CAMBRIDGE
UNIVERSITY PRESS

Unusually large components in near-critical Erdős–Rényi graphs via ballot theorems

Umberto De Ambroggio* and Matthew I. Roberts

Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK
*Corresponding author. Email: umbidea@gmail.com

Abstract

We consider the near-critical Erdős–Rényi random graph $G(n, p)$ and provide a new probabilistic proof of the fact that, when p is of the form $p = p(n) = 1/n + \lambda/n^{4/3}$ and A is large,

$$\mathbb{P}(|\mathcal{C}_{\max}| > An^{2/3}) \asymp A^{-3/2} e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}},$$

where $\mathcal{C}_{\max}$ is the largest connected component of the graph. Our result allows A and λ to depend on n. While this result is already known, our proof relies only on conceptual and adaptable tools such as ballot theorems, whereas the existing proof relies on a combinatorial formula specific to Erdős–Rényi graphs, together with analytic estimates.

1. Introduction

The Erdős–Rényi random graph, denoted by $G(n, p)$, is obtained from the complete graph with vertex set $[n]$ by independently retaining each edge with probability $p \in [0, 1]$ and deleting it with probability $1 - p$. We are interested in the size of the largest connected component $\mathcal{C}_{\max}$, or a typical connected component $\mathcal{C}(v)$ for $v \in [n]$. It is well known (see e.g [7, 32] or [15] for more details) that, if $p = p(n) = \gamma/n$ for constant γ, then $G(n, p)$ undergoes a phase transition as γ passes 1:

i. if $\gamma < 1$ (the *subcritical* case), then $|\mathcal{C}_{\max}|$ is of order $\log n$;
ii. if $\gamma = 1$ (the *critical* case), then $|\mathcal{C}_{\max}|$ is of order $n^{2/3}$;
iii. if $\gamma > 1$ (the *supercritical* case), then $|\mathcal{C}_{\max}|$ is of order n.

Motivated by the lack of a simple proof of (ii), Nachmias and Peres [24] used a martingale argument to prove that for any $n > 1000$ and $A > 8$,

$$\mathbb{P}(|\mathcal{C}(v)| > An^{2/3}) \leq 4n^{-1/3} \exp\{-A^2(A - 4)/32\}$$

and

$$\mathbb{P}(|\mathcal{C}_{\max}| > An^{2/3}) \leq \frac{4}{A} \exp\{-A^2(A - 4)/32\}.$$

CrossMark

They also gave bounds when $p = \frac{1+\lambda n^{-1/3}}{n}$ for fixed $\lambda \in \mathbb{R}$. The best known bound on the latter quantity is due originally to Pittel [26] who showed that for p of this form,

$$\lim_{n \to \infty} A^{3/2} e^{\frac{A^3}{8} - \frac{\lambda A^2}{2} + \frac{\lambda^2 A}{2}} \mathbb{P}(|\mathcal{C}_{\max}| > An^{2/3})$$

converges as $A \to \infty$ to a specific constant, which is stated to be $(2\pi)^{-1/2}$ but should be $(8/9\pi)^{1/2}$ due to a small oversight in the proof. More details, and a stronger result that allows A and λ to depend on n, are available in [28]. Both [26] and [28] rely on a combinatorial formula for the expected number of components with exactly k vertices and $k + \ell$ edges, which is specific to Erdős–Rényi graphs and appears difficult to adapt to other models, together with analytic approximations.

We provide a new proof of asymptotics for $\mathbb{P}(|\mathcal{C}(v)| > An^{2/3})$ and $\mathbb{P}(|\mathcal{C}_{\max}| > An^{2/3})$ that combines the strengths of the results mentioned above:

- it gives accurate bounds for large A as $n \to \infty$;
- it allows A and λ to depend on n;
- it uses only robust probabilistic tools and therefore has the potential to be adapted to other models of random graphs.

This is the purpose of our main theorem, which we now state.

Theorem 1.1 *There exists $A_0 > 0$ such that if $A = A(n)$ satisfies $A_0 \leq A = o(n^{1/30})$ and $p = p(n) = 1/n + \lambda/n^{4/3}$ with $\lambda = \lambda(n)$ such that $|\lambda| \leq A/3$, then for sufficiently large n and any vertex $v \in [n]$, we have*

$$a. \frac{c_1}{A^{1/2}n^{1/3}} e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}} \leq \mathbb{P}(|\mathcal{C}(v)| > An^{2/3}) \leq \frac{c_2}{A^{1/2}n^{1/3}} e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}}$$

and

$$b. \frac{c_1}{A^{3/2}} e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}} \leq \mathbb{P}(|\mathcal{C}_{\max}| > An^{2/3}) \leq \frac{c_2}{A^{3/2}} e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}}$$

for some constants $0 < c_1 \leq c_2 < \infty$.

Although our methods are not accurate enough to give the correct constant factor in the asymptotic, identified in [28], we believe that the substantially more robust approach is worth the small sacrifice in precision. Indeed, the probabilistic arguments in [24] have been adapted to critical random d-regular graphs by Nachmias and Peres [23], the configuration model with bounded degrees by Riordan [27], and more recently a particular model of inhomogeneous random graphs by the first author and Pachon [11].

At the end of Sections 3 and 4 (see Remarks 3.7 and 4.15) we will comment further on the problem of finding the right constant in the asymptotic using methods similar to the ones in this paper. We believe that it may be possible but would require significantly more technical work. We also remark that Nachmias and Peres [24] gave specific values of n and A for which their bounds hold, namely $n > 1000$ and $A > 8$. We could do the same, but it would again require significantly more technical details given the more precise nature of our bounds.

We remark that our proofs of the upper bounds in Theorem 1.1 are particularly straightforward, perhaps even more so than those in [24], despite giving a much more accurate bound. A key part of the argument will be the following simple *ballot-type* result, which may be of independent interest.

Lemma 1.2 *Fix $n \in \mathbb{N}$. Let $X_1, \ldots, X_n$ be $\mathbb{Z}-$valued random variables and suppose that the law of $(X_1, \ldots, X_n)$ is invariant under rotations (it may depend on n). Define $S_0 = 0$ and $S_t = \sum_{i=1}^{t} X_i$, for $t \in [n]$. Then for any $j \in \mathbb{N}$,*

$$\mathbb{P}(S_t > 0 \ \forall t \in [n], \ S_n = j) \leq \frac{j}{n}\mathbb{P}(S_n = j).$$

Our proofs of the lower bounds in Theorem 1.1 will be more complicated than those for the upper bounds, although they still use only robust probabilistic techniques such as a generalised ballot theorem, Poisson approximations for the binomial distribution and Brownian approximations to random walks. In future work we intend to demonstrate the adaptability of our new approach by applying our methods to other random graph models. As the first step in this direction, for applications of Lemma 1.2 to a random intersection graph, an inhomogeneous random graph and percolation on a d-regular graph, see [10].

Structure of the paper. We start by introducing ballot-type results in Section 2, where we prove Lemma 1.2 and a corollary which will be the main tool to obtain the upper bounds in Theorem 1.1. We also state a generalised ballot theorem due to Addario-Berry and Reed [2] that will be used for our lower bounds. Subsequently, in Section 3 we prove the upper bounds in (*a*) and (*b*) of Theorem 1.1, whereas the corresponding lower bounds will be proved in Section 4.

Notation. We write $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $[n] = \{1, 2, \ldots, n\}$, and $[\![a, b]\!] = [a, b] \cap \mathbb{Z}$. The abbreviation i.i.d. means 'independent and identically distributed'. The empty sum is defined to be 0, and the empty product is defined to be 1. In particular we use the convention that $\sum_{i=n+1}^{n} a_i$ is zero, for any n and any sequence (a_i). For brevity we simply write A rather than $A(n)$, λ instead of $\lambda(n)$, and p in place of $p(n) = 1/n + \lambda n^{-4/3}$. We will often write c to mean a constant in $(0, \infty)$ and use c many times in a single proof even though the constant may change from line to line.

1.1. Related work

Besides his Proposition 2, which gives asymptotics for $\mathbb{P}(|\mathcal{C}_{\max}| > An^{2/3})$ in the case of constant (large) A and λ, Pittel [26] includes several other results which we make no attempt to rework. These include asymptotics for $\mathbb{P}(|\mathcal{C}_{\max}| < an^{2/3})$ when a is small. Nachmias and Peres [24] also gave a simple but inaccurate upper bound on this quantity, and it would be interesting to give an intuitive probabilistic proof of more accurate asymptotics. Pittel's paper is partially based on an earlier article by Luczak, Pittel and Wierman [20].

For $G(n, p)$ outside the critical scaling window, i.e. when λ is not bounded in n, $n^{2/3}$ is not the most likely size for the largest component of the graph, and therefore our results—while still true, at least provided $|\lambda| \leq A/3 = o(n^{1/30})$—appear less natural than those by Nachmias and Peres [22], Bollobás and Riordan [8] or Riordan [27].

A local limit theorem for the size of the k largest components (for arbitrary k) was given by Van der Hofstad et al. [34]. See also Van der Hofstad et al. [36], where similar results to those established by Pittel [26] are proved in the context of inhomogeneous random graphs.

Aldous [3] used a *breadth-first* search algorithm to explore $G(n, p)$ for p within the critical window and showed that the sizes of the largest components, if rescaled by $n^{2/3}$, converge (in an appropriate sense) to some limit, which he described in detail. The same type of argument has been used by Van der Hofstad [33] to investigate critical SIR epidemics. The work of Aldous was then developed by Addario-Berry, Broutin and Goldschmidt [1] who showed that the rescaled components themselves converge to metric spaces characterised by excursions of Brownian motion with parabolic drift, decorated by a Poisson point process.

There are several other models that share similar properties with the near-critical Erdős–Rényi graph. For instance, there are many critical models whose component sizes, when suitably rescaled, converge to the lengths of excursions of Brownian motion with parabolic drift just as for

the Erdős–Rényi graph. Some examples include inhomogeneous random graphs (see e.g. [5] and [6]), the configuration model (see [13], [16] and [27]), and quantum random graphs (see [12]).

In another direction we mention [25], where a large deviations rate function is provided for the size of the maximal component divided by n, valid for the $G(n, \gamma/n)$ model with $\gamma > 0$. For a very recent work in this direction, see [4].

Finally, the results of [28] were used to show the existence of times when a dynamical version of the Erdős–Rényi graph has an unusually large connected component. Related results about the structure of dynamical Erdős–Rényi graphs were given by Rossignol [29].

2. Ballot-style results

Let $X_1, \ldots, X_n \in \{-1, 1\}$ be i.i.d. random variables taking values in $\{-1, 1\}$, with $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = 1/2$, and let $S_t = \sum_{i=1}^{t} X_i$. In its simplest form the ballot theorem concerns the probability that S_t stays positive for all times $t \in [n]$, given that $S_n = k \in \mathbb{N}$, and says that the answer is k/n; see e.g. [2, 17, 19, 35] and references therein. However, we will be interested in evaluating probabilities of the following type:

$$\mathbb{P}\left(1 + S_t > 0 \; \forall t \in [n], \; 1 + S_n = k\right),$$

where $k \geq 1$ and $X_1, \ldots, X_n$ are i.i.d. random variables taking values in $\{-1, 0, 1, 2, \ldots\}$. A possible solution might be to apply the following generalised ballot theorem.

Theorem 2.1 (Addario-Berry and Reed [2]) *Suppose X is a random variable satisfying $\mathbb{E}[X] = 0$, $Var(X) > 0$, $\mathbb{E}[X^{2+\alpha}] < \infty$ for some $\alpha > 0$, and X is a lattice random variable with period d (meaning that dX is an integer random variable and d is the smallest positive real number for which this holds). Then given independent random variables $X_1, X_2, \ldots$ distributed as X with associated partial sums $S_t = \sum_{i=1}^{t} X_i$, for all j such that $0 \leq j = O\left(\sqrt{n}\right)$ and such that j is a multiple of $1/d$ we have*

$$\mathbb{P}\left(S_t > 0 \; \forall t \in [n], \; S_n = j\right) = \Theta\left(\frac{j+1}{n^{3/2}}\right).$$

This result will indeed be useful in the proof of the lower bounds in our Theorem 1.1. However, for the upper bound we will need a result that holds when j is much larger than $\sqrt{n}$. Our Lemma 1.2 shows that the upper bound remains true more generally. We now aim to prove that result.

Fix $n \in \mathbb{N}$. Let $X = (X_1, \ldots, X_n)$ be random variables taking values in $\mathbb{Z}$. Define $S_0 = 0$ and $S_t = \sum_{i=1}^{t} X_i$ for all $t \in [n]$. Given $r \in [n]$, define the *rotation* of $S = (S_0, S_1, \ldots, S_n)$ by r as the walk $S^r = (S_0^r, S_1^r, \ldots, S_n^r)$ corresponding to the rotated sequence $X^r = (X_{r+1}, \ldots, X_n, X_1, \ldots, X_r)$. That is,

- if $0 \leq t \leq n - r$, then $S_t^r = S_{t+r} - S_r = \sum_{i=r+1}^{t+r} X_i$;
- if $n - r < t \leq n$, then $S_t^r = S_n + S_{t+r-n} - S_r = \sum_{i=r+1}^{n} X_i + \sum_{i=1}^{t+r-n} X_i$.

In particular, $S_n^r = \sum_{i=r+1}^{n} X_i + \sum_{i=1}^{r} X_i = S_n$ (for every $r \in [n]$) and $S^n = S$.

Definition 2.2 We say that $r \in [n]$ is *favourable* if $S_t^r > 0$ for every $t \in [n]$.

The following lemma contains the key observation needed to prove Lemma 1.2.

Lemma 2.3 *Fix $j \in \mathbb{N}$. If $S_n = j$, then*

$$|\{r \in [n] : r \text{ is favourable}\}| \leq j.$$

Proof. The idea is simply that between each two favourable vertices, the partial sum must increase by at least one. To see this, let $1 \leq I_1 < \cdots < I_L \leq n$ denote the indices (if any) such that I_k is favourable for $1 \leq k \leq L$. We need to show that $L \leq j$. Observe that for each $1 \leq k \leq L - 1$, since I_k is favourable and $X_1, \ldots, X_n$ are integer-valued, we have $S_t^{I_k} \geq 1$ for all $t \in [n]$; in particular $S_{I_{k+1} - I_k}^{I_k} \geq 1$. For the same reason, $S_{(I_1 + n) - I_L}^{I_L} \geq 1$. Consequently

$$S_n = \sum_{k=1}^{L-1} S_{I_{k+1} - I_k}^{I_k} + S_{(I_1+n)-I_L}^{I_L} \geq \sum_{k=1}^{L-1} 1 + 1 = L.$$

But our assumption is that $S_n = j$, so we must have $L \leq j$ as required. $\qquad\square$

Proof of Lemma 1.2. For any $r \in [n]$, since $(X_1, X_2, \ldots, X_n)$ is invariant under rotations,

$$\mathbb{P}(S_t > 0 \; \forall t \in [n], \; S_n = j) = \mathbb{P}(S_t^r > 0 \; \forall t \in [n], \; S_n^r = j) = \mathbb{P}(r \text{ is favourable}, \; S_n^r = j)$$

and since $S_n^r = S_n$, we obtain that

$$\mathbb{P}(S_t > 0 \; \forall t \in [n], \; S_n = j) = \mathbb{P}(r \text{ is favourable}, \; S_n = j).$$

Summing over $r \in [n]$ and applying Lemma 2.3 we have

$$n\mathbb{P}(S_t > 0 \; \forall t \in [n], \; S_n = j) = \sum_{r=1}^n \mathbb{E}[\mathbb{1}_{\{r \text{ is favourable}\}} \mathbb{1}_{\{S_n = j\}}]$$

$$= \mathbb{E}\left[\mathbb{1}_{\{S_n = j\}} \sum_{r=1}^n \mathbb{1}_{\{r \text{ is favourable}\}} \right]$$

$$\leq \mathbb{E}\left[\mathbb{1}_{\{S_n = j\}} j \right] = j\mathbb{P}(S_n = j)$$

which completes the proof. $\qquad\square$

The following corollary will be used to prove the upper bounds of Theorem 1.1.

Corollary 2.4 *Fix $n \in \mathbb{N}$ and let $(X_i)_{i \geq 1}$ be i.i.d. random variables taking values in $\mathbb{Z}$, whose distribution may depend on n. Let $h \in \mathbb{N}$, and suppose that $\mathbb{P}(X_1 = h) > 0$. Define $S_t = \sum_{i=1}^t X_i$ for $t \in \mathbb{N}_0$. Then for any $j \geq 1$ we have*

$$\mathbb{P}(h + S_t > 0 \; \forall t \in [n], \; h + S_n = j) \leq \mathbb{P}(X_1 = h)^{-1} \frac{j}{n+1} \mathbb{P}(S_{n+1} = j).$$

Proof. Let X_0 be an independent copy of X_1. Define $S_t^* = X_0 + S_t$ for $0 \leq t \leq n$. Then

$$\mathbb{P}(h + S_t > 0 \; \forall t \in [n], \; h + S_n = j)$$

$$= \mathbb{P}(X_0 = h)^{-1} \mathbb{P}(h + S_t > 0 \; \forall t \in [n], \; h + S_n = j, \; X_0 = h)$$

$$= \mathbb{P}(X_1 = h)^{-1} \mathbb{P}(S_t^* > 0 \; \forall t \in [n], \; S_n^* = j, \; S_0^* = h)$$

$$\leq \mathbb{P}(X_1 = h)^{-1} \mathbb{P}(S_t^* > 0 \; \forall t \in \{0\} \cup [n], \; S_n^* = j). \tag{1}$$

Now since $(S_0^*, S_1^*, \ldots, S_n^*) \stackrel{d}{=} (S_1, S_2, \ldots, S_{n+1})$, applying Lemma 1.2 we obtain that

$$\mathbb{P}(S_t^* > 0 \; \forall t \in \{0\} \cup [n], \; S_n^* = j) = \mathbb{P}(S_t > 0 \; \forall t \in [n+1], \; S_{n+1} = j)$$

$$\leq \frac{j}{n+1} \mathbb{P}(S_{n+1} = j),$$

and substituting this into (1) gives the result. $\qquad\square$

3. Proof of the upper bounds in Theorem 1.1

A main ingredient in our analysis is an *exploration process*, which is a procedure to sequentially discover the component containing a given vertex, and which reduces the study of component sizes to the analysis of the trajectory of a stochastic process. Such exploration processes are well known, dating back at least to [21], and several variants exist. Our description closely follows the one appearing in [28]; see also [24].

Let G be any (undirected) graph with vertex set $[n]$, and let $v \in [n]$ be any given vertex. Fix an ordering of the n vertices with v first. At each time $t \in \{0\} \cup [n]$ of the exploration, each vertex will be *active*, *explored* or *unseen*; the number of explored vertices will be t whereas the (possibly random) number of active vertices will be denoted by Y_t. At time $t = 0$, vertex v is declared to be active whereas all other vertices are declared unseen, so that $Y_0 = 1$. At each step $t \in [n]$ of the procedure, if $Y_{t-1} > 0$ then we let u_t be the first active vertex; if $Y_{t-1} = 0$, we let u_t be the first unseen vertex (here the term *first* refers to the ordering that we fixed at the beginning of the procedure). Note that at time $t = 1$ we have $u_1 = v$. Denote by η_t the number of unseen neighbours of u_t in G and change the status of these vertices to active. Then, set u_t itself explored. From this description we see that:

- $Y_t = Y_{t-1} + \eta_t - 1$, if $Y_{t-1} > 0$;
- $Y_t = \eta_t$, if $Y_{t-1} = 0$.

We now specialise to the Erdős–Rényi random graph, i.e. we now take $G = G(n, p)$. Let us denote by $U_t = n - Y_t - t$ the number of unseen vertices in $G(n, p)$ at time t, and define $\mathcal{F}_0 = \{\Omega, \emptyset\}$ and $\mathcal{F}_t = \sigma(\{\eta_j : 1 \leq j \leq t\})$ for $t \in [n]$. Then for $t \in [n]$, given $\mathcal{F}_{t-1}$, we see that $\eta_t \sim \mathrm{Bin}(U_{t-1}, p)$. Since $U_t \leq n - t$, we can couple the process $(\eta_i)_{i \in [n]}$ with a sequence $(\tau_i)_{i \in [n]}$ of independent $\mathrm{Bin}(n - i, p)$ random variables such that $\tau_i \geq \eta_i$ for all i. It follows that, for any $k \in [n]$,

$$
\mathbb{P}(|\mathcal{C}(v)| > k) = \mathbb{P}(Y_t > 0 \,\forall t \in [k])
$$

$$
= \mathbb{P}\left(1 + \sum_{i=1}^{t}(\eta_i - 1) > 0 \,\forall t \in [k]\right)
$$

$$
\leq \mathbb{P}\left(1 + \sum_{i=1}^{t}(\tau_i - 1) > 0 \,\forall t \in [k]\right). \tag{2}
$$

We would like to apply Corollary 2.4 to the sequence $(1 + \sum_{i=1}^{t}(\tau_i - 1))_{t \in [k]}$, and to this end we need to turn the latter process into a random walk with identically distributed increments. This is achieved in Lemma 3.1 below.

Lemma 3.1 *There exists a finite constant c such that for any $k \in [n]$,*

$$
\mathbb{P}\left(1 + \sum_{i=1}^{t}(\tau_i - 1) > 0 \,\forall t \in [k]\right) \leq c\mathbb{P}\left(1 + R_t > 0 \,\forall t \in [k], \ 1 + R_k \geq \frac{k^2 p}{2} - \frac{k}{n^{1/2}}\right),
$$

where $(R_t)_{t \geq 0}$ is a random walk with $R_0 = 0$ and i.i.d. steps each having distribution $\mathrm{Bin}(n, p) - 1$.

The idea behind this lemma is that by adding an independent $Bin(i, p)$ random variable to τ_i, we transform it into a $Bin(n, p)$ random variable which forms one of the steps of the random walk R_t appearing on the right-hand side. If the sum of the τ_i up to t remains positive then R_t, which is larger, must certainly also remain positive; but also the final value R_k must be larger than the sum of the additional contributions from the $Bin(i, p)$ random variables. A standard bound shows that these additional contributions are concentrated about their mean, which is approximately $k^2 p / 2$.

We postpone the details until Section 3.1 and continue with the proof of the upper bounds in Theorem 1.1. By summing over the possible values of R_k, we can apply Corollary 2.4 with $h = 1$ to the quantity on the right-hand side of Lemma 3.1: it is at most

$$\frac{c}{k+1} \sum_{j=h(k,n)}^{(k+1)(n-1)} j\mathbb{P}\left(R_{k+1}=j\right), \tag{3}$$

where $h(k, n) = \lceil \frac{k^2}{2}p - \frac{k}{n^{1/2}} \rceil$, and the upper limit on the sum is due to the fact that $R_{k+1} \leq (k+1)(n-1)$ (because each step of R_t is at most $n-1$, and in R_{k+1} we are summing $k+1$ of them).

We now rewrite the above sum in a way that is easier to analyse, using the following elementary observation. If X is a random variable taking values in $\mathbb{Z} \cap (-\infty, N]$ for some $N \in \mathbb{N}$, then for any $h \geq 1$, we have

$$\mathbb{E}[X\mathbb{1}_{\{X\geq h\}}] = \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{1}_{\{i\leq X\}}\mathbb{1}_{\{X\geq h\}}\right] = \mathbb{E}\left[\sum_{i=1}^{h} \mathbb{1}_{\{X\geq h\}}\right] + \mathbb{E}\left[\sum_{i=h+1}^{N} \mathbb{1}_{\{X\geq i\}}\right]$$

$$= h\mathbb{P}(X \geq h) + \sum_{i=h+1}^{N} \mathbb{P}(X \geq i).$$

Applying this to R_{k+1}, and using that $h(k, n)/(k + 1) \leq k/n$ when n is large, we have

$$\frac{1}{k+1} \sum_{j=h(k,n)}^{(k+1)(n-1)} j\mathbb{P}\left(R_{k+1}=j\right) \leq \frac{k}{n}\mathbb{P}(R_{k+1} \geq h(k, n)) + \frac{1}{k+1} \sum_{j=h(k,n)+1}^{(k+1)(n-1)} \mathbb{P}(R_{k+1} \geq j),$$

and putting this together with (2), Lemma 3.1 and (3), we have shown that

$$\mathbb{P}\left(|\mathcal{C}(v)| > k\right) \leq \frac{ck}{n}\mathbb{P}(R_{k+1} \geq h(k, n)) + \frac{c}{k+1} \sum_{j=h(k,n)+1}^{(k+1)(n-1)} \mathbb{P}(R_{k+1} \geq j).$$

The next two lemmas conclude the proof of the upper bound in part (a) of Theorem 1.1 by showing that, when we take $k = \lceil An^{2/3} \rceil$ with $A \geq 1$, the right-hand side above is bounded by $cA^{-1/2}n^{-1/3} \exp\{-A^3/8 + \lambda A^2/2 - \lambda^2 A/2\}$. Let

$$H(A, n) = h(\lceil An^{2/3}\rceil, n) = \left\lceil \frac{\lceil An^{2/3}\rceil^2}{2}p - \frac{\lceil An^{2/3}\rceil}{n^{1/2}} \right\rceil.$$

Lemma 3.2 *Suppose that $1 \leq A = o\left(n^{1/12}\right)$, $\lambda = o(n^{1/12})$ and $\lambda \leq A/3$. There exists a finite constant c such that*

$$\frac{\lceil An^{2/3}\rceil}{n}\mathbb{P}\left(R_{\lceil An^{2/3}\rceil+1} \geq H(A, n)\right) \leq \frac{c}{A^{1/2}n^{1/3}}e^{-A^3/8+\lambda A^2/2-\lambda^2 A/2}.$$

Lemma 3.3 *Suppose that $1 \leq A = o\left(n^{1/12}\right)$, $\lambda = o(n^{1/12})$ and $\lambda \leq A/3$. There exists a finite constant c such that*

$$\frac{1}{\lceil An^{2/3}\rceil+1} \sum_{j=H(A,n)+1}^{(\lceil An^{2/3}\rceil+1)(n-1)} \mathbb{P}(R_{\lceil An^{2/3}\rceil+1} \geq j) \leq \frac{c}{A^2 n^{1/3}}e^{-A^3/8+\lambda A^2/2-\lambda^2 A/2}.$$

Since $R_{\lceil An^{2/3}\rceil+1}$ is simply a binomial random variable, the proofs of Lemmas 3.2 and 3.3 are exercises in applying standard estimates to binomial random variables. We carry out the details in Section 3.1. Subject to these and the proof of Lemma 3.1, the proof of the upper bound in part (a) of Theorem 1.1 is complete.

The upper bound of part (b) in Theorem 1.1 is deduced from the upper bound in part (a) using the following standard procedure, used for example in [24]. For any $k \in [n]$, denote by

$$N_k = \sum_{i=1}^{n} \mathbb{1}_{\{|\mathcal{C}(v_i)| > k\}}$$

the number of vertices that are contained in components of size larger than k. If u is any fixed vertex in $G(n, p)$, we have

$$\mathbb{P}(|\mathcal{C}_{\max}| > k) = \mathbb{P}(N_k > k) \leq \frac{1}{k}\mathbb{E}[N_k] = \frac{n}{k}\mathbb{P}(|\mathcal{C}(u)| > k)$$

and then taking $k = \lceil An^{2/3} \rceil$ and applying part (a), this is at most

$$\frac{n}{\lceil An^{2/3} \rceil} \frac{c}{A^{1/2}n^{1/3}} e^{-A^3/8 + \lambda A^2/2 - \lambda^2 A/2} \leq cA^{-3/2} e^{-A^3/8 + \lambda A^2/2 - \lambda^2 A/2},$$

as required. This concludes the proof for the upper bounds (a) and (b) in Theorem 1.1, subject to proving Lemmas 3.1, 3.2 and 3.3.

3.1. Proofs of Lemmas 3.1, 3.2 and 3.3

To prove Lemmas 3.1, 3.2 and 3.3 we will make use of the following two preliminary results on the concentration of Binomial random variables about their mean. The first of these results is Theorem 1.3 in [7], while the second is Theorem 2.1 in [15].

Lemma 3.4 *Let $S \sim Bin(n, p)$, and suppose that $np \geq 1$ and $1 \leq h \leq (1 - p)n/3$. Then*

$$\mathbb{P}(S \geq np + h) \leq \frac{(p(1-p)n)^{1/2}}{h\sqrt{2\pi}} \exp\left\{-\frac{h^2}{2p(1-p)n} + \frac{h}{p(1-p)n} + \frac{h^3}{2(1-p)^3n^3}\right\}.$$

Lemma 3.5 *Let $S \sim Bin(n, p)$ and define $\phi(x) = (1 + x)\log(1 + x) - x$ for $x \geq -1$. Then for every $t \geq 0$ we have that*

  a. $\mathbb{P}(S \geq \mathbb{E}[S] + t) \leq \exp\{-\mathbb{E}[S]\phi(t/\mathbb{E}[S])\} \leq \exp\left\{-t^2/2(\mathbb{E}[S] + t/3)\right\};$
  b. $\mathbb{P}(S \leq \mathbb{E}[S] - t) \leq \exp\{-t^2/2\mathbb{E}[S]\}.$

We are now ready to start with the proofs of the lemmas stated in the previous section.

Proof of Lemma 3.1. We want to bound

$$\mathbb{P}\left(1 + \sum_{i=1}^{t} (\tau_i - 1) > 0 \ \forall t \in [k]\right)$$

from above, where $\tau_i \sim Bin(n - i, p)$ are independent. We do this by adding extra terms to the sum $\sum_{i=1}^{t} (\tau_i - 1)$ to create a random walk with identically distributed steps. To this end, let $(B_i)_{i \in [n]}$ be a sequence of independent random variables, also independent from $(\tau_i)_{i \in [n]}$, and such that $B_i \sim Bin(i, p)$ for every $i \in [n]$. Moreover, define $S_t = \sum_{i=1}^{t} B_i$ for $t \in [n]$. Let

$$P = \mathbb{P}\left(S_k \geq \frac{k^2}{2}p - \frac{k}{n^{1/2}}\right).$$

Since $S_k \sim Bin\left(k(k + 1)/2, p\right)$, an application of Lemma 3.5(b) with $t = kn^{-1/2} + kp/2$ yields that $P \geq c$ for some $c > 0$. Now using the independence of $(\tau_i)_{i \in [n]}$ and $(B_i)_{i \in [n]}$ we obtain that

$$\mathbb{P}\left(1 + \sum_{i=1}^{t} (\tau_i - 1) > 0 \ \forall t \in [k]\right) = P^{-1}\mathbb{P}\left(1 + \sum_{i=1}^{t} (\tau_i - 1) > 0 \ \forall t \in [k], S_k \geq \frac{k^2}{2}p - \frac{k}{n^{1/2}}\right).$$

Setting $R_t = \sum_{i=1}^{t} (\tau_i + B_i - 1)$, we see that the last quantity is bounded from above by

$$c^{-1} \mathbb{P}\left(1 + R_t > 0 \; \forall t \in [k], \; 1 + R_k \geq \frac{k^2}{2} p - \frac{k}{n^{1/2}}\right)$$

so noting that $\tau_i + B_i \overset{iid}{\sim} \mathrm{Bin}(n, p)$ for every $i \in [n]$ completes the proof. $\qquad\square$

Proof of Lemma 3.2. Write

$$K = K(A, n) = \lceil An^{2/3} \rceil + 1$$

and recall that

$$H = H(A, n) = \left\lceil \frac{\lceil An^{2/3} \rceil^2}{2} p - \frac{\lceil An^{2/3} \rceil}{n^{1/2}} \right\rceil.$$

We want to use Lemma 3.4 to bound from above the quantity

$$\frac{K - 1}{n} \mathbb{P}(R_K \geq H),$$

where we recall that $(R_t)_{t \geq 0}$ is a random walk with $R_0 = 0$ and i.i.d. steps each having distribution $\mathrm{Bin}(n, p) - 1$. Letting $B_{j,p}$ be a binomial random variable with parameters j and p, setting $h := H - K\lambda/n^{1/3}$ (which is of order $A^2 n^{1/3}$ since $\lambda < A/3$) and observing that

$$\frac{h}{p(1-p)nK} + \frac{h^3}{2(1-p)^3(nK)^3} = O(1),$$

applying Lemma 3.4 we obtain

$$\begin{aligned}
\mathbb{P}(R_K \geq H) &= \mathbb{P}(B_{nK,p} \geq K + H) \\
&= \mathbb{P}(B_{nK,p} \geq nKp + H - K\lambda/n^{1/3}) \\
&\leq \frac{(p(1-p)nK)^{1/2}}{h\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{h^2}{p(1-p)nK} + \frac{h}{p(1-p)nK} + \frac{h^3}{2(1-p)^3(nK)^3}\right\} \\
&\leq c\frac{K^{1/2}}{A^2 n^{1/3}} \exp\left\{-\frac{1}{2}\frac{h^2}{p(1-p)nK}\right\} \\
&\leq \frac{c}{A^{3/2}} \exp\left\{-\frac{1}{2}\frac{h^2}{p(1-p)nK}\right\}.
\end{aligned} \tag{4}$$

Define

$$x(A, n, \lambda) = \frac{h}{\sqrt{p(1-p)nK}} = \frac{H - K\lambda/n^{1/3}}{\sqrt{p(1-p)nK}}.$$

Elementary estimates using the fact that $A = o(n^{1/12})$ and $\lambda = o(n^{1/12})$ show that

$$x(A, n, \lambda) = \frac{A^{3/2}}{2} - \lambda A^{1/2} + o(A^{1/2}n^{-1/6}).$$

From (4) we therefore have, for large n,

$$\begin{aligned}
\frac{K - 1}{n} \mathbb{P}(R_K \geq H) &\leq c\frac{A}{n^{1/3}} \frac{1}{A^{3/2}} e^{-x(A,n,\lambda)^2/2} \\
&\leq \frac{c}{\sqrt{A}n^{1/3}} \exp\left(-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}\right),
\end{aligned}$$

which completes the proof of Lemma 3.2. $\qquad\square$

Before we prove Lemma 3.3, we will need the following bound, which is an easy application of Lemma 3.5.

Lemma 3.6 *Suppose that $B_{N,p}$ is a binomial random variable with parameters $N \geq 1$ and $p \in [0, 1]$. Let $C \in (0, \infty)$ be constant. Then for all $x \in (0, C(Np)^{2/3}]$ we have that*

$$\mathbb{P}\left(B_{N,p} \geq Np + x\right) \leq c \exp\left(-\frac{x^2}{2Np}\right)$$

where c is another finite constant.

Proof. Applying Lemma 3.5, we have

$$\mathbb{P}\left(B_{N,p} \geq Np + x\right) \leq \exp\left(-Np\left[\left(1 + \frac{x}{Np}\right)\log\left(1 + \frac{x}{Np}\right) - \frac{x}{Np}\right]\right),$$

and since $\log(1 + t) > t - t^2/2$ for every $t > 0$,

$$\mathbb{P}\left(B_{N,p} \geq Np + x\right) \leq \exp\left(-Np\left[\left(1 + \frac{x}{Np}\right)\left(\frac{x}{Np} - \frac{x^2}{2(Np)^2}\right) - \frac{x}{Np}\right]\right)$$

$$= \exp\left(-Np\left[\frac{x^2}{2(Np)^2} - \frac{x^3}{2(Np)^3}\right]\right)$$

$$= \exp\left(-\frac{x^2}{2Np} + \frac{x^3}{2(Np)^2}\right),$$

which establishes the result with $c = \exp(C^3/2)$. $\square$

Proof of Lemma 3.3. Writing

$$K = K(A, n) = \lceil An^{2/3} \rceil + 1 \text{ and } H = H(A, n) = \left\lceil \frac{\lceil An^{2/3} \rceil^2}{2n} - \frac{\lceil An^{2/3} \rceil}{n^{1/2}} \right\rceil,$$

we aim to bound

$$\frac{1}{K} \sum_{j=H+1}^{K(n-1)} \mathbb{P}(R_K \geq j)$$

from above. We first note that

$$\frac{1}{K} \sum_{j=H+1}^{K(n-1)} \mathbb{P}(R_K \geq j) \leq \frac{1}{K} \sum_{j=H+1}^{\lfloor K^{2/3} \rfloor} \mathbb{P}(R_K \geq j) + n\mathbb{P}(R_K \geq K^{2/3}). \tag{5}$$

To bound the second term on the right-hand side of (5) observe that, since $A = o(n^{1/12})$ and $\lambda = o(n^{1/12})$, we have $K \geq nKp - K^{2/3}/2$ when n is large. Thus, when n is large,

$$n\mathbb{P}(R_K \geq K^{2/3}) = n\mathbb{P}(B_{nK,p} \geq K + K^{2/3})$$

$$\leq n\mathbb{P}(B_{nK,p} \geq nKp + K^{2/3}/2). \tag{6}$$

Using the second inequality in part (a) of Lemma 3.5 we obtain

$$(6) \leq n \exp\left\{-\frac{K^{4/3}}{8(nKp + \frac{1}{6}K^{2/3})}\right\} \leq n \exp\left\{-cA^{1/3}n^{2/9}\right\} \tag{7}$$

and for sufficiently large n we have that

$$(7) \leq \frac{1}{A^2 n^{1/3}} e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}}.$$

Next, for the first term on the right-hand side of (5), note that

$$\frac{1}{K} \sum_{j=H+1}^{\lfloor K^{2/3} \rfloor} \mathbb{P}(R_K \geq j) = \frac{1}{K} \sum_{j=H+1}^{\lfloor K^{2/3} \rfloor} \mathbb{P}(B_{nK,p} \geq K+j)$$

$$= \frac{1}{K} \sum_{j=H+1}^{\lfloor K^{2/3} \rfloor} \mathbb{P}(B_{nK,p} \geq nKp + j - K\lambda n^{-1/3}). \tag{8}$$

Since $A = o(n^{1/12})$ and $\lambda = o(n^{1/12})$, we have $K\lambda n^{-1/3} = o(K^{2/3})$, and therefore we may apply Lemma 3.6 to obtain

$$(8) \leq \frac{c}{K} \sum_{j=H+1}^{\lfloor K^{2/3} \rfloor} \exp\left\{ -\frac{(j - K\lambda n^{-1/3})^2}{2nKp} \right\} \leq \frac{c}{\sqrt{K}} \mathbb{P}\left(G \geq \frac{H + 1 - K\lambda/n^{1/3}}{\sqrt{nKp}} \right),$$

where G denotes a Gaussian random variable with mean zero and unit variance. Recalling the standard bound $\mathbb{P}(G \geq t) \leq (t\sqrt{2\pi})^{-1} e^{-t^2/2}$, which is valid for every $t > 0$, we obtain

$$\mathbb{P}\left(G \geq \frac{H + 1 - K\lambda/n^{1/3}}{\sqrt{nKp}} \right) \leq \frac{1}{\sqrt{2\pi}} \frac{\sqrt{nKp}}{H + 1 - K\lambda/n^{1/3}} \exp\left(-\frac{(H + 1 - K\lambda/n^{1/3})^2}{2nKp} \right).$$

An easy computation reveals that

$$\frac{(H + 1 - K\lambda/n^{1/3})^2}{2nKp} \geq \frac{A^3}{8} - \lambda \frac{A^2}{2} + \lambda^2 \frac{A}{2} + o(1),$$

and consequently we obtain

$$\frac{c}{\sqrt{K}} \mathbb{P}\left(G \geq \frac{H + 1 - K\lambda/n^{1/3}}{\sqrt{nKp}} \right) \leq \frac{c}{A^2 n^{1/3}} \exp\left\{ -\frac{A^3}{8} + \lambda \frac{A^2}{2} - \lambda^2 \frac{A}{2} \right\},$$

as required. □

Remark 3.7 Here we comment on the possibility of obtaining the correct constant factor in the upper bound of the asymptotic for $\mathbb{P}(|\mathcal{C}_{\max}| > An^{2/3})$ using the above methodology. We will give a corresponding comment for the lower bound in Remark 4.15.

There are three places in our proof of the upper bounds stated in Theorem 1.1 where unspecified constants appear. One of these is from Lemma 3.1. The constant appearing within this result could be eliminated by amending the argument to use k/n^γ rather than $k/n^{1/2}$, for e.g. $\gamma = 5/12$, and strengthening our assumption on A appropriately to ensure that the error coming from Lemma 3.5 is smaller than our desired bound.

Another step requiring a more detailed analysis comes from the application of Lemma 3.4 in the proof of Lemma 3.2. Similarly to the first case above, this should be possible but would require stronger conditions on the relationship between A and λ, or error bounds that depend on A/λ.

Finally, a further constant appears when we apply Corollary 2.4 in (3). Finding the correct constant here would require a different approach; either an improvement to Corollary 2.4, or an attempt to apply other results about random walks to $(R_t)_{t \geq 0}$. The latter option would be complicated by the fact that the step distribution of R_t depends on n.

4. Proof of the lower bounds in Theorem 1.1

Let $v \in [n]$ be any vertex in $G(n, p)$ from which we start running the exploration process described at the beginning of Section 3. We write $T_2 = \lceil An^{2/3} \rceil$; we will in due course also have a time T_1 which is smaller than T_2.

Recall from Section 3 that η_i denotes the number of unseen vertices which become active during the ith step of the exploration process, and $Y_t = 1 + \sum_{i=1}^{t} (\eta_i - 1)$ is the number of active vertices at step t of the procedure. Moreover, recall that

$$\eta_i | \mathcal{F}_{i-1} \sim \mathrm{Bin}(n - i + 1 - Y_{i-1}, p), \tag{9}$$

where $\mathcal{F}_i = \sigma(\{\eta_1, \ldots, \eta_i\})$ and $p = p(n) = 1/n + \lambda n^{-4/3}$. We will start by proving the lower bound in part (a) of Theorem 1.1; that is, by bounding from below the probability

$$\mathbb{P}\left(|\mathcal{C}(v)| > T_2\right) = \mathbb{P}\left(1 + \sum_{i=1}^{t} (\eta_i - 1) > 0 \ \forall t \in [T_2]\right).$$

We note that the random variables η_i are not independent, which makes our analysis more difficult. The first part of our argument, therefore, consists of replacing the η_i with a sequence of independent binomial random variables which are easier to analyse. The idea is that η_i, which is the number of neighbours of our ith vertex that are unseen, is roughly $\mathrm{Bin}(n - i, p)$, but the first parameter is slightly smaller due to the (random) number of active vertices that are present at the beginning of the ith step of the exploration process. If we can bound the number of active vertices above by some deterministic value K with high probability, then we can remove this source of randomness and obtain a sequence of independent increments in place of the η_i. To this end, fix $K \in \mathbb{N}$ and suppose that $(\delta_i)_{i \in [T_2]}$ is a sequence of independent random variables with $\delta_i \sim \mathrm{Bin}(n - K - i, p)$ and set $R_t = 1 + \sum_{i=1}^{t} (\delta_i - 1)$. We note that the definitions of δ_i and R_t depend implicitly on K; sometimes for clarity we will write $\delta_i^{(K)}$ and $R_t^{(K)}$. We will soon fix $K = \lfloor n^{2/5} \rfloor$, but the following lemma works for any $K < n - T_2$. We postpone the proof, which constructs a coupling between η_i and δ_i, until Section 4.3.

Lemma 4.1 *Suppose that $K + T_2 < n$. Define $\tau_0 := \min\{t \geq 1 : Y_t = 0\}$, the first time at which the set of active vertices becomes empty. Then*

$$\mathbb{P}\left(|\mathcal{C}(v)| > T_2\right) \geq \mathbb{P}\left(R_t^{(K)} > 0 \ \forall t \in [T_2]\right) - \mathbb{P}\left(\exists i \in [\tau_0 \wedge T_2] : Y_i \geq K\right). \tag{10}$$

Our next result shows that if we choose $K = \lfloor n^{2/5} \rfloor$, then we do not have to worry about the last probability on the right-hand side of (10).

Lemma 4.2 *Let τ_0 be as in the previous lemma. As $n \to \infty$,*

$$\mathbb{P}\left(\exists i \in [\tau_0 \wedge T_2] : Y_i \geq \lfloor n^{2/5} \rfloor\right) = o\left(A^{-1/2} n^{-1/3} e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}}\right). \tag{11}$$

The proof of Lemma 4.2, which easily follows from Lemma 3.5, is again postponed to Section 4.3.

Given (11), we can now fix $K = \lfloor n^{2/5} \rfloor$ and focus on providing a lower bound for

$$\mathbb{P}\left(R_t^{(K)} > 0 \ \forall t \in [T_2]\right). \tag{12}$$

Observe that, although we now have a process with independent increments, obtaining a lower bound for (12) remains a non-trivial task, because the δ_i that are used to define R_t are not identically distributed. We consider two options to produce a random walk with i.i.d. increments from $(R_t)_{t \in [T_2]}$. The first is to view δ_i as a sum of i.i.d. Bernoulli random variables, with δ_1 summing more Bernoullis than δ_2 and so on; and then to rearrange the same Bernoullis amongst sums

δ_i' that all have equal length. The second is simply to add an independent $\text{Bin}(K + i, p)$ random variable to δ_i for each i.

It turns out that neither of these two options works on its own. The first has problems if we try to cover too many values of i, since the more Bernoullis that we have to rearrange, the less accurate our estimates become. The second has problems when i is small, as the variance of the added $\text{Bin}(K + i, p)$ random variables is too large when our random walk is near the origin.

We therefore combine the two techniques. We take $T_1 \in [T_2]$ and carry out the first strategy for times $t \in [T_1]$, and the second strategy for $t \in [\![T_1, T_2]\!]$.

We note first that for any deterministic $H \in \mathbb{N}$ and $T_1 \in [T_2]$,

$$\mathbb{P}\big(R_t > 0 \; \forall t \in [T_2]\big)$$
$$\geq \mathbb{P}\big(R_t > 0 \; \forall t \in [T_1], \; R_{T_1} \in [H, 2H], \; R_t > 0 \; \forall t \in [\![T_1, T_2]\!]\big)$$
$$\geq \mathbb{P}\big(R_t > 0 \; \forall t \in [T_1], \; R_{T_1} \in [H, 2H]\big)\mathbb{P}\big(R_t > 0 \; \forall t \in [\![T_1, T_2]\!] \; \big| \; R_{T_1} = H\big). \quad (13)$$

We now fix $T_1 = 2\lfloor n^{2/3}/A^2 \rfloor - 1$ and $H = \lceil n^{1/3}/A \rceil$.

Proposition 4.3 *There exists $c > 0$ such that for sufficiently large n and A,*

$$\mathbb{P}\left(R_t > 0 \; \forall t \in [T_1], R_{T_1} \in [H, 2H]\right) \geq c\frac{A}{n^{1/3}}.$$

Proposition 4.4 *There exists $c > 0$ such that for sufficiently large n and A,*

$$\mathbb{P}\big(R_t > 0 \; \forall t \in [\![T_1, T_2]\!] \; \big| \; R_{T_1} = H\big) \geq \frac{c}{A^{3/2}}e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}}.$$

We will prove Proposition 4.3 in Section 4.1 and Proposition 4.4 in Section 4.2. For now we show how these results can be used to complete the proof of the lower bounds in Theorem 1.1.

Proof of lower bounds in Theorem 1.1. By Lemmas 4.1 and 4.2,

$$\mathbb{P}\left(|\mathcal{C}(v)| > T_2\right) \geq \mathbb{P}\left(R_t > 0 \; \forall t \in [T_2]\right) - o\left(A^{-1/2}n^{-1/3}e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}}\right).$$

In light of (13), it then follows from Propositions 4.3 and 4.4 that

$$\mathbb{P}\left(|\mathcal{C}(v)| > T_2\right) \geq \frac{c}{A^{1/2}n^{1/3}}e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}}. \quad (14)$$

This concludes the proof of the lower bound in part (*a*) of Theorem 1.1. In order to prove the lower bound in part (*b*), we will need to use the fact that for any $\mathbb{N}_0$-valued random variable X,

$$\mathbb{P}(X \geq 1) \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}. \quad (15)$$

This can be proved by applying the Cauchy–Schwarz inequality to $X\mathbb{1}_{\{X \geq 1\}}$.

To proceed with the proof of the lower bound in part (*b*), let us denote by $X = \sum_{i=1}^{n} \mathbb{1}_{\{|\mathcal{C}(i)| \in [T_2, 2T_2]\}}$ the number of components of size between T_2 and $2T_2$. Observe that $X \geq 1$ implies $|\mathcal{C}_{\max}| \geq T_2$. Therefore using (15) we obtain

$$\mathbb{P}\left(|\mathcal{C}_{\max}| \geq T_2\right) \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}. \quad (16)$$

For the numerator, we have

$$\mathbb{E}[X]^2 = n^2\mathbb{P}\left(|C(1)| \in [T_2, 2T_2]\right)^2. \quad (17)$$

Next we bound the denominator from above. Given vertices $i, j \in [n]$, write $i \leftrightarrow j$ if there exists a path of opens edges between i and j. Then we can write

$$\mathbb{E}[X^2] \leq n\mathbb{P}\left(|\mathcal{C}(1)| \in [T_2, 2T_2]\right) + S_1 + S_2, \quad (18)$$

where

$$S_1 = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{1}_{\{|\mathcal{C}(i)|\in[T_2,2T_2]\}}\mathbb{1}_{\{|\mathcal{C}(j)|\in[T_2,2T_2]\}}\mathbb{1}_{\{i\nleftrightarrow j\}}\right]$$

and

$$S_2 = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j\neq i}\mathbb{1}_{\{|\mathcal{C}(i)|\in[T_2,2T_2]\}}\mathbb{1}_{\{|\mathcal{C}(j)|\in[T_2,2T_2]\}}\mathbb{1}_{\{i\leftrightarrow j\}}\right].$$

For S_1 we have

$$S_1 \le n^2\sum_{k=T_2}^{2T_2}\mathbb{P}\Big(|\mathcal{C}(1)| = k, 1 \nleftrightarrow 2\Big)\mathbb{P}\Big(|C(2)| \in [T_2, 2T_2]\ \Big|\ |\mathcal{C}(1)| = k, 1 \nleftrightarrow 2\Big)$$

$$\le n^2\mathbb{P}\left(|\mathcal{C}(1)| \in [T_2, 2T_2]\right)\mathbb{P}\left(|\mathcal{C}(2)| \ge T_2\right).$$

For S_2 we have

$$S_2 = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{k=T_2}^{2T_2}\mathbb{1}_{\{|\mathcal{C}(i)|=k\}}\sum_{j\neq i}\mathbb{1}_{\{j\in\mathcal{C}(i)\}}\right] \le \mathbb{E}\left[\sum_{i=1}^{n}\sum_{k=T_2}^{2T_2}\mathbb{1}_{\{|\mathcal{C}(i)|=k\}}k\right]$$

$$\le 2T_2 n\mathbb{P}\left(|C(1)| \in [T_2, 2T_2]\right).$$

Returning to (18) and recalling that $T_2 = \lceil An^{2/3}\rceil$, we obtain

$$\mathbb{E}[X^2] \le n\mathbb{P}\left(|\mathcal{C}(1)| \in [T_2, 2T_2]\right) + n^2\mathbb{P}\left(|\mathcal{C}(1)| \in [T_2, 2T_2]\right)\mathbb{P}\left(|\mathcal{C}(2)| \ge T_2\right)$$
$$+ 3An^{5/3}\mathbb{P}\left(|\mathcal{C}(1)| \in [T_2, 2T_2]\right).$$

By the upper bound in part (*a*) of Theorem 1.1,

$$\mathbb{P}\left(|\mathcal{C}(2)| \ge T_2\right) \le \frac{c_2}{A^{1/2}n^{1/3}}$$

and therefore

$$\mathbb{E}[X^2] \le cAn^{5/3}\mathbb{P}\left(|C(1)| \in [T_2, 2T_2]\right).$$

Substituting this and (17) into (16) and then applying (14), we obtain

$$\mathbb{P}\left(|\mathcal{C}_{\max}| \ge \lceil An^{2/3}\rceil\right) \ge \frac{n^2\mathbb{P}\left(|\mathcal{C}(1)| \in [T_2, 2T_2]\right)^2}{cAn^{5/3}\mathbb{P}\left(|\mathcal{C}(1)| \in [T_2, 2T_2]\right)}$$

$$= c\frac{n^{1/3}}{A}\mathbb{P}\left(|\mathcal{C}(1)| \in [T_2, 2T_2]\right). \tag{19}$$

Observe that

$$\mathbb{P}\left(|\mathcal{C}(1)| \in [T_2, 2T_2]\right) = \mathbb{P}(|\mathcal{C}(1)| \ge T_2) - \mathbb{P}(|\mathcal{C}(1)| > 2T_2),$$

and by applying the upper bound in Theorem 1.1(a) with $2A$ in place of A, together with (14), for A sufficiently large we may ensure that $\mathbb{P}(|\mathcal{C}(1)| > 2T_2) \le \mathbb{P}(|\mathcal{C}(1)| \ge T_2)/2$. Therefore (19) is at least

$$\frac{c}{A^{3/2}}e^{-\frac{A^3}{8}+\frac{\lambda A^2}{2}-\frac{\lambda^2 A}{2}},$$

as required. This completes the proof of Theorem 1.2, subject to the proofs of Lemmas 4.1 and 4.2 and Propositions 4.3 and 4.4. □

4.1. Rearranging Bernoullis and applying the ballot theorem: proof of Proposition 4.3

We first introduce a technical result which will be used to transform $(R_t)_{t \in [T_1]}$ into a process with i.i.d. increments.

Lemma 4.5 *Suppose that $N \in \mathbb{N}$, and that $L \in [N]$ is odd. Let $(I_j^i)_{i,j \geq 1}$ be i.i.d. non-negative random variables and set $X_i = \sum_{j=1}^{N-i} I_j^i$ for $i = 1, \ldots, L$. Then there exist i.i.d. random variables $(\tilde{I}_j^i)_{i,j \geq 1}$ with the same distribution as I_j^i such that if we set $\tilde{X}_i = \sum_{j=1}^{N-(L+1)/2} \tilde{I}_j^i$ then*

- $\sum_{i=1}^t \tilde{X}_i \leq \sum_{i=1}^t X_i$ *for all $1 \leq t \leq L$;*
- $\sum_{i=1}^L \tilde{X}_i = \sum_{i=1}^L X_i$.

The reader can think of the I_j^i as Bernoulli(p)-distributed, so that $X_i \sim \text{Bin}(N - i, p)$ and $\tilde{X}_i \sim$ Bin$(N - (L+1)/2, p)$. The idea behind the proof is that X_1 has more summands than X_L, so if we transfer some of the summands from X_1 to X_L, we do not change the value of $\sum_{i=1}^L X_i$ but we decrease X_1. Then we move on to X_2 and transfer some of its summands to X_{L-1}, which decreases $\sum_{i=1}^2 X_i$ without changing $\sum_{i=1}^L X_i$; and so on. We postpone the details until Section 4.3.

Before we can proceed with the proof of Proposition 4.3, we need one more tool. We can use Lemma 4.5 to transform $(R_t)_{t \in [T_1]}$ into a process with i.i.d. increments, but in order to apply the generalised ballot theorem, Theorem 2.1, we need our increments also to have mean zero and for their distribution not to depend on n. The following lemma is slightly more general than we will need.

Lemma 4.6 *Take $n \in \mathbb{N}$, $h_n \geq 0$, $a_n \in (-1, \infty)$ satisfying $na_n \in \mathbb{Z}$, $b_n \in (-1, n-1)$ and $t_n \in \mathbb{N}$. Suppose that $M_t = 1 + \sum_{i=1}^t (W_i - 1)$ where the W_i are independent Bin$(n(1 + a_n), (1 + b_n)/n)$ random variables. Let $\mu_n = (1 + a_n)(1 + b_n)$. Then*

$$\mathbb{P}\big(M_t > 0 \ \ \forall t \in [t_n], M_{t_n} \in [h_n, 2h_n]\big)$$

$$\geq (\mu_n \wedge 1)^{2h_n} \mu_n^{t_n - 1} e^{(1 - \mu_n)t_n} \mathbb{P}\big(\hat{M}_t > 0 \ \ \forall t \in [t_n], \hat{M}_{t_n} \in [h_n, 2h_n]\big) - \frac{t_n}{n}(1 + a_n)(1 + b_n)^2$$

where $\hat{M}_t = 1 + \sum_{i=1}^t (\hat{W}_i - 1)$, and $(\hat{W}_i)_{i=1}^{t_n}$ is a sequence of independent Poisson random variables with mean one.

We delay the proof, which uses a fairly standard Poisson approximation for the binomial distribution and then a simple change of measure to remove the drift, until Section 4.4 and proceed with the proof of Proposition 4.3.

Proof of Proposition 4.3. As previously mentioned, we want to bound

$$\mathbb{P}\left(R_t > 0 \ \forall t \in [T_1], R_{T_1} \in [H, 2H]\right)$$

by means of the generalised ballot theorem, Theorem 2.1. To this end, we first need to turn the process $(R_t)_{t \in [T_1]}$ into a random walk with i.i.d. steps having mean zero. In order to obtain identically distributed steps we will make use of Lemma 4.5.

Recall that $H = \lceil n^{1/3}/A \rceil$ and $R_t = 1 + \sum_{i=1}^t (\delta_i - 1)$, where each δ_i is the sum of $n - \lfloor n^{2/5} \rfloor - i$ i.i.d. *Ber*(p) random variables. It follows from Lemma 4.5, with $N = n - \lfloor n^{2/5} \rfloor$ and $L = T_1$, that there exists a sequence $(\tilde{\delta}_i)_{i \in [T_1]}$ of i.i.d random variables with $\tilde{\delta}_i \sim$ Bin$(n - \lfloor n^{2/5} \rfloor - (T_1 + 1)/2, p)$ for which, setting $\tilde{R}_t = 1 + \sum_{i=1}^t (\tilde{\delta}_i - 1)$, we obtain

$$\mathbb{P}\left(R_t > 0 \ \forall t \in [T_1], R_{T_1} \in [H, 2H]\right) \geq \mathbb{P}\left(\tilde{R}_t > 0 \ \forall t \in [T_1], \tilde{R}_{T_1} \in [H, 2H]\right). \tag{20}$$

In order to evaluate the probabilities appearing in the above sum by means of the generalised ballot theorem, we still have to turn $(\tilde{R}_t)_{t \in [T_1]}$ into a process whose increments have mean zero.

We do this by applying Lemma 4.6 with $h_n = H$, $t_n = T_1 = 2\lfloor n^{2/3}/A^2 \rfloor - 1$, $a_n = -\lfloor n^{2/5} \rfloor/n - (T_1 + 1)/(2n)$ and $b_n = \lambda/n^{1/3}$. Since $|\lambda| \leq A/3$, it is easy to see that there exists a constant $c > 0$ (not depending on n) such that

$$(\mu_n \wedge 1)^{2h_n} \geq c.$$

Also, using the inequality $1 + x \geq e^{x - x^2}$ valid for $x \geq -1/2$, for sufficiently large n we have

$$\mu_n^{t_n - 1} e^{(1 - \mu_n)t_n} = \mu_n^{-1}(1 + (\mu_n - 1))^{t_n} e^{(1 - \mu_n)t_n} \geq \mu_n^{-1} e^{-(\mu_n - 1)^2 t_n} \geq c$$

for some constant $c > 0$. Finally, since

$$\frac{t_n}{n}(1 + a_n)(1 + b_n)^2 \asymp \frac{t_n}{n} \asymp \frac{1}{A^2 n^{1/3}},$$

from Lemma 4.6 we obtain that

$$\mathbb{P}\left(\tilde{R}_t > 0 \; \forall t \in [T_1], \; \tilde{R}_{T_1} \in [H, 2H]\right)$$
$$\geq c\mathbb{P}\left(\hat{R}_t > 0 \; \forall t \in [T_1], \; \hat{R}_{T_1} \in [H, 2H]\right) - \frac{C}{A^2 n^{1/3}} \tag{21}$$

for some constants $c > 0$ and $C < \infty$, where $\hat{R}_t = 1 + \sum_{i=1}^{t} (\hat{\delta}_i - 1)$ and $(\hat{\delta}_i)_{i=1}^{T_1}$ is a sequence of independent Poisson random variables with parameter 1.

We are now in a position to apply Theorem 2.1. Recalling that $H = \lceil n^{1/3}/A \rceil$ and $T_1 = 2\lfloor n^{2/3}/A^2 \rfloor - 1$, for all $k \in [H - 1, 2H - 1]$ we have $k \leq 2H = O(\sqrt{T_1})$. We can therefore conclude from Theorem 2.1 that

$$\mathbb{P}\left(\hat{R}_t > 0 \; \forall t \in [T_1], \; \hat{R}_{T_1} \in [H, 2H]\right)$$
$$\geq \mathbb{P}\left(\hat{R}_t - 1 > 0 \; \forall t \in [T_1], \; \hat{R}_{T_1} - 1 \in [H - 1, 2H - 1]\right)$$
$$\geq c \sum_{k=H-1}^{2H-1} \frac{k+1}{T_1^{3/2}},$$

which is of order $An^{-1/3}$. Substituting this bound into (21) gives

$$\mathbb{P}\left(\tilde{R}_t > 0 \; \forall t \in [T_1], \; \tilde{R}_{T_1} \in [H, 2H]\right) \geq \frac{cA}{n^{1/3}} - \frac{C}{A^2 n^{1/3}}.$$

Taking A sufficiently large that the first term dominates, and then recalling (20), gives the result. □

4.2. Adding independent binomials and approximating with Brownian motion: Proof of Proposition 4.4

Recall that $R_t = 1 + \sum_{i=1}^{t} (\delta_i - 1)$ where $(\delta_i)_{i=1}^{T_2}$ is a sequence of independent $\mathrm{Bin}(n - K - i, p)$ random variables. Recall also that $H = \lceil n^{1/3}/A \rceil$, $K = \lfloor n^{2/5} \rfloor$, $T_1 = 2\lfloor n^{2/3}/A^2 \rfloor - 1$ and $T_2 = \lceil An^{2/3} \rceil$. Throughout this section we write $T = T_2 - T_1$.

Our first task in this section is to replace R_t with a sum of i.i.d. random variables. We do this by adding an independent $\mathrm{Bin}(K + i, p)$ random variable to δ_i for each i, and checking that the sum of these additional random variables cannot be too large using Lemma 3.5.

Lemma 4.7 *For $t \in [0, \infty)$, define*

$$g(t) = -\frac{n^{1/3}}{2A} + \frac{9t}{A^2 n^{1/3}} + \frac{pt^2}{2}.$$

Then there exists $c > 0$ such that for all large n,

$$\mathbb{P}(R_t > 0 \ \forall t \in [\![T_1, T_2]\!] \mid R_{T_1} = H) \geq \mathbb{P}(S_t > g(t) \ \forall t \in [\![1, T]\!]) - Te^{-cAn^{1/6}},$$

where $S_t = \sum_{i=1}^{t} (\Delta_i - 1)$, and $(\Delta_i)_{i=1}^{T}$ is a sequence of independent Bin(n,p) random variables.

The proof of Lemma 4.7 is in Section 4.3. Next, in order to apply a Brownian approximation to our random walk, we would like the step distribution not to depend on *n*.

Lemma 4.8 *For $t \in [0, \infty)$, define*

$$\gamma(t) = -\frac{n^{1/3}}{4A} + \frac{9t}{A^2 n^{1/3}} + \frac{pt^2}{2}.$$

Let g and $(S_t)_{t=0}^{T}$ be as in Lemma 4.7. Then there exist constants $c, C \in (0, \infty)$ such that

$$\mathbb{P}(S_t > g(t) \ \forall t \in [\![1, T]\!])$$
$$\geq ce^{\lambda A^2/2 - \lambda^2 A/2} \mathbb{P}(\hat{S}_t > \gamma(t) \ \forall t \in [\![1, T]\!], \ \hat{S}_T \leq \gamma(T) + \tfrac{3n^{1/3}}{8A}) - C\exp\left(-\frac{n^{1/3}}{4A}\right),$$

where $\hat{S}_t = \sum_{i=1}^{t} (\hat{\Delta}_i - 1)$ and $(\hat{\Delta}_i)_{i=1}^{T}$ is a sequence of independent Poisson random variables of parameter 1.

Again we delay the proof, which is similar to the proof of Lemma 4.6, until Section 4.4 and proceed with the proof of Proposition 4.4. As mentioned above, our strategy is to approximate the random walk $(\hat{S}_t)_{t=0}^{T}$ appearing in Lemma 4.8 with Brownian motion. We will use an accurate bound of Komlós, Major and Tusnády due to its ease of application, although we will not really need the additional precision gained over the earlier result of Strassen [31, Theorem 1.5]. The following rephrasing of the original theorem is from [9].

Theorem 4.9 (Komlós et al. [18]). *Let $(\xi_i)_{i\geq 1}$ be a sequence of i.i.d. random variables with $\mathbb{E}[\xi_1] = 0$ and $\mathbb{E}[\xi_1^2] = 1$. Suppose that there exists $\theta > 0$ such that $\mathbb{E}[e^{\theta|\xi_1|}] < \infty$. For each $k \in \{0\} \cup \mathbb{N}$, let $U_k = \sum_{i=1}^{k} \xi_i$. Then for every $N \in \mathbb{N}$ it is possible to construct a version of $(U_k)_{k=0}^{N}$ and a standard Brownian Motion $(B_s)_{s \in [0,N]}$ on the same probability space such that, for every $x \geq 0$,*

$$\mathbb{P}\left(\max_{k \leq N} |U_k - B_k| > M \log N + x\right) \leq Ce^{-cx},$$

where M, C and $c > 0$ do not depend on N.

Applying this with $N = T$ and $U_k = \hat{S}_k$, we immediately obtain the following corollary.

Corollary 4.10 *Suppose that $\hat{S}_t = \sum_{i=1}^{t} \hat{\Delta}_i$, where $(\hat{\Delta}_i)_{i=1}^{T}$ is a sequence of independent Poisson random variables of parameter 1, and that $(B_s)_{s\geq 0}$ is a standard Brownian motion. There exist constants $c, C \in (0, \infty)$ such that for any $x_n \geq 0$ and any function $\gamma:[0, \infty) \to \mathbb{R}$,*

$$\mathbb{P}(\hat{S}_t > \gamma(t) \ \forall t \in [\![1, T]\!], \ \hat{S}_T \leq \gamma(T) + \tfrac{3n^{1/3}}{8A})$$
$$\geq \mathbb{P}(B_s > \gamma(s) + M\log T + x_n \ \forall s \in [0, T], \ B_T \leq \gamma(T) + \tfrac{3n^{1/3}}{8A} - M\log T - x_n) - Ce^{-cx_n}.$$

We have now reduced our task to bounding the probability that a Brownian motion remains above a curve up to time T and is not too far above the curve at time T.

Proposition 4.11 *There exists a constant $c > 0$ such that for any x_n satisfying $A^3 \ll x_n \ll n^{1/3}/A$, any constant M (not depending on n) and γ as in Lemma 4.8, for large n,*

$$\mathbb{P}(B_s > \gamma(s) + M\log T + x_n \ \forall s \in [0, T], \ B_T \leq \gamma(T) + \tfrac{3n^{1/3}}{8A} - M\log T - x) \geq \frac{c}{A^{3/2}} e^{-A^3/8}.$$

The proof of Proposition 4.11 involves considering two time intervals, $[0, T/2]$ and $[T/2, T]$, and approximating $\gamma(T)$ by a straight line on each of these intervals. We carry out the details in Section 4.5.

We now have all the ingredients to prove Proposition 4.4 and therefore Theorem 1.1.

Proof of Proposition 4.4. We simply combine Lemmas 4.7 and 4.8, Corollary 4.10 and Proposition 4.11. □

4.3. Proofs of Lemmas 4.1, 4.2, 4.5 and 4.7: creating i.i.d. sequences

We first prove Lemma 4.1, which replaces η_i, the number of unseen vertices that become active at the ith step of the exploration process, with an independent Binomial random variable that does not depend on the history of the exploration process.

Proof of Lemma 4.1. From the description of the exploration process provided at the beginning of Section 3, recall that u_t is the vertex that is explored at step t. Let us denote by $\mathcal{A}_t^*$ the set of unseen vertices that become active at step $t - 1$ of the process (with $\mathcal{A}_1^* = \{u_1\}$), and let $\mathcal{A}_t = \bigcup_{i=0}^t \mathcal{A}_i^*$, the set of all active or explored vertices after step $t - 1$. Also, write X_v^t for the indicator that u_t is a neighbour of vertex v.

For each $t = 1, 2, \ldots, n - K$, if $|\mathcal{A}_t| < K + t$ then let $\mathcal{B}_t^*$ be any subset of the vertices $[n]$ such that

- $\mathcal{A}_t^* \subset \mathcal{B}_t^*$;
- $\mathcal{B}_t^* \cap \mathcal{A}_{t-1} = \emptyset$;
- $\left| \mathcal{B}_t^* \cup \mathcal{A}_{t-1} \right| = K + t$.

If $|\mathcal{A}_t| \geq K + t$ then let $\mathcal{B}_t^* = \mathcal{A}_t^*$. Then let

$$r_t = \left| \mathcal{B}_t^* \cup \mathcal{A}_{t-1} \right| - K - t \geq 0.$$

Take a sequence $\hat{X}_1^t, \hat{X}_2^t, \ldots$ of independent Bernoulli random variables of parameter p, also independent of everything else.

Note that

$$\eta_t = \sum_{v \notin \mathcal{A}_t} X_v^t$$

and define

$$\delta_t = \sum_{v \notin \mathcal{B}_t^* \cup \mathcal{A}_{t-1}} X_v^t + \sum_{i=1}^{r_t} \hat{X}_j^t.$$

Then, since

$$\left| \left(\mathcal{B}_t^* \cup \mathcal{A}_{t-1} \right)^c \right| + r_t = n - K - t,$$

and the random variables $\{X_v^i : v \in \mathcal{A}_{i-1}^c\}$ are independent and independent of $\{X_v^j : v \in \mathcal{A}_{j-1}^c\}$ for any $j \neq i$, we see that $(\delta_t)_{t=1}^{n-K}$ is a sequence of independent random variables such that $\delta_t \sim \text{Bin}(n - K - t, p)$.

We also observe that if $|\mathcal{A}_t| < K + t$, then $|\mathcal{B}_t^* \cup \mathcal{A}_{t-1}| = K + t$ and so $r_t = 0$. Since we also have $\mathcal{A}_t^* \subset \mathcal{B}_t^*$, we see that if $|\mathcal{A}_t| < K + t$ then $\eta_t \geq \delta_t$. Thus, recalling that τ_0 denotes the first time at which the set of active vertices is empty and since $\{\tau_0 > T_2\} = \{1 + \sum_{i=1}^t (\eta_i - 1) > 0 \ \forall t \in [T_2]\}$ we obtain

$$\mathbb{P}\left(|\mathcal{C}(\nu)| > T_2\right) = \mathbb{P}\left(1 + \sum_{i=1}^{t}(\eta_i - 1) > 0 \forall t \in [T_2]\right)$$

$$\geq \mathbb{P}\left(1 + \sum_{i=1}^{t}(\eta_i - 1) > 0 \forall t \in [T_2] \text{ and } |\mathcal{A}_t| < K + t \forall t \in [\tau_0 \wedge T_2]\right)$$

$$\geq \mathbb{P}\left(1 + \sum_{i=1}^{t}(\delta_i - 1) > 0 \forall t \in [T_2] \text{ and } |\mathcal{A}_t| < K + t \forall t \in [\tau_0 \wedge T_2]\right),$$

where the last inequality follows from the fact that on the event $\{|\mathcal{A}_t| < K + t \forall t \in [T_2]\}$ we have $\eta_t \geq \delta_t$ for $t \in [T_2]$ (and $\tau_0 \wedge T_2 = T_2$ on $\{1 + \sum_{i=1}^{t}(\eta_i - 1) > 0 \ \forall t \in [T_2]\}$). Now

$$\mathbb{P}\left(1 + \sum_{i=1}^{t}(\delta_i - 1) > 0 \forall t \in [T_2] \text{ and } |\mathcal{A}_t| < K + t \forall t \in [\tau_0 \wedge T_2]\right)$$

$$\geq \mathbb{P}\left(1 + \sum_{i=1}^{t}(\delta_i - 1) > 0 \forall t \in [T_2]\right) - \mathbb{P}\left(\exists t \in [\tau_0 \wedge T_2] : |\mathcal{A}_t| \geq K + t\right).$$

Since $|\mathcal{A}_t| = Y_t + t$, the result follows. □

Next we prove Lemma 4.2, which ensures that the probability that the number of active vertices becomes too large is small.

Proof of Lemma 4.2. Let $\zeta_i \overset{i.i.d.}{\sim} \text{Bin}(n, p)$. A union bound gives

$$\mathbb{P}\left(\exists i \leq \tau_0 \wedge T_2 : Y_i \geq \lfloor n^{2/5} \rfloor\right) \leq \sum_{i=1}^{T_2} \mathbb{P}\left(Y_i \geq \lfloor n^{2/5} \rfloor, i \leq \tau_0\right)$$

$$\leq \sum_{i=1}^{T_2} \mathbb{P}\left(1 + \sum_{j=1}^{i}(\zeta_j - 1) \geq \lfloor n^{2/5} \rfloor, i \leq \tau_0\right) \tag{22}$$

$$\leq \sum_{i=1}^{T_2} \mathbb{P}\left(1 + \sum_{j=1}^{i}(\zeta_j - 1) \geq \lfloor n^{2/5} \rfloor\right). \tag{23}$$

Denoting by $B_{N,q}$ a binomial random variable with parameters N and q, by Lemma 3.5 we see that for $i \in [T_2]$,

$$\mathbb{P}\left(1 + \sum_{j=1}^{i}(\zeta_j - 1) \geq \lfloor n^{2/5} \rfloor\right) = \mathbb{P}\left(B_{in,p} \geq inp - i\lambda n^{-1/3} + \lfloor n^{2/5} \rfloor - 1\right)$$

$$\leq \mathbb{P}\left(B_{in,p} \geq inp + n^{2/5}\left(1 - c\frac{A|\lambda|}{n^{1/15}}\right)\right). \tag{24}$$

Now, $A = o(n^{1/30})$ and $|\lambda| \leq A/3$, so $A|\lambda| = o\left(n^{1/15}\right)$ and hence for large enough n we obtain

$$(24) \leq \mathbb{P}\left(B_{in,p} \geq inp + n^{2/5}/2\right) \leq \exp\left\{-\frac{n^{4/5}/4}{2inp + \frac{1}{3}n^{2/5}}\right\}. \tag{25}$$

Since $i \leq T_2$ we see that $(25) \leq e^{-cn^{2/15}/A}$ for some positive constant $c > 0$. Finally, since $A, \lambda = o(n^{1/30})$ as $n \to \infty$ we conclude that

$$e^{-cn^{2/15}/A} = o\left(A^{-1/2}n^{-1/3}e^{-\frac{A^3}{8} + \frac{\lambda A^2}{2} - \frac{\lambda^2 A}{2}}\right). \qquad □$$

Lemma 4.5 involves rearranging Bernoulli random variables to produce an i.i.d. sequence.

Proof of Lemma 4.5. Recall the convention that the empty sum is zero. By hypothesis,

$$X_i = \sum_{j=1}^{N-i} I_j^i, \tag{26}$$

where $(I_j^i)_{i,j \geq 1}$ are i.i.d. non-negative random variables. Let $\ell = (L+1)/2$; recall that L is odd, so $\ell \in \mathbb{N}$. Define

$$\tilde{X}_i = \begin{cases} \displaystyle\sum_{j=1}^{N-\ell} I_j^i, & 1 \leq i \leq \ell \\[6mm] \displaystyle\sum_{j=1}^{N-i} I_j^i + \sum_{j=N-\ell+1}^{N-(L+1-i)} I_j^{L+1-i}, & \ell < i \leq L. \end{cases}$$

Observe that $(\tilde{X}_i)_{i=1}^L$ is a sequence of i.i.d. random variables and $\tilde{X}_i \stackrel{d}{=} \sum_{j=1}^{N-\ell} I_j^1$, $1 \leq i \leq L$. Next we claim that

$$\sum_{i=1}^{t} \tilde{X}_i \leq \sum_{i=1}^{t} X_i \tag{27}$$

for all $1 \leq t \leq L$. To see this, observe first that when $1 \leq t \leq \ell$ we have that

$$\sum_{i=1}^{t} \tilde{X}_i = \sum_{i=1}^{t} \sum_{j=1}^{N-\ell} I_j^i \leq \sum_{i=1}^{t} \sum_{j=1}^{N-i} I_j^i = \sum_{i=1}^{t} X_i.$$

Next, for $\ell < t \leq L$ we have that

$$\sum_{i=1}^{t} \tilde{X}_i = \sum_{i=1}^{\ell} \tilde{X}_i + \sum_{i=\ell+1}^{t} \tilde{X}_i = \sum_{i=1}^{\ell} \sum_{j=1}^{N-\ell} I_j^i + \sum_{i=\ell+1}^{t} \sum_{j=1}^{N-i} I_j^i + \sum_{i=\ell+1}^{t} \sum_{j=N-\ell+1}^{N-(L+1-i)} I_j^{L+1-i}.$$

Making the change of variable $k = L + 1 - i$, we see that

$$\sum_{i=\ell+1}^{t} \sum_{j=N-\ell+1}^{N-(L+1-i)} I_j^{L+1-i} = \sum_{k=L+1-t}^{\ell-1} \sum_{j=N-\ell+1}^{N-k} I_j^k = \sum_{k=L+1-t}^{\ell} \sum_{j=N-\ell+1}^{N-k} I_j^k,$$

where the last equality follows from the fact that the term corresponding to $k = \ell$ is zero. Therefore

$$\sum_{i=1}^{t} \tilde{X}_i = \sum_{i=1}^{\ell} \sum_{j=1}^{N-\ell} I_j^i + \sum_{i=\ell+1}^{t} \sum_{j=1}^{N-i} I_j^i + \sum_{i=L+1-t}^{\ell} \sum_{j=N-\ell+1}^{N-i} I_j^i \tag{28}$$

$$\leq \sum_{i=1}^{\ell} \sum_{j=1}^{N-\ell} I_j^i + \sum_{i=\ell+1}^{t} \sum_{j=1}^{N-i} I_j^i + \sum_{i=1}^{\ell} \sum_{j=N-\ell+1}^{N-i} I_j^i$$

$$= \sum_{i=1}^{t} X_i$$

as claimed. The second statement of the lemma simply follows by taking $t = L$ in (28), in which case the subsequent inequality is an equality. $\square$

Lemma 4.7 provides an alternative way of producing i.i.d. sequences of Binomial random variables, by adding independent binomials to the original sequence.

Proof of Lemma 4.7. Recall that $R_t = 1 + \sum_{i=1}^{t} (\delta_i - 1)$. As we did in Section 4.2, write $T = T_2 - T_1$. Let $(L_i)_{i \in [T_2]}$ be a sequence of independent random variables, also independent of $(\delta_i)_{i \in [T_2]}$ and such that $L_i \sim \text{Bin}(K + i, p)$. Then, setting

$$S_t = \sum_{i=T_1+1}^{T_1+t} (\delta_i + L_i - 1), \tag{29}$$

we see that $\delta_i + L_i \overset{i.i.d.}{\sim} \text{Bin}(n, p)$. Let $\mathcal{L}_t = \sum_{i=T_1+1}^{T_1+t} L_i$. Then for any $f : \mathbb{N} \to \mathbb{R}$,

$$\mathbb{P}(R_t > 0 \ \forall t \in [\![T_1, T_2]\!] \mid R_{T_1} = H) = \mathbb{P}(R_{T_1+t} - R_{T_1} > -H \ \forall t \in [T])$$
$$= \mathbb{P}(S_t - \mathcal{L}_t > -H \ \forall t \in [T])$$
$$\geq \mathbb{P}(S_t > \mathcal{L}_t - H \ \forall t \in [T], \ \mathcal{L}_t \leq f(t) + H \ \forall t \in [T])$$
$$\geq \mathbb{P}(S_t > f(t) \ \forall t \in [T]) - \mathbb{P}(\exists t \in [T] : \mathcal{L}_t > f(t) + H). \tag{30}$$

We now let $f(t) = \mathbb{E}[\mathcal{L}_t] + \frac{A^{1/2}}{n^{5/12}}(T_1 + t) - H$ and aim to show that

$$\mathbb{P}(\exists t \in [T] : \mathcal{L}_t > f(t) + H) \leq T e^{-cAn^{1/6}}. \tag{31}$$

Indeed, a union bound gives

$$\mathbb{P}(\exists t \in [T] : \mathcal{L}_t > f(t) + H) \leq \sum_{t=1}^{T} \mathbb{P}\left(\mathcal{L}_t \geq \mathbb{E}[\mathcal{L}_t] + \frac{A^{1/2}}{n^{5/12}}(T_1 + t)\right),$$

and then applying Lemma 3.5 yields

$$\mathbb{P}(\exists t \in [T] : \mathcal{L}_t > f(t) + H) \leq \sum_{t=1}^{T} \exp\left(-\frac{A(T_1 + t)^2 n^{-5/6}}{2(K + T_1 + 1/2)tp + t^2 p + \frac{2A^{1/2}(T_1+t)}{3n^{5/12}}}\right). \tag{32}$$

One may easily check that for some finite constant C, we have

$$2(K + T_1 + 1/2)tp \leq CT_1 t/n \leq C(T_1 + t)^2/n,$$
$$t^2 p \leq Ct^2/n \leq C(T_1 + t)^2/n$$

and

$$\frac{2A^{1/2}(T_1 + t)}{3n^{5/12}} \leq C\frac{T_1}{n}(T_1 + t) \leq C(T_1 + t)^2/n.$$

Thus, for some $c > 0$,

$$\exp\left(-\frac{A(T_1 + t)^2 n^{-5/6}}{2(K + T_1 + 1/2)tp + t^2 p + \frac{2A^{1/2}(T_1+t)}{3n^{5/12}}}\right) \leq \exp\left(-cAn^{1/6}\right)$$

which combines with (32) to give (31).

Substituting (31) into (30) proves the Lemma with f in place of g. It therefore suffices to check that $f(t) \leq g(t)$ for all t when n is large. This holds since

$$\mathcal{L}_t \sim \text{Bin}\left(\sum_{i=T_1+1}^{T_1+t} (K + i), p\right) = \text{Bin}((K + T_1)t + t(t+1)/2, p),$$

and we have

$$p(K + T_1 + 1/2)t \le \frac{8t}{n^{1/3}A^2}, \quad A^{1/2}n^{-5/12}T_1 \le \frac{n^{1/3}}{2A} \quad \text{and} \quad A^{1/2}n^{-5/12}t \le \frac{t}{n^{1/3}A^2}.$$

These estimates show that $f(t) \le g(t)$ and complete the proof. □

4.4. *Proofs of Lemmas 4.6 and 4.8: Poisson approximation and a change of measure*

The proof of Lemma 4.6 uses two standard ingredients: a coupling between Binomial and Poisson random variables, and a change of measure to remove the drift from a random walk.

Proof of Lemma 4.6. Note that $\mathbb{E}[W_i] = (1 + a_n)(1 + b_n) = \mu_n$ for each i. By [32, Theorem 2.10] we can construct a coupling between $(W_i)_{i \in \mathbb{N}}$ and a sequence W_i' of i.i.d. Poisson random variables with parameter μ_n, such that

$$\mathbb{P}(W_i \ne W_i') \le \sum_{i=1}^{n(1+a_n)} \left(\frac{1 + b_n}{n}\right)^2 = \frac{(1 + a_n)(1 + b_n)^2}{n}.$$

Let $M_t' = 1 + \sum_{i=1}^{t}(W_i' - 1)$. Then

$$\mathbb{P}(M_t > 0 \quad \forall t \in [t_n], \, M_{t_n} \in [h_n, 2h_n])$$
$$\ge \mathbb{P}(M_t' > 0 \quad \forall t \in [t_n], \, M_{t_n}' \in [h_n, 2h_n]) - \mathbb{P}(\exists i \in [t_n] : W_i \ne W_i'), \quad (33)$$

and a union bound gives that

$$\mathbb{P}(\exists i \in [t_n] : W_i \ne W_i') \le t_n \frac{(1 + a_n)(1 + b_n)^2}{n}. \quad (34)$$

We now seek to remove the drift from the sequence M_t' by using a change of measure. Define a new probability measure $\mathbb{Q}$ by setting, for $B \in \sigma(W_1', \ldots, W_{t_n}')$,

$$\mathbb{Q}(B) = \mathbb{E}\left[\mathbb{1}_B \prod_{i=1}^{t_n} \mu_n^{-W_i'}\right] \mathbb{E}\left[\mu_n^{-W_1'}\right]^{-t_n} = \mathbb{E}\left[\mathbb{1}_B \mu_n^{-\sum_{i=1}^{t_n} W_i'}\right] \mathbb{E}\left[\mu_n^{-W_1'}\right]^{-t_n}.$$

Note that $\sum_{i=1}^{t_n} W_i' = M_{t_n}' + t_n - 1$ and, since W_1' is Poisson with parameter μ_n, $\mathbb{E}[\mu_n^{-W_1'}] = e^{\mu_n(1/\mu_n - 1)} = e^{1 - \mu_n}$. Thus

$$\mathbb{Q}(B) = \mathbb{E}\left[\mathbb{1}_B \mu_n^{-M_{t_n}' - t_n + 1}\right] e^{(\mu_n - 1)t_n}. \quad (35)$$

We write $\mathbb{E}_{\mathbb{Q}}$ for expectation with respect to the probability measure $\mathbb{Q}$. It is straightforward to check that, under $\mathbb{Q}$, the W_i' are independent Poisson random variables of parameter 1.

By the definition of $\mathbb{Q}$, we have

$$\mathbb{P}(M_t' > 0 \ \forall t \in [\![0, t_n]\!], \, M_{t_n}' \in [h_n, 2h_n])$$
$$= \mathbb{E}_{\mathbb{Q}}\left[\mu_n^{M_{t_n}' + t_n - 1} \mathbb{1}_{\{M_t' > 0 \ \forall t \in [\![0, t_n]\!], \, M_{t_n}' \in [h_n, 2h_n]\}}\right] e^{(1 - \mu_n)t_n}$$
$$\ge (\mu_n \wedge 1)^{2h_n} \mu_n^{t_n - 1} \mathbb{Q}(M_t' > 0 \ \forall t \in [\![0, t_n]\!], \, M_{t_n}' \in [h_n, 2h_n]) e^{(1 - \mu_n)t_n}.$$

Since $(W_i')_{i=1}^{t_n}$ is a sequence of independent Poisson random variables of parameter 1 under $\mathbb{Q}$, substituting this and (34) into (33) gives the result. □

The proof of Lemma 4.8 is similar to that of Lemma 4.6, but we will need to delve deeper into the details of the coupling between the Binomial and Poisson random variables.

Proof of Lemma 4.8. We follow almost the same proof as Lemma 4.6, noting that $\mathbb{E}[\Delta_i] = np$ for each i. By [32, Theorem 2.10] we can couple $(\Delta_i)_{i=1}^T$ with a sequence $(\Delta_i')_{i=1}^T$ of i.i.d. Poisson random variables with parameter np. Write $S_t' = \sum_{i=1}^t (\Delta_i' - 1)$. Then we obtain

$$\mathbb{P}\big(S_t > g(t) \quad \forall t \in [\![0, T]\!]\big)$$

$$\geq \mathbb{P}\big(S_t > g(t) \ \forall t \in [\![0, T]\!], \ S_t' > g(t) + \tfrac{n^{1/3}}{4A} \ \forall t \in [\![0, T]\!]\big)$$

$$= \mathbb{P}\big(S_t' > g(t) + \tfrac{n^{1/3}}{4A} \quad \forall t \in [\![0, T]\!]\big)$$

$$- \mathbb{P}\big(S_t' > g(t) + \tfrac{n^{1/3}}{4A} \quad \forall t \in [\![0, T]\!], \ \exists t \in [\![0, T]\!] : S_t \leq g(t)\big)$$

$$\geq \mathbb{P}\big(S_t' > g(t) + \tfrac{n^{1/3}}{4A} \quad \forall t \in [\![0, T]\!]\big) - \mathbb{P}\big(\max_{t \leq T} |S_t - S_t'| > \tfrac{n^{1/3}}{4A}\big), \tag{36}$$

where the last inequality follows from the fact that

$$\mathbb{P}\big(S_t' > g(t) + \tfrac{n^{1/3}}{4A} \quad \forall t \in [\![0, T]\!], \ \exists t \in [\![0, T]\!] : S_t \leq g(t)\big)$$

$$\leq \mathbb{P}\big(\exists t \in [\![0, T]\!] : |S_t' - S_t| > \tfrac{n^{1/3}}{4A}\big) = \mathbb{P}\big(\max_{t \leq T} |S_t - S_t'| > \tfrac{n^{1/3}}{4A}\big).$$

To estimate the last probability, we see that

$$\mathbb{P}\big(\max_{t \leq T} |S_t - S_t'| > \tfrac{n^{1/3}}{4A}\big) \leq \mathbb{P}\big(\sum_{i=1}^T |\Delta_i - \Delta_i'| > \tfrac{n^{1/3}}{4A}\big) \leq \mathbb{E}[e^{|\Delta_1 - \Delta_1'|}]^T e^{-n^{1/3}/(4A)}, \tag{37}$$

where for the last inequality we used the i.i.d. property of the increments $\Delta_i - \Delta_i'$. To continue our bounds we need some more detail about the coupling of Δ_1 and Δ_1' from the proof of [32, Theorem 2.10]. We break Δ_1 up into a sum of n i.i.d. Bernoulli random variables of parameter p, which we call $(\beta_j)_{j=1}^n$, and couple these with n Poisson random variables $(\beta_j')_{j=1}^n$ of parameter p, so that

$$\Delta_1 = \sum_{j=1}^n \beta_j \text{ and } \Delta_1' = \sum_{j=1}^n \beta_j'.$$

The coupling is arranged so that for each i and j,

- $\mathbb{P}(\beta_j = \beta_j' = 0) = 1 - p$,
- $\mathbb{P}(\beta_j = \beta_j' = 1) = pe^{-p}$,
- $\mathbb{P}(\beta_j = 1, \ \beta_j' = 0) = e^{-p} - (1 - p)$ and
- $\mathbb{P}(\beta_j = 1, \ \beta_j' = k) = \mathbb{P}(\beta_j' = k) = \tfrac{e^{-p} p^k}{k!}$ for $k \geq 2$.

We deduce (using the inequality $e^{-x} \leq 1 - x + x^2/2$, valid for all $x \geq 0$) that

$$\mathbb{E}[e^{|\beta_j - \beta_j'|}] = 1 - p + pe^{-p} + e(e^{-p} - (1 - p)) + \sum_{k=2}^\infty e^{k-1} \frac{e^{-p} p^k}{k!}$$

$$\leq 1 + e\frac{p^2}{2} + p^2 e e^{-p} \sum_{k=0}^\infty \frac{(ep)^k}{(k+2)!} \leq 1 + cp^2$$

for some finite constant c. Thus

$$\mathbb{E}[e^{|\Delta_1 - \Delta_1'|}] \leq (1 + cp^2)^n \leq \exp(cp^2 n),$$

and substituting this into (37) gives

$$\mathbb{P}\left(\max_{t \leq T} |S_t - S'_t| > \frac{n^{1/3}}{4A}\right) \leq \exp\left(cp^2 nT - \frac{n^{1/3}}{4A}\right) \leq C \exp\left(-\frac{n^{1/3}}{4A}\right) \tag{38}$$

for some finite constant C.

We now consider the first quantity on the right-hand side of (36) and use the same change of measure as in (35) with $\mu_n = pn$ to remove the drift from S'_t. Noting that $g(t) + \frac{n^{1/3}}{4A} = \gamma(t)$, by the definition of $\mathbb{Q}$, for any $\ell \geq 0$,

$$\begin{aligned}
&\mathbb{P}\left(S'_t > g(t) + \tfrac{n^{1/3}}{4A} \quad \forall t \in [\![0, T]\!]\right) \\
&\geq \mathbb{P}\left(S'_t > \gamma(t) \quad \forall t \in [\![0, T]\!], \ S'_T \leq \gamma(t) + \ell\right) \\
&= \mathbb{E}_{\mathbb{Q}}\left[(pn)^{S'_T + T} \mathbb{1}_{\{S'_t > \gamma(t) \ \forall t \in [\![0, T]\!], \ S'_T \leq \gamma(t) + \ell\}}\right] e^{(1-pn)T} \\
&\geq (pn \wedge 1)^\ell (pn)^{\gamma(T) + T} e^{(1-pn)T} \mathbb{Q}\left(S'_t > \gamma(t) \quad \forall t \in [\![0, T]\!], \ S'_T \leq \gamma(t) + \ell\right). \tag{39}
\end{aligned}$$

Taking $\ell = \frac{3n^{1/3}}{8A}$ and recalling that $pn = 1 + \lambda n^{-1/3}$ and

$$\gamma(T) = -\frac{n^{1/3}}{4A} + \frac{9T}{A^2 n^{1/3}} + \frac{pT^2}{2} = \frac{A^2 n^{1/3}}{2} + O\left(\frac{n^{1/3}}{A}\right),$$

and using that $|\lambda| \leq A/3$, $|A| = o(n^{1/30})$ and $1 + x \geq e^{x - x^2}$ for all $x > -1/2$ we have

$$(pn \wedge 1)^\ell \geq c$$

and

$$(pn)^{\gamma(T)} = (1 + \lambda n^{-1/3})^{A^2 n^{1/3}/2 + O(n^{1/3}/A)} \geq c e^{\lambda A^2/2},$$

and using the more precise inequality $1 + x \geq e^{x - x^2/2 - |x|^3}$ for all $x > -1/2$,

$$(pn)^T e^{(1-pn)T} = ((1 + \lambda n^{-1/3}) e^{-\lambda n^{-1/3}})^T \geq e^{-\lambda^2 n^{-2/3} T/2 - O(\lambda^3 T/n)} \geq c e^{-\lambda^2 A/2}.$$

Substituting these estimates into (39), we obtain

$$\begin{aligned}
&\mathbb{P}\left(S'_t > g(t) + \tfrac{n^{1/3}}{4A} \quad \forall t \in [\![0, T]\!]\right) \\
&\geq c e^{\lambda A^2/2 - \lambda^2 A/2} \mathbb{Q}\left(S'_t > \gamma(T) \quad \forall t \in [\![0, T]\!], \ S'_T \leq \gamma(T) + \tfrac{3n^{1/3}}{8A}\right).
\end{aligned}$$

Since $(S'_t)_{t=1}^T$ is a sum of independent Poisson random variables of parameter 1 under $\mathbb{Q}$, substituting this and (38) into (36) gives the result. $\qquad\square$

4.5. The probability a Brownian motion stays above a curve: proof of Proposition 4.11

Recall that

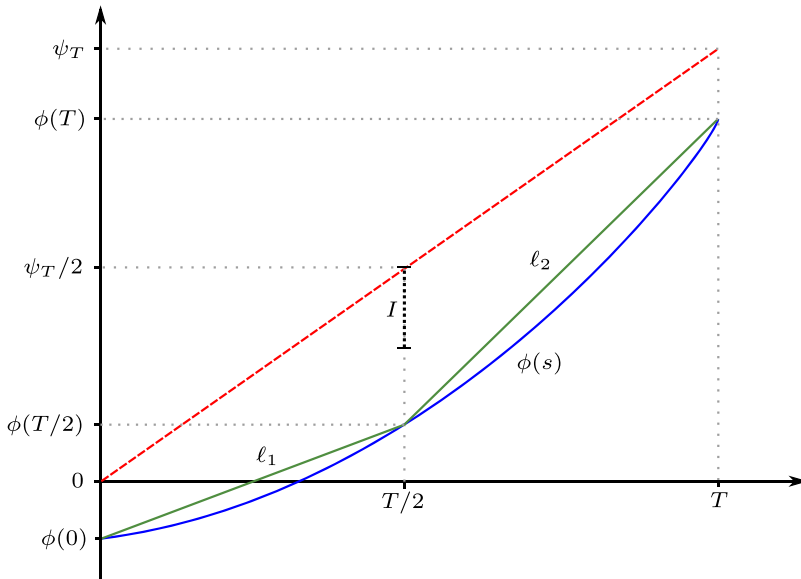$$\gamma(s) = -\frac{n^{1/3}}{4A} + \frac{9s}{A^2 n^{1/3}} + \frac{ps^2}{2}$$

and write

$$P_n(T) = \mathbb{P}\left(B_s > \gamma(s) + M \log T + x_n \ \forall s \in [0, T], \ B_T \leq \gamma(T) + \tfrac{3n^{1/3}}{8A} - M \log T - x\right);$$

our aim in this section is to bound $P_n(T)$ from below.

Define

$$\phi(s) = \gamma(s) + \frac{n^{1/3}}{8A} = -\frac{n^{1/3}}{8A} + \frac{9s}{A^2 n^{1/3}} + \frac{ps^2}{2}$$

Figure 1. We want our Brownian motion to stay above the blue curve, and the two green lines ℓ_1 and ℓ_2 show linear approximations to this curve on the two half-intervals. The dashed red line shows roughly where we expect our Brownian motion to be, given that it stays above the curve. This is a caricature of the true picture, and not to scale.

and

$$\psi_T = \gamma(T) + \frac{n^{1/3}}{4A} = \frac{9T}{A^2 n^{1/3}} + \frac{pT^2}{2}.$$

Note that since $x_n \ll n^{1/3}/A$, for large n we have $M \log T + x_n \le n^{1/3}/(8A)$, so

$$P_n(T) \ge \mathbb{P}\big(B_s > \phi(s) \ \forall s \in [0, T], \ B_T \le \psi_T\big). \tag{40}$$

We approximate the curve $\phi(s)$ given above with two straight lines defined, for $s \in [0, T/2]$, by

$$\ell_1(s) = \phi(0) + \Big(\frac{\phi(T/2) - \phi(0)}{T/2}\Big)s = -\frac{n^{1/3}}{8A} + \Big(\frac{9}{A^2 n^{1/3}} + \frac{pT}{4}\Big)s$$

and

$$\ell_2(s) = \phi(T/2) + \Big(\frac{\phi(T) - \phi(T/2)}{T/2}\Big)s = -\frac{n^{1/3}}{8A} + \frac{9T}{2A^2 n^{1/3}} + \frac{pT^2}{8} + \Big(\frac{9}{A^2 n^{1/3}} + \frac{3pT}{4}\Big)s.$$

Also define

$$I = \Big[\psi_T/2 - A^{1/2} n^{1/3}, \ \psi_T/2\Big]$$
$$= \Big[\frac{9T}{2A^2 n^{1/3}} + \frac{pT^2}{4} - A^{1/2} n^{1/3}, \ \frac{9T}{2A^2 n^{1/3}} + \frac{pT^2}{4}\Big].$$

See Figure 1 for reference. Note that $\psi_T/2 - A^{1/2} n^{1/3} > \phi(T/2)$ when n is large, so the interval I falls entirely above the curve ϕ.

Since ϕ is convex, the linear interpolations ℓ_1 and ℓ_2 fall above the curve and therefore (40) is at least

$$\mathbb{P}\big(B_s > \ell_1(s) \ \forall s \in [0, T/2], \ B_s > \ell_2(s - T/2) \ \forall s \in [T/2, T], \ B_T \le \psi_T\big).$$

For a lower bound, we may also insist that at time $T/2$ our Brownian motion falls within the interval I. Moreover, it will be convenient for us to know that at time T, our Brownian motion is

some distance above the curve. To this end define

$$I' := \left[\psi_T - \frac{n^{1/3}}{16A}, \psi_T \right] = \left[l_2(T/2) + \frac{n^{1/3}}{16A}, l_2(T/2) + \frac{n^{1/3}}{8A} \right].$$

Putting all this together we obtain

$$P_n(T) \geq \int_I \mathbb{P}\big(B_s > \ell_1(s) \ \forall s \in [0, T/2], \ B_{T/2} \in \mathrm{d}w\big)$$
$$\cdot \mathbb{P}_w\big(B_s > \ell_2(s) \ \forall s \in [0, T/2], \ B_{T/2} \in I'\big). \tag{41}$$

Here $\mathbb{P}_w$ denotes a probability measure under which our Brownian motion starts from w rather than 0.

Lemma 4.12 *For any $\mu, y \in \mathbb{R}$, $t > 0$, $x > y$ and $z > y + \mu t$,*

$$\mathbb{P}_x(B_s > y + \mu s \ \forall s \leq t, \ B_t \in \mathrm{d}z)$$
$$= \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(z-x)^2}{2t} \right)\left(1 - \exp\left(\frac{2(x-y)(\mu t + y - z)}{t} \right) \right) \mathrm{d}z.$$

Proof. We begin with an exponential change of measure to balance the drift μ. Letting $(\mathcal{F}_s)_{s \geq 0}$ be the natural filtration of our Brownian motion, define $\mathcal{P}_x$, with expectation operator $\mathcal{E}_x$, by setting

$$\left. \frac{\mathrm{d}\mathcal{P}_x}{\mathrm{d}\mathbb{P}_x} \right|_{\mathcal{F}_t} = e^{\mu B_t - \mu x - \mu^2 t/2}.$$

Then under $\mathcal{P}_x$, $(B_s)_{s \geq 0}$ is a Brownian motion with drift μ started from x, and therefore

$$\mathbb{P}_x(B_s > y + \mu s \ \forall s \leq t, \ B_t \in \mathrm{d}z) = \mathcal{E}_x\big[e^{-\mu B_t + \mu^2 t/2 + \mu x} \mathbb{1}_{\{B_s > y + \mu s \ \forall s \leq t, \ B_t \in \mathrm{d}z\}} \big]$$
$$= e^{-\mu z + \mu^2 t/2 + \mu x} \mathcal{P}_x(B_s > y + \mu s \ \forall s \leq t, \ B_t \in \mathrm{d}z)$$
$$= e^{-\mu z + \mu^2 t/2 + \mu x} \mathbb{P}_x(B_s > y \ \forall s \leq t, \ B_t + \mu t \in \mathrm{d}z).$$

We now recall that, as a consequence of the reflection principle for Brownian motion, for $x > y$ and $w > y$,

$$\mathbb{P}_x(B_s > y \ \forall s \leq t, \ B_t \in dw) = \frac{1}{\sqrt{2\pi t}}\left(\exp\left(-\frac{(w-x)^2}{2t} \right) - \exp\left(-\frac{(w+x-2y)^2}{2t} \right) \right) dw.$$

Since the ratio within the second exponential expands as

$$-\frac{(w+x-2y)^2}{2t} = -\frac{(w-x)^2}{2t} + \frac{2(yx - wx + yw - y^2)}{t},$$

grouping by $-(w-x)^2/2t$ we obtain

$$\mathbb{P}_x(B_s > y \ \forall s \leq t, \ B_t \in dw)$$
$$= \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(w-x)^2}{2t} \right)\left(1 - \exp\left(\frac{2(yx - wx + yw - y^2)}{t} \right) \right).$$

Taking $w = z - \mu t$ and substituting into the expression above, and then simplifying, gives the result. $\qquad \square$

We now use Lemma 4.12 to obtain a lower bound for the probability that B_t stays above the line $l_1(s)$ and finishes near $w \in I$ at time $T/2$.

Corollary 4.13 *For $w \in I$,*

$$\mathbb{P}\big(B_s > \ell_1(s) \ \forall s \in [0, T/2], \ B_{T/2} \in \, dw\big) \geq \frac{c}{\sqrt{T}} e^{-w^2/T} \, dw.$$

Proof. We apply Lemma 4.12 with $x = 0$, $z = w$, $y = -n^{1/3}/(8A)$, $\mu = \frac{9}{A^2 n^{1/3}} + \frac{pT}{4}$ and $t = T/2$. With these parameters, $w > y + \mu t$ and hence Lemma 4.12 tells us that

$$\mathbb{P}\big(B_s > \ell_1(s) \ \forall s \in [0, T/2], \ B_{T/2} \in \, dw\big)$$
$$= \frac{1}{\sqrt{\pi T}} e^{-w^2/T} \Big(1 - \exp\Big(-2\Big(w + \frac{n^{1/3}}{8A} - \frac{9T}{2A^2 n^{1/3}} - \frac{pT^2}{8}\Big)\frac{n^{1/3}}{4AT}\Big)\Big) \, dw.$$

Since $w \in I$, we have $w \geq \frac{9T}{2A^2 n^{1/3}} + \frac{pT^2}{4} - A^{1/2} n^{1/3}$ and therefore

$$\Big(w + \frac{n^{1/3}}{8A} - \frac{9T}{2A^2 n^{1/3}} - \frac{pT^2}{8}\Big)\frac{n^{1/3}}{4AT} \geq \Big(\frac{pT^2}{8} - A^{1/2} n^{1/3}\Big)\frac{n^{1/3}}{4AT} \geq c$$

for some $c > 0$, and the result follows.   □

Next we bound from below the second probability that appears in the integral (41), again by means of Lemma 4.12.

Corollary 4.14 *For $w \in I$ and A sufficiently large,*

$$\mathbb{P}_w\big(B_s > \ell_2(s) \ \forall s \in [0, T/2], \ B_{T/2} \in I'\big) \geq \frac{c}{\sqrt{T}} \int_{I'} e^{-(z-w)^2/T} \, dz.$$

Proof. We now apply Lemma 4.12 with $x = w$, $y = -\frac{n^{1/3}}{8A} + \frac{9T}{2A^2 n^{1/3}} + \frac{pT^2}{8}$, $\mu = \frac{9}{A^2 n^{1/3}} + \frac{3pT}{4}$ and $t = T/2$. This tells us that

$$\mathbb{P}_w\big(B_s > \ell_2(s) \ \forall s \in [0, T/2], \ B_{T/2} \in I'\big)$$
$$= \frac{1}{\sqrt{\pi T}} \int_{I'} e^{-(z-w)^2/T} \Big(1 - \exp\Big(\frac{2(w - y)(\mu t + y - z)}{t}\Big)\Big) \, dz, \tag{42}$$

we want an upper bound for the exponential term

$$\exp\Big(\frac{2(w - y)(\mu t + y - z)}{t}\Big).$$

To this end observe that, since $z \geq \phi(T) + n^{1/3}/16A$ and

$$\phi(T) = -\frac{n^{1/3}}{8A} + \frac{9T}{A^2 n^{1/3}} + \frac{pT^2}{2},$$

then

$$\mu t + y - z = \frac{9T}{A^2 n^{1/3}} + \frac{pT^2}{2} - \frac{n^{1/3}}{8A} - z \leq -\frac{n^{1/3}}{16A}.$$

Also, since $w \in I$ we obtain

$$(w - y)(\mu t + y - z) \leq -\frac{n^{1/3}}{16A}(w - y) \leq -\frac{n^{1/3}}{16A}\Big(\frac{pT^2}{8} + \frac{n^{1/3}}{8A} - A^{1/2} n^{1/3}\Big)$$
$$= -\frac{n^{1/3}}{16A} \frac{pT^2}{8}(1 - o(1)),$$

whence

$$\frac{2(w-y)(\mu t + y - z)}{t} = \frac{4(w-y)(\mu t + y - z)}{T} \leq -\frac{4}{T}\frac{n^{1/3}}{16A}\frac{pT^2}{8}(1 - o(1)) \sim -\frac{1}{32}.$$

Plugging this estimate into (42) we obtain the desired result. ☐

Proof of Proposition 4.11. Substituting Corollaries 4.13 and 4.14 into (41), and recalling the definition of I' we obtain

$$P_n(T) \geq \frac{c}{T} \int_{\psi_T - \frac{n^{1/3}}{16A}}^{\psi_T} \int_{\psi_T/2 - A^{1/2}n^{1/3}}^{\psi_T/2} e^{-w^2/T - (z-w)^2/T} \, dw \, dz. \tag{43}$$

Using the substitutions $u = w - \psi_T/2$ and $v = z - \psi_T$, the term on the right-hand side of the inequality in (43) equals

$$\frac{c}{T} \int_{-\frac{n^{1/3}}{16A}}^{0} \int_{-A^{1/2}n^{1/3}}^{0} e^{-(u+\psi_T/2)^2/T - (v-u+\psi_T/2)^2/T} \, du \, dv$$

which, after multiplying out the quadratic terms in the exponent, becomes

$$\frac{c}{T} \int_{-\frac{n^{1/3}}{16A}}^{0} \int_{-A^{1/2}n^{1/3}}^{0} e^{-2u^2/T + 2uv/T - \psi_T^2/(2T) - v^2/T - v\psi_T/T} \, du \, dv.$$

Since $u, v \leq 0$, we have $2uv/T \geq 0$ and therefore the integral over u is at least

$$\int_{-A^{1/2}n^{1/3}}^{0} e^{-2u^2/T} \, du \geq cA^{1/2}n^{1/3} \geq c\sqrt{T}.$$

We deduce that

$$P_n(T) \geq \frac{c}{\sqrt{T}} \int_{-\frac{n^{1/3}}{16A}}^{0} e^{-\psi_T^2/(2T) - v^2/T - v\psi_T/T} \, dv. \tag{44}$$

Now

$$\frac{\psi_T}{T} = \frac{9}{A^2 n^{1/3}} + \frac{pT}{2} = \frac{A}{2n^{1/3}} + O\left(\frac{1}{A^2 n^{1/3}}\right),$$

so the exponent on the right-hand side of (44) is $e^{-\psi_T^2/2T - O(1)}$; thus (44) becomes

$$P_n(T) \geq \frac{c}{\sqrt{T}} \cdot \frac{n^{1/3}}{16A} e^{-\psi_T^2/(2T)} \geq \frac{c}{A^{3/2}} e^{-\psi_T^2/(2T)}.$$

It then remains only to note that

$$\frac{\psi_T^2}{2T} = \frac{1}{2T}\left(\frac{9T}{A^2 n^{1/3}} + \frac{pT}{2}\right)^2 = \frac{81T}{2A^2 n^{1/3}} + \frac{9pT^2}{2A^2 n^{1/3}} + \frac{p^2 T^3}{8} = \frac{A^3}{8} + O(1),$$

and the proof is complete. ☐

Remark 4.15 Here we continue the discussion started with Remark 3.7, where we highlighted at which steps within the proof of the upper bounds stated in Theorem 1.1 we saw the appearance of constant factors in front of the probability $\mathbb{P}(|\mathcal{C}_{\max}| > An^{2/3})$. For the lower bound, the problem of finding the right constant is similarly involved. For instance, with Proposition 4.3 we see the appearance of a constant which comes not only from some involved computations, which similarly to Remark 3.7 could possibly be tightened subject to further conditions on A and λ, but also from an application of Theorem 2.1 (the generalised ballot theorem). Therefore, in order to properly identify this constant, one would need to either go through the proof of the generalised ballot theorem in our setting and check whether it would be possible to keep track of all the constants

appearing there; or to apply other results on random walks, for example the probability for a random walk to stay positive [30] together with a convergence in law of random walks conditioned to stay positive, e.g. from [14].

Acknowledgements

References

[1] Addario-Berry, L., Broutin, N. and Goldschmidt, C. (2012) The continuum limit of critical random graphs. *Probab. Theory Relat. Fields* **152**(3–4) 367–406.

[2] Addario-Berry, L. and Reed, B. A. (2008) Ballot theorems, old and new. In *Horizons of Combinatorics*. Springer, pp. 9–35.

[3] Aldous, D. (1997) Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.* **25**(2) 812–854.

[4] Andreis, L., König, W. and Patterson, R. I. A. (2021) A large-deviations principle for all the cluster sizes of a sparse Erdös-Rényi graph. *Random Struct. Alg.* **59** 522–553.

[5] Bhamidi, S., van der Hofstad, R. and van Leeuwaarden, J. S. H. (2012) Novel scaling limits for critical inhomogeneous random graphs. *Ann. Probab.* **40**(6) 2299–2361.

[6] Bhamidi, S., van der Hofstad, R. and van Leeuwaarden, J. S. H. (2010) Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electron. J. Probab.* **15** 1682–1702.

[7] Bollobás, B. (2001) *Random Graphs*, 2nd ed., Vol. 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

[8] Bollobás, B. and Riordan, O. (2012) Asymptotic normality of the size of the giant component via a random walk. *J. Comb. Theory Ser. B* **102**(1) 53–61.

[9] Chatterjee, S. (2012) A new approach to strong embeddings. *Probab. Theory Relat. Fields* **152**(1–2) 231–264.

[10] De Ambroggio, U. (2021) *An elementary approach to component sizes in critical random graphs*. Preprint: https://arxiv.org/abs/2101.06625.

[11] De Ambroggio, U. and Pachon, A. (2020) *Simple upper bounds for the largest components in critical inhomogeneous random graphs*. Preprint: http://arxiv.org/abs/2012.09001.

[12] Dembo, A., Levit, A. and Vadlamani, S. (2019) Component sizes for large quantum Erdös-Rényi graph near criticality. *Ann. Probab.* **47**(2) 1185–1219.

[13] Dhara, S., van der Hofstad, R., van Leeuwaarden, J. S. H. and Sen, S. (2017) Critical window for the configuration model: finite third moment degrees. *Electron. J. Probab.* **22**(16) 1–33.

[14] Durrett, R. (1978) Conditioned limit theorems for some null-recurrent Markov processes. *Ann. Probab.* **6**(5) 798–828.

[15] Janson, S., Luczak, T. and Rucinski, A. (2011) *Random Graphs*, Vol. **45**. John Wiley & Sons, New York.

[16] Joseph, A. (2014) The component sizes of a critical random graph with given degree sequence. *Ann. Appl. Probab.* **24**(6) 2560–2594.

[17] Kager, W. (2011) The hitting time theorem revisited. *Am. Math. Mon.* **118**(8) 735–737.

[18] Komlós, J., Major, P. and Tusnády, G. (1975) An approximation of partial sums of independent RV'-s, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **32**(1–2) 111–131.

[19] Konstantopoulos, T. (1995) Ballot theorems revisited. *Stat. Probab. Lett.* **24**(4) 331–338.

[20] Łuczak, T., Pittel, B. and Wierman, J. C. (1994) The structure of a random graph at the point of the phase transition. *Trans. Am. Math. Soc.* **341**(2) 721–748.

[21] Martin-Löf, A. (1986) Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *J. Appl. Probab.* **23**(2) 265–282.

[22] Nachmias, A. and Peres, Y. (2007) Component sizes of the random graph outside the scaling window. *ALEA Latin Am. J. Probab. Math. Stat.* **3** 133–142.

[23] Nachmias, A. and Peres, Y. (2010) Critical percolation on random regular graphs. *Random Struct. Alg.* **36**(2) 111–148.

[24] Nachmias, A. and Peres, Y. (2010) The critical random graph, with martingales. *Israel J. Math.* **176** 29–41.

[25] O'Connell, N. (1998) Some large deviation results for sparse random graphs. *Probab. Theory Related Fields* **110**(3) 277–285.

[26] Pittel, B. (2001) On the largest component of the random graph at a nearcritical stage. *J. Comb. Theory Ser. B* **82**(2) 237–269.

[27] Riordan, O. (2012) The phase transition in the configuration model. *Comb. Probab. Comput.* **21**(1–2) 265–299.

[28] Roberts, M. I. (2017) The probability of unusually large components in the near-critical Erdös-Rényi graph. *Adv. Appl. Probab.* **50**(1) 245–271.

[29] Rossignol, R. (2021) Scaling limit of dynamical percolation on critical Erdös-Rényi random graphs. *Ann. Probab.* **49**(1) 322–399.

[30] Spitzer, F. (1960) A Tauberian theorem and its probability interpretation. *Trans. Am. Math. Soc.* **94**(1) 150–179.

[31] Strassen, V. (1967) Almost sure behavior of sums of independent random variables and martingales. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume* **2:** *Contributions to Probability Theory, Part 1*. The Regents of the University of California.

[32] van der Hofstad, R. (2016) *Random Graphs and Complex Networks*, Vol. **1**. Cambridge University Press, Cambridge.

[33] van der Hofstad, R., Janssen, A. J. E. M. and van Leeuwaarden, J. S. H. (2010) Critical epidemics, random graphs, and Brownian motion with a parabolic drift. *Adv. Appl. Probab.* **42**(4) 1187–1206.

[34] van der Hofstad, R., Kager, W. and Müller, T. (2009) A local limit theorem for the critical random graph. *Electron. Commun. Probab.* **14** 122–131.

[35] van der Hofstad, R. and Keane, M. (2008) An elementary proof of the hitting time theorem. *Am. Math. Mon.* **115**(8) 753–756.

[36] van der Hofstad, R., Kliem, S. and van Leeuwaarden, J. S. H. (2018) Cluster tails for critical power-law inhomogeneous random graphs. *J. Stat. Phys.* **171**(1) 38–95.