Critical window for the configuration model: finite third moment degrees

Souvik Dhara, Remco van der Hofstad, Johan S.H. van Leeuwaarden and Sanchayan Sen

Department of Mathematics and Computer Science, Eindhoven University of Technology

May 11, 2016

Abstract

We investigate the component sizes of the critical configuration model, as well as the related problem of critical percolation on a supercritical configuration model. We show that, at criticality, the finite third moment assumption on the asymptotic degree distribution is enough to guarantee that the component sizes are $O(n^{2/3})$ and the re-scaled component sizes converge to the excursions of an inhomogeneous Brownian Motion with a parabolic drift. We use percolation to study the evolution of these component sizes while passing through the critical window and show that the vector of percolation cluster-sizes, considered as a process in the critical window, converge to the multiplicative coalescent process in finite dimensions. This behavior was first observed for Erdős-Rényi random graphs by Aldous (1997) and our results provide support for the empirical evidences that the nature of the phase transition for a wide array of random-graph models are universal in nature. Further, we show that the re-scaled component sizes and surplus edges converge jointly under a strong topology, at each fixed location of the scaling window.

1 Introduction

Random graphs are the main vehicles to study complex networks that go through a radical change in their connectivity, often called the phase-transition. A large body of literature aims at understanding the properties of random graphs that experience this phase-transition in the sizes of the large connected components for various models. The behavior is well understood for the Erdős-Rényi random graphs, thanks to a plethora of results [2, 15]. However, these graphs are often inadequate for modeling real-world networks since the real-world network data often show a power-law behavior of the asymptotic degrees whereas the degree distribution of the Erdős-Rényi random graphs has exponentially decaying tails. See [10, 22] for interesting discussions about internet topology. Therefore, many alternative models have been proposed to capture this power-law tail behavior. An interesting fact, however, is that the behavior, in most of these models, is quite universal in the sense that there is a critical value where the graphs experience a phase-transition and the nature of this phase-transition is insensitive to the microscopic descriptions of the model.
In this work, we focus on the configuration model, the canonical model for generating a random multi-graph with a prescribed degree sequence. This model was introduced by Bollobás [8] to choose a uniform simple $d$-regular graph on $n$ vertices, when $dn$ is even. The idea was later generalized for general degree sequences $d$ by Molloy and Reed [19] and others. We denote by $\text{CM}_n(d)$ the multi-graph generated by the configuration model on the vertex set $[n] = \{1, 2, \ldots, n\}$ with the degree sequence $d$. The configuration model, conditioned on simplicity, yields a uniform simple graph with the same degree sequence. Various features related to the emergence of the giant component phenomenon for this model have been studied recently [11, 12, 14, 16, 19, 21].

We shall give a brief overview of the relevant literature in Section 4.1. Our aim is to obtain precise asymptotics for the component sizes of $\text{CM}_n(d)$ in the critical window of the phase transition under the optimal assumptions on the degree sequence involving a finite third-moment condition. The scaled limiting vector of component sizes is shown to be distributed as the excursions of certain reflected inhomogeneous Brownian motions with a parabolic drift. This shows that $\text{CM}_n(d)$, for a large collection of possible $d$, has similar component sizes as the Erdős-Rényi random graphs [2] and the inhomogeneous random graphs [6]. We use percolation on a super-critical configuration model to show the joint convergence of the scaled vectors of component sizes at multiple locations of the percolation scaling window. We also obtain the asymptotic distribution of the number of surplus edges in each component and show that the sequence of vectors consisting of the re-scaled component sizes and surplus converges to a suitable limit under a strong topology as discussed in [5]. These results give very strong evidence in favor of the structural similarity of the component sizes of $\text{CM}_n(d)$ and Erdős-Rényi random graphs at criticality.

Our contribution

The main innovation of this paper is that we derive the strongest results in the literature under only finite third-moment assumption on the degrees. This finite third-moment assumption is also believed to be necessary for Erdős-Rényi type scaling limits, amongst others since the third moment appears in the limiting random variable. In work in progress [9], we consider the infinite third-moment case where the degrees follow a power-law and show that the scaling limit of the cluster sizes is quite different. The joint convergence of the component sizes and the surplus edges has not been proved under such a strong topology for the configuration model, which makes our results the most general in the existing literature. We also study percolation on the configuration model to gain insight about the evolution of the configuration model over the critical scaling window. This is achieved by studying a dynamic process that generates the percolated graphs with different values of the percolation parameter, a problem that is interesting in its own right.

Before stating our main results, we need to introduce some notation and concepts.

2 Definitions and notation

We shall use the standard notation $\mathbb{P} \to$, $\mathcal{L} \to$ to denote convergence in probability and in distribution or law, respectively. The topology needed for the distributional convergence will always be specified unless it is clear from the context. A sequence of events $(\mathcal{E}_n)_{n \geq 1}$ is said to occur with high probability (whp) with respect to probability measures $(\mathbb{P}_n)_{n \geq 1}$ if $\mathbb{P}_n(\mathcal{E}_n) \to 1$. Denote $f_n = O_p(g_n)$ if $(|f_n|/|g_n|)_{n \geq 1}$ is tight; $f_n = o_p(g_n)$ if $f_n/g_n \mathbb{P} \to 0$; $f_n = \Theta_p(g_n)$ if $f_n = O_p(g_n)$ and $g_n = O_p(f_n)$. We write $f_n = O_E(a_n)$ (respectively $f_n = o_E(a_n)$) to denote that $\sup_{n \geq 1} \mathbb{E} [a_n^{-1} f_n] < \infty$ (respectively $\lim_{n \to \infty} \mathbb{E} [a_n^{-1} f_n] = 0$). Denote by

$$\ell_1^2 := \{ \mathbf{x} = (x_1, x_2, x_3, \ldots) : x_1 \geq x_2 \geq x_3 \geq \ldots \text{ and } \sum_{i=1}^{\infty} x_i^2 < \infty \},$$

(2.1)
the subspace of non-negative, non-increasing sequences of real numbers with square norm metric 
\( d(x, y) = (\sum_{i=1}^{\infty} (x_i - y_i)^2)^{1/2} \) and let \((\ell^2)^k\) denote the \( k \)-fold product space of \( \ell^2 \). With \((\ell^2)^2 \times \mathbb{N}^\infty\), we 
declare the product topology of \( \ell^2 \) and \( \mathbb{N}^\infty \), where \( \mathbb{N}^\infty \) denotes the collection of the sequences on \( \mathbb{N} \), endowed with the product topology. Define also

\[
U_i := \{ ((x_i, y_i))_{i=1}^{\infty} \in (\ell^2) \times \mathbb{N}^\infty : \sum_{i=1}^{\infty} x_i y_i < \infty \text{ and } y_i = 0 \text{ whenever } x_i = 0 \forall i \} \tag{2.2}
\]

with the metric

\[
d_U((x_1, y_1), (x_2, y_2)) := \left( \sum_{i=1}^{\infty} (x_{1i} - x_{2i})^2 \right)^{1/2} + \sum_{i=1}^{\infty} |x_{1i} y_{1i} - x_{2i} y_{2i}|. \tag{2.3}
\]

Further, we introduce \( U_i^0 \subset U_i \) as

\[
U_i^0 := \{ ((x_i, y_i))_{i=1}^{\infty} \in U_i : \text{ if } x_k = x_m, k \leq m, \text{ then } y_k \geq y_m \}. \tag{2.4}
\]

We usually use the boldface notation \( \mathbf{X} \) for a time-dependent stochastic process \( (X(s))_{s \geq 0} \). Unless stated otherwise, \( \mathbb{C}[0, t] \) denotes the set of all continuous functions from \([0, t]\) to \( \mathbb{R} \) equipped with the topology induced by sup-norm \( | \cdot | \). Similarly, \( \mathbb{D}[0, t] \) (resp. \( \mathbb{D}[0, \infty) \)) denotes the set of all càdlàg functions from \([0, t]\) (resp. \([0, \infty)\)) to \( \mathbb{R} \) equipped with the Skorohod \( J_1 \) topology. The inhomogeneous Brownian motion with a parabolic drift is given by

\[
B^\lambda_{\mu, \eta}(s) = \frac{\sqrt{\eta}}{\mu} B(s) + \frac{\lambda s}{2} - \frac{\eta s^2}{2\mu^3}, \tag{2.5}
\]

where \( B = (B(s))_{s \geq 0} \) is a standard Brownian motion, and \( \mu > 0, \eta > 0 \) and \( \lambda \in \mathbb{R} \) are constants. Define the reflected version of \( B^\lambda_{\mu, \eta} \) as

\[
W^\lambda(s) = B^\lambda_{\mu, \eta}(s) - \min_{0 \leq t \leq s} B^\lambda_{\mu, \eta}(t). \tag{2.6}
\]

For a function \( f \in \mathbb{C}[0, \infty) \), an interval \( \gamma = (l, r) \) is called an excursion above past minima or simply an excursion of \( f \) if \( f(l) = f(r) = \min_{u \leq r} f(u) \) and \( f(x) > f(r) \) for all \( l < x < r \). \( |\gamma| = r(\gamma) - l(\gamma) \) will denote the length of the excursion \( \gamma \).

Also, define the counting process of marks \( \mathbf{N}^\lambda = (N^\lambda(s))_{s \geq 0} \) to be a process that has intensity \( \beta W^\lambda(s) \) at time \( s \) conditional on \( (W^\lambda(u))_{u \leq s} \), so that

\[
N^\lambda(s) = \int_0^s \beta W^\lambda(u) du \tag{2.7}
\]

is a martingale (see [2]). For an excursion \( \gamma \), let \( N(\gamma) \) denote the number of marks in the interval \([l(\gamma), r(\gamma)]\).

**Remark 1.** By [2, Lemma 25] and the Cameron-Martin theorem, almost surely, the excursions of the paths of \( B^\lambda_{\mu, \eta} \) can be rearranged in decreasing order of length and the ordered excursion lengths can be considered as a vector in \( \ell^2 \). Let \( \gamma^\lambda = (|\gamma^\lambda_j|)_{j \geq 1} \) be the ordered excursion lengths of \( B^\lambda_{\mu, \eta} \). Then, \( ((|\gamma^\lambda_j|, N(\gamma^\lambda_j)))_{j \geq 1} \) can be ordered as an element of \( U_i^0 \) almost surely by [5, Theorem 4.1]. Denote this element of \( U_i^0 \) by \( \mathbf{Z}(\lambda) = ((Y^\lambda_j, N^\lambda_j))_{j \geq 1} \) obtained from \( ((|\gamma^\lambda_j|, N(\gamma^\lambda_j)))_{j \geq 1} \).

Finally, we define a Markov process \( \mathbf{X} : = (X(s))_{-\infty < s < \infty} \text{ on } \mathbb{D}((-\infty, \infty), \ell^2) \), called the multiplicative coalescent process. Think of \( \mathbf{X}(s) \) as a collection of masses of some objects in a system at time \( s \). Thus the \( i^{th} \) object has mass \( X_i(s) \) at time \( s \). The evolution of the system take place according to the following rule at time \( s \): At rate \( X_i(s)X_j(s) \), objects \( i \) and \( j \) merge into a new component of mass \( X_i(s) + X_j(s) \). This process has been extensively studied in [2, 3]. In particular, Aldous [2, Proposition 5] showed that this is a Feller process.


3 Main results

Consider \( n \) vertices labeled by \( [n] := \{1, 2, \ldots, n\} \) and a sequence of degrees \( d = (d_i)_{i \in [n]} \) such that \( \ell_n = \sum_{i \in [n]} d_i \) is even. For convenience we suppress the dependence of the degree sequence on \( n \) in the notation. The configuration model on \( n \) vertices with degree sequence \( d \) is constructed as follows:

Equip vertex \( j \) with \( d_j \) stubs, or half-edges. Two half-edges create an edge once they are paired. Therefore, initially we have \( \ell_n = \sum_{i \in [n]} d_i \) half-edges. Obviously, \( \ell_n \) is even. Pick any one half-edge and pair it with a uniformly chosen half-edge from the remaining unpaired half-edges and keep repeating the above procedure until we exhaust all the unpaired half-edges.

Note that the graph constructed by the above procedure may contain self-loops or multiple edges. It can be shown [23, Proposition 7.13] that, conditionally on \( CM_n(d) \) being simple, the law of such graphs is uniform over all possible simple graphs with degree sequence \( d \).

In this section, we discuss the main results in this paper. As discussed in the introduction, our results are twofold: (i) General \( CM_n(d) \) at criticality, and (ii) Critical percolation on a super-critical configuration model, both under a finite third moment assumption.

3.1 Configuration model results

Let us assume that we have a configuration model \( CM_n(d) \) for each \( n \) satisfying the following:

**Assumption 1.** Let \( D_n \) denote the degree of a vertex chosen uniformly at random independently of the graph. Then,

(i) (weak convergence of \( D_n \))

\[
D_n \xrightarrow{D} D
\]

for some random variable \( D \) such that \( \mathbb{E}[D^3] < \infty \).

(ii) (uniformly integrability of \( D_n^3 \))

\[
\mathbb{E} \left[ D_n^3 \right] = \frac{1}{n} \sum_{i \in [n]} d_i^3 \to \mathbb{E}[D^3].
\]

(iii) (critical window)

\[
\nu_n := \frac{\sum_{i \in [n]} d_i(d_i-1)}{\sum_{i \in [n]} d_i} = 1 + \lambda n^{-1/3} + o(n^{-1/3}),
\]

for some \( \lambda \in \mathbb{R} \).

(iv) \( \mathbb{P}(D = 1) > 0 \).

Suppose that \( \mathcal{C}_{(1)}, \mathcal{C}_{(2)}, \ldots \) are the connected components of \( CM_n(d) \) in decreasing order of size. In case of a tie, order the components according to the values of the minimal indices of those components. For a connected graph \( G \), let \( SP(G) := (\text{number of edges in } G) - (|G| - 1) \) denote the number of surplus edges. Intuitively, this measures the deviation of \( G \) from a tree-like structure. Let \( \sigma_r = \mathbb{E}[D^r] \) and consider the reflected Brownian motion, the excursions, and the counting process \( N^\lambda \) as defined in Section 2 with parameters

\[
\mu := \sigma_1, \quad \eta := \sigma_3 \mu - \sigma_2^2, \quad \beta := 1/\mu.
\]

Let \( \gamma^\lambda \) denote the vector of excursion lengths of the process \( B_{\mu,\eta}^\lambda \), arranged in non-increasing order. Our main results are as follows:
Theorem 1. Fix any $\lambda \in \mathbb{R}$. Under Assumption 1,
\[ n^{-2/3} \left( |\mathcal{C}^{(i)}| \right)_{j \geq 1} \xrightarrow{\mathcal{L}} \gamma^\lambda \] (3.5)
with respect to the $\ell_2$ topology.

Recall the definition of $Z(\lambda)$ from Remark 1. Order the vector $(n^{-2/3}|\mathcal{C}^{(i)}|, \text{SP}(\mathcal{C}^{(i)}))_{j \geq 1}$ as an element of $U_0^n$ and denote it by $Z_n(\lambda)$.

Theorem 2. Fix any $\lambda \in \mathbb{R}$. Under Assumption 1,
\[ Z_n(\lambda) \xrightarrow{\mathcal{L}} Z(\lambda) \] (3.6)
with respect to the $U_0^n$ topology.

In words, Theorem 1 gives the precise asymptotic distribution of the component sizes re-scaled by $n^{2/3}$ and Theorem 2 gives the asymptotic number of surplus edges in each component jointly with their sizes.

Remark 2. The strength of Theorems 1 and 2 lies in Assumption 1. Clearly, Assumption 1 is satisfied when the distribution of $D$ satisfies an asymptotic power-law relation with finite third moment, i.e., $\mathbb{P}(D \geq x) \sim x^{-(\tau - 1)}(1 + o(1))$ for some $\tau > 4$. Also, if one has a random degree-sequence that satisfies Assumption 1 with high probability, then Theorems 1 and 2 hold conditionally on the degrees. In particular, when the degree sequence consists of an i.i.d. sample from a distribution with $\mathbb{E}[D^3] < \infty$ [16], then Assumption 1 is satisfied almost surely. We will later see that degree sequences in the percolation scaling window also satisfy Assumption 1.

3.2 Percolation results

Bond percolation on a graph $G$ refers to deleting edges of $G$ independently with equal probability $p$. Consider bond percolation on $\text{CM}_n(d)$ with probability $p_n$, yielding $\text{CM}_n(d, p_n)$. We assume the following:

Assumption 2. (i) Assumption 1 (i) and (ii) hold for the degree sequence and the $\text{CM}_n(d)$ is super-critical, i.e.
\[ \nu_n = \frac{\sum_{i \in [n]} d_i (d_i - 1)}{\sum_{i \in [n]} d_i} \rightarrow \nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]} > 1. \] (3.7)

(ii) (critical window for percolation) For some $\lambda \in \mathbb{R}$,
\[ p_n = p_n(\lambda) := \frac{1}{\nu_n} \left( 1 + \frac{\lambda}{n^{1/3}} \right). \] (3.8)

Note that $p_n(\lambda)$, as defined in Assumption 2 (ii), is non-negative for $n$ sufficiently large. Now, suppose $d_i \sim \text{Bin}(\hat{d}_i, \sqrt{p_n})$, $n_+ := \sum_{i \in [n]} (d_i - \hat{d}_i)$ and $\hat{n} = n + n_+$. Consider the degree sequence $\hat{d}$ consisting of $\hat{d}_i$ for $i \in [n]$ and $n_+$ additional vertices of degree 1, i.e. $\hat{d}_i = 1$ for $i \in [\hat{n}] \setminus [n]$. We shall show later that the degree $\hat{D}_n$ of a random vertex from this degree sequence satisfies Assumption 1 (i), (ii) almost surely for some random variable $\hat{D}$ with $\mathbb{E}[\hat{D}^3] < \infty$. Now, using the notation in Section 2, define $\tilde{\gamma}^\lambda_j$ to be the ordered excursions of the inhomogeneous Brownian motion $B^\lambda_{\mu, \eta, \beta}$ and $N^\lambda$ with the parameters
\[ \mu = \mathbb{E}[\hat{D}], \quad \eta = \mathbb{E}[\hat{D}^2] = \mathbb{E}[\hat{D}] - \mathbb{E}[\hat{D}^2], \quad \beta = 1/\mathbb{E}[\hat{D}]. \] (3.9)

Let $p = 1/\nu$. Denote the $j^{th}$ largest cluster of $\text{CM}_n(d, p_n(\lambda))$ by $\mathcal{C}^{(i)}(\lambda)$. Also, let $Z_n^p(\lambda)$ denote the vector in $U_0^n$ obtained by rearranging critical percolation clusters and their surplus edges and $Z(\lambda)$ denote the vector in $U_0^n$ obtained by rearranging $((\sqrt{\nu})\tilde{\gamma}^\lambda_j, N(\tilde{\gamma}^\lambda_j))_{j \geq 1}$. 


Theorem 3. Under Assumption 2,
\[ Z^n_t(\lambda) \xrightarrow{\mathcal{L}} \tilde{Z}(\lambda) \]  
(3.10)
with respect to the \( \mathcal{U}^0_i \) topology.

Next we consider the percolation cluster for multiple values of \( \lambda \). There is a very natural way to couple \((CM_n(d, p_n(\lambda)))_{\lambda \in \mathbb{R}}\) described as follows: for \( \lambda < \lambda' \), perform bond-percolation on \( CM_n(d, p_n(\lambda')) \) with probability \( p_n(\lambda)/p_n(\lambda') \). The resulting graph is distributed as \( CM_n(d, p_n(\lambda)) \). This can be used to couple \((CM_n(d, p_n(\lambda_i)))_{i=0}^{k-1}\) for any fixed \( k \geq 1 \). The next theorem shows that the convergence of the component sizes holds jointly in finitely many locations within the critical window, under the above described coupling:

Theorem 4. Under Assumption 2 and with \( C_n(\lambda) = \langle n^{-2/3} |\mathcal{C}_i^n(\lambda)| \rangle_{i \geq 1} \), for any fixed \( k \in \mathbb{N} \) and \(-\infty < \lambda_0 < \lambda_1 < \cdots < \lambda_{k-1} < \infty\),
\[ (C_n(\lambda_0), C_n(\lambda_1), \ldots, C_n(\lambda_{k-1})) \xrightarrow{\mathcal{L}} \sqrt{p(\tilde{\gamma}^\lambda_0, \tilde{\gamma}^\lambda_1, \ldots, \tilde{\gamma}^\lambda_{k-1})} \]  
(3.11)
with respect to the \((\ell^k)\) topology where \( p = 1/\nu \).

Remark 3. The coupling for the limiting process in Theorem 4 is given by the multiplicative coalescent process described in Section 2. This will become more clear when we describe the ideas of the proof. An intuitive picture is that as we change the value of the percolation parameter from \( p_n(\lambda) \) to \( p_n(\lambda + d\lambda) \), exactly one edge is added to the graph and the two endpoints \( i, j \) are chosen approximately proportional to the number of half-edges of \( i \) and \( j \) that were not retained in percolation. Define the degree deficiency \( D_i \) of a component \( \mathcal{C}_i \) to be the total number of half-edges in a component that were not retained in percolation. Think of \( D_i \) as the mass of \( \mathcal{C}_i \). By the above heuristics, \( \mathcal{C}_i \) and \( \mathcal{C}_j \) merge at rate proportional to \( D_i D_j \) and creates a cluster of mass \( D_i + D_j - 2 \). Later, we shall show that the degree deficiency of a component is approximately proportional to the component size. Therefore, the component sizes merge approximately like the multiplicative coalescent over the critical scaling window.

Remark 4. Janson [12] studied the phase transition of the maximum component size for percolation on a super-critical configuration model. The critical value turns out to be \( p = 1/\nu \). This is precisely the reason behind taking \( p_n \) of the form given by Assumption 2 (ii). The reason behind the width of the critical window being of the order \( n^{1/3} \) does not come as a surprise in light of the existing literature [2, 6, 20, 21].

Remark 5. Theorem 1 and Theorem 2 also hold for configuration models conditioned on simplicity. We do not give a proof here. The arguments in [16, Section 7] can be followed verbatim to obtain a proof of this fact. As a result, Theorem 3 and Theorem 4 also hold, conditioned on simplicity.

The rest of the paper is organized as follows: In Section 4.1, we give a brief overview of the relevant literature. This will enable the reader to understand better the relation of this work to the large body of literature already present. Also, it will become clear why the choices of the parameters in Assumption 1 (iii) and Assumption 2 (ii) should correspond to the critical scaling window. We prove Theorems 1 and 2 in Section 5. In Section 6 we find the asymptotic degree distribution in each component. This is used along with Theorem 2 to establish Theorem 3 in Section 7. In Section 8, we analyze the evolution of the component sizes over the percolation critical window and prove Theorem 4.


4 Discussion

4.1 Literature overview

Erdős-Rényi type behavior. We first explain what ‘Erdős-Rényi type behavior’ means. The study of critical window for random graphs started with the seminal paper [2] on the Erdős-Rényi random graphs with \( p = n^{-1}(1 + \lambda n^{-1/3}) \). Aldous showed in this regime that all the components are of asymptotic size \( n^{2/3} \) and the ordered component sizes have the same distribution as the excursions of a Brownian motion with a parabolic drift. Aldous also described the component sizes as a dynamic process as \( \lambda \) varies and showed that the dynamic process can be described by a process called the standard multiplicative coalescent. In Theorem 4, we show that similar results hold for the configuration model under a very general set of assumptions. Of course, for general configuration models, there is no obvious way to couple the graphs such that the location parameter in the scaling window varies and percolation seems to be the most natural way to achieve this. By [11, 12], percolation on a configuration model can be viewed as a configuration model with a random degree sequence and this is precisely the reason for studying percolation in this paper.

Universality and optimal assumptions. In [6] it was shown that, inside the critical scaling window, the scaled component sizes of an inhomogeneous random graph with

\[
p_{ij} = 1 - \exp\left(\frac{-\lambda n^{-1/3} w_i w_j}{\sum_{k \in [n]} w_k}\right)
\]

converge to excursions of an inhomogeneous Brownian motion with a parabolic drift under only finite third-moment assumption on the weight distribution. We establish a counterpart of this for the configuration model in Theorem 1. Later Nachmias and Peres [20] studied the case of percolation scaling window on the random regular graph; for percolation on the configuration model similar results were obtained by Riordan [21] for bounded maximum degrees. Joseph [16] obtained the same scaling limits when the degrees are i.i.d. samples from a distribution having finite third moment. Theorem 2 and Theorem 3 prove stronger versions of all these existing results for the configuration model under less stringent and possibly optimal assumptions. Further, in Theorem 4, we give a dynamic picture for percolation cluster sizes in the critical window and show that this dynamics can be approximated by the multiplicative coalescent.

Comparison to branching processes. In [14, 19] the phase transition for the component sizes configuration model was identified in terms of the parameter \( \nu = \frac{\mathbb{E}[(D - 1)^2]/\mathbb{E}[D]}{\mathbb{E}[D]} \). Janson and Luczak [14], showed that the configuration model can be locally approximated by a branching process \( \mathcal{X} \) which has \( \nu \) as its expected progeny and thus, when \( \nu > 1 \), \( \text{CM}_n(d) \) has a component \( C_{\text{max}} \) of approximate size \( \rho n \), where \( \rho \) is the survival probability of \( \mathcal{X} \). Further, the progeny distribution of \( \mathcal{X} \) has finite variance when \( \mathbb{E}[D^3] < \infty \). Now, for a branching process with mean \( \approx 1 + \varepsilon \) and finite variance \( \sigma^2 \), the survival probability is approximately \( 2\sigma^{-2}\varepsilon \) for small \( \varepsilon > 0 \). This seems to suggest that the maximum component size under Assumption 1 should be of the order \( n^{2/3} \) since \( \varepsilon = \Theta(n^{-1/3}) \). Theorem 1 formalizes this intuition and shows that in fact all the maximal component sizes are of the order \( n^{2/3} \).

4.2 Proof ideas

The proof of Theorem 1 uses standard functional central limit theorem argument. Indeed we associate a suitable semi-martingale with the graph obtained from an exploration process used to explore the connected components of \( \text{CM}_n(d) \). The martingale part is then shown to converge to an inhomogeneous Brownian motion, and the drift part is shown to converge to a parabola. The fact that the component sizes can be expressed in terms of the hitting times of the semi-martingale implies the finite-dimensional convergence of the component sizes. The convergence with respect
to $\ell_1^2$ is then concluded using size-biased point process arguments by Aldous [2]. Theorem 2 requires a careful estimate of the tail probability of the distribution of surplus edges when the component size is small and we obtain this using martingale estimates. Theorem 3 is proved by showing that the percolated degree sequence satisfies Assumption 1 almost surely. Finally, we prove Theorem 4 using a coupling argument. Key challenges here are that, for each fixed $n$, the components do not merge according to their component size, and that the components do not merge exactly like a multiplicative coalescent over the scaling window. We shall deal with these in Section 8.

4.3 Open problems

(a) Theorem 4 proves the joint convergence at finitely many locations of the scaling window. However, the tightness of $(C_n(\lambda))_{\lambda \in \mathbb{R}}$ in $\mathcal{D}((\infty, \infty), \ell_1^2)$ should also hold so that we have convergence of the whole process.

(b) It is also believed that the connected components, considered as metric spaces with the graph distance, converge to a suitable limiting metric structure under the finite third-moment condition only. This result was proved under exponential moment conditions in [4, Theorem 4.7].

(c) A reason for studying percolation in this paper is to understand the minimal spanning tree of the giant component. For a super-critical configuration model with i.i.d. edge weights, it should be the case that the minimal spanning tree can be described by the critically percolated graph at a very high location of the scaling window. Such results were obtained in [1] for the Erdős-Rényi random graph.

5 Proofs of Theorems 1 and 2

5.1 The exploration process

Let us explore the graph sequentially using a natural approach outlined in [21]. At step $k$, divide the set of half-edges into three groups; sleeping half-edges $S_k$, active half-edges $A_k$, and dead half-edges $D_k$. The depth-first exploration process can be summarized in the following algorithm:

Algorithm 1 (DFS exploration). At $k = 0$, $S_k$ contains all the half-edges and $A_k, D_k$ are empty. While $(S_k \neq \emptyset$ or $A_k \neq \emptyset$) we do the following at stage $k + 1$:

S1 If $A_k \neq \emptyset$, then take the smallest half-edge $a$ from $A_k$.

S2 Take the half-edge $b$ from $S_k$ that is paired to $a$. Suppose $b$ is attached to a vertex $w$ (which is necessarily not discovered yet). Declare $w$ to be discovered, let $r = d_w - 1$ and $b_{w1}, b_{w2}, \ldots b_{wr}$ be the half-edges of $w$ other than $b$. Declare $b_{w1}, b_{w2}, \ldots b_{wr}, b$ to be smaller than all other half-edges in $A_k$. Also order the half-edges of $w$ among themselves as $b_{w1} > b_{w2} > \cdots > b_{wr} > b$.

Now identify $B_k \subset A_k \cup \{b_{w1}, b_{w2}, \ldots, b_{wr}\}$ as the collection of all half-edges in $A_k$ paired to one of the $b_{wi}$’s and the corresponding $b_{wi}$’s. Similarly identify $C_k \subset \{b_{w1}, b_{w2}, \ldots, b_{wr}\}$ which is the collection of loops incident to $w$. Finally, declare $A_{k+1} = A_k \cup \{b_{w1}, b_{w2}, \ldots, b_{wr}\} \setminus (B_k \cup C_k)$, $D_{k+1} = D_k \cup \{a, b\} \cup B_k \cup C_k$ and $S_{k+1} = S_k \setminus \{b_{w1}, b_{w2}, \ldots, b_{wr}\}$. Go to stage $k + 2$.

S3 If $A_k = \emptyset$ for some $k$, then take out one half-edge $a$ from $S_k$ uniformly at random and identify the vertex $v$ incident to it. Declare $v$ to be discovered. Let $r = d_v - 1$ and assume that $a_{v1}, a_{v2}, \ldots a_{vr}$ are the half-edges of $v$ other than $a$ and identify the collection of half-edges involved in a loop/multiple edge/cycle $C_k$ as in Step 2. Order the half-edges of $v$ as $a_{v1} > a_{v2} > \cdots > a_{vr} > a$. Set $A_{k+1} = \{a, a_{v1}, a_{v2}, \ldots, a_{vr}\} \setminus C_k$, $D_{k+1} = D_k \cup C_k$, and $S_{k+1} = S_k \setminus \{a, a_{v1}, a_{v2}, \ldots, a_{vr}\}$. Go to stage $k + 2$.
In words, we explore a new vertex at each stage and throw away all the half-edges involved in a loop/multiple edge/cycle with the vertex set already discovered before proceeding to the next stage. The ordering of the half-edges is such that the connected components of $CM_n(d)$ are explored in the depth-first way. We call the half-edges of $B_k \cup C_k$ cycle half-edges because they create loops, cycles or multiple edges in the graph. Let

$$A_k := |A_k|, \quad c(k+1) := (|B_k| + |C_k|)/2, \quad U_k := |S_k|. \quad (5.1)$$

Let $d_{(j)}$ be the degree of the $j^{th}$ explored vertex and define the following process:

$$S_n(0) = 0, \quad S_n(i) = \sum_{j=1}^{i} (d_{(j)} - 2 - 2c_{(j)}). \quad (5.2)$$

The process $S_n = (S_n(i))_{i \in [n]}$ “encodes the component sizes as lengths of path segments above past minima” as discussed in [2]. Suppose $C_i$ is the $i^{th}$ connected component explored by the above exploration process. Define

$$\tau_k = \inf \{ i : S_n(i) = -2k \}. \quad (5.3)$$

Then $C_k$ is discovered between the times $\tau_{k-1} + 1$ and $\tau_k$ and $C_k$ has size $\tau_k - \tau_{k-1}$.

5.2 Size-biased exploration

The vertices are explored in a size-biased manner with sizes proportional to their degrees, i.e., if we denote by $v_{(i)}$ the $i^{th}$ explored vertex in Algorithm 1 and by $d_{(i)}$ the degree of $v_{(i)}$ then

$$P(v_{(i)} = j|v_{(1)}, v_{(2)}, ..., v_{(i-1)}) = \frac{d_j}{\sum_{k \not\in \mathcal{Y}_{i-1}} d_k} = \frac{d_j}{\sum_{k \in [n]} d_k - \sum_{k=1}^{j-1} d_{(k)}}, \quad \forall j \in \mathcal{Y}_{i-1}, \quad (5.4)$$

where $\mathcal{Y}_i$ denotes the first $i$ vertices to be discovered in the above exploration process. The following lemma will be used in the proof of Theorem 1:

Lemma 5. Suppose that Assumption 1 holds and denote $\sigma_r = \mathbb{E}[D^r]$ and $\mu = \mathbb{E}[D]$. Then for all $t > 0$, as $n \to \infty$,

$$\sup_{u \leq t} \left| n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3} u \rfloor} d_{(i)} - \frac{\sigma_2}{\mu} \right| \overset{P}{\to} 0, \quad (5.5)$$

and

$$\sup_{u \leq t} \left| n^{-2/3} \sum_{i=1}^{\lfloor n^{2/3} u \rfloor} d_{(i)}^2 - \frac{\sigma_3}{\mu} \right| \overset{P}{\to} 0. \quad (5.6)$$

The proof of this lemma follows from a more general result stated in Proposition 29 and the following observation:

Lemma 6. Assumption 1 implies

$$\lim_{k \to \infty} \lim_{n \to \infty} \frac{1}{n} \sum_{j \in [n]} 1\{d_j > k\} d_j^r = 0, \quad r = 1, 2, 3. \quad (5.7)$$

For $r = 3$, in particular, this implies $d_{\text{max}}^3 = o(n)$. 

9
5.3 Estimate of cycle half-edges

The following lemma gives an estimate of the number of cycle half-edges created up to time \( t \). This result is proved in [21] for bounded degrees. In our case, it follows from Lemma 5 as we show below:

**Lemma 7.** For Algorithm 1, if \( A_k = \|A_k\|, B_k := \|B_k\|, \) and \( C_k := \|C_k\|, \) then

\[
\mathbb{E}[B_k|\mathcal{F}_k] = (1 + o_p(1)) \frac{2A_k}{U_k} + O_p(n^{-2/3}) \tag{5.8}
\]

and

\[
\mathbb{E}[C_k|\mathcal{F}_k] = O_p(n^{-1}) \tag{5.9}
\]

uniformly for \( k \leq t^{2/3} \) and any \( t > 0 \), where \( \mathcal{F}_k \) is the sigma-field generated by the information revealed up to stage \( k \). Further, all the \( O_p \) and \( o_p \) terms in (5.8) and (5.9) can be replaced by \( O_E \) and \( o_E \).

**Proof.** Suppose \( U_k := |S_k| \). First note that by (5.5)

\[
\frac{U_k}{n} = \frac{1}{n} \sum_{j \in [n]} d_j - \frac{1}{n} \sum_{j=1}^k d_{(j)} = \mathbb{E}[D] + o_p(1) \tag{5.10}
\]

uniformly over \( k \leq t^{2/3} \). Now let \( a \) be the half-edge that is being explored at stage \( k + 1 \). Now each of the \( (A_k - 1) \) half-edges of \( A_k \setminus \{a\} \) is equally likely to be paired with a half-edge of \( v_{(k+1)} \), thus creating two elements of \( B_k \). Also, given \( \mathcal{F}_k \) and \( v_{(k+1)} \), the probability that a half-edge of \( A_k \setminus \{a\} \) is paired to one of the half-edges of \( v_{(k+1)} \) is \( (d_{(k+1)} - 1)/(U_k - 1) \). Therefore,

\[
\mathbb{E}[B_k|\mathcal{F}_k, v_{(k+1)}] = 2(A_k - 1) \frac{d_{(k+1)} - 1}{U_k - 1} = 2(d_{(k+1)} - 1) \frac{A_k}{U_k - 1} - 2 \frac{d_{(k+1)} - 1}{U_k - 1}. \tag{5.11}
\]

Hence,

\[
\mathbb{E}[B_k|\mathcal{F}_k] = 2\mathbb{E}[d_{(k+1)} - 1|\mathcal{F}_k] \frac{A_k}{U_k - 1} - 2 \mathbb{E}[d_{(k+1)} - 1|\mathcal{F}_k] \frac{A_k}{U_k - 1}. \tag{5.12}
\]

Now, using (5.5) and (5.6),

\[
\mathbb{E}[d_{(k+1)} - 1|\mathcal{F}_k] = \frac{\sum_{j \notin \mathcal{R}_k} d_j (d_j - 1)}{\sum_{j \notin \mathcal{R}_k} d_j} = \frac{\sum_{j \in [n]} d_j^2}{\sum_{j \in [n]} d_j} - 1 + o_p(1) = 1 + o_p(1). \tag{5.13}
\]

uniformly over \( k \leq t^{2/3} \), where the last step follows from Assumption 1. Further, since \( p_0 < 1 \), \( U_k \geq c_0 n \) uniformly over \( k \leq o(n) \). Thus, (5.12) gives (5.8). The fact that all the \( O_p \), \( o_p \) can be replaced by \( O_E \), \( o_E \) follows from \( \sum_{j \in [n]} d_j^2 - kd_{\text{max}}^2 \leq \sum_{j \notin \mathcal{R}_k} d_j^2 \leq \sum_{j \in [n]} d_j^2 \) for \( r = 1, 2 \), together with \( d_{\text{max}} = o(n^{1/3}) \). To prove (5.9), note that

\[
\mathbb{E}[C_k|\mathcal{F}_k, v_{(k+1)}] = 2(d_{(k+1)} - 1) \frac{d_{(k+1)} - 1}{U_k - 1}. \tag{5.14}
\]

By Assumption 1 and (5.5)

\[
\mathbb{E}[d_{(k+1)}^2|\mathcal{F}_k] = \frac{\sum_{j \notin \mathcal{R}_k} d_j^2}{\sum_{j \notin \mathcal{R}_k} d_j} \leq \frac{\sum_{j \in [n]} d_j^2}{\sum_{j \in [n]} d_j + o_p(n^{2/3})} = O_p(1), \tag{5.15}
\]

uniformly for \( k \leq t^{2/3} \). Therefore,

\[
\mathbb{E}[C_k|\mathcal{F}_k] = O_p(n^{-1}) \tag{5.16}
\]

uniformly over \( k \leq t^{2/3} \). Again, \( O_p \) term can be replaced by \( O_E \), as argued before. \( \square \)
5.4 Key ingredients

For any $\mathbb{D}[0, \infty)$-valued process $X_n$ define $\bar{X}_n(u) := n^{-1/3}X_n(\lfloor n^{2/3}u \rfloor)$ and $\bar{X}_n := (\bar{X}_n(u))_{u \geq 0}$.

The following result is the main ingredient for proving Theorem 1. Recall the definition of $B^\lambda_{\mu, \eta}$ from (2.5) with parameters given in (3.4).

**Theorem 8** (Convergence of the exploration process). Under Assumption 1, as $n \to \infty$,

$$\bar{S}_n \overset{\mathcal{L}}{\longrightarrow} B^\lambda_{\mu, \eta}$$

with respect to the Skorohod $J_1$ topology.

As in [16], we shall prove this by approximating $S_n$ by a simpler process defined as

$$s_n(0) = 0, \quad s_n(i) = \sum_{j=1}^{i}(d_{(j)} - 2).$$

(5.18)

Note that the difference between the processes $S_n$ and $\bar{s}_n$ is due to the cycles, loops, and multiple-edges encountered during the exploration. Following the approach of [16], it will be enough to prove the following:

**Proposition 9.** Under Assumption 1, as $n \to \infty$,

$$\bar{s}_n \overset{\mathcal{L}}{\longrightarrow} B^\lambda_{\mu, \eta}$$

(5.19)

with respect to the Skorohod $J_1$ topology.

**Remark 6.** It will be shown that Proposition 9 implies Theorem 8 by showing that the distributions of $\bar{S}_n$ and $\bar{s}_n$ are very close as $n \to \infty$. This is achieved by proving that we shall not see too many cycle half-edges up to the time $\lfloor n^{2/3}u \rfloor$ for any fixed $u > 0$.

From here onwards we shall look at the continuous versions of the processes $\bar{S}_n$ and $\bar{s}_n$ by linearly interpolating between the values at the jump points and write it using the same notation. It is easy to see that these continuous versions differ from their càdlàg versions by at most $n^{-1/3}d_{\max} = o(1)$ uniformly on $[0, T]$, for any $T > 0$. Therefore, the convergence in law of the continuous versions implies the convergence in law of the càdlàg versions and vice versa. Before proceeding to show that Theorem 8 is a consequences of Proposition 9, we shall need to bound the difference of these two processes in a suitable way. We need the following lemma. Recall the definition of $c_{(i+1)} := (B_i + C_i)/2$ from (5.1).

**Lemma 10.** Fix $t > 0$ and $M > 0$ (large). Define $E_n(t, M) := \{ \max_{s \leq t} \{ \bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u) \} < M \}$. Then

$$\limsup_{n \to \infty} \sum_{k \leq tn^{2/3}} \mathbb{E}[c_{(i)} 1_{E_n(t, M)}] < \infty.$$  

(5.20)

**Proof.** Lemma 10 is similar to [16, Lemma 6.1]. We add a brief proof here. Note that, for all large $n$, $A_k \leq Mn^{1/3}$ on $E_n(t, M)$, because

$$A_k = S_n(k) - \min_{j \leq k} S_n(j) + 2 = s_n(k) - \sum_{j=1}^{k} c_{(j)} - \min_{j \leq k} S_n(j) + 2 \leq s_n(k) - \min_{j \leq k} s_n(j) + O(1),$$

(5.21)

where the last step follows by noting that $\min_{j \leq k} s_n(j) \leq \min_{j \leq k} S_n(j) + 2 \sum_{j=1}^{k} c_{(j)}$. By Lemma 7

$$\mathbb{E}[c_{(i)} 1_{E_n(t, M)}] \leq \frac{Mn^{1/3}}{\mu n} + o(n^{-2/3}) = \frac{M}{\mu} - \frac{2/3}{n^{2/3}} + o(n^{-2/3})$$

(5.22)

uniformly for $k \leq tn^{2/3}$. Summing over $1 \leq k \leq tn^{2/3}$ and taking the $\limsup$ completes the proof.
Proof of Theorem 8. The argument is a standard and we include the proof for the sake of completeness (see [16, Section 6.2]). For \( t > 0 \) and \( M > 0 \) define the event \( E_n(t, M) \) as in Lemma 10. Also, denote by \( C_L[0, t] \) the set of all bounded Lipschitz functions from \( [0, t] \) to \( \mathbb{R} \). Since \( C_L[0, t] \) separates the points in \( [0, t] \), and \( [0, t] \) is compact, \( C_L[0, t] \) is dense in \( C[0, t] \). Therefore, to prove Theorem 8, it suffices to prove as \( n \to \infty \)

\[
\mathbb{E}[f(S_n)] - \mathbb{E}[\bar{f}(\bar{S}_n)] \to 0, \tag{5.23}
\]

for all \( f \in C_L[0, t] \). Choose and fix \( b > 0 \) such that

\[
z, z_1, z_2 \in C[0, t] \implies |f(z)| \leq b \quad \text{and} \quad |f(z_1) - f(z_2)| \leq b||z_1 - z_2||_\infty. \tag{5.24}
\]

Now,

\[
\begin{align*}
|\mathbb{E}[f(S_n)] - \mathbb{E}[\bar{f}(\bar{S}_n)]| &\leq |(\mathbb{E}[f(S_n)1_{E_n(t, M)}] - \mathbb{E}[f(\bar{S}_n)1_{E_n(t, M)}])| + |(\mathbb{E}[f(S_n)1_{E_n(t, M)^c}] - \mathbb{E}[f(\bar{S}_n)1_{E_n(t, M)^c}])| \\
&\leq b\mathbb{E}[|\bar{S}_n - S_n|1_{E_n(t, M)}] + 2b\mathbb{P}\left( \max_{s \leq t} \{ \bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u) \} \geq M \right) \\
&\leq 2bn^{-1/3} \sum_{k \leq tn^{2/3}} \mathbb{E}[c(k)1_{E_n(t, M)}] + 2b\mathbb{P}\left( \max_{s \leq t} \{ \bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u) \} \geq M \right). \\
&\leq 2bn^{-1/3} \sum_{k \leq tn^{2/3}} \mathbb{E}[c(k)1_{E_n(t, M)}] + 2b\mathbb{P}\left( \max_{s \leq t} \{ \bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u) \} \geq M \right). \tag{5.25}
\end{align*}
\]

The first term in the above sum tends to zero as \( n \to \infty \), by Lemma 10. The fact that the reflection of a process is a continuous map from \( \mathbb{D}([0, \infty), \mathbb{R}) \) to \( \mathbb{D}([0, \infty), \mathbb{R}) \) (see [24, Theorem 13.5.1]), Proposition 9 implies

\[
(\bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u))_{s \geq 0} \overset{L}{\to} W^\lambda, \tag{5.26}
\]

where \( W^\lambda \) is as defined in (5.2). By the Portmanteau lemma

\[
\limsup_{n \to \infty} \mathbb{P}\left( \max_{s \leq t} \{ \bar{s}_n(s) - \min_{u \leq s} \bar{s}_n(u) \} \geq M \right) \leq \mathbb{P}\left( \max_{s \leq t} W^\lambda(s) \geq M \right), \tag{5.27}
\]

for any \( t > 0 \). The proof follows by taking the limit \( M \to \infty \).

From here onward the main focus of this section will be to prove Proposition 9. We use the martingale functional central limit theorem in a similar manner as Aldous.

Proof of Proposition 9. Let \( \{ \mathcal{F}_i \}_{i \in [n]} \) be the natural filtration defined in Lemma 7. Recall the definition of \( s_n(i) \) from (5.18). By the Doob-Meyer decomposition [17, Theorem 4.10] we can write

\[
s_n(i) = M_n(i) + A_n(i), \quad s_n^2(i) = H_n(i) + B_n(i), \tag{5.28}
\]

where

\[
M_n(i) = \sum_{j=1}^{i} \left( d_j - \mathbb{E}[d_j|\mathcal{F}_{j-1}] \right), \tag{5.29a}
\]

\[
A_n(i) = \sum_{j=1}^{i} \mathbb{E}[d_j - 2|\mathcal{F}_{j-1}], \tag{5.29b}
\]

\[
B_n(i) = \sum_{j=1}^{i} \left( \mathbb{E}[d_j^2|\mathcal{F}_{j-1}] - \mathbb{E}[d_j^2|\mathcal{F}_{j-1}] \right). \tag{5.29c}
\]
Recall that for a discrete time process \((X_n(i))_{i \geq 1}\), we write \(\bar{X}_n(t) = n^{-1/3} X_n(\lfloor tn^{2/3} \rfloor)\). Our result follows from the martingale functional central limit theorem [25, Theorem 2.1] if we can prove the following four conditions: For any \(u > 0\),

\[
\sup_{s \leq u} |A_n(s) - st + \frac{\eta s^2}{2\mu^2}| \mathbb{P} \to 0, \tag{5.30a}
\]

\[
n^{-1/3} B_n(u) \mathbb{P} \to \frac{\eta}{\mu^2} u, \tag{5.30b}
\]

\[
\mathbb{E} \left[ \sup_{s \leq u} |\overline{M}_n(s) - \overline{M}_n(s-)|^2 \right] \to 0, \tag{5.30c}
\]

and

\[
n^{-1/3} \mathbb{E} \left[ \sup_{s \leq u} |\overline{B}_n(s) - \overline{B}_n(s-)| \right] \to 0. \tag{5.30d}
\]

Indeed (5.30a) gives rise to the quadratic drift term of the limiting distribution. Conditions (5.30b), (5.30c), (5.30d) are the same as [25, Theorem 2.1, Condition (iii)]. The facts that the jumps of both the martingale and the quadratic-variation process go to zero and that the quadratic variation process is converging to the quadratic variation of an inhomogeneous Brownian Motion, together imply the convergence of the martingale term. The validation of these conditions are given separately in the subsequent part of this section.

**Lemma 11.** The conditions (5.30b), (5.30c), (5.30d) hold.

**Proof.** Denote by \(\sigma_r(n) = \frac{1}{n} \sum_{i \in [n]} d_i r, r = 2, 3\) and \(\mu(n) = \frac{1}{n} \sum_{i \in [n]} d_i\). To prove (5.30b), it is enough to prove that

\[
n^{-2/3} B_n(\lfloor un^{2/3} \rfloor) \mathbb{P} \to \frac{\sigma_3 \mu - \sigma_3^2}{\mu^2} u. \tag{5.31}
\]

Recall that \(\mathbb{E}[d_{(2)}^2 | \mathcal{F}_{i-1}] = \sum_{j \notin \Gamma_{i-1}} d_j^2 / \sum_{j \notin \Gamma_{i-1}} d_j\). Furthermore, uniformly over \(i \leq un^{2/3}\),

\[
\sum_{j \notin \Gamma_{i-1}} d_j = \sum_{j \in [n]} d_j + O_p(d_{\max}) = \ell_n + o_p(n). \tag{5.32}
\]

Assume that, without loss of generality, \(j \mapsto d_j\) is non-increasing. We have, uniformly over \(i \leq un^{2/3}\),

\[
\left| \sum_{j \notin \Gamma_{i-1}} d_j^2 - n \sigma_3(n) \right| \leq \sum_{j=1}^{un^{2/3}} d_j^3. \tag{5.33}
\]

For each fixed \(k\),

\[
\frac{1}{n} \sum_{j=1}^{un^{2/3}} d_j^3 \leq \frac{1}{n} \sum_{j=1}^{un^{2/3}} 1_{\{d_j \leq k\}} d_j^3 + \frac{1}{n} \sum_{j \in [n]} 1_{\{d_j > k\}} d_j^3 \leq k^3 un^{-1/3} + \frac{1}{n} \sum_{j \in [n]} 1_{\{d_j > k\}} d_j^3 = o(1), \tag{5.34}
\]

where we first let \(n \to \infty\) and then \(k \to \infty\) and use Lemma 6. Therefore, the right-hand side of (5.33) is \(o(n)\) and we conclude that, uniformly over \(i \leq un^{2/3}\),

\[
\mathbb{E}[d_{(2)}^2 | \mathcal{F}_{i-1}] = \frac{\sigma_3}{\mu} + o_p(1). \tag{5.35}
\]

A similar argument gives

\[
\mathbb{E}[d_{(3)} | \mathcal{F}_{i-1}] = \frac{\sigma_3}{\mu} + o_p(1), \tag{5.36}
\]

Therefore, the right-hand side of (5.33) is \(o(n)\) and we conclude that, uniformly over \(i \leq un^{2/3}\),

\[
\mathbb{E}[d_{(2)}^2 | \mathcal{F}_{i-1}] = \frac{\sigma_3}{\mu} + o_p(1). \tag{5.35}
\]

A similar argument gives

\[
\mathbb{E}[d_{(3)} | \mathcal{F}_{i-1}] = \frac{\sigma_3}{\mu} + o_p(1), \tag{5.36}
\]
and (5.30b) follows by noting that the error term is $o_B(1)$, since we are summing $n^{2/3}$ terms, scaling by $n^{-2/3}$ and using the uniformity of errors over $i \leq un^{2/3}$. The proofs of (5.30c) and (5.30d) are rather short and we present them now. For (5.30c), we bound

$$
\mathbb{E}\left[ \sup_{s \leq u} |\mathcal{M}_n(s) - \mathcal{M}_n(s-)| \right] = n^{-2/3} \mathbb{E}\left[ \sup_{k \leq un^{2/3}} |M_n(k) - M_n(k-1)| \right] = n^{-2/3} \mathbb{E}\left[ \sup_{k \leq un^{2/3}} |d_{(k)} - \mathbb{E}[d_{(k)}|\mathcal{F}_{k-1}]|^2 \right] 
$$

$$(5.37)$$

Similarly, (5.30d) gives

$$
n^{-1/3} \mathbb{E}\left[ \sup_{s \leq u} |\mathcal{B}_n(s) - \mathcal{B}_n(s-)| \right] = n^{-2/3} \mathbb{E}\left[ \sup_{k \leq un^{2/3}} |B_n(k) - B_n(k-1)| \right] = n^{-2/3} \mathbb{E}\left[ \sup_{k \leq un^{2/3}} \text{var}(d_{(k)}|\mathcal{F}_{k-1}) \right] 
$$

$$(5.38)$$

$$
\leq 2n^{-2/3} d_{\max}^2,
$$

and Conditions (5.30c) and (5.30d) follow from Lemma 6 using $d_{\max} = o(n^{1/3})$. □

Next, we prove Condition (5.30a) which requires some more work. Note that

$$
\mathbb{E}[d_{(i)} - 2|\mathcal{F}_{i-1}] = \frac{\sum_{j \notin \mathcal{Y}_{i-1}} d_{j}(d_{j} - 2)}{\sum_{j \notin \mathcal{Y}_{i-1}} d_{j}} = \frac{\sum_{j \in [n]} d_{j}(d_{j} - 2)}{\sum_{j \in [n]} d_{j}} - \frac{\sum_{j \notin \mathcal{Y}_{i-1}} d_{j}(d_{j} - 2)}{\sum_{j \notin \mathcal{Y}_{i-1}} d_{j}} + \frac{\sum_{j \notin \mathcal{Y}_{i-1}} d_{j}(d_{j} - 2) \sum_{j \in \mathcal{Y}_{i-1}} d_{j}}{\sum_{j \notin \mathcal{Y}_{i-1}} d_{j} \sum_{j \in [n]} d_{j}}
$$

$$\lambda \left( n^{1/3} - \sum_{j \in \mathcal{Y}_{i-1}} d_{j}^2 \right) = \frac{\sum_{j \notin \mathcal{Y}_{i-1}} d_{j}^2 \sum_{j \in \mathcal{Y}_{i-1}} d_{j}}{\sum_{j \notin \mathcal{Y}_{i-1}} d_{j} \sum_{j \in [n]} d_{j}} + o(n^{-1/3}),
$$

(5.39)

where the last step follows from Assumption 1 (iii). Therefore,

$$
A_n(k) = \sum_{i=1}^{k} \mathbb{E}[d_{(i)} - 2|\mathcal{F}_{i-1}] = \frac{k\lambda}{n^{1/3}} - \sum_{i=1}^{k} \sum_{j \notin \mathcal{Y}_{i-1}} d_{j}^2 \sum_{j \in \mathcal{Y}_{i-1}} d_{j} + o(kn^{-1/3}).
$$

(5.40)

The following lemma estimates the sums on the right-hand side of (5.40):

**Lemma 12.** For all $u > 0$, as $n \to \infty$,

$$
\sup_{s \leq u} \left| n^{-1/3} \sum_{i=1}^{[sn^{2/3}]} \sum_{j=1}^{i-1} \frac{d_{(j)}^2}{\ell_n} - \frac{\sigma_3 s^2}{2\mu^2} \right| \overset{\mathbb{P}}{\to} 0
$$

and

$$
\sup_{s \leq u} \left| n^{-1/3} \sum_{i=1}^{[sn^{2/3}]} \sum_{j=1}^{i-1} \frac{d_{(j)}^2}{\ell_n} - \frac{\sigma_3 s^2}{2\mu^2} \right| \overset{\mathbb{P}}{\to} 0.
$$

(5.41)

(5.42)
Consequently,

\[
\sup_{s \leq u} \left| n^{-1/3} \sum_{i=1}^{\lfloor n^{2/3} \rfloor} \frac{\sum_{j \notin \mathcal{Y}_{i-1}} d_j^2 \sum_{j \in \mathcal{Y}_{i-1}} d_j}{\sum_{j \notin \mathcal{Y}_{i-1}} d_j \sum_{j \in [n]} d_j} - \frac{\sigma_3^2}{2\mu^2} \right| \mathbb{P} \to 0. \tag{5.43}
\]

**Proof.** Notice that

\[
\sup_{s \leq u} \left| n^{-1/3} \sum_{i=1}^{\lfloor n^{2/3} \rfloor} \frac{\sum_{j = 1}^{i-1} d_{(j)}^2}{\ell_n} - \frac{\sigma_3^2}{2\mu^2} \right| = \sup_{k \leq un^{2/3}} \left| n^{-1/3} \sum_{i=1}^{k} \frac{\sum_{j = 1}^{i-1} d_{(j)}^2}{\ell_n} - \frac{\sigma_3^2}{2\mu^2} \right|
\leq \frac{1}{\ell_n} \sup_{k \leq un^{2/3}} \left| n^{-1/3} \sum_{i=1}^{k} \frac{\sum_{j = 1}^{i} d_{(j)}^2}{\ell_n} - \frac{\sigma_3^2}{2\mu^2} \right| + \sup_{k \leq un^{2/3}} \left| \frac{k^2 \sigma_3}{2\mu\ell_n n^{1/3}} - \frac{k^2 \sigma_3}{2\mu n^{2/3}} \right|
\leq \frac{1}{\ell_n} \left| n^{-1/3} \sum_{i=1}^{n^{2/3}} \frac{\sum_{j = 1}^{i} d_{(j)}^2}{\ell_n} - \frac{\sigma_3^2}{\mu} \right| + o(1) + \frac{\sigma_3 n^{-1/3}}{2\mu} \left| \frac{1}{\ell_n} - \frac{1}{n \mu} \right| n^{2/3}
\leq \frac{u}{\mu + o(1)} \sup_{s \leq u} \left( n^{-2/3} \sum_{j=1}^{\lfloor n^{2/3} \rfloor} d_{(j)}^2 - \frac{\sigma_3^2}{\mu} \right) + o(1).
\]

and the (5.41) follows from (5.6) in Lemma 5. The proof of (5.42) is similar and it follows from (5.5). We now show (5.43). Recall that \(\sigma_2(n) = \frac{1}{n} \sum_{i \in [n]} d_i^2\) and observe

\[
\frac{1}{n} \sum_{j \notin \mathcal{Y}_{i-1}} d_j^2 = \sigma_2(n) - \frac{1}{n} \sum_{j \in \mathcal{Y}_{i-1}} d_j^2 = \sigma_2(n) + o_p(1) \tag{5.45}
\]

uniformly over \(i \leq un^{2/3}\) where we use Lemma 5 to conclude the uniformity. Similarly, (5.32) implies that \(\sum_{j \notin \mathcal{Y}_{i-1}} d_j = \ell_n + o_p(n)\) uniformly over \(i \leq un^{2/3}\). Therefore,

\[
n^{-1/3} \sum_{i=1}^{k} \frac{\sum_{j \notin \mathcal{Y}_{i-1}} d_j^2 \sum_{j \in \mathcal{Y}_{i-1}} d_j}{\sum_{j \notin \mathcal{Y}_{i-1}} d_j \sum_{j \in [n]} d_j} = \frac{n\sigma_2(n) + o_p(n)}{\ell_n + o_p(n)} n^{-1/3} \sum_{i=1}^{k} \frac{\sum_{j \in \mathcal{Y}_{i-1}} d_j}{\ell_n} \tag{5.46}
\]

and Assumption 1, combined with (5.42), complete the proof. \(\square\)

**Lemma 13.** Condition (5.30a) holds.

**Proof.** The proof follows by using Lemma 12 in (5.40). \(\square\)

### 5.5 Finite dimensional convergence of the ordered component sizes

Note that the convergence of the exploration process in Theorem 8 implies that, for any large \(T > 0\), the \(k\)-largest components explored up to time \(T n^{2/3}\) converge to the \(k\)-largest excursions above past minima of \(B^\lambda_{\alpha,\eta}\) up to time \(T\). Therefore, we can conclude the finite dimensional convergence of the ordered components sizes in the whole graph if we can show that the large components are explored early by the exploration process. The following lemma formalizes the above statement:

**Lemma 14.** Let \(\mathcal{C}^{\geq T}_{\text{max}}\) denote the largest component which is started exploring after time \(T n^{2/3}\). Then, for any \(\delta > 0\),

\[
\lim_{T \to \infty} \limsup_{n \to \infty} \mathbb{P}\left( |\mathcal{C}^{\geq T}_{\text{max}}| > \delta n^{2/3} \right) = 0. \tag{5.47}
\]
Let us first state the two main ingredients to complete the proof of Lemma 14:

**Lemma 15** ([13, Lemma 5.2]). Consider $CM_n(d)$ with $\nu_n < 1$ and let $\mathcal{C}(V_n)$ denote the component containing the vertex $V_n$ where $V_n$ is a random vertex chosen independently of the graph. Then,

$$\mathbb{E}[|\mathcal{C}(V_n)|] \leq 1 + \frac{\mathbb{E}[D_n]}{1 - \nu_n}. \quad (5.48)$$

**Lemma 16.** Define, $\nu_{n,i} = \sum_{j\in \mathcal{V} \backslash \mathcal{V}_{i-1}} d_j (d_j - 1)/\sum_{j\in \mathcal{V} \backslash \mathcal{V}_{i-1}} d_j$. There exists some constant $C_0 > 0$ such that for any $T > 0$,

$$\nu_{n,Tn^{2/3}} = \nu_n - C_0 T n^{-1/3} + o_p(n^{-1/3}). \quad (5.49)$$

**Proof.** Using a similar split up as in (5.39), we have

$$\nu_{n,i} = \nu_n + \frac{\sum_{j\in \mathcal{V} \backslash \mathcal{V}_{i-1}} d_j (d_j - 1)}{\ell_n} - \frac{\sum_{j\in \mathcal{V} \backslash \mathcal{V}_{i-1}} d_j (d_j - 1) \sum_{j\in \mathcal{V} \backslash \mathcal{V}_{i-1}} d_j}{\ell_n \sum_{j\in \mathcal{V} \backslash \mathcal{V}_{i-1}} d_j}. \quad (5.50)$$

Now, (5.5) and (5.6) give that, uniformly over $i \leq Tn^{2/3}$,

$$\frac{\sum_{j\in \mathcal{V} \backslash \mathcal{V}_{i-1}} d_j (d_j - 1)}{\sum_{j\in \mathcal{V} \backslash \mathcal{V}_{i-1}} d_j} = \frac{\sum_{j\in [n]} d_j (d_j - 1) + o_p(n^{2/3})}{\sum_{j\in [n]} d_j + o_p(n^{2/3})} = 1 + o_p(n^{-1/3}), \quad (5.51a)$$

$$\sum_{j\in \mathcal{V} \backslash \mathcal{V}_{i-1}} d_j (d_j - 2) = \left(\frac{\sigma_3}{\mu} - 2\right)(i - 1) + o_p(n^{2/3}). \quad (5.51b)$$

Further, note that $\sigma_3 - 2\mu = \mathbb{E}[D(D - 1)(D - 2)] + \mathbb{E}[D(D - 2)] > 0$, by Assumption 1(iii) and (iv). Therefore, (5.50) gives (5.49) for some constant $C_0 > 0$. ☐

**Proof of Lemma 14.** Let $i_T := \inf\{i \geq Tn^{2/3} : S_n(i) = \inf_{j \leq i} S_n(j)\}$. Thus, $i_T$ denotes the first time we finish exploring a component after time $Tn^{2/3}$. Note that, conditional on the explored vertices up to time $i_T$, the remaining graph $\hat{G}$ is still a configuration model. Let $\nu_n = \sum_{i \in \hat{G}} d_i (d_i - 1)/\sum_{i \in \hat{G}} d_i$ be the criticality parameter of $\hat{G}$. Then, using (5.49), we can conclude that

$$\nu_n \leq \nu_n - C_0 T n^{-1/3} + o_p(n^{-1/3}). \quad (5.52)$$

Take $T > 0$ such that $\lambda - C_0 T < 0$. Thus, with high probability, $\nu_n < 1$. Denote the component corresponding to a randomly chosen vertex from $\hat{G}$ by $\mathcal{C}^{zT}(V_n)$, and the $i$th largest component of $\hat{G}$ by $\mathcal{C}^{zT}_{(i)}$. Also, let $\hat{P}$ denote the probability measure conditioned on $\mathcal{F}_{i_T}$, and $\nu_n < 1$ and let $\hat{E}$ denote the corresponding expectation. Now, for any $\delta > 0$,

$$\hat{P}\left(\sum_{i \geq 1} |\mathcal{C}^{zT}_{(i)}|^2 > \delta^2 n^{4/3}\right) \leq \frac{1}{\delta^2 n^{4/3}} \sum_{i \geq 1} \hat{E}(|\mathcal{C}^{zT}_{(i)}|^2) = \frac{1}{\delta^2 n^{4/3}} \hat{E}(|\mathcal{C}^{zT}(V_n)|) \leq \frac{1}{\delta^2 (-\lambda + C_0 T + o_p(1))}. \quad (5.53)$$

where the second step follows from the Markov inequality and the last step follows by combining Lemma 15 and (5.52). Noting that $\nu_n < 1$ with high probability, we get

$$\limsup_{n \to \infty} \hat{P}\left(|\mathcal{C}_{\text{max}}^{zT}| > \delta n^{2/3}\right) \leq \frac{C}{\delta^2 T}, \quad (5.54)$$

for some constant $C > 0$ and the proof follows. ☐

**Theorem 17.** The convergence in Theorem 1 holds with respect to the product topology.

**Proof.** The proof follows from Theorem 8 and Lemma 14. ☐
5.6 Proof of Theorem 1

The proof of Theorem 1 follows using similar argument as [2, Section 3.3]. However, the proof is a bit tricky since the components are explored in a size-biased manner with sizes being the total degree in the component (not the component sizes as in [2]). For a sequence of random variables \( Y = (Y_i)_{i \geq 1} \) satisfying \( \sum_{i \geq 1} Y_i^2 < \infty \) almost surely, define \( \xi := (\xi_i)_{i \geq 1} \) such that \( \xi_i | Y \sim \text{Exp}(Y_i) \) and the coordinates of \( \xi \) are independent conditional on \( Y \). For \( a \geq 0 \), let \( \mathcal{F}(a) := \{ \xi_i \leq a \} \). Then the size biased point process is defined to be the random collection of points \( \Xi := \{ (\mathcal{F}(\xi_i), Y_i) \} \) (see [2, Section 3.3]). We shall use Lemma 8, Lemma 14 and Proposition 15 from [2]. Let \( \mathcal{C} := \{ \mathcal{C} : \mathcal{C} \) is a component of \( \text{CM}_n(d) \} \). Consider the collection \( \xi := (\xi(\mathcal{C}))_{\mathcal{C} \in \mathcal{C}} \) such that conditional on the values \( (\sum_{k \in \mathcal{C}} d_k, |\mathcal{C}|)_{\mathcal{C} \in \mathcal{C}}, \xi(\mathcal{C}) \) has an exponential distribution with rate \( n^{-2/3} \sum_{k \in \mathcal{C}} d_k \) independently over \( \mathcal{C} \). Then the order in which Algorithm 1 explores the components can be obtained by ordering the components according to their \( \xi \)-value. Recall that \( \mathcal{C}_i \) denotes the \( i \)th explored component by Algorithm 1 and let \( D_i := \sum_{k \in \mathcal{C}_i} d_k \). Define the size biased point process

\[
\Xi_n := \left(n^{-2/3} \sum_{i=1}^n D_i, n^{-2/3} D_i \right)_{i \geq 1}.
\]  

Also define the point processes

\[
\Xi'_n := \left(n^{-2/3} \sum_{i=1}^n |\mathcal{C}_i|, n^{-2/3} |\mathcal{C}_i| \right)_{i \geq 1}, \quad \Xi_{\infty} := \{(l(\gamma), |\gamma|) : \gamma \) an excursion of \( B_{\mu/\eta}^1 \},
\]

where we recall that \( l(\gamma) \) are the left endpoints of the excursions of \( B_{\mu/\eta}^1 \) and \( |\gamma| \) is the length of the excursion \( \gamma \) (see (2.6)). Note that \( \Xi'_n \) is not a size biased point process. However, applying [2, Lemma 8] and Theorem 8, we get \( \Xi'_n \xrightarrow{\ell^2} \Xi_{\infty} \). We claim that

\[
\Xi_n \xrightarrow{\ell^2} 2\Xi_{\infty}.
\]  

To verify the claim, note that (5.5) and Assumption 1 (iii) together imply

\[
\sup_{u \leq t} \left| n^{-2/3} \sum_{i=1}^{\lfloor un^{2/3} \rfloor} d_{i(t)} - \frac{\sigma_2}{\mu} u \right| = \sup_{u \leq t} \left| n^{-2/3} \sum_{i=1}^{\lfloor un^{2/3} \rfloor} d_{i(t)} - 2u \right| \xrightarrow{P} 0,
\]

for any \( t > 0 \) since \( \sigma_2/\mu = \mathbb{E}[D^2]/\mathbb{E}[D] = 2 \). Therefore,

\[
\sum_{\xi(\mathcal{C}) \leq s} D(\mathcal{C}) - 2 \sum_{\xi(\mathcal{C}) \leq s} |\mathcal{C}| = o_P(n^{2/3})
\]

Thus, (5.57) follows using (5.58) and (5.59). Now, the point process \( 2\Xi_{\infty} \) satisfies all the conditions of [2, Proposition 15] as shown by Aldous. Thus, [2, Lemma 14] gives

\[
\{D_{i(t)}\}_{i \geq 1} \text{ is tight in } \ell^2.
\]

This implies that \( (n^{-2/3} |\mathcal{C}_{i(t)}|)_{i \geq 1} \) is tight in \( \ell^2 \) by simply observing that \( |\mathcal{C}_i| \leq \sum_{k \in \mathcal{C}_i} d_k + 1 \). Therefore, the proof of Theorem 1 is complete using Theorem 17.

5.7 Proof of Theorem 2

The proof of Theorem 2 is completed in two separate lemmas. In Lemma 18 we first show that the convergence in Theorem 2 holds with respect to the \( \ell^2 \times \mathbb{N}^\infty \) topology. The tightness of \( (\mathbb{Z}_n)_{n \geq 1} \) with respect to the \( \mathbb{U}_\mathbb{Z}^0 \) topology is ensured in Lemma 19 and Theorem 2 follows.
Lemma 18. Let $N^\lambda_n(k)$ be the number of surplus edges discovered up to time $k$ and $\bar{N}^\lambda_n(u) = N^\lambda_n(\lfloor un^{2/3} \rfloor)$. Then, as $n \to \infty$,
\[ \bar{N}^\lambda_n \overset{d}{\to} N^\lambda, \] (5.61)
where $N^\lambda$ is defined in (2.7).

Proof. Recall the definitions of $a, b, A_k, B_k, C_k, S_k$ from Section 5.1. Recall also that $A_k := |A_k|$, $B_k := |B_k|$, $C_k := |C_k|$, $U_k := |S_k|$, $c_{(k+1)} := ((B_k) + |C_k|)/2$ as in Section 5.1. We have $A_k = S_n(k) - \min_{j \leq k} S_n(j) + 2$. From Lemma 7 we can conclude that
\[ \mathbb{E}[c_{(k+1)}|\mathcal{F}_k] = \frac{A_k}{\mu n} + O_p(n^{-1}). \] (5.62)

The counting process $\bar{N}^\lambda_n$ has conditional intensity (conditioned on $\mathcal{F}_{k-1}$) given by (5.62). Writing the conditional intensity in (5.62) in terms of $\bar{S}_n$, we get that the conditional intensity of the rescaled process $\bar{N}^\lambda_n$ is given by
\[ \frac{1}{\mu} |\bar{S}_n(u) - \min_{\bar{u} \leq u} \bar{S}_n(\bar{u})| + o_p(1). \] (5.63)

Denote by $\bar{W}_n(u) := \bar{S}_n(u) - \min_{\bar{u} \leq u} \bar{S}_n(\bar{u})$ which is the reflected version $\bar{S}_n$. By Theorem 1,
\[ \bar{W}_n \overset{d}{\to} W^\lambda, \] (5.64)
where $W^\lambda$ is as defined in (2.6). Therefore, we can assume that there exists a probability space such that $\bar{W}_n \to W^\lambda$ almost surely. Using [18, Theorem 1; Chapter 5.3], the continuity of the sample paths of $W^\lambda$, we conclude that
\[ \bar{N}^\lambda_n \overset{d}{\to} N^\lambda, \] (5.65)
where $N^\lambda$ is defined in (2.7).

Lemma 19. The vector $(Z_n)_{n \geq 1}$ is tight with respect to the $\mathcal{U}_\lambda^0$ topology.

The proof of Lemma 19 makes use of the following crucial estimate of the probability that a component with small size has very large number of surplus edges.

Lemma 20. Assume that $\lambda < 0$. Let $V_n$ denote a vertex chosen uniformly at random, independent of the graph $\operatorname{CM}_n(d)$ and let $\mathcal{C}(V_n)$ denote the component containing $V_n$. Let $\delta_k = \delta k^{-0.12}$. Then, for $\delta > 0$ (small),
\[ \Pr(\operatorname{SP}(\mathcal{C}(V_n)) \geq K, |\mathcal{C}(V_n)| \in (\delta K n^{2/3}, 2\delta K n^{2/3})) \leq \frac{C \sqrt{\delta}}{n^{1/3} K^{1/3}}, \] (5.66)
where $C$ is a fixed constant independent of $n, \delta, K$.

Proof of Lemma 19. To simplify the notation, we write $Y^n_i = n^{-2/3}|\mathcal{C}(i)|$ and $N^n_i := \# \{\text{surplus edges in } \mathcal{C}(i)\}$. Let $Y^n, N^n$ denote the distributional limits of $Y^n_i$ and $N^n_i$ respectively. Recall from Remark 1 that $Z(\lambda)$ is almost surely $\mathcal{U}^0$-valued. Using the definition of $d_\mathcal{U}$ from (2.3) and Lemma 18, the proof of Lemma 19 is complete if we can show that, for any $\eta > 0$
\[ \lim_{\varepsilon \to 0} \lim_{n \to \infty} \Pr( \sum_{Y^n_i \leq \varepsilon} Y^n_i N^n_i > \eta) = 0. \] (5.67)
First, consider the case $\lambda < 0$. For every $\eta, \varepsilon > 0$ sufficiently small

$$
P\left( \sum_{i=1}^{n} Y_i^n N_i^n > \eta \right) \leq \frac{1}{\eta} \mathbb{E} \left[ \sum_{i=1}^{n} Y_i^n N_i^n 1\{Y_i^n \leq \varepsilon \} \right] = \frac{n^{-2/3}}{\eta} \mathbb{E} \left[ \sum_{i=1}^{n} |G_i| N_i^n 1\{|G_i| \leq \varepsilon n^{2/3} \} \right]
$$

$$
= \frac{n^{1/3}}{\eta} \mathbb{E} \left[ \text{SP}(\mathcal{G}(V_n)) 1\{|\mathcal{G}(V_n)| \leq \varepsilon n^{2/3} \} \right]
$$

$$
= \frac{n^{1/3}}{\eta} \sum_{k=1}^{\infty} \sum_{i \geq \log_2(1/(k0.12\varepsilon))} \mathbb{P} \left( \text{SP}(\mathcal{G}(V_n)) \geq k, |\mathcal{G}(V_n)| \in (2^{-(i+1)} k^{-0.12} n^{2/3}, 2^{-i} k^{-0.12} n^{2/3}] \right)
$$

$$
\leq C \frac{1}{\eta} \sum_{k=1}^{\infty} \frac{1}{k^{1.1}} \sum_{i \geq \log_2(1/(k0.12\varepsilon))} 2^{-(1/2)i} \leq C \frac{\sqrt{\varepsilon}}{\eta} \sum_{k=1}^{\infty} \frac{\sqrt{\varepsilon}}{k^{1.04}} = O(\sqrt{\varepsilon}).
$$

where the last but one step follows from Lemma 20. Therefore, (5.67) holds when $\lambda < 0$. Now consider the case $\lambda > 0$. For $T > 0$ (large), let

$$
\mathcal{K}_n := \{ i : Y_i^n \leq \varepsilon, \mathcal{G}(i) \text{ is explored before } Tn^{2/3} \}.
$$

Then, by applying the Cauchy-Schwarz inequality,

$$
\sum_{i \in \mathcal{K}_n} Y_i^n N_i^n \leq \left( \sum_{i \in \mathcal{K}_n} (Y_i^n)^2 \right)^{1/2} \times \left( \sum_{i \in \mathcal{K}_n} (N_i^n)^2 \right)^{1/2}
$$

$$
\leq \left( \sum_{i \in \mathcal{K}_n} (Y_i^n)^2 \right)^{1/2} \times \left( \text{# surplus edges explored before } Tn^{2/3} \right)
$$

Using similar ideas as the proof of Lemma 14, we can run the exploration process till $Tn^{2/3}$ and the unexplored graph becomes a configuration model with negative criticality parameter for large $T > 0$, by (5.49). Thus, the proof can be completed using (5.70), the $l^2$ convergence of the component sizes given by Theorem 1, and Lemma 18.

**Proof of Lemma 20.** To complete the proof of Lemma 20, we shall use martingale techniques coupled with Lemma 15. Fix $\delta > 0$ (small). First we describe another way of exploring $\mathcal{G}(V_n)$ which turns out to be convenient to work with.

**Algorithm 2 (Exploring $\mathcal{G}(V_n)$).** Consider the following exploration of $\mathcal{G}(V_n)$:

(S0) Initialize all half-edges to be alive. Choose a vertex from $[n]$ uniformly at random and declare all its half-edges active.

(S1) In the next step, take any active half-edge and pair it uniformly with another alive half-edge. Kill these paired half-edges. Declare all the half-edges corresponding to the new vertex (if any) active. Keep repeating (S1) until the set of active half-edges is empty.

Unlike Algorithm 1, we need not see a new vertex at each stage and we explore only two half-edges at each stage. Recall that we denote by $D_n$, the degree of a vertex chosen uniformly at random independently of the graph. Define the exploration process $s'_n$ by

$$
s'_n(0) = D_n, \quad s'_n(l) = \sum_{i \in [n]} d_i I_i^n(l) - 2l,
$$

where $I_i^n(l) = 1_{\{i \in \mathcal{V}_l\}}$ and $\mathcal{V}_l$ is the vertex set discovered up to time $l$. Therefore, $s'_n(l)$ counts the number of active half-edges at time $l$, until $\mathcal{G}(V_n)$ is explored. Note that $\mathcal{G}(V_n)$ is explored when $s'_n$ hits zero. We shall use $C$ to denote a positive constant that can be different in different equations.
In this proof, $\mathcal{F}_l$ shall be used to denote the sigma-field containing information revealed up to stage $l$ by Algorithm 2. For $H > 0$, let
\[
\gamma := \inf \{ l \geq 1 : s'_n(l) \geq H \text{ or } s'_n(l) = 0 \} \land 2\delta n^{2/3}. \tag{5.72}
\]

Note that
\[
\mathbb{E} \left[ s'_n(l + 1) - s'_n(l) \mid (I_i^n(l))_{i=1}^n \right] = \sum_{i \in [n]} d_i \mathbb{P} \left( i \notin \mathcal{F}_l, i \in \mathcal{F}_{l+1} \mid (I_i^n(l))_{i=1}^n \right) - 2
\]
\[
= \frac{\sum_{i \notin \mathcal{F}_l} d_i^2}{\ell_n - 2l - 1} - 2 \leq \frac{\sum_{i \notin [n]} d_i^2}{\ell_n - 2l - 1} - 2 \leq \frac{2l + 1}{\ell_n - 2l - 1} \sum_{i \in [n]} d_i^2 \leq 0 \tag{5.73}
\]
uniformly over $l \leq 2\delta n^{2/3}$ for all small $\delta > 0$ and large $n$, where the last step follows from the fact that $\lambda < 0$. Therefore, $\{s'_n(l)\}_{l=1}^{2\delta n^{2/3}}$ is a super-martingale. The optional stopping theorem now implies
\[
\mathbb{E} [D_n] \geq \mathbb{E} [s'_n(\gamma)] \geq H \mathbb{P} \left( s'_n(\gamma) \geq H \right). \tag{5.74}
\]
Thus,
\[
\mathbb{P} \left( s'_n(\gamma) \geq H \right) \leq \frac{\mathbb{E} [D_n]}{H}. \tag{5.75}
\]

We shall put $H = n^{1/3}K^{1-1/\sqrt{\delta}}$. To simplify the writing, we write $s'_n[0, t] \in A$ to denote that $s'_n(l) \in A$, for all $l \in [0, t]$. Notice that
\[
\mathbb{P} \left( \text{SP}(\mathcal{E}(V_n)) \geq K, |\mathcal{E}(V_n)| \in (\delta_K n^{2/3}, 2\delta_K n^{2/3}) \right)
\leq \mathbb{P} \left( s'_n(\gamma) \geq H \right) + \mathbb{P} \left( \text{SP}(\mathcal{E}(V_n)) \geq K, s'_n[0, 2\delta_K n^{2/3}] < H, s'_n[0, \delta_K n^{2/3}] > 0 \right). \tag{5.76}
\]

Now,
\[
\mathbb{P} \left( \text{SP}(\mathcal{E}(V_n)) \geq K, s'_n[0, 2\delta_K n^{2/3}] < H, s'_n[0, \delta_K n^{2/3}] > 0 \right)
\leq \sum_{1 \leq l_1 < \ldots < l_K \leq 2\delta_K n^{2/3}} \mathbb{P} \left( \text{surprises occur at times } l_1, \ldots, l_K, s'_n[0, 2\delta_K n^{2/3}] < H, s'_n[0, \delta_K n^{2/3}] > 0 \right)
\leq \sum_{1 \leq l_1 < \ldots < l_K \leq 2\delta_K n^{2/3}} \mathbb{E} \left[ \mathbb{1}_{\{0 < s'_n(l_k-1) < H, \text{SP}(l_k-1) = K-1\}} \right], \tag{5.77}
\]
where
\[
Y = \mathbb{P} \left( K^{th} \text{ surplus occurs at time } l_K, s'_n[l_K, 2\delta_K n^{2/3}] < H, s'_n[l_K, \gamma] > 0 \mid \mathcal{F}_{l_K-1} \right)
\leq \frac{CK^{1-1/3}n^{1/3}}{\ell_n \sqrt{\delta}} \leq \frac{CK^{1-1}}{n^{2/3} \sqrt{\delta}}. \tag{5.78}
\]

Therefore, using induction, (5.76) yields,
\[
\mathbb{P} \left( \text{SP}(\mathcal{E}(V_n)) \geq K, s'_n[0, 2\delta_K n^{2/3}] < H, s'_n[0, \delta_K n^{2/3}] > 0 \right)
\leq C \left( K^{1-1/3} / \delta n^{2/3} \right)^K \mathbb{P} \left( |\mathcal{E}(V_n)| \geq l_1 \right) \leq C \frac{\delta^{K/2}}{K^{1-1}n^{2/3}} \mathbb{E} [ |\mathcal{E}(V_n)| ], \tag{5.79}
\]
where we have used the fact that \( \# \{ 1 \leq l_2, \ldots, l_k \leq 2\delta n^{2/3} \} = (2\delta n^{2/3})^{K-1} / (K - 1)! \) and have used the Stirling approximation for \((K - 1)!\) in the last step. Since \( \lambda \leq 0 \), we can use Lemma 15 to conclude that for all sufficiently large \( n \)

\[
\mathbb{E} \left[ \left| \mathcal{E}(V_n) \right| \right] \leq C n^{1/3},
\]

for some constant \( C > 0 \) and we get the desired bound for (5.76). The proof of Lemma 20 is now complete .

6 Vertices of degree \( k \)

In this section, we compute the number of vertices of degree \( k \) in each connected component at criticality. This will be useful in Section 7 and 8. Such an estimate was proved in Janson and Luczak [14, Theorem 2.4] for supercritical graphs under a stronger moment assumption.

**Lemma 21.** Denote by \( N_k(t) \) the number of vertices of degree \( k \) discovered up to time \( t \). For any \( t > 0 \),

\[
\sup_{u \leq t} \left| n^{-2/3} N_k(un^{2/3}) - \frac{kn_k}{\ell_n} u \right| = O_p((kn^{1/3})^{-1}).
\]

**Proof.** By setting \( \alpha = 2/3 \) and \( f_n(i) = 1_{\{d_i = k\}} \) in Proposition 29 (see Appendix A) we can directly conclude that

\[
\sup_{u \leq t} \left| n^{-2/3} N_k(un^{2/3}) - \frac{kn_k}{\ell_n} u \right| \xrightarrow{P} 0.
\]

However, if we repeat the same arguments leading to the proof of Proposition 29 (see in particular (A.7)) we shall get

\[
\mathbb{P} \left( \sup_{u \leq t} \left| n^{-2/3} N_k(un^{2/3}) - \frac{kn_k}{\ell_n} u \right| > A \right) \leq \frac{3(k^3 s^2 r_k^2 (\mathbb{E}[d])^2 + \sqrt{k^{1/3} s (\mathbb{E}[d])})}{A} + o(1).
\]

Now, we can use the finite third-moment assumption to conclude that the numerator in the right hand side can be taken to be uniform over \( k \). Thus, the proof follows.

Define \( v_k(G) := \) the number of vertices of degree \( k \) in the connected graph \( G \). As a corollary to Lemma 21 and (5.47) we can deduce that

\[
v_k(\mathcal{E}(\Omega)) = \frac{k r_k}{\mathbb{E}[d]} |\mathcal{E}(\Omega)| + O_p((k^{-1} n^{1/3})).
\]

Moreover, the following also holds:

**Lemma 22.** For each \( k \geq 1 \) denote by \( V_k^n := (n^{-2/3} v_k(\mathcal{E}_j))_{j \geq 1} \). Then, \( \{\text{ord}(V_k^n)\}_{n \geq 1} \) is tight in \( l_2^2 \).

**Proof.** Note that for any \( j \geq 1 \), \( v_k(\mathcal{E}_j) \leq |\mathcal{E}_j| \) uniformly over \( k \). The proof now follows from (6.4) and \( l_2^2 \) tightness of the component sizes given in Theorem 1.

**Remark 7.** Denote by \( V^n := (n^{-2/3} v_k(\mathcal{E}_j))_{k,j \geq 1} \). Then \( \{\text{ord}(V^n)\}_{n \geq 1} \) is also tight in \( l_2^2 \).


7 Critical Percolation

7.1 Percolation on Configuration Model

Let \( p = p_n \in (0, 1) \) be the percolation parameter. Recall the notation \( \text{CM}_n(d, p) \) for the random graph obtained after deleting edges of \( \text{CM}_n(d) \) independently with probability \( 1 - p \). Suppose, \( d' \) is the random degree sequence obtained after percolation. Fountoulakis [11] showed that, given \( d' \), the law of \( \text{CM}_n(d, p) \) is same as the law of \( \text{CM}_n(d') \).

Algorithm 3. We shall use the following construction of \( \text{CM}_n(d, p) \) due to Janson [12]:

(S1) For each half-edge \( e \), let \( v_e \) be the vertex to which \( e \) is attached. With probability \( 1 - \sqrt{p} \), one detaches \( e \) from \( v_e \) and associate \( e \) to a new vertex \( v' \). Color the new vertex \( \text{red} \). This is done independently for every existing half-edge. Let \( n_+ \) be the number of red vertices created and \( \tilde{n} = n + n_+ \). Suppose, \( \tilde{d} = (\tilde{d}_i)_{i \in [\tilde{n}]} \) is the new degree sequence obtained by the above procedure, i.e. \( \tilde{d}_i \sim \text{Bin}(d_i, \sqrt{p}) \) for \( i \in [n] \) and \( \tilde{d}_i = 1 \) for \( i \in [\tilde{n}] \setminus [n] \).

(S2) Construct \( \text{CM}_\tilde{n}(\tilde{d}) \).

(S3) Delete all the red vertices.

The obtained multigraph has the law of \( \text{CM}_n(d, p) \).

Remark 8. It was argued in [12] that the obtained multigraph also has the same distribution as \( \text{CM}_n(d, p) \) if we replace (S3) by

(S3') Instead of deleting red vertices, choose any \( n_+ \) degree one vertices uniformly at random and delete them.

Remark 9. The construction of \( \text{CM}_n(d) \) in Algorithm 3 consists of two stages of randomization, the first one is described by (S1), and the second one by (S2). We shall consider the following probability space to describe the randomization arising from Algorithm 3 (S1): Suppose we have a sequence of degree sequences \( \{\tilde{d}_i\}_{i \geq 1} \). Let \( P^n_p \) denote the probability measure induced on \( \mathbb{N}^{\infty} \) by Algorithm 3 (S1). Denote the product measure of \( \{P^n_p\}_{n \geq 1} \) by \( \mathbb{P}_p \). Thus (S1) is performed independently on \( d = d(n) \) as \( n \) varies. All the almost sure statements in this section will be with respect to the probability measure \( \mathbb{P}_p \).

7.2 Proof of Theorem 3

We now consider the critical window corresponding to percolation. The goal is to prove Theorem 3. Let \( \tilde{n}_j \) and \( \tilde{n}_j \) be the number of vertices of degree \( j \) before and after performing Algorithm 3 (S1) respectively. Further let

\[
\tilde{\nu}_n = \frac{\sum_{i \in [\tilde{n}]} \tilde{d}_i \left( \tilde{d}_i - 1 \right)}{\sum_{i \in [\tilde{n}]} \tilde{d}_i}.
\]  

(7.1)

For convenience we write \( r_j = \mathbb{P}(D = j) \). Denote by \( \tilde{n}_{jl} \), the number of vertices that had degree \( l \) before and have degree \( j \) after Algorithm 3 (S1). Therefore, \( \tilde{n}_{jl} \sim \text{Bin}(n_l, b_{lj} (\sqrt{p_n})) \), where \( b_{lj} (\sqrt{p_n}) = \binom{l}{j} (\sqrt{p_n})^j (1 - \sqrt{p_n})^{l-j} \). Using the strong law of large numbers for triangular arrays, note that \( \mathbb{P}_p \) almost surely, \( \tilde{n}_{jl} = n_l b_{lj} (\sqrt{p_n}) + o(n_l) = n r_l b_{lj} (\sqrt{p_n}) + o(n_l) \). Now, \( \sum_{l \geq 1} |n_l / n - r_l| \to 0 \) and therefore, for all \( j \geq 2, \mathbb{P}_p \) almost surely

\[
\frac{\tilde{n}_j}{n} = \sum_{l=j}^{\infty} \frac{\tilde{n}_{jl}}{n} = \sum_{l=j}^{\infty} r_l b_{lj} (\sqrt{p_n}) + o(1).
\]  

(7.2)
Also \( n_+ = \sum_{i \in [n]} (d_i - \hat{d}_i) \sim \text{Bin}(\ell_n, 1 - \sqrt{p_n}) \). Therefore, using the similar arguments as (7.2) again, \( \mathbb{P}_p \) almost surely,

\[
\frac{n_+}{n} = \mathbb{E}(D)(1 - \sqrt{p_n}) + o(1),
\]

(7.3)

\[
\frac{n_1}{n} = \sum_{i=1}^{\infty} \frac{n_{1i}}{n} + \frac{n_+}{n} = \sum_{i=1}^{\infty} \frac{n_{1i}}{n} + \mathbb{E}(D)(1 - \sqrt{p_n}) + o(1),
\]

(7.4)

and

\[
\frac{n}{n} = 1 + \frac{n_+}{n} = 1 + \mathbb{E}(D)(1 - \sqrt{p_n}) + o(1).
\]

(7.5)

Denote \( \tilde{r}_l = \mathbb{P}(\tilde{D} = l) = \lim_{n \to \infty} \frac{n_l}{n} \). We shall denote by \( \hat{G}_{(j)} \), the \( j^{th} \) largest component of \( \text{CM}_{n}(\hat{d}) \).

**Remark 10.** The idea of the proof of Theorem 3 is as follows. We show that \( \hat{d} \), under Assumption 2, satisfies Assumption 1 \( \mathbb{P}_p \) almost surely and then estimate the number of vertices to be deleted from each component using Lemma 21. Since deleting a degree one vertex does not break up any component, we can just subtract this from the component sizes of \( \text{CM}_{n}(\hat{d}) \) to get the component sizes of \( \text{CM}_{n}(d,p) \). Since the degree one vertices do not get involved in surplus edges, deleting degree one vertices do not change the surplus edges also.

**Lemma 23.** The statements below are true \( \mathbb{P}_p \) almost surely:

1. **Under Assumption 2 (i) and for** \( r = 1, 2, 3 \),

\[
\frac{1}{n} \sum_{i \in [n]} \frac{d_i^r}{n} \xrightarrow{\mathbb{P}} \mathbb{E}[\hat{D}^r].
\]

(7.6)

2. **Under Assumption 2,**

\[
\tilde{\nu}_n = 1 + \frac{\lambda n^{-1/3} + o(n^{-1/3})}{n}.
\]

(7.7)

**Proof.** We shall make use of [15, Corollary 2.27]. Suppose \( Z_1, Z_2, ..., Z_N \) are independent random variables with \( Z_i \) taking values in \( \Lambda_i \) and \( f : \prod_{i=1}^{N} \Lambda_i \to \mathbb{R} \) satisfies the following: If two vectors \( z, z' \in \prod_{i=1}^{N} \Lambda_i \) differ only in the \( j^{th} \) coordinate, then \( |f(z) - f(z')| \leq c_i \) for some constant \( c_i \). Then, for any \( t > 0 \), the random variable \( X = f(Z_1, Z_2, \ldots, Z_N) \) satisfies

\[
\mathbb{P}\left( |X - \mathbb{E}[X]| > t \right) \leq 2 \exp\left( -\frac{t^2}{2 \sum_{i=1}^{N} c_i^2} \right).
\]

(7.8)

Now let \( I_{ij} \) denote the indicator of the \( j^{th} \) half-edge corresponding to vertex \( i \) to be kept after the explosion. Then \( I_{ij} \sim \text{Ber}(\sqrt{p_n}) \) independently for \( j \in [d_i], i \in [n] \). Let

\[
I := (I_{ij})_{j \in [d_i], i \in [n]} \quad \text{and} \quad f_1(I) := \sum_{i \in [n]} d_i (d_i - 1).
\]

(7.9)

Note that \( f_1(I) = \sum_{i \in [n]} d_i (d_i - 1) \) since the degree one vertices do not contribute to the sum. One can check that, by changing the status of one half-edge corresponding to vertex \( k \), we can change \( f_1(\cdot) \) by at most \( 2(d_k + 1) \). Therefore (7.8) yields

\[
\mathbb{P}_p \left( \sum_{i \in [n]} d_i (d_i - 1) - p_n \sum_{i \in [n]} d_i (d_i - 1) > t \right) \leq 2 \exp\left( -\frac{t^2}{8 \sum_{i \in [n]} d_i (d_i + 1)^2} \right).
\]

(7.10)
By setting \( t = n^{1/2+\varepsilon} \) for some suitably small \( \varepsilon > 0 \), using the finite third moment conditions and the Borel-Cantelli lemma we conclude that \( \mathbb{P}_p \) almost surely,

\[
\sum_{i \in [n]} \hat{d}_i(d_i - 1) = p_n \sum_{i \in [n]} d_i(d_i - 1) + O(n^{1/2+\varepsilon}), \tag{7.11}
\]

in particular,

\[
\sum_{i \in [n]} \hat{d}_i(d_i - 1) = \sum_{i \in [n]} \hat{d}_i(d_i - 1) = p_n \sum_{i \in [n]} d_i(d_i - 1) + o(n^{2/3}). \tag{7.12}
\]

Similarly, take \( f_2(I) = \sum_{i \in [n]} \hat{d}_i(d_i - 1)(d_i - 2) \) and note that changing the status of one bond changes \( f_2(\cdot) \) by at most \( |2(d_k + 1)|^2 \). Thus, (7.8) gives

\[
\mathbb{P}_p \left( \left| f_2(I) - p_n^{3/2} \sum_{i \in [n]} d_i(d_i - 1)(d_i - 2) \right| > t \right) \leq 2 \exp \left( - \frac{t^2}{32 \sum_{i \in [n]} d_i(d_i + 1)^2} \right) \leq \exp \left( - \frac{t^2}{32d_{\max}(d_{\max} + 1) \sum_{i \in [n]} (d_i + 1)^3} \right),
\]

which implies that, \( \mathbb{P}_p \) almost surely,

\[
\sum_{i \in [n]} \hat{d}_i(d_i - 1)(d_i - 2) = \sum_{i \in [n]} d_i(d_i - 1)(d_i - 2) = p_n^{3/2} \sum_{i \in [n]} d_i(d_i - 1)(d_i - 2) + o(n), \tag{7.14}
\]

since \( d_{\max}^2 \sum_{i \in [n]} (d_i + 1)^3 = o(n^{5/3}) \). Now, to prove Lemma 23 (1), note that the case \( r = 1 \) follows by simply observing that \( \sum_{i \in \mathcal{N}} \hat{d}_i = \sum_{i \in [n]} d_i \). The cases \( r = 2, 3 \) follow from (7.12) and (7.14). Finally, to see Lemma 23 (2), note that

\[
\hat{\nu}_n = \frac{\sum_{i \in [n]} \hat{d}_i(d_i - 1)}{\sum_{i \in [n]} d_i} = \frac{p_n \sum_{i \in [n]} d_i(d_i - 1) + o(n^{2/3})}{\sum_{i \in [n]} d_i} \tag{7.15}
\]

by (7.12) and this completes the proof of Lemma 23.

To conclude Theorem 3 we also need to estimate the number of deleted vertices from each component. Recall from Remark 8 that \( \text{CM}_n(d, p_n(\lambda)) \) can be obtained from \( \text{CM}_n(d) \) by deleting relevant number of degree one vertices uniformly at random. Let \( \nu_1^d(\mathcal{G}_\ell(\lambda)) \) be the number of degree one vertices of \( \mathcal{G}_\ell(\lambda) \) that are deleted while creating \( \text{CM}_n(d, p_n(\lambda)) \) from \( \text{CM}_n(d) \). Since the vertices are to be chosen uniformly from all degree one vertices, the number of vertices to be deleted from \( \mathcal{G}_\ell(\lambda) \) is asymptotically the total number of degree one vertices in \( \mathcal{G}_\ell(\lambda) \) times the proportion of degree one vertices to be deleted. Therefore,

\[
\nu_1^d(\mathcal{G}_\ell(\lambda)) = \frac{n+\nu_1(\mathcal{G}_\ell(\lambda)) + o_p(n^{2/3})}{n_\ell} = \frac{n_\ell^1}{n_\ell} \sum_{k=0}^{\infty} \lambda k n_k \left| \mathcal{G}_\ell(\lambda) \right| + o_p(n^{2/3})
\]

\[
= \frac{n_\ell^1}{n_\ell} \left| \mathcal{G}_\ell(\lambda) \right| + o_p(n^{2/3}) = \frac{\mathbb{E}[D](1 - \sqrt[3]{p_n})}{\mathbb{E}[D]} \left| \mathcal{G}_\ell(\lambda) \right| + o_p(n^{2/3}) \tag{7.16}
\]

where the third equality follows from (6.4). The proof of Theorem 3 is now complete by using the \( \ell_2^2 \) convergence in Lemma 22, (7.16) and Remark 10.
8 Convergence at multiple locations

We shall prove Theorem 4 in this section. For each percolation cluster let the degree deficiency of the cluster be the number of half-edges of the component that were deleted by percolation. Firstly, we show that the degree deficiency of each component is approximately proportional to the component sizes. We couple the graphs $\text{CM}_n(d, p_n(\lambda))$ for $\lambda \geq \lambda_0$ such that the degree deficiencies evolve like an approximate multiplicative coalescent. Then, we describe an exact multiplicative coalescent that is close to the original process. Finally, we conclude the proof of Theorem 4 using known properties of the multiplicative coalescent.

8.1 Estimate of the degree deficiency

Recall that $\text{CM}_n(\bar{d})$ is the graph obtained before deleting the red vertices while constructing $\text{CM}_n(d, p_n(\lambda))$ as described in Algorithm 3. $\tilde{G}^{(i)}$ denotes the $i^{th}$ largest component of $\text{CM}_n(\bar{d})$ and let $\tilde{C}^{p,(i)}$ denote the residual part of $\tilde{G}^{(i)}$ after applying Algorithm 3 (S3). Denote by $(d'_k)_{k \in [n]}$ the degree sequence of $\text{CM}_n(d, p_n(\lambda))$ and define the degree-deficiency $\mathcal{D}_i$ of component $\tilde{C}^{p,(i)}$ to be $\sum_{k \in \tilde{C}^{p,(i)}} (d_k - d'_k)$. Whenever a half-edge is detached from a vertex in Algorithm 3 (S1), we say that a hole is created on that vertex. Note that

$$\mathcal{D}_i = \sum_{k \in \tilde{C}^{p,(i)}} (d_k - d'_k) = H_i + R_i, \quad (8.1)$$

where $R_i$ denotes the number of degree one vertices deleted from $\tilde{G}^{(i)}$ to obtain $\tilde{C}^{p,(i)}$ and $H_i$ denotes the number of holes in $\tilde{C}^{p,(i)}$. We shall now show that the deficiency of a component is approximately proportional to its size. For that we need to estimate $H_i$ and $R_i$, as in the following lemma:

Lemma 24. There exist constants $\kappa_1, \kappa_2 > 0$ such that

$$H_i = \kappa_1 |\tilde{C}^{p,(i)}| + o_p(n^{2/3}), \quad (8.2a)$$

and

$$R_i = \kappa_2 |\tilde{C}^{p,(i)}| + o_p(n^{2/3}). \quad (8.2b)$$

Therefore, for some constant $\kappa > 0$,

$$\mathcal{D}_i = \kappa |\tilde{C}^{p,(i)}| + o_p(n^{2/3}). \quad (8.3)$$

Proof. We use the following notation:

- $n_{k,l} := \#\{v \in [n] : d_v = k, \tilde{d}_v = l\}$,
- $n_{k,l}(\mathcal{C}) := \#\{v \in [n] : d_v = k, \tilde{d}_v = l, v \in \mathcal{C}\}$,
- $\tilde{n}_l(\mathcal{C}) := \#\{v \in [n] : \tilde{d}_v = l, v \in \mathcal{C}\}$.

First, note that $n_{k,l} = \sum_{i=1}^{\infty} \mathbb{I}(X_i = l)$ where $X_i \sim \text{Bin}(k, \sqrt{p_n})$. Therefore,

$$\frac{n_{k,l}}{n} \xrightarrow{p} p_k \left(\begin{array}{c} k \\ l \end{array} \right) (1 - \sqrt{p})^{r-1} = p_{k,l}, \quad (8.4)$$

where $p = 1/\nu$. Fix any $l \geq 1$. Now, conditioned on $\tilde{n}_l(\tilde{G}^{(i)})$ and $(n_{k,l})_{k \geq 1}$, the vector $(n_{k,l}(\tilde{G}^{(i)}))_{k \geq 1}$ has a multivariate hypergeometric distribution with sample of size $\tilde{n}_l(\tilde{G}^{(i)})$, population of size $\tilde{n}_l = \sum_k n_{k,l}$ that is partitioned into parts of sizes $(n_{k,l})_{k \geq 1}$ and $n_{k,l}(\tilde{G}^{(i)})$ corresponds to an occurrence from the $(k,l)^{th}$ element of the partition which has size $n_{k,l}$. Denote by $\mathbb{P}_2(\cdot), \mathbb{E}_2(\cdot), \text{Var}_2(\cdot)$ the
the red vertices of a component can be generated uniformly
imply the finite-dimensional convergence of
where the error term can be chosen to be independent over $k, l$ since $n_{k,l}/\bar{n}_l \leq 1$. By Chebyshev’s inequality,
where the error term is over $k, l$. Therefore,
where the last equality follows by using the Cauchy-Schwart inequality, $d_{\max} = o(n^{1/3})$ and the tightness of $n^{-2/3} |\tilde{\gamma}(i)|$. Now, using (6.4), we get
Also $\sum_{l=1}^{d_{\max}} l^{-1} \sum_{j=0}^{d_{\max}} (k-l)p_{k,l} = O(d_{\max}) = o(n^{1/3})$, where we have used the fact that $\sum_{k\geq l}(k-l)p_{k,l} = O(l)$, Thus,
for some $\kappa’ > 0$. Now, the proof of (8.2a) follows from (7.16). To see (8.2b), let $\tilde{n}_1$ be the number of degree one vertices in $\text{CM}_n(d)$, $n_{1(i)}$ the number of degree one vertices in $\tilde{\gamma}(i)$, $n_+ \text{ the number of red vertices in } \text{CM}_n(d)$. By Remark 8 the red vertices of a component can be generated uniformly at random. Thus the central limit theorem yields
Using (6.4) to estimate the number of degree one vertices, we conclude
for some $\kappa’_2 > 0$ and the proof follows by applying (7.16) again.

From here onwards, we shall augment $\lambda$ with some previously defined notation to emphasize the location parameter in the critical window.

**Theorem 25.** Denote by $\mathbf{D}_n(\lambda) := (\mathcal{G}_j(\lambda))_{j \geq 1}$ the ordered version of $(\mathcal{G}_j(\lambda))_{j \geq 1}$. Then, for every $\lambda \in \mathbb{R}$,

$$n^{-2/3} \mathbf{D}_n(\lambda) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \kappa \sqrt{\mathbf{P}(|\tilde{\gamma}_j^2|)})_{j \geq 1}$$

(8.12)

with respect to the $\ell^2$ topology, where $(|\tilde{\gamma}_j^2|)_{j \geq 1}$ are the lengths of the Brownian excursions defined in (3.10) and $p = 1/\nu$.

**Proof.** Observe that Lemma 24 and Theorem 3 imply the finite-dimensional convergence of $\mathbf{D}_n(\lambda)$. The proof of $\ell^2$ tightness is similar to Section 5.6. □
8.2 Coupling

In this section, we describe the following dynamic process which generates \((CM_n(d, p_n(\lambda)))_{\lambda \in \mathbb{R}}\):

(C1) Suppose that, between every two free half-edges \(e \neq f\), there is an \(\text{Exp}(1/(2\ell_n - 3))\) clock \(\xi(e, f)\) and that all these clocks across pairs are independent. Whenever a clock rings, the corresponding half-edges are paired and all the other exponential clocks corresponding to these two half-edges are discarded. Let \(G_n(\lambda)\) be the graph obtained by upto time \(-\log(1 - p_n(\lambda))\).

The following lemma ensures that this dynamic process preserves the coupling described before Theorem 4.

Lemma 26. The graph \(G_n(\lambda)\) is distributed as \(CM_n(d, p_n(\lambda))\) for any \(\lambda\). Further, for any \(\lambda < \lambda'\), \(G_n(\lambda)\) is distributed as the graph obtained by doing percolation on \(G_n(\lambda')\) with probability \(p_n(\lambda)/p_n(\lambda')\).

Proof. We shall use the following property of the exponential random variables: Let \(E_1, E_2, \ldots, E_r\) be an i.i.d. sequence of exponential random variables with rate \(\lambda\). Then, conditional on the event \(\{E_1 = \min\{E_1, \ldots, E_r\}\}\), \(E_1\) is distributed as an exponential random variable with rate \(r\lambda\). Note that, by construction, \(G_n(\infty)\) is distributed as \(CM_n(d)\). Moreover, for any pair of half-edges \(e, f\)

\[
\mathbb{P}(\{(e, f)\text{ is an edge in } G_n(\lambda)\} | \{(e, f)\text{ is an edge in } G_n(\infty)\}) = \mathbb{P}(\xi(e, f) < -\log(1 - p_n(\lambda)) | \xi(e, f) = \min_{e' = e, or f' = f} \xi(e', f'))
\]

\[
= p_n(\lambda),
\]

where we have used the fact stated at the beginning of the proof and the fact that \(|\{(e', f') : e' = e, or f' = f, e' \neq f'\}| = 2\ell_n - 3\). Further, the events \(\{(e, f) \in G_n(\lambda)\}\) and \(\{(e', f') \in G_n(\lambda)\}\) are independent conditioned on \(G_n(\infty)\), because they depend on disjoint sets of \(\xi\)-values. To see the second assertion, note that for a random variable \(X\) and an event \(A\), satisfying \(X|A \sim \text{Exp}(1), \mathbb{P}(X \leq x | A, X \leq y) = (1 - e^{-x})/(1 - e^{-y})\) for any \(x \leq y\). Therefore,

\[
\mathbb{P}(\{(e, f)\text{ is an edge in } G_n(\lambda)\} | \{(e, f)\text{ is an edge in } G_n(\infty)\}) = \mathbb{P}(\xi(e, f) < -\log(1 - p_n(\lambda)) | \xi(e, f) = \min_{e' = e, or f' = f} \xi(e', f'), \xi(e, f) < -\log(1 - p_n(\lambda')))
\]

\[
= \frac{p_n(\lambda)}{p_n(\lambda')},
\]

The independence across the edges follow similarly as before and the proof is now complete. \(\Box\)

8.3 Proof of Theorem 4

We shall consider the case \(k = 2\) only, since the case for general \(k\) can be proved inductively. Fix \(-\infty < \lambda_0 < \lambda_1 < \infty\). Define,

\[
D_n(\lambda) := a_n^{-1} D_n(\lambda),
\]

where \(a_n = ((\nu_n - 1)(2\ell_n - 3)n^{1/3})^{1/2}\). Thus \(n^{-2/3}a_n \rightarrow a > 0\). \(D_n(\lambda)\) denotes the ordered vector of the number of free half-edges in each component of \(G_n(\lambda_0)\) (suitably re-scaled), where by free half-edges we mean the half-edges that were deleted in percolation. Also, the above construction gives a coupling of the graphs \((G_n(\lambda_0))_{\lambda_0 \in \mathbb{R}}\). Moreover, at time \(\lambda\), the \(i^{th}\) and the \(j^{th}\) coordinate of \(D_n(\lambda) = (D^i_n(\lambda))_{i \geq 1}\) merge at rate

\[
\frac{1}{2\ell_n - 3} \times \frac{1}{(1 - p_n(\lambda))\nu_n n^{1/3}} \approx D^i_n(\lambda) \times \frac{1}{(2\ell_n - 3)n^{1/3}} \times \frac{1}{\nu_n - 1} = \hat{D}^i_n(\lambda) \hat{D}^j_n(\lambda),
\]
and the merged component has size \( \mathcal{D}^\lambda_{(o)} + \mathcal{D}^\lambda_{(j)} - 2a_n^{-1} \approx \mathcal{D}^\lambda_{(o)} + \mathcal{D}^\lambda_{(j)} \). Thus, \((\mathcal{D}_n(\lambda))_{\lambda \geq \lambda_0} \) is not an exact multiplicative coalescent but it is close. Firstly notice that

\[
- \log(1 - p_n(\lambda)) = - \log(1 - 1/\nu_n) - \log \left( 1 - \frac{\lambda}{n^{1/3}(\nu_n - 1)} + o(n^{-1/3}) \right). \tag{8.17}
\]

Using the fact that \( x - x^2/2 \leq \log(1 - x) \leq -x \) for small \( x > 0 \), we obtain, for all sufficiently large \( n \),

\[
- \log(1 - p_n(\lambda)) = - \log(1 - 1/\nu_n) + \frac{\lambda}{n^{1/3}(\nu_n - 1)} + o(n^{-1/3}), \tag{8.18}
\]

where the error term may depend on \( \lambda \). But we restrict our attention to \( \lambda \in [\lambda_0, \lambda_1] \). Thus, we can choose \( \varepsilon_n \) (only depending upon \( \lambda_0, \lambda_1 \)) with \( n^{1/3}\varepsilon_n \to 0 \) such that

\[
- \log(1 - p_n(\lambda)) \leq - \log(1 - 1/\nu_n) + \frac{\lambda}{n^{1/3}(\nu_n - 1)} + \varepsilon_n. \tag{8.19}
\]

Define the graph \( \hat{\mathcal{G}}_n(\lambda) \) to be the graph obtained by keeping the edge \((e, f)\) if \( \xi(e, f) \leq - \log(1 - 1/\nu_n) + \lambda n^{-1/3}(\nu_n - 1)^{-1} + \varepsilon_n \). If \( \hat{D}'_n(\lambda) \) for \( \hat{G}_n(\lambda) \) is defined as the same quantity as \( D_n(\lambda) \) for \( G_n(\lambda) \). We can use (8.18) to conclude that \( \hat{G}_n(\lambda_0 - \eta_n) \subset \hat{G}_n(\lambda_0) \subset \hat{G}_n(\lambda_0 + \eta_n) \) for some \( \eta_n = o(n^{-1/3}) \) and therefore,

\[
\hat{D}'_n(\lambda_0), \text{ and } D_n(\lambda_0) \text{ have the same distributional limit}. \tag{8.20}
\]

Note that the co-ordinates of \( \hat{D}'_n(\lambda) \) merge exactly according to the product of their size, as \( \lambda \) varies. However, \( (D'_n(\lambda))_{\lambda_0 \leq \lambda \leq \lambda_1} \) is not a multiplicative coalescent yet due to the depletion of the sizes with each merge. We define an exact multiplicative coalescent below. Let \( D_n(\lambda_0) = D'_n(\lambda_0) \) and consider the following modification of the coupling (C1) defined in Section 8.2 for all the edges that appear after \( \lambda_0 \).

After pairing a half-edge \( e \) with its neighbors, we do not throw away the exponential clocks between \( e \) and other half-edges in the other components.

Let us denote by \( (D_n(\lambda))_{\lambda_0 \leq \lambda \leq \lambda_1} \) the exact multiplicative coalescent version of the above process starting with the initial distribution \( D'_n(\lambda_0) \). The processes \( (D_n(\lambda))_{\lambda_0 \leq \lambda \leq \lambda_1}, (D'_n(\lambda))_{\lambda_0 \leq \lambda \leq \lambda_1} \text{ and } (D_n(\lambda))_{\lambda_0 \leq \lambda \leq \lambda_1} \) are all coupled by using the same set of \( \xi \)-values. Then, using Theorem 25, (8.20) and the Feller property of the multiplicative coalescent [2, Proposition 5], we conclude

\[
(D_n(\lambda_0), D_n(\lambda_1)) \xrightarrow{L} \kappa \sqrt{p_a^{-1}(\gamma^\lambda_0, \gamma^\lambda_1)} \tag{8.21}
\]

with respect to the \((\ell^2_1)^2\) topology, where \( \gamma^\lambda = (|\gamma^\lambda_i|)_{i \geq 1} \). Write \( \hat{D}_n(\lambda) = (\hat{G}^\lambda_{(i)})_{i \geq 1} \). Now observe that, under the above coupling, for each \( R \geq 1 \), we have

\[
\sum_{i \leq R} (\hat{G}^\lambda_{(i)})^2 \leq \sum_{i \leq R} (\hat{G}^\lambda_{(i)})^2 \tag{8.22}
\]

for all \( \lambda \geq \lambda_0 \), where we have used (8.19). To complete the proof, we need the following two facts, stated separately as lemmas:

**Lemma 27.** Suppose \( X_n, Y_n \) are non-negative random variables such that \( X_n \leq Y_n \) a.s. and \( X_n \xrightarrow{L} X, Y_n \xrightarrow{L} X \). Then,

\[
Y_n - X_n \xrightarrow{P} 0.
\]

28
Proof. Note that \((X_n, Y_n)_{n\geq 1}\) is tight in \(\mathbb{R}^2\). Thus, for any \((n'_i)_{i\geq 1}\) there exists a subsequence \((n_i)_{i\geq 1} \subset (n'_i)_{i\geq 1}\) such that \((X_{n_i}, Y_{n_i}) \xrightarrow{D} (Z_1, Z_2)\). Using the marginal distributional limits we get \(Z_1 \overset{D}\equiv X, Z_2 \overset{D}\equiv X\). Also the joint distribution of \((Z_1, Z_2)\) is concentrated on the line \(y = x\) in the \(xy\) plane. Thus, \((X_{n_i}, Y_{n_i}) \xrightarrow{D} (X, X)\). This limiting distribution does not depend on the subsequence \((n_i)_{i\geq 1}\). Thus the tightness of \((X_n, Y_n)_{n\geq 1}\) implies \((X_n, Y_n) \xrightarrow{D} (X, X)\). The proof is now complete. \(\square\)

Lemma 28. Suppose \(X_n := (X_{ni})_{i\geq 1}\), and \(Y_n := (Y_{ni})_{i\geq 1}\) are tight random variables. Suppose \(Y_{ni} - X_{ni} \xrightarrow{P} 0\) for each \(i\). Then, \(Y_n \overset{D}\rightarrow X\) implies \(X_n \overset{D}\rightarrow X\).

Now, we have all the ingredients to complete the proof of Theorem 4 and we stitch the proof as follows: Note that (8.21), (8.22), and Lemma 27 implies that the joint vector consisting of only the first \(R\) co-ordinates of \(D_n(\lambda_0)\), and \(D_n(\lambda_1)\) has the right limit. Recall the definition of \(D_n(\lambda)\) involving \(D_n(\lambda)\) from (8.15). Applying Lemma 28 and Theorem 25, we get
\[
n^{-2\lambda/3}(D_n(\lambda_0), D_n(\lambda_1)) \xrightarrow{D} \kappa\sqrt{p}(\tilde{\gamma}^{\lambda_0}, \tilde{\gamma}^{\lambda_1}),
\]
for any \(\lambda_0 < \lambda_1\). Now we can use Lemma 24 together with the \(\ell^2\) tightness of \(C_n(\lambda)\) from Theorem 2 to conclude that
\[
(C_n(\lambda_0), C_n(\lambda_1)) \xrightarrow{D} \sqrt{p}(\tilde{\gamma}^{\lambda_0}, \tilde{\gamma}^{\lambda_1}).
\]
Thus, we have proved Theorem 4 for \(k = 2\). As remarked at the beginning of this Section, an inductive argument gives the proof for any \(k > 2\). \(\square\)

A Appendix

In this section we state and prove a general version of Lemma 5 dealing with the size-biased degrees.

Proposition 29. Suppose \(f_n : [n] \rightarrow \mathbb{R}\) is a function with \(n^{-1} \sum_{i\in[n]} f_n(i) d_i \rightarrow \sigma_f (\text{say})\), where \((d_i)_{i\in[n]}\) satisfies Assumption 1. Suppose \(\langle f_n(i)\rangle_{i\in[n]}\) is the size-biased ordering of \((f_n(i))_{i\in[n]}\) with size \((d_i/\ell_n)_{i\in[n]}\).

Consider \(H_n(u) := n^{-\alpha} \sum_{i=1}^{[un^\alpha]} f_n(i)\) for some \(\alpha \in (0, \frac{2}{3})\) and \(u > 0\). Then,
\[
\sup_{u \leq t_n} |H_n(u) - \frac{\sigma_f u}{\mu}| = O_P(n^{-1/3}),
\]
where \(a_n = (t_n n^{-\alpha} \max_{i\in[n]} f_n(i))^{1/2}\) and \(b_n = d_{\text{max}} t_n^2 n^{\alpha - 1}\), as long as \(t_n = o(n^{\beta/2})\) for some \(\beta\) satisfying \(\max \{\alpha/2, 2\alpha - 1\} \leq \beta < \alpha\).

Proof. Let \(\ell_n = \sum_{i\in[n]} d_i\). Recall the definition of \(H_n(u)\). Consider \(n\) independent exponential random variables \(T_i \sim \exp(d_i/\ell_n)\). Define
\[
\tilde{H}_n(u) = n^{-\alpha} \sum_{i\in[n]} f_n(i) \mathbb{1}_{\{T_i \leq u n^\alpha u\}}.
\]
Therefore, \(\tilde{H}_n(u) = \sum_{i=1}^{N(n^\alpha u)} f_n(i) = H_n(n^\alpha u)\) where \(N(u) := \#\{j : T_j \leq un^\alpha\}\). Consider \(Y_0(s) = n^{-\beta}(N(sn^\alpha) - sn^\alpha)\) for some \(\alpha/2, 2\alpha - 1\) \(\leq \beta < \alpha\). Define \(\nu'_s := \{j : T_j \leq sn^\alpha\}\). We have
\[
\mathbb{E}[Y_0(u)]_{\mathcal{F}_s} = Y_0(s) + \mathbb{E}[Y_0(u) - Y_0(s)]_{\mathcal{F}_s}
\]
\[
= Y_0(s) + \frac{1}{n^\beta} \left[ \mathbb{E}\left[ \#\{j : T_j \in ([sn^\alpha, [un^\alpha])]\} | \mathcal{F}_s \right] - (u-s)n^\alpha \right]
\]
\[
= Y_0(s) + \frac{1}{n^\beta} \left[ \sum_{j \notin \nu'_s} (1 - \exp(-d_j(t-s)n^\alpha \ell_n)) - (u-s)n^\alpha \right] \leq Y_0(s),
\]
(3.4)
where the last step follows because $1 - e^{-x} \leq x$. Therefore, \(\{Y_0(s)\}_{s \geq 0}\) is a supermartingale. Also noting that \(\mathbb{E}[Y_0(0)] = 0\) and \(e^{-x} \leq 1 - x + x^2/2\),

\[
|\mathbb{E}[Y_0(u)]| = -\mathbb{E}[Y_0(u)] = \frac{1}{n^\beta} \left[ un^\alpha - \sum_{i=1}^n (1 - \exp(-un^\alpha d_i \ell_n^{-1})) \right] \\
= \frac{1}{n^\beta} \left[ \sum_{i=1}^n \left( un^\alpha d_i \ell_n^{-1} - (1 - \exp(-un^\alpha d_i \ell_n^{-1})) \right) \right] \leq n^{-\beta} \frac{u^2 n^{2\alpha} \sum_{i=1}^n d_i^2}{2 \ell_n^2} \tag{A.4}
\]

and

\[
\text{Var}(Y_0(u)) = n^{-2\beta} \text{var}(N(un^\alpha)) = n^{-2\beta} \sum_{i=1}^n \mathbb{P}(T_j \leq un^\alpha)(1 - \mathbb{P}(T_j \leq un^\alpha)) \\
\leq n^{-2\beta} \sum_{i=1}^n \frac{d_i^2}{\ell_n^2} = un^\alpha - 2\beta. \tag{A.5}
\]

Using the maximal inequality in [2, Lemma 12], for any \(\varepsilon > 0\) and \(T > 0\),

\[
\varepsilon \mathbb{P}\left( \sup_{s \leq T} |Y_0(s)| > 3\varepsilon \right) \leq 3 \left( |\mathbb{E}(Y_0(T))| + \sqrt{\text{Var}(Y_0(T))} \right). \tag{A.6}
\]

Therefore,

\[
\sup_{u \leq t_n} \left| n^{-\alpha} N(un^\alpha) - u \right| = O_P(t_n^{-2\beta}). \tag{A.7}
\]

This, in particular, implies \(N(2t_n n^\alpha) \geq t_n n^\alpha\) whp when \(t_n = o(n^{\beta/2})\). Therefore,

\[
\sup_{u \leq t_n} \left| H_n(u) - \frac{\sigma f u}{\mu} \right| \leq \sup_{u \leq 2t_n} \left| n^{-\alpha} \sum_{i=1}^n f_n(i) - \frac{\sigma f}{\mu} n^{-\alpha} N(n^\alpha u) \right| \\
\leq \sup_{u \leq 2t_n} \left| n^{-\alpha} \tilde{H}_n(u) - \frac{\sigma f u}{\mu} \right| + \frac{\sigma f}{\mu} \sup_{u \leq 2t_n} \left| n^{-\alpha} N(n^\alpha u) - u \right|. \tag{A.8}
\]

Define \(Y_1(u) = n^{-\alpha} \sum_{i=1}^n f_n(i) - \sigma_f^{(n)} u\), where \(\sigma_f^{(n)} = \sum_{i=1}^n f_n(i) d_i / \ell_n = \sigma_f / \mu + o(1)\) and hence

\[
\mathbb{E}[Y_1(t)| \mathcal{F}_s] = Y_1(s) + \mathbb{E}[Y_1(t) - Y_1(s)| \mathcal{F}_s] \\
= Y_1(s) + \frac{1}{n^\alpha} \sum_{j \notin Y_s'} f_n(j) \left( 1 - \exp(-d_j (t-s)n^{2/3} \ell_n^{-1}) \right) - (t-s) \sigma_f^{(n)} \leq Y_1(s). \tag{A.9}
\]

Thus, \(Y_1(u)_{u \geq 0}\) is also a super-martingale and by noting that \(\mathbb{E}[Y_1(0)] = 0\) we have,

\[
|\mathbb{E}[Y_1(t)]| = -\mathbb{E}[Y_1(t)] = \sigma_f^{(n)} t - n^{-\alpha} \sum_{i=1}^n f_n(i) (1 - \exp(-tn^\alpha d_i \ell_n^{-1})) \\
= n^{-\alpha} \sum_{i=1}^n f_n(i) (\exp(-tn^\alpha d_i \ell_n^{-1}) - 1 + tn^\alpha d_i \ell_n^{-1}) \leq n^{-\alpha} \frac{t^2}{2} n^{2\alpha} \frac{\sum_{i=1}^n f_n(i) d_i^2}{\ell_n^2} \tag{A.10}
\]

\[
= \frac{t^2}{2} n^{-2/3} d_{\text{max}} \frac{1}{n^{1/3}} \sum_{i=1}^n f_n(i) d_i \left( \frac{ \ell_n}{n} \right)^2,
\]

30
where the last step follows by conditions on \( f_n, d_i \) and the fact that \( \alpha \leq 2/3 \). Also,

\[
\text{Var}(Y_1(t)) = n^{-2\alpha} \sum_{i \in [n]} f_n^2(i) \exp \left( -\frac{tn^\alpha d_i}{\ell_n} \right) \left( 1 - \exp \left( -\frac{tn^\alpha d_i}{\ell_n} \right) \right)
\]

\[
\leq n^{-2\alpha} \max_{i \in [n]} \frac{t n^\alpha \sum_{i \in [n]} f_n(i) d_i}{\ell_n} 
\]

(A.11)

Recalling the assumptions on \( f_n \) from the statement of Proposition 29, another application of (A.6) yields

\[
\sup_{u \leq t} \left| \frac{n^{-\alpha} \hat{H}_n(u) - \frac{\alpha f_n u}{\mu}}{\mu} \right| = O_p(a_n \vee b_n) 
\]

(A.12)

where \( a_n, b_n \) are as stated in Proposition 29. Thus, (A.8) together with (A.12) and (A.7) completes the proof.

\[ \square \]

**Acknowledgement**

This research has been supported by the Netherlands Organisation for Scientific Research (NWO) through Gravitation Networks grant 024.002.003. In addition, RvdH has been supported by VICI grant 639.033.806 and JvL has been supported by the European Research Council (ERC).

**References**

[1] Addario-Berry, L., Broutin, N., Goldschmidt, C., and Miermont, G. (2013). The scaling limit of the minimum spanning tree of the complete graph. *arXiv:1301.1664*.

[2] Aldous, D. (1997). Brownian excursions, critical random graphs and the multiplicative coalescent. *Annals of Probability*, 25(2):812–854.

[3] Aldous, D. and Limic, V. (1998). The entrance boundary of the multiplicative coalescent. *Electronic Journal of Probability*, 3(3):1–59.

[4] Bhamidi, S., Broutin, N., Sen, S., and Wang, X. (2014). Scaling limits of random graph models at criticality: Universality and the basin of attraction of the Erdos-Rényi random graph. *arXiv:1411.3417*.

[5] Bhamidi, S., van der Hofstad, R., and van Leeuwaarden, J. S. H. (2010). Scaling Limits for Critical Inhomogeneous Random Graphs with Finite Third Moments. *Electronic Journal of Probability*, 15(6):1682–1702.

[6] Bhamidi, S., van der Hofstad, R., and van Leeuwaarden, J. S. H. (2012b). Novel scaling limits for critical inhomogeneous random graphs. *Annals of Probability*, 40(6):2299–2361.

[7] Bollobás, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316.

[9] Dhara, S., van der Hofstad, R., van Leeuwaarden, J. S. H., and Sen, S. Critical window for the configuration model: infinite third moment degrees.
[10] Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the Internet topology. *Computer Communication Review*, 29(4):251–262.

[11] Fountoulakis, N. (2007). Percolation on sparse random graphs with given degree sequence. *Internet Mathematics*, 4(1):329–356.

[12] Janson, S. (2009a). On percolation in random graphs with given vertex degrees. *Electronic Journal of Probability*, 14:87–118.

[13] Janson, S. (2009b). Susceptibility of random graphs with given vertex degrees. *arXiv:0911.2636*.

[14] Janson, S. and Luczak, M. J. (2009). A new approach to the giant component problem. *Random Structures and Algorithms*, 34(2):197–216.

[15] Janson, S., Łuczak, T., and Rucinski, A. (2000). *Random Graphs*. Wiley, New York.

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

[17] Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.

[18] Lipster, R. S. and Shiryayev, A. N. (1989). *Theory of Martingales*. Springer, Dordrecht.

[19] Molloy, M. and Reed, B. (1995). A critical-point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6(2-3):161–179.

[20] Nachmias, A. and Peres, Y. (2010). Critical percolation on random regular graphs. *Random Structures and Algorithms*, 36(2):111–148.

[21] Riordan, O. (2012). The phase transition in the configuration model. *Combinatorics, Probability and Computing*, 21:265—299.

[22] Siganos, G., Faloutsos, M., Faloutsos, P., and Faloutsos, C. (2003). Power laws and the AS-level internet topology. *IEEE/ACM Transactions on Networking*, 11(4):514–524.

[23] van der Hofstad, R. (2014). *Random Graphs and Complex Networks*, volume I.

[24] Whitt, W. (2002). *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer-Verlag, New York.

[25] Whitt, W. (2007). Proofs of the martingale FCLT. *Probability Surveys*, 4:269–302.