

Моя это телеграм сеть

Выполнила
Мария Сухарева

Данные

- **Данные:** несколько моих групповых чатов Telegram, собраны через API (Telethon) в CSV
- **Разметка чатов:** вручную по типам (семья, друзья, учёба, работа, школа и т.п.)
- **Ноды:** уникальные пользователи (кроме меня), состоящие хотя бы в одном выбранном чате
- **Рёбра:** общие чаты; вес ребра = число общих чатов
- **Тип сети:** ненаправленная, взвешенная, однородная (только пользователи и связи “состоят в одном чате”)
- **Атрибуты:**
 - у нод: `user_id`, имя, `username`, типы чатов
 - у рёбер: `weight` (сколько общих чатов)

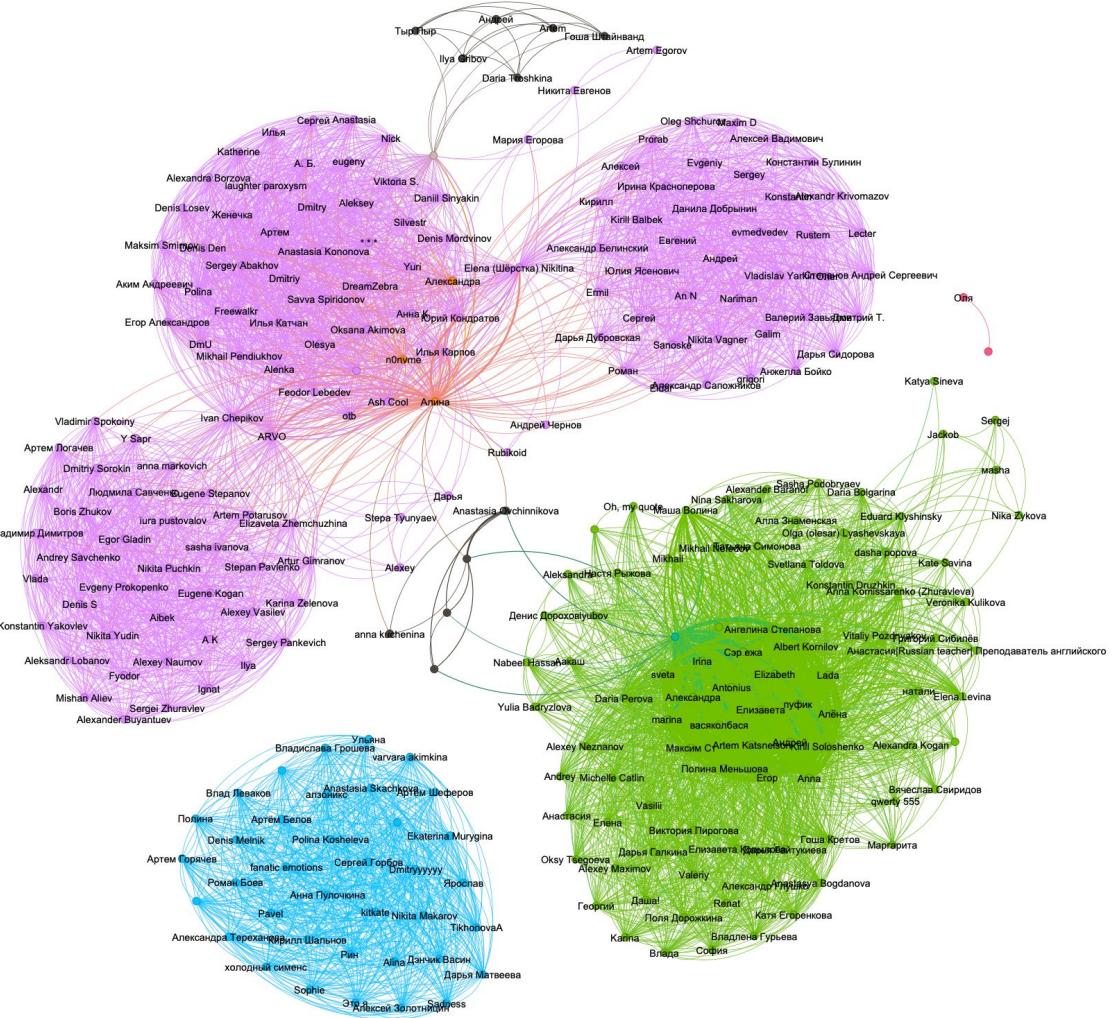
	<code>user_id</code>	<code>name</code>	<code>username</code>	<code>group_idx</code>	<code>group_tg_id</code>	<code>group_name</code>	<code>class</code>
0	1293695386	❤️	irinabarabash2	0	-787553900	Тефачки	семья
1	948057117	Оля	ishk_yapts	0	-787553900	Тефачки	семья
2	1354905725	Будущий 💚	Halcyon999	1	-1002891225908	Alt-right party	друзья
3	469145391	Алиса Комрад 🤝	wonderland_13	1	-1002891225908	Alt-right party	друзья
4	1125600016	Салимат 🌸	Zaydieva	1	-1002891225908	Alt-right party	друзья

Граф

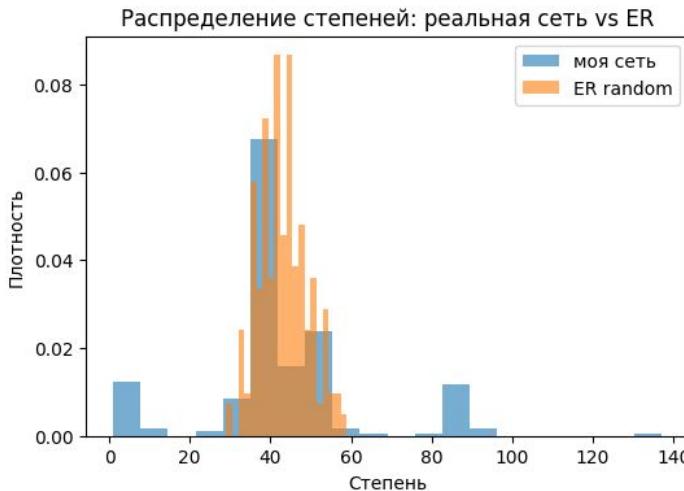
Количество нод: **277**

Количество рёбер: **5947**

цвет ноды – тип группы
(друзья / работа / учёба /
школа / семья)

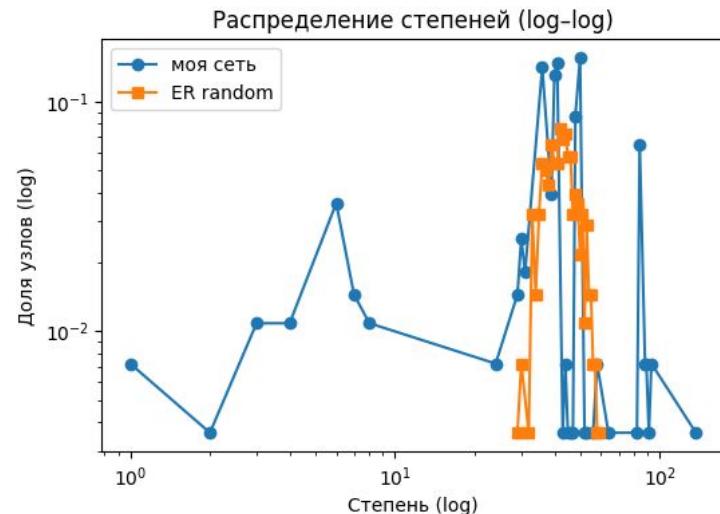


Распределение степеней вершин



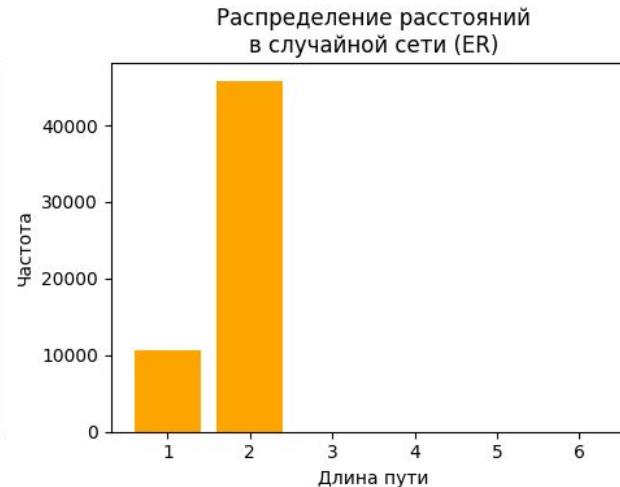
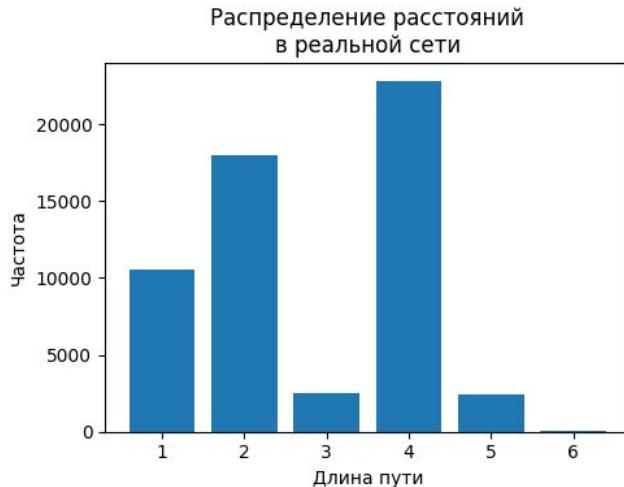
Хвост всё-таки тяжелее, чем у ER-графа: в сети есть несколько узлов с аномально высокой степенью, но это единичные центральные участники, а не системный степенной хвост

Распределение степеней не степенное: на log-log графике нет прямой линии, поэтому сеть не является классической scale-free



Средняя длина пути

- Большинство пар вершин соединяются очень короткими путями (2–4 шага) → сеть ведёт себя как small-world
- В сравнении с ER-графом с теми же N и E пути лишь немного длиннее, что связано с наличием устойчивых сообществ и кластерной структуры, отсутствующей в полностью случайной модели



Сравнение с ER, BA, WS

	Вершины	Ребра	Средняя степень	Максимальная степень	Средний коэф. кластеризации	Средняя длина пути
REAL	277	5947	42.94	137	0.94	2.80
ER	277	5947	42.94	59	0.16	1.85
BA	277	5376	38.82	116	0.23	1.87
WS	277	6094	44.00	52	0.55	1.97

- **ER:** кластеризация и max degree сильно ниже → в моей сети гораздо сильнее выражены группы и отдельные очень центральные узлы
- **BA:** как и в BA, есть вершины с очень большой степенью, но у BA кластеризация намного ниже, тогда как у меня связи собраны в плотные чаты-клики
- **WS:** по коротким путям и высокой кластеризации одновременно сеть ближе всего к WS, но она ещё более small-world из-за групповых чатов

Centrality measures

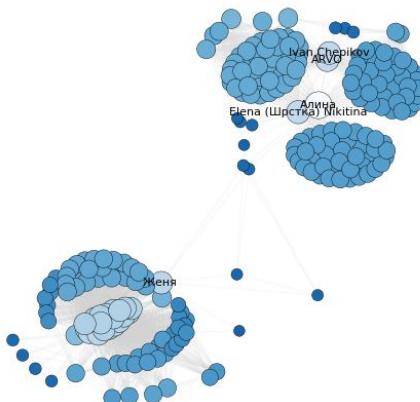
Degree: в топе в основном коллеги – они состоят сразу в нескольких самых больших рабочих чатах, поэтому у них больше всего связей

Betweenness / Closeness: мосты между рабочими, учебными и дружескими кругами, через них проходят многие кратчайшие пути

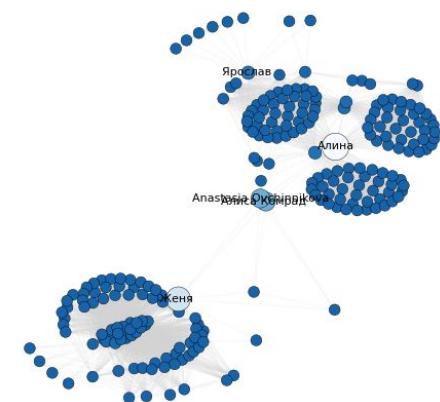
Eigenvector: ядро учебного кластера (несколько общих проектных и предметных чатов), поэтому они связаны с другими важными узлами

После 5-го места у многих людей очень похожие значения центральностей, потому что они состоят в почти одинаковых наборах чатов

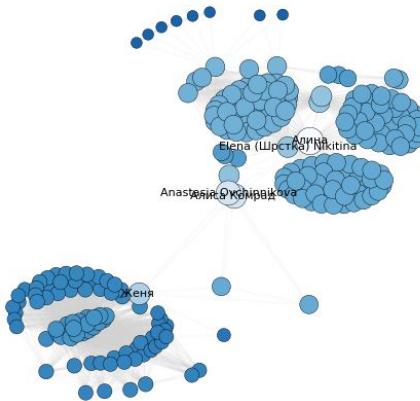
Degree centrality



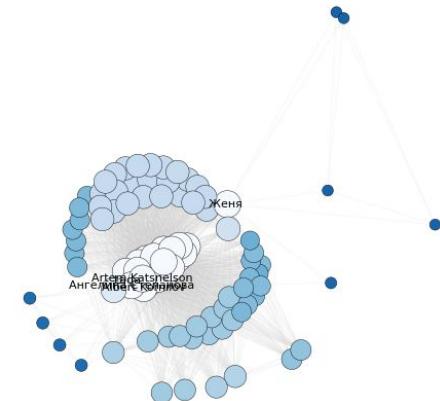
Betweenness centrality



Closeness centrality

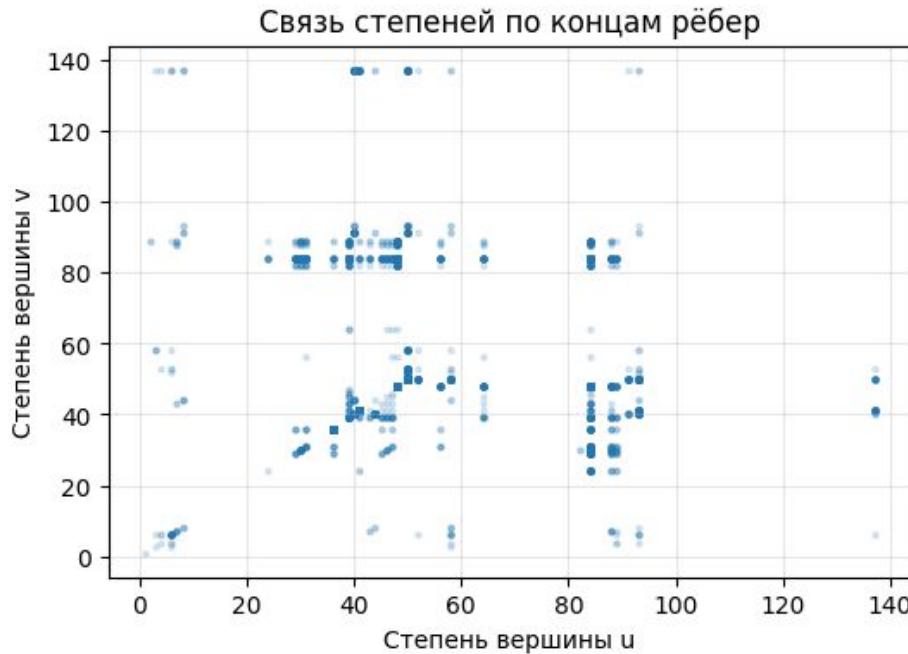


Eigenvector centrality

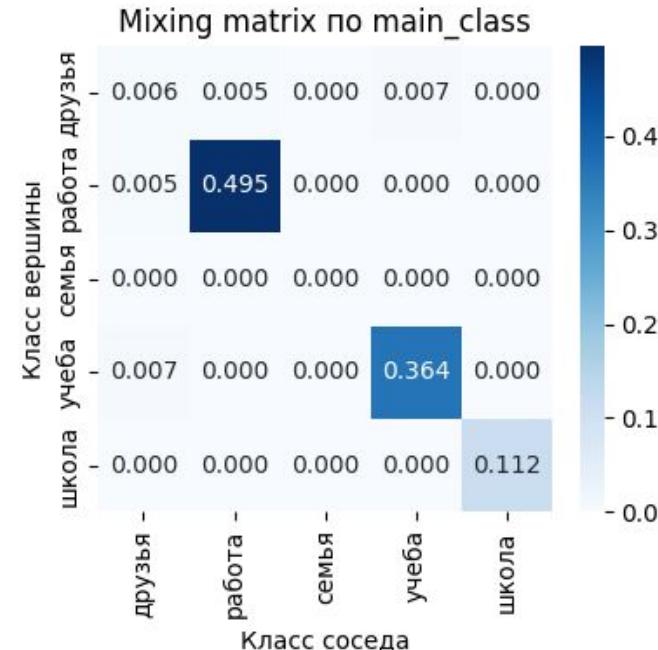


Смешивание по степени и по атрибутам

Коэффициент асортативности по степени: **0.0053**

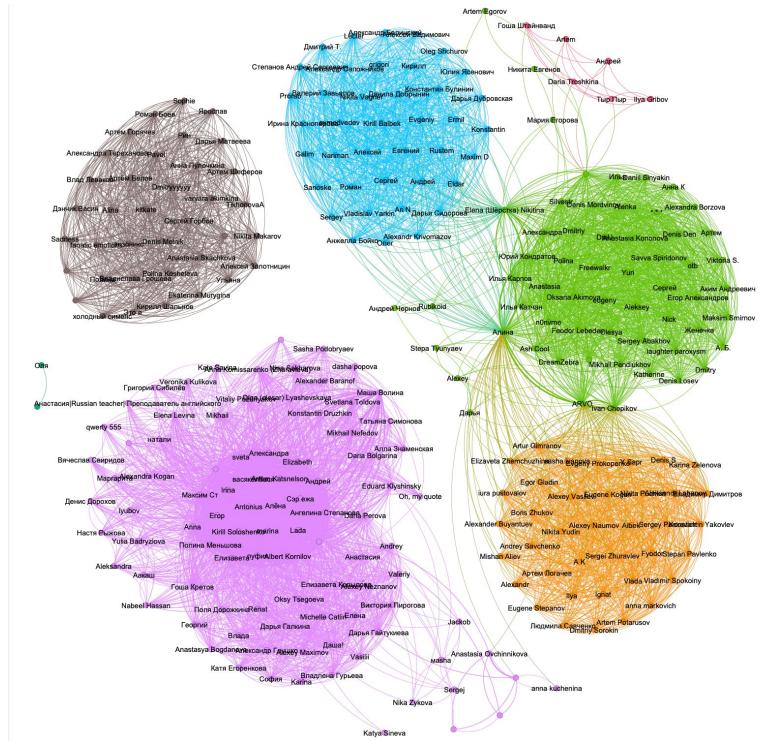


Асортативность по типу группы: **0.9608**

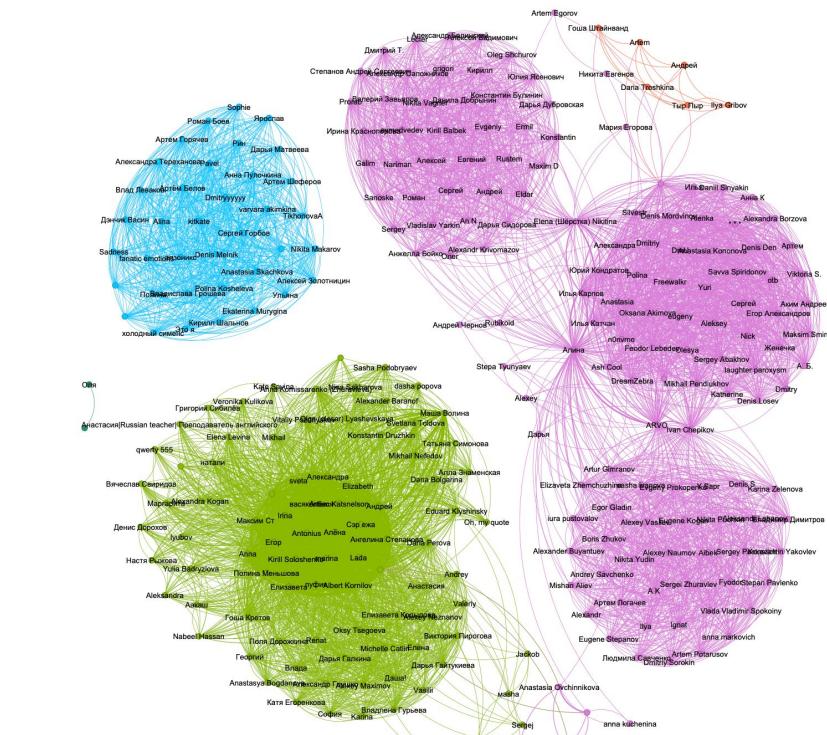


Сообщества

Greedy: сообществ = 7, modularity = **0.6269**

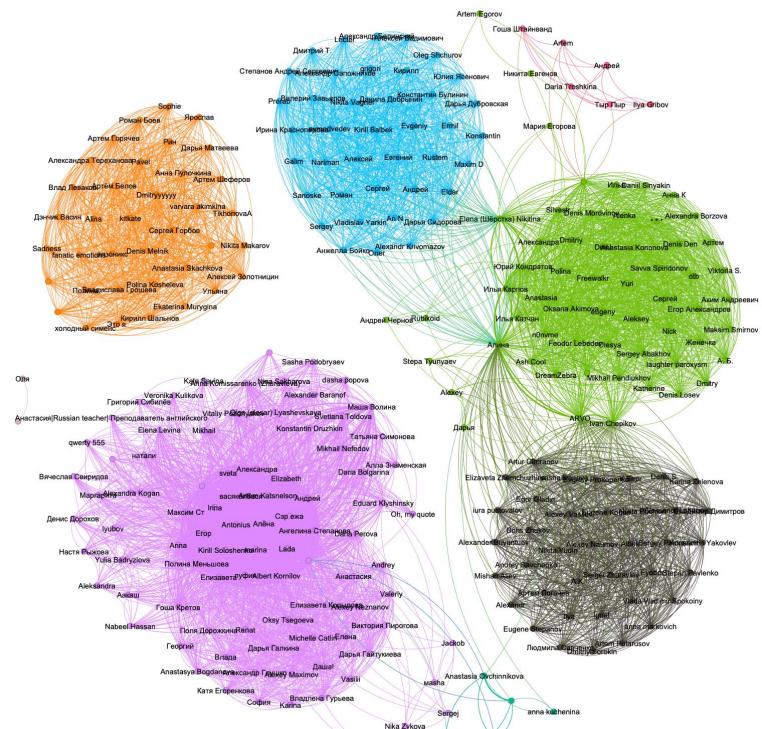


Label propagation: сообществ = 5, modularity = **0.5615**



Сообщества

Louvain: сообществ = 8, modularity = 0.6287



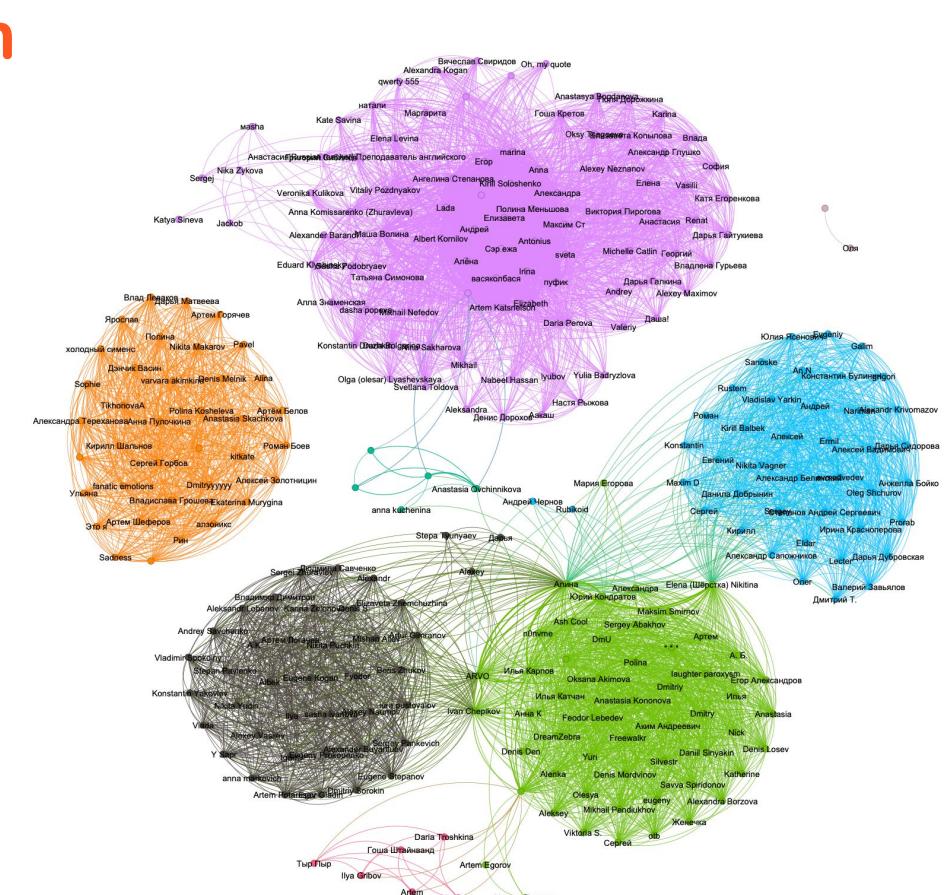
- **3 алгоритма: Greedy, Louvain**
(оптимизируют модулярность) и **Label propagation** (распространение меток)
- У всех алгоритмов модулярность высокая → в сети чётко выражены сообщества
- **Louvain** даёт наилучший результат и почти совпадает с **Greedy**; **Label propagation** даёт меньше сообществ и заметно более низкую модулярность

Сообщества: ground truth

Ground truth-сообщества задала вручную по смыслу чатов: 3 рабочих (основной + 2 проекта), 2 группы друзей, школа, учёба, семья – всего **8** кластеров.

По NMI лучше всего их восстанавливает **Louvain** (чуть хуже **Greedy**), а **Label propagation** часто склеивает соседние сообщества в один крупный кластер.

NMI (ground truth, Louvain)	0.9585
NMI (ground truth, Greedy)	0.9366
NMI (ground truth, Label prop.)	0.7801



Клики

Clique 1 (size=51)

Ярослава Никитина
Юрий Кондратов
София Смирнова
Анна Смирнова
van Gorpov
Elena (Шреттер) Nikitina

Clique 2 (size=49)

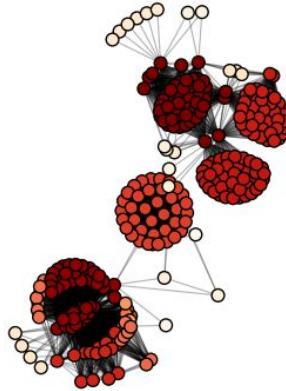
Валерий Григорьев
Анна Смирнова
Ангелина Бородинова

22 максимальные клики

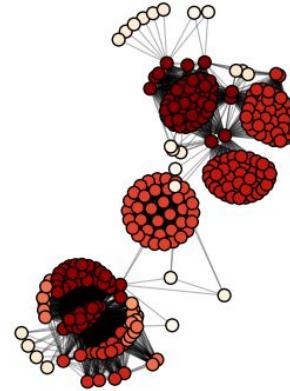
Крупные клики
соответствуют большим
рабочим и учебным чатам

K-core декомпозиция

k-shells on 1-core



k-shells on 2-core

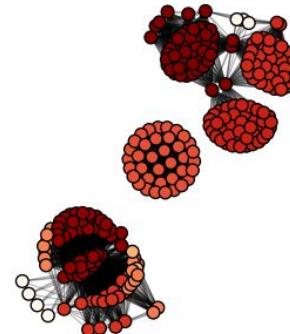


Вершины из глубоких k-cores часто совпадают с узлами с наибольшими центральностями и степенями, то есть это те же ключевые люди сети

K-core декомпозиция показывает структуру ядро-периферия:

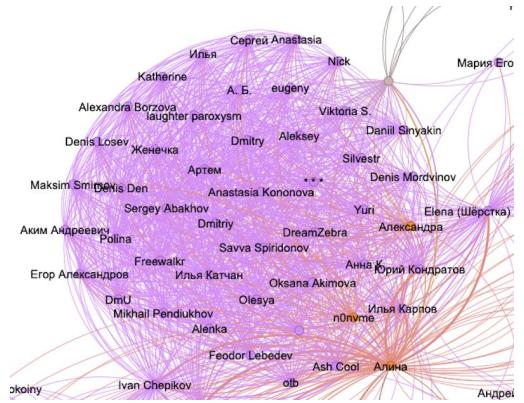
- внешние слои → мелкие и слабо связанные чаты,
- глубокие k-cores → плотное ядро из крупных рабочих и учебных чатов

k-shells on 7-core



k-shells on 8-core





Спасибо за внимание!

