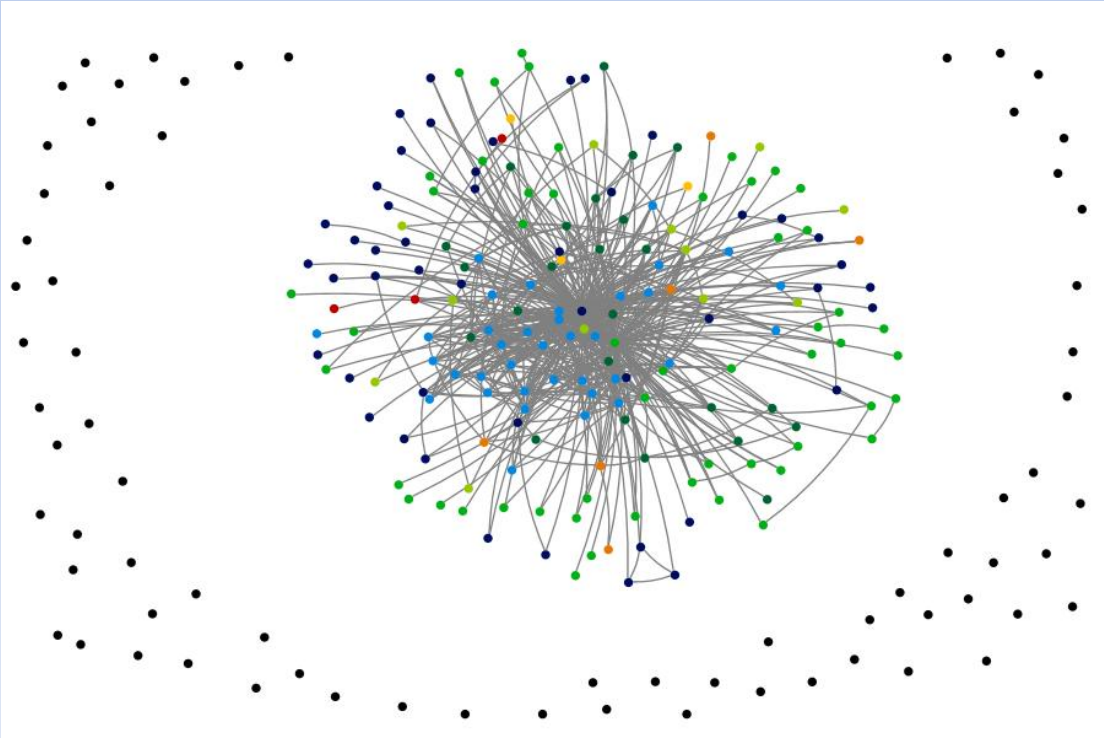
ЛАБОРАТОРИЯ ИНТЕРНЕТ-ИССЛЕДОВАНИЙ
ЛИНИС

ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
САНКТ-ПЕТЕРБУРГ

Сети "дружбы" в онлайновых сообществах социальных движений: влияет ли активность и размеры группы на структуру сети?



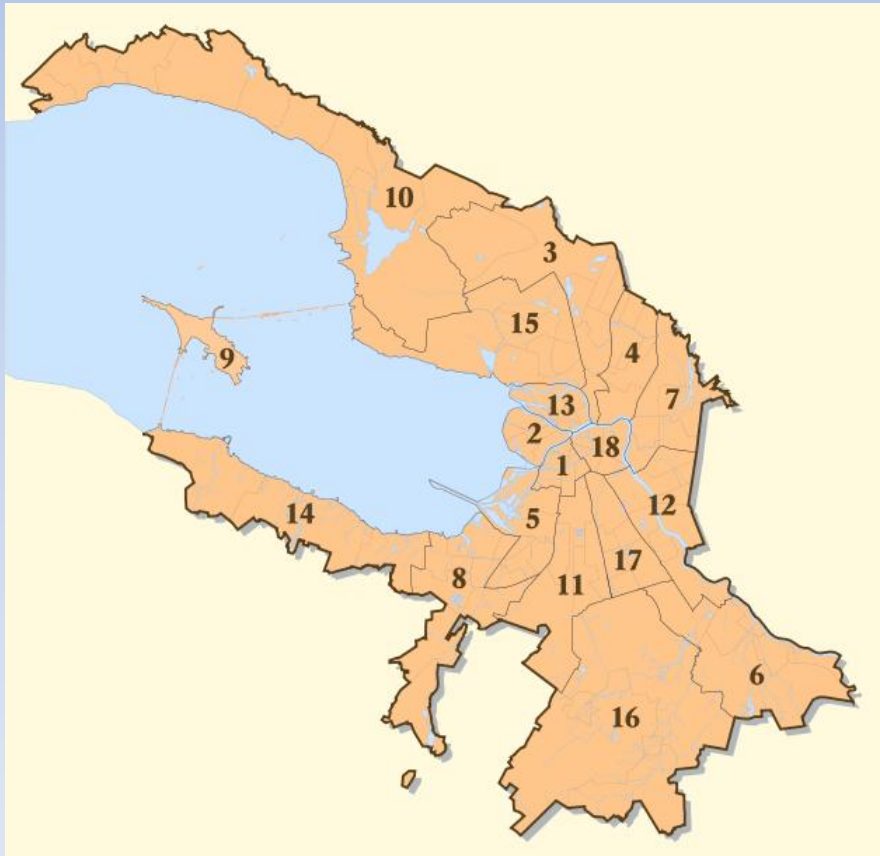
Автор:

Рыков Юрий

Объект

- Онлайновые сообщества в социальной сети «Вконтакте» социального движения «Наблюдатели Петербурга», объединяющие наблюдателей на политических выборах.

- Всего = 17 сообществ (соответствуют 17 административным районам Санкт-Петербурга)
- Выборочная совокупность охватывает все 17 сообществ



Открытые сообщества (для участия не обязательно быть членом движения / наблюдателем)

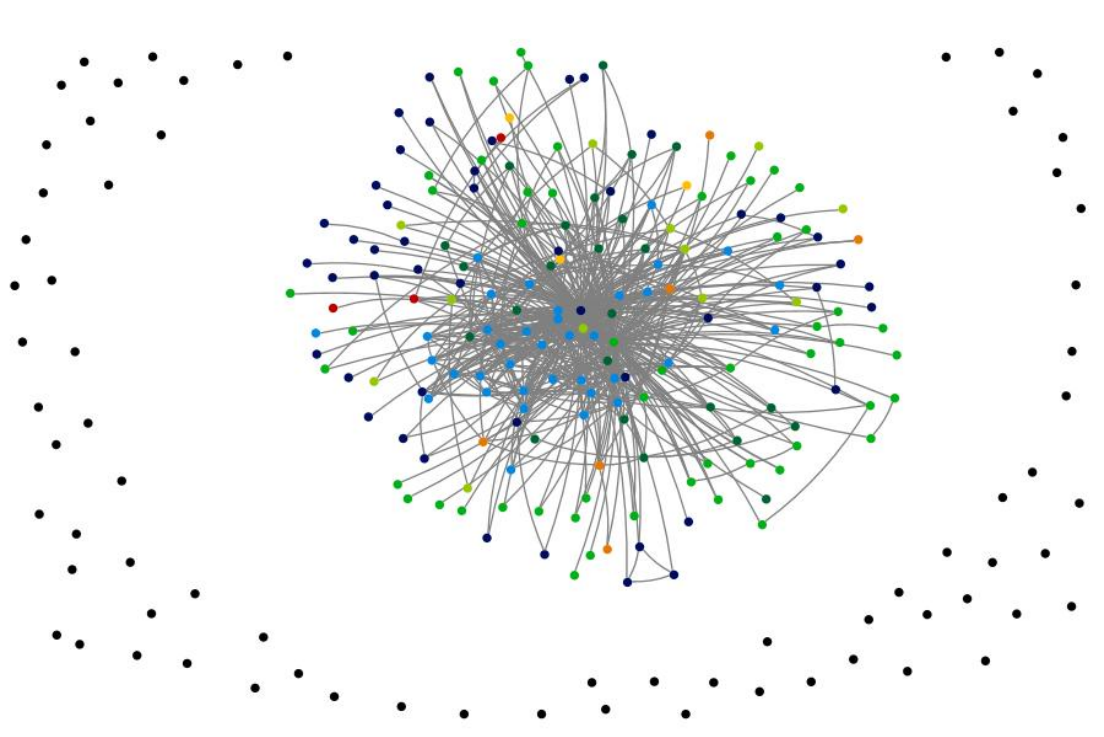
Предмет

- Теоретически интерес восходит к проблеме социальной структуры современного общества: **сети VS сообщества.**
- Членство в сообществе и феномен виртуальной «дружбы» в Интернет-сервисах социальных сетей существуют независимо друг от друга.
- В фокусе исследования - отношения виртуальной «дружбы», связывающие членов онлайнового сообщества в сеть.
- Анализируются сети «дружбы», существующие «внутри» каждого изучаемого сообщества по отдельности (в каждом сообществе существует собственная сеть «дружбы»).

Единицы анализа

- Вершина сети = пользователь (член сообщества).
- Ребро сети = наличие отношения «дружбы» между пользователями.

Пример: сеть «дружбы» сообщества наблюдателей Красносельского района



Задачи исследования

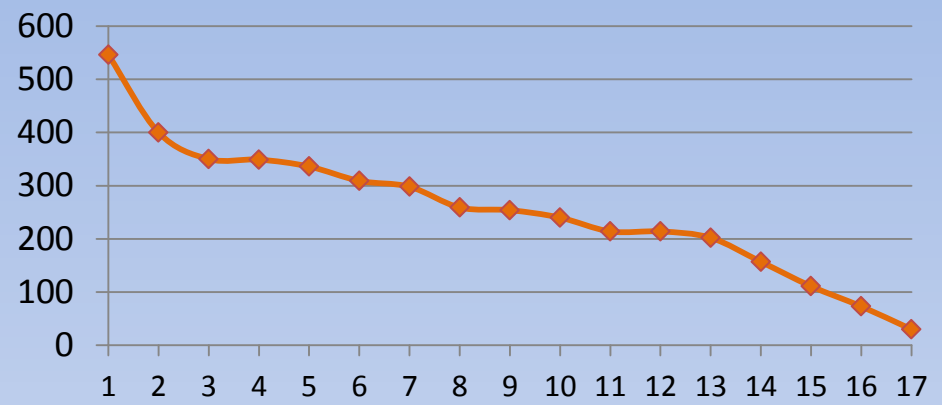
1. Сравнить сети «дружбы», построенные для каждого сообщества, между собой и выявить схожее и различное в их структуре.
2. Выявить зависимость структурных характеристик сетей «дружбы» от коммуникативной активности и от численности сообществ.

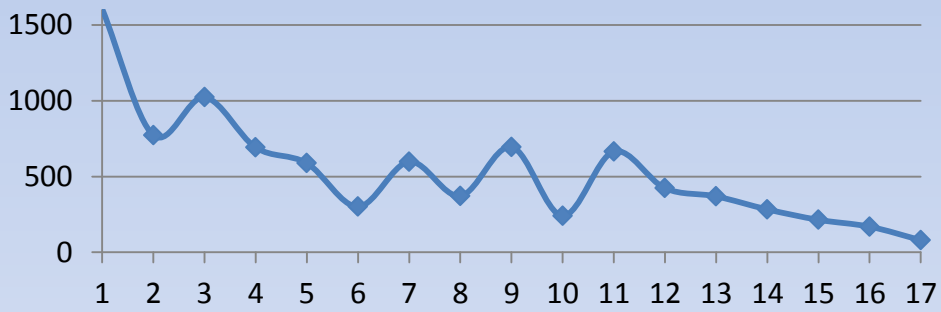
Метод: Social Network Analysis (SNA)

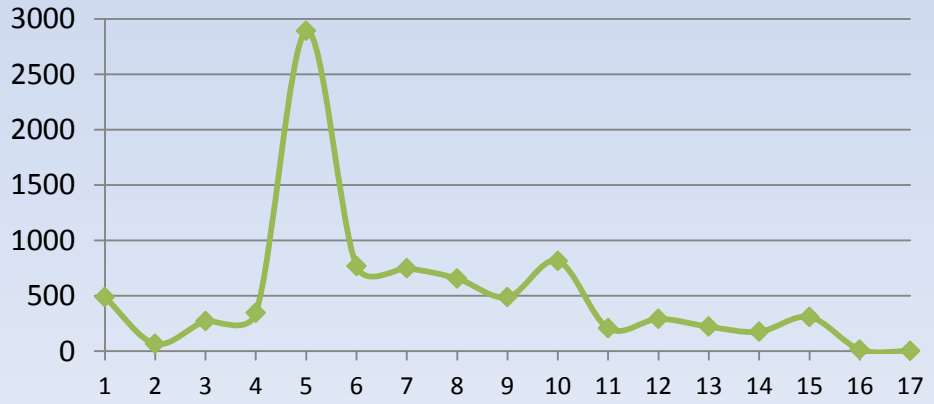
Предварительные данные

Численность и коммуникативная активность сообществ

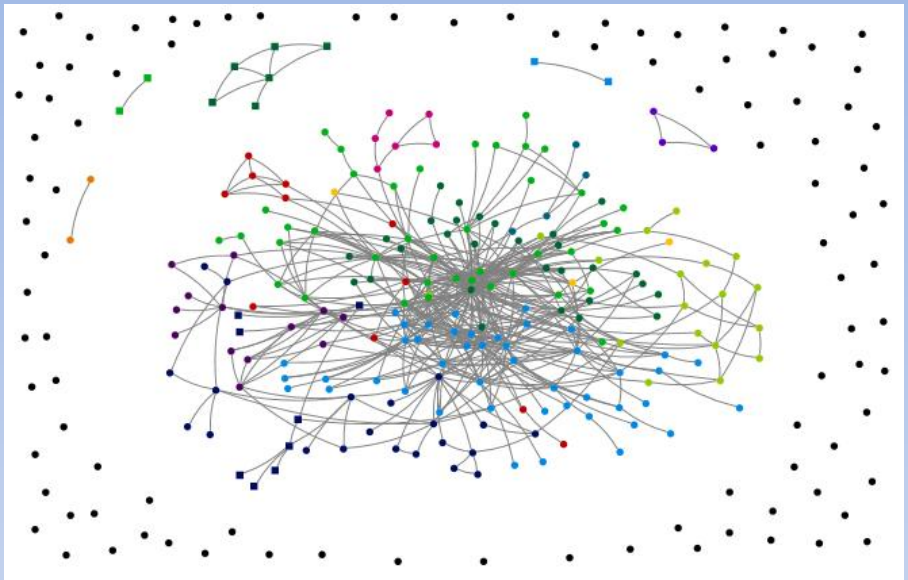
Сообщества по районам	Численность	Посты на стене (всего)	Количество сообщений во всех темах
Пушкин	546	1606	488
Приморский	400	773	67
Васильевский	350	1023	272
Адмиралтейский	349	692	346
Кировский	336	588	2891
Калининский	309	301	769
Невский	298	598	748
Центральный	259	372	656
Красносельский	254	695	488
Выборгский	240	241	814
Петроградский	214	665	207
Московский	214	423	291
Красногвардейский	202	369	221
Фрунзенский	157	283	176
Петродворец	111	214	309
Колпинский	73	169	15
Курортный	30	80	1



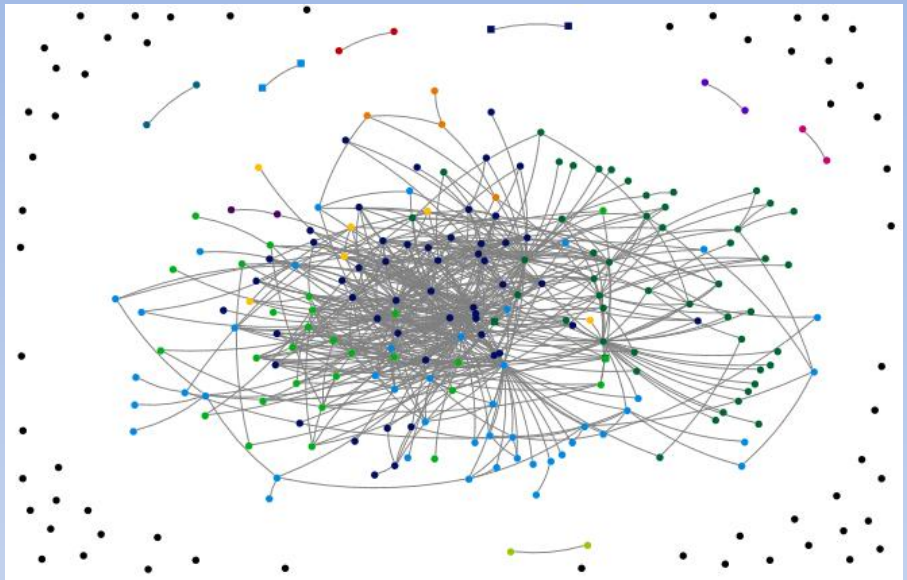




Как сравнить эти сети?



Калининский район



Центральный район

Многоуровневый анализ:

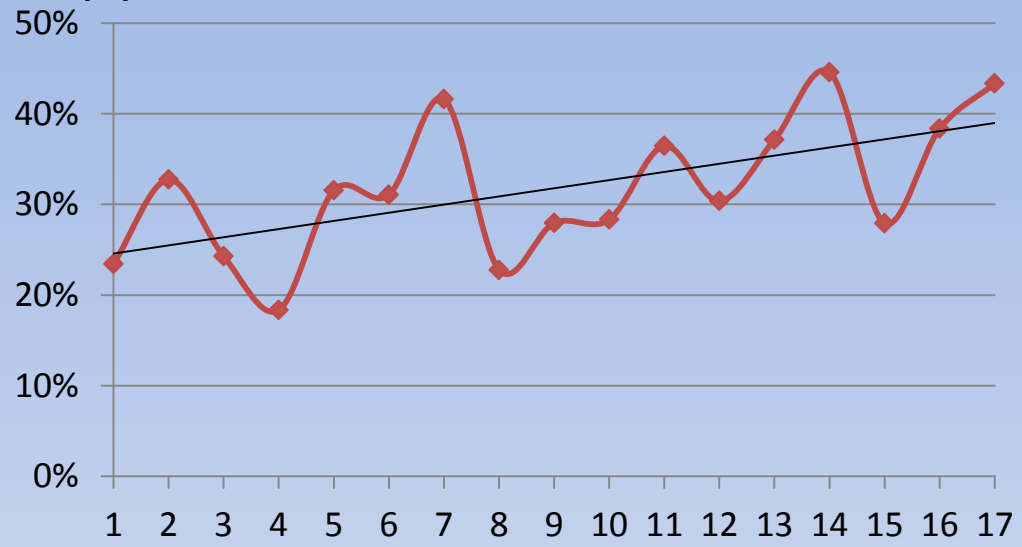
1. Макро структуры сети – соединенные компоненты
2. Мезо структуры сети – кластеры (применение алгоритмов выявления сообществ)
3. Микро элементы сети – отдельные вершины

I уровень анализа: показатели

1. Доля изолянтов / связанных вершин
2. Количество соединенных компонентов
3. Распределение соединенных компонентов по числу вершин и ребер
4. Доля вершин и ребер максимального соединенного компонента от общего числа связанных вершин и ребер
5. Плотность графа (с изолянтами)
6. Диаметр и среднее значение геодезической дистанции
7. Средние значения степени и betweenness centrality в полном графе

I уровень анализа: результаты

1. Доля изолянтов



Размах вариации = 44,6% - 18,3% = 26,3%

Среднее = 31,8%

Стандартное отклонение = 7,6%

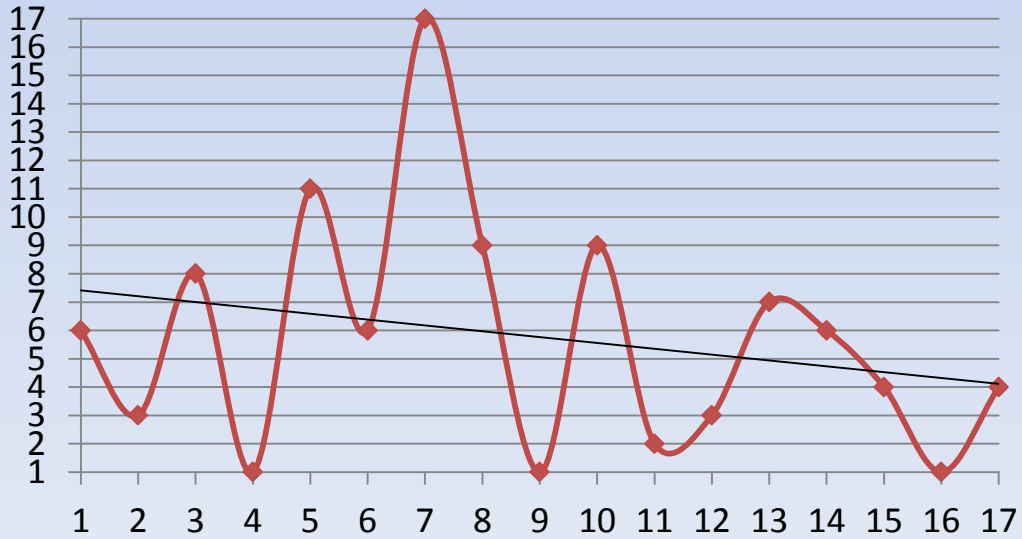
Коэффициент вариации = 24% < 33%, => вариация данных не значительна (совокупность однородна)

Коэффициент детерминации парной регрессии:

R^2 **(число участников)** = 0,329

R^2 **(число постов)** = 0,227

2. Количество соединенных компонентов



Размах вариации = 16

Среднее = 6

Стандартное отклонение = 4,22

Коэффициент вариации = 74% > 33%, => вариация данных значительна (совокупность неоднородна)

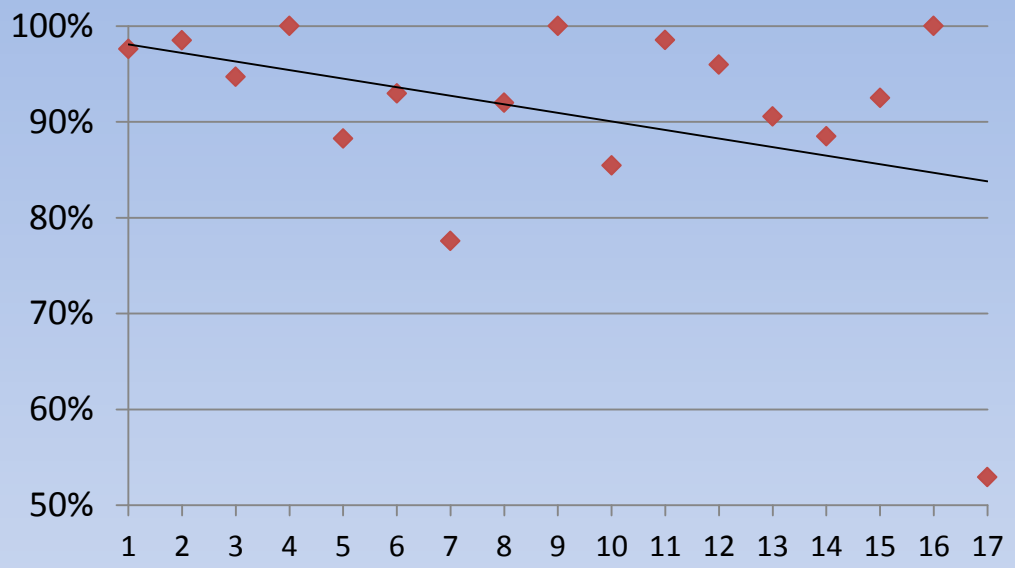
Коэффициент детерминации парной регрессии:

R^2 **(число участников)** = 0,056

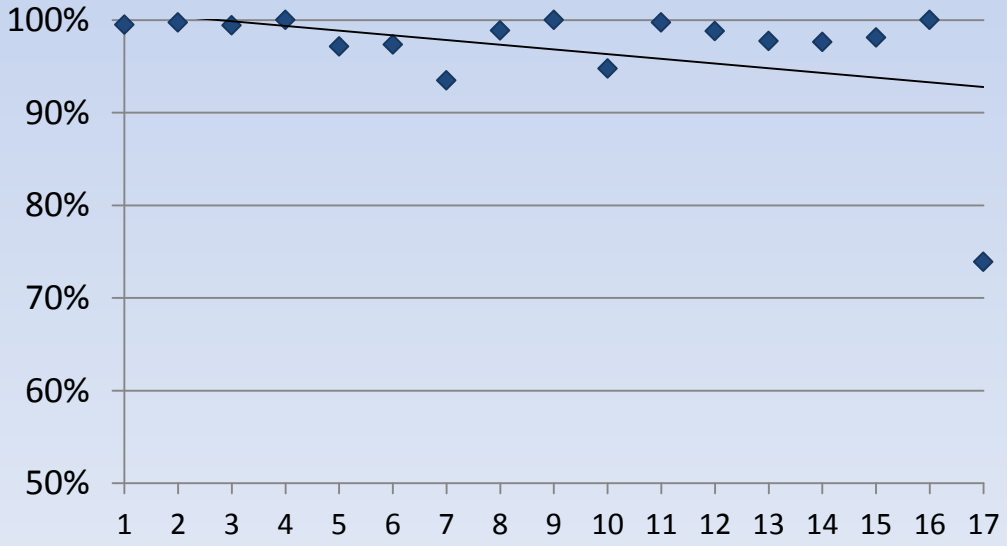
R^2 **(число сообщений)** = 0,266 (r = 0,516)

I уровень анализа: результаты

3. Max соединенный компонент



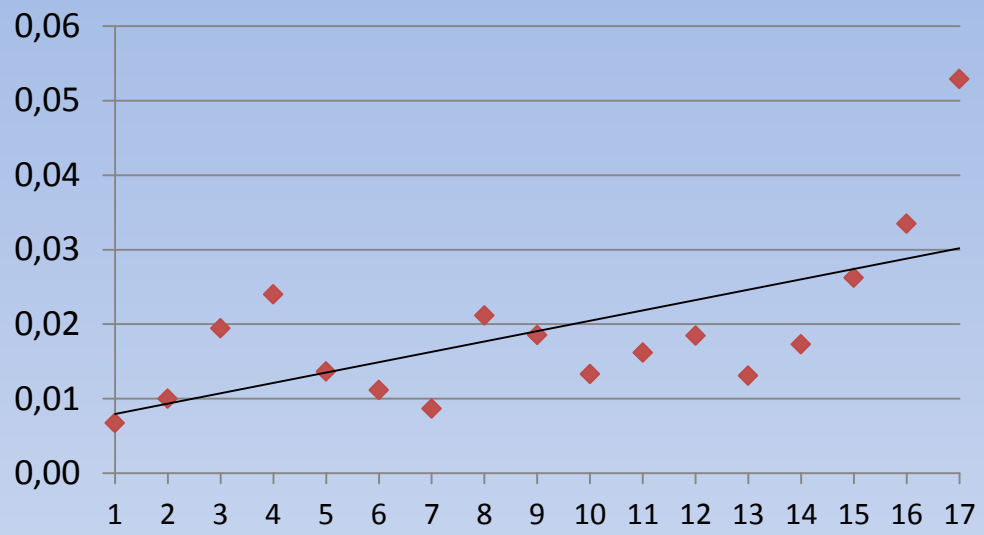
Доля **вершин** в max компоненте от числа всех связанных вершин
Размах вариации = 43% (22,4%)
Среднее = 91%
Стандартное отклонение = 11,5% (6,2%)
Коэффициент вариации = 13% < 33%,
=> вариация данных не значительна (совокупность однородна)
R^2 **(число участников)** = 0,187
R^2 **(число постов)** = 0,168



Доля **ребер** в max компоненте от числа всех ребер сети
Размах вариации = 26% (6,5%)
Среднее = 96,8%
Стандартное отклонение = 6,2% (1,9%)
Коэффициент вариации = 6% < 33%,
=> вариация данных не значительна (совокупность однородна)
R^2 **(число участников)** = 0,211
R^2 **(число постов)** = 0,152

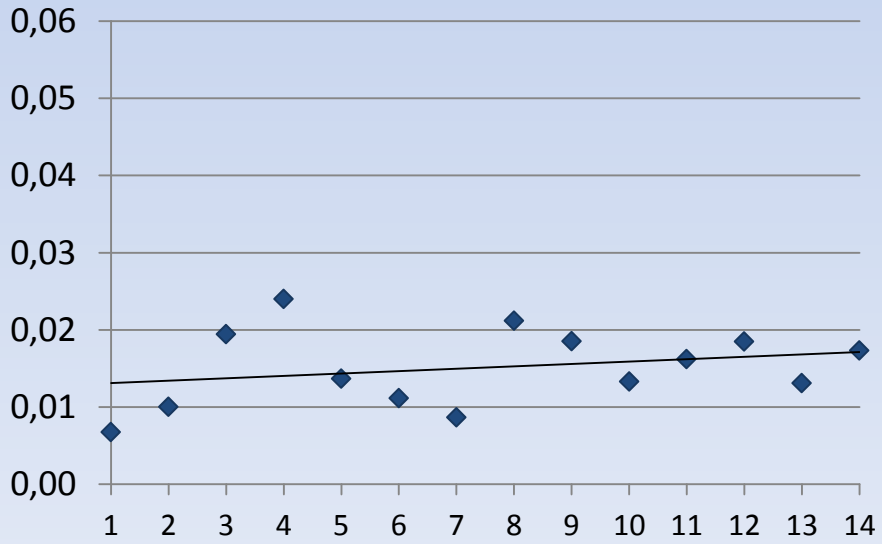
I уровень анализа: результаты

4.1. Плотность графа (с изолянтами)



Размах вариации = 0,046
Среднее = 0,019
Стандартное отклонение = 0,011
Коэффициент вариации = 58% > 33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = **0,531**
R^2 **(число постов)** = 0,267

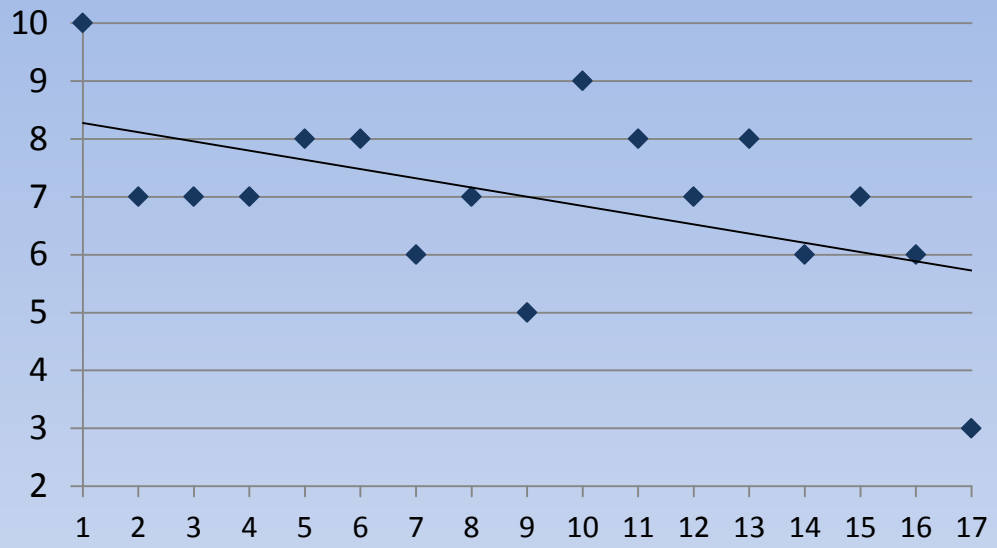
4.2. Плотность графа (выборка из 14 сообществ)



(**без** Курортного, Колпинского, Петродоврца)
Размах вариации = 0,017
Среднее = 0,015
Стандартное отклонение = 0,005
Коэффициент вариации = 33%
=> вариация данных не значительна
(совокупность однородна)
R^2 **(число участников)** = 0,185
R^2 **(число постов от им. сообщ.)** = 0,210

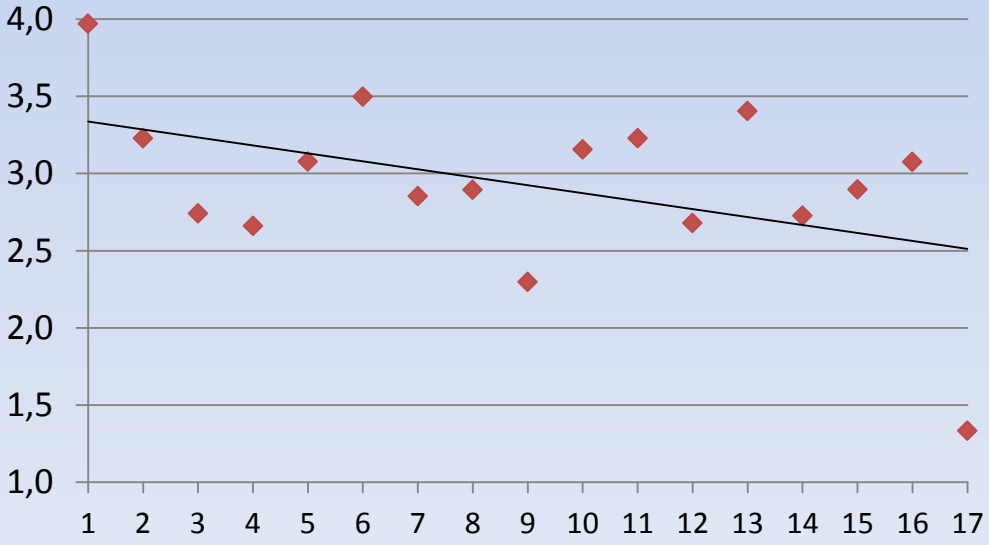
I уровень анализа: результаты

5. Диаметр графа



Размах вариации = 7
Среднее = 7
Стандартное отклонение = 1,6
Коэффициент вариации = 23% < 33%,
=> вариация данных не значительна
(совокупность однородна)
R^2 **(число участников)** = 0,401
R^2 **(число постов от им. сообщ.)** = 0,285

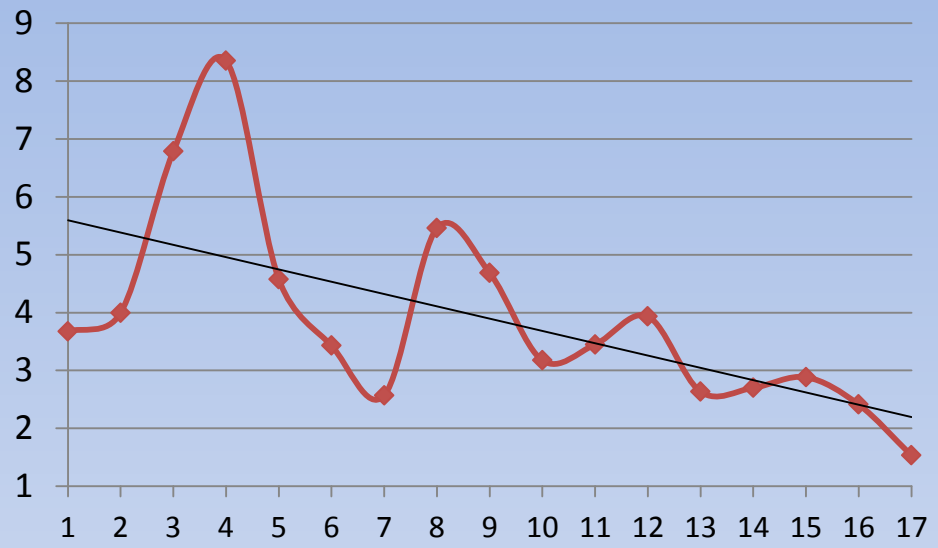
6. Средняя геодезическая дистанция



Размах вариации = 2,636
Среднее = 2,924
Стандартное отклонение = 0,563
Коэффициент вариации = 19% < 33%,
=> вариация данных не значительна
(совокупность однородна)
R^2 **(число участников)** = 0,346
R^2 **(число постов от им. сообщ.)** = 0,245

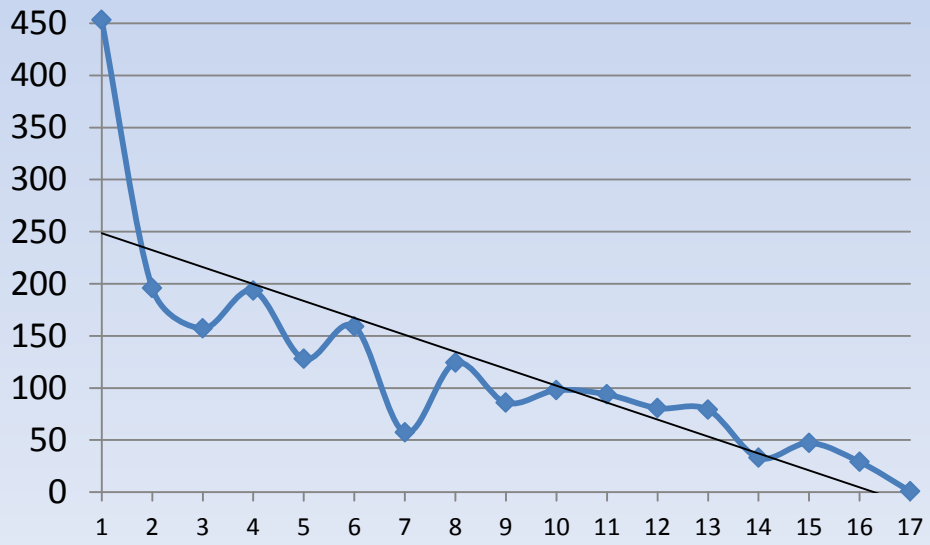
I уровень анализа: результаты

6. Среднее значение степени



Размах вариации = 6,817
Среднее = 3,897
Стандартное отклонение = 1,701
Коэффициент вариации = 44% > 33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = 0,268 (r = 0,518)
R^2 **(число постов)** = 0,190

7. Среднее значение betweenness centrality



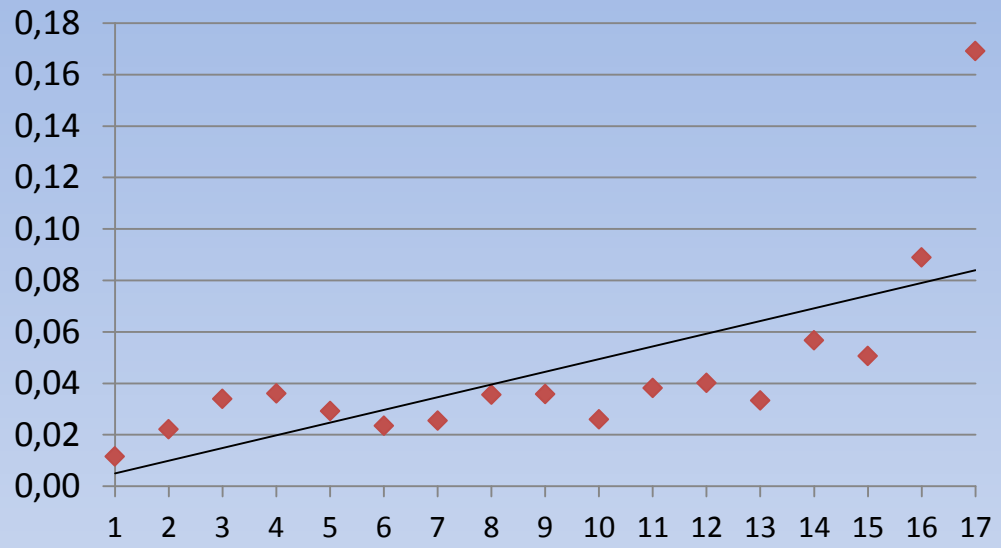
Размах вариации = 452,258
Среднее = 118,566
Стандартное отклонение = 103,047
Коэффициент вариации = 87% > 33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = **0,798**
R^2 **(число постов)** = **0,749**
R^2 **(число сообщений)** = 0,013

II уровень анализа: показатели

1. Плотность графа (без изолянтов)
2. Максимальные, средние и медианные значения степени и betweenness centrality в графе (без изолянтов)
3. **Число кластеров всего графа**
4. **Число кластеров максимального соединенного компонента**
5. **Распределение кластеров по числу вершин и ребер**
6. **Плотность самого крупного (в вершинах) кластера**
7. **Доля вершин самого крупного (в вершинах) кластера от всех связанных вершин**
8. **Перепад между max степенью в кластере и средней / медианой по кластеру**
   => топологические типы кластеров

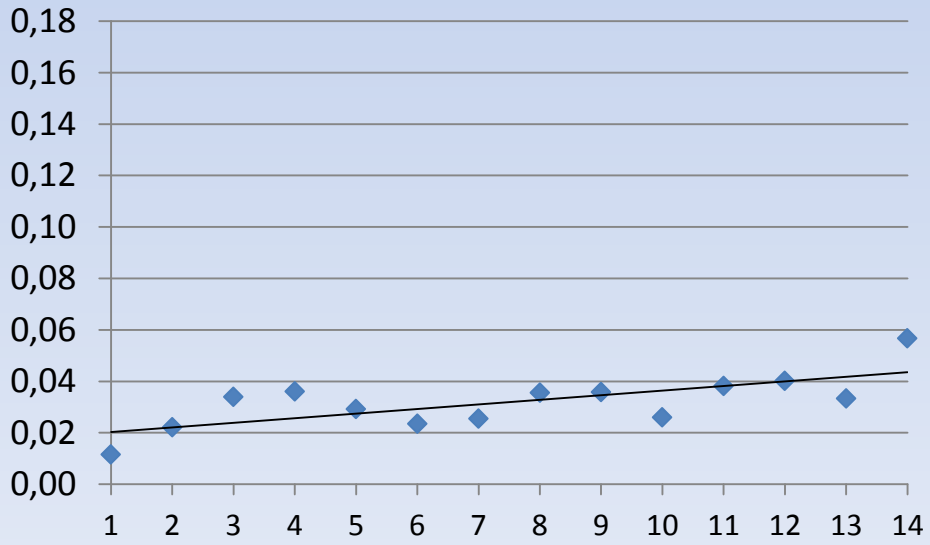
II уровень анализа: результаты

1.1. Плотность графа (без изолянтов)



Размах вариации =0,1576
Среднее = 0,0445
Стандартное отклонение = 0,0364
Коэффициент вариации = 81,8% > 33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = **0,554**
R^2 **(число постов)** = 0,278

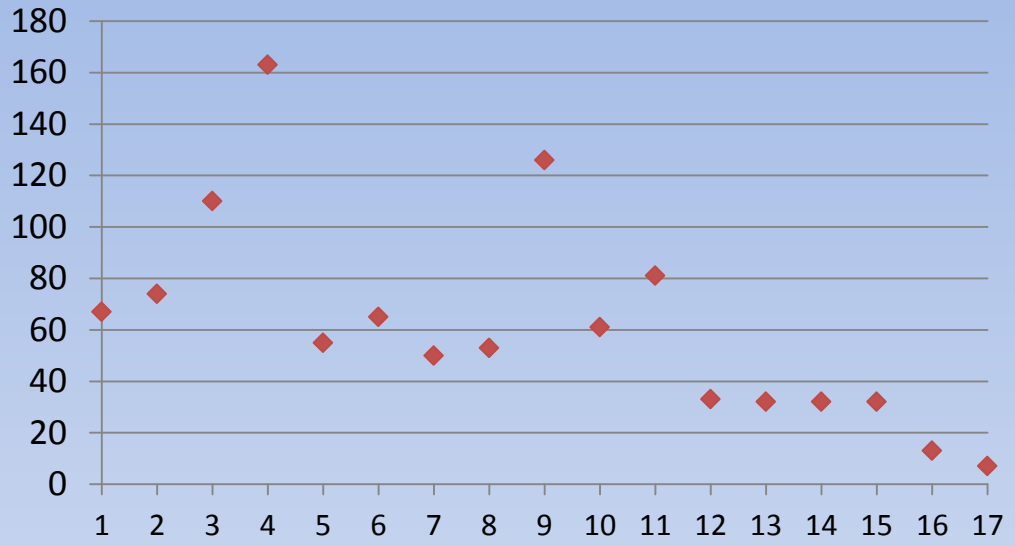
1.2. Плотность графа (выборка из 14 сообществ)



(**без** Курортного, Колпинского, Петродоврца)
Размах вариации = 0,0452
Среднее = 0,0319
Стандартное отклонение = 0,01
Коэффициент вариации = 32,9% < 33%,
=> вариация данных не значительна
(совокупность однородна)
R^2 **(число участников)** = **0,623**
R^2 **(число постов от им. сообщ.)** = 0,332

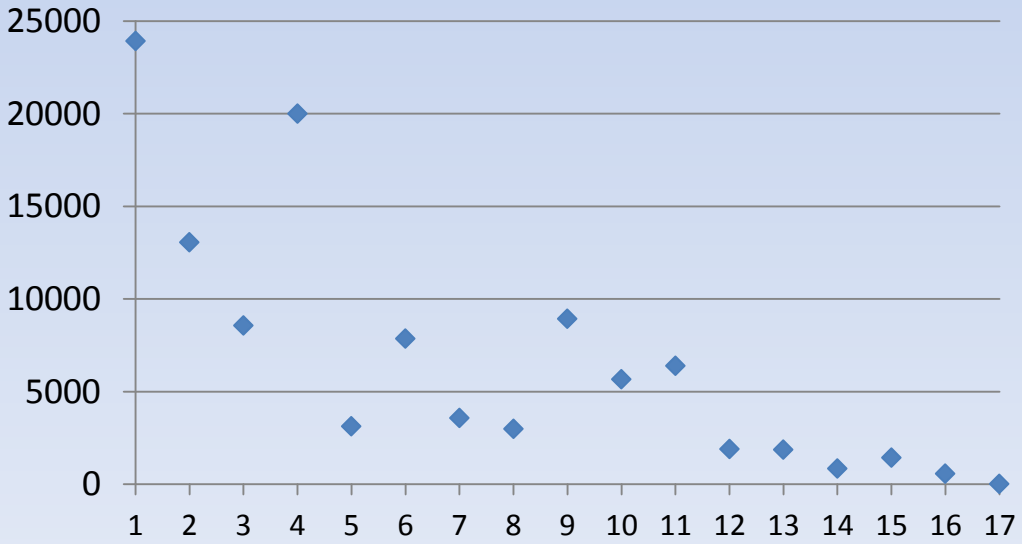
II уровень анализа: результаты

2.1. Max значение степени



Размах вариации = 156
Среднее = 62
Стандартное отклонение = 40,6
Коэффициент вариации = 65,5% >33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = 0,325
R^2 **(число постов)** = 0,232

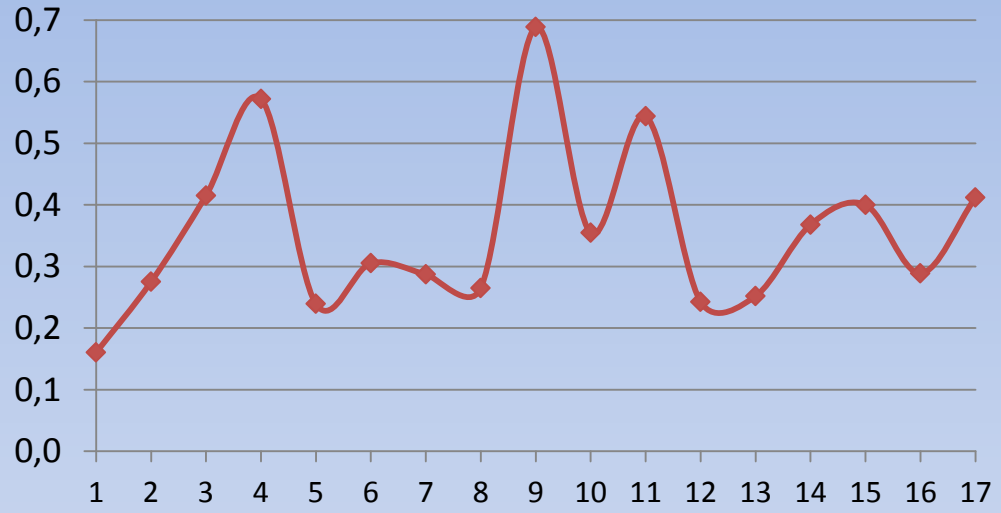
2. 2. Max значение betweenness centrality



Размах вариации = 23905,5
Среднее = 6506,1
Стандартное отклонение = 6851,7
Коэффициент вариации = 105% > 33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = **0,679**
R^2 **(число постов)** = **0,62**

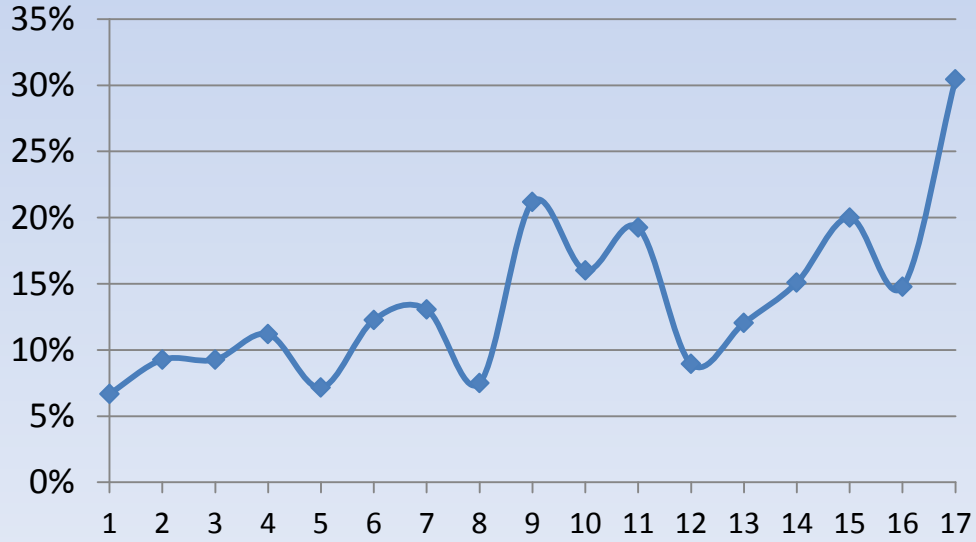
II уровень анализа: результаты

2.3. Max значение степени нормализованное на число связанных вершин



Размах вариации = 0,528
Среднее = 0,357
Стандартное отклонение = 0,138
Коэффициент вариации = 38,6% >33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = 0,115
R^2 **(число сообщений)** = 0,375

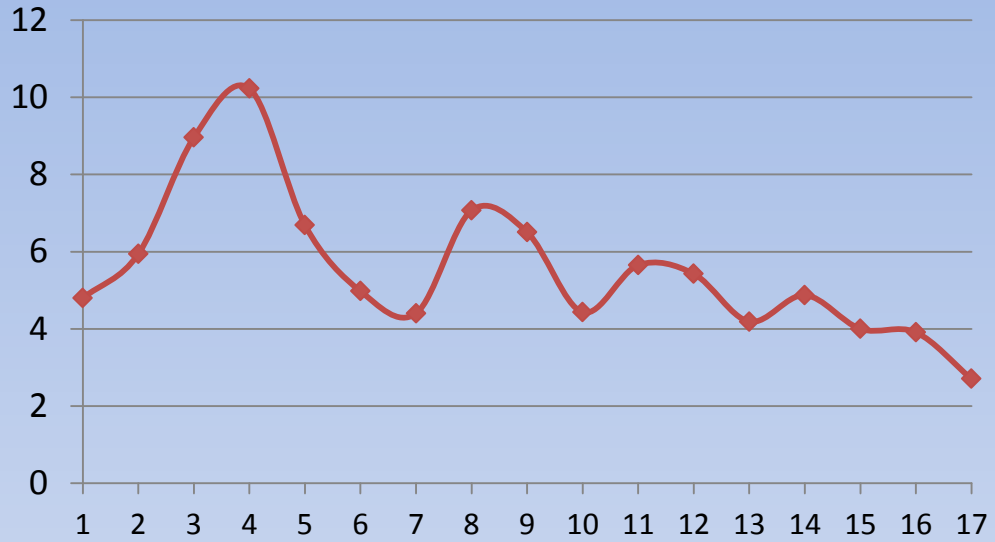
2.4. Доля max значения степени от всех ребер



Размах вариации = 23,8%
Среднее = 13,8%
Стандартное отклонение = 6,2%
Коэффициент вариации = 45,2% >33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = **0,512**
R^2 **(число постов)** = 0,319

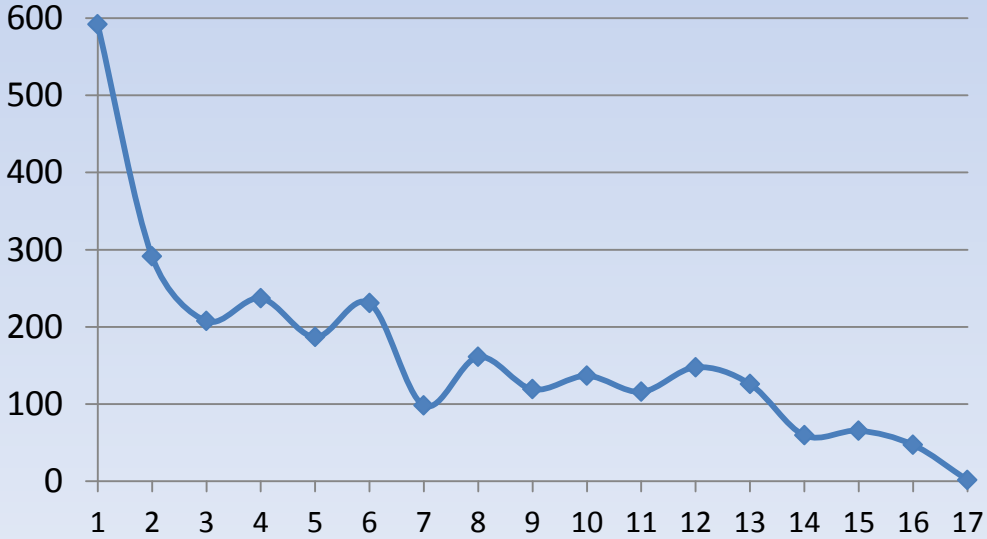
II уровень анализа: результаты

3.1. Среднее значение степени



Размах вариации = 7,519
Среднее = 5,573
Стандартное отклонение = 1,889
Коэффициент вариации = 33,3% >33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = 0,257
R^2 **(число постов)** = 0,181

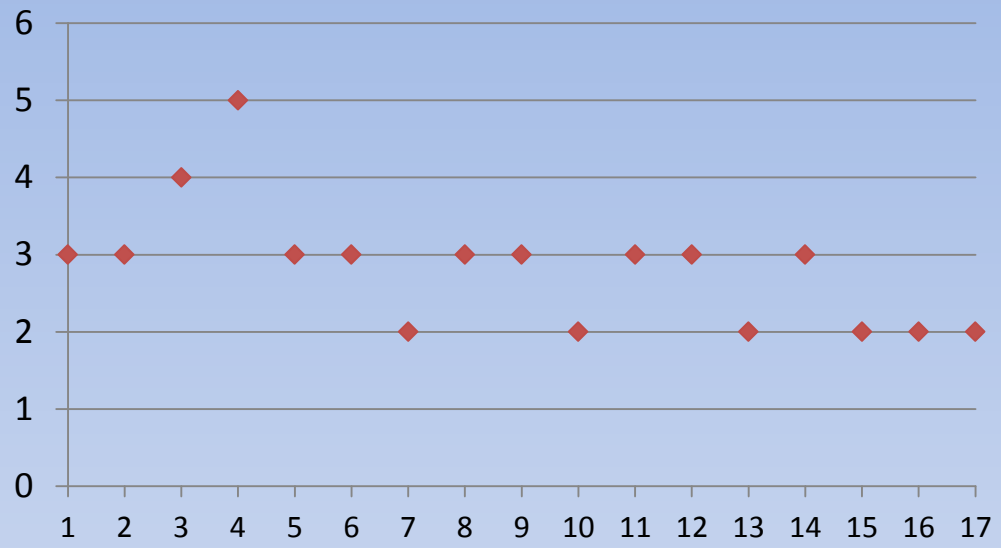
3.2. Среднее значение betweenness centrality



Размах вариации = 590,4
Среднее = 166,1
Стандартное отклонение = 133
Коэффициент вариации = 80% > 33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = **0,822**
R^2 **(число постов)** = **0,750**

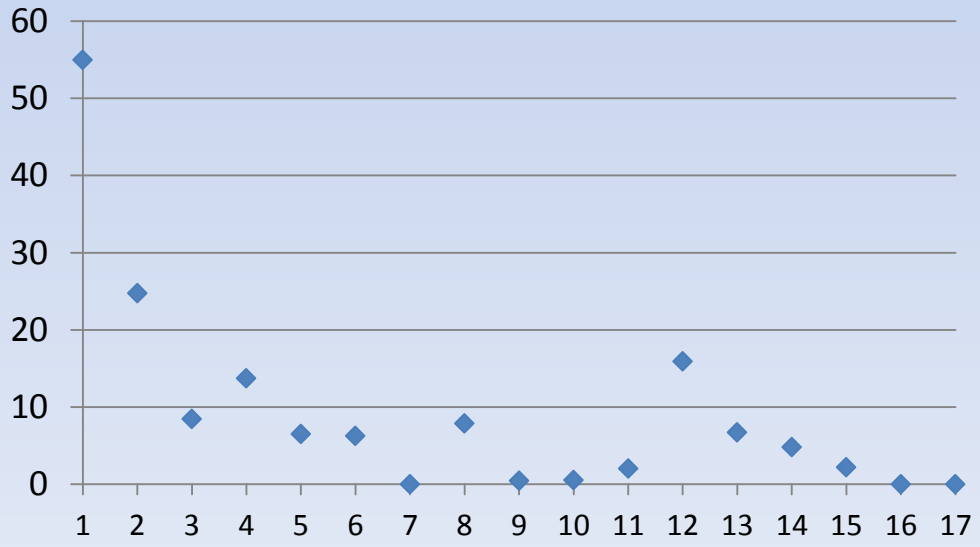
II уровень анализа: результаты

4.1. Медиана степени



Размах вариации = 3
Среднее = 2,8
Стандартное отклонение = 0,8
Коэффициент вариации = 28,7% < 33%,
=> вариация данных не значительна
(совокупность однородна)
R^2 **(число участников)** = 0,277
R^2 **(число постов)** = 0,234

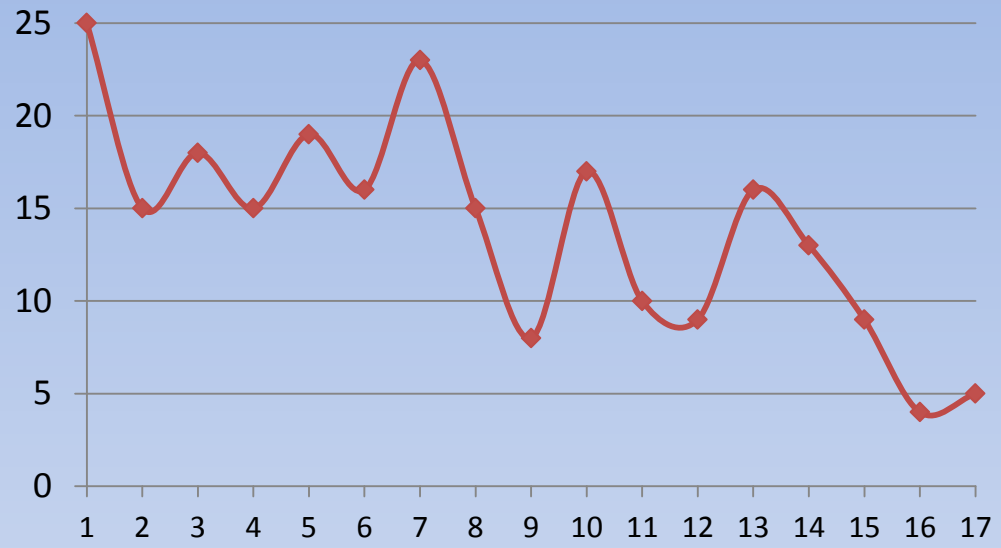
4.2. Медиана betweenness centrality



Размах вариации = 55
Среднее = 9,1
Стандартное отклонение = 13,6
Коэффициент вариации = 149% > 33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = **0,564**
R^2 **(число постов)** = **0,690**
R^2 **(число сообщений)** = 0,003

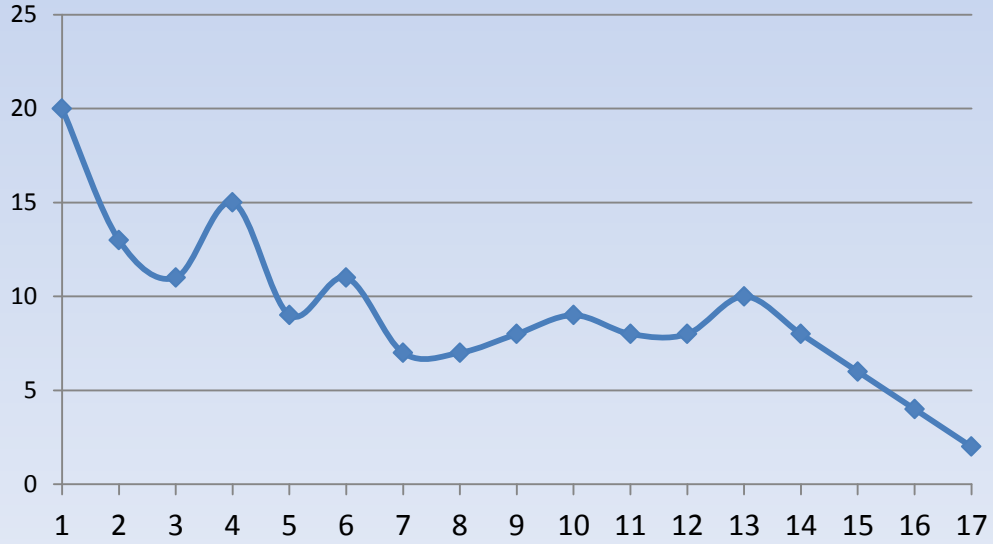
II уровень анализа: результаты

5.1. Число кластеров всего



Размах вариации = 21
Среднее = 13,9
Стандартное отклонение = 5,8
Коэффициент вариации = 42% > 33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = **0,661**
R^2 **(число постов от им. сообщ.)** = 0,419

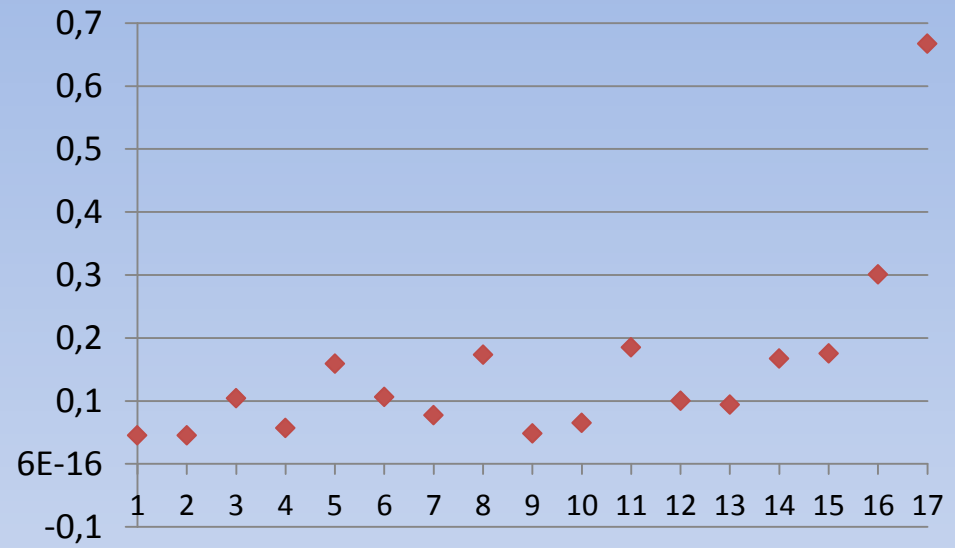
5.2. Число кластеров max соединенного компонента



Размах вариации = 18
Среднее = 9,2
Стандартное отклонение = 4,2
Коэффициент вариации = 45,3% > 33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = **0,816**
R^2 **(число постов)** = **0,666**
R^2 **(число сообщений)** = 0,006

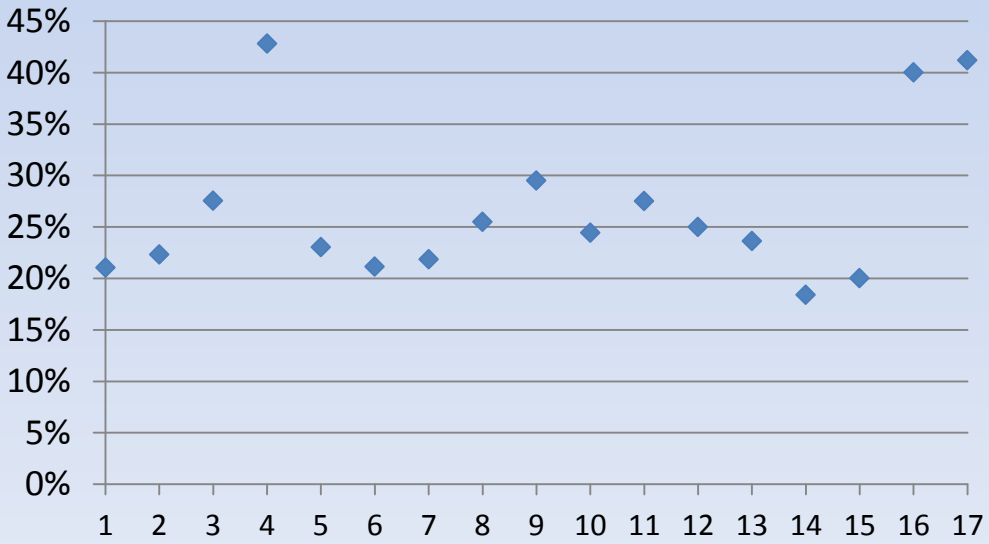
II уровень анализа: результаты

6.1. Плотность max кластера



Размах вариации = 0,622
Среднее = 0,151
Стандартное отклонение = 0,149
Коэффициент вариации = 99% > 33%,
=> вариация данных значительна
(совокупность неоднородна)
R^2 **(число участников)** = 0,474
R^2 **(число постов)** = 0,275

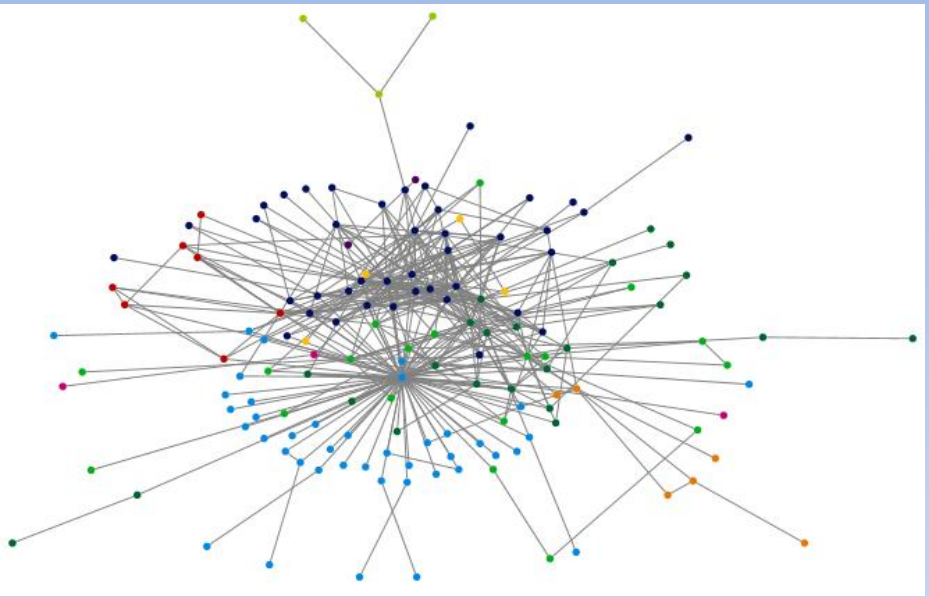
6.2. Доля вершин max кластера от всех связанных вершин

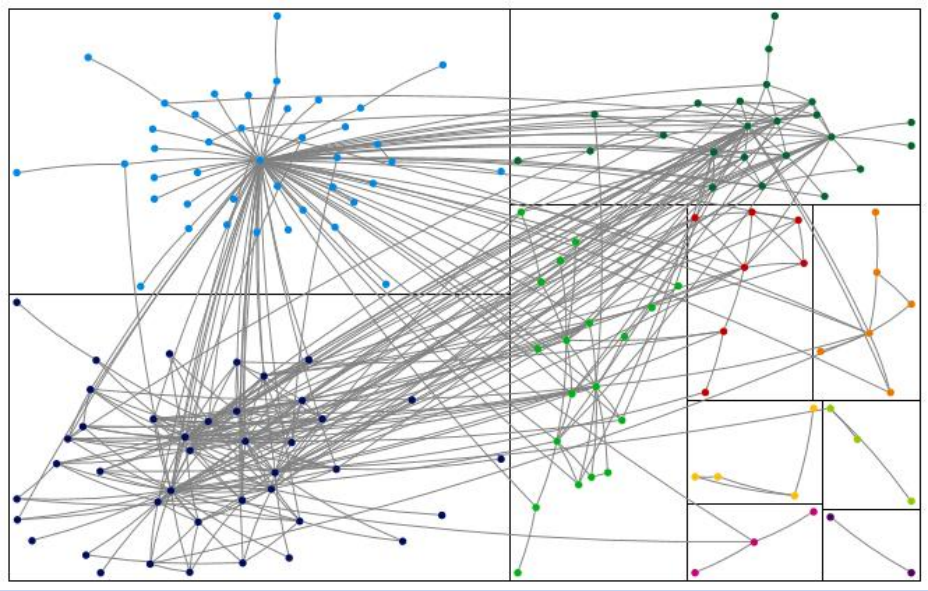


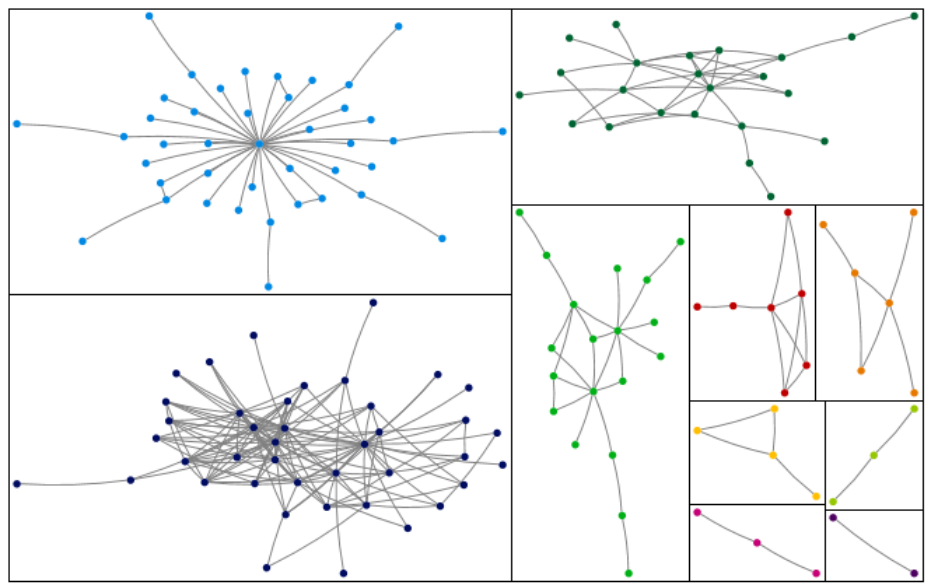
Размах вариации = 24,4%
Среднее = 26,8%
Стандартное отклонение = 7,5%
Коэффициент вариации = 28% < 33%,
=> вариация данных не значительна
(совокупность однородна)
R^2 **(число участников)** = 0,115
R^2 **(число обсуждений)** = 0,137

II уровень анализа: результаты

Топологические типы кластеров: сеть группы Московского р-на



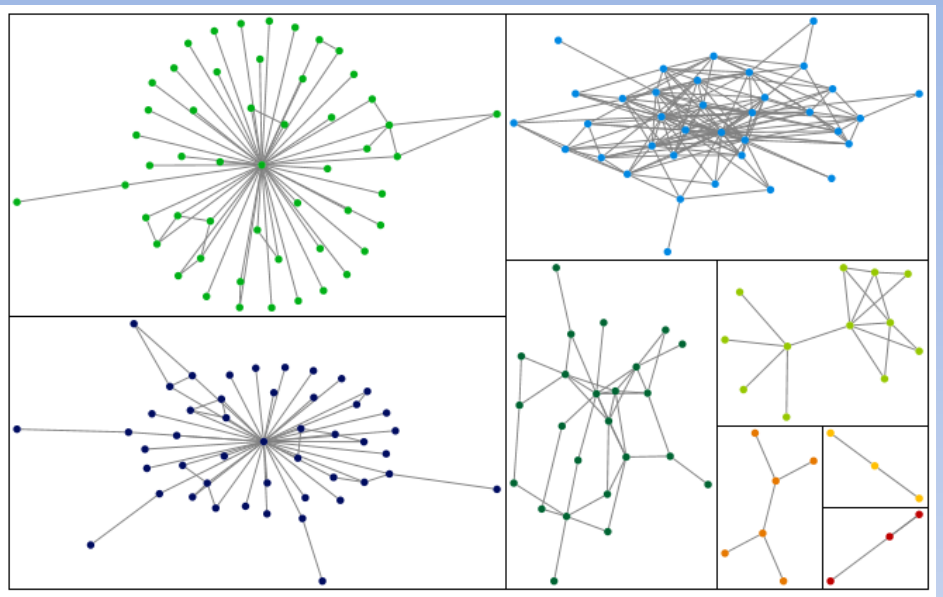




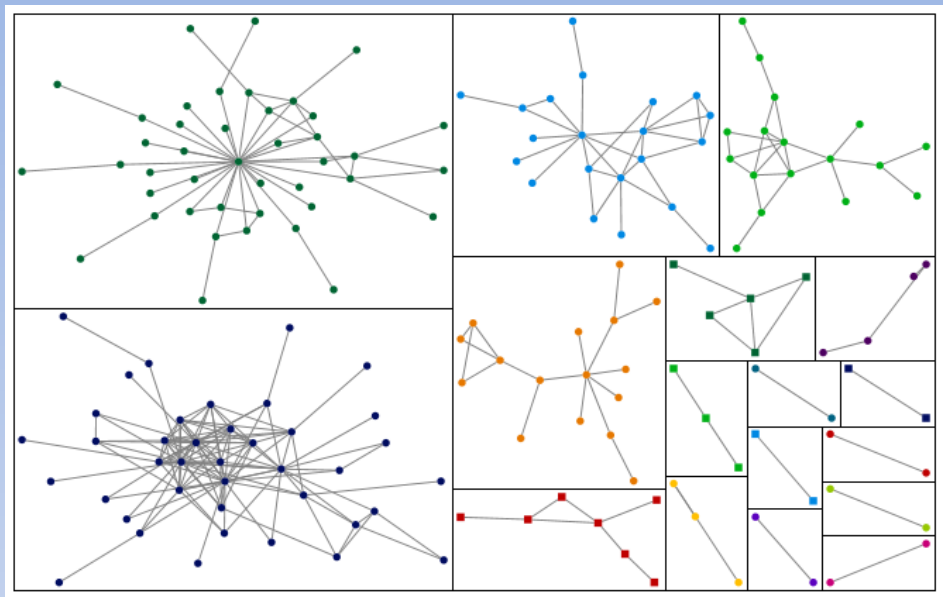
Неравномерное распределение степени внутри кластера: можно измерить отношением max степени к средней или медиане по кластеру.

UD = max degree / average (median) degree

II уровень анализа: результаты



Красносельского район



Выборгский район

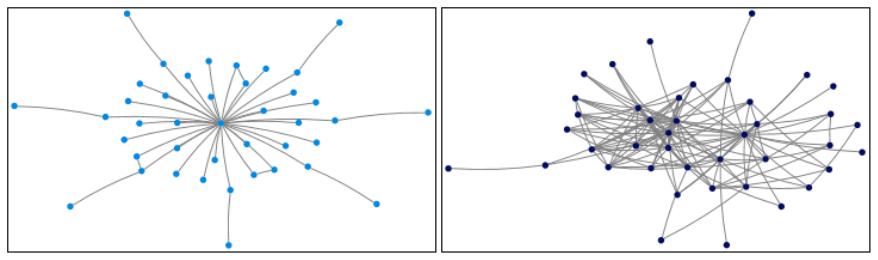
Сообщества, где наблюдается подобный феномен – **наибольшие кластеры сети принадлежат разным топологическим типам («звездчатым» и равномерно связанным)**:

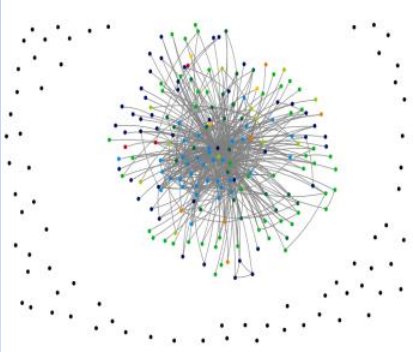
Василеостровский
Выборгский
Калининский
Красносельский
Московский
Невский
Петродворец
Приморский
Фрунзенский **= 9 сообществ**

Выводы

1. На I уровне «приближения», связанном с анализом макро структур сети (соединенных компонентов и изолированных участников) обнаружены сходства, что свидетельствует об образовании сетевой структуры *по **единой модели*** в сообществах различных по уровню активности и своим размерам.

   Данную модель структуры образно можно сравнить со строением ***атома***: соединенные вершины тяготеют к объединению в одно ядро + относительно постоянная доля изолянтов (20-40%).





2. На II уровне «приближения», связанном с анализом кластеров, также обнаружены некоторые сходства: наибольшие кластеры - разных топологических типов (*равномерное* VS *неравномерное* распределение степени внутри кластера).
   Требует статистически обоснованного подтверждения.

Перспективы и проблемы

1. III уровень анализа – элементарные единицы сети (отдельные вершины): объяснение позиции каждого пользователя в сети «дружбы» через индивидуальные показатели его коммуникативной активности.
2. Локализация модераторов сообщества в графе
3. Кросс-сетевой анализ сообществ (всех сообществ как одного) с целью выявления общих подгрупп участников
4. Теоретическая интерпретация результатов исследования

1. Методологические проблемы: обоснование предложенных мер для сравнения сетей
2. Необходимость разработки специальных модулей ПО для закачки данных о коммуникации пользователей.

Благодарности и используемое ПО

- Сергею Кольцову за написание ПО для закачки сетей «дружбы»
- Олесе Кольцовой за плодотворное сотрудничество
- Social Media Research Foundation за разработку и свободное распространение ПО для анализа социальных сетей

NodeXL Network Graphs
The Social Media Research Foundation

- NodeXL http://nodexl.codeplex.com/

  Все графы визуализированы с помощью NodeXL
- Vkontakte Networks

ЛАБОРАТОРИЯ ИНТЕРНЕТ-ИССЛЕДОВАНИЙ

ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
САНКТ-ПЕТЕРБУРГ

Спасибо за внимание!