UDC 519.2

ON ONE STATISTICAL ISSUE FOR POWER-LAW RANDOM GRAPHS

[1] Marina M. Leri
[2] Irina A. Cheplyukova

[1] Karelian Research Center of the Russian Academy of Sciences
185910, Petrozavodsk, Pushkinskaya st., 11
PhD (technical)
E-mail: leri@krc.karelia.ru
[2] Karelian Research Center of the Russian Academy of Sciences
185910, Petrozavodsk, Pushkinskaya st., 11
PhD (physical and mathematical)
E-mail: chia@krc.karelia.ru

In the paper there are considered random graphs of Internet-type, i.e. graph node degrees are drawn independently from power-law distributions. We study the simulated power of the Chi-square and Kolmogorov-Smirnov goodness-of-fit tests in checking a hypothesis that graph node degrees are identically distributed.

Keywords: power-law random graphs, Chi-square goodness-of-fit test, Kolmogorov-Smirnov goodness-of-fit test, simulation modeling.

We consider random graphs with node degrees $\xi_1, \xi_2, \ldots, \xi_N$ drawn independently from the following distribution:

$$\mathbf{P}\{\xi_i \geq k\} = k^{-\tau_i},\ k = 1,2,\ldots;\ i = 1,2,\ldots,N;\ \tau_i > 0, \quad (1)$$

where N is the number of nodes numbered from 1 to N and τ_i are distribution parameters. Such graphs are being widely studied since the past decade due to the wide use of these models for the description of complex networks (see e.g. [1-3]). It had been shown [1, 2] that in real networks (in particular, Internet) parameters τ_i are most likely to be equal $\tau_1 = \tau_2 = \ldots = \tau_N = \tau$ and laying in the interval (1,2), that is why they sometimes are called Internet graphs [4]. In this case the distribution (1) has finite expectation and infinite variance.

In our work we explore competence of using the Chi-square and Kolmogorov-Smirnov goodness-of-fit tests for testing a hypothesis that graph node degrees are identically distributed and follow a power-law (1). Previously there was built a computer model of random graphs of Internet-type [5] and analyzed the behavior of some structure characteristics of these graphs when $\tau_1 = \tau_2 = \ldots = \tau_N = \tau$. Here by Monte-Carlo simulations we study the power of considered goodness of fit tests when checking a simple null hypothesis H_0 that node degrees follow the distribution (1) with $\tau_1 = \tau_2 = \ldots = \tau_N = \tau$ along with an alternative H_1 that the data is drawn from a different distribution or that not all τ_i are equal to τ.

At first, graphs were simulated for seven graph sizes N from 1000 to 10000 and 9 values of parameter $\tau_1 = \tau_2 = \ldots = \tau_N = \tau^* \in (1,2)$ with a step 0.1 (for each pair (N, τ^*) we generated 100 graphs). For each graph there was checked a hypothesis H_0 at the standard significance level 0.05. There were built regression models for the dependence of powers of

Chi-square (μ_χ) and Kolmogorov-Smirnov (μ_{KS}) goodness-of-fit tests on N and $\delta = |\tau - \tau^*|$ (multiple correlation coefficients of the models are 0.83 and 0.78 correspondingly):

$$\mu_\chi = \begin{cases} 0.12 \ln N + 0.19 \ln \delta + 0.05, & 0.01 \le \delta \le \min\{0.8; 148.4 \cdot N^{-0.632}\} \\ 1, & \delta > \min\{0.8; 148.4 \cdot N^{-0.632}\} \end{cases};$$

$$\ln \mu_{KS} = \begin{cases} 0.5 \ln N + 0.7 \ln \delta - 3.7, & 0.01 \le \delta \le \min\{0.8; 197.5 \cdot N^{-0.715}\} \\ 1, & \delta > \min\{0.8; 197.5 \cdot N^{-0.715}\} \end{cases}.$$

It's important to note that the graph structure is such that 50-70% of nodes have degree 1, 14-16% – have degree 2, 5-8% – degree 3, etc. Following this observation and the results of power study we showed that it is enough to use 3 bins in the Chi-square goodness-of-fit test.

Secondly, we simulated random graphs with two types of nodes such that node degrees follow a power-law (1) with τ_1^* and τ_2^* ($\tau_1^* < \tau_2^*$). Percentages of nodes of each type were taken to be 50/50%, 20/80%, 80/20%, 2/98% and 98/2%. We generated random graphs of 4 sizes N from 1000 to 10000 and 8 values of $\delta^* = |\tau_1^* - \tau_2^*| \in [0.1, 0.8]$ with a step 0.1. In accordance with calculations of simulated power there were built models of the dependence of powers of goodness-of-fit tests on N and δ^*. Models show that the power of goodness-of-fit tests increases with the growth of both graph size N and δ^* along with the growth of the difference between percentage of nodes of various types.

So, in accordance with the results of simulated power study of the Chi-square and Kolmogorov-Smirnov goodness-of-fit tests we conclude that the use of these tests in checking a hypothesis that graph vertex degrees are identically distributed and follow a power-law (1) is lawful, though sufficiently large power of tests could be ensured for graphs consisting of more than 10^4 nodes. Furthermore the rejection of the null hypothesis does not fully show that graph node degrees are identically distributed with a defined parameter, it doesn't exclude a case that graph contains nodes of different type, but their proportion is rather small.

References:

1. Faloutsos C. On power-law relationships of the internet topology / C. Faloutsos, P. Faloutsos, M. Faloutsos // Computer Communications Rev. 1999. Vol. 29, № 4. P. 251–262.

2. Reittu H. On the power-law random graph model of massive data networks / H. Reittu, I. Norros // Performance Evaluation. 2004. Vol. 55, № 1-2. P. 3–23.

3. Clauset A. Power-law distributions in empirical data / A. Clauset, C. R. Shalizi, M. E. J. Newman // SIAM Review. 2009. Vol. 51, № 4. P. 661–703.

4. Pavlov Yu. L. The limit distribution of the size of a giant component in an Internet-type random graph / Discrete Mathematics and Applications. 2007. Vol. 17, № 5. P. 425–437.

5. Leri M. M. Modeling of random graphs of Internet-type / Surveys in Applied and Industrial Mathematics. 2009. Vol. 16, № 5. P. 737–744. (in russian).

УДК 519.2

ОБ ОДНОЙ СТАТИСТИЧЕСКОЙ ЗАДАЧЕ ДЛЯ СЛУЧАЙНЫХ ИНТЕРНЕТ ГРАФОВ

[1] Марина Муксумовна Лери
[2] Ирина Александровна Чеплюкова

[1] Учреждение Российской академии наук Карельский научный центр РАН
185910, г. Петрозаводск, ул. Пушкинская, 11
Кандидат технических наук
E-mail: leri@krc.karelia.ru
[2] Учреждение Российской академии наук Карельский научный центр РАН
185910, г. Петрозаводск, ул. Пушкинская, 11
Кандидат физико-математических наук
E-mail: chia@krc.karelia.ru

Рассматриваются случайные графы Интернет-типа, т.е. графы, степени вершин которых независимы и имеют степенные распределения. С помощью методов имитационного моделирования проведено исследование мощности критериев согласия Пирсона и Колмогорова-Смирнова при проверке гипотезы о том, что степени вершин графа одинаково распределены.

Ключевые слова: случайные графы, критерий согласия Пирсона, критерий согласия Колмогорова-Смирнова, имитационное моделирование.