Труды Карельского научного центра РАН
№ 7. 2018. С. 57–61
DOI: 10.17076/mat780

УДК 519.179.4

О КЛАСТЕРИЗАЦИИ УСЛОВНОГО КОНФИГУРАЦИОННОГО ГРАФА

Ю. Л. Павлов

Институт прикладных математических исследований КарНЦ РАН, ФИЦ «Карельский научный центр РАН», Петрозаводск, Россия

Рассматриваются конфигурационные графы с N вершинами. Степени вершин являются независимыми одинаково распределенными ограниченными случайными величинами. Они равны числу занумерованных в произвольном порядке полуребер. Граф строится путем попарного равновероятного соединения полуребер для образования ребер. Изучается подмножество таких случайных графов при условии, что сумма степеней известна и равна n. Важной характеристикой топологии графа является локальный кластерный коэффициент. При $N, n \to \infty$ найдены предельные распределения локального кластерного коэффициента и числа треугольников одной вершины. Мы рассмотрели также предельное поведение их математических ожиданий.

Ключевые слова: случайный конфигурационный граф; условный граф; кластерный коэффициент; предельные теоремы.

Yu. L. Pavlov. ON CLUSTERING OF CONDITIONAL CONFIGURATION GRAPHS

We consider configuration graphs with N vertices. The degrees of the vertices are independent identically distributed limited random variables. They are equal to the number of vertex semiedges that are numbered in an arbitrary order. The graph is constructed by joining all of the semiedges pairwise equiprobably to form edges. We study the subset of such random graphs under the condition that the sum of vertex degrees is known and it is equal to n. An important characteristic of the topology of a graph is the local clustering coefficient. We obtained the limit distributions of the local clustering coefficient and the number of triangles of each vertex as N and n tend to infinity. We also considered the limit behaviour of their mathematical expectations.

Keywords: random configuration graph; conditional graph; clustering coefficient; limit theorems.

Введение

В работе [4] были введены конфигурационные графы, широко использующиеся в настоящее время для моделирования сложных сетей коммуникаций (см., например, [5]). Мы будем рассматривать конфигурационные графы с N вершинами, степени которых являются независимыми одинаково распределенными случайными величинами $\xi_1, \ldots, \xi_N$. Обозначим ξ случайную величину, равную степени любой вершины графа, и пусть

$$p_k = \mathbf{P}\{\xi = k\}, \quad k = 1, 2, \ldots \qquad (1)$$

Степень вершины равна числу выходящих из нее различимых полуребер. Понятно, что сумма степеней вершин любого графа должна быть четной, поэтому в случае необходимости в граф вводится дополнительная вершина единичной степени. В статье [6] отмечается, что появление такой вспомогательной вершины не влияет на асимптотические свойства графа при стремящемся к бесконечности числе вершин. Граф строится путем попарного равновероятного соединения полуребер друг с другом для образования ребер. Легко видеть, что такие графы могут содержать петли и кратные ребра.

Далее мы будем изучать условные конфигурационные графы при условии, что сумма степеней известна и равна n. Такие условные графы впервые рассматривались в [2]. В последние годы опубликован ряд работ, в которых исследовалось предельное поведение степенной структуры условных конфигурационных графов при различном характере стремления N и n к бесконечности (см. [1, 3] и библиографию в них).

Важной характеристикой топологии графа является кластерный коэффициент, впервые введенный в [7]. Локальным кластерным коэффициентом вершины i, имеющей степень $d_i, 1 \leqslant i \leqslant N$, называется величина

$$C(i) = \frac{2T_i}{d_i(d_i-1)}, \tag{2}$$

где T_i – число пар соседей вершины i, соединенных друг с другом хотя бы одним ребром. Понятно, что под соседом вершины понимается любая другая смежная с ней вершина. Заметим, что T_i равно числу различных треугольников, одной из вершин которых служит вершина i, а стороны – ребра графа, которые могут быть и кратными. Коэффициент $C(i)$ можно рассматривать как вероятность того, что два различных соседа вершины i тоже являются соседями друг другу. Кроме локального кластерного коэффициента (2) используется также средний локальный кластерный коэффициент:

$$C = \frac{1}{N}\sum_{i=1}^{N} C(i). \tag{3}$$

В данной статье при $N, n \to \infty$ изучается предельное поведение $T_i, C(i)$ и C. В теореме 1 найдены асимптотики распределения T_i, и $\mathbf{E}T_i$, а в теореме 2 рассматриваются $C(i)$ и $\mathbf{E}C$. В следующем разделе приводятся формулировки теорем 1 и 2, а в последнем разделе статьи собраны доказательства этих теорем. Полученные результаты являются только первым шагом в исследовании процессов кластеризации в условных конфигурационных графах.

Результаты

Введем обозначения:

$$m = \mathbf{E}\xi, \quad \sigma^2 = \mathbf{D}\xi. \tag{4}$$

Ниже доказаны следующие утверждения.

Теорема 1. *Пусть $N, n \to \infty, n > N, \max\xi = O(1)$. Тогда для любого $i, 1 \leqslant i \leqslant N$, справедливы следующие соотношения:*

$$\mathbf{P}\{T_i = 0\} =$$
$$= 1 - \frac{(m-1)^2(\sigma^2+m^2-m)}{2n^3}\left(1 + O\left(\frac{1}{n}\right)\right);$$
$$\mathbf{P}\{T_i = 1\} =$$
$$= \frac{(m-1)^2(\sigma^2+m^2-m)}{2n^3}\left(1 + O\left(\frac{1}{n}\right)\right);$$
$$\mathbf{P}\{T_i = k\} = O(n^{-3k}), \quad k > 1;$$
$$\mathbf{E}T_i = \frac{(m-1)^2(\sigma^2+m^2-m)}{2n^3}\left(1 + O\left(\frac{1}{n}\right)\right).$$

Теорема 2. *Пусть выполнены условия теоремы 1. Тогда для любого $i, 1 \leqslant i \leqslant N$, и всех $j, 2 \leqslant j \leqslant \max\xi$, справедливы следующие соотношения:*

$$\mathbf{P}\{C(i) = 0\} =$$
$$= 1 - \frac{(m-1)^2(\sigma^2+m^2-m)}{2n^3}\left(1 + O\left(\frac{1}{n}\right)\right);$$
$$\mathbf{P}\left\{C(i) = 1/\binom{j}{2}\right\} =$$
$$= \binom{j}{2}\frac{(m-1)^2}{n^3}p_j\left(1 + O\left(\frac{1}{n}\right)\right);$$
$$\mathbf{P}\left\{C(i) = k/\binom{j}{2}\right\} = O(n^{-6}), \quad 2 \leqslant k \leqslant \binom{j}{2};$$
$$\mathbf{E}C = \frac{(m-1)^2(1-p_1)}{n^3}\left(1 + O\left(\frac{1}{n}\right)\right).$$

Доказательства теорем

Пусть выполнены условия теоремы 1. По формуле полной вероятности

$$\mathbf{P}\{T_i = k\} = \sum_{j \geqslant 1}\mathbf{P}\{T_i = k | \xi_i = j\}p_j. \tag{5}$$

Очевидно, что

$$\mathbf{P}\{T_i = 0 | \xi_i = 1\} = 1. \tag{6}$$

Рассмотрим пару смежных ребер, соединенных вершиной i. Обозначим q вероятность события S, состоящего в том, что другие концы этих ребер являются смежными, т. е. соединенными хотя бы одним ребром. Пусть d_1 и d_2 – степени этих концевых вершин. Понятно, что

$$q = \sum_{j\geqslant 2} \sum_{l\geqslant 2} \mathbf{P}\{S|d_1 = j, d_2 = l\} \times \times \mathbf{P}\{d_1 = j, d_2 = l\}, \tag{7}$$

поскольку

$$\mathbf{P}\{S|d_1 = 1, d_2 = l\} = 0,$$

$$\mathbf{P}\{S|d_1 = j, d_2 = 1\} = 0.$$

Нетрудно видеть, что число всех возможных различных способов построения рассматриваемого графа равно $(n-1)(n-3)\cdot\ldots\cdot 3\cdot 1 = (n-1)!!$ и все такие реализации графа равновероятны. Поэтому

$$\mathbf{P}\{S|d_1 = 2, d_2 = 2\} = \frac{(n-7)!!}{(n-1)!!} = \frac{1}{(n-1)(n-3)(n-5)}.$$

Рассуждая аналогично, получаем, что

$$\mathbf{P}\{S|d_1 = 2, d_2 = 3\} = \frac{2}{(n-1)(n-3)(n-5)}$$

и

$$\mathbf{P}\{S|d_1 = 2, d_2 = l\} = \frac{l-1}{(n-1)(n-3)(n-5)}.$$

Тогда, в силу независимости степеней вершин,

$$\sum_{l\geqslant 2} \mathbf{P}\{S|d_1 = 2, d_2 = l\}\mathbf{P}\{d_1 = 2, d_2 = l\} = = \frac{p_2(p_2 + 2p_3 + 3p_4 + ...)}{(n-1)(n-3)(n-5)} = \frac{p_2(m-1)}{(n-1)(n-2)(n-5)}.$$

Таким же образом приходим к выводу, что

$$\sum_{l\geqslant 2} \mathbf{P}\{S|d_1 = j, d_2 = l\}\mathbf{P}\{d_1 = j, d_2 = l\} = = \frac{(j-1)p_j(m-1)}{(n-1)(n-2)(n-5)}.$$

Отсюда и из (1), (4) и (7) следует:

$$q = \frac{(m-1)(p_2 + 2p_3 + \ldots)}{(n-1)(n-3)(n-5)} = = \frac{(m-1)^2}{(n-1)(n-3)(n-5)}. \tag{8}$$

Вернемся к случайной величине T_i. Если степень вершины i равна двум, то у этой вершины может быть петля. Если это не так, то возможны два случая: соседи вершины i могут быть или не быть смежными. Поэтому

$$\mathbf{P}\{T_i = 0|\xi_i = 2\} = \frac{(n-3)!!}{(n-1)!!} + + \frac{(n-2)(n-3)!!}{(n-1)!!}(1-q) = 1 - q + \frac{q}{n-1}.$$

Отсюда и из (8) получаем, что

$$\mathbf{P}\{T_i = 0|\xi_i = 2\} = = \frac{(m-1)^2}{n^3}\left(1 + O\left(\frac{1}{n}\right)\right). \tag{9}$$

Далее находим:

$$\mathbf{P}\{T_i = 0|\xi_i = 3\} = \frac{\binom{3}{2}}{n-1} + + \frac{(n-3)(n-4)(n-5)!!}{(n-1)!!}(1-q)^{\binom{3}{2}} = = 1 - \binom{3}{2}\frac{(m-1)^2}{n^3}\left(1 + O\left(\frac{1}{n}\right)\right). \tag{10}$$

Здесь и далее мы используем факт, что, в силу (8), для любого натурального $t > 2$

$$(1-q)^{\binom{t}{2}} = 1 - \binom{t}{2}q + O(n^{-6}).$$

Если $\xi_i = 4$, то вершина i может иметь две петли, или одну петлю, или ни одной. Учитывая это, с помощью (8) выводим:

$$\mathbf{P}\{T_i = 0|\xi_i = 4\} = = 1 - \binom{4}{2}\frac{(m-1)^2}{n^3}\left(1 + O\left(\frac{1}{n}\right)\right). \tag{11}$$

Повторяя эти рассуждения, из (6), (8)–(11) и условия $\max\xi = O(1)$ по индукции заключаем:

$$\mathbf{P}\{T_i = 0|\xi_i = j\} = = 1 - \binom{j}{2}\frac{(m-1)^2}{n^3}\left(1 + O\left(\frac{1}{n}\right)\right), \tag{12}$$

поэтому из (1), (4) и (5) при $k = 0$ следует:

$$\mathbf{P}\{T_i = 0\} =$$

$$= 1 - \sum_{j\geqslant 1} \binom{j}{2} \frac{(m-1)^2}{n^3} \left(1 + O\left(\frac{1}{n}\right)\right) p_j = \tag{13}$$

$$= 1 - \frac{(m-1)^2(\sigma^2 + m^2 - m)}{2n^3} \left(1 + O\left(\frac{1}{n}\right)\right),$$

что и доказывает первое утверждение теоремы 1.

Положим теперь $k = 1$ в (5). Очевидно, что

$$\mathbf{P}\{T_i = 1 | \xi_i = 1\} = 0. \tag{14}$$

Легко видеть также, что, в силу (8),

$$\mathbf{P}\{T_i = 1 | \xi_i = 2\} = \frac{(n-2)(n-3)!!}{(n-1)!!} q =$$

$$= q - \frac{q}{n-1} = \frac{(m-1)^2}{n^3} \left(1 + O\left(\frac{1}{n}\right)\right),$$

$$\mathbf{P}\{T_i = 1 | \xi_i = 3\} = \binom{3}{2} \frac{(n-4)(n-3)!!}{(n-1)!!} \times$$

$$\times\, q(1-q)^2 = \binom{3}{2} \frac{(m-1)^2}{n^3} \left(1 + O\left(\frac{1}{n}\right)\right)$$

и, в более общем виде, с учетом (14), при $j \geqslant 1$

$$\mathbf{P}\{T_i = 1 | \xi_i = j\} = \tag{15}$$

$$= \binom{j}{2} \frac{(m-1)^2}{n^3} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Тогда, согласно (5),

$$\mathbf{P}\{T_i = 1\} = \tag{16}$$

$$= \frac{(m-1)^2(\sigma^2 + m^2 - m)}{2n^3} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Следуя той же схеме доказательства, получаем, что

$$\mathbf{P}\{T_i = 2 | \xi_i = 3\} =$$

$$= \frac{(n-4)(n-3)!!}{(n-1)!!} q^2 = O\left(\frac{1}{n^6}\right)$$

и, аналогично, для всех $j > 3$

$$\mathbf{P}\{T_i = 2 | \xi_i = j\} = O\left(\frac{1}{n^6}\right). \tag{17}$$

Поэтому

$$\mathbf{P}\{T_i = 2\} = O\left(\frac{1}{n^6}\right).$$

Общий случай, как легко видеть, теперь выглядит так:

$$\mathbf{P}\{T_i = k\} = O(n^{-3k}), \quad k \geqslant 2, \tag{18}$$

что и приводит к третьему утверждению теоремы. Последнее утверждение очевидным образом следует из равенства

$$\mathbf{E}T_i = \sum_{k\geqslant 0} k\mathbf{P}\{T_i = k\}$$

и из (13), (16) и (18). Теорема 1 доказана.

Пусть выполнены условия теоремы 2. Из (12) следует, что

$$\mathbf{P}\{C(i) = 0\} = \sum_{j\geqslant 1} \mathbf{P}\{T_i = 0 | \xi_i = j\} p_j =$$

$$= \sum_{j\geqslant 1} \left(1 - \binom{j}{2} \frac{(m-1)^2}{n^3} + O\left(\frac{1}{n^4}\right)\right) p_j = \tag{19}$$

$$= 1 - \frac{(m-1)^2(\sigma^2 + m^2 - m)}{2n^3} \left(1 + O\left(\frac{1}{n}\right)\right),$$

поэтому первое утверждение теоремы доказано. Используя (2), (15) и (18), легко понять, что

$$\mathbf{P}\{C(i) = 1\} = \sum_{j\geqslant 2} \mathbf{P}\left\{T_i = \binom{j}{2} \middle| \xi_i = j\right\} p_j = \tag{20}$$

$$= \frac{(m-1)^2}{n^3} \left(1 + O\left(\frac{1}{n}\right)\right) p_2 + O\left(\frac{1}{n^9}\right).$$

Опять используя (15), получаем, что

$$\mathbf{P}\left\{C(i) = 1/\binom{3}{2}\right\} = \mathbf{P}\{T_i = 1 | \xi_i = 3\} p_3 =$$

$$= \binom{3}{2} \frac{(m-1)^2}{n^3} p_3 \left(1 + O\left(\frac{1}{n}\right)\right)$$

и, обобщая на все $j, 3 \leqslant j \leqslant \max \xi$,

$$\mathbf{P}\left\{C(i) = 1/\binom{j}{2}\right\} = \tag{21}$$

$$= \binom{j}{2} \frac{(m-1)^2}{n^3} p_j \left(1 + O\left(\frac{1}{n}\right)\right),$$

откуда и вытекает второе утверждение теоремы. Третье очевидным образом следует из (17). Переходим наконец к оценке математического ожидания $C(i)$ с помощью (19)–(21):

$$\mathbf{E}C(i) = \sum_{j\geqslant 2} \frac{(m-1)^2}{n^3} p_j \left(1 + O\left(\frac{1}{n}\right)\right) =$$

$$= \frac{(m-1)^2(1-p_1)}{n^3} p_j \left(1 + O\left(\frac{1}{n}\right)\right).$$

Отсюда и из (3) очевидным образом следует последнее утверждение теоремы 2.

Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН) и при финансовой поддержке РФФИ (проект № 16-01-0005а).

Литература

1. *Павлов Ю. Л.* Условные конфигурационные графы со случайным параметром степенного распределения степеней // Математический сборник. 2018. Т. 209, вып. 2. С. 120–137. doi: 10.4213/sm8832

2. *Павлов Ю. Л., Чеплюкова И. А.* Случайные графы интернет-типа и обобщенная схема размещения // Дискретная математика. 2008. Т. 20, вып. 3. С. 3–18. doi: 10.4213/dm1008

3. *Павлов Ю. Л., Чеплюкова И. А.* Об асимптотике степенной структуры конфигурационных графов с ограничениями на число ребер // Дискретная математика. 2018. Т. 30, вып. 1. С. 21–39. doi: 10.4213/dm1445

4. *Bollobas B.* A probabilistic proof of an asymptotic formula for the number of labelled regular graphs // European J. Comb. 1980. Vol. 1, iss. 4. P. 311–316. doi: 10.1016/S0195-6698(80)80030-8

5. *Hofstad R.* Random Graphs and Complex Networks. Vol. 1. Cambridge University Press, 2017. 337 p.

6. *Reittu H., Norros I.* On the power-law random graph model of massive data networks // Performance Evaluation. 2004. Vol. 55, iss. 4. P. 3–23. doi: 10.1016/S0166-53/6(3)00097-x

7. *Watts D. J., Strogats S.* Collective dynamics of "small-world"networks // Nature. 1998. Vol. 393(6684). P. 440–442. doi: 10.1038/30918

Поступила в редакцию 19.02.2018

References

1. *Pavlov Yu. L.* Conditional configuration graphs with random parameter of the power-law degree distribution. *SB MATH*. 2018. Vol. 209. doi: 10.1070/sm8832 (in print).

2. *Pavlov Yu. L., Cheplyukova I. A.* Random graphs of Internet type and the generalised allocation scheme. *Discrete Mathematics and Applications*. 2008. Vol. 18, iss. 5. P. 447–464. doi: 10.1515/DMA.2008.033

3. *Pavlov Yu. L., Cheplyukova I. A.* Ob asimptotike stepennoi struktury konfiguratsionnykh grafov s ogranicheniyami na chislo reber [On the asymptotics of degree structure of configuration graphs with restrictions on the number of edges]. *Discrete Mathematics and Applications* (in print).

4. *Bollobas B.* A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Comb.* 1980. Vol. 1, iss. 4. P. 311–316. doi: 10.1016/S0195-6698(80)80030-8

5. *Hofstad R.* Random Graphs and Complex Networks. Vol. 1. Cambridge University Press, 2017. 337 p.

6. *Reittu H., Norros I.* On the power-law random graph model of massive data networks. *Performance Evaluation*. 2004. Vol. 55, no. 4. P. 3–23. doi: 10.1016/S0166-53/6(3)00097-x

7. *Watts D. J., Strogats S.* Collective dynamics of "small-world" networks. *Nature*. 1998. Vol. 393(6684). P. 440–442. doi: 10.1038/30918

Received February 19, 2018

СВЕДЕНИЯ ОБ АВТОРЕ:

Павлов Юрий Леонидович
главный научный сотрудник, д. ф.-м. н., проф.
Институт прикладных математических исследований КарНЦ РАН, Федеральный исследовательский центр «Карельский научный центр РАН»
ул. Пушкинская, 11, Петрозаводск,
Республика Карелия, Россия, 185910
эл. почта: pavlov@krc.karelia.ru
тел.: (8142) 781218

CONTRIBUTOR:

Pavlov, Yury
Institute of Applied Mathematical Research,
Karelian Research Centre,
Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk,
Karelia, Russia
e-mail: pavlov@krc.karelia.ru
tel.: (8142) 781218