

УДК 519.179.4

О МАКСИМАЛЬНОЙ МОДУЛЯРНОСТИ СЛУЧАЙНЫХ КОНФИГУРАЦИОННЫХ ГРАФОВ

Ю. Л. Павлов

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

Рассматриваются конфигурационные графы со случайными независимыми одинаково распределенными степенями вершин. Эти степени равны числу полуребер вершин, занумерованных в произвольном порядке. Граф строится путем попарного равновероятного соединения полуребер для образования ребер. Такие модели можно использовать для адекватного описания топологии транспортных, электрических, социальных сетей и Интернета. Важной характеристикой структуры графа является модулярность. Это мера кластеризации графа в случае разделения вершин на группы (кластеры). Графы с высокой модулярностью обладают высокой плотностью ребер между вершинами внутри кластеров, но слабыми связями между вершинами разных кластеров. В статье обсуждаются понятие модулярности и его свойства в случайных конфигурационных графах. Максимальная модулярность графа используется для описания уровня его кластеризации и для нахождения наилучшего разделения вершин. Доказана предельная теорема для максимальной модулярности при стремлении числа вершин к бесконечности.

Ключевые слова: случайный конфигурационный граф; модулярность; предельная теорема.

Yu. L. Pavlov. ON THE MAXIMUM OF THE MODULARITY OF RANDOM CONFIGURATION GRAPHS

Configuration graphs with random independent identically distributed vertex degrees are considered. The degrees are equal to the number of vertex semiedges that are numbered in an arbitrary order. The graph is constructed by joining all of semiedges pairwise equiprobably to form edges. Such models can be used to adequately describe the topology of transport, electricity, social networks and the Internet. An important characteristic of the structure of a graph is its modularity. It is a measure for graph clustering in the case vertices are divided into groups (clusters). Graphs with high modularity have dense edges between the vertices within clusters but sparse connections between vertices of different clusters. The notion of modularity and its properties in random configuration graphs are discussed. The maximum modularity of a graph is used to describe the level of graph clustering and to find the best division of vertices. The limit theorem for the maximum modularity as the number of vertices tends to infinity is proved.

Key words: configuration random graph; modularity; limit theorem.

ВВЕДЕНИЕ

В последние годы значительное внимание исследователей привлечено к проблеме моделирования сложных сетей коммуникаций, таких как Интернет, социальные, телефонные, транспортные, электрические сети и т. д. Естественно, что такими моделями часто служат графы, в которых вершины интерпретируются как узлы сети, а ребра означают связи между узлами. Эти модели позволяют изучать структуру и динамику развития сетей. Одной из важных возникающих здесь задач является обнаружение групп (сообществ) узлов, тесно связанных между собой, но значительно слабее взаимодействующих с узлами из других сообществ. Понятно, что сети могут отличаться друг от друга количеством и составом групп. Более того, даже внутри одной сети возможны разные разбиения узлов на сообщества. В связи с этим в [8] была предложена мера качества разбиения сети на сообщества, получившая название модулярности. Естественно считать, что максимальное значение модулярности графа, моделирующего сеть, соответствует оптимальному разбиению узлов на сообщества, наилучшим образом характеризующему групповой состав сети. Далее группы вершин такого графа, моделирующие соответствующие сообщества сети, будем называть кластерами. Большое число работ, начиная с [8, 9], посвящено исследованию и разработке алгоритмов нахождения оптимальных разбиений вершин графов на кластеры, обеспечивающих максимальные значения модулярности. В следующем разделе статьи дается определение модулярности графа. Далее это определение переносится на случайные графы, и формулируется в виде теоремы основной результат. Доказательству этой теоремы посвящен последний раздел статьи.

МОДУЛЯРНОСТЬ ГРАФА

Пусть граф G содержит N вершин. Обозначим $D = \{d_{ij}\}_{i,j=1}^N$ матрицу смежностей графа, где d_{ij} равно числу ребер, соединяющих вершины i и j . Мы рассматриваем графы, в которых могут быть кратные ребра и петли, что должно быть отражено в матрице D . Если число ребер, соединяющих вершины i и j ($i \neq j$), равно $s > 1$, то $d_{ij} = d_{ji} = s$. Если вершина i имеет петли, то d_{ii} равно удвоенному числу петель этой вершины. Таким образом, хотя каждую петлю можно рассматривать как одно ребро, она дает вклад, равный двум, в степень вершины, т. е. в число ребер, инцидентных этой вершине. Это значит, что сумма элементов каждой строки матрицы смеж-

ностей равна степени соответствующей вершины.

Обозначим k_i степень вершины i , $i = 1, \dots, N$, и пусть L равно числу ребер графа, тогда

$$\sum_{i=1}^N k_i = 2L. \quad (1)$$

Для описания процесса построения графа удобно использовать понятие полуребра, т. е. ребра, инцидентного данной вершине, но для которого смежная вершина еще не определена. Ясно, что степень каждой вершины равна числу выходящих из нее полуребер. Все полуребра различимы, например, занумерованы в произвольном порядке. Построение графа заключается в образовании ребер путем попарного соединения полуребер друг с другом в соответствии с некоторым правилом.

Если вершины графа образуют непересекающиеся кластеры, то можно оценить, насколько плотность связей внутри групп отличается от плотности межгрупповых связей. Идея измерения модулярности графа состоит в сравнении такой оценки с оценкой плотности связей в графе с теми же вершинами и их степенями, но при равновероятном соединении полуребер. Большая разница между такими оценками говорит о значительном отличии кластерной структуры рассматриваемого графа от структуры графа со случайным (равновероятным) образованием ребер.

В [8, 10] предложена следующая формула для вычисления модулярности Q графа с фиксированным разбиением вершин на кластеры:

$$Q = \frac{1}{2L} \sum_{i,j=1}^N \left(d_{ij} - \frac{k_i k_j}{2L} \right) \delta_{ij}, \quad (2)$$

где $\delta_{ij} = 1$, если вершины i и j принадлежат одному кластеру, и $\delta_{ij} = 0$ иначе.

Заметим, что $Q = 0$, если разбиения на кластеры нет, т. е. граф содержит единственный кластер, включающий все вершины. Заметим еще, что Q не превышает единицы. Так, например, если граф представляет собой совокупность кластеров, каждый из которых является паросочетанием с одним ребром, то, как следует из (2), с ростом числа таких паросочетаний модулярность неограниченно приближается к единице.

МОДУЛЯРНОСТЬ СЛУЧАЙНЫХ ГРАФОВ

Современными моделями сложных сетей в большинстве случаев являются случайные графы. Наиболее полно, по нашему мнению, соответствующая теория изложена в книге [6].

Вполне естественно, что представляет интерес задача изучения свойств модулярности в случайных графах. В предыдущем разделе обсуждалось понятие модулярности для фиксированного графа, но следует заметить, что интерпретации модулярности в фиксированных и случайных графах несколько отличаются. В последнем случае степени вершин могут быть случайными величинами и образование ребер также может происходить случайно. Поэтому, даже если зафиксировать степени вершин и составы кластеров, модулярность является случайной величиной и может принимать разные значения в зависимости от того, как образуются ребра. Отсюда следует, что область возможных значений модулярности зависит от выбора функций распределения степеней вершин и от алгоритмов образования ребер. Это позволяет при моделировании реальных сетей осуществлять выбор законов распределения и алгоритмов, наиболее подходящих для описания структуры моделируемой сети с помощью случайного графа. Данные о модулярности некоторых реальных сетей можно найти в литературе (см., например, [8, 10]).

В статье [5] свойства модулярности рассматривались для случайных графов Эрдеша–Реньи. Для них в [7] доказано несколько теорем о предельном поведении максимума модулярности при стремлении числа вершин к бесконечности. Хорошо известно (см., например, [6]), что случайные графы Эрдеша–Реньи не подходят для моделирования современных сложных сетей. Можно выделить два основных класса случайных графов, на основе которых строятся адекватные модели сетей коммуникаций. Это так называемые графы предпочтительного присоединения, идея которых была изложена в [2], и конфигурационные графы, введенные в [3]. Статья [11] содержит ряд результатов о максимальной модулярности регулярных графов и графов предпочтительного присоединения. Для их доказательства в некоторых случаях использовались также и свойства конфигурационных графов с фиксированными степенями вершин. Исследования многих авторов (см. [6]) показали, что сложные сети целесообразно моделировать с помощью конфигурационных графов, степени вершин которых являются независимыми одинаково распределенными случайными величинами, а ребра образованы путем попарного равновероятного соединения полуребер друг с другом. Поскольку сумма степеней вершин любого графа должна быть четной, в [12] предложено в случае нечетной суммы вводить в граф дополнительную вершину единичной

степени. При этом, согласно [12], появление такой вершины не влияет на асимптотические свойства графа, если число вершин стремится к бесконечности.

В настоящей статье, по-видимому, впервые рассматривается модулярность конфигурационных графов со случайными степенями вершин. Доказана теорема о предельном поведении максимальной модулярности таких графов. Доказательство теоремы приводится для случая четной суммы степеней вершин, но, как легко проверить, внося незначительные изменения в доказательство, утверждения теоремы остаются в силе и при появлении вспомогательной вершины в случае нечетной суммы степеней основных вершин.

ОСНОВНОЙ РЕЗУЛЬТАТ

Рассмотрим конфигурационный граф G с N вершинами. Обозначим ξ_1, \dots, ξ_N случайные величины, равные степеням вершин $1, \dots, N$. Все они независимы и одинаково распределены. Пусть

$$p_k = \mathbf{P}\{\xi = k\}, k = 1, 2, \dots; i = 1, \dots, N, \quad (3)$$

при этом распределение (3) имеет конечное математическое ожидание

$$m = \sum_{k=1}^{\infty} k p_k. \quad (4)$$

Введем производящую функцию

$$F(z) = \sum_{k=1}^{\infty} p_k z^k. \quad (5)$$

Далее нам понадобится начинающийся с одной частицы вспомогательный ветвящийся процесс Гальтона–Ватсона, в котором распределение числа прямых потомков начальной частицы совпадает с (3), но распределение случайной величины μ , равной числу прямых потомков всех остальных частиц, задано следующим образом:

$$\mathbf{P}\{\mu = k\} = \frac{(k+1)p_{k+1}}{m}, \quad k = 0, 1, 2, \dots$$

Математическое ожидание μ равно

$$M = \mathbf{E}\mu = \frac{1}{m} \sum_{k=2}^{\infty} k(k-1)p_k.$$

Обозначим g вероятность вырождения этого ветвящегося процесса. Рассмотрим множество Λ всех возможных разбиений вершин графа G на кластеры. Пусть $Q_\lambda(G)$ означает модулярность графа G с разбиением $\lambda \in \Lambda$, а Q_Λ^*

– максимальное значение модулярности среди всех возможных разбиений:

$$Q_{\Lambda}^*(G) = \max_{\lambda \in \Lambda} Q_{\lambda}(G).$$

В следующем разделе доказан такой результат.

Теорема. Пусть $N \rightarrow \infty$. Тогда асимптотически почти наверное справедливы следующие утверждения.

1. Если $M \leq 1$, то $Q_{\Lambda}^*(G) \rightarrow 1$.
2. Если $M > 1$, то $Q_{\Lambda}^*(G) \geq 1 - (1 - F(g))^2 + o(1)$.

Доказательство. Прежде всего заметим, что если $M \leq 1$, то вспомогательный ветвящийся процесс является докритическим или критическим. В этом случае, как хорошо известно, $g = 1$ и из (5) следует, что $1 - F(g) = 0$. Если же $M > 1$, то ветвящийся процесс надкритический, поэтому $g < 1$ и $1 - F(g) > 0$. Пусть $\lambda \in \Lambda$ – некоторое разложение вершин графа G , состоящее из l кластеров $\lambda_1, \dots, \lambda_l$. Обозначим $\zeta_N = \xi_1 + \dots + \xi_N$. В рассматриваемой модели модулярность $Q_{\lambda}(G)$, в силу (2), примет вид:

$$Q_{\lambda}(G) = \frac{1}{\zeta_N} \sum_{i,j=1}^N \left(\alpha_{ij} - \frac{\xi_i \xi_j}{\zeta_N} \right) \delta_{ij},$$

где $\alpha_{ij}, i, j = 1, \dots, N$, – случайные элементы матрицы смежностей. Это соотношение можно представить таким образом:

$$Q_{\lambda}(G) = \frac{2}{\zeta_N} \sum_{i=1}^l \beta(\lambda_i) - \frac{1}{\zeta_N^2} \sum_{i=1}^l \nu_i^2, \quad (6)$$

где $\beta(\lambda_i)$ и ν_i равны соответственно числу ребер и сумме степеней вершин кластера λ_i . Очевидно, что

$$\sum_{i=1}^l \nu_i = \zeta_N.$$

Построим вариационный ряд, расположив ν_1, \dots, ν_l в неубывающем порядке, и пусть $\nu_{(l-1)}$ и $\nu_{(l)}$ являются последними двумя членами этого ряда. Допустим, без ограничения общности, что $\nu_{(l)} = \nu_l$. Следуя идее доказательств лемм 2.1 и 2.2 работы [7], находим, что

$$\sum_{i=1}^{l-1} \nu_i^2 \leq (\zeta_N - \nu_l) \nu_{(l-1)}.$$

Отсюда следует, что

$$\frac{1}{\zeta_N^2} \sum_{i=1}^l \nu_i^2 \leq \left(\frac{\nu_l}{\zeta_N} \right)^2 + \frac{\nu_{(l-1)}}{\zeta_N}. \quad (7)$$

Обозначим Ω подмножество множества Λ всех возможных разбиений графа, состоящих только из связных компонент, и пусть $Q_{\Omega}^*(G)$ означает максимальное значение модулярности среди всех таких разбиений. Обозначим $\nu^{(1)}(\omega)$ и $\nu^{(2)}(\omega)$ соответственно максимальную и вторую по величине суммы степеней вершин компонент связности разбиения $\omega \in \Omega$. Поскольку

$$2 \sum_{i=1}^l \beta(\lambda_i) = \zeta_N,$$

из (6) и (7) вытекает, что для любого ω

$$Q_{\Lambda}^*(G) \geq 1 - (\nu^{(1)}(\omega)/\zeta_N)^2 - \nu^{(2)}(\omega)/\zeta_N. \quad (8)$$

Степень каждой вершины графа G имеет конечное математическое ожидание m , поэтому из локальной предельной теоремы 4.2.1 [1] следует, с учетом теоремы 2.1.1 для выбора нормирующих постоянных, что при $N \rightarrow \infty$

$$\zeta_N \sim mN. \quad (9)$$

Обозначим $V^{(1)}$ и $V^{(2)}$ соответственно максимальный объем (число вершин) и второй по величине объем компонент связности графа G . Пусть $M < 1$. Согласно теореме 3.3.2 [4]

$$V^{(1)} = O(\sqrt{N}). \quad (10)$$

Очевидно, что

$$Q_{\Lambda}^*(G) \geq Q_{\Omega}^*(G), \quad (11)$$

поэтому первое утверждение теоремы для случая $M < 1$ следует из (8)–(11). Если $M = 1$, то, как показано в [12, стр. 74–75],

$$V^{(1)} = O(N^{2/3}),$$

следовательно, из (8), (9) и (11) получаем, что первое утверждение теоремы верно и для этого случая.

Пусть теперь $M > 1$. Тогда вспомогательный ветвящийся процесс является надкритическим и $1 - F(g) > 0$. В теореме 3.2.2 [4] показано, что в этом случае граф имеет единственную гигантскую компоненту связности, объем которой при $N \rightarrow \infty$ эквивалентен $(1 - F(g))N$, при этом

$$V^{(2)} = O(\ln N).$$

Отсюда и из (8), (9), (11) легко следует и второе утверждение теоремы. \square

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

ЛИТЕРАТУРА

1. Ибрагимов И. А., Линник Ю. В. Независимые и стационарно связанные величины. М.: Наука, 1965. 524 с.
2. Barabasi A.-L., Albert R. Emergence of scaling in random networks // *Science*. 1999. Vol. 286. P. 509–512. doi: 10.1126/science286.5499.509
3. Bollobas B. A probabilistic proof of an asymptotic formula for the number of regular graphs // *European. J. Combin.* 1980. Vol. 1, iss. 4. P. 311–316. doi: 10.1016/S0195-6698(80)80030-8
4. Durrett R. *Random Graph Dynamics*. Cambridge: Cambridge Univ. Press, 2007. 223 p. doi: 10.1017/CBO9780511546594
5. Guimera R., Sales-Prado M., Amaral L. A. N. Modularity from fluctuation in random graphs and complex networks // *Physical Review E*70. 2004. 025101. doi: 10.1103/PhysRevE.70.025101
6. Hofstad R. *Random Graphs and Complex Networks*. Vol. One. Cambridge: Cambridge Univ. Press, 2017. 337 p. doi: 10.1017/9781316779422

REFERENCES

1. Ibragimov I. A., Linnik Yu. V. Independent and stationary sequences of random variables. Groningen: Wolters Neordhoff Publ., 1971. 438 p.
2. Barabasi A.-L., Albert R. Emergence of scaling in random networks. *Science*. 1999. Vol. 286. P. 509–512. doi: 10.1126/science286.5499.509
3. Bollobas B. A probabilistic proof of an asymptotic formula for the number of regular graphs. *European. J. Combin.* 1980. Vol. 1, iss. 4. P. 311–316. doi: 10.1016/S0195-6698(80)80030-8
4. Durrett R. *Random Graph Dynamics*. Cambridge: Cambridge Univ. Press, 2007. 223 p. doi: 10.1017/CBO9780511546594
5. Guimera R., Sales-Prado M., Amaral L. A. N. Modularity from fluctuation in random graphs and complex networks. *Physical Review E*70. 2004. 025101. doi: 10.1103/PhysRevE.70.025101
6. Hofstad R. *Random Graphs and Complex Networks*. Volume One. Cambridge: Cambridge Univ. Press, 2017. 337 p. doi: 10.1017/9781316779422
7. McDiarmid C., Sherman F. Modularity of Erdos-Renyi random graphs. *29th International*

7. McDiarmid C., Sherman F. Modularity of Erdos-Renyi random graphs // *29th Int. conf. on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis and Algorithms*. LIPIcs. 2018. Vol. 110. P. 31.1–31.18. doi: 10.4230/LIPIcs.AofA.2018.31

8. Newman M. E. J., Girvan M. Finding and evaluating community structure in networks // *Physical Review E*69. 2004. 026113. doi: 10.1103/PhysRevE.69.026113

9. Newman M. E. J. Fast algorithm for detecting community structure in networks // *Physical Review E*69. 2004. 066133. doi: 10.1103/PhysRevE.69.066133

10. Newman M. E. J. Modularity and community structure in networks // *PNAS*. 2006. Vol. 103, iss. 23. P. 8577–8582. doi: 10.1073/pnas.0602103

11. Prokhorenkova L., Pralat P., Raigorodskii A. V. Modularity in several random graph models // *Electronic Notes in Discrete Math*. 2017. Vol. 61. P. 941–953. doi: 10.1016/j.endm.2017.07.058

12. 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

Поступила в редакцию 12.01.2019

Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis and Algorithms. LIPIcs. 2018. Vol. 110. P. 31.1–31.18. doi: 10.4230/LIPIcs.AofA.2018.31

8. Newman M. E. J., Girvan M. Finding and evaluating community structure in networks. *Physical Review E*69. 2004. 026113. doi: 10.1103/PhysRevE.69.026113

9. Newman M.E.J. Fast algorithm for detecting community structure in networks. *Physical Review E*69. 2004. 066133. doi: 10.1103/PhysRevE.69.066133

10. Newman M. E. J. Modularity and community structure in networks. *PNAS*. 2006. Vol. 103, iss. 23. P. 8577–8582. doi: 10.1073/pnas.0602103

11. Prokhorenkova L., Pralat P., Raigorodskii A. V. Modularity in several random graph models. *Electronic Notes in Discrete Math*. 2017. Vol. 61. P. 941–953. doi: 10.1016/j.endm.2017.07.058

12. 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

Received January 12, 2019

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