

# Тестирование стационарности сообществ в динамических реальных графах

Васильев Михаил Владимирович

11 мая 2024 г.

## 1 Введение

Рассматриваются эволюция реальных ненаправленных сетей и описывающих их графов. Обозначим граф на шаге эволюции  $t$ , как  $G_t = (V_t, E_t)$ ,  $V_t$ - множество вершин,  $E_t$ - множество ребер. Эволюция, т.е. присоединение новых узлов новыми связями к существующим узлам, начинается с начального графа  $G_0$  и имеет сложную природу. Существуют некоторые методы моделирования эволюции графа такие как предпочтительное присоединение (ПП, preferential attachment), кластерное присоединение (КП, clustering attachment) и их смеси. [9].

Изучаются методы разбиения реальных графов на сообщества, такие как алгоритм Leuven [8].

В работе данные взяты из открытого графового репозитория [2], а именно динамический граф fb-messages.csv. Граф составлен на основе сообщений студентов Калифорнийского университета в Ирвине в социальной сети, похожей на Facebook. 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.

Цель работы состоит в разработке и сравнении методов тестирования стационарности распределений сообществ при различных методах разбиения на сообщества.

## 2 Описательные характеристики

Зафиксируем состояние графа на последний момент времени, считаем для каждого его узла PageRank и проведём исследование рас-

спределения значений PageRank. Исследование состоит в описании распределения и оценивании индекса экстримальной величины  $\gamma$  разными методами. Методы оценивания включают в себя: оценку Хилла (Hill), оценку отношения (Ratio), оценку моментов (Moment) [10] и смешанную оценку моментов (Mixed Moment) [11]. Так же интерес представляет исследование тяжести хвоста распределения.

### 3 Тестирование стационарности

Пусть  $(V'_n, E'_n)$  и  $(V''_n, E''_n)$  два непересекающихся сообщества в графе  $(V_n, E_n)$  с соответствующим набором входящих ребер  $I'_n = \{I_n(v), v \in V'_n\}$  и  $I''_n = \{I_n(v), v \in V''_n\}$ . Положим нулевую гипотезу, как

$$H_0 : \alpha_1 = \alpha_2 = \alpha, \quad (1)$$

где  $\alpha_1$  и  $\alpha_2$  - хвостовые индексы входящих ребер в  $I'_n$  и  $I''_n$  соответственно. Модифицированная статистика Филлипса и Лоретана представлено в [12] и выглядит как:

$$S = \frac{k_1^*(\hat{\alpha}_2)^2(\hat{\alpha}_1/\hat{\alpha}_2 - 1)^2}{(\hat{\alpha}_1)^2 + (k_1^*/k_2^*)(\hat{\alpha}_2)^2} \quad (2)$$

Где  $\hat{\alpha}_1$  и  $\hat{\alpha}_2$  - значения хвостового индекса вычисленные по средствам оценки Хилла с оптимальным выбором наибольших порядковых статистик  $k_1^*$  и  $k_2^*$ . При нулевой гипотезе статистика  $S$  сходится в распределении к случайной величине, имеющей распределение хи-квадрат с одной степенью свободы. Например, при уровне значимости 5% областью отклонения будет интервал  $[3.841, +\infty]$ .

Можно рассмотреть PageRank узлов сообществ для ориентированных графов или степени узлов для неориентированных вместо входящих степеней, а затем проверить нулевую гипотезу для пар сообществ.

### 4 Выводы

Произведено описание динамического реального графа. Приведены некоторые подходы для проверки стационарности сообществ. В дальнейшем эти подходы будут применены к реальным данным.

## Список литературы

- [1] Райгородский А. М. Модели случайных графов и их применения // Труды МФТИ. 2010.
- [2] Dynamic Networks // Network Repository URL: <https://networkrepository.com> (дата обращения: 05.05.2024).
- [3] Маркович Н.М., Вайсиулюс М.Р. Extreme Value Statistics for Evolving Random Networks // Mathematics. 2023. 11(9). С. 2171 <https://www.mdpi.com/2227-7390/11/9/2171>. FRAGA ALVES M.I., GOMES M.I., DE HAAN L. Mixed moment estimator and location invariant alternatives //Extremes. – 2009 – No. 12 – P. 149–185.
- [4] Nicolas Dugué, Anthony Perez. Directed Louvain : maximizing modularity in directed networks. [Research Report] Université d’Orléans. 2015.
- [5] BEIRLANT J., GOEGEBEUR Y., TEUGELS J., SEGERS J. Applications. – Chichester, West Sussex: Wiley, 2004 – 504 p.
- [6] BAGROW J., BROCKMANN D. Natural Emergence of Clusters and Bursts in Network Evolution // Physical Review X. – 2012 –V. 3 – No. 2 –P. 21016
- [7] Leadbetter, M.R., Lingren, G. Rootz’n, H. (1983). Extremes and Related Properties of Random Sequence and Processes. ch.3, New York: Springer.
- [8] Nicolas Dugué, Anthony Perez. Directed Louvain : maximizing modularity in directed networks. [Research Report] Université d’Orléans. 2015.
- [9] ARNOLD N.A., MONDRAG Likelihood-based approach to discriminate mixtures of network models that vary in time // Sci. Rep. –2021. –No.11 –P. 5205
- [10] DEKKERS A. L. M., EINMAHL J. H. J., DE HAAN L. A Moment Estimator for the Index of an Extreme-Value Distribution // Ann. Statist. –1989. –No. 17 –P. 1833–1855
- [11] FRAGA ALVES M.I., GOMES M.I., DE HAAN L. Mixed moment estimator and location invariant alternatives //Extremes. – 2009 – No. 12 – P. 149–185.

- [12] Quintos, C., Fan Z. and P. C. B. Phillips. Structural Change Tests in Tail Behaviour and the Asian Crisis. *The Review of Economic Studies* 2001, 68, 633–663.