Distance in power-law configuration graphs

Marina Leri*

June 30, 2022

We consider configuration graphs [1] consisting of N nodes with node degrees $\xi_1, \dots, \xi_N$ being independent identically distributed random variables following the power-law distribution (see e. g. [2]):

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

where $\tau > 1$ is a parameter of the distribution (1) with the expectation equal to the Riemann zeta function in τ: $\zeta(\tau)$ and variance being infinite when $\tau \in (1, 2)$ and finite when $\tau > 2$.

Graph construction [1] starts by giving each node a certain degree in conformity with the distribution (1). Node degrees form stubs [2] (or semiedges) incident to the node. Pairing all stubs equiprobably, we form graph edges. Obviously, the total number of stubs in a graph has to be even. If not, we add a stub to an equiprobably chosen node increasing its degree by 1 to form the lacking edge. Obviously, configuration graphs may have loops, multiple edges and cycles.

In this work we consider average distance in power-law configuration graphs. Under a distance $d(v, u)$ between a pair of graph nodes v and u is understood a minimal number of edges (the shortest path) between these nodes. The distance between two nodes lying in different connected components is assumed to be ∞. Thus, the average distance in such graphs is calculated as the average of all non-infinite distances between all pairs of nodes:

$$dist = \frac{\sum_{v \neq u} d(v, u)}{k},$$

where k is a number of $d(v, u) \neq \infty$.

*Institute of Applied Mathematical Research, KarRC RAS, leri@krc.karelia.ru

The aim of the work was to find the dependencies of *dist* on the graph size N and the parameter of node degree distribution τ for the power-law configuration graphs. The study was performed using simulation methods followed by statistical data processing. The Dijkstra's algorithm [4] was used to find the shortest path between a pair of graph nodes.

In [3] the author shows that for typical nodal distance in graphs $d(v,u)$ the following holds:

$$d(v,u) \sim \frac{2\ln\ln N}{|\ln(\tau-1)|}, \tag{2}$$

when $1 < \tau < 2$ and $N \to \infty$. And if $\tau > 2$ then

$$d(v,u) \sim \frac{\ln N}{\ln \nu} \tag{3}$$

for $N \to \infty$, where $\nu = \frac{\mathbf{E}\xi(\xi-1)}{\mathbf{E}\xi}$ for $\nu > 1$. It is easy to show that in power-law configuration graphs $\nu = \frac{2\zeta(\tau-1)}{\zeta(\tau)} - 2$ and $\nu > 1$ for $2 < \tau \le \tau^*$ where $\tau^* = 2.81063\ldots$

Therefore, an attempt was made to built regression relations in the following forms: $dist = \frac{2\ln\ln N}{|\ln(\tau-1)|} + b_1$ on the interval $1.1 \le \tau \le 1.99$ and $dist = \frac{\ln N}{\ln\left(\frac{2\zeta(\tau-1)}{\zeta(\tau)}-2\right)} + b_2$ on the interval $2.01 \le \tau \le 2.8$. The sizes of the simulated graphs varied from 10 to 7000 nodes 100 graph realizations were generated for each pair (N,τ). The right bound of the first interval and the left bound of the second one are explained by limitations in denominators. Coefficients b_1 and b_2 were found using the least squares method. However, not only the determination coefficients of the obtained models showed low values (0.01 and 0.0005, respectively) but the models themselves also showed poor fit with experimental data.

For this reason, other regression relations were made for the dependence of average distance on N and τ, which we offer to use for power-law configuration graphs that are smaller or equal to 7000 nodes (sizes lying in the pre-asymptotic range):

$$dist = \frac{(8.977 - 6.154\tau + 0.834\tau^2)\ln\ln N}{|\ln(\tau-1)|} \tag{4}$$

for $1.1 \le \tau \le 1.99$ and

$$dist = \frac{(31,706 - 22,076\tau + 3,841\tau^2)\ln N}{\ln\left(\frac{2\zeta(\tau-1)}{\zeta(\tau)} - 2\right)} \tag{5}$$

for $2.01 \leq \tau \leq 2.8$ with the determination coefficients of 0.88 and 0.74, respectively.

Thus, although the relations (2) and (3) hold for graphs where $N \to \infty$ (asymptotic range) [3], our study shows that they do not work for the pre-asymptotic range, where the relations (4) and (5) better represent the real situation.

References

[1] Bollobas B. A probabilistic proof of an asymptotic formula of the number of labelled regular graphs. *Eur. J. Combin.* 1980. Vol. 1, iss. 4, 311–316.

[2] Reittu H., Norros I. On the power-law random graph model of massive data networks. *Performance Evaluation*. 2004. Vol. 55, iss. 1-2, 3–23.

[3] Hofstad R. Random graphs and complex networks. 2018. Vol. 2. 314 p. URL: https://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf

[4] Dijkstra E. W. A note on two problems in connexion with graphs. *Numer. Math.*. 1959. Vol. 1, iss. 1, 269–271.