Тестирование стационарности сообществ в динамических случайных графах

Васильев Михаил Владимирович

11 мая 2024 г.

1 Введение

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

2 Задание графа

2.1 Задание изначального графа

Зададим случайный граф G_0 в модели Эрдёша - Реньи [1]. Пусть дано множество $V_n = \{1, ..., n\}$, элементы которого мы назовем вершинами. Именно на этом множестве мы будем "строить" случайный граф. Понятно, стало быть, что случайным будет множество ребер графа. Мы не хотим сейчас рассматривать графы с кратными ребрами (мультиграфы), графы с петлями (псевдографы) и ориентированные графы (орграфы). Поэтому мы считаем, что потенциальных ребер у графа не больше, чем $C_n^{(2)}$ штук. Будем соединять любые две вершины i и j ребром с некоторой вероятностью $p \in [0, 1]$ независимо от всех остальных $C_n^{(2)} - 1$ пар вершин. Иными словами, ребра появляются в соответствии со стандартной схемой Бернулли, в которой $C_n^{(2)} - 1$ испытаний и "вероятность успеха" p. Обозначим через E случайное множество ребер, которое возникает в результате реализации такой схемы. Положим $G_0 = (V_n, E)$.

2.2 Эволюция случайного графа

2.2.1 Обзор моделей эволюции случайного графа

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. Рост числа узлов и связей в реальных сетях происходит по различным моделям эволюции таким, как предпочтительное присоединение (ПП, preferential attachment), кластерное присоединение (КП, clustering attachment) и их смеси. Смена модели присоединения происходит из-за внешних факторов, например, появление новой области знаний в сети цитирования. Параметры модели могут меняться во времени. Поскольку присоединение новых узлов в моделях ПП и КП имеет вероятностную природу, можно определить G_t, как динамический случайный граф.

2.2.2 Preferential attachment

ПП моделирует присоединение новых узлов с большей вероятностью к наиболее влиятельным узлам, так как вероятность присоединения пропорциональна числу связей узла. Новый узел присоединяется к существующему узлу i ненаправленного графа с вероятностью:

$$P_{PA}(i,t) = \frac{d_{i,t}}{\sum\limits_{s \in V_t} d_{s,t}}, \tag{1}$$

где $d_{i,t}$, - число связей узла $i \in V_t$. ПП был предложен для обоснования тяжелых хвостов распределений характеристик влиятельности узлов таких, как число их связей и Пейджранги, что характерно для Веб-графов.

2.2.3 Clustering attachment

КП предложен в [6] для моделирования локальных сетей, когда присоединение нового узла происходит лишь к узлам, вовлеченным в треугольники. Например, в социальных сетях индивиды могут иметь друзей не среди популярных людей, а в своем окружении, формируя тесные сообщества. Вместо гигантских узлов с большими числом связей, как для ПП, КП ведет к тесно-связанным сообществам. В показано без доказательства, что КП порождает легкие хвосты распределений числа связей узлов. КП моделирует эволюцию социальных сетей и сетей, где новые узлы присоединяются не к узлам с большим числом связей, а к сильно связанным сообществам. Вероятность присоединения нового узла к узлу $i \in V_t$ определяется как

$$P_{CA}(i,t) = (C^{\alpha}_{i,t} + \varepsilon)/(\sum_{s \in V_t} c^{\alpha}_{s,t} + \epsilon \|V_t\|), \quad \alpha > 0, \tag{2}$$

$$P_{CA}(i,t) = \frac{\mathbb{1}\left\{c_{i,t} > 0\right\} + \epsilon}{\sum\limits_{j \in V_t} \mathbb{1}\left\{c_{j,t} > 0\right\} + \|V_t\|\epsilon}, \quad \alpha = 0, \tag{3}$$

где

$$c_{i,t} = \begin{cases} 0, & d_{i,t} = 0 \text{ или } d_{i,t} = 1, \\ 2 \bigtriangleup_{i,t} /(d_{i,t}(d_{i,t}-1)), & d_{i,t} \geq 2, \end{cases}$$

-кластерный коэффициент, $0 \leq c_{i,t} \leq 1$, $\bigtriangleup_{i,t}$ число треугольников, в которые вовлечен узел i на шаге эволюции t, $\alpha, \epsilon \geq 0$ - параметры модели. Модель КП не позволяет моделировать параллельные связи между парами узлов и петли узлов.

3 Выявление сообществ в графе

Сообщество - это группа тесно взаимодействующих узлов, которые слабо связаны с остальным графом.

4 Стационарность сообществ

4.1 Метод случайных последовательностей

Одним из подходов для тестирования сообществ на стационарность распределения может быть сведение к тестированию случайных последовательностей, возникающих на случайных графах. для этого могут быть использованы случайные блуждания для сбора информации о характеристиках влиятельности (например число входящих связей узла или PageRank) узлов сообщества, которые определяют соответствующие случайные последовательности.

4.2 Метод разбиения на подграфы

Ещё один подход - это разбиение сообщества на более мелкие подграфы, рассматривая их как блоки данных. В каждом блоке можно оценить индекс экстремальной величины или его обратную величину, хвостовой индекс, и понять, насколько сильно меняется его величина. Далее в статье будут описаны методы оценки индекса экстримальной величины.

5 Методы оценки индекса экстримальной величины

Методология оценивания индекса экстримальной величины состоит в следующем: рассматривается стационарный случайный граф G, для каждой его вершины V_i рассчитывается значение X_i, как число связей вершины или её PageRank. Таким образом последовательность значений $X_i, 1 \leq i \leq n$ пораждает некоторое распределение с тяжёлым или лёгким хвостом.

Для последовательности сл.в $\{X_n\}_{n \geq 1}$ обозначим соответствующие порядковые статистики, расположенные в порядке возрастания, как $X_{i,n}$

Оценку ИЭВ $\gamma \in R$ будем производить методом Mixed Moment estimator:

$$\hat{\gamma}^{MM}(n,k) = \frac{\hat{\varphi}_n(k) - 1}{1 + 2min(\hat{\varphi}_n(k) - 1, 0)}, \tag{4}$$

$$\hat{\varphi}_n(k) = \frac{M_n^{(1)}(k) - L_n^{(1)}(k)}{(L_n^{(1)}(k))^2}, \tag{5}$$

$$L_n^{(1)}(k) = 1 - \frac{1}{k}\sum_{i=1}^{k} \frac{X_{n-k,n}}{X_{n-i+1,n}}, \quad M_n^{(1)}(k) = \frac{1}{k}\sum_{i=1}^{k} ln\frac{X_{n-i+1,n}}{X_{n-k,n}}. \tag{6}$$

Оценка предназначена для $\gamma \in R$. $\gamma \leq 0$ указывает на лёгкий хвост распределения, а $\gamma > 0$ на тяжёлый по теореме Фишера-Типпета-Гнеденко, [5].

6 Выводы

На данный момент произведено изначальное описание методологии проверки стационарности сообществ.

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.

[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