Труды Карельского научного центра РАН. 2023. № 4. С. 50–53
Transactions of the Karelian Research Centre RAS. 2023. No. 4. P. 50–53
DOI: 10.17076/mat1765

ОРИГИНАЛЬНЫЕ СТАТЬИ
Original articles

УДК 519.212.2+519.179.4

О ГЛОБАЛЬНОМ КЛАСТЕРНОМ КОЭФФИЦИЕНТЕ ИНТЕРНЕТ-ГРАФА

Ю. Л. Павлов

Институт прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910)

Рассматривается конфигурационный граф с N вершинами, степени которых независимы и одинаково распределены по степенному закону, зависящему от медленно меняющейся функции. Конфигурационные графы широко используются для моделирования сложных сетей коммуникаций, таких как Интернет. Параметр степенного распределения обычно выбирается так, что распределение степеней имеет конечное математическое ожидание и бесконечную дисперсию. Важной характеристикой топологии случайного графа является глобальный кластерный коэффициент. Он измеряет, в какой степени соседи вершин сами могут быть соседями друг другу. При $N \to \infty$ доказана предельная теорема для кластерного коэффициента.

Ключевые слова: случайный конфигурационный граф; кластерный коэффициент; предельная теорема

Для цитирования: Павлов Ю. Л. О глобальном кластерном коэффициенте Интернет-графа // Труды Карельского научного центра РАН. 2023. № 4. С. 50–53. doi: 10.17076/mat1765

Финансирование. Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН).

Yu. L. Pavlov. ON THE INTERNET-GRAPH CLUSTERING COEFFICIENT

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia)

We consider a configuration graph with N vertices whose degrees are independent and identically distributed according to the power law depending on a slowly varying function. Configuration graphs are widely used for modeling complex communication networks such as the Internet. The parameter of the power-law distribution is usually selected so that the vertex degree distribution has a finite expectation and infinite variance. An important characteristic of the topology of a configuration graph is the global clustering coefficient. Clustering measures the extent to which neighbours of vertices are also each other's neighbours. We prove the limit theorem for the clustering coefficient as N tends to infinity.

Keywords: random configuration graph; clustering coefficient; limit theorem

F o r c i t a t i o n: Pavlov Yu. L. On the Internet-graph clustering coefficient. *Trudy Karel'skogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS.* 2023;4:50–53. doi: 10.17076/mat1765

F u n d i n g. The studies were funded from the federal budget through state assignment to the Karelian Research Centre RAS (Institute of Applied Mathematical Research, Karelian Research Centre RAS).

Введение

Конфигурационные графы часто используются в качестве моделей современных сложных сетей коммуникаций, таких как Интернет, системы мобильной связи, социальные сети и т. п. В таких графах предполагается, что степени вершин являются независимыми одинаково распределенными случайными величинами. Обозначим ξ случайную величину, равную степени любой вершины. Будем предполагать, что эта случайная величина имеет распределение

$$p_k = \mathbf{P}\{\xi = k\} = \frac{h(k)}{k^\tau}, \qquad (1)$$

где $k = 1, 2, \ldots$, $\tau > 1$, а $h(x)$ – медленно меняющаяся на бесконечности функция. Допустим также, что функция $h(x)$ локально ограничена, начиная с некоторого $X > 0$, т. е. при $x \in [X, \infty)$. Напомним, что функция называется локально ограниченной на некотором множестве, если она ограничена на любом компакте из этого множества [4].

Наблюдения за реальными сетями показали (см., например, [5]), что в большинстве случаев $\tau \in (2, 3)$, в том числе это верно и для сети Интернет, поэтому иногда такие модели называют Интернет-графами (например, в [3]). Степень каждой вершины равна числу выходящих из нее занумерованных полуребер. Ребра графа образуются путем попарного равновероятного соединения полуребер друг с другом. Сумма степеней вершин графа должна быть четной, поэтому в случае необходимости в граф вводится дополнительная вершина единичной степени, что не влияет на асимптотическое поведение основных числовых характеристик, если число вершин стремится к бесконечности. Очевидно, что такая конструкция графа допускает появление кратных ребер и петель.

В современной литературе, посвященной случайным графам, значительное внимание уделяется изучению их структуры (см. книгу [5] и библиографию в ней). Важными характеристиками структуры служат, в частности, различные кластерные коэффициенты. Обычно выделяют глобальный, локальный и средний кластерные коэффициенты [6, 7]. Рассмотрим конфигурационный граф $G = G(V, E)$, в котором множество V состоит из N вершин, а E – множество ребер. В статье [1] при $N \to \infty$ изучалось предельное поведение глобального кластерного коэффициента графа G, в котором распределение случайной величины ξ задавалось так, что при $k \to \infty$

$$h(k) \sim (\ln k)^{-\alpha}, \qquad (2)$$

где $\alpha \geqslant 0$. Легко видеть, что распределение со свойством (2) является частным случаем (1). В настоящей статье будет найдена асимптотика кластерного коэффициента в общем случае распределения (1), в котором $\tau \in (2, 3)$.

Дадим определение глобального кластерного коэффициента, следуя [5]. Обозначим i, j, t три разные вершины графа и будем отождествлять вершины с их номерами. Обозначим $S(i, j, t)$ событие, состоящее в том, что в E присутствуют ребра, соединяющие i с j и j с t. Тогда эти два ребра являются смежными и имеют общую вершину j. Пусть W_G означает число всех таких пар смежных ребер. Обозначим $I(U)$ индикатор события U. Тогда

$$W_G = \sum_{1 \leqslant i,j,t \leqslant N} I(S(i, j, t)) = 2 \sum_{1 \leqslant i<j<t \leqslant N} I(S(i, j, t)).$$

Пусть $T(i, j, t)$ – событие, состоящее в том, что произошло событие $S(i, j, t)$ и, кроме того, в E есть ребро, соединяющее i и t. Понятно, что событие $T(i, j, t)$ означает, что вершины i, j, t образуют треугольник. Тогда число треугольников в графе равно

$$\Delta_G = \sum_{1 \leqslant i,j,t \leqslant N} I(T(i, j, t)) = 6 \sum_{1 \leqslant i<j<t \leqslant N} I(T(i, j, t)).$$

Глобальным кластерным коэффициентом C_G графа G называется отношение утроенного числа различных треугольников к числу различных пар смежных ребер. Следовательно,

$$C_G = \frac{\Delta_G}{W_G}. \qquad (3)$$

Это равенство позволяет рассматривать кластерный коэффициент как вероятность того, что если ребро, соединяющее вершины i и j, смежно с ребром, соединяющим j и t, то вершины i и t тоже соединены ребром. Заметим, что число треугольников, образованных тремя выбранными вершинами, может быть больше одного, если ребра между вершинами окажутся кратными. Это число равно произведению кратностей всех ребер треугольника.

В следующем разделе будет доказана предельная теорема о поведении кластерного коэффициента при $N \to \infty$. В последнем разделе статьи обсуждается следствие этой теоремы.

Асимптотика глобального кластерного коэффициента

В дальнейшем нам потребуется знать предельное распределение максимальной степени вершины. Обозначим $\xi_1, \ldots, \xi_N$ случайные величины, равные степеням вершин $1, \ldots, N$ соответственно, и пусть

$$\xi_{(N)} = \max_{1 \leqslant i \leqslant N} \xi_i. \tag{4}$$

Справедливо следующее утверждение.

Лемма. *Пусть $N \to \infty$, z – фиксированное положительное число, а u – решение уравнения*

$$u = h\left(\left(\frac{Nu}{(\tau-1)z}\right)^{\frac{1}{\tau-1}}\right). \tag{5}$$

Тогда

$$\mathbf{P}\left\{\xi_{(N)} < \left(\frac{Nu}{(\tau-1)z}\right)^{\frac{1}{\tau-1}}\right\} \to e^{-z}.$$

Доказательство. Нетрудно видеть, что

$$\begin{aligned}\mathbf{P}\{\xi_{(N)} < x\} &= \mathbf{P}\{\xi_1 < x, \ldots, \xi_{(N)} < x\} \\ &= (1 - \mathbf{P}\{\xi_1 \geqslant x\})^N.\end{aligned} \tag{6}$$

Из (1) и [2, гл. VIII, § 9, теорема 1, а] следует, что

$$\mathbf{P}\{\xi_1 \geqslant x\} = \frac{h(x)}{(\tau-1)x^{\tau-1}}(1 + o(1)). \tag{7}$$

Из (5) и (6) следует утверждение леммы, если в (7) положить

$$x = \left(\frac{Nu}{(\tau-1)z}\right)^{\frac{1}{\tau-1}}.$$

$\square$

Поскольку $\tau \in (2,3)$, очевидно, что распределение (1) имеет конечное математическое ожидание:

$$m = \mathbf{E}\xi. \tag{8}$$

Пусть в графе G существует ребро, соединяющее вершины i и j, $i \neq j$. Степень вершины i задается случайной величиной ξ_i с распределением (1). Предположим, что степень вершины j равна $k+1$, при этом как минимум одно ребро соединяет i с j. Поскольку это ребро может быть образовано любым из $k+1$-го полуребра вершины j, в [7] доказано, с учетом (4) и (8), что распределение степени вершины j без учета этого ребра имеет вид:

$$q_k = \frac{(k+1)p_{k+1}}{m}, \quad k = 0, 1, \ldots, \xi_{(N)} - 1. \tag{9}$$

Далее заметим, что если вершины i и j имеют степени k_i и k_j соответственно, то число способов образовать соединяющие их ребра равно $k_i k_j$. Опираясь на это свойство, в [7] удалось доказать, что кластерный коэффициент (3) можно выразить следующим образом:

$$C_G = \frac{\left(\sum_{k=1}^{\xi_{(N)}-1} k q_k\right)^2}{Nm}. \tag{10}$$

Теперь мы можем доказать основной результат статьи, который сформулирован ниже в виде теоремы. В ней учтено, что, как следует из леммы, максимальная степень вершины в графе G пропорциональна $(Nu)^{1/(\tau-1)}$.

Теорема. *Пусть $N \to \infty$, $\tau \in (2,3)$, $\xi_{(N)} = v(Nu)^{1/(\tau-1)}$, $0 < v < \infty$. Тогда*

$$C_G \sim \left(\frac{(vu^{\frac{1}{\tau-1}})^{3-\tau} h(v(Nu)^{\frac{1}{\tau-1}})}{3-\tau}\right)^2 \frac{N^{\frac{7-3\tau}{\tau-1}}}{m^3}.$$

Доказательство. Из (1), (9) следует, что при достаточно большом $A > 0$

$$\sum_{k=1}^{\xi_{(N)}-1} k q_k = C + \frac{1+o(1)}{m} \sum_{k=A}^{\xi_{(N)}-1} \frac{h(k)}{k^{\tau-2}}, \tag{11}$$

где C – некоторая положительная постоянная. Из [4, утверждение 1.5.8] следует, что

$$\sum_{k=A}^{\xi_{(N)}-1} \frac{h(k)}{k^{\tau-2}} \sim \frac{(\xi_{(N)})^{3-\tau} h(\xi_{(N)})}{3-\tau}. \tag{12}$$

Хорошо известно свойство медленно меняющихся функций, состоящее в том, что при достаточно больших x и любом $\delta > 0$

$$x^{-\delta} < h(x) < x^{\delta}.$$

Используя это свойство, из (5) нетрудно получить, что для любого $\delta > 0$ существует $\theta = \theta(\delta)$ такое, что $-\delta < \theta < \delta$ и

$$u = N^{\theta}. \tag{13}$$

Поскольку $\xi_{(N)} = v(Nu)^{1/(\tau-1)}$, из (12), (13) видим, что при $N \to \infty$

$$\sum_{k=A}^{\xi_{(N)}-1} \frac{h(k)}{k^{\tau-2}} \to \infty,$$

поэтому из (10) и (11) получаем утверждение теоремы. □

Следствие теоремы

Сравнивая доказанную в предыдущем разделе теорему с утверждением 4 теоремы из статьи [1], приходим к выводу, что кластерный коэффициент C_G в графе с распределением (1) степеней вершин ведет себя аналогично кластерному коэффициенту графа, в котором медленно меняющаяся функция удовлетворяет соотношению (2). Это значит, что если $\tau > 7/3$, то $C_G \to 0$, а если $\tau < 7/3$, то $C_G \to \infty$. Эта особенность объясняется тем, что при малых значениях τ в графе образуется много кратных ребер и, следовательно, появится много дополнительных треугольников. Из доказанного следует, что поведение C_G зависит от медленно меняющейся функции $h(x)$ и, как в [1], для каждой $h(x)$ существует такое значение τ, при котором кластерный коэффициент стремится к константе. По-видимому, впервые такие свойства глобального кластерного коэффициента конфигурационного графа были замечены в [7], где распределение (1) степени любой вершины задавалось простейшим образом путем замены медленно меняющейся функции $h(x)$ на нормирующую константу $1/\zeta(\tau)$, где $\zeta(\tau)$ – значение дзета-функции Римана в точке τ.

Литература

1. *Павлов Ю. Л.* Об асимптотике кластерного коэффициента конфигурационного графа с неизвестным распределением степеней вершин // Информатика и ее применения. 2019. Т. 13, вып. 3. С. 9–13. doi: 10.14357/19922264190302

2. *Феллер В.* Введение в теорию вероятностей и ее применения. Т. 2. М.: Мир, 1984. 738 с.

3. *Alvares-Hamelin J. I., Dall'Astra L., Barrat A., Vespignani A.* K-core decomposition of Internet-graphs: hierarchies, self-similarity and measurement biases // Networks and heterogeneous media. 2008. Vol. 3, iss. 2. P. 371–393. doi: 10.3934/nhm.2008.3.371

4. *Bingman N. H., Goldie C. M., Teugels J. L.* Regular variation. Cambridge: Cambridge Univ. Press, 1987. 513 p. doi: 10.1017/CBO9780511721434

5. *Hofstad R.* Random graphs and complex networks. Vol. One. Cambridge: Cambridge Univ. Press, 2017. 337 p. doi: 10.1017/9781316779422

6. *Newman M. E. J.* Networks. An Introduction. Oxford: Oxford Univ. Press, 2010. 772 p. doi: 10.1162/artl_r_00062

7. *Newman M. E. J.* The structure and function of complex networks // SIAM Rev. 2003. Vol. 45, iss. 2. P. 167–256. doi: 10.1137/S003614450342480

References

1. *Pavlov Yu. L.* On the asymptotics of clustering coefficient in a configuration graph with unknown distribution of vertex degrees. *Informatika i ee primeneniya = Informatics and applications.* 2019;13(3):9–13. doi: 10.14357/19922264190302 (In Russ.)

2. *Feller W.* An introduction to probability theory and its applications. Vol. 2. New York: John Wiley; 1991. 704 p.

3. *Alvares-Hamelin J. I., Dall'Astra L., Barrat A., Vespignani A.* K-core decomposition of Internet-graphs: hierarchies, self-similarity and measurenment biases. *Networks and heterogeneous media.* 2008;3(2):371–393. doi: 10.3934/nhm.2008.3.371

4. *Bingman N. H., Goldie C. M., Teugels J. L.* Regular variation. Cambridge: Cambridge Univ. Press; 1987. 513 p. doi: 10.1017/CBO9780511721434

5. *Hofstad R.* Random graphs and complex networks. Vol. One. Cambridge: Cambridge Univ. Press; 2017. 337 p. doi: 10.1017/9781316779422

6. *Newman M. E. J.* Networks. An Introduction. Oxford: Oxford Univ. Press; 2010. 772 p. doi: 10.1162/artl_r_00062

7. *Newman M. E. J.* The structure and function of complex networks. *SIAM Review.* 2003;45(2): 167–256. doi: 10.1137/S003614450342480

Поступила в редакцию / received: 09.04.2023; принята к публикации / accepted: 12.05.2023.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Павлов Юрий Леонидович
д-р физ.-мат. наук, профессор, главный научный сотрудник

e-mail: pavlov@krc.karelia.ru

CONTRIBUTOR:

Pavlov, Yury
Dr. Sci. (Phys.-Math.), Professor, Chief Researcher