

A moment-generating formula for Erdős-Rényi component sizes

Balázs Ráth*

Abstract

We derive a simple formula characterizing the distribution of the size of the connected component of a fixed vertex in the Erdős-Rényi random graph which allows us to give elementary proofs of some results of [9] and [12] about the susceptibility in the subcritical graph and the CLT [17] for the size of the giant component in the supercritical graph.

Keywords: Erdős-Rényi graph; generating function; susceptibility; giant component; central limit theorem.

AMS MSC 2010: 60C05; 60F05; 82B26; 05A15.

Submitted to ECP on July 30, 2017, final version accepted on March 12, 2018.

Supersedes arXiv:1707.05169.

1 Introduction

The Erdős-Rényi graph $\mathcal{G}_{n,p}$, introduced in [8], is the random graph on n vertices where each pair of vertices is connected with probability p , independently from each other. For an introduction to this fundamental mathematical model of large networks, see [6, 13, 11].

We denote by $\mathbb{P}_{n,p}$ the law of $\mathcal{G}_{n,p}$ and $\mathbb{E}_{n,p}$ the corresponding expectation.

We assume that the vertex set of $\mathcal{G}_{n,p}$ is $[n] = \{1, \dots, n\}$ and we denote by \mathcal{C} the connected component in $\mathcal{G}_{n,p}$ of the vertex indexed by 1. We denote by $|\mathcal{C}|$ the number of vertices of \mathcal{C} .

For any $n \in \mathbb{N}$, $p \in [0, 1]$, $j \in \mathbb{Z} \cap (-n, +\infty)$, and $k \in [n]$ we define

$$g_{n,p}(j, k) = (1-p)^{jk} \prod_{i=0}^{k-1} \frac{n-i+j}{n-i}. \quad (1.1)$$

The central result of this short note is the following formula:

Proposition 1.1. *For any $n \in \mathbb{N}$, $j \in \mathbb{Z} \cap (-n, +\infty)$ and $p \in [0, 1]$ we have*

$$\mathbb{E}_{n,p} [g_{n,p}(j, |\mathcal{C}|)] = \frac{n+j}{n} (1 - \mathbb{P}_{n+j,p} [|\mathcal{C}| > n]). \quad (1.2)$$

Note that if $j \leq 0$ then the r.h.s. is simply $\frac{n+j}{n}$. We prove Proposition 1.1 in Section 2.

*MTA-BME Stochastics Research Group, Budapest University of Technology and Economics, Hungary.
E-mail: rathb@math.bme.hu

Remark 1.2. (i) Define the $n \times n$ matrix M by $M_{j,k} = g_{n,p}(j, k)$ for $j \in \mathbb{Z} \cap (-n, 0]$ and $k \in [n]$. The matrix M is triangular with non-zero diagonal entries, hence it is invertible. Therefore, Proposition 1.1 uniquely characterizes the distribution of $|\mathcal{C}|$ under $\mathbb{P}_{n,p}$.

(ii) A generalization of Proposition 1.1 appears in Proposition 1.6 of the recent preprint [14], see also [14, Remark 1.7]. The random graph process studied in [14] can be informally defined as follows: starting from the empty graph on the vertex set $[n]$, cliques are added with a rate that only depends on their size (the dynamical Erdős-Rényi graph is the special case when only cliques of size two are added).

Proposition 1.1 allows us to give short and self-contained proofs of some delicate results about the sizes of connected components of the Erdős-Rényi graph in the subcritical (see Theorem 1.4) as well as the supercritical (see Theorem 1.6) cases. First, we give a short non-rigorous demonstration of how our formula is used in Remark 1.3.

When we study the phase transition of the Erdős-Rényi graph, it is natural to introduce a parameter $t \in \mathbb{R}_+$ and to study $\mathcal{G}_{n,p}$ for

$$p = p(t, n) = 1 - e^{-t/n}. \tag{1.3}$$

We will fix this relation between p and t throughout this paper.

For any $n \in \mathbb{N}$, $\lambda \in \mathbb{R}$, and $k \in [n]$ we define

$$f_{n,t}(\lambda, k) = \prod_{i=0}^{k-1} e^{-\lambda t} \cdot \left(1 + \frac{\lambda}{1 - \frac{i}{n}} \right), \tag{1.4}$$

so that we have $f_{n,t}(\frac{j}{n}, k) = g_{n,p}(j, k)$ if $j \in \mathbb{Z} \cap (-n, +\infty)$ and thus

$$\mathbb{E}_{n,p} [f_{n,t}(\lambda, |\mathcal{C}|)] \stackrel{(1.2)}{=} (1 + \lambda) (1 - \mathbb{P}_{(1+\lambda)n,p}[|\mathcal{C}| > n]), \quad \lambda \in \frac{\mathbb{Z}}{n} \cap (-1, +\infty). \tag{1.5}$$

Remark 1.3. If we fix $t < 1$ and (non-rigorously) denote $G_t(z) = \lim_{n \rightarrow \infty} \mathbb{E}_{n,p(t,n)}[z^{|\mathcal{C}|}]$ for any $z \in [0, 1]$, then for $\lambda = z - 1$ we (non-rigorously) obtain

$$z = 1 + \lambda \stackrel{(1.5)}{=} \lim_{n \rightarrow \infty} \mathbb{E}_{n,p(t,n)} [f_{n,t}(\lambda, |\mathcal{C}|)] \stackrel{(1.4)}{=} G_t(e^{-\lambda t} \cdot (1 + \lambda)) = G_t(e^{(1-z)t} z). \tag{1.6}$$

Thus $G_t(z) = -W(-e^{-t}tz)/t$, where W is the Lambert-W function. Now it is known that if $p = 1 - e^{-t/n}$ and $n \rightarrow \infty$ then $|\mathcal{C}|$ converges in distribution to the total number of offspring in a subcritical Galton-Watson branching process with $\text{POI}(t)$ offspring distribution (see [4, Theorem 11.6.1]), i.e., $|\mathcal{C}|$ has Borel distribution with parameter t (see [2, Section 2.2] or [12, Section 7]). The generating function G_t of the Borel distribution with parameter t is known to be characterized by the identity $G_t(z) \equiv ze^{(G_t(z)-1)t}$ (see [3, Section 10.4]), which is in turn equivalent to $G_t(z) = -W(-e^{-t}tz)/t$, therefore a more rigorous version of (1.6) can be used to show that the distribution $|\mathcal{C}|$ weakly converges to the Borel distribution with parameter t as $n \rightarrow \infty$.

Now we state our rigorous results. We will use the Bachmann-Landau big O notation: we write $f(n, t) = \mathcal{O}(g(n, t))$ if there exists a universal constant C such that $f(n, t) \leq Cg(n, t)$ for any $n \in \mathbb{N}$ and any t in an explicitly specified domain. We write $f(n) = \mathcal{O}(g(n))$ if there exists a constant C (that may depend on t) such that $f(n) \leq Cg(n)$ for any $n \in \mathbb{N}$.

We will give a short and self-contained proof of some results of [9] and [12]:

Theorem 1.4. For any $t \in [0, 1 - n^{-1/3}]$ we have

$$\mathbb{E}_{n,p}(|\mathcal{C}|) = \frac{1}{1-t} + \frac{t^2 - t}{(1-t)^4} \frac{1}{n} + \mathcal{O}\left(\frac{1}{(1-t)^7} \frac{1}{n^2}\right), \quad (1.7)$$

$$\mathbb{E}_{n,p}(|\mathcal{C}|^2) = \frac{1}{(1-t)^3} + \mathcal{O}\left(\frac{1}{(1-t)^6} \frac{1}{n}\right). \quad (1.8)$$

We will prove Theorem 1.4 in Section 3.

Remark 1.5. $\mathbb{E}_{n,p}(|\mathcal{C}|)$ is often called the *susceptibility* of the Erdős-Rényi graph.

(i) Equation (1.15) of [9, Theorem 1.2] states that if $p = \frac{\mu}{n-1}$ and $0 < \mu < 1$ then

$$\mathbb{E}_{n,p}(|\mathcal{C}|) = \frac{1}{1-\mu} - \frac{2\mu^2 - \mu^4}{2(1-\mu)^4} \frac{1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right). \quad (1.9)$$

Now (1.9) follows from (1.7) if we take into account that $\mu = (n-1)(1 - e^{-t/n})$. The proof of (1.9) in [9, Section 2] uses a coupling of the breadth-first exploration process of \mathcal{C} and a process related to a branching random walk. Our proof of (1.7) is completely different as it only uses Proposition 1.1.

(ii) Equation (1.3) of [12, Theorem 1.1] follows from our (1.7). In fact it already follows from our short Lemma 3.2, see (3.9). Our (1.8) is equivalent to one of the statements about S_3 in [12, Theorem 3.4]. The proofs of these results in [12, Section 3] use differential equations (in the variable t) and are completely different from ours.

(iii) Both statements of Theorem 1.4 give something meaningful in the whole subcritical regime outside the critical window, e.g., the first term of the r.h.s. of (1.7) is much bigger than the second one, which is much bigger than the third one if $(1-t)^3 n \gg 1$.

We also give a short and self-contained proof of the central limit theorem proved in [17] for the size of the giant connected component of $\mathcal{G}_{n,p}$ (see also [5], [11] and [15] for alternative proofs). Our proof only uses Proposition 1.1, see Theorem 1.6 below. We begin with some notation.

Given some $t > 1$ let us define the function $\varphi : [0, 1) \rightarrow \mathbb{R}$ by

$$\varphi(x) = -xt - \ln(1-x). \quad (1.10)$$

Then φ is a convex function satisfying $\varphi(0) = 0$, $\varphi'(0) < 0$ and $\varphi(1_-) = +\infty$.

Given $t > 1$ define $\theta = \theta(t) \in (0, 1)$ to be unique number for which

$$\varphi(\theta) = 0, \quad \text{or, equivalently} \quad e^{t\theta}(1-\theta) = 1. \quad (1.11)$$

Note that $\theta(t)$ is the survival probability of a branching process with $\text{POI}(t)$ offspring distribution (however, our proof of Theorem 1.6 below does not make use of this fact).

We also note that it follows from $\varphi(0) = 0$, $\varphi'(0) < 0$ and $\varphi''(x) > 0$, $x \in [0, 1)$ that

$$0 < \varphi'(\theta) = -t + \frac{1}{1-\theta}. \quad (1.12)$$

Recall the notion of $p = p(t, n) = 1 - e^{-t/n}$ from (1.3).

Theorem 1.6. Let us denote by $|\mathcal{C}_{max}|$ the size of the largest connected component of $\mathcal{G}_{n,p}$. For any $t > 1$ we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} \left[\frac{|\mathcal{C}_{max}| - \theta n}{\sigma \sqrt{n}} \leq x \right] = \Phi(x), \quad \text{where} \quad \sigma = \frac{\sqrt{\theta}}{\varphi'(\theta) \sqrt{1-\theta}} \quad (1.13)$$

and $\Phi(x)$ is the c.d.f. of the standard normal distribution.

We prove Theorem 1.6 in Section 4. Our proof is different from earlier proofs, which use the joint CLT for tree components of various sizes [17], stochastic differential equations which arise in the context of epidemics [15], and exploration processes [5, 11].

Remark 1.7. We believe that Proposition 1.1 can also be used to give elementary alternative proofs of some results of [1] on the sizes of connected components in the critical Erdős-Rényi graph. In particular, let ρ_0^u denote the sigma-finite excursion length measure of the “first” excursion of the Brownian motion with parabolic drift which encodes the block sizes of the standard multiplicative coalescent process at time $u \in \mathbb{R}$ (see [1, (64)]). We believe that if $t := 1 + un^{-1/3}$ and $X_n := |\mathcal{C}|/n^{2/3}$ then the formula

$$\lim_{n \rightarrow \infty} n^{1/3} \mathbb{P}_{n,p(t,n)}[X_n > x] = \rho_0^u(x, +\infty), \quad x \in (0, +\infty) \tag{1.14}$$

can be proved using the methods of this paper, as we now argue. If we fix some $\beta \in \mathbb{R}$ and plug $\lambda := \lfloor \beta n^{2/3} \rfloor / n$ into (1.5) then we obtain (after some calculation) the formula

$$\lim_{n \rightarrow \infty} n^{1/3} \mathbb{E}_{n,p(t,n)} \left[\exp \left(-\beta u X_n - \frac{1}{2} \beta^2 X_n + \frac{1}{2} \beta X_n^2 \right) - 1 \right] = \beta, \quad \beta \in \mathbb{R}. \tag{1.15}$$

Now one can use stochastic calculus to show

$$\int_0^\infty \left(\exp \left(-\beta u x - \frac{1}{2} \beta^2 x + \frac{1}{2} \beta x^2 \right) - 1 \right) d\rho_0^u(x) = \beta \quad \text{for any } \beta \in \mathbb{R}. \tag{1.16}$$

We conjecture that (1.14) can be derived from (1.15) and (1.16).

We discuss the origins of (1.2) in Remark 2.2(i) and an extension of (1.2) to the *stochastic block model* in Remark 2.2(ii).

2 Proof of Proposition 1.1

The proof of Proposition 1.1 will easily follow from the *change of measure* formula (2.1). An idea similar to (2.1) has already been used in the proof of [4, Theorem 11.6.1].

Lemma 2.1. *For any $M, N \in \mathbb{N}$, $p \in [0, 1]$, and $k \in \{1, \dots, N\}$ we have*

$$\mathbb{P}_{M,p}[|\mathcal{C}| = k] = \mathbb{P}_{N,p}[|\mathcal{C}| = k] \cdot (1-p)^{(M-N)k} \prod_{i=1}^{k-1} \frac{M-i}{N-i}. \tag{2.1}$$

Proof. If $k > M$ then both sides of (2.1) are zero. Thus w.l.o.g. we can assume $k \leq M \wedge N$. Now we observe that if we prove (2.1) for some $M \leq N$, then we also obtain (2.1) for $M' = N$ and $N' = M$ by rearranging the formula (2.1), thus we may assume w.l.o.g. that $k \leq M \leq N$. In order to prove (2.1) it is enough to show

$$\binom{M-1}{k-1}^{-1} \mathbb{P}_{M,p}[|\mathcal{C}| = k] \cdot (1-p)^{k(N-M)} = \binom{N-1}{k-1}^{-1} \mathbb{P}_{N,p}[|\mathcal{C}| = k]. \tag{2.2}$$

Now if we denote by $V(\mathcal{C})$ the vertex set of \mathcal{C} then

$$\mathbb{P}_{N,p}[|\mathcal{C}| = k] = \binom{N-1}{k-1} \mathbb{P}_{N,p}[V(\mathcal{C}) = [k]], \tag{2.3}$$

since $\mathbb{P}_{N,p}$ is invariant under the permutation of vertices and there are $\binom{N-1}{k-1}$ subsets of $[N]$ with cardinality k that contain the vertex indexed by 1. Using (2.3) for $\mathbb{P}_{N,p}$ as well as $\mathbb{P}_{M,p}$, the formula (2.2) reduces to showing

$$\mathbb{P}_{M,p}[V(\mathcal{C}) = [k]] \cdot (1-p)^{k(N-M)} = \mathbb{P}_{N,p}[V(\mathcal{C}) = [k]]. \tag{2.4}$$

Now (2.4) holds since $V(\mathcal{C}) = [k]$ in $\mathcal{G}_{N,p}$ if and only if $V(\mathcal{C}) = [k]$ in $\mathcal{G}_{M,p}$ and there are no edges in $\mathcal{G}_{N,p}$ between $[k]$ and $[N] \setminus [M]$. This completes the proof of Lemma 2.1. \square

Proof of Proposition 1.1. For any $n \in \mathbb{N}$, $j \in \mathbb{Z} \cap (-n, +\infty)$ and $p \in [0, 1]$ we have

$$\begin{aligned} \mathbb{E}_{n,p} [g_{n,p}(j, |\mathcal{C}|)] &\stackrel{(1.1)}{=} \frac{n+j}{n} \sum_{k=1}^n \mathbb{P}_{n,p}[|\mathcal{C}| = k] \cdot (1-p)^{jk} \prod_{i=1}^{k-1} \frac{n+j-i}{n-i} \\ &\stackrel{(*)}{=} \frac{n+j}{n} \sum_{k=1}^n \mathbb{P}_{n+j,p}[|\mathcal{C}| = k] = \frac{n+j}{n} (1 - \mathbb{P}_{n+j,p}[|\mathcal{C}| > n]), \end{aligned} \quad (2.5)$$

where in $(*)$ we used (2.1) with $n = N$ and $M = n + j$. The proof of (1.2) is complete. \square

Remark 2.2. (i) Our original proof of Proposition 1.1 used the so-called *rigid representation* of the time evolution of the component size structure of the Erdős-Rényi graph, see [16, Section 6.1.1, Case 1]. In a nutshell, if $Y_k = t - X_k$, $k \in [n]$, where X_1, X_2, \dots, X_n denote independent exponentially distributed random variables $X_k \sim \text{EXP}(1 - \frac{k}{n})$, then $\tau = \min\{k : Y_1 + \dots + Y_k < 0\}$ has the same distribution as $|\mathcal{C}|$ under $\mathbb{P}_{n,p}$, $p = 1 - e^{-t/n}$. We chose to include an elementary proof instead in order to keep the paper self-contained.

(ii) It is possible to extend Proposition 1.1 to the *stochastic block model*, as we now briefly explain. Consider a random graph in which each vertex has a label, where the set of labels is $\{1, \dots, \ell\}$. Let $\underline{n} = (n_1, \dots, n_\ell)$ and $n = n_1 + \dots + n_\ell$. We uniformly choose a labelling of the vertex set $[n]$ from the set of labellings where the number of vertices with label j is n_j for each $j = 1, \dots, \ell$. Given the labels, we add edges independently: a vertex with label i and a vertex with label j is connected with probability $p_{i,j}$. Let $\underline{p} = (p_{i,j})_{i,j=1}^\ell$. Denote by $\mathbb{P}_{\underline{n},\underline{p}}$ the law of the resulting random graph $\mathcal{G}_{\underline{n},\underline{p}}$ and $\mathbb{E}_{\underline{n},\underline{p}}$ the corresponding expectation. This random graph model is often called the *stochastic block model* and it is also a special case of the *inhomogeneous random graph model* of [7].

Denote by $\mathcal{K}(\underline{n})$ the set of vectors $\underline{k} = (k_1, \dots, k_\ell)$ for which $0 \leq k_j \leq n_j$ and $k_1 + \dots + k_\ell \geq 1$. Denote by $\mathcal{J}(\underline{n})$ the set of vectors $\underline{J} = (J_1, \dots, J_\ell)$ for which $-n_j \leq J_j$ and $-n < J_1 + \dots + J_\ell$. Denote by $\mathcal{J}^{\leq 0}(\underline{n})$ the subset of $\mathcal{J}(\underline{n})$ which consists of vectors $\underline{J} = (J_1, \dots, J_\ell)$ for which $J_j \leq 0$ for any $j = 1, \dots, \ell$. Let us define

$$g_{\underline{n},\underline{p}}(\underline{J}, \underline{k}) := \prod_{i,j=1}^\ell (1 - p_{i,j})^{k_i J_j} \cdot \prod_{j=1}^\ell \prod_{i=0}^{k_j-1} \frac{n_j + J_j - i}{n_j - i}, \quad \underline{J} \in \mathcal{J}(\underline{n}), \underline{k} \in \mathcal{K}(\underline{n}). \quad (2.6)$$

Let \mathcal{C} denote the connected component of the vertex indexed by 1 in $\mathcal{G}_{\underline{n},\underline{p}}$. Denote by $|\mathcal{C}|_j$ the number of vertices with label j in \mathcal{C} and let $|\underline{\mathcal{C}}| = (|\mathcal{C}|_1, \dots, |\mathcal{C}|_\ell)$. The generalization of the formula (1.2) to the stochastic block model is

$$\mathbb{E}_{\underline{n},\underline{p}} [g_{\underline{n},\underline{p}}(\underline{J}, |\underline{\mathcal{C}}|)] = \frac{\sum_{j=1}^\ell (n_j + J_j)}{\sum_{j=1}^\ell n_j} \mathbb{P}_{\underline{n}+\underline{J},\underline{p}} [|\mathcal{C}|_1 \leq n_1, \dots, |\mathcal{C}|_\ell \leq n_\ell], \underline{J} \in \mathcal{J}(\underline{n}). \quad (2.7)$$

In order to prove (2.7), one needs the following analogue of (2.1), valid for $\underline{k} \in \mathcal{K}(\underline{N})$:

$$\mathbb{P}_{\underline{M},\underline{p}} [|\underline{\mathcal{C}}| = \underline{k}] = \mathbb{P}_{\underline{N},\underline{p}} [|\underline{\mathcal{C}}| = \underline{k}] \prod_{i,j=1}^\ell (1 - p_{i,j})^{(M_j - N_j)k_j} \frac{\sum_{j=1}^\ell N_j}{\sum_{j=1}^\ell M_j} \prod_{j=1}^\ell \prod_{i=0}^{k_j-1} \frac{M_j - i}{N_j - i} \quad (2.8)$$

Note that if $\underline{J} \in \mathcal{J}^{\leq 0}(\underline{n})$ then the r.h.s. of (2.7) is simply $\frac{\sum_{j=1}^\ell (n_j + J_j)}{\sum_{j=1}^\ell n_j}$. Also note that the analogue of the property stated in Remark 1.2(i) holds: the system of equations (2.7) indexed by $\underline{J} \in \mathcal{J}^{\leq 0}(\underline{n})$ uniquely characterizes the distribution of $|\underline{\mathcal{C}}|$ under $\mathbb{P}_{\underline{n},\underline{p}}$.

3 Proof of Theorem 1.4

The basic idea is to treat $\mathbb{E}_{n,p}[g_{n,p}(j, |\mathcal{C}|)]$ as the generating function of $|\mathcal{C}|$, c.f. Remark 1.3. Thus if we want to obtain information about the first and second moments of $|\mathcal{C}|$, we have to “differentiate” with respect to the variable j twice. Since j can only take integer values, we have to consider the first order discrete differences $g_{n,p}(j, |\mathcal{C}|) - g_{n,p}(0, |\mathcal{C}|)$ for $j = -1$ and $j = -2$ in the proof of Lemmas 3.2 and 3.4, and the second order discrete difference (i.e., the difference of the first order differences) in the proof of Lemma 3.5.

The statement of Lemma 3.1 is equivalent to [12, Lemma 3.2] (which is proved using differential equations), moreover it also classically follows from the fact that $|\mathcal{C}|$ is stochastically dominated by a subcritical branching process if $t < 1$. Despite of this, we chose to include a proof of Lemma 3.1 which only uses Proposition 1.1 in order to keep the paper self-contained.

Recall our convention $p = 1 - e^{-t/n}$ from (1.3).

Lemma 3.1. *If $t \in (0, 1)$ then*

$$\mathbb{E}_{n,p}(|\mathcal{C}|^i) = \mathcal{O}\left(\frac{1}{(1-t)^{2i-1}}\right), \quad i \in \mathbb{N}. \tag{3.1}$$

Proof. W.l.o.g. we assume $t \in [\frac{1}{2}, 1)$ and $\frac{2}{1-t} \leq n$. For any $j \geq 0$ we have

$$\mathbb{E}_{n,p}\left[\left(e^{-tj/n}\left(1 + \frac{j}{n}\right)\right)^{|\mathcal{C}|}\right] \stackrel{(1.4)}{\leq} \mathbb{E}_{n,p}\left[f_{n,t}\left(\frac{j}{n}, |\mathcal{C}|\right)\right] \stackrel{(1.5)}{\leq} 1 + \frac{j}{n}. \tag{3.2}$$

Note that if we let $\tilde{\lambda} := \frac{1}{t} - 1$ then we have

$$\max_{\lambda} e^{-\lambda t}(1 + \lambda) = e^{-\tilde{\lambda}t}(1 + \tilde{\lambda}) = \frac{1}{t}e^{t-1} \stackrel{(*)}{>} e^{\frac{1}{2}(1-t)^2} \quad t \in (0, 1), \tag{3.3}$$

where $(*)$ follows from $-\ln(t) + (t-1) - \frac{1}{2}(t-1)^2 = \int_t^1 \int_s^1 \left(\frac{1}{u^2} - 1\right) du ds > 0$.

Next we show that if we choose $j^* := \lfloor n\tilde{\lambda} \rfloor = \lfloor n \cdot (\frac{1}{t} - 1) \rfloor$ then we have

$$e^{-tj^*/n}\left(1 + \frac{j^*}{n}\right) \geq e^{\frac{1}{4}(1-t)^2}. \tag{3.4}$$

Indeed, if we let $f(x) := -tx + \ln(1+x)$, then we have $f'(\tilde{\lambda}) = 0$ and thus

$$f(\tilde{\lambda}) - f\left(\frac{j^*}{n}\right) = \int_{j^*/n}^{\tilde{\lambda}} \int_x^{\tilde{\lambda}} -f''(y) dy dx = \int_{j^*/n}^{\tilde{\lambda}} \frac{y - j^*/n}{(1+y)^2} dy \stackrel{(**)}{\leq} \frac{1}{n^2} \stackrel{(***)}{\leq} \frac{1}{4}(1-t)^2, \tag{3.5}$$

where $(**)$ follows from $0 \leq j^*$ and $0 \leq \tilde{\lambda} - j^*/n \leq 1/n$, and $(***)$ follows from $\frac{2}{1-t} \leq n$. Now (3.4) follows from (3.3) and (3.5). We are now ready to prove (3.1):

$$\begin{aligned} 1 + \frac{1}{i!} \frac{1}{4^i} (1-t)^{2i} \mathbb{E}_{n,p}(|\mathcal{C}|^i) &\leq \mathbb{E}_{n,p}\left[\sum_{\ell=0}^{\infty} \frac{(\frac{1}{4}(1-t)^2|\mathcal{C}|)^\ell}{\ell!}\right] \\ &= \mathbb{E}_{n,p}\left[e^{\frac{1}{4}(1-t)^2|\mathcal{C}|}\right] \stackrel{(3.4)}{\leq} \mathbb{E}_{n,p}\left[\left(e^{-tj^*/n}\left(1 + \frac{j^*}{n}\right)\right)^{|\mathcal{C}|}\right] \stackrel{(3.2)}{\leq} 1 + \frac{j^*}{n} \leq \frac{1}{t} \quad i \in \mathbb{N}, \end{aligned} \tag{3.6}$$

from which (3.1) follows if $t \in [\frac{1}{2}, 1)$. □

Lemma 3.2. *For any $t \in [0, 1)$ we have*

$$1 = (1-t)\mathbb{E}_{n,p}(|\mathcal{C}|) + \left(t - \frac{t^2}{2}\right) \frac{\mathbb{E}_{n,p}(|\mathcal{C}|^2)}{n} + \left(\frac{t^2}{2} - \frac{t^3}{6}\right) \frac{\mathbb{E}_{n,p}(|\mathcal{C}|^3)}{n^2} + \mathcal{O}\left(\frac{1}{(1-t)^7 n^3}\right). \tag{3.7}$$

Before we prove Lemma 3.2, let us state an immediate corollary.

Corollary 3.3. Applying (3.1) to $\mathbb{E}_{n,p}(|\mathcal{C}|^3)$ in (3.7) we obtain

$$\mathbb{E}_{n,p}(|\mathcal{C}|) = \frac{1}{1-t} + \frac{\frac{t^2}{2} - t}{1-t} \frac{\mathbb{E}_{n,p}(|\mathcal{C}|^2)}{n} + \mathcal{O}\left(\frac{1}{(1-t)^6 n^2}\right), \quad t \in [0, 1 - n^{-1/3}]. \quad (3.8)$$

Applying (3.1) to $\mathbb{E}_{n,p}(|\mathcal{C}|^2)$ in (3.8) we obtain

$$\mathbb{E}_{n,p}(|\mathcal{C}|) = \frac{1}{1-t} + \mathcal{O}\left(\frac{1}{(1-t)^4 n}\right), \quad t \in [0, 1 - n^{-1/3}]. \quad (3.9)$$

Proof of Lemma 3.2. Let $k \in [n]$. We begin by observing that (1.1) is a telescopic product if $j = -1$ and then we apply Taylor expansion:

$$\begin{aligned} g_{n,p}(-1, k) &\stackrel{(1.1),(1.3)}{=} e^{tk/n} \left(1 - \frac{k}{n}\right) = \left(\sum_{i=0}^3 \frac{1}{i!} \frac{t^i k^i}{n^i} + \mathcal{O}\left(\frac{k^4}{n^4}\right)\right) \left(1 - \frac{k}{n}\right) \\ &= 1 + (t-1)\frac{k}{n} + \left(\frac{t^2}{2} - t\right)\frac{k^2}{n^2} + \left(\frac{t^3}{6} - \frac{t^2}{2}\right)\frac{k^3}{n^3} + \mathcal{O}\left(\frac{k^4}{n^4}\right). \end{aligned} \quad (3.10)$$

Combining (3.10) with Proposition 1.1 we obtain

$$\begin{aligned} 1 - \frac{1}{n} &= 1 + (t-1)\frac{\mathbb{E}_{n,p}(|\mathcal{C}|)}{n} \\ &\quad + \left(\frac{t^2}{2} - t\right)\frac{\mathbb{E}_{n,p}(|\mathcal{C}|^2)}{n^2} + \left(\frac{t^3}{6} - \frac{t^2}{2}\right)\frac{\mathbb{E}_{n,p}(|\mathcal{C}|^3)}{n^3} + \mathcal{O}\left(\frac{\mathbb{E}_{n,p}(|\mathcal{C}|^4)}{n^4}\right). \end{aligned} \quad (3.11)$$

Subtracting one from both sides of (3.11), multiplying the result by $-n$ and applying (3.1) to $\mathbb{E}_{n,p}(|\mathcal{C}|^4)$, we obtain (3.7). \square

Lemma 3.4. For any $t \in [0, 1 - n^{-1/3}]$ we have

$$\begin{aligned} -2 &= (2t-2)\mathbb{E}_{n,p}[|\mathcal{C}|] + (1-4t+2t^2)\frac{\mathbb{E}_{n,p}[|\mathcal{C}|^2]}{n} - \frac{\mathbb{E}_{n,p}[|\mathcal{C}|]}{n} \\ &\quad + (2t-4t^2+\frac{4}{3}t^3)\frac{\mathbb{E}_{n,p}[|\mathcal{C}|^3]}{n^2} + \mathcal{O}\left(\frac{1}{(1-t)^7 n^3}\right). \end{aligned} \quad (3.12)$$

Proof. Let $k \in [n]$. We begin with a calculation similar to (3.10):

$$\begin{aligned} g_{n,p}(-2, k) &\stackrel{(1.1),(1.3)}{=} e^{2tk/n} \left(1 - \frac{k}{n}\right) \left(1 - \frac{k}{n-1}\right) \\ &= \left(\sum_{i=0}^3 \frac{1}{i!} \frac{2^i t^i k^i}{n^i} + \mathcal{O}\left(\frac{k^4}{n^4}\right)\right) \left(1 - \frac{k}{n}\right) \left(1 - \frac{k}{n} \left(\sum_{j=0}^2 \frac{1}{n^j} + \mathcal{O}\left(\frac{1}{n^3}\right)\right)\right) \\ &= 1 + (2t-2)\frac{k}{n} + (1-4t+2t^2)\frac{k^2}{n^2} - \frac{k}{n^2} + (2t-4t^2+\frac{4}{3}t^3)\frac{k^3}{n^3} + (1-2t)\frac{k^2}{n^3} - \frac{k}{n^3} + \mathcal{O}\left(\frac{k^4}{n^4}\right). \end{aligned} \quad (3.13)$$

From (3.13) and Proposition 1.1 we obtain

$$\begin{aligned} 1 - \frac{2}{n} &= 1 + (2t-2)\frac{\mathbb{E}_{n,p}[|\mathcal{C}|]}{n} + (1-4t+2t^2)\frac{\mathbb{E}_{n,p}[|\mathcal{C}|^2]}{n^2} - \frac{\mathbb{E}_{n,p}[|\mathcal{C}|]}{n^2} \\ &\quad + (2t-4t^2+\frac{4}{3}t^3)\frac{\mathbb{E}_{n,p}[|\mathcal{C}|^3]}{n^3} + (1-2t)\frac{\mathbb{E}_{n,p}[|\mathcal{C}|^2]}{n^3} - \frac{\mathbb{E}_{n,p}[|\mathcal{C}|]}{n^3} + \mathcal{O}\left(\frac{\mathbb{E}_{n,p}[|\mathcal{C}|^4]}{n^4}\right). \end{aligned} \quad (3.14)$$

Subtracting one from both sides of (3.14), multiplying the result by n and applying (3.1) to the last three terms of (3.14), we obtain (3.12). \square

Lemma 3.5. For any $t \in [0, 1 - n^{-1/3}]$ we have

$$\mathbb{E}_{n,p}(|\mathcal{C}|^2) = \frac{\mathbb{E}_{n,p}(|\mathcal{C}|)}{(1-t)^2} + \mathcal{O}\left(\frac{1}{(1-t)^6 n}\right). \quad (3.15)$$

Proof. Adding (3.12) to twice (3.7) we obtain

$$0 = (1 - 2t + t^2) \frac{\mathbb{E}_{n,p}(|\mathcal{C}|^2)}{n} - \frac{\mathbb{E}_{n,p}(|\mathcal{C}|)}{n} + (t^3 - 3t^2 + 2t) \frac{\mathbb{E}_{n,p}(|\mathcal{C}|^3)}{n^2} + \mathcal{O}\left(\frac{1}{(1-t)^7 n^3}\right). \quad (3.16)$$

Rearranging (3.16) and multiplying by n we obtain

$$\mathbb{E}_{n,p}(|\mathcal{C}|) = (1-t)^2 \mathbb{E}_{n,p}(|\mathcal{C}|^2) + t(t-1)(t-2) \frac{\mathbb{E}_{n,p}(|\mathcal{C}|^3)}{n} + \mathcal{O}\left(\frac{1}{(1-t)^7 n^2}\right). \quad (3.17)$$

Dividing both sides of (3.17) by $(1-t)^2$ we use (3.1) to obtain (3.15). \square

Proof of Theorem 1.4. From (3.9) and (3.15) we obtain (1.8).

Plugging (1.8) into (3.8) we obtain (1.7). \square

4 Proof of Theorem 1.6

We will deduce Theorem 1.6 (i.e., the CLT for $|\mathcal{C}_{max}|$) from Lemma 4.1 (i.e., the CLT for $|\mathcal{C}|$) using the idea of [10, Lemma 2.1]. We deduce Lemma 4.1 from Lemmas 4.2 and 4.3 using that convergence of moment generating functions implies weak convergence of probability distributions. We prove Lemmas 4.2 and 4.3 by viewing (1.5) as a moment generating function identity. The crux of the proof of Lemma 4.2 is (4.19) and the crux of the proof of Lemma 4.3 is (4.33).

Throughout this section we fix $t > 1$. Recall the notion of $\varphi : [0, 1] \rightarrow \mathbb{R}$ from (1.10) and $\theta = \theta(t) \in (0, 1)$ from (1.11). Recall the notion of $p = p(t, n) = 1 - e^{-t/n}$ from (1.3).

We will often use the shorthand \mathbb{P} for $\mathbb{P}_{n,p(t,n)}$ and \mathbb{E} for $\mathbb{E}_{n,p(t,n)}$.

If X is a random variable and A is an event, we will denote $\mathbb{E}(X; A) := \mathbb{E}(X\mathbb{1}_A)$.

Lemma 4.1. Let us define σ as in (1.13). For any $x \in \mathbb{R}$ we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} \left[\frac{|\mathcal{C}| - \theta n}{\sigma \sqrt{n}} \leq x \right] = (1 - \theta) + \theta \Phi(x). \quad (4.1)$$

Before we prove Lemma 4.1, we use it to prove Theorem 1.6.

Proof of Theorem 1.6. Denote by $|\mathcal{C}_1|, |\mathcal{C}_2|, \dots$ the non-increasing rearrangement of the sequence of component sizes of the graph $\mathcal{G}_{n,p}$. Thus $|\mathcal{C}_1| = |\mathcal{C}_{max}|$ and $|\mathcal{C}_2|$ is the size of the second largest component. Note that $|\mathcal{C}_1| = |\mathcal{C}_2|$ is possible, but we will show that $|\mathcal{C}_2| < |\mathcal{C}_1|$ with high probability.

For any $a \in \mathbb{R}$ let us denote

$$k_{n,a} = \lfloor \theta n + a \cdot \sigma \sqrt{n} \rfloor. \quad (4.2)$$

We will show that for $a \leq b \in \mathbb{R}$ we have

$$\lim_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} [|\mathcal{C}_1| \in [k_{n,a}, k_{n,b}], |\mathcal{C}_2| < k_{n,a}] = \frac{1}{\theta} \lim_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} [|\mathcal{C}| \in [k_{n,a}, k_{n,b}]]. \quad (4.3)$$

Now by Lemma 4.1 the right-hand side of (4.3) is $\Phi(b) - \Phi(a)$. This equation readily implies $\liminf_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} [|\mathcal{C}_2| < k_{n,a}] \geq \Phi(b) - \Phi(a)$ for any $a \leq b \in \mathbb{R}$, which in turn implies $\lim_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} [|\mathcal{C}_2| < k_{n,a}] = 1$ for any $a \in \mathbb{R}$. Combining this with Lemma 4.1 and (4.3) we obtain that Theorem 1.6 indeed holds.

In order to prove (4.3) we observe that if $k \in [k_{n,a}, k_{n,b}]$, then

$$\mathbb{P}_{n,p} [|\mathcal{C}_1| = k, |\mathcal{C}_2| < k_{n,a}] = \mathbb{P}_{n-k,p} [|\mathcal{C}_1| < k_{n,a}] \frac{n}{k} \mathbb{P}_{n,p} [|\mathcal{C}| = k]. \quad (4.4)$$

Equation (4.4) is essentially a special case of [10, Lemma 2.1], but we include the proof of (4.4) here for completeness: if $v \in [n]$ and $\mathcal{C}(v)$ denotes the connected component of v in $\mathcal{G}_{n,p}$ and $|\mathcal{C}^*(v)|$ denotes the size of the largest connected component of $\mathcal{G}_{n,p} \setminus \mathcal{C}(v)$ then

$$\begin{aligned} \mathbb{P}_{n,p} [|\mathcal{C}_1| = k, |\mathcal{C}_2| < k_{n,a}] &= \frac{1}{k} \sum_{v=1}^n \mathbb{P}_{n,p} [|\mathcal{C}_1| = k, v \in \mathcal{C}_1, |\mathcal{C}_2| < k_{n,a}] \\ &= \frac{1}{k} \sum_{v=1}^n \mathbb{P}_{n,p} [|\mathcal{C}(v)| = k, |\mathcal{C}^*(v)| < k_{n,a}] = \frac{n}{k} \mathbb{P}_{n,p} [|\mathcal{C}(1)| = k, |\mathcal{C}^*(1)| < k_{n,a}] \\ &= \frac{n}{k} \mathbb{P}_{n,p} [|\mathcal{C}| = k] \mathbb{P}_{n-k,p} [|\mathcal{C}_1| < k_{n,a}]. \end{aligned}$$

This proves (4.4). Next we show that

$$\lim_{n \rightarrow \infty} \min_{k \in [k_{n,a}, k_{n,b}]} \mathbb{P}_{n-k,p(t,n)} [|\mathcal{C}_1| < k_{n,a}] = 1. \quad (4.5)$$

Let us denote $\tilde{n} = n - k_{n,a}$. For any $k \in [k_{n,a}, k_{n,b}]$ we have

$$\mathbb{P}_{n-k,p} [|\mathcal{C}_1| \geq k_{n,a}] \leq \frac{n-k}{k_{n,a}} \mathbb{P}_{n-k,p} [|\mathcal{C}| \geq k_{n,a}] \leq \frac{n}{k_{n,a}} \mathbb{P}_{\tilde{n},p} [|\mathcal{C}| \geq k_{n,a}] \leq \frac{n \mathbb{E}_{\tilde{n},p} [|\mathcal{C}|]}{(k_{n,a})^2}. \quad (4.6)$$

Now we observe that $\mathcal{G}_{\tilde{n},p(t,n)}$ is a *subcritical* Erdős-Rényi graph, since

$$\lim_{n \rightarrow \infty} \tilde{n} \cdot p(t,n) \stackrel{(1.3),(4.2)}{=} \lim_{n \rightarrow \infty} (n - \lfloor \theta n + a \cdot \sigma \sqrt{n} \rfloor) \cdot (1 - e^{-t/n}) = (1 - \theta)t \stackrel{(1.12)}{<} 1.$$

Note that $\mathbb{E}_{\tilde{n},p} [|\mathcal{C}|]$ remains bounded as $n \rightarrow \infty$ by (3.1), hence (4.5) follows from (4.6).

We are now ready to prove (4.3):

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} [|\mathcal{C}_1| \in [k_{n,a}, k_{n,b}], |\mathcal{C}_2| < k_{n,a}] \\ \stackrel{(4.4)}{=} \lim_{n \rightarrow \infty} \sum_{k=k_{n,a}}^{k_{n,b}} \mathbb{P}_{n-k,p(t,n)} [|\mathcal{C}_1| < k_{n,a}] \frac{n}{k} \mathbb{P}_{n,p(t,n)} [|\mathcal{C}| = k] \\ \stackrel{(4.2),(4.5)}{=} \frac{1}{\theta} \lim_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} [|\mathcal{C}| \in [k_{n,a}, k_{n,b}]]. \quad (4.7) \end{aligned}$$

This completes the proof of Theorem 1.6 given Lemma 4.1. □

We will deduce Lemma 4.1 from Lemmas 4.2 and 4.3 below.

Let us subdivide the interval $[n]$ into five disjoint sub-intervals:

$$I_n = [1, n^{1/4}), \quad J_n = [n^{1/4}, n^{3/4}), \quad K_n = [n^{3/4}, \theta n - n^{5/8}), \quad (4.8)$$

$$\tilde{I}_n = [\theta n - n^{5/8}, \theta n + n^{5/8}), \quad \tilde{K}_n = [\theta n + n^{5/8}, n]. \quad (4.9)$$

Note that the choice of the exponents $\frac{1}{4}$, $\frac{3}{4}$ and $\frac{5}{8}$ above is somewhat arbitrary. Also note that I_n and \tilde{I}_n are the important intervals, while J_n , K_n and \tilde{K}_n are insignificant, i.e., we will see that $|\mathcal{C}| \in I_n \cup \tilde{I}_n$ with high probability. The only reason behind the distinction between J_n and K_n is that we will use different methods to show that J_n and K_n are insignificant.

Lemma 4.2. We have

$$\lim_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} (|\mathcal{C}| \in I_n) = 1 - \theta. \tag{4.10}$$

Lemma 4.3. For any $\alpha \in \mathbb{R}$ we have

$$\lim_{n \rightarrow \infty} \left(\mathbb{P}_{n,p(t,n)} (|\mathcal{C}| \in I_n) + \mathbb{E}_{n,p(t,n)} \left(\exp \left(\alpha \varphi'(\theta) \frac{|\mathcal{C}| - \theta n}{\sqrt{n}} - \frac{\alpha^2}{2} \frac{\theta}{1 - \theta} \right); |\mathcal{C}| \in \tilde{I}_n \right) \right) = 1. \tag{4.11}$$

Before we prove Lemmas 4.2 and 4.3, let us deduce Lemma 4.1 from them.

Proof of Lemma 4.1. First note that $\lim_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} (|\mathcal{C}| \in I_n \cup \tilde{I}_n) = 1$ follows from the $\alpha = 0$ case of (4.11). Combining this with (4.10) we obtain

$$\lim_{n \rightarrow \infty} \mathbb{P}_{n,p(t,n)} (|\mathcal{C}| \in \tilde{I}_n) = \theta. \tag{4.12}$$

Denote by μ_n the conditional distribution of $\frac{|\mathcal{C}| - \theta n}{\sqrt{n}}$ given $|\mathcal{C}| \in \tilde{I}_n$. We have

$$\lim_{n \rightarrow \infty} \int \exp(\alpha \varphi'(\theta)x) d\mu_n(x) \stackrel{(4.10),(4.11),(4.12)}{=} \exp \left(\frac{\alpha^2}{2} \frac{\theta}{1 - \theta} \right), \quad \alpha \in \mathbb{R}. \tag{4.13}$$

The r.h.s. of (4.13) is the moment generating function of $\mathcal{N} \left(0, \frac{\theta}{1 - \theta} \right)$, thus it classically follows from (4.13) that μ_n weakly converges to $\mathcal{N} (0, \sigma^2)$ as $n \rightarrow \infty$, where σ appears in (1.13). Together with (4.10) and (4.12) this implies Lemma 4.1, given Lemmas 4.2 and 4.3. \square

We will prove Lemma 4.2 in Section 4.1 and Lemma 4.3 in Section 4.2. The proofs will make excessive use of (1.5). Let us now introduce some notation that will be used throughout.

For any $\lambda \in (-1, +\infty)$ and any $n \in \mathbb{N}$ let us define

$$\lambda_n^* = \frac{1}{n} \lfloor n\lambda \rfloor. \tag{4.14}$$

Now $\lambda_n^* \in \frac{\mathbb{Z}}{n} \cap (-1, +\infty)$, which is required if we want to use (1.5).

Having fixed $t > 1$, we note that λ_n^* approximates λ well, i.e., we have

$$f_{n,t}(\lambda_n^*, k) \stackrel{(1.4)}{=} f_{n,t}(\lambda, k) \exp \left(\mathcal{O} \left(\frac{k}{n} \right) \right), \quad e^{-t} - 1 \leq \lambda \leq 1, \quad 1 \leq k \leq \frac{e^{-t}}{2} n. \tag{4.15}$$

We will often implicitly use that for any $\lambda > -1$ we have

$$f_{n,t}(\lambda_n^*, k) = 0 \quad \text{if } n + \lfloor n\lambda \rfloor < k \leq n \quad \text{and} \quad f_{n,t}(\lambda_n^*, k) \geq 0 \quad \text{if } k \in [n]. \tag{4.16}$$

Having fixed $t > 1$, we note that if we let

$$\tilde{\lambda} := \frac{1}{t} - 1 \quad \text{then we have} \quad \tilde{x} := \max_{\lambda} e^{-\lambda t} (1 + \lambda) = e^{-\tilde{\lambda} t} (1 + \tilde{\lambda}) = \frac{1}{t} e^{t-1} \stackrel{(*)}{>} 1, \tag{4.17}$$

where $(*)$ follows from the inequality $e^x > 1 + x$ applied to $x = t - 1$.

In Sections 4.1 and 4.2 we will dominate $f_{n,t}(\lambda, k)$ by $f_{n,t}(\tilde{\lambda}_n^*, k)$ for $k \in J_n$ (defined in (4.8)) in order to show that “nothing interesting happens” in the interval J_n .

We will write $f(n) = \Omega(g(n))$ if there exists a constant $c > 0$ (that may depend on t) such that $f(n) \geq cg(n)$ for any $n \in \mathbb{N}$.

4.1 Proof of Lemma 4.2

Before we outline the strategy of the proof of Lemma 4.2 in the paragraph below (4.18), we need to introduce some notation. Let us abbreviate

$$X = f_{n,t}(-\theta, |\mathcal{C}|) \quad \text{and} \quad X^* \stackrel{(4.14)}{=} f_{n,t}((-\theta)_n^*, |\mathcal{C}|).$$

Recalling the definition of the intervals I_n and J_n from (4.8), we have

$$1 + (-\theta)_n^* \stackrel{(1.5)}{=} \mathbb{E}[X^*; |\mathcal{C}| \in I_n] + \mathbb{E}[X^*; |\mathcal{C}| \in J_n] + \mathbb{E}[X^*; n^{3/4} \leq |\mathcal{C}|]. \quad (4.18)$$

We will estimate the three terms on the r.h.s. of (4.18). We will show that the first term approximates $\mathbb{P}(|\mathcal{C}| \in I_n)$ as $n \rightarrow \infty$, while the second and third terms vanish as $n \rightarrow \infty$.

Proof of Lemma 4.2. Before we start estimating the three terms of (4.18), we observe

$$f_{n,t}(-\theta, k) \stackrel{(1.4),(1.11)}{=} \prod_{i=0}^{k-1} \left(1 - \frac{\theta}{1-\theta} \frac{\frac{i}{n}}{1-\frac{i}{n}} \right), \quad k \in [n]. \quad (4.19)$$

Note that $(-\theta)_n^* > e^{-t} - 1$ for large enough n , since $\theta < 1 - e^{-t}$ by (1.10) and (1.11), so we can apply (4.15) in (4.20) and (4.24) below. Now we bound the three terms of (4.18).

First term:

$$\mathbb{E}[X^*; |\mathcal{C}| \in I_n] \stackrel{(4.8),(4.15)}{=} \mathbb{E}\left[X e^{\mathcal{O}(n^{-3/4})}; |\mathcal{C}| \in I_n\right] \stackrel{(4.8),(4.19)}{=} \mathbb{P}(|\mathcal{C}| \in I_n) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right). \quad (4.20)$$

Second term ($\mathbb{E}[X^*; |\mathcal{C}| \in J_n]$):

$$e^{-(\theta)_n^* t} \left(1 + \frac{(-\theta)_n^*}{1-\frac{i}{n}} \right) \stackrel{(1.11),(4.14),(4.17)}{\leq} \left(\frac{1+\tilde{x}}{2} \right)^{-1} e^{-\tilde{\lambda}_n^* t} \left(1 + \frac{\tilde{\lambda}_n^*}{1-\frac{i}{n}} \right), \quad 1 \leq i \leq n^{3/4}, \quad (4.21)$$

$$\begin{aligned} \mathbb{E}[X^*; |\mathcal{C}| \in J_n] &\stackrel{(1.4),(4.8),(4.21)}{\leq} \mathbb{E}\left[\left(\frac{1+\tilde{x}}{2}\right)^{-|\mathcal{C}|} f_{n,t}(\tilde{\lambda}_n^*, |\mathcal{C}|); |\mathcal{C}| \in J_n\right] \\ &\stackrel{(4.8)}{\leq} \left(\frac{1+\tilde{x}}{2}\right)^{-n^{1/4}} \mathbb{E}\left[f_{n,t}(\tilde{\lambda}_n^*, |\mathcal{C}|)\right] \stackrel{(1.5)}{=} \left(\frac{1+\tilde{x}}{2}\right)^{-n^{1/4}} (1+\tilde{\lambda}_n^*) \stackrel{(4.14),(4.17)}{\leq} e^{-\Omega(n^{1/4})}. \end{aligned} \quad (4.22)$$

Third term ($\mathbb{E}[X^*; n^{3/4} \leq |\mathcal{C}|]$):

$$e^{-(\theta)_n^* t} \left(1 + \frac{(-\theta)_n^*}{1-\frac{i}{n}} \right) \leq e^{\theta t} (1-\theta) \stackrel{(1.11)}{=} 1 \quad \text{for any } i \geq \lceil n^{3/4} \rceil, \quad (4.23)$$

$$\begin{aligned} \mathbb{E}[X^*; n^{3/4} \leq |\mathcal{C}|] &\stackrel{(1.4),(4.23)}{\leq} f_{n,t}((-\theta)_n^*, \lceil n^{3/4} \rceil) \stackrel{(4.15)}{=} f_{n,t}(-\theta, \lceil n^{3/4} \rceil) e^{\mathcal{O}(n^{-1/4})} \\ &\stackrel{(4.19)}{\leq} \exp\left(-\frac{\theta}{1-\theta} \sum_{i=0}^{\lceil n^{3/4} \rceil - 1} \frac{i}{n}\right) e^{\mathcal{O}(n^{-1/4})} \leq e^{-\Omega(\sqrt{n})}. \end{aligned} \quad (4.24)$$

The statement of Lemma 4.2 follows from (4.18), (4.20), (4.22) and (4.24). □

4.2 Proof of Lemma 4.3

Before we outline the strategy of the proof of Lemma 4.3 in the paragraph below (4.26), we need to introduce some notation. If we define

$$\alpha_n^{**} := \frac{\lfloor \sqrt{n}\alpha \rfloor}{\sqrt{n}} \quad \text{then} \quad \left(\frac{\alpha}{\sqrt{n}}\right)_n^* \stackrel{(4.14)}{=} \frac{\alpha_n^{**}}{\sqrt{n}} \quad \text{and} \quad |\alpha_n^{**} - \alpha| \leq \frac{1}{\sqrt{n}}. \quad (4.25)$$

Let us abbreviate

$$Y^* = f_{n,t} \left(\frac{\alpha_n^{**}}{\sqrt{n}}, |\mathcal{C}| \right).$$

Recall the definitions of the five intervals from (4.8) and (4.9). We have

$$\begin{aligned} & \left(1 + \frac{\alpha_n^{**}}{\sqrt{n}}\right) (1 - \mathbb{P}_{n+\lfloor \sqrt{n}\alpha \rfloor, p} [|\mathcal{C}| > n]) \stackrel{(1.5)}{=} \mathbb{E}[Y^*; |\mathcal{C}| \in I_n] + \mathbb{E}[Y^*; |\mathcal{C}| \in J_n] \\ & \quad + \mathbb{E}[Y^*; |\mathcal{C}| \in K_n] + \mathbb{E}[Y^*; |\mathcal{C}| \in \tilde{I}_n] + \mathbb{E}[Y^*; |\mathcal{C}| \in \tilde{K}_n]. \end{aligned} \quad (4.26)$$

We will estimate the five terms on the r.h.s. of (4.26). We will show that the terms corresponding to I_n and \tilde{I}_n in (4.26) approximate the terms corresponding to I_n and \tilde{I}_n in (4.11) as $n \rightarrow \infty$, while the terms corresponding to J_n, K_n and \tilde{K}_n in (4.26) vanish as $n \rightarrow \infty$.

Proof of Lemma 4.3. Before we start estimating the five terms of (4.26), we note that if $k \in I_n \cup J_n \cup K_n \cup \tilde{I}_n$ then we can use Taylor expansion of $\ln(1+x)$ to obtain for any $\alpha \in \mathbb{R}$ the formula

$$\begin{aligned} f_{n,t} \left(\frac{\alpha}{\sqrt{n}}, k \right) & \stackrel{(1.4)}{=} \exp \left(-\frac{\alpha}{\sqrt{n}} kt + \sum_{i=0}^{k-1} \ln \left(1 + \frac{\alpha}{1 - \frac{i}{n}} \right) \right) \\ & = \exp \left(-\frac{\alpha}{\sqrt{n}} kt + \sum_{i=0}^{k-1} \frac{\alpha}{\sqrt{n}} \frac{1}{1 - \frac{i}{n}} - \frac{1}{2} \sum_{i=0}^{k-1} \frac{\alpha^2}{n} \frac{1}{(1 - \frac{i}{n})^2} + \mathcal{O} \left(\frac{1}{\sqrt{n}} \right) \right) \\ & = \exp \left(\frac{\alpha}{\sqrt{n}} \left(-kt + n \int_0^{\frac{k}{n}} \frac{1}{1-x} dx \right) - \frac{\alpha^2}{2} \int_0^{\frac{k}{n}} \frac{1}{(1-x)^2} dx + \mathcal{O} \left(\frac{1}{\sqrt{n}} \right) \right) \\ & \stackrel{(1.10)}{=} \exp \left(\alpha \sqrt{n} \varphi \left(\frac{k}{n} \right) - \frac{\alpha^2}{2} \frac{\frac{k}{n}}{1 - \frac{k}{n}} + \mathcal{O} \left(\frac{1}{\sqrt{n}} \right) \right). \end{aligned} \quad (4.27)$$

Now we can estimate the five terms on the r.h.s. of (4.26).

First term:

$$\mathbb{E}[Y^*; |\mathcal{C}| \in I_n] \stackrel{(1.10), (4.8), (4.27)}{=} \mathbb{P}(|\mathcal{C}| \in I_n) + \mathcal{O} \left(n^{-1/4} \right). \quad (4.28)$$

Second term: The bound

$$\mathbb{E}[Y^*; |\mathcal{C}| \in J_n] \leq e^{-\Omega(n^{1/4})} \quad (4.29)$$

can be deduced analogously to (4.22) using that for large enough n we have

$$e^{-\frac{\alpha_n^{**}}{\sqrt{n}} t} \left(1 + \frac{\frac{\alpha_n^{**}}{\sqrt{n}}}{1 - \frac{i}{n}} \right) \stackrel{(4.14), (4.17)}{\leq} \left(\frac{1 + \tilde{x}}{2} \right)^{-1} e^{-\tilde{\lambda}_n^* t} \left(1 + \frac{\tilde{\lambda}_n^*}{1 - \frac{i}{n}} \right), \quad 1 \leq i \leq n^{3/4}. \quad (4.30)$$

Third term ($\mathbb{E}[Y^*; |\mathcal{C}| \in K_n]$): We note

$$\frac{f_{n,t} \left(\frac{\alpha_n^{**}}{\sqrt{n}}, k \right)}{f_{n,t} \left(\frac{(\alpha-1)n^{**}}{\sqrt{n}}, k \right)} \stackrel{(4.27)}{=} \exp \left(\sqrt{n} \varphi \left(\frac{k}{n} \right) + \mathcal{O}(1) \right) \stackrel{(1.10), (1.11), (4.8)}{\leq} e^{-\Omega(n^{1/8})}, \quad k \in K_n, \quad (4.31)$$

$$\mathbb{E}[Y^*; |\mathcal{C}| \in K_n] \stackrel{(4.31)}{\leq} e^{-\Omega(n^{1/8})} \mathbb{E} \left[f_{n,t} \left(\frac{(\alpha-1)_n^{**}}{\sqrt{n}}, |\mathcal{C}| \right); |\mathcal{C}| \in K_n \right] \stackrel{(1.5)}{\leq} 2e^{-\Omega(n^{1/8})}. \quad (4.32)$$

Fourth term ($\mathbb{E}[Y^*; |\mathcal{C}| \in \tilde{I}_n]$): If $x \in [-n^{1/8}, n^{1/8}]$, i.e., if $k = \lfloor \theta n + x\sqrt{n} \rfloor \in \tilde{I}_n$ then

$$f_{n,t} \left(\frac{\alpha}{\sqrt{n}}, k \right) \stackrel{(1.11),(4.27)}{=} \exp \left(\alpha \varphi'(\theta)x - \frac{\alpha^2}{2} \frac{\theta}{1-\theta} + \mathcal{O}(n^{-1/4}) \right), \quad (4.33)$$

$$f_{n,t} \left(\frac{\alpha_n^{**}}{\sqrt{n}}, k \right) \stackrel{(4.25),(4.33)}{=} \exp \left(\alpha \varphi'(\theta)x - \frac{\alpha^2}{2} \frac{\theta}{1-\theta} \right) + \mathcal{O} \left(n^{-1/4} f_{n,t} \left(\frac{\alpha_n^{**}}{\sqrt{n}}, k \right) \right), \quad (4.34)$$

$$\mathbb{E}[Y^*; |\mathcal{C}| \in \tilde{I}_n] \stackrel{(1.5),(4.9),(4.34)}{=} \mathbb{E} \left[\exp \left(\alpha \varphi'(\theta) \frac{|\mathcal{C}| - \theta n}{\sqrt{n}} - \frac{\alpha^2}{2} \frac{\theta}{1-\theta} \right); |\mathcal{C}| \in \tilde{I}_n \right] + \mathcal{O}(n^{-1/4}). \quad (4.35)$$

Fifth term ($\mathbb{E}[Y^*; |\mathcal{C}| \in \tilde{K}_n]$): We observe that

$$\frac{f_{n,t} \left(\frac{\alpha_n^{**}}{\sqrt{n}}, \lfloor \theta n + n^{1/8} \sqrt{n} \rfloor \right)}{f_{n,t} \left(\frac{(\alpha+1)_n^{**}}{\sqrt{n}}, \lfloor \theta n + n^{1/8} \sqrt{n} \rfloor \right)} \stackrel{(4.33)}{=} \exp \left(-\varphi'(\theta)n^{1/8} + \mathcal{O}(1) \right) \stackrel{(1.12)}{\leq} e^{-\Omega(n^{1/8})}, \quad (4.36)$$

$$\exp \left(-\frac{\alpha_n^{**}}{\sqrt{n}} t \right) \left(1 + \frac{\frac{\alpha_n^{**}}{\sqrt{n}}}{1 - \frac{i}{n}} \right) \stackrel{(1.12)}{\leq} \exp \left(-\frac{(\alpha+1)_n^{**}}{\sqrt{n}} t \right) \left(1 + \frac{\frac{(\alpha+1)_n^{**}}{\sqrt{n}}}{1 - \frac{i}{n}} \right), \quad \lfloor \theta n + n^{1/8} \sqrt{n} \rfloor \leq i, \quad (4.37)$$

$$f_{n,t} \left(\frac{\alpha_n^{**}}{\sqrt{n}}, k \right) \stackrel{(1.4),(4.36),(4.37)}{\leq} e^{-\Omega(n^{1/8})} f_{n,t} \left(\frac{(\alpha+1)_n^{**}}{\sqrt{n}}, k \right), \quad \lfloor \theta n + n^{1/8} \sqrt{n} \rfloor \leq k, \quad (4.38)$$

$$\mathbb{E}[Y^*; |\mathcal{C}| \in \tilde{K}_n] \stackrel{(4.9),(4.38)}{\leq} e^{-\Omega(n^{1/8})} \mathbb{E} \left[f_{n,t} \left(\frac{(\alpha+1)_n^{**}}{\sqrt{n}}, |\mathcal{C}| \right); |\mathcal{C}| \in \tilde{K}_n \right] \stackrel{(1.5)}{\leq} 2e^{-\Omega(n^{1/8})}. \quad (4.39)$$

Finally, the proof of the fact that the error term $\mathbb{P}_{n+\lfloor \sqrt{n}\alpha \rfloor, p} [|\mathcal{C}| > n]$ that appears on the l.h.s. of (4.26) goes to zero as $n \rightarrow \infty$ is analogous to the $\alpha = 0$ case of (4.39). The statement of Lemma 4.3 follows from (4.26), (4.28), (4.29), (4.32), (4.35) and (4.39). \square

References

- [1] D. J. Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *The Annals of Probability* 812–854, 1997. MR-1434128
- [2] D. J. Aldous. Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists. *Bernoulli* 5.1: 3–48, 1999. MR-1673235
- [3] N. Alon, J. H. Spencer. *The probabilistic method (Second edition)*. Wiley, 2000. MR-1885388
- [4] N. Alon, J. H. Spencer. *The probabilistic method (Fourth edition)*. Wiley, 2016. MR-3524748
- [5] D. Barraez, S. Boucheron, W. F. De La Vega. On the fluctuations of the giant component. *Combinatorics, Probability and Computing* 9.4, 28–304, 2000. MR-1786919
- [6] B. Bollobás. *Random graphs*. No. 73. Cambridge university press, 2001. MR-1864966
- [7] B. Bollobás, S. Janson, O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms* 31.1: 3–122, 2007. MR-2337396
- [8] P. Erdős, and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* 5.1: 17–60, 1960. MR-0125031
- [9] L. Federico, R. van der Hofstad, F. den Hollander, T. Hulshof. Expansion of percolation critical points for Hamming graphs. arXiv:1701.02099, 2017.
- [10] R. van der Hofstad., W. Kager, T. Müller. A local limit theorem for the critical random graph. *Electr. Comm. in Probab.* 14: 122–131, 2009. MR-2481672

- [11] R. van der Hofstad. Random graphs and complex networks. *Cambridge Series in Statistical and probabilistic Mathematics* 43, 2016.
- [12] S. Janson, M. J. Luczak. Susceptibility in subcritical random graphs. *Journal of Mathematical Physics* 49.12, 125207, 2008. MR-2484338
- [13] S. Janson, T. Luczak, A. Rucinski. Random graphs. Vol. 45. John Wiley and Sons, 2011. MR-1782847
- [14] S. Lemaire. A multiplicative coalescent with asynchronous multiple mergers. arXiv:1709.01896, 2017.
- [15] A. Martin-Löf. The final size of a nearly critical epidemic, and the first passage time of a Wiener process to a parabolic barrier. *Journal of Applied probability* 35.3: 671–682, 1998. MR-1659544
- [16] J. B. Martin, B. Ráth. Rigid representations of the multiplicative coalescent with linear deletion. *Electron. J. Probab.* Vol. 22, paper no. 83, 2017. MR-3718711
- [17] B. Pittel. On tree census and the giant component in sparse random graphs. *Random Structures and Algorithms* 1.3: 311–342, 1990. MR-1099795

Acknowledgments. The author thanks James Martin, Dominic Yeo and Bálint Tóth for valuable discussions and an anonymous referee for thoroughly reading the manuscript and for suggesting to use [10, Lemma 2.1] in the proof of Theorem 1.6. This work is partially supported by OTKA (Hungarian National Research Fund) grant K109684, the Postdoctoral Fellowship of NKFI (National Research, Development and Innovation Office) and the Bolyai Research Scholarship of the Hungarian Academy of Sciences.