

УДК 519.179.2 + 519.876.5

## ОБ АССОРТАТИВНОСТИ И КЛАСТЕРИЗАЦИИ СТЕПЕННОЙ МОДЕЛИ КОНФИГУРАЦИОННЫХ ГРАФОВ

М. М. Лери

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

Рассматриваются случайные конфигурационные графы с распределением степеней вершин, являющимся дискретным степенным распределением. Параметр распределения либо принимает фиксированные значения, либо является случайным, имеющим либо равномерное, либо усеченное нормальное распределение на заданном интервале  $(a, b)$ . Посредством имитационного моделирования проводится анализ ассортативности и кластеризации рассматриваемых графов. Находятся оценки таких числовых структурных характеристик, как коэффициенты ассортативности и кластеризации. Проводится сравнение полученных результатов с имеющимися в литературе данными для реальных сетей.

**Ключевые слова:** конфигурационный граф; степенное распределение; ассортативность; кластеризация; имитационное моделирование.

### M. M. Leri. ON ASSORTATIVITY AND CLUSTERING OF A POWER-LAW CONFIGURATION GRAPH MODEL

The paper deals with configuration graphs with vertex degrees distributed independently according to the power law. The distribution parameter is either fixed or randomly distributed according to either uniform or truncated normal distribution on a predefined interval  $(a, b)$ . The assortativity and clustering of these graphs are studied by computer simulation. The estimates of numerical structural characteristics are found, such as the assortativity and clustering coefficients. The obtained results are compared with those known from the literature for real world complex networks.

**Key words:** configuration graph; power-law; assortativity; clustering; simulations.

### ВВЕДЕНИЕ

С появлением, быстрым развитием и усложнением структуры глобальных сетей исследование случайных графов, как моделей, широко применяемых для описания таких объектов, приобретает всевозрастающую популярность. Рассмотрение структуры и функционирования реальных телекоммуникационных сетей показало (см., например, [6, 7]), что случайные графы, степени вершин которых являются независимыми одинаково распределенными случайными величинами с общим законом распределения, хорошо подходят для

их описания. Однако с ростом размерности таких сетей стало понятно, что при построении их моделей недостаточно учитывать только распределение степеней вершин в соответствующем случайном графе, но также необходимо принимать в рассмотрение и другие не менее важные числовые характеристики графовой структуры, такие как ассортативность, кластеризация, модулярность, центральность и пр.

В настоящей работе рассматриваются случайные графы, распределение степеней вершин которых является дискретным анало-

гом распределения Парето. Обозначим через  $\xi_1, \xi_2, \dots, \xi_N$  степени вершин случайного графа размерности  $N$ , представляющие собой независимые случайные величины со следующим общим распределением [11]:

$$P\{\xi = k\} = k^{-\tau} - (k+1)^{-\tau}, \quad k = 1, 2, \dots, \quad (1)$$

где  $\tau > 1$  является параметром распределения. Рассматриваемые в большинстве работ графы (см., например, [5, 7, 11]) имеют фиксированный параметр  $\tau$ , однако с ростом размеров глобальных сетей все большее внимание исследователей стали привлекать модели случайных графов, у которых распределения степеней вершин либо меняются с ростом размерности графа, либо являются случайными (см., например, [2, 3]). Поэтому в настоящей работе рассматриваются случайные графы, степени вершин которых имеют распределение (1) как с фиксированным параметром  $\tau$ , так и с параметром  $\tau$ , имеющим либо равномерное, либо усеченное нормальное распределение на заданном интервале  $(a, b)$ , где  $1 < a < b < \infty$  [1]. Вследствие того, что параметр  $\tau$  выбирается отдельно для каждой вершины из заданного распределения (равномерного или усеченного нормального), можно говорить о том, что эволюция графа происходит в случайной среде. Также немаловажное значение имеет собственно формирование графа. В работе рассматривается так называемая конфигурационная модель случайного графа, впервые предложенная Б. Боллобашем в [4]. Степень вершины графа определяет различимые полуредра [11], занумерованные в произвольном порядке. Далее все полуредра соединяются между собой попарно и равновероятно, образуя ребра графа. Ясно, что при таком построении сумма степеней вершин должна быть четной, поэтому в противном случае к равновероятно выбранной вершине добавляется одно недостающее полуредра, тем самым увеличив степень этой вершины на 1. Конфигурационные графы могут иметь кратные ребра, циклы и петли.

## ХАРАКТЕРИСТИКИ СТРУКТУРЫ

Наряду с рассмотрением распределения степеней узлов реальных сетей все большее внимание уделяется исследованиям других важных числовых характеристик сетевой структуры, отражающих как глобальные, так и локальные свойства сетей. В настоящей работе рассматриваются три таких характеристики: глобальный и локальный коэффициенты кластеризации графа, а также коэффициент ассортативности.

Коэффициент ассортативности  $A$  используется для оценки взаимозависимости между

степенями смежных вершин. В [10] было предложено использовать для этой цели коэффициент корреляции Пирсона. Таким образом, значение коэффициента  $A$  будет положительным в том случае, когда вершины с большими степенями соединяются ребрами в основном тоже с вершинами с большими степенями, и тогда граф считают «ассортативным». В противном случае значение  $A$  будет отрицательным и граф называют «дисассортативным». Для оценки степени кластеризации графа применяются соответствующие кластерные коэффициенты. В работе рассматриваются два таких коэффициента: глобальный кластерный коэффициент  $C_G$  [10]

$$C_G = \frac{3 \times \text{число треугольников}}{\text{число пар смежных ребер}}$$

и средний локальный кластерный коэффициент  $C_L$  [10]

$$C_L = \frac{1}{N} \sum_{i=1}^N C_i,$$

где

$$C_i = \frac{\text{число треугольников при вершине } i}{\text{число пар смежных ребер с центром в } i},$$

где «пара смежных ребер» означает, что вершина соединена ребрами с двумя другими вершинами [10]. Как уже было упомянуто выше, в конфигурационном графе могут быть петли и кратные ребра. При вычислении кластерных коэффициентов петли не учитываются, а кратное ребро считается за одно. Значение глобального кластерного коэффициента отражает вероятность того, что в равновероятно выбранной паре смежных ребер, имеющих общую вершину, другие концы ребер будут тоже соединены ребром.

## РЕЗУЛЬТАТЫ

Целью настоящей работы была оценка введенных выше структурных характеристик конфигурационных графов посредством методов имитационного моделирования и нахождения зависимостей этих характеристик от объема графа и параметра распределения степеней вершин. Как в случае фиксированного параметра  $\tau$  распределения (1), так и при случайном  $\tau$  (равномерно или нормально распределенном на заданном интервале  $(a, b)$ ) рассматривались графы размерности  $100 \leq N \leq 10000$  с шагом 500. В случае фиксированного  $\tau$  значения параметра изменялись от 1, 1 до 2, 5 с шагом 0, 1. На основе полученных статистических данных были построены следующие регрессионные зависимости: коэффициентов  $A$ ,

$C_G$  и  $C_L$  от  $N$  и  $\tau$  (рис. 1–3):

$$A = -0,581N^{-0,888+0,896/\tau}, \quad R^2 = 0,94,$$

$$C_G = 0,457N^{-1,215+1,091/\tau}, \quad R^2 = 0,93,$$

$$C_L = 0,394N^{-0,853+0,986/\tau}, \quad R^2 = 0,95.$$

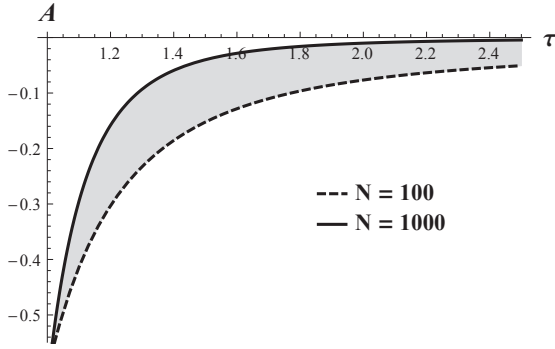


Рис. 1. Зависимость  $A$  от  $N$  и  $\tau$  для фиксированных значений  $\tau$

Fig. 1. Dependence of  $A$  on  $N$  and  $\tau$  for fixed  $\tau$

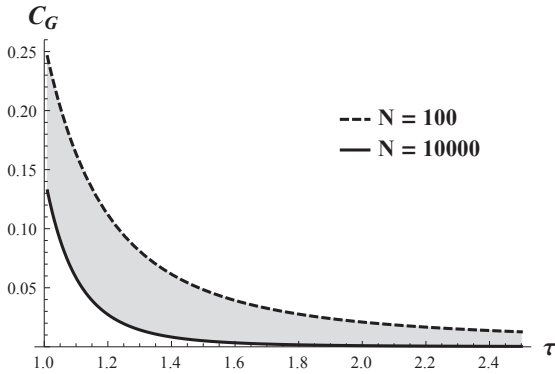


Рис. 2. Зависимость  $C_G$  от  $N$  и  $\tau$  для фиксированных значений  $\tau$

Fig. 2. Dependence of  $C_G$  on  $N$  and  $\tau$  for fixed  $\tau$

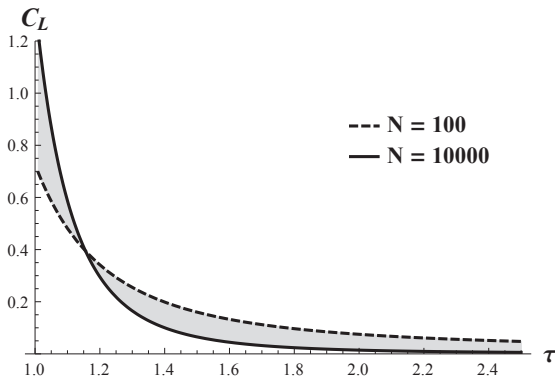


Рис. 3. Зависимость  $C_L$  от  $N$  и  $\tau$  для фиксированных значений  $\tau$

Fig. 3. Dependence of  $C_L$  on  $N$  and  $\tau$  for fixed  $\tau$

В случае, когда параметр  $\tau$  распределения (1) имел либо равномерное, либо усеченное нормальное распределение, рассматривались три интервала  $(a, b)$ :  $(1, 2)$ ,  $(2, 3)$  и  $(1, 3)$ . Выбор интервалов, как и ранее [1, 8, 9], был основан на том, что конфигурационные графы с параметром  $\tau \in (1, 2)$  считаются наиболее подходящими для описания сложных сетей телекоммуникаций [5, 7, 11], графы с  $\tau \in (2, 3)$  показывают устойчивость к «лесному пожару» [8, 9], а интервал  $(1, 3)$  был выбран как обобщающий. По полученным результатам были построены следующие регрессионные модели зависимостей коэффициентов  $A$ ,  $C_G$  и  $C_L$  от  $N$  (обозначения  $\tau \sim \mathbf{U}(a, b)$  и  $\tau \sim \mathbf{N}(a, b)$  означают, что параметр  $\tau$  имеет соответственно равномерное или усеченное нормальное распределение на  $(a, b)$ ):

$$\tau \sim \mathbf{U}(1, 2) : \quad A = -0,54 \cdot N^{-0,16},$$

$$\tau \sim \mathbf{U}[2, 3] : \quad A = -0,38 \cdot N^{-0,41},$$

$$\tau \sim \mathbf{U}(1, 3) : \quad A = -0,23 \cdot N^{-0,10},$$

$$\tau \sim \mathbf{N}(1, 2) : \quad A = -0,24 \cdot N^{-0,12},$$

$$\tau \sim \mathbf{N}(2, 3) : \quad A = -0,31 \cdot N^{-0,39},$$

$$\tau \sim \mathbf{N}(1, 3) : \quad A = -0,18 \cdot N^{-0,18}.$$

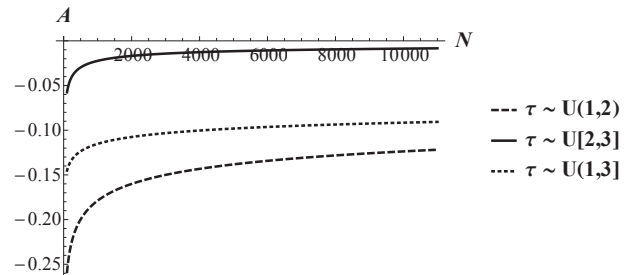


Рис. 4. Зависимость  $A$  от  $N$  при  $\tau \sim \mathbf{U}(a, b)$  для трех интервалов  $(a, b)$

Fig. 4. Dependence of  $A$  on  $N$  when  $\tau \sim \mathbf{U}(a, b)$  for the three intervals  $(a, b)$

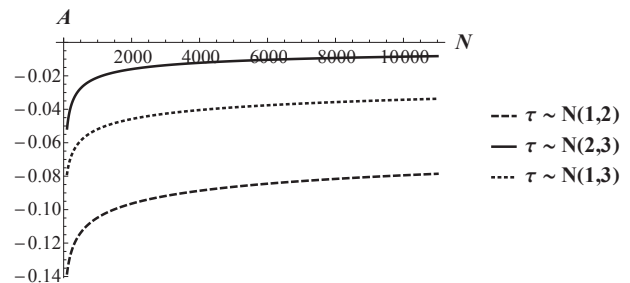


Рис. 5. Зависимость  $A$  от  $N$  при  $\tau \sim \mathbf{N}(a, b)$  для трех интервалов  $(a, b)$

Fig. 5. Dependence of  $A$  on  $N$  when  $\tau \sim \mathbf{N}(a, b)$  for the three intervals  $(a, b)$

Коэффициенты детерминации для всех моделей зависимости  $A$  от  $N$  были не менее 0,93.

Необходимо отметить, что коэффициент ассортативности  $A$  для всех рассмотренных видов конфигурационных графов является отрицательным, что означает, что данные графы можно использовать для моделирования только дисассортативных сетей. Заметим также (рис. 4 и 5), что самые большие по модулю значения коэффициент ассортативности принимает тогда, когда значения параметра  $\tau$  лежат в интервале (1, 2), а самые маленькие – в интервале (2, 3).

Для коэффициентов  $C_G$  и  $C_L$  были получены следующие их зависимости от  $N$ :

$$\begin{aligned} \tau \sim \mathbf{U}(1, 2) : & \quad C_G = 0,41 \cdot N^{-0,38}, \\ \tau \sim \mathbf{U}[2, 3] : & \quad C_G = 0,71 \cdot N^{-0,76}, \\ \tau \sim \mathbf{U}(1, 3) : & \quad C_G = 0,21 \cdot N^{-0,31}, \\ \tau \sim \mathbf{N}(1, 2) : & \quad C_G = 0,28 \cdot N^{-0,36}, \\ \tau \sim \mathbf{N}(2, 3) : & \quad C_G = 0,11 \cdot N^{-0,58}, \\ \tau \sim \mathbf{N}(1, 3) : & \quad C_G = 0,15 \cdot N^{-0,44}, \end{aligned}$$

$$\begin{aligned} \tau \sim \mathbf{U}(1, 2) : & \quad C_L = 0,20 \cdot N^{-0,10}, \\ \tau \sim \mathbf{U}[2, 3] : & \quad C_L = 0,75 \cdot N^{-0,96}, \\ \tau \sim \mathbf{U}(1, 3) : & \quad C_L = 0,11 \cdot N^{-0,05}, \\ \tau \sim \mathbf{N}(1, 2) : & \quad C_L = 0,23 \cdot N^{-0,14}, \\ \tau \sim \mathbf{N}(2, 3) : & \quad C_L = 0,14 \cdot N^{-0,71}, \\ \tau \sim \mathbf{N}(1, 3) : & \quad C_L = 0,12 \cdot N^{-0,27} \end{aligned}$$

с коэффициентами детерминации  $R^2 \geq 0,92$  для всех моделей. Результаты показывают, что значения глобального кластерного коэффициента  $C_G$  при фиксированных  $\tau$  всегда меньше значений локального  $C_L$ . Однако, когда параметр  $\tau$  случаен и распределен равномерно или нормально на интервале (2, 3), значения коэффициента  $C_L$  оказываются меньше значений  $C_G$  (рис. 7 и 10). Кроме того, наибольшие значения оба коэффициента принимают в случае, когда значения параметра  $\tau$  лежат в интервале (1, 2), а при  $\tau \in (2, 3)$  коэффициенты  $C_G$  и  $C_L$  принимают самые маленькие значения.

Предложенные в этой работе модели могут быть использованы при моделировании реальных сетей, когда необходимо учесть не только распределение степеней вершин, но и другие числовые характеристики сети, такие как ее ассортативность и кластеризация.

Кроме того, сравнение полученных значений коэффициентов  $A$ ,  $C_G$  и  $C_L$  для конфигурационных графов со значениями, вычисленными для некоторых реальных сетей [10], по-

казало, что графы с параметром распределения степеней вершин  $1,02 \leq \tau \leq 1,17$  наилучшим образом подходят для моделирования сети Интернет (когда узлы сети – автономные системы), а для моделирования некоторых дисассортативных социальных сетей лучше использовать графы с  $2,2 \leq \tau \leq 2,6$ .

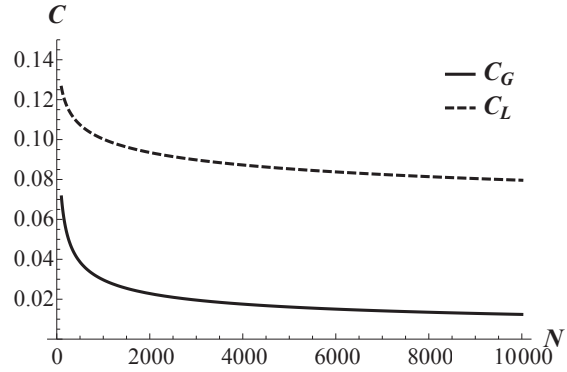


Рис. 6. Зависимости  $C_G$  и  $C_L$  от  $N$  при  $\tau \sim \mathbf{U}(1, 2)$   
Fig. 6. Dependencies of  $C_G$  and  $C_L$  on  $N$  when  $\tau \sim \mathbf{U}(1, 2)$

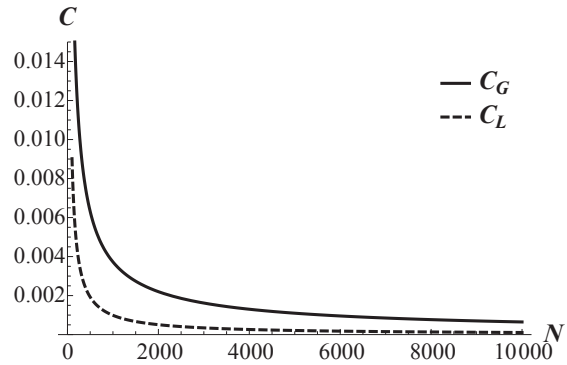


Рис. 7. Зависимости  $C_G$  и  $C_L$  от  $N$  при  $\tau \sim \mathbf{U}[2, 3]$   
Fig. 7. Dependencies of  $C_G$  and  $C_L$  on  $N$  when  $\tau \sim \mathbf{U}[2, 3]$

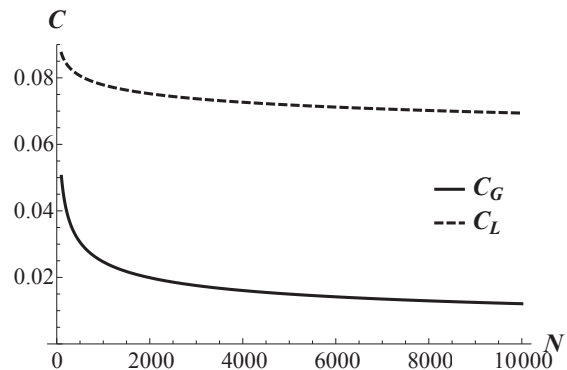


Рис. 8. Зависимости  $C_G$  и  $C_L$  от  $N$  при  $\tau \sim \mathbf{U}(1, 3)$   
Fig. 8. Dependencies of  $C_G$  and  $C_L$  on  $N$  when  $\tau \sim \mathbf{U}(1, 3)$

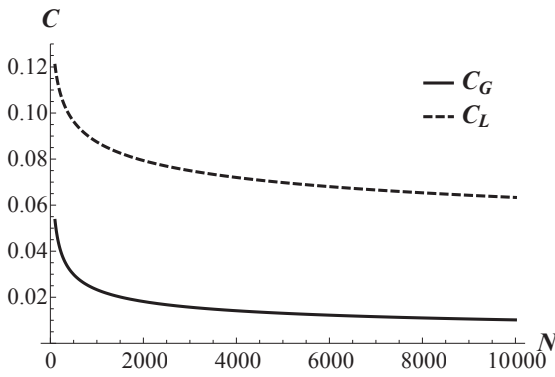


Рис. 9. Зависимости  $C_G$  и  $C_L$  от  $N$  при  $\tau \sim \mathbf{N}(1, 2)$   
 Fig. 9. Dependencies of  $C_G$  and  $C_L$  on  $N$  when  $\tau \sim \mathbf{N}(1, 2)$

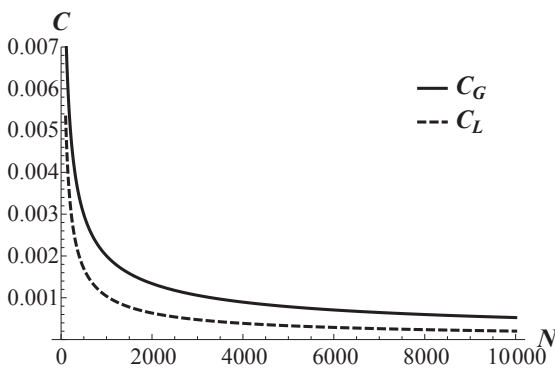


Рис. 10. Зависимости  $C_G$  и  $C_L$  от  $N$  при  $\tau \sim \mathbf{N}[2, 3]$   
 Fig. 10. Dependencies of  $C_G$  and  $C_L$  on  $N$  when  $\tau \sim \mathbf{N}[2, 3]$

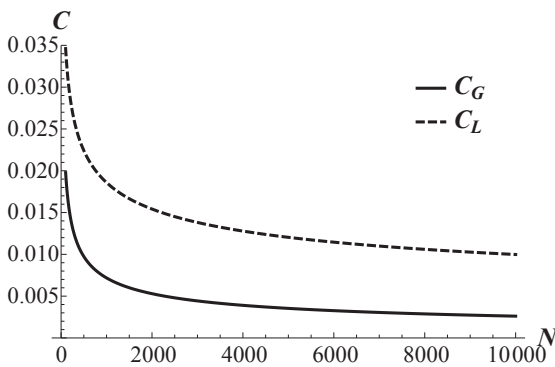


Рис. 11. Зависимости  $C_G$  и  $C_L$  от  $N$  при  $\tau \sim \mathbf{N}(1, 3)$   
 Fig. 11. Dependencies of  $C_G$  and  $C_L$  on  $N$  when  $\tau \sim \mathbf{N}(1, 3)$

Финансовое обеспечение исследований осуществлялось из средств федерального бюджета

та на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН).

Исследования выполнены на научном оборудовании Центра коллективного пользования Федерального исследовательского центра «Карельский научный центр Российской академии наук».

## ЛИТЕРАТУРА

1. Лери М. М. Сравнение устойчивости конфигурационных графов в случайной среде // Труды КарНЦ РАН. 2018. № 7. С. 46–56. doi: 10.17076/mat817
2. Afanasyev V. I., Boinghoff C., Kersting G., Vatutin V. A. Limit theorems for weekly branching processes in random environment // J. Theoretical Probability. 2012. Vol. 25, iss. 3. P. 703–732. doi: 10.1007/s10959-010-0331-6
3. Bianconi G., Barabasi A.-L. Bose-Einstein condensation in complex networks // Physical Review Letters. 2001. Vol. 86, iss. 24. P. 5632–5635. doi: 10.1103/PhysRevLett.86.5632
4. Bollobas B. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs // European J. Combin. 1980. Vol. 1, iss. 4. P. 311–316. doi: 10.1016/S0195-6698(80)80030-8
5. Durrett R. Random Graph Dynamics. Cambridge: Cambridge Univ. Press, 2007. 221 p.
6. Faloutsos C., Faloutsos P., Faloutsos M. On power-law relationships of the internet topology // Comp. Comm. Rev. 1999. Vol. 29. P. 251–262. doi: 10.1145/316194.316229
7. Hofstad R. Random Graphs and Complex Networks. Vol. 1. Cambridge: Cambridge Univ. Press, 2017. 337 p.
8. Leri M., Pavlov Y. Forest fire models on configuration random graphs // Fundamenta Informaticae. 2016. Vol. 145, iss. 3. P. 313–322. doi: 10.3233/FI-2016-1362
9. Leri M., Pavlov Yu. Random graphs' robustness in random environment // Austrian J. Stat. 2017. Vol. 46, iss. 3-4. P. 89–98. doi: 10.17713/ajs.v46i3-4.674
10. 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
11. Reittu H., Norros I. On the power-law random graph model of massive data networks // Performance Evaluation. 2004. Vol. 55, iss. 1-2. P. 3–23. doi: 10.1016/S0166-5316(03)00097-X

Поступила в редакцию 22.03.2019

## REFERENCES

1. Leri M. M. Sravnenie ustoichivosti konfiguratsionnykh grafov v sluchainoi srede [On comparing configuration graphs robustness in a random environment]. *Trudy KarNTs RAN*. [Trans. KarRC RAS]. 2018. No. 7. P. 46–56. doi: 10.17076/mat817
2. Afanasyev V. I., Boinghoff C., Kersting G., Vatutin V. A. Limit theorems for weekly branching processes in random environment. *J. Theoretical Probability*. 2012. Vol. 25, iss. 3. P. 703–732. doi: 10.1007/s10959-010-0331-6
3. Bianconi G., Barabasi A.-L. Bose-Einstein condensation in complex networks. *Physical Review Letters*. 2001. Vol. 86, iss. 24. P. 5632–5635. doi: 10.1103/PhysRevLett.86.5632
4. Bollobas B. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combinatorics*. 1980. Vol. 1, iss. 4. P. 311–316. doi: 10.1016/S0195-6698(80)80030-8
5. Durrett R. *Random Graph Dynamics*. Cambridge: Cambridge Univ. Press, 2007. 221 p.
6. Faloutsos C., Faloutsos P., Faloutsos M. On power-law relationships of the internet topology. *Comp. Comm. Rev.* 1999. Vol. 29. P. 251–262. doi: 10.1145/316194.316229
7. Hofstad R. *Random Graphs and Complex Networks*. Vol. 1. Cambridge: Cambridge Univ. Press, 2017. 337 p.
8. Leri M., Pavlov Y. Forest fire models on configuration random graphs. *Fundamenta Informaticae*. 2016. Vol. 145, iss. 3. P. 313–322. doi: 10.3233/FI-2016-1362
9. Leri M., Pavlov Yu. Random graphs' robustness in random environment. *Austrian J. Stat.* 2017. Vol. 46, iss. 3-4. P. 89–98. doi: 10.17713/ajs.v46i3-4.674
10. 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
11. Reittu H., Norros I. On the power-law random graph model of massive data networks. *Performance Evaluation*. 2004. Vol. 55, iss. 1-2. P. 3–23. doi: 10.1016/S0166-5316(03)00097-X

Received March 22, 2019

## СВЕДЕНИЯ ОБ АВТОРЕ:

### Лери Марина Муксумовна

научный сотрудник, к. т. н.  
Институт прикладных математических исследований  
КарНЦ РАН, Федеральный исследовательский центр  
«Карельский научный центр РАН»  
ул. Пушкинская, 11, Петрозаводск,  
Республика Карелия, Россия, 185910  
эл. почта: leri@krc.karelia.ru  
тел.: (8142) 781218

## CONTRIBUTOR:

### Leri, Marina

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