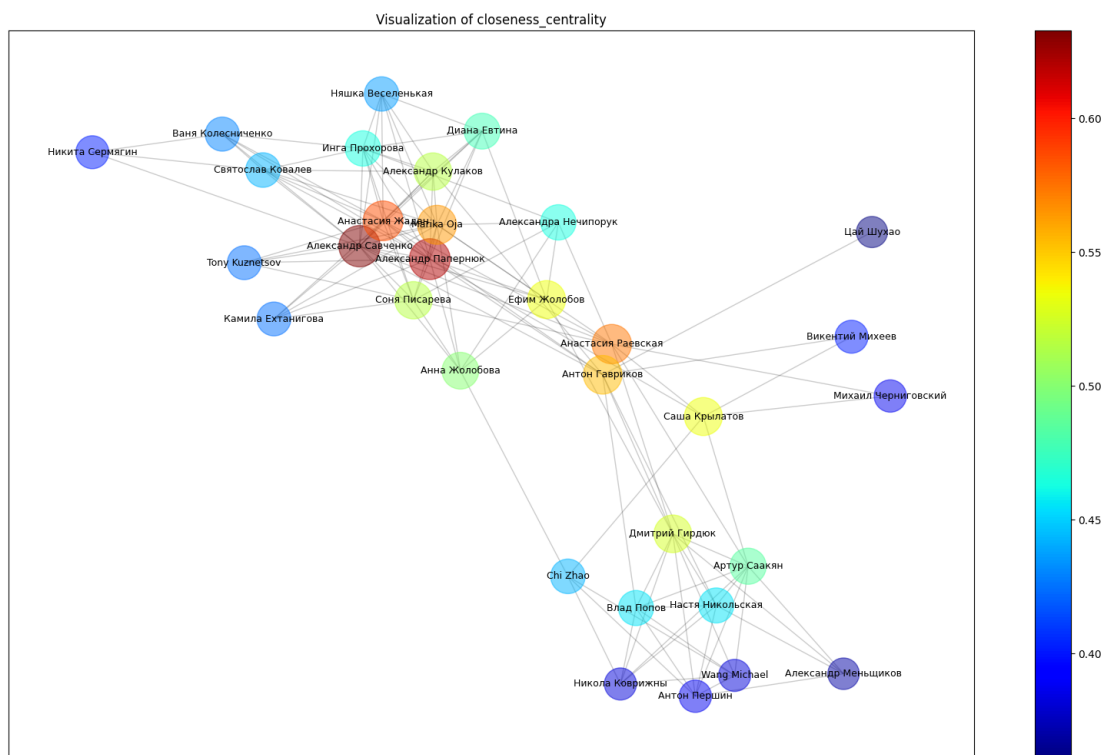


Рис. 27: Центральность по близости

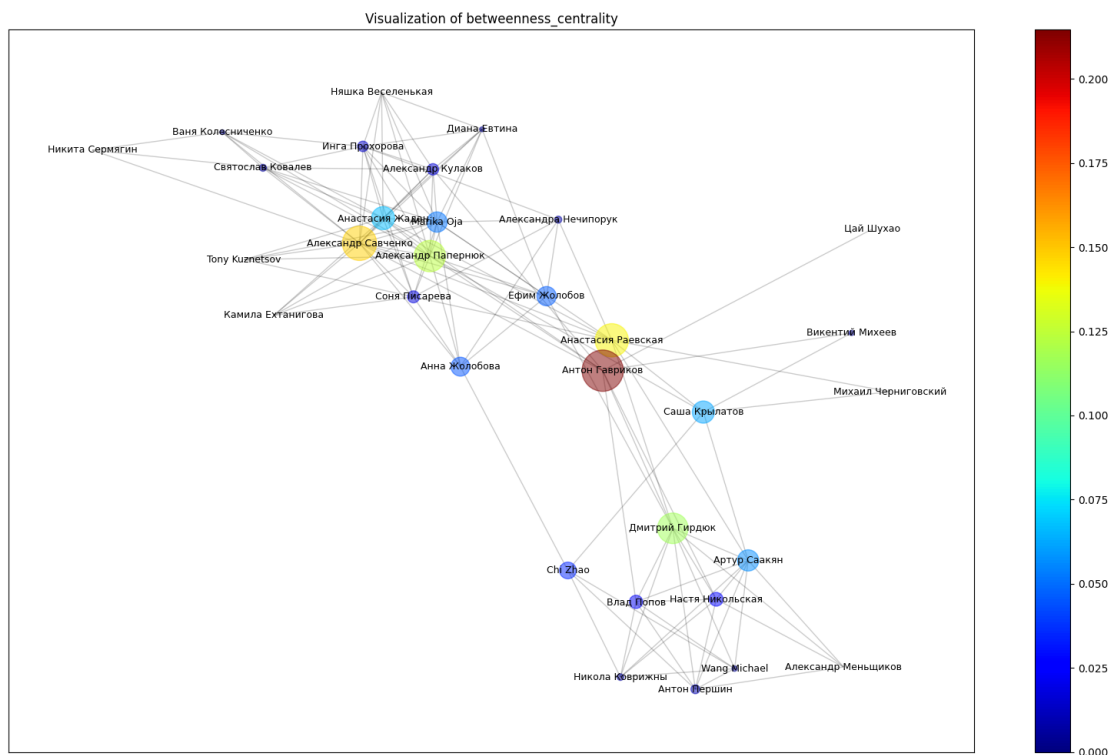


Рис. 28: Центральность по посредничеству

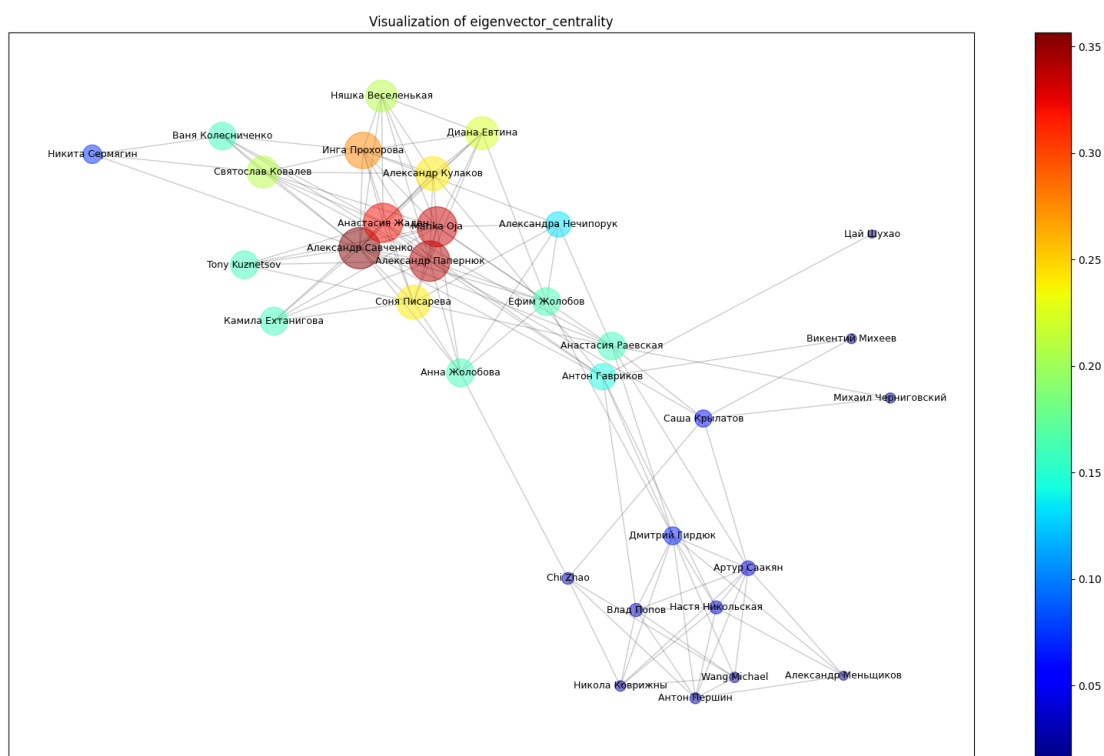


Рис. 29: Центральность по собственному вектору

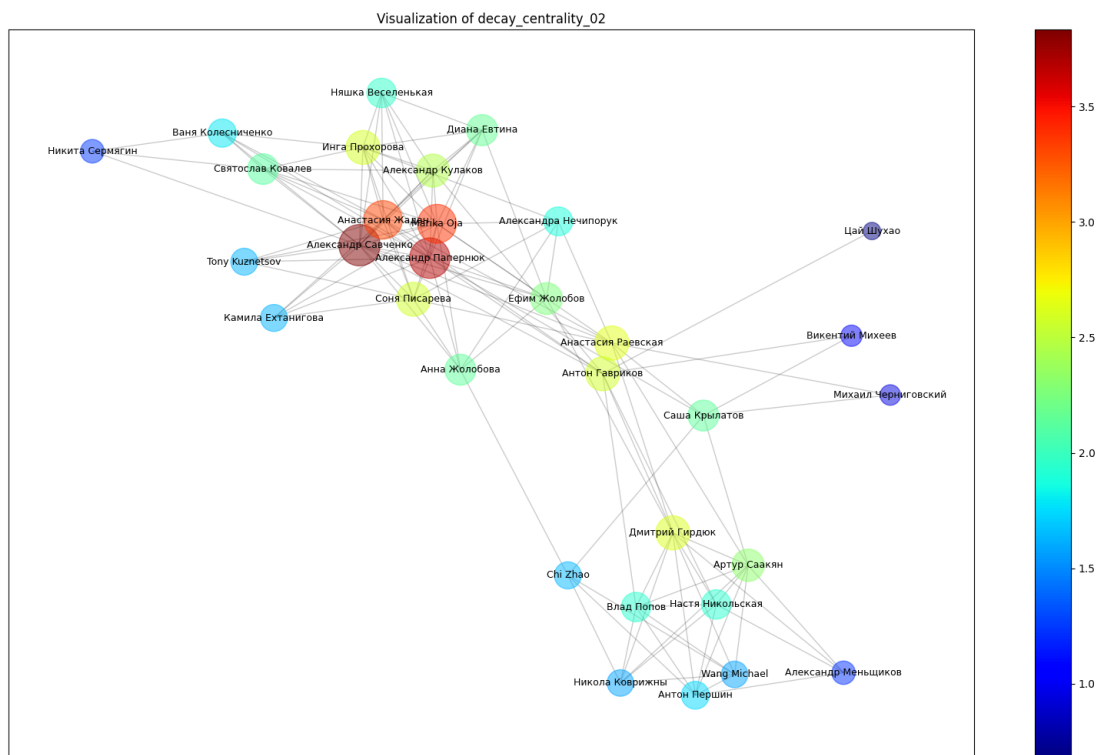


Рис. 30: Decay Centrality ($\delta=0.2$)

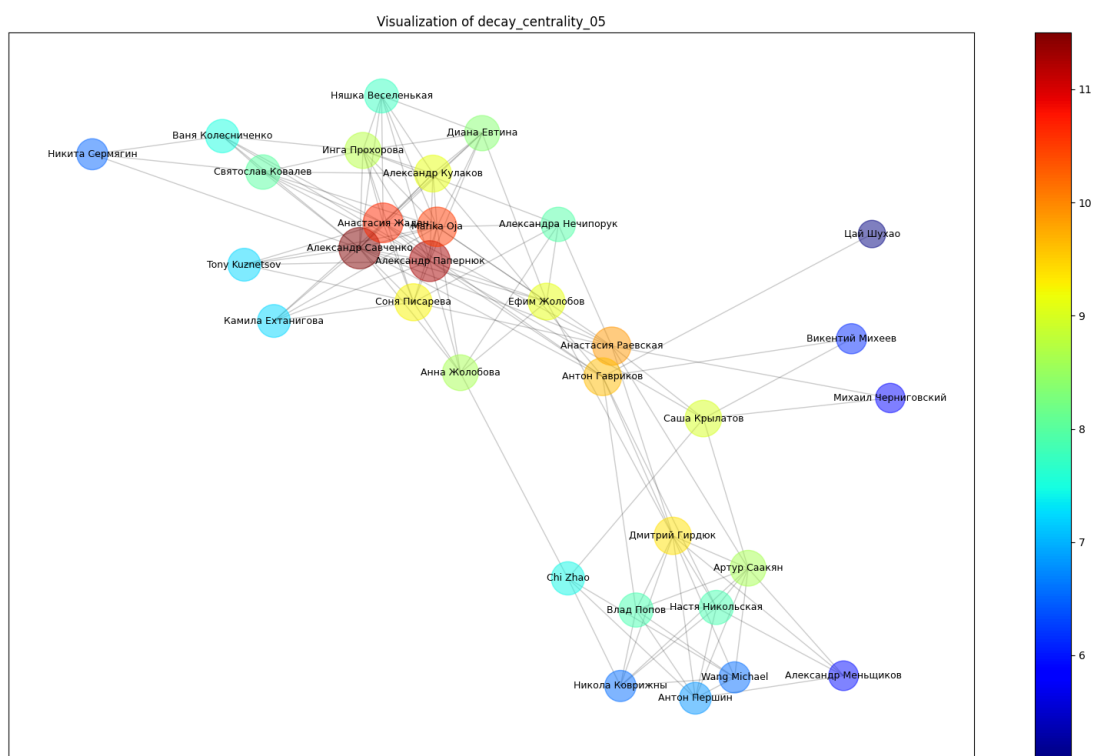


Рис. 31: Decay Centrality ($\delta=0.5$)

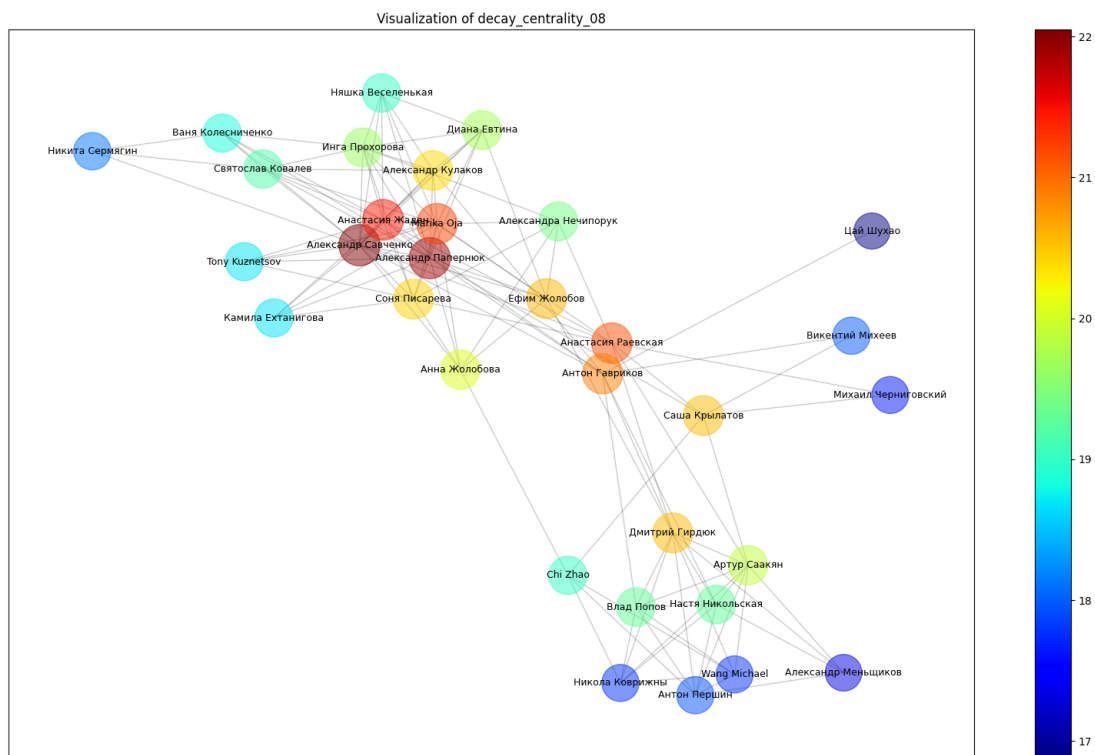


Рис. 32: Decay Centrality ($\delta=0.8$)

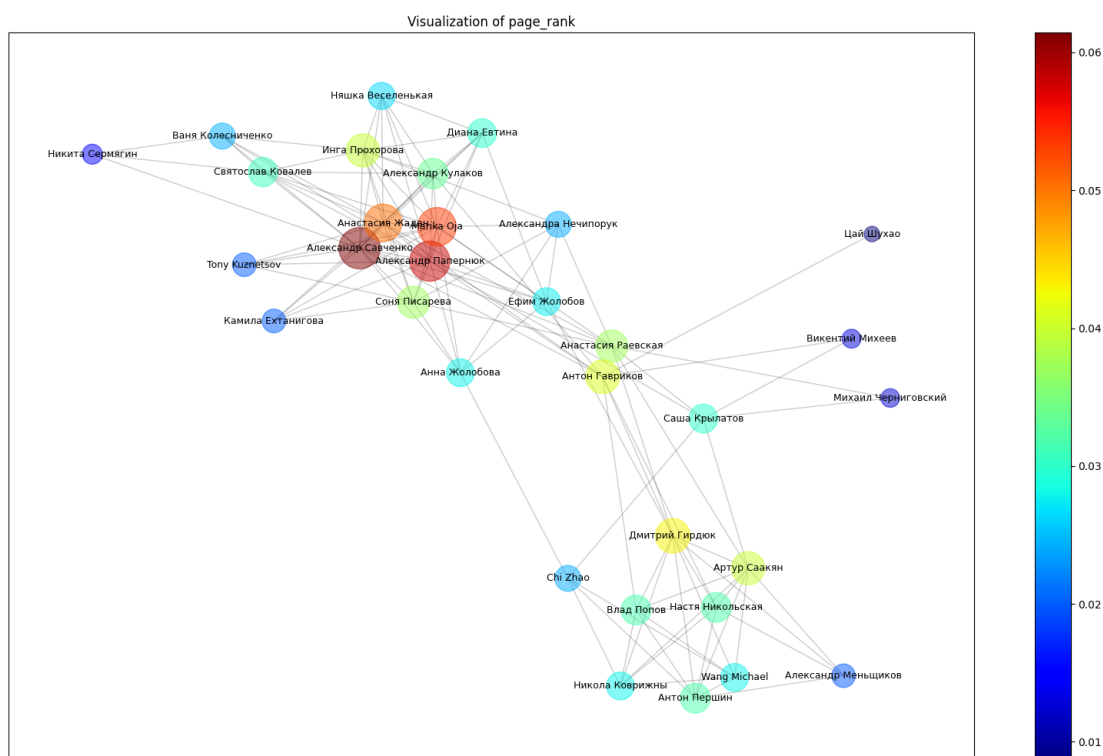


Рис. 33: PageRank

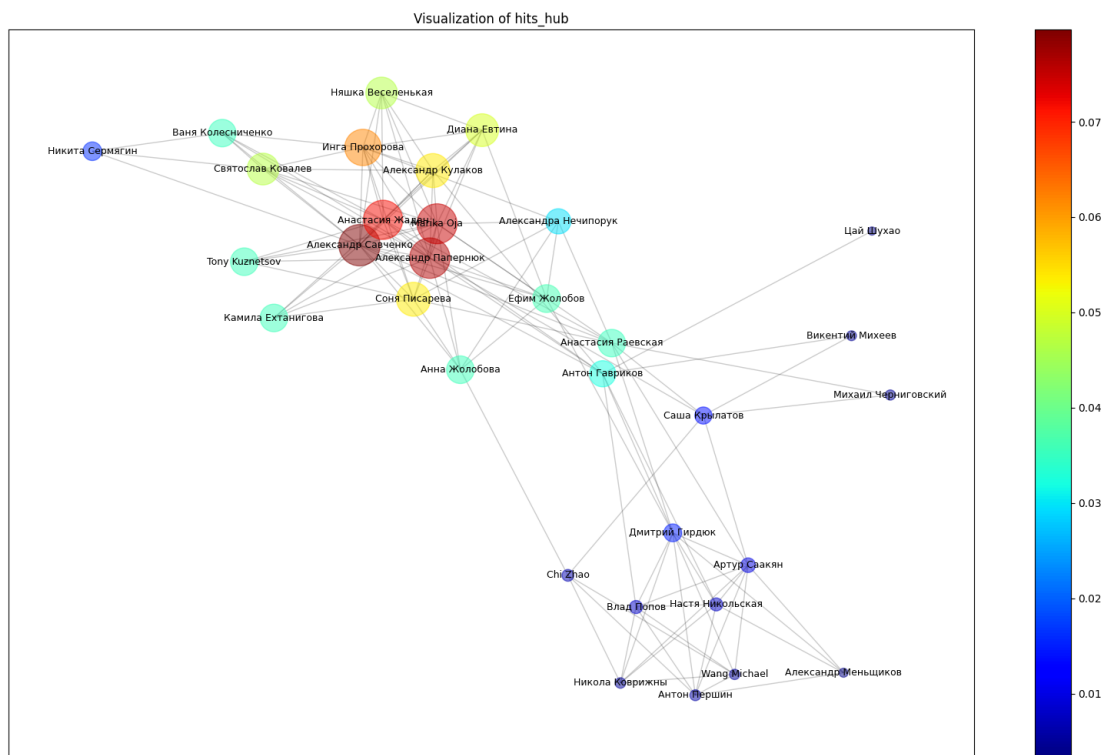


Рис. 34: Метрика Hub

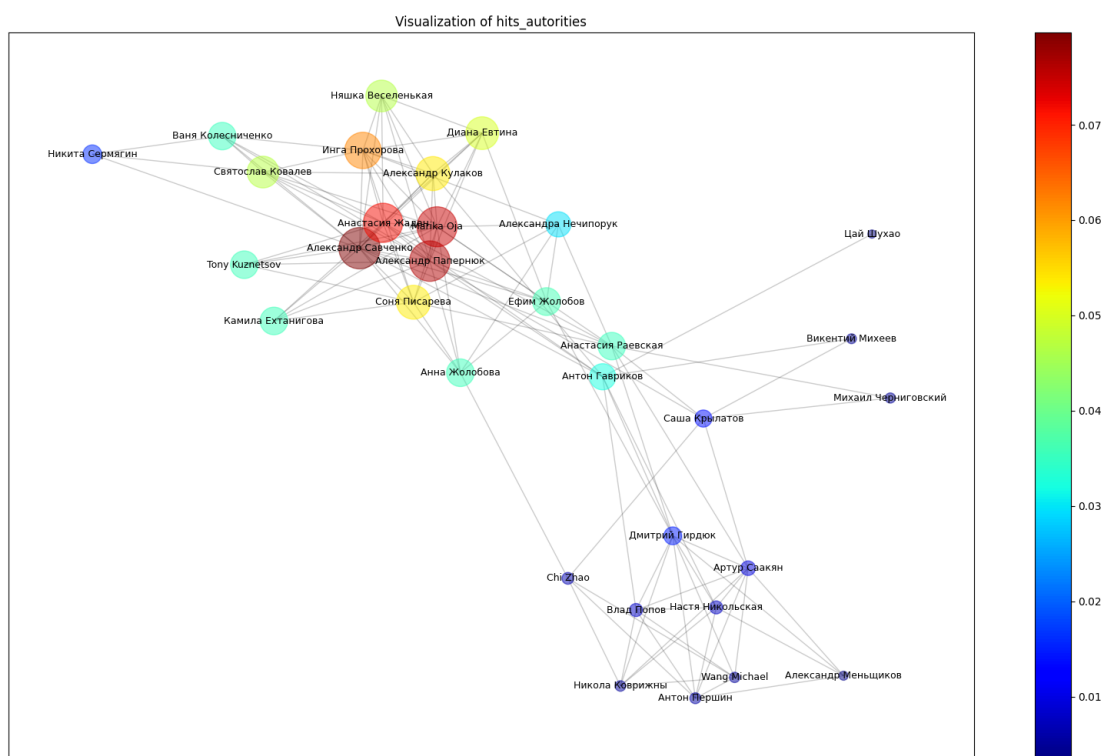


Рис. 35: Метрика Authorities

Несмотря на неявное выделение двух сообществ в данном подграфе, наибольшие значения почти всех метрик имеют одни и те же узлы. Это связано как раз с тем, что одно из двух сообществ плотнее (имеет внутри себя больше связей), чем оставшееся. Также данное почти полное совпадение связано с тем, что подграф мал – 10 узлов с наибольшими значениями мет-

рик составляют примерно 30% от вершин в подграфе, что повышает вероятность попадания в выборку. Так как подграф является неориентированным, то метрики *hub* и *authorities* совпадают.