Scale-free graphs with many edges

Clara Stegehuis and Bert Zwart

December 15, 2022

Abstract

We develop tail estimates for the number of edges in a Chung-Lu random graph with regularly varying weight distribution. Our results show that the most likely way to have an unusually large number of edges is through the presence of one or more hubs, i.e. edges with degree $O(n)$.

1 Introduction and main results

We analyze a sequence of random graphs introduced by [5, 12] which is constructed as follows. Let n be the number of vertices and let $X_i, i \geq 1$, be an i.i.d. sequence of non-negative random variables with mean μ and a right tail which is regularly varying with index $-\alpha < -1$:

$$\mathbb{P}(X_1 > x) = L(x)x^{-\alpha},$$

$\alpha > 1$, with $L(yx)/L(x) \to 1$ as $x \to \infty$. X_i can be interpreted as a weight for vertex i, and we denote $\mu = \mathbb{E}[X_i]$. A vertex with a high weight tends to have more edges: the probability p_{ij} that an edge is present between vertices i and j equals

$$p_{ij} = p_{ij}^n(X_i, X_j) := \min\left\{\frac{X_i X_j}{\mu n}, 1\right\}. \tag{1}$$

Given i.i.d. uniform $[0,1]$ random variables $U_{ij}, i \geq 1, j \geq 1$, we define the total number of edges E_n in the graph as

$$E_n := \sum_{i=1,j=1,i\neq j}^{n} \mathbb{1}(U_{ij} \leq \min\{X_i X_j/n, 1\}), \tag{2}$$

where $\mathbb{1}$ denotes the indicator function. The mean of E_n grows as μn. The specific purpose of this study is to investigate the probability that E_n has significantly more edges than usual, i.e.

$$\mathbb{P}(E_n > (\mu + a)n)$$

for some fixed $a > 0$. Our broader aim is to contribute to a better understanding of large-deviations properties of random graphs with power-law degrees. In the past decade there has been increased activity in establishing large deviations for random graphs. There now exist various large-deviations results for dense graphs and sparse graphs with light-tailed degrees [6, 8, 9, 13, 17], which do not cover scale-free graphs. The typical behavior of scale-free graphs is subject to intense research activity [21, 22], while their large-deviations analysis is so far restricted to the Pagerank functional [11, 18] or the cluster sizes for critical random graphs [23].

To describe our main results, we introduce additional notation. Denote the mean M_n of E_n, conditional on the weights $X_1, \ldots, X_n$ by

$$M_n := \sum_{i=1,j=1,i\neq j}^{n} \min\{X_i X_j/(\mu n), 1\}, \tag{3}$$

and set $S_n = \mu n M_n$, i.e.

$$S_n := \sum_{i=1,j=1,i\neq j}^{n} \min\{X_i X_j, \mu n\}. \tag{4}$$

We now give a description of our main results. A key parameter is

$$k(a) := \lceil a/(2\mu) \rceil. \tag{5}$$

Assuming that $a/(2\mu)$ is not an integer, we show that the most likely way for S_n to reach a value exceeding $(\mu^2 + a)n$ is by k large (of order n) values of X_i, an event which has probability $O(n^k \mathbb{P}(X_1 > n)^k)$. In particular, if $X_1, \ldots, X_k$ equal $a_1 n, \ldots, a_k n$ and the remaining X_i have a typical value, S_n is approximately equal to $n^2(\mu^2 + 2\sum_{i=1}^k \mathbb{E}[\min\{a_i X_j, \mu\})$ (ignoring terms that are of lower order in n). Following the intuition from large deviations for heavy-tailed random variables (see e.g. [20]) we need to choose k as the smallest number such that there exist constants $a_1, \ldots, a_k$ to get $2\sum_{i=1}^k \mathbb{E}[\min\{a_i X_j, \mu\}) > a$. This leads to the choice $k = k(a)$. To state our results formally, we define

$$C(a_1, \ldots, a_k) := 2\sum_{i=1}^{k} \mathbb{E}[\min\{a_i X_j, \mu\}] \tag{6}$$

and we let, for $b > 0$, $X_i^b, i \geq 1$, be an i.i.d. sequence such that $\mathbb{P}(X_i^b > x) = (x/b)^{-\alpha}, x \geq b$. With $f(n) \sim g(n)$ we denote that the ratio of f and g converges to 1 as $n \to \infty$. We first state our main result on S_n:

Proposition 1.1. *Assume that $a/2\mu$ is not an integer. Set $\eta(a)$ as the smallest number η for which $2((k(a) - 1)\mu + \mathbb{E}[\min\{\eta X_1, \mu\}) \geq a$. Then*

$$\mathbb{P}(S_n > (\mu^2 + a)n^2) \sim \eta(a)^{-k(a)\alpha} \mathbb{P}(C(X_1^{\eta(a)}, \ldots, X_{k(a)}^{\eta(a)} \geq a)(n\mathbb{P}(X_1 > n))^{k(a)}. \tag{7}$$

S_n only involves randomness from the vertex weights X_i, while E_n also involves randomness from the uniform random variables in (2). Our main result, derived from Proposition 1.1, shows that the tail of E_n behaves the same as the one of M_n:

Theorem 1.2. *Suppose that $a/2$ is not an integer. Then*

$$\mathbb{P}(E_n > (\mu + a)n) \sim \mathbb{P}(M_n > (\mu + a)n) = \mathbb{P}(S_n > (\mu^2 + a\mu)n^2). \tag{8}$$

Thus, $\mathbb{P}(E_n > (\mu + a)n)$ is regularly varying of index $-\lceil a/2 \rceil(\alpha - 1)$ if $a/2$ is non-integer. The intuition behind this result is similar to the intuition given for S_n, combined with the insight that the additional randomness generated by the uniform random variables $U_{ij}, i, j \geq 1$ is of lesser importance: the event that the number of edges exceeds $(\mu + a)n$ is caused by $k = \lceil a/2 \rceil$ hubs, i.e. vertices with nodes of degree of order n. More in particular, our proofs give the insight that the degrees of the k hubs, normalized by n, converge weakly to $(X_1^{\eta(a/\mu)}, \ldots, X_k^{\eta(a/\mu)})$ conditioned upon $C(X_1^{\eta(a/\mu)}, \ldots, X_k^{\eta(a/\mu)}) \geq a$.

To prove Theorem 1.2, we use well-known concentration bounds for non-identically distributed Bernoulli random variables to show that E_n and M_n are close, facilitated by an estimate for the lower tail of S_n. It is difficult to get rid on the integrality condition in Theorem 1.2, as this is where a transition occurs between the number of hubs that are needed. We are able to derive a weaker result, namely a large-deviations principle. Define $I(x) = (\alpha - 1)\lceil x/2 \rceil$ if $x \geq 0$ and ∞ otherwise. Although I is discontinuous on its effective domain, it is lower semi-continuous, so that I is a rate function. Define $\bar{E}_n = E_n/n - \mu$.

Corollary 1.3. *$\bar{E}_n, n \geq 1$, satisfies a large-deviations principle with speed $\log n$ and rate function I, i.e.*

$$-\inf_{x \in \mathring{A}} I(x) \leq \liminf_{n\to\infty} \frac{\log \mathbb{P}(\bar{E}_n \in A)}{\log n} \leq \limsup_{n\to\infty} \frac{\log \mathbb{P}(\bar{E}_n \in A)}{\log n} \leq -\inf_{x \in \overline{A}} I(x). \tag{9}$$

Our results constitute another case where a rare event in the presence of heavy tails is caused by multiple big jumps. Other heavy-tailed systems exhibiting rare events with multiple big jumps are exit problems [4], fluid networks [10, 24], multi-server queues [3, 14, 15], and reinsurance problems [1]. For sample-path large deviations of heavy-tailed random walks, see [20].

In the random geometric graph, large deviations of the number of edges are caused by one large clique, due to the geometric nature of the model [9]. We here show that power-law random graphs on the other hand, are more likely to contain a large amount of edges due to the presence of hubs. It would therefore be interesting to investigate large deviations of edge counts for models with both geometry and power-law degrees, such as the hyperbolic random graph [16] or geometric inhomogeneous random graphs [7].

The rest of this article is organized as follows. In Section 2 we gather some preliminary results from the literature needed for our proofs. The proof of Proposition 1.1 is developed in Section 3. The proof of Theorem 1.2 is presented in Section 4. The proof of Corollary 9 is given in Section 5.

2 Preliminary results

The following lemma is a key estimate for sums of truncated heavy-tailed random variables, which is a reformulation of Lemma 3 in [19].

Lemma 2.1. *For every $\delta > 0$ and $\beta < \infty$ there exists an $\epsilon > 0$ such that*

$$\mathbb{P}(\sum_{i=1}^{n} X_i > (\mu + \delta)n, X_i \leq \epsilon n, i = 1, \ldots, n) = o(n^{-\beta}). \tag{10}$$

We proceed by stating a version of Chernoff's bound for sums of independent Bernoulli random variables. The statement is a variation of Theorem A.1.4 in [2].

Lemma 2.2. *Let $B_i, i \geq 1$ be a sequence of independent Bernoulli random variables with $p_i = \mathbb{P}(B_i = 1) = 1 - \mathbb{P}(B_i = 0)$. Set $\mu_n = \sum_{i=1}^{n} p_i$. For every $b > 0$ we have*

$$\mathbb{P}(\sum_{i=1}^{n} B_i > (1 + b)\mu_n) \leq e^{-\mu_n I_B(b)}, \qquad \mathbb{P}(\sum_{i=1}^{n} B_i < (1 - b)\mu_n) \leq e^{-\mu_n I_B(-b)}, \tag{11}$$

with $I_B(b) = (1 + b)\log(1 + b) - b$.

We finally state an elementary tail bound for binomially distributed random variables.

Lemma 2.3. *Suppose $B(n, p)$ has a binomial distribution with parameters n and p. Then*

$$\mathbb{P}(B(n, p) \geq m) \leq e^{np}(np)^m. \tag{12}$$

Proof.

$$\sum_{i=m}^{n} \binom{n}{i} p^i (1 - p)^{n-i} \leq \sum_{i=m}^{\infty} \frac{1}{i!}(np)^i = (np)^m \sum_{i=m}^{\infty} \frac{1}{i!}(np)^{i-m} \leq (np)^m e^{np}.$$

□

3 Proof of Proposition 1.1

Throughout this section, we fix a such that $a/(2\mu)$ is not an integer and write $k(a) = k, \eta(a) = \eta$. Define for $\epsilon > 0$:

$$N_{n,\epsilon} := \#\{i \leq n : X_i > \epsilon n\}. \tag{13}$$

The idea of the proof is to subsequently rule out the events $N_{n,\epsilon} < k$ and $N_{n,\epsilon} > k$. After that, we condition on $N_{n,\epsilon} = k$ to work out the remaining technical details. This will be the focus of the next three lemmas which together form the proof of Proposition 1.1.

Lemma 3.1. *For ϵ sufficiently small*

$$\mathbb{P}(S_n > (\mu^2 + a)n^2; N_{n,\epsilon} \leq k-1) = o((n\mathbb{P}(X_1 > n))^k). \tag{14}$$

Proof. We prove this lemma by suitably upper bounding S_n in order to invoke Lemma 2.1. Let $m \leq k$. Set for fixed ϵ the event $A_m := \{X_i > \epsilon n, i < m; X_i \leq \epsilon n, i \geq m\}$. Write

$$\mathbb{P}(S_n > (\mu^2 + a)n^2; N_{n,\epsilon} = m-1) = \binom{n}{m-1}\mathbb{P}(S_n > (\mu^2 + a)n^2; A_m) \tag{15}$$

On the event A_m, we have

$$\begin{aligned} S_n &\leq \mu n(m-1) + \sum_{i,j \geq m, i \neq j} \min\{X_i X_j, \mu n\} + 2 \sum_{i<m, j \geq m} \min\{X_i X_j, \mu n\} \\ &\leq \mu n(m-1) + (\sum_{i \geq m} X_i)^2 + 2(m-1)n^2\mu. \end{aligned}$$

Thus,

$$\mathbb{P}(S_n > (\mu^2 + a)n^2; A_m) \leq \mathbb{P}(\sum_{i \geq m} X_i > \sqrt{\mu^2 n^2 + (a - 2(m-1)\mu)n^2 - \mu n(m-1)}; A_m). \tag{16}$$

Recalling $k = \lceil a/(2\mu) \rceil$, we obtain $a/(2\mu) > k-1 \geq m-1$, and therefore $(a - 2(m-1)\mu) > 0$. Consequently, there exists a $\zeta > 0$ such that for sufficiently large n,

$$\sqrt{\mu^2 n^2 + (a - 2(m-1)\mu)n^2 - \mu n(m-1)} > (\mu + \zeta)n.$$

We can now bound (16) for n large, by

$$\mathbb{P}(S_n > (\mu^2 + a)n^2; A_m) \leq \mathbb{P}(\sum_{i \geq m} X_i > (\mu + \zeta)n) = o(\mathbb{P}(X_1 > n)^k), \tag{17}$$

for suitably small ϵ, where we applied Lemma 2.1 in the last equality. Invoking (15) and summing the estimates over $m = 1, \ldots, k$ gives the desired result. $\square$

Lemma 3.2. *For ϵ sufficiently small,*

$$\mathbb{P}(S_n > (\mu^2 + a)n^2; N_{n,\epsilon} \geq k+1) = o((n\mathbb{P}(X_1 > n))^k). \tag{18}$$

Proof. We observe that $N_{n,\epsilon}$ has Binomial distribution with parameters n and $\mathbb{P}(X_1 > \epsilon n)$ and invoke Lemma 2.3:

$$\mathbb{P}(S_n > (\mu^2 + a)n^2; N_{n,\epsilon} \geq k+1) \leq \mathbb{P}(N_{n,\epsilon} \geq k+1) \leq e^{n\mathbb{P}(X_1 > n)}(n\mathbb{P}(X_1 > n))^{k+1}.$$

The RHS is $o((n\mathbb{P}(X_1 > n))^k)$. $\square$

We are left to consider $\mathbb{P}(S_n > (\mu^2 + a)n^2; N_{n,\epsilon} = k)$. Recall that η is the smallest number for which $2((k-1)\mu + \mathbb{E}[\min\{\eta X_1, \mu\}) \geq a$.

Lemma 3.3.

$$\lim_{n \to \infty} \frac{\mathbb{P}(S_n > (\mu^2 + a)n^2; N_{n,\epsilon} = k)}{(n\mathbb{P}(X_1 > n))^k} = \eta^{-k\alpha}\mathbb{P}(C(X_1^\eta, \ldots, X_k^\eta \geq a). \tag{19}$$

Proof. Write

$$\mathbb{P}(S_n > (\mu^2 + a)n^2; N_{n,\epsilon} = k) = \binom{n}{k}\mathbb{P}(S_n > (\mu^2 + a)n^2; A_{k+1}). \tag{20}$$

To analyze $\mathbb{P}(S_n > (\mu^2+a)n^2; A_{k+1})$, define the random variable $S_n(x_1,\ldots,x_k)$ as S_n conditioned on $X_i = x_i n, i = 1,..,k$. Recall that $C(a_1,\ldots,a_k) = 2\sum_{i=1}^k \mathbb{E}[\min\{a_i X_j, \mu\}]$. From the weak law of large numbers, it follows that

$$\mathbb{P}(S_n(x_1,\ldots,x_k) > (\mu^2+a)n^2; A_{k+1}) \to I(C(x_1,\ldots,x_k) \geq a) \tag{21}$$

for all points $(x_1,\ldots,x_k)$ at which the RHS is continuous.

Next, recall that $X_i^\epsilon, i \geq 1$ are i.i.d. random variables with support on $[\epsilon,\infty)$ such that $\mathbb{P}(X_i^\epsilon > y) = (y/\epsilon)^{-\alpha}$. Write $\mathbb{P}(S_n > (\mu^2+a)n^2; A_{k+1})$ as

$$\mathbb{P}(X_1 > \epsilon n)^k \int_{(\epsilon,\infty)^k} \mathbb{P}(S_n(x_1,\ldots,x_k) > (\mu^2+a)n^2; X_i < \epsilon n, i > k) d\prod_{i=1}^k \mathbb{P}(X_i/n \leq x_i \mid X_i > \epsilon n).$$

Since $\mathbb{P}(X_i/n \leq x_i \mid X_i > \epsilon n)$ converges to the continuous distribution $\mathbb{P}(X_i^\epsilon \leq x_i)$, and using (21), we see that the integral in the last display converges to $\mathbb{P}(C(X_1^\epsilon,\ldots,X_k^\epsilon) \geq a)$.

Because $a/2k$ is non-integer, it follows that $C(\epsilon,\infty,\ldots,\infty) = 0$ for all $\epsilon < \eta$. Since C is symmetric, a similar property holds for the other coordinates. Therefore, if $\epsilon < \eta$,

$$\mathbb{P}(C(X_1^\epsilon,\ldots,X_k^\epsilon) \geq a) = (\eta/\epsilon)^{-k\alpha}\mathbb{P}(C(X_1^\eta,\ldots,X_k^\eta) \geq a). \tag{22}$$

Furthermore, by regular variation,

$$\mathbb{P}(X_1 > \epsilon n)^k \sim (\eta/\epsilon)^{k\alpha}\mathbb{P}(X_1 > \eta n)^k. \tag{23}$$

Putting everything together, we conclude that

$$\mathbb{P}(S_n > (\mu^2+a)n^2; A_{k+1}) \sim \mathbb{P}(C(X_1^\eta,\ldots,X_k^\eta) \geq a)\mathbb{P}(X_i > \eta n)^k. \tag{24}$$

The lemma now follows from (20) and the fact that $\binom{n}{k} \sim n^k$. □

4 Proof of Theorem 1.2

The proof of Theorem 1.2 is based on suitably bounding the difference between E_n and its conditional mean $M_n = S_n/\mu n$, using the concentration bounds in Lemma 2.2. For this procedure to work, we need an asymptotic estimate for the lower tail of S_n. Since $X_i, i \geq 1$, are non-negative random variables, this estimate is considerably easier to obtain than the upper tail.

Lemma 4.1. *For each $a > 0$, there exists a $\delta > 0$ such that*

$$\mathbb{P}(S_n \leq (\mu^2-a)n^2) = O(e^{-\delta n}). \tag{25}$$

Proof. Define $X_i^M = \min\{X_i, M\}$. Let $M < \infty$ be large enough such that $(\mathbb{E}[\min\{X_i, M\}])^2 \geq \mu^2 - a/2$. Observe that, for sufficiently large n,

$$S_n = \sum_{i,j,i\neq j} \min\{X_i X_j, \mu n\} \geq \sum_{i,j,i\neq j} X_i^M X_j^M \geq (\sum_{i=1}^n X_i^M)^2 - nM^2. \tag{26}$$

The estimate (25) now follows by an application of Chernoff's bound to $\sum_{i=1}^n -X_i^M$. □

Proof of Theorem 1.2. Conditional on $X_1,\ldots,X_n$, the variables $B_{ij}, i \neq j$, indicating whether there is an edge between node i and j, are independent. Therefore, observing $M_n = S_n/(n\mu) = \mathbb{E}[E_n \mid X_1,\ldots,X_n]$, we can apply Lemma 2.2 to obtain

$$\mathbb{P}(|E_n - M_n| > bM_n \mid X_1,\ldots,X_n) \leq 2e^{-M_n J(b)} \tag{27}$$

almost surely, with $J(b) = \min\{I_B(b), I_B(-b)\}$. Now, write for fixed $\epsilon > 0$,

$$\mathbb{P}(E_n > (\mu + a)n) = \mathbb{P}(E_n > (\mu + a)n; |E_n - M_n| \leq \epsilon M_n) + \mathbb{P}(E_n > (\mu + a)n; |E_n - M_n| > \epsilon M_n). \tag{28}$$

Invoking Lemma 4.1, the second term on the RHS of (28) is smaller than

$$\mathbb{P}(|E_n - M_n| > \epsilon M_n; M_n > \eta n) + \mathbb{P}(M_n \leq \eta n) \leq 2e^{-\eta n J(b)} + O(e^{-\delta n}) \tag{29}$$

for some $\delta > 0$ depending on $\eta > 0$, the latter chosen suitably small. We conclude that (making δ smaller than $\eta J(b)$ if needed)

$$\mathbb{P}(E_n > (\mu + a)n) = \mathbb{P}(E_n > (\mu + a)n; |E_n - M_n| \leq \epsilon M_n) + O(e^{-\delta n}). \tag{30}$$

We use this identity to prove asymptotic lower and upper bounds which together complete the proof of Theorem 1.2. Invoking (30) and Proposition 1.1, we see that

$$\limsup_{n\to\infty} \frac{\mathbb{P}(E_n > (\mu + a)n)}{\mathbb{P}(M_n > (\mu + a)n)} \leq \limsup_{n\to\infty} \frac{\mathbb{P}(M_n > (\mu + a - \epsilon)n)}{\mathbb{P}(M_n > (\mu + a)n)} = \left(\frac{\mu + a}{\mu + a - \epsilon}\right)^{k(\alpha-1)}, \tag{31}$$

which converges to 1 if $\epsilon \downarrow 0$, providing the upper bound. The lower bound uses that

$$\mathbb{P}(E_n > (\mu + a)n; |E_n - M_n| < \epsilon M_n) \geq \mathbb{P}(M_n > (\mu + a + \epsilon)n) - \mathbb{P}(|E_n - M_n| > \epsilon M_n). \tag{32}$$

The second term on the RHS is exponentially small in n, as shown in (29). Consequently, invoking Proposition 1.1,

$$\liminf_{n\to\infty} \frac{\mathbb{P}(E_n > (\mu + a)n)}{\mathbb{P}(M_n > (\mu + a)n)} \geq \limsup_{n\to\infty} \frac{\mathbb{P}(M_n > (\mu + a + \epsilon)n)}{\mathbb{P}(M_n > (\mu + a)n)} = \left(\frac{\mu + a}{\mu + a + \epsilon}\right)^{k(\alpha-1)}, \tag{33}$$

which converges to 1 if $\epsilon \downarrow 0$, providing the lower bound. □

5 Proof of Corollary 1.3

As a first step we show that the left tail of E_n is lighter than polynomial.

Lemma 5.1. *For each $a > 0$, there exists a $\delta > 0$ such that*

$$\mathbb{P}(E_n \leq (\mu - a)n) = O(e^{-\delta n}). \tag{34}$$

Proof. Write

$$\mathbb{P}(E_n \leq (\mu - a)n) \leq \mathbb{P}(M_n - E_n \geq na/2; M_n \geq (\mu - a/2)n) + \mathbb{P}(M_n \leq (\mu - a/2)n) \tag{35}$$

The second term is exponentially small in n due to Lemma 4.1. To analyze the first term, apply Lemma 2.2 to obtain

$$\mathbb{P}(E_n - M_n \leq -na/2/M_n; M_n \geq (\mu - a/2)n) \leq E[e^{-M_n I(-na2/M_n)} I(M_n \geq (\mu - a/2)n)] \tag{36}$$

Observe that

$$yI(-na2/y) \geq n[(\mu = a)\log(1 - (a/2)/(\mu - a/2)) + a/2] =: na_0$$

if $y \geq (\mu - a/2)n$, and we can take a sufficiently small if needed to have $a_0 > 0$. □

Proof of Corollary 1.3. Consider first A closed. If $0 \in A$, the upper bound is trivial. If $0 \notin A$ we can write $A = A_- \cup A_+$, with $a_- = \sup A_- < 0$ and $a_+ = \inf A_+ > 0$. Since A is closed and $0 \notin A$, both a_- and a_+ are elements of A, and $a_- < 0 < a_+$. Next, note that

$$\mathbb{P}(\bar{E}_n \in A) \leq \mathbb{P}(\bar{E}_n \leq a_-) + \mathbb{P}(\bar{E}_n \geq a_+).$$

Invoking Lemma 5.1, the first term is exponentially small in n. By Theorem 1.2, the second term is regularly varying with exponent $(\alpha-1)\lceil a_+/2\rceil$ if $a_+/2$ is not an integer. If $a_+/2$ is an integer, we can make a_+ a bit smaller, while keeping $\lceil a_+/2\rceil$ fixed, preserving the upper bound for $\mathbb{P}(\bar{E}_n \in A)$. This yields, using $\log L(n)/\log n \to 0$, and abbreviating the constant in Theorem 1.2 with $a = a_+$ by K,

$$\begin{aligned}\limsup_{n\to\infty} \frac{\log \mathbb{P}(\bar{E}_n \in A)}{\log n} &\leq \limsup_{n\to\infty} \frac{\log[\mathbb{P}(\bar{E}_n \leq a_-) + \mathbb{P}(\bar{E}_n \geq a_+)]}{\log n} \\ &\leq \limsup_{n\to\infty} \frac{\log[O(e^{-\delta n}) + K(nP(X_1 > a_+ n))^k]}{\log n} \\ &= -(\alpha-1)\lceil a_+/2\rceil = -\inf_{x\in A} I(x).\end{aligned}$$

Assume now that A is open. If $\sup A \leq 0$ the result is straightforward, so assume that $\sup A > 0$. For every $\epsilon > 0$, we can pick the following subset of A: take a such that $a \in A$; and $\inf_{x\in A} I(x) \geq I(a) - \epsilon$. Since A is open, we may modify the constant a slightly such that $a/2$ is non-integer. Next, take a sufficiently small constant b such that the ball around a with radius b is in A, such that $a - b/2$ and $a + b/2$ are both non-integer, and have the same integer part. Now, observe that

$$\mathbb{P}(\bar{E}_n \in A) \geq \mathbb{P}(\bar{E}_n \in (a-b/2, a+b/2)) = \mathbb{P}(E_n > n(\mu+a-b/2)) - \mathbb{P}(E_n \geq n(\mu+a+b/2)). \quad (37)$$

Due to Theorem 1.2, the RHS is regularly varying with index $I(a)$. Taking logarithms, the lim inf and letting $\epsilon \downarrow 0$ completes the proof. ☐

References

[1] H. Albrecher, B. Chen, E. Vatamidou, and B. Zwart. Finite-time ruin probabilities under large-claim reinsurance treaties for heavy-tailed claim sizes. *Journal of Applied Probability*, 57(2):513–530, 2020.

[2] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2016.

[3] M. Bazhba, J. Blanchet, C.-H. Rhee, and B. Zwart. Queue length asymptotics for the multiple-server queue with heavy-tailed Weibull service times. *Queueing Systems*, 93(3-4):195–226, 2019.

[4] A. Bhattacharya, Z. Palmowski, and B. Zwart. Persistence of heavy-tailed sample averages: principle of infinitely many big jumps. *Electronic Journal in Probability*, 27:Paper No. 50, 25, 2022.

[5] M. Boguñá and R. Pastor-Satorras. Class of correlated random networks with hidden variables. *Phys. Rev. E*, 68:036112, 2003.

[6] C. Bordenave and P. Caputo. Large deviations of empirical neighborhood distribution in sparse random graphs. *Probab. Theory Related Fields*, 163(1-2):149–222, 2015.

[7] K. Bringmann, R. Keusch, and J. Lengler. Geometric inhomogeneous random graphs. *Theoretical Computer Science*, 760:35–54, feb 2019.

[8] S. Chatterjee. *Large deviations for random graphs*, volume 2197 of *Lecture Notes in Mathematics*. Springer, Cham, 2017. Lecture notes from the 45th Probability Summer School held in Saint-Flour, June 2015, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].

[9] S. Chatterjee and M. Harel. Localization in random geometric graphs with too many edges. *The Annals of Probability*, 48(2):574–621, 2020.

[10] B. Chen, J. Blanchet, C.-H. Rhee, and B. Zwart. Efficient rare-event simulation for multiple jump events in regularly varying random walks and compound Poisson processes. *Mathematics of Operations Research*, 44(3):919–942, 2019.

[11] N. Chen, N. Litvak, and M. Olvera-Cravioto. Generalized PageRank on directed configuration networks. *Random Structures Algorithms*, 51(2):237–274, 2017.

[12] F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *PNAS*, 99(25):15879–15882, 2002.

[13] K. Doku-Amponsah and P. Mörters. Large deviation principles for empirical measures of colored random graphs. *The Annals of Applied Probability*, 20(6).

[14] S. Foss and D. Korshunov. Heavy tails in multi-server queue. *Queueing Systems*, 52(1):31–48, 2006.

[15] S. Foss and D. Korshunov. On large delays in multi-server queues with heavy tails. *Mathematics of Operations Research*, 37(2):201–218, 2012.

[16] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguná. Hyperbolic geometry of complex networks. *Phys. Rev. E*, 82(3):036106, 2010.

[17] Q. Liu and Z. Dong. Large deviations for empirical measures of generalized random graphs. *Communications in Statistics - Theory and Methods*, 51(8):2676–2687.

[18] M. Olvera-Cravioto. PageRank's behavior under degree correlations. *The Annals of Applied Probability*, 31(3):1403–1442, 2021.

[19] S. Resnick and G. Samorodnitsky. Activity periods of an infinite server queue and performance of certain heavy tailed fluid queues. volume 33, pages 43–71. 1999. Queues with heavy-tailed distributions.

[20] C.-H. Rhee, J. Blanchet, and B. Zwart. Sample path large deviations for Lévy processes and random walks with regularly varying increments. *The Annals of Probability*, 47(6):3551–3605, 2019.

[21] R. van der Hofstad. *Random graphs and complex networks. Vol. 1.* Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge, 2017.

[22] R. van der Hofstad. *Random graphs and complex networks. Vol. 2.* Cambridge University Press, Cambridge, 2022+.

[23] R. van der Hofstad, S. Kliem, and J. S. H. van Leeuwaarden. Cluster tails for critical power-law inhomogeneous random graphs. *Journal of Statistical Physics*, 171(1):38–95.

[24] B. Zwart, S. Borst, and M. Mandjes. Exact asymptotics for fluid queues fed by multiple heavy-tailed on–off flows. *The Annals of Applied Probability*, 14(2):903 – 957, 2004.