

GLOBAL CLUSTERING COEFFICIENT IN SCALE-FREE WEIGHTED AND UNWEIGHTED NETWORKS

Liudmila Ostroumova Prokhorenkova

Yandex, Moscow, Russia and Moscow State University, Moscow, Russia

Abstract *In this article, we present a detailed analysis of the global clustering coefficient in scale-free graphs. Many observed real-world networks of diverse nature have a power-law degree distribution. Moreover, the observed degree distribution usually has an infinite variance. Therefore, we are especially interested in such degree distributions. In addition, we analyze the clustering coefficient for both weighted and unweighted graphs.*

There are two well-known definitions of the clustering coefficient of a graph: the global and the average local clustering coefficients. There are several models proposed in the literature for which the average local clustering coefficient tends to a positive constant as a graph grows. However, there are no models of scale-free networks with an infinite variance of the degree distribution and with an asymptotically constant global clustering coefficient. Models with constant global clustering and finite variance were also proposed. Therefore, in this work we focus only on the most interesting case: we analyze the global clustering coefficient for graphs with an infinite variance of the degree distribution.

For unweighted graphs, we prove that the global clustering coefficient tends to zero with high probability and we also estimate the largest possible clustering coefficient for such graphs. On the contrary, for weighted graphs, the constant global clustering coefficient can be obtained even for the case of an infinite variance of the degree distribution.

1. INTRODUCTION

In this study, we analyze the global clustering coefficient of graphs with a power-law degree distribution. Namely, we consider a sequence of graphs with degree distributions following a regularly varying distribution F . We assume that the degrees of the vertices are randomly generated according to F . Then, for a given outcome of the degree sequence, a graph can be built in any arbitrary way. Thus, there is no distribution on the graphs.

It was previously shown in [9] that if F has an infinite variance, then the global clustering coefficient for any such sequence of graphs tends to zero with high probability. Namely, an upper bound for the number of triangles is obtained [9]. Again, the argument for the upper bound uses information only on the degree sequence and does not assume any random graph model. In addition, the constructing procedure that allows us to obtain the sequence of graphs with a superlinear number of triangles is presented in [9]. However, the number of triangles in the constructed graphs grows slower than the upper bound obtained.

A preliminary version of this work was presented at the 11th Workshop on Algorithms and Models for the Web Graph (WAW '14), Beijing, China, December 2014.

Address correspondence to Liudmila Ostroumova Prokhorenkova, Department of Theoretical and Practical Research, L'va Tolstogo Street, 16, Moscow, Russian Federation. Email: ostroumova-la@yandex.ru

In this study, we close this gap by improving the upper bound obtained in [9]. Moreover, we also analyze graphs with multiple edges and show that weighted scale-free graphs with asymptotically constant global clustering coefficient and with an infinite variance of the degree distribution do exist.

The rest of this article is organized as follows. In the next section, we discuss several definitions of the clustering coefficient for weighted and unweighted graphs. Then, in Section 3, we formally define our restriction on a sequence of graphs. In Sections 5 and 6, we analyze the global clustering coefficient for the unweighted and the weighted case, respectively. Section 7 concludes the article.

2. CLUSTERING COEFFICIENTS

There are two well-known definitions of the clustering coefficient [2, 5] of an unweighted graph. The *global clustering coefficient* $C_1(G_n)$ is the ratio of three times the number of triangles to the number of pairs of adjacent edges in G_n . The *average local clustering coefficient* is defined as follows: $C_2(G_n) = \frac{1}{n} \sum_{i=1}^n C(i)$, where $C(i)$ is the local clustering coefficient for a vertex i : $C(i) = \frac{T^i}{P_2^i}$, where T^i is the number of edges between the neighbors of the vertex i and P_2^i is the number of pairs of neighbors. Note that both clustering coefficients equal 1 for a complete graph.

It was mentioned in [2, 5] that in research papers, either the average local or the global clustering coefficients are considered, and it is not always clear which definition is used. However, these two clustering coefficients differ: e.g., it was demonstrated [7] that, for networks based on the idea of preferential attachment, the difference between these two clustering coefficients is crucial.

It is also reasonable to study the global clustering coefficient for graphs with multiple edges. This agrees well with reality; for example, the Web host graph has many multiple edges: there can be several edges between the pages of two hosts. And even in the Internet graph (vertices are webpages and edges are links between them) multiple edges occur.

We refer to the study [6] for the definition of the global clustering coefficient for weighted graphs. They propose the following generalization of the global clustering coefficient to multigraphs:

$$C_1(G) = \frac{\text{total value of closed triplets}}{\text{total value of triplets}}.$$

A triplet is a group of three nodes u, v, w such that the pairs u, v , and u, w are connected. A triplet is called closed if v and w are also connected. Note that every triangle consists of three closed triplets. There are several ways to define the value of a triplet. First, the triplet value can be defined as the *arithmetic mean* of the weights of two edges (u, v) and (u, w) that make up the triplet. Second, it can be defined as the *geometric mean* of the weights of the edges. Third, it can be defined as the *maximum or minimum value* of the weights of the edges. In addition to these methods proposed in [6], we also propose the following natural definition of the weight: the weight of a triplet is the *product* of the weights of the edges. This definition agrees with the following property: the total value of all triplets located in a vertex is close to its degree squared.

In Section 6 we consider graphs with multiple edges. In this case, the weight of an edge is equal to its multiplicity.

3. SCALE-FREE GRAPHS

We consider a sequence of graphs $\{G_n\}$. Each graph G_n has n vertices. As in [9], we assume that the degrees of the vertices are independent random variables following a *regularly varying* distribution with a cumulative distribution function F satisfying

$$1 - F(x) = L(x)x^{-\gamma}, \quad x > 0, \quad (3.1)$$

where $L(\cdot)$ is a slowly varying function, that is, for any fixed constant $t > 0$

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1.$$

There is another obvious restriction on the function $L(\cdot)$: the function $1 - L(x)x^{-\gamma}$ must be a cumulative distribution function of a random variable taking positive integer values with probability 1.

Note that (3.1) describes a broad class of heavy-tailed distributions without imposing the rigid Pareto assumption. The power-law distribution with parameter $\gamma + 1$ corresponds to the cumulative distribution $1 - F(x) = L(x)x^{-\gamma}$. Further, by ξ, ξ_1, ξ_2, \dots we denote random variables with the distribution F . Note that for any $\alpha < \gamma$, the moment $\mathbb{E}\xi^\alpha$ is finite.

Models with $\gamma > 2$ and with the global clustering coefficient tending to some positive constant were already proposed (see, e.g., [7]). Therefore, in this work we consider only the case $1 < \gamma < 2$.

One small problem remains: we can construct a graph with a given degree distribution only if the sum of all degrees is even. This problem is easy to solve: we can either regenerate the degrees until their sum is even or we can add 1 to the last variable if their sum is odd [3]. For the sake of simplicity, we choose the second option, i.e., if $\sum_{i=1}^n \xi_i$ is odd, then we replace ξ_n by $\xi_n + 1$. It is easy to see that this modification does not change any of our results, therefore, we do not focus on the evenness further.

In Sections 5 and 6 of this article, we state that some results hold *with high probability*. Let us emphasize that the probability refers only to the randomness defining the degree sequence, and the obtained bounds, e.g., $O(n^{-\alpha})$ with some $\alpha > 0$, hold uniformly with respect to any sequence of graphs $\{G_n\}$ with the given degree sequence.

4. AUXILIARY RESULTS

In this section, we prove several auxiliary lemmas. These lemmas generalize several results from [9]. In order to prove these lemmas we use the following theorem (see, e.g., [1]).

Theorem 4.1. (Karamata's theorem.) *Let L be slowly varying and locally bounded in $[x_0, \infty]$ for some $x_0 \geq 0$. Then*

1. for $\alpha > -1$

$$\int_{x_0}^x t^\alpha L(t) dt = (1 + o(1))(\alpha + 1)^{-1} x^{\alpha+1} L(x), \quad x \rightarrow \infty.$$

2. for $\alpha < -1$

$$\int_x^\infty t^\alpha L(t) dt = -(1 + o(1))(\alpha + 1)^{-1} x^{\alpha+1} L(x), \quad x \rightarrow \infty.$$

We also use the following known lemma (its proof can be found, e.g., in [8]).

Lemma 4.2. *Let ξ_1, \dots, ξ_n be mutually independent random variables, $\mathbb{E}\xi_i = 0$, $\mathbb{E}|\xi_i|^\alpha < \infty$, $1 \leq \alpha \leq 2$, then*

$$\mathbb{E}(|\xi_1 + \dots + \xi_n|^\alpha) \leq 2^\alpha (\mathbb{E}(|\xi_1|^\alpha) + \dots + \mathbb{E}(|\xi_n|^\alpha)).$$

We need the following notation:

$$S_{n,c}(x) = \sum_{i=1}^n \xi_i^c I[\xi_i > x], \quad \bar{S}_{n,c}(x) = \sum_{i=1}^n \xi_i^c I[\xi_i \leq x], \quad \text{here } c, x \geq 0.$$

Lemma 4.3. *Fix any c such that $0 \leq c < \gamma$, any β such that $1 < \beta < \gamma/c$ and $\beta \leq 2$, and any $\varepsilon > 0$. Then, for any $x = x(n) > 0$ such that $x(n) \rightarrow \infty$, we have*

$$\begin{aligned} \mathbb{E}S_{n,c}(x) &= \frac{\gamma}{\gamma - c} n x^{c-\gamma} L(x) (1 + o(1)), \quad n \rightarrow \infty, \\ \mathbf{P}(|S_{n,c}(x) - \mathbb{E}S_{n,c}(x)| > \varepsilon \mathbb{E}S_{n,c}(x)) &= O\left(\left(\frac{x^\gamma}{n L(x)}\right)^{\beta-1}\right). \end{aligned}$$

Proof. We now assume that $c > 0$.

First, we estimate the expectation of $S_{n,c}(x)$:

$$\begin{aligned} \mathbb{E}S_{n,c}(x) &= n \int_x^\infty t^c dF(t) = -n \int_x^\infty t^c d(1 - F(t)) \\ &= -n t^c (1 - F(t)) \Big|_x^\infty + n c \int_x^\infty t^{c-1} (1 - F(t)) dt \\ &= n x^{c-\gamma} L(x) + n c \int_x^\infty t^{c-1-\gamma} L(t) dt \\ &\sim n x^{c-\gamma} L(x) - n c (c - \gamma)^{-1} x^{c-\gamma} L(x) = \frac{\gamma}{\gamma - c} n x^{c-\gamma} L(x). \end{aligned}$$

Then, we estimate

$$\mathbb{E}(\xi^c I[\xi > x])^\beta = \frac{1}{n} \mathbb{E}S_{n,c\beta}(x) \sim \frac{\gamma}{\gamma - \beta c} x^{\beta c - \gamma} L(x)$$

and get

$$\begin{aligned} \mathbf{P}(|S_{n,c}(x) - \mathbb{E}S_{n,c}(x)| > \varepsilon \mathbb{E}S_{n,c}(x)) &\leq \frac{\mathbb{E}|S_{n,c}(x) - \mathbb{E}S_{n,c}(x)|^\beta}{(\varepsilon \mathbb{E}S_{n,c}(x))^\beta} \\ &= O\left(\frac{n \mathbb{E}(\xi^c I[\xi > x])^\beta}{(\mathbb{E}S_{n,c}(x))^\beta}\right) = O\left(\frac{n x^{\beta c - \gamma} L(x)}{n^\beta x^{\beta(c-\gamma)} (L(x))^\beta}\right) \\ &= O(n^{1-\beta} x^{-(1-\beta)\gamma} (L(x))^{1-\beta}). \end{aligned}$$

Here we applied Lemma 4.2 in order to estimate $\mathbb{E}|S_{n,c}(x) - \mathbb{E}S_{n,c}(x)|^\beta$.

The case $c = 0$ can be considered similarly:

$$\begin{aligned} \mathbb{E}S_{n,0}(x) &= n\mathbf{P}(\xi > x) = nx^{-\gamma}L(x), \\ \mathbf{P}(|S_{n,0}(x) - \mathbb{E}S_{n,0}(x)| > \varepsilon \mathbb{E}S_{n,0}(x)) &= O\left(\frac{nx^{-\gamma}L(x)}{(nx^{-\gamma}L(x))^\beta}\right) \\ &= O\left(\left(\frac{x^\gamma}{nL(x)}\right)^{\beta-1}\right). \end{aligned}$$

□

Lemma 4.4. *Fix any c such that $c > \gamma$ and any $\varepsilon > 0$. Then, for any $x = x(n) > 0$ such that $x(n) \rightarrow \infty$, we have*

$$\mathbb{E}\bar{S}_{n,c}(x) = \frac{\gamma}{c-\gamma}nx^{c-\gamma}L(x)(1+o(1)), \quad n \rightarrow \infty,$$

$$\mathbf{P}(|\bar{S}_{n,c}(x) - \mathbb{E}\bar{S}_{n,c}(x)| > \varepsilon \mathbb{E}\bar{S}_{n,c}(x)) = O\left(\frac{x^\gamma}{nL(x)}\right).$$

Proof. Again, first we estimate the expectation of $\bar{S}_{n,c}(x)$:

$$\begin{aligned} \mathbb{E}\bar{S}_{n,c}(x) &= n \int_0^x t^c dF(t) = -n \int_0^x t^c d(1-F(t)) \\ &= -nt^c(1-F(t))\Big|_0^x + nc \int_0^x t^{c-1}(1-F(t)) dt \\ &= -nx^{c-\gamma}L(x) + nc \int_0^x t^{c-1-\gamma}L(t) dt \\ &\sim -nx^{c-\gamma}L(x) + nc(c-\gamma)^{-1}x^{c-\gamma}L(x) = \frac{\gamma}{c-\gamma}nx^{c-\gamma}L(x). \end{aligned}$$

Then, we estimate

$$\mathbb{E}(\xi^c I[\xi \leq x])^2 = \frac{1}{n}\mathbb{E}\bar{S}_{n,2c}(x) \sim \frac{\gamma}{2c-\gamma}x^{2c-\gamma}L(x)$$

and get

$$\begin{aligned} \mathbf{P}(|\bar{S}_{n,c}(x) - \mathbb{E}\bar{S}_{n,c}(x)| > \varepsilon \mathbb{E}\bar{S}_{n,c}(x)) &\leq \frac{\mathbb{E}|\bar{S}_{n,c}(x) - \mathbb{E}\bar{S}_{n,c}(x)|^2}{(\varepsilon \mathbb{E}\bar{S}_{n,c}(x))^2} \\ &= O\left(\frac{n\mathbb{E}(\xi^c I[\xi \leq x])^2}{(\mathbb{E}\bar{S}_{n,c}(x))^2}\right) = O\left(\frac{nx^{2c-\gamma}L(x)}{n^2x^{2(c-\gamma)}(L(x))^2}\right) = O\left(\frac{x^\gamma}{nL(x)}\right). \end{aligned}$$

□

We prove two more lemmas. Put $\xi_{max} = \max\{\xi_1, \dots, \xi_n\}$.

Lemma 4.5. For any $\varepsilon > 0$ and any $\alpha > 0$,

$$\mathbf{P}\left(\xi_{\max} > n^{\frac{1}{\gamma}-\varepsilon}\right) = 1 - O\left(n^{-\alpha}\right).$$

Also, for any $\delta < \gamma\varepsilon$

$$\mathbf{P}\left(\xi_{\max} \leq n^{\frac{1}{\gamma}+\varepsilon}\right) = 1 - O\left(n^{-\delta}\right).$$

Proof.

$$\begin{aligned} \mathbf{P}\left(\xi_{\max} \leq n^{\frac{1}{\gamma}-\varepsilon}\right) &= \left[\mathbf{P}\left(\xi \leq n^{\frac{1}{\gamma}-\varepsilon}\right)\right]^n = \exp\left(n \log\left(1 - \mathbf{P}\left(\xi > n^{\frac{1}{\gamma}-\varepsilon}\right)\right)\right) \\ &= \exp\left(n \log\left(1 - L\left(n^{\frac{1}{\gamma}-\varepsilon}\right) n^{-\gamma\left(\frac{1}{\gamma}-\varepsilon\right)}\right)\right) \\ &= \exp\left(-n L\left(n^{\frac{1}{\gamma}-\varepsilon}\right) n^{-\gamma\left(\frac{1}{\gamma}-\varepsilon\right)}(1 + o(1))\right) \\ &= \exp\left(-L\left(n^{\frac{1}{\gamma}-\varepsilon}\right) n^{\gamma\varepsilon}(1 + o(1))\right) = O\left(n^{-\alpha}\right), \\ \mathbf{P}\left(\xi_{\max} > n^{\frac{1}{\gamma}+\varepsilon}\right) &\leq n \mathbf{P}\left(\xi > n^{\frac{1}{\gamma}+\varepsilon}\right) \leq n L\left(n^{\frac{1}{\gamma}+\varepsilon}\right) n^{-\gamma\left(\frac{1}{\gamma}+\varepsilon\right)} = O\left(n^{-\delta}\right). \end{aligned}$$

□

Lemma 4.6. For any $\varepsilon > 0$ and any $\delta < \frac{\gamma\varepsilon}{\gamma+2}$,

$$\mathbf{P}\left(\bar{\mathcal{S}}_{n,2}(\infty) \leq n^{\frac{2}{\gamma}+\varepsilon}\right) = 1 - O\left(n^{-\delta}\right).$$

Proof. Choose φ such that $\frac{\delta}{\gamma} < \varphi < \frac{\varepsilon}{\gamma+2}$. From Lemma 4.5 we get

$$\mathbf{P}\left(\xi_{\max} \leq n^{\frac{1}{\gamma}+\varphi}\right) = 1 - O\left(n^{-\delta}\right).$$

From Lemma 4.3 and Lemma 4.4, with probability

$$1 - O\left(\frac{n^{\gamma\left(\frac{1}{\gamma}-\varphi\right)}}{n L\left(n^{\frac{1}{\gamma}-\varphi}\right)}\right) = 1 - O\left(n^{-\delta}\right),$$

we have

$$\begin{aligned} \bar{\mathcal{S}}_{n,2}\left(n^{\frac{1}{\gamma}-\varphi}\right) &\leq (1 + \varepsilon) \frac{\gamma}{2 - \gamma} n^{\frac{2}{\gamma}+\varphi\gamma-2\varphi} L\left(n^{\frac{1}{\gamma}-\varphi}\right), \\ \mathcal{S}_{n,0}\left(n^{\frac{1}{\gamma}-\varphi}\right) &\leq (1 + \varepsilon) n^{\varphi\gamma} L\left(n^{\frac{1}{\gamma}-\varphi}\right). \end{aligned}$$

In this case,

$$\begin{aligned} \bar{S}_{n,2} \left(n^{\frac{1}{\gamma} + \varphi} \right) &\leq \bar{S}_{n,2} \left(n^{\frac{1}{\gamma} - \varphi} \right) + \xi_{\max} S_{n,0} \left(n^{\frac{1}{\gamma} - \varphi} \right) \\ &\leq (1 + \varepsilon) \frac{\gamma}{2 - \gamma} n^{\frac{2}{\gamma} + \varphi \gamma - 2\varphi} L \left(n^{\frac{1}{\gamma} - \varphi} \right) + n^{\frac{2}{\gamma} + 2\varphi} (1 + \varepsilon) n^{\varphi \gamma} L \left(n^{\frac{1}{\gamma} - \varphi} \right) \leq n^{\frac{2}{\gamma} + \varepsilon} \end{aligned}$$

for large enough n . This concludes the proof. □

Note that we estimated only the upper bound for $\bar{S}_{n,2}(\infty)$, because the lower bound can be obtained using the lower bound for ξ_{\max} . Here, we may use the inequality $\bar{S}_{n,2}(\infty) \geq \xi_{\max}^2$.

5. CLUSTERING IN UNWEIGHTED GRAPHS

5.1. Previous Results

The behavior of the global clustering coefficient in scale-free unweighted graphs was considered in [9]. In the case of an infinite variance, the reasonable question is whether there exists a simple graph (i.e., a graph without loops and multiple edges) with a given degree distribution. The following theorem is proved in [9].

Theorem 5.1. *With high probability there exists a simple graph on n vertices with the degree distribution defined in Section 3.*

So, with high probability such a graph exists and it is reasonable to discuss its global clustering coefficient. The following upper bound on the global clustering coefficient is obtained in [9].

Theorem 5.2. *For any $\varepsilon > 0$, with high probability the global clustering coefficient satisfies the following inequality*

$$C_1(G_n) \leq n^{-\frac{(\gamma-2)^2}{2\gamma} + \varepsilon}.$$

Taking small enough ε , one can see that with high probability $C_1(G_n) \rightarrow 0$ as n grows.

In addition, using simulations and empirical observations, the authors of [9] claimed that with high probability there exists a graph with $\sim n^{\frac{3}{\gamma+1}}$ triangles and with the degree distribution defined in Section 3, whereas the theoretical upper bound on the number of triangles is $n^{2-\frac{\gamma}{2}}$. For the considered case $1 < \gamma < 2$, we have $\frac{3}{\gamma+1} < 2 - \frac{\gamma}{2}$, and there is a gap between the number of constructed triangles and the obtained upper bound.

Further in this section we close this gap by improving the upper bound. We also rigorously prove the lower bound.

5.2. Upper Bound

We prove the following theorem.

Theorem 5.3. *For any $\varepsilon > 0$ and any α such that $0 < \alpha < \frac{1}{\gamma+1}$, with probability $1 - O(n^{-\alpha})$, the global clustering coefficient of G_n satisfies the following inequality*

$$C_1(G_n) \leq n^{-\frac{(2-\gamma)}{\gamma(\gamma+1)} + \varepsilon}.$$

Proof. The global clustering coefficient is

$$C_1(G_n) = \frac{3 \cdot T(n)}{P_2(n)},$$

where $T(n)$ is the number of triangles and $P_2(n)$ is the number of pairs of adjacent edges in G_n .

Note that $P_2(n) \geq \xi_{\max}(\xi_{\max} - 1)/2$. Therefore, from Lemma 4.5 we get that for any $\delta > 0$ with probability $1 - O(n^{-\alpha})$,

$$P_2(n) > n^{\frac{2}{\gamma} - \delta}.$$

It remains to estimate $T(n)$. Obviously, for any x

$$T(n) \leq |\{i : \xi_i > x\}|^3 + \sum_{i:\xi_i \leq x} \xi_i^2. \quad (5.1)$$

The first term in (5.1) is the upper bound for the number of triangles with all vertices among the set $\{i : \xi_i > x\}$. The second term is the upper bound for the number of triangles with at least one vertex among $\{i : \xi_i \leq x\}$.

From Lemma 4.3 and Lemma 4.4 we get

$$\begin{aligned} |\{i : \xi_i > x\}| &= S_{n,0}(x) \leq (1 + \varepsilon)n x^{-\gamma} L(x), \\ \sum_{i:\xi_i \leq x} \xi_i^2 &= \bar{S}_{n,2}(x) \leq (1 + \varepsilon) \frac{\gamma}{2 - \gamma} n x^{2-\gamma} L(x), \end{aligned}$$

with probability $1 - O(\frac{x^\gamma}{n L(x)})$.

Now we can fix $x = n^{\frac{1}{\gamma+1}}$. So, with probability

$$1 - O\left(\frac{n^{-\frac{1}{\gamma+1}}}{L\left(n^{\frac{1}{\gamma+1}}\right)}\right) = 1 - O(n^{-\alpha}),$$

we have

$$T(n) \leq n^{\frac{3}{\gamma+1} + \delta}$$

Taking small enough δ , we obtain

$$C_1(G_n) \leq n^{\varepsilon - \frac{2-\gamma}{\gamma(\gamma+1)}}.$$

This concludes the proof. □

5.3. Lower Bound

We prove the following theorem.

Theorem 5.4. *For any $\varepsilon > 0$ and any α such that $0 < \alpha < \min\{\frac{\gamma\varepsilon}{\gamma+2}, \frac{1}{\gamma+1}, \gamma - 1\}$, with probability $1 - O(n^{-\alpha})$, there exists a graph with the degree distribution defined in Section 3 and the global clustering coefficient satisfying the following inequality*

$$C_1(G_n) \geq n^{-\frac{(2-\gamma)}{\gamma(\gamma+1)} - \varepsilon}.$$

Proof. Again,

$$C_1(G_n) = \frac{3 \cdot T(n)}{P_2(n)}.$$

The upper bound for $P_2(n)$ follows from Lemma 4.6. Fix ε' such that $\frac{\alpha(\gamma+2)}{\gamma} < \varepsilon' < \varepsilon$. Then,

$$\mathbf{P}\left(P_2(n) \leq n^{\frac{2}{\gamma} + \varepsilon'}\right) \geq \mathbf{P}\left(\bar{S}_{n,2}(\infty) \leq n^{\frac{2}{\gamma} + \varepsilon'}\right) = 1 - O(n^{-\alpha}).$$

□

Now, we present the lower bound for $T(n)$. Fix any δ such that $0 < \gamma\delta < \min\{\frac{1}{\gamma+1} - \alpha, \frac{\varepsilon - \varepsilon'}{3}\}$. It follows from Lemma 4.3 that, with probability $1 - O(n^{-\alpha})$,

$$S_{n,0}\left(n^{\frac{1}{\gamma+1} + \delta}\right) \leq (1 + \varepsilon)n^{\frac{1}{\gamma+1} - \gamma\delta} L\left(n^{\frac{1}{\gamma+1} + \delta}\right) \leq n^{\frac{1}{\gamma+1} + \delta}.$$

Let us denote by A the set of vertices whose degrees are greater than $n^{\frac{1}{\gamma+1} + \delta}$. The size of A equals $S_{n,0}\left(n^{\frac{1}{\gamma+1} + \delta}\right)$. Because the number of vertices in A is not greater than the minimum degree in A , a clique on A can be constructed. Therefore, with probability $1 - O(n^{-\alpha})$,

$$S_{n,0}\left(n^{\frac{1}{\gamma+1} + \delta}\right) \geq (1 - \varepsilon)n^{\frac{1}{\gamma+1} - \gamma\delta} L\left(n^{\frac{1}{\gamma+1} + \delta}\right)$$

and

$$3T(n) \geq 3\binom{S_{n,0}\left(n^{\frac{1}{\gamma+1} + \delta}\right)}{3} \geq n^{\frac{3}{\gamma+1} - (\varepsilon - \varepsilon')}.$$

Finally, we get

$$C_1(G_n) = \frac{3 \cdot T(n)}{P_2(n)} \geq \frac{n^{\frac{3}{\gamma+1} - (\varepsilon - \varepsilon')}}{n^{\frac{2}{\gamma} + \varepsilon'}} = n^{-\frac{(2-\gamma)}{\gamma(\gamma+1)} - \varepsilon}.$$

It remains to be proven that after we constructed a clique on the set A , with high probability we still can construct a graph without loops and multiple edges. This can be easily proved similarly to Theorem 5.1. Namely, we use the following theorem by Erdős and T. Gallai [4].

Theorem 5.5. (Erdős–Gallai.) *A sequence of nonnegative integers $d_1 \geq \dots \geq d_n$ can be represented as the degree sequence of a finite simple graph on n vertices if and only if*

1. $d_1 + \dots + d_n$ is even;
2. $\sum_{i=1}^k d_i \leq k(k-1) + \sum_{i=k+1}^n \min(d_i, k)$ holds for $1 \leq k \leq n$.

Let us order the realized values of the random variables ξ_1, \dots, ξ_n and obtain the ordered sequence $d_1 \geq \dots \geq d_n$. In order to apply the theorem of Erdős and Gallai, we assume that the set A is now a single vertex with the degree

$$\deg(A) = S_{n,1} \left(n^{\frac{1}{\nu+1} + \delta} \right) - 2 \binom{S_{n,0} \left(n^{\frac{1}{\nu+1} + \delta} \right)}{2} < S_{n,1} \left(n^{\frac{1}{\nu+1} + \delta} \right).$$

It is sufficient to prove that with probability $1 - O(n^{-\alpha})$, the following condition is satisfied

$$\deg(A) + \sum_{i=|A|+1}^k d_i \leq (k - |A|)(k - |A| + 1) + \sum_{i=k+1}^n \min(d_i, k - |A| + 1) \quad (5.2)$$

for all $k \geq |A|$.

Let us now prove that with probability $1 - O(n^{-\alpha})$, the condition (5.2) is satisfied. For some large enough C if $k > C\sqrt{n}$, then

$$\deg(A) + \sum_{i=|A|+1}^k d_i \leq (k - |A|)(k - |A| + 1).$$

This holds because with probability $1 - O(n^{-\alpha})$,

$$|A| = S_{n,0} \left(n^{\frac{1}{\nu+1} + \delta} \right) \leq (1 + \varepsilon) n^{\frac{1}{\nu+1} - \delta} L \left(n^{\frac{1}{\nu+1} + \delta} \right) \leq n^{\frac{1}{\nu+1}},$$

and the sum of all degrees grows linearly with n :

$$\mathbf{P} \left(|S_{n,1}(0) - n\mathbb{E}\xi| > \frac{n}{2} \mathbb{E}\xi \right) \leq \frac{4^{\alpha+1} n \mathbb{E}|\xi - \mathbb{E}\xi|^{\alpha+1}}{n^{\alpha+1} (\mathbb{E}\xi)^{\alpha+1}} = O(n^{-\alpha}).$$

Here, we used that $\alpha + 1 < \gamma$.

Finally, consider the case $k \leq C\sqrt{n}$. Note that $\min(d_i, k - |A| + 1) > 1$, so

$$\sum_{i=k+1}^n \min(d_i, k - |A| + 1) \geq n - C\sqrt{n}.$$

It remains to show that with probability $1 - O(n^{-\alpha})$,

$$\deg(A) + \sum_{i=|A|+1}^k d_i \leq n - C\sqrt{n}.$$

It is sufficient to show that

$$\sum_{i=1}^{\lfloor C\sqrt{n} \rfloor} d_i \leq n - C\sqrt{n}.$$

This inequality is easy to prove using Lemma 4.3. For any $\frac{1}{3\gamma} < \delta < \frac{1}{2\gamma}$, with probability $1 - O(n^{-1/2})$, we have

$$S_{n,0}(n^\delta) > C\sqrt{n}$$

and

$$S_{n,1}(n^\delta) \leq n^{\frac{2\gamma+1}{3\gamma}} \leq n - C\sqrt{n}.$$

Therefore, the condition (5.2) is satisfied.

6. CLUSTERING IN WEIGHTED GRAPHS

In this section, we analyze the global clustering coefficient of graphs with multiple edges. First, let us note that the case when we allow both loops and multiple edges is not very interesting: we can get a high clustering coefficient just by avoiding triplets. Namely, we can construct several triangles and then just create loops in all vertices. Then, we can connect the remaining half-edges for the vertices with odd degrees. Therefore, we assume that loops are not allowed. We show that even with this restriction it is possible to obtain a constant global clustering coefficient.

Several definitions of the global clustering coefficient for graphs with multiple edges are presented in Section 2. The following theorem holds for any definition of the global clustering coefficient $C_1(G_n)$.

Theorem 6.1. *Fix any $\delta > 0$. For any α such that $0 < \alpha < \frac{\gamma-1}{\gamma+1}$, with probability $1 - O(n^{-\alpha})$, there exists a loopless multigraph with the degree distribution defined in Section 3 and the global clustering coefficient satisfying the following inequality*

$$C_1(G_n) \geq \frac{2-\gamma}{2+\gamma} - \delta.$$

Proof. Fix some $\varepsilon > 0$. From Lemma 4.3, with $c = 0$ it follows that with probability $1 - O\left(\frac{x^\gamma}{nL(x)}\right)$

$$(1 - \varepsilon)n x^{-\gamma} L(x) \leq S_{n,0}(x) \leq (1 + \varepsilon)n x^{-\gamma} L(x). \quad (6.1)$$

Let us prove that, for large enough n , there always exists such x_0 that

$$(1 + \varepsilon)n x_0^{-\gamma} L(x_0) \leq x_0 \leq (1 + 2\varepsilon)n x_0^{-\gamma} L(x_0). \quad (6.2)$$

In other words, we want to find such x_0 that

$$\frac{1}{(1 + 2\varepsilon)n} \leq \frac{x_0^{-\gamma} L(x_0)}{x_0} \leq \frac{1}{(1 + \varepsilon)n}.$$

Recall that $x^{-\gamma}L(x) = 1 - F(x)$, where $F(x)$ is a cumulative distribution function. Therefore, $f(x) = \frac{x^{-\gamma}L(x)}{x}$ monotonically decreases to zero on $(0, \infty)$. The only problem is that $f(x)$ is a discontinuous function. In order to guarantee the existence of the required value x_0 , we have to prove that (for large enough n) if $f(x) < \frac{1}{n}$, then $f(x+) - f(x-) < \frac{\varepsilon}{(1+\varepsilon)(1+2\varepsilon)n}$. This can be proved as follows. For the function $F(x)$ it is obvious that if $1 - F(x) < \frac{1}{n}$, then $|F(x+) - F(x-)| < \frac{1}{n}$. Therefore, in this case, $|f(x+) - f(x-)| < \frac{1}{xn}$. For large enough n (and this leads to large enough x), we have $\frac{1}{xn} \leq \frac{\varepsilon}{(1+\varepsilon)(1+2\varepsilon)n}$. This concludes the proof of the fact that the required x_0 exists.

We take any value that satisfies (6.2) and further denote it by x_0 . Note that, up to a slowly varying multiplier, x_0 is of order $n^{\frac{1}{\gamma+1}}$. Therefore, $O(\frac{x_0^\gamma}{nL(x_0)}) = O(n^{-\alpha})$. From (6.1) and (6.2) it follows that with probability $1 - O(\frac{x_0^\gamma}{nL(x_0)})$, the number of vertices with degree greater than x_0 (i.e., $S_{n,0}(x_0)$) is not larger than x_0 . Denote this set of vertices by A_{x_0} . In this case, a clique on A_{x_0} can be constructed.

In addition, we want all vertices from the set A_{x_0} to be connected only to each other. This can be possible, because multiple edges are allowed (the proof of the fact that it is actually possible is given later in the text). If the sum of degrees in A_{x_0} is odd, then we allow one edge (from the vertex with the smallest degree in A_{x_0}) to go outside this set.

We are ready to estimate the global clustering coefficient:

$$C_1(G_n) = \frac{\text{total value of closed triplets}}{\text{total value of triplets}}.$$

The total value of closed triplets is at least $3\binom{S_{n,0}(x_0)}{3}$ regardless of the definition of the value of a triplet. With probability $1 - O(\frac{x_0^\gamma}{nL(x_0)})$,

$$3\binom{S_{n,0}(x_0)}{3} \geq \frac{1}{2}(1 - \varepsilon)^3 n^3 x_0^{-3\gamma} L^3(x_0).$$

The total value of all triplets includes:

- The total value of closed triplets on A_{x_0} estimated above,
- The total value of triplets on the remaining vertices, which is not greater than $\bar{S}_{n,2}(x_0)$,
- (optionally) Some unclosed triplets on the vertex with the smallest degree in A_{x_0} , if the sum of degrees in A_{x_0} is odd.

Because the smallest degree in the set A_{x_0} is of order x_0 , we can estimate the last two summands in the total value of triplets by

$$\bar{S}_{n,2}(x_0) + O(x_0^2) \leq (1 + \varepsilon) \frac{\gamma}{2 - \gamma} n x_0^{2-\gamma} L(x_0).$$

By Lemma 4.4, this holds with probability $1 - O(\frac{x_0^\gamma}{nL(x_0)})$.

¹ $f(x+) = \lim_{y \rightarrow x+} f(y)$, $f(x-) = \lim_{y \rightarrow x-} f(y)$.

Finally, with probability $1 - O(\frac{x_0^\gamma}{n L(x_0)})$, we have

$$\begin{aligned} C_1(G_n) &\geq \frac{\frac{1}{2}(1-\varepsilon)^3 n^3 x_0^{-3\gamma} L^3(x_0)}{\frac{1}{2}(1-\varepsilon)^3 n^3 x_0^{-3\gamma} L^3(x_0) + (1+\varepsilon)\frac{\gamma}{2-\gamma} n x_0^{2-\gamma} L(x_0)} \\ &\geq \frac{\frac{1}{2}(1-\varepsilon)^3 n^2 x_0^{-2\gamma} L^2(x_0)}{\frac{1}{2}(1-\varepsilon)^3 n^2 x_0^{-2\gamma} L^2(x_0) + (1+\varepsilon)\frac{\gamma}{2-\gamma} (1+2\varepsilon)^2 n^2 x_0^{-2\gamma} L^2(x_0)} \\ &= \frac{\frac{1}{2}(1-\varepsilon)^3}{\frac{1}{2}(1-\varepsilon)^3 + (1+\varepsilon)\frac{\gamma}{2-\gamma} (1+2\varepsilon)^2} \geq \frac{2-\gamma}{2+\gamma} - \delta. \end{aligned}$$

for sufficiently small ε . Here, in the second inequality we used (6.2).

Recall that the loops are not allowed. Therefore, it remains to prove that (1) a mult clique on A_{x_0} can be constructed; (2) a graph on the remaining vertices can be constructed. Note that a multigraph without loops can always be constructed if the maximum degree is not larger than the sum of the other degrees.

A mult clique on A_{x_0} can be constructed if

$$\xi_{\max} \leq S_{n,1}(x_0) - \xi_{\max} - x_0^2. \quad (6.3)$$

Here, x_0^2 is the upper bound for the number of half edges already involved in the required clique. From Lemma 4.3, with probability $1 - O(n^{-\alpha})$,

$$S_{n,1}(x_0) > (1-\varepsilon)\frac{\gamma}{\gamma-1} n x_0^{1-\gamma} L(x_0). \quad (6.4)$$

Fix some ε' such that $0 < \varepsilon' < \frac{1}{\gamma} \left(\frac{\gamma-1}{\gamma+1} - \alpha \right)$. In this case we have $\alpha < \frac{\gamma-1}{\gamma+1} - \varepsilon'\gamma$, therefore, Lemma 4.5 gives that

$$\mathbf{P} \left(\xi_{\max} \leq n^{\frac{2}{\gamma+1}-\varepsilon'} \right) = \mathbf{P} \left(\xi_{\max} \leq n^{\frac{1}{\gamma} + \frac{\gamma-1}{\gamma(\gamma+1)} - \varepsilon'} \right) = 1 - O(n^{-\alpha}). \quad (6.5)$$

Now (6.3) follows immediately from (6.4), (6.5), and the fact that x_0 is of order $n^{\frac{1}{\gamma+1}}$.

Similarly, it is easy to show that the graph on the remaining vertices can be constructed:

$$x_0 \leq \bar{S}_{n,1}(x_0) - x_0$$

because $\bar{S}_{n,1}(x_0) = S_{n,1}(0) - S_{n,1}(x_0)$ grows linearly with n . □

7. CONCLUSION

In this study, we fully analyzed the behavior of the global clustering coefficient in scale-free graphs with an infinite variance of the degree distribution. We considered both unweighted graphs and graphs with multiple edges. For the unweighted case, we first obtained the upper bound for the global clustering coefficient. In particular, we proved that the global clustering coefficient tends to zero with high probability. We also presented the

constructing procedure that allows us to reach the obtained upper bound. The situation turns out to be different for graphs with multiple edges. In this case, it is possible to construct a sequence of graphs with an asymptotically constant clustering coefficient.

FUNDING

This work was supported by the grant of RFBR №15-01-03530.

REFERENCES

- [1] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge, UK: Cambridge University Press, 1987.
- [2] B. Bollobás and O. M. Riordan. Mathematical Results on Scale-Free Random Graphs. In *Handbook of Graphs and Networks: From the Genome to the Internet* edited by S. Bornholdt and H. G. Schuster, pp. 1–34. Weinheim: Wiley-VCH, 2003.
- [3] T. Britton, M. Deijfen, and A. Martin-Löf. “Generating Simple Random Graphs with Prescribed Degree Distribution.” *J. Stat. Phys.* 124:6 (2006), 1377–1397.
- [4] P. Erdős and T. Gallai. “Graphs with Given Degrees of Vertices.” *Mat. Lapok* 11 (1960), 264–274.
- [5] M. E. J. Newman. “The Structure and Function of Complex Networks.” *SIAM Review* 45 (2003), 167256.
- [6] T. Opsahl and P. Panzarasa. “Clustering in Weighted Networks.” *Social Networks* 31:2 (2009), 155–163.
- [7] L. Ostroumova, A. Ryabchenko, and E. Samosvat. “Generalized Preferential Attachment: Tunable Power-Law Degree Distribution and Clustering Coefficient.” In *Algorithms and Models for the Web Graph*, vol. 8305, pp. 185–202, Lecture Notes in Computer Science. Berlin: Springer, 2013.
- [8] L. Ostroumova and E. Samosvat. “Recency-Based Preferential Attachment Models.” *arXiv*. Available online <http://arxiv.org/abs/1406.4308>, 2014.
- [9] L. Ostroumova Prokhorenkova and E. Samosvat. *Global Clustering Coefficient in Scale-Free Networks*, Lecture Notes in Computer Science, 8882. Berlin: Springer, 2014.