

# Тестирование стационарности сообществ

Васильев Михаил Владимирович  
Студент 5 курса факультета ФРКТ

(Московский физико-технический институт)  
(Dated: 11 мая 2024 г.)

## I. ВВЕДЕНИЕ

Постановка задачи состоит в исследовании стационарности динамического графа во времени. Для этого используется датасет графов из открытых источников таких как [1]. Одним из подходов исследования стационарности сообществ является сведение к тестированию случайных последовательностей, возникающих на графе. Для этого используются случайные блуждания для сбора информации о характеристиках влияния (PageRank), узлов сообщества, которые определяют соответствующие случайные последовательности. Далее производится вычисление хвостового индекса на элементах этих последовательностей и посредством визуального анализа делается вывод о стационарности графа во времени. Ключевым критерием анализа является значительное изменение хвостового индекса между соседними временными промежутками.

## II. DATA

Данные взяты из открытого графового репозитория [1], а именно динамический граф fb-messages.csv. Описание: "Социальная сеть, похожая на Facebook, создана на основе онлайн-сообщества студентов Калифорнийского университета в Ирвине. В набор данных входят пользователи, которые отправили или получили хотя бы одно сообщение." Граф имеет три столбца: source, target, time. Количество вершин - 1900, количество рёбер - 61732.

## III. ОПИСАТЕЛЬНЫЕ ХАРАКТЕРИСТИКИ

Рассмотрим граф как статистический на последний момент времени, рассчитаем для каждого его узла PageRank и проведём исследование распределения. Исследование состоит в оценивании индекса экстремального значения  $\gamma$  разными методами и изучении хвоста распределения.

## IV. ОПИСАТЕЛЬНЫЕ ХАРАКТЕРИСТИКИ

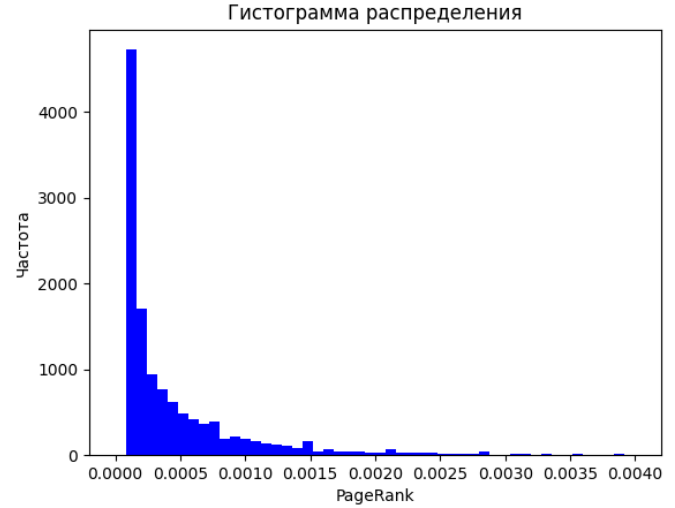


Рис. 1: Распределение PageRank

### A. Гистограмма распределения

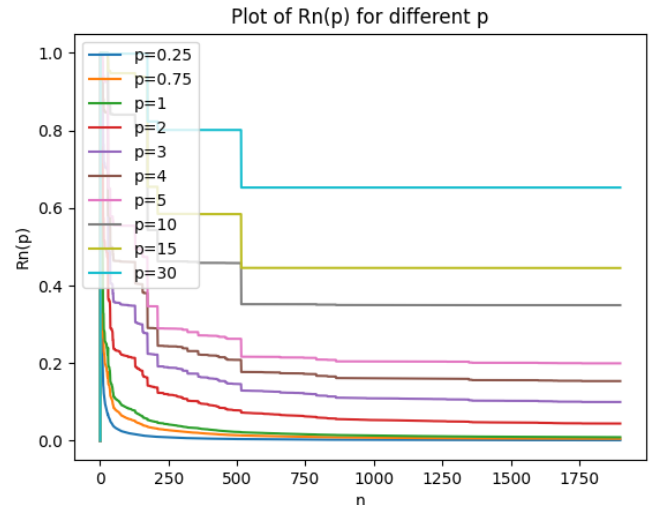


Рис. 2: Зависимость отношения максимума к сумме элементов для различных значений хвостового индекса

По итогам исследования можно сказать, Для  $p \in \{0.25, 0.75, 1\}$   $R_n(p)$  по всей видимости стремиться к нулю при возрастании  $n$ . Для  $p \in \{2, 3, 4, 5, 10, 15, 30\}$   $R_n(p)$  по всей видимости стремиться к положительной константе при возрастании  $n$ .

Вывод:  $E|X|^p < \infty$  для  $p \leq 1$  только,  $E|X|^p = \infty$  для  $p > 1$ . Значит распределение имеет единственный момент.

### В. Оценка распределения

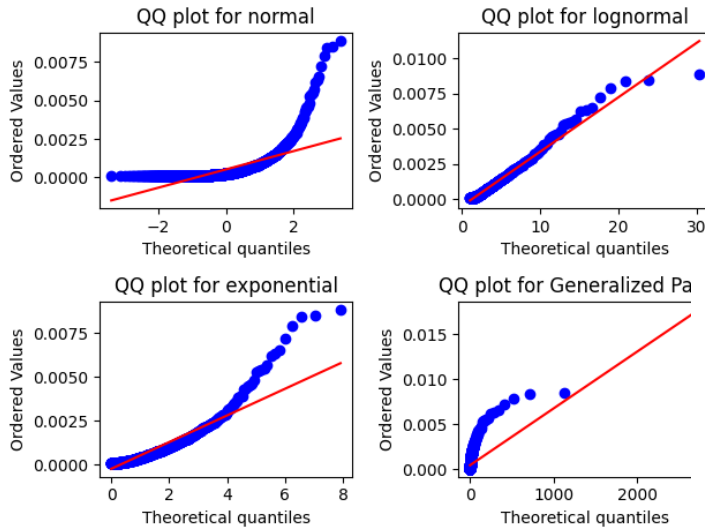


Рис. 3: QQ-plot для нормального распределения

Как видно нормальное распределение не подходит для выборки.

Как видно логнормальное распределение относительно подходит для выборки.

Как видно экспоненциальное распределение не подходит для выборки.

Как видно Парето обобщённое распределение не подходит для выборки.

Вывод: логнормальное распределение лучше всего описывает распределение PageRank графа.

### С. Оценки индекса экстремального значения

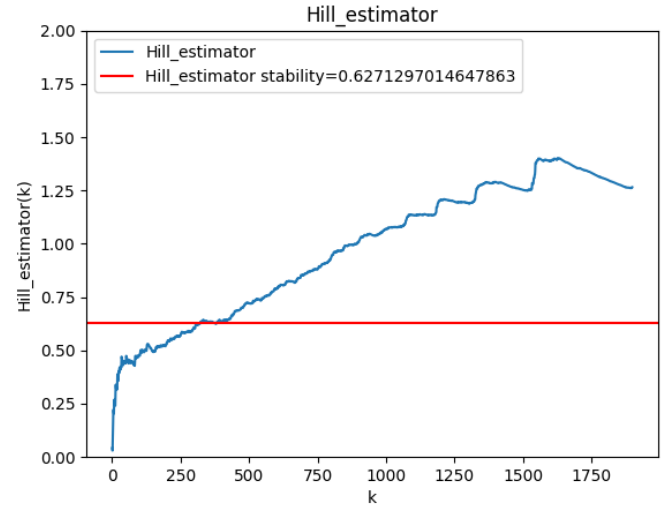


Рис. 4: Hills estimator plot

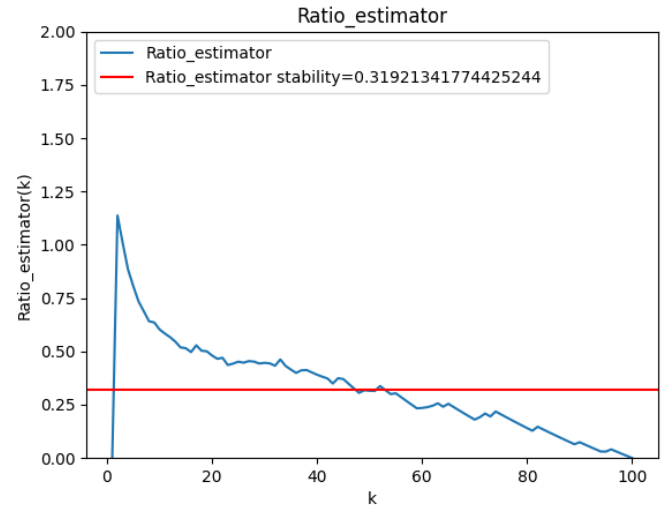


Рис. 5: Ratio estimator plot

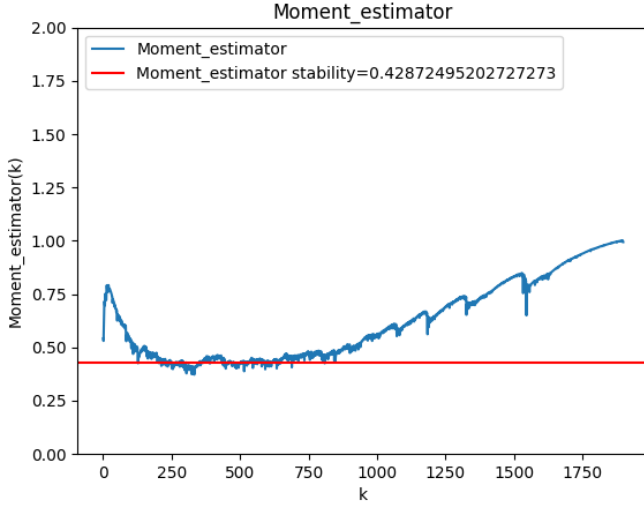


Рис. 6: Moment estimator plot

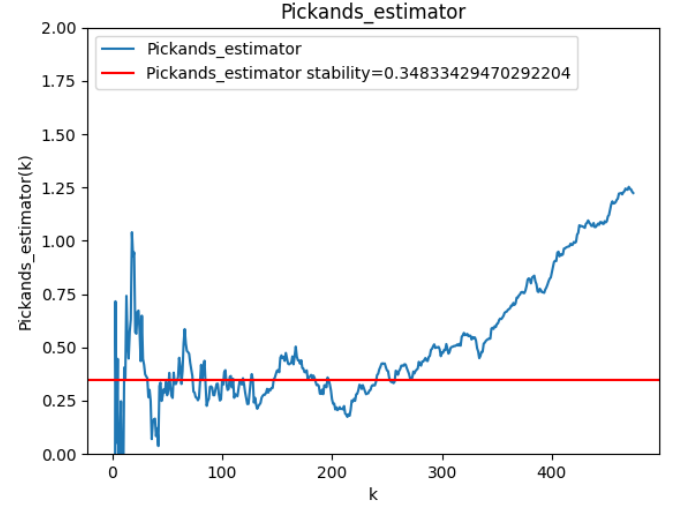


Рис. 8: Pickands estimator plot

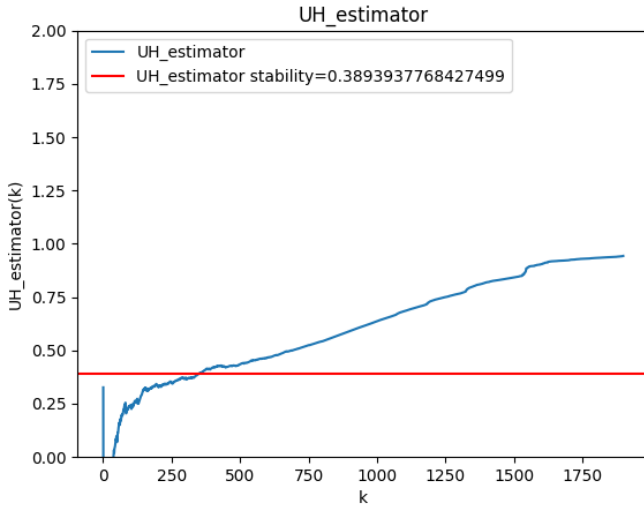


Рис. 7: UH estimator plot

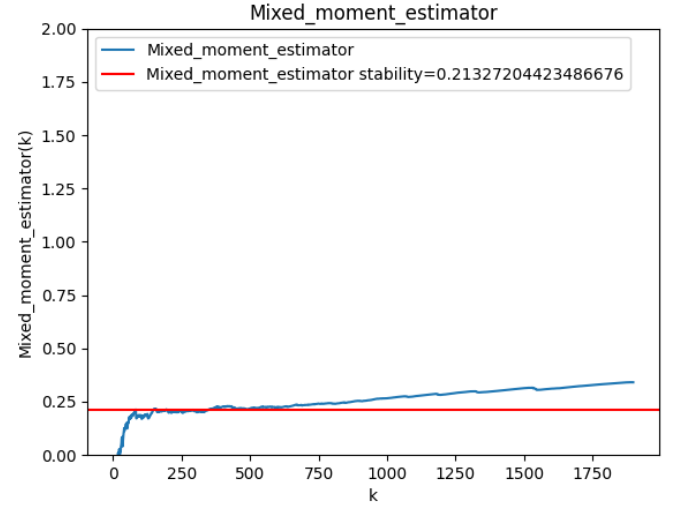


Рис. 9: Mixed estimator plot

Из графиков видно, что оценки Hill, Ratio и UH на предоставленных данных работают плохо. Тогда как оценки Moment, Pickands и Mixed Moment имеют интервалы стабильности. При взятии среднего по этим оценкам по интервалу стабильности получаются значения 0.42, 0.34 и 0.21 соответственно. Для дальнейших вычислений хвостового индекса будем использовать Mixed Moment estimator в связи с относительной стабильностью оценки на всей последовательности.

Вывод: по всей видимости хвостовой индекс можно положить примерно равным  $\gamma \approx 0.21$ .

## V. РАЗБИЕНИЕ НА СООБЩЕСТВА

Библиотека NetworkX позволяет разделить граф на сообщества посредством Louvain Community Detection

Algorithm. Выявлено 13 сообществ разного размера.

Номер сообщества	Количество вершин
0	125
1	2
2	236
3	367
4	122
5	41
6	331
7	248
8	29
9	287
10	107
11	2
12	2

## VI. ТЕСТИРОВАНИЕ СТАЦИОНАРНОСТИ

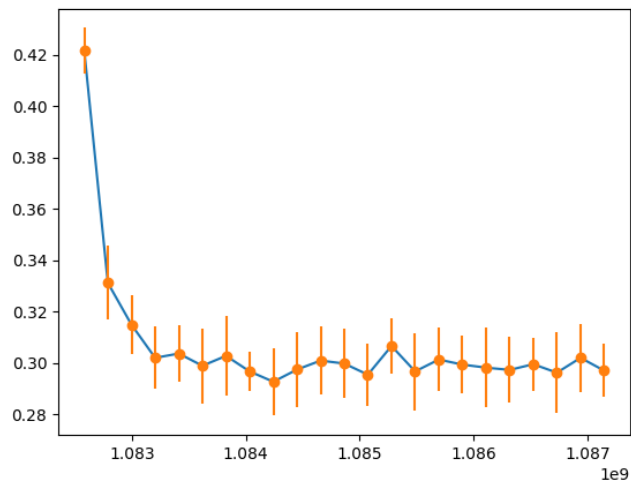


Рис. 10: Hills estimator plot

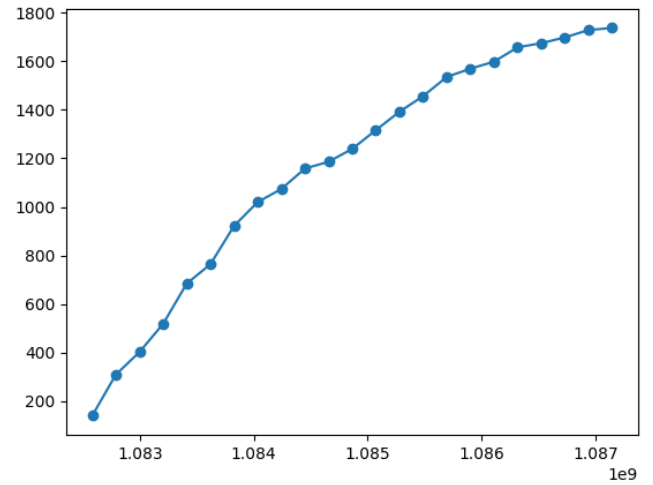


Рис. 11: Hills estimator plot

## VII. ЛИТЕРАТУРА

1) <https://networkrepository.com>