



Московский институт  
электроники и математики  
имени А. Н. Тихонова

Департамент прикладной математики

Москва  
2022

№931

# Топологический анализ данных в исследовании сложных сетей

Выполнил(а):

Забродина Татьяна Николаевна

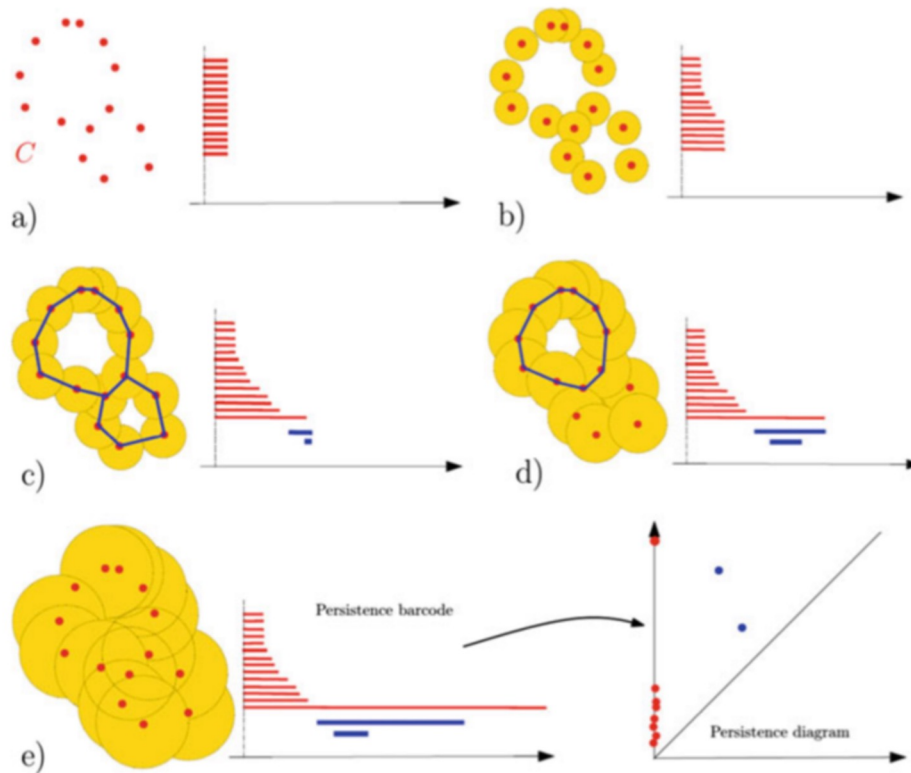
MCMT213

Научный руководитель:

Вальба О.В., доцент ДПМ, к.ф.-м.н

## Актуальность

- Топологический анализ данных (TDA) – современный подход статистического анализа данных;
- применение TDA к анализу сложных сетей позволит выявить «скрытую» топологию данных, трудно устанавливаемую стандартными структурными методами.





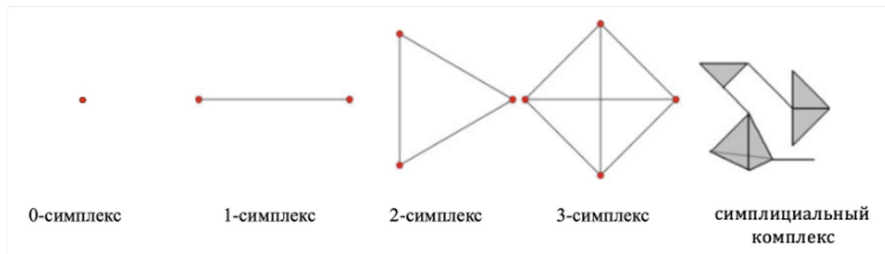
## Цель

исследование применимости  
методов топологического анализа данных к сложным сетям

## Задачи

- разработка алгоритмов вычисления и анализа топологических характеристик для сетевых данных
  - моделирование различных сценариев изменения сетей
  - анализ сетевых данных – сеть свободных ассоциаций

## Подробнее о TDA



Правила:

1. Все грани симплекса лежат внутри симплициального комплекса.
2. Склеивание симплексов происходит по общим граням симплексов.

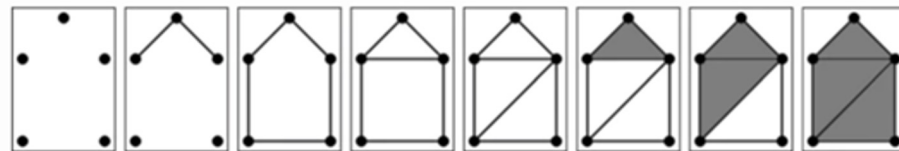
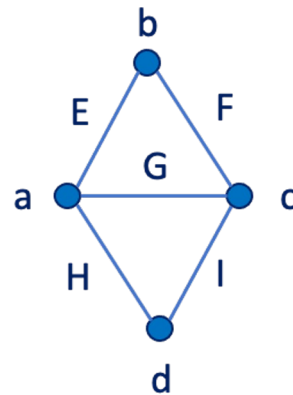
**Симплициальный комплекс:**

0-симплекс:  $a, b, c, d$

1-симплекс:  $E, F, G, I, H$

2-симплекс:  $abc, acd$

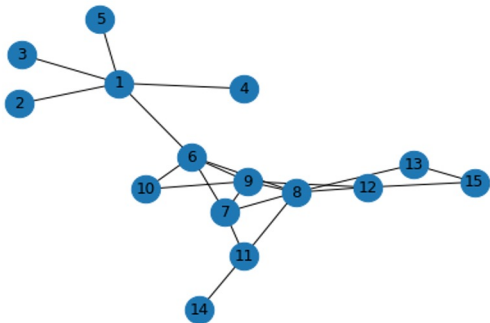
Числа Бетти:  $1, 2, 0, \dots$



**Пример фильтрации**

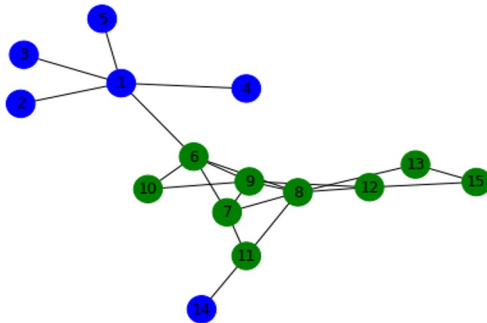
## Разложение по k-core

**k-core (к-ядро)** графа  $G$  – максимальный связный подграф  $G$ , в котором все вершины имеют степень по меньшей мере  $k$ .



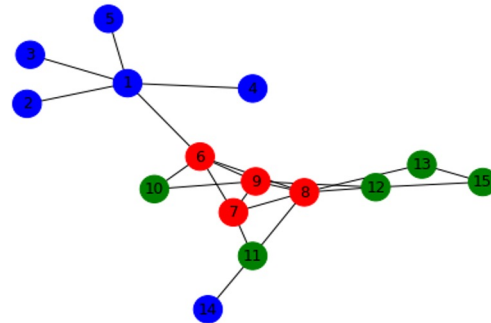
1-core

$$k = 1, f(k) = 1$$



2-core

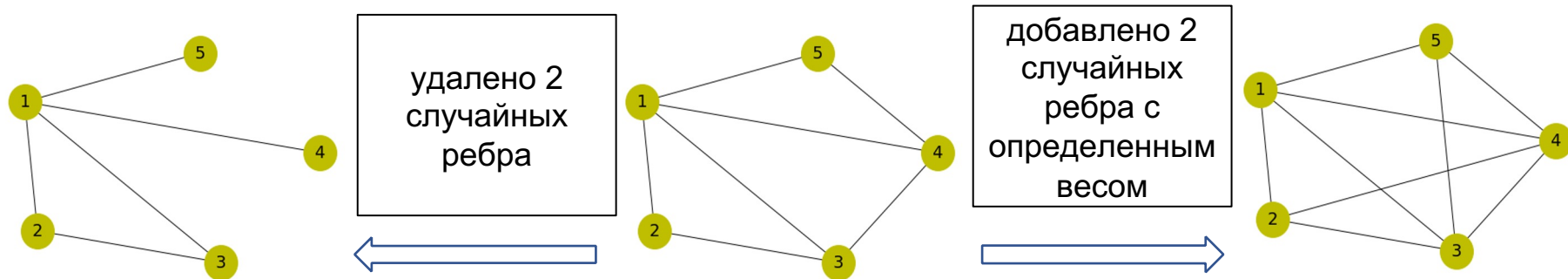
$$k = 2, f(k) = 9/15$$



3-core

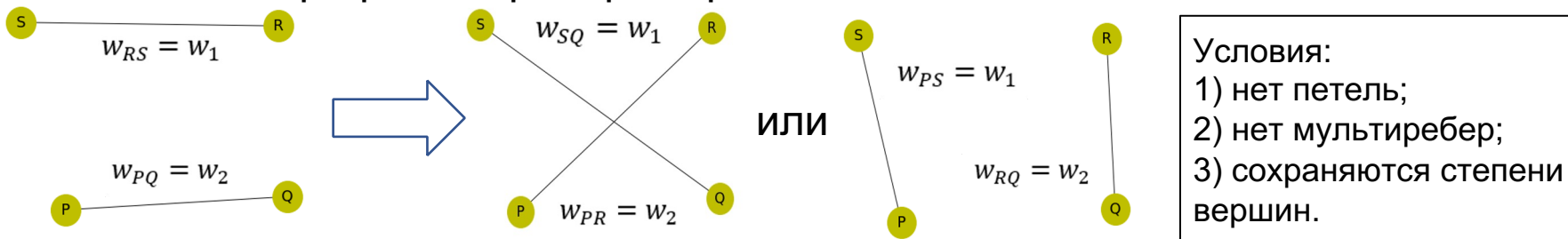
$$k = 3, f(k) = 4/15$$

## Удаление и добавление ребер



## Рандомизация с сохранением степеней вершин

Взвешенный граф: веса ребер сохраняются



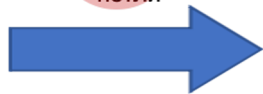


# Семантические сети

## Описание данных

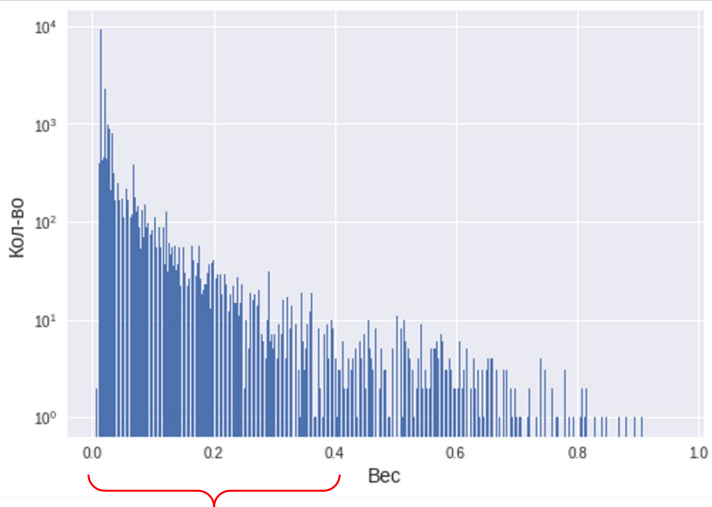
Взвешенный направленный граф на 5019 вершин и 63625 ребра

- направленные мультиребра ребра
  - петли

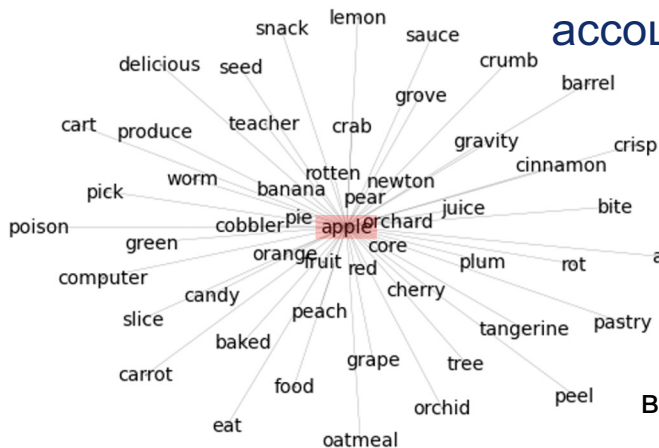


Взвешенный **ненаправленный**  
граф на 5019 вершин и 55242 ребра

## Распределение весов



## Пример из сети свободных ассоциаций



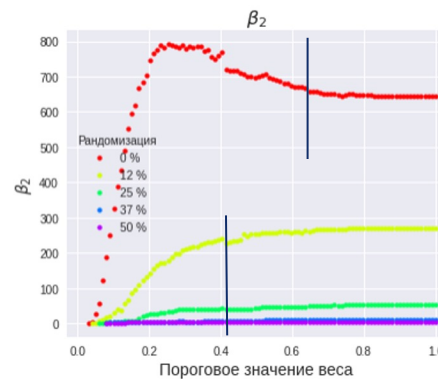
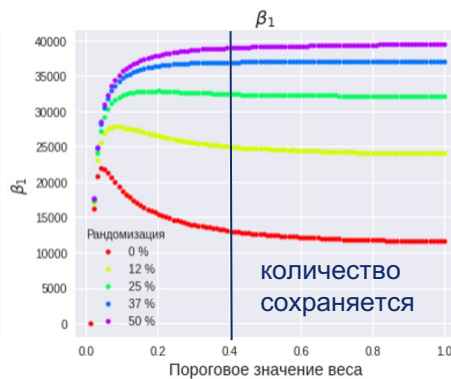
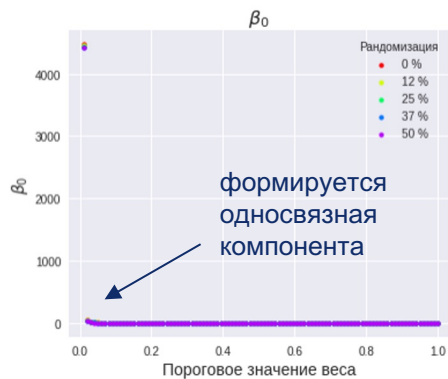
**вершина** - СЛОВО

**ребро** - возникновение ассоциации

try **вес ребра** - доля  
респондентов, у которых  
возникала такая ассоциация

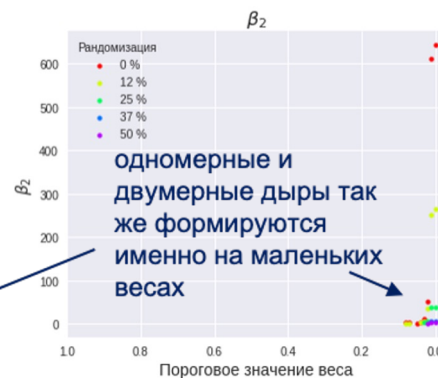
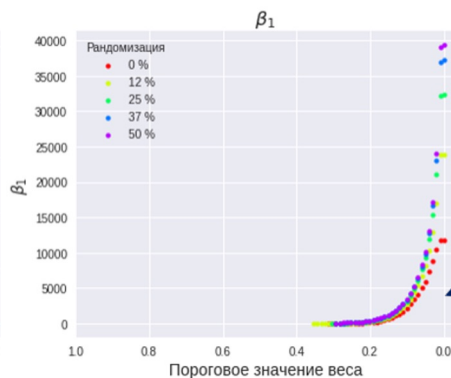
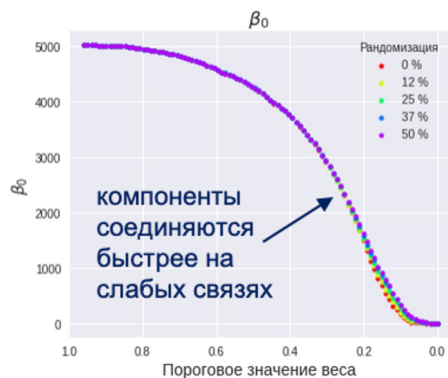


# Рандомизация и числа Бетти



фильтрация  
Вьеториса-Рипса  
( $G[\text{веса}] < \text{порог}$   
порог : 0  $\rightarrow$  1)

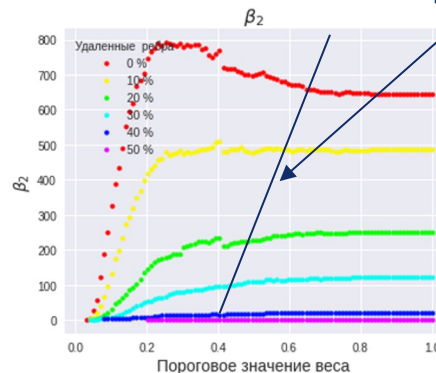
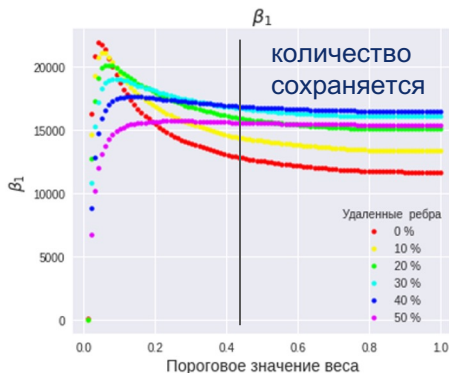
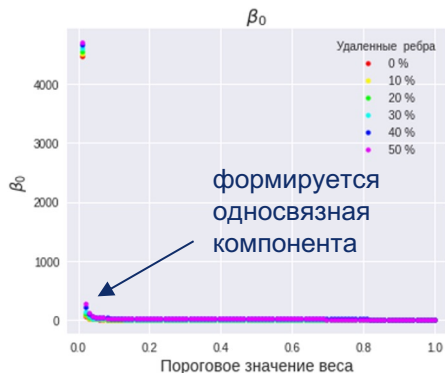
- граф без изменений



обратная фильтрация  
Вьеториса-Рипса  
( $G[\text{веса}] > \text{порог}$   
порог : 1  $\rightarrow$  0)

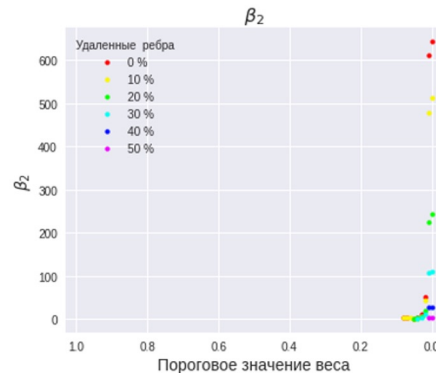
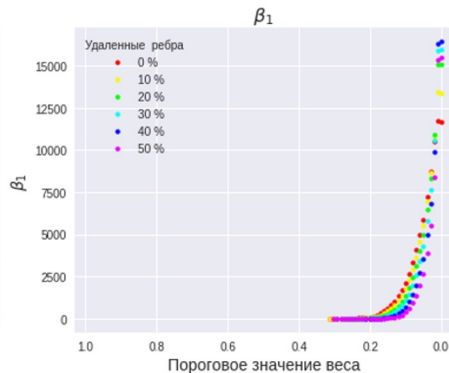
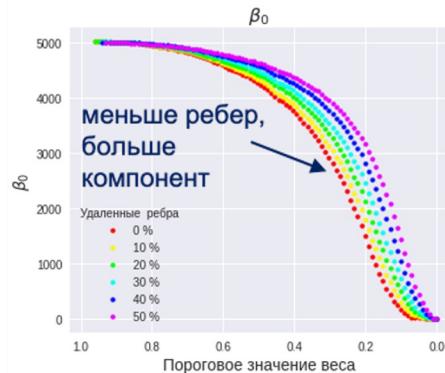
# Удаление ребер и числа Бетти

меньше ребер, быстрее  
завершается  
формирование пустот



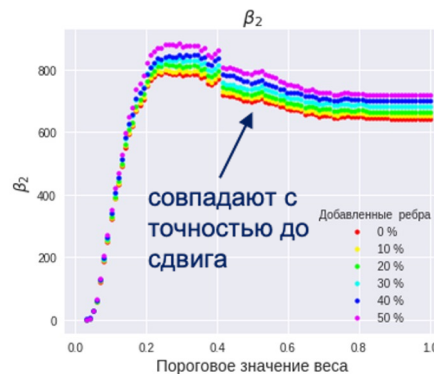
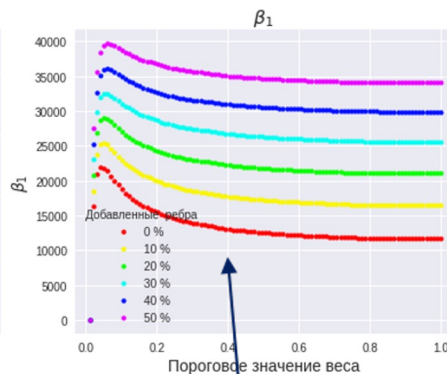
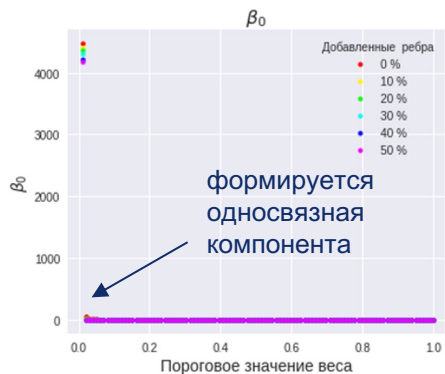
фильтрация  
Вьеториса-Рипса  
( $G[\text{веса}] < \text{порог}$   
порог : 0  $\rightarrow$  1)

- граф без изменений

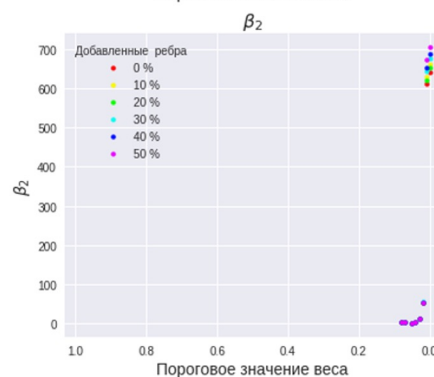
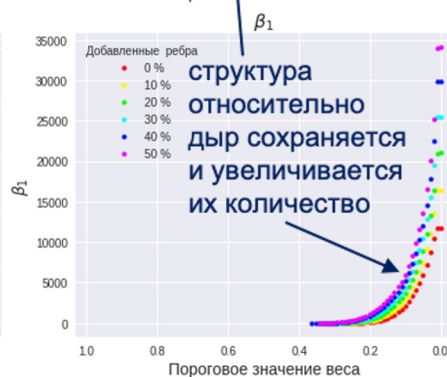
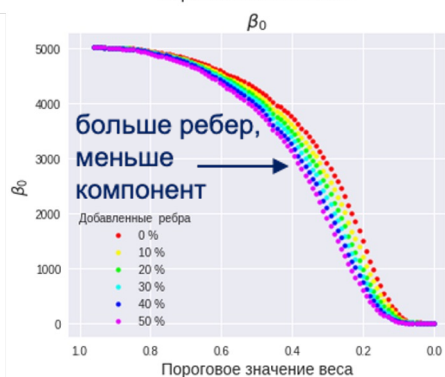


обратная фильтрация  
Вьеториса-Рипса  
( $G[\text{веса}] > \text{порог}$   
порог : 1  $\rightarrow$  0)

# Добавление ребер и числа Бетти



фильтрация  
Вьеториса-Рипса  
( $G[\text{веса}] < \text{порог}$   
порог : 0 -> 1)



- граф без изменений

обратная фильтрация  
Вьеториса-Рипса  
( $G[\text{веса}] > \text{порог}$   
порог : 1 -> 0)



# k-core и возмущения сети

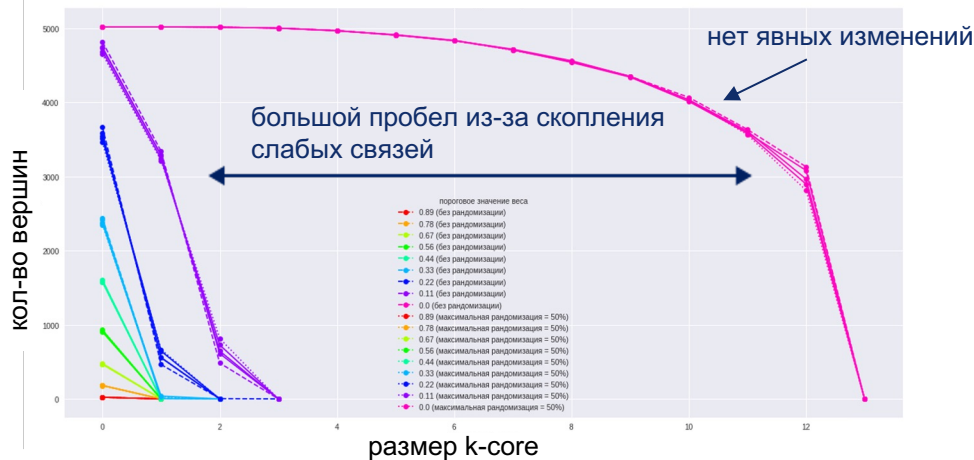
обратная фильтрация Вьеториса-Рипса

( $G[\text{веса}] > \text{порог}$ )

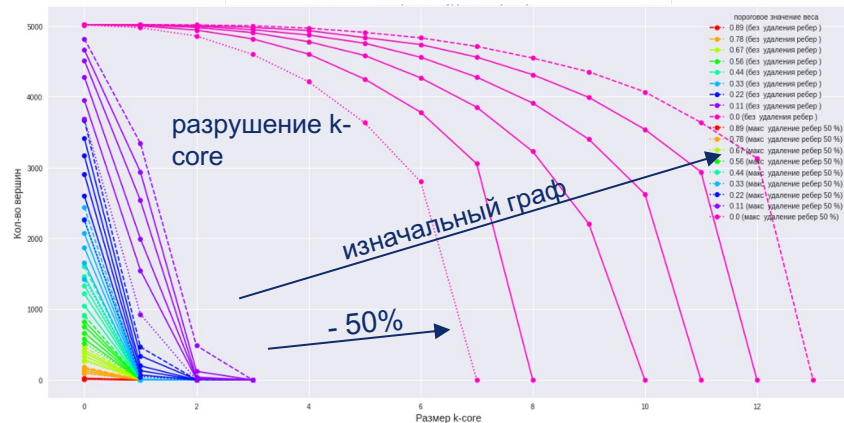
порог : 1  $\rightarrow$  0)

— изначальный граф    ..... макс процент изменений

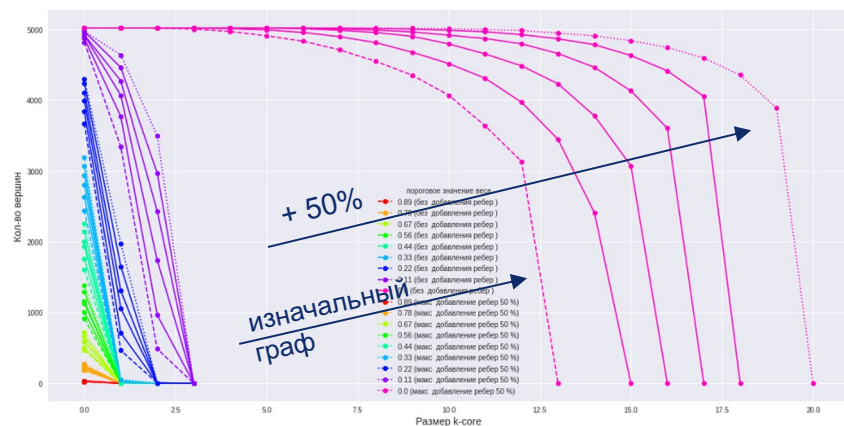
рандомизация



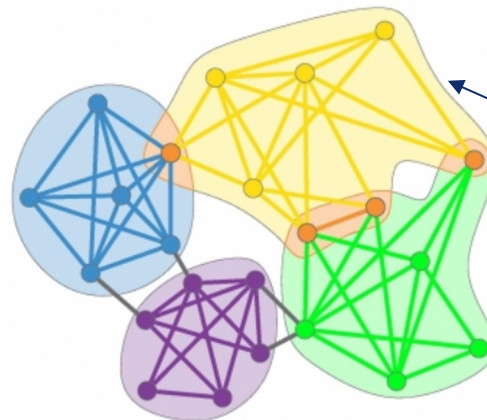
удаление ребер



добавление ребер

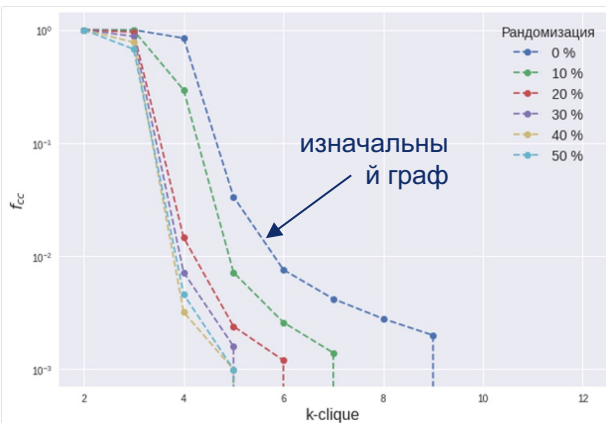


# Кликовые кластеры

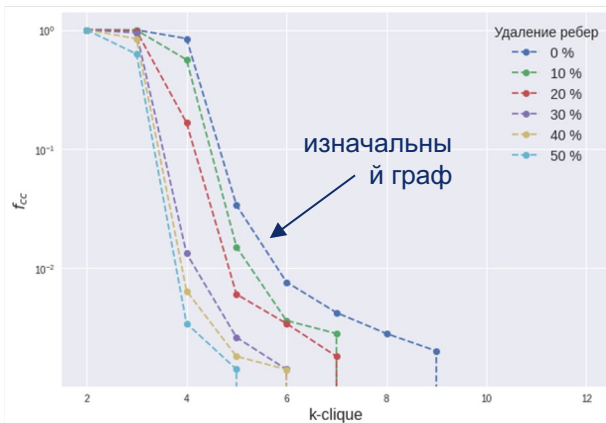


кластер  
максимального  
размера, построенный  
на 4-кликах

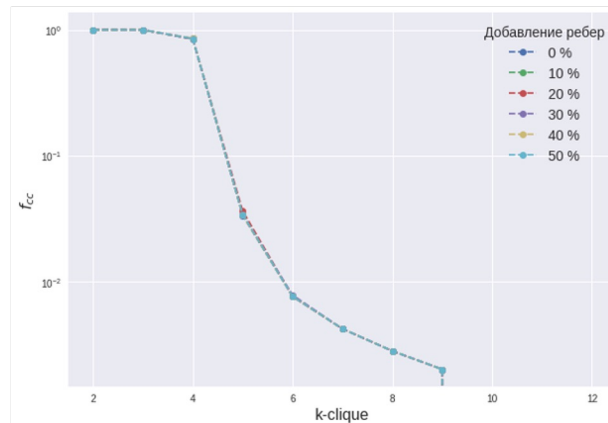
## Рандомизация



## Удаление ребер

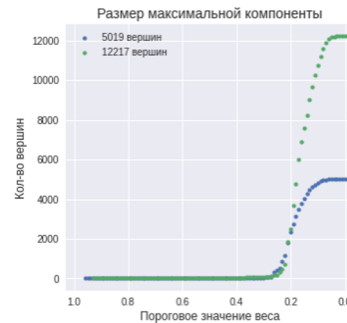
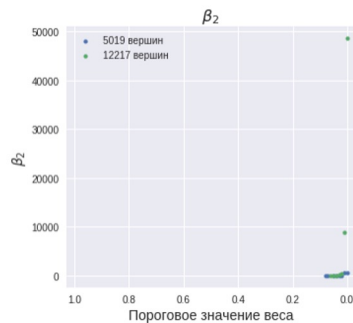
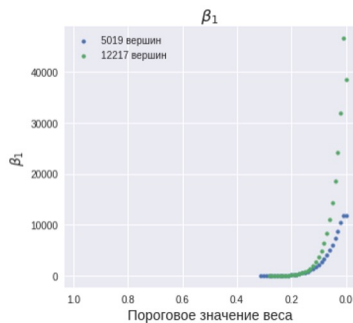
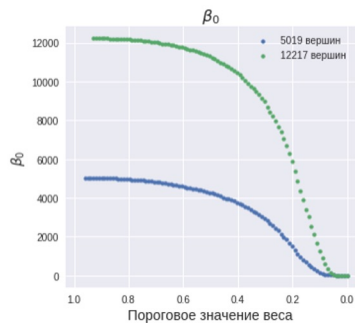


## Добавление ребер





# Сравнение двух семантических сетей

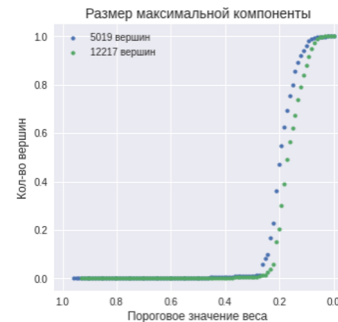
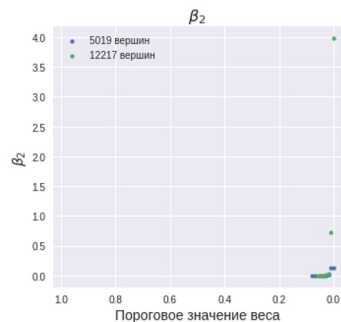
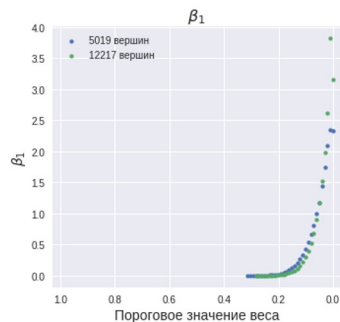
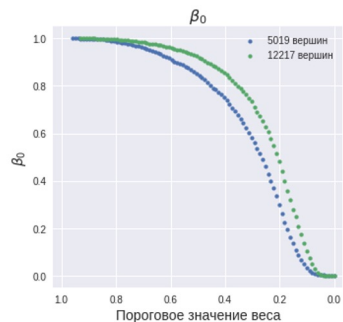


**основная сеть:**  
5 019 вершин  
55 242 ребер

**сеть для  
сравнения:**  
12 217 вершин  
352 403 ребер



нормировка на количество вершин



обратная фильтрация  
Вьеториса-Рипса  
( $G[\text{веса}] > \text{порог}$   
порог : 1  $\rightarrow$  0)



## Итоги и выводы

1. адаптация методов топологического анализа для исследования сложных сетей;
2. анализ топологических характеристик сетей различной природы;
3. анализ устойчивости топологий сетей к различным изменениям в сети.



Коньков Сергей  
sikonkov@edu.hse.ru

Забродина Татьяна  
tnzabrodina@edu.hse.ru

Ригвава Владимир  
vgrigvava@miem.hse.ru