ANATOMY OF A YOUNG GIANT
COMPONENT IN THE RANDOM GRAPH

JIAN DING, JEONG HAN KIM, EYAL LUBETZKY AND YUVAL PERES

ABSTRACT. We provide a complete description of the giant component of the Erdős-Rényi random graph $\mathcal{G}(n, p)$ as soon as it emerges from the scaling window, i.e., for $p = (1 + \varepsilon)/n$ where $\varepsilon^3 n \to \infty$ and $\varepsilon = o(1)$.

Our description is particularly simple for $\varepsilon = o(n^{-1/4})$, where the giant component $\mathcal{C}_1$ is contiguous with the following model (i.e., every graph property that holds with high probability for this model also holds w.h.p. for $\mathcal{C}_1$). Let Z be normal with mean $\frac{2}{3}\varepsilon^3 n$ and variance $\varepsilon^3 n$, and let $\mathcal{K}$ be a random 3-regular graph on $2\lfloor Z \rfloor$ vertices. Replace each edge of $\mathcal{K}$ by a path, where the path lengths are i.i.d. geometric with mean $1/\varepsilon$. Finally, attach an independent Poisson$(1 - \varepsilon)$-Galton-Watson tree to each vertex.

A similar picture is obtained for larger $\varepsilon = o(1)$, in which case the random 3-regular graph is replaced by a random graph with N_k vertices of degree k for $k \geq 3$, where N_k has mean and variance of order $\varepsilon^k n$.

This description enables us to determine fundamental characteristics of the supercritical random graph. Namely, we can infer the asymptotics of the diameter of the giant component for any rate of decay of ε, as well as the mixing time of the random walk on $\mathcal{C}_1$.

1. INTRODUCTION

The Erdős and Rényi random graph $\mathcal{G}(n, p)$ has been studied extensively since its introduction in 1959 [15]. Much of the analysis of this fundamental random graph model has focused on its behavior near the critical point $p = 1/n$. Nevertheless, a few key features, such as the diameter and the mixing time of the random walk on the largest component, have remained unknown in a regime just beyond criticality.

In their seminal papers from the 1960's, Erdős and Rényi established a phenomenon known as *the double jump*. For $p = c/n$ where $c < 1$ is fixed, the largest component $\mathcal{C}_1$ has size $O(\log n)$ with high probability (w.h.p.). When $c > 1$, the size of $\mathcal{C}_1$ is linear in n, and at the critical $c = 1$ it has order $n^{2/3}$ (this latter fact was fully established much later by Bollobás [10] and Łuczak [24]). As discovered in [10], the critical behavior extends throughout the *critical window*, the regime where $p = (1 \pm \varepsilon)/n$ for $\varepsilon = O(n^{-1/3})$.

Up to the critical point, the structure of $\mathcal{C}_1$ is relatively well understood. For instance, in the fully subcritical regime ($p = (1 - \varepsilon)/n$ for $\varepsilon > 0$ fixed), $\mathcal{C}_1$ is a tree of known (logarithmic) size and diameter. In the critical window

($\varepsilon = O(n^{-1/3})$ the distribution of $|\mathcal{C}_1|$ was determined in [1, 26], and the diameter was found in [28]. See [9, 19] for further information.

In the supercritical regime ($p = (1 + \varepsilon)/n$ with $\varepsilon^3 n \to \infty$), a variety of methods can determine key features of $\mathcal{C}_1$ up to some continuous functions of ε. While these functions remain bounded in the fully supercritical case ($\varepsilon > 0$ fixed), the situation becomes much more delicate as ε approaches the critical window.

For example, one can deduce that the diameter of the fully supercritical $\mathcal{C}_1$ has order $\log n$ merely by analyzing certain (weak) expansion properties of its 2-core (formally defined in Section 2). More precise results on the diameter were obtained in [27, 32], but they still do not give the asymptotic diameter in the whole supercritical regime.

In the fully supercritical case, it is known that the giant component consists of an expander, "decorated" using paths and trees of at most logarithmic size (see [6] for a concrete example of such a statement, used there to obtain the order of the mixing time on the fully supercritical $\mathcal{C}_1$). However, the existing decompositions of the giant component are not precise enough to handle the case where $\varepsilon \to 0$ (e.g., in [32] Riordan and Wormald point out that this is the most difficult regime for determining the diameter).

In this work, we obtain a complete characterization of the supercritical giant component. Rather than merely describing its properties, we present a simple construction whose distribution is contiguous with that of $\mathcal{C}_1$. This construction is particularly elegant when the giant component is "young", namely when $\varepsilon = o(n^{-1/4})$. Since this is the hardest regime for alternative approaches, we start by describing this special case.

Let $\mathcal{N}(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2, and let $\text{Geom}(\varepsilon)$ denote the geometric distribution with mean $1/\varepsilon$.

Theorem 1. *Let $\mathcal{C}_1$ be the largest component of the random graph $\mathcal{G}(n, p)$ for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon^3 n \to \infty$ and $\varepsilon = o(n^{-1/4})$. Then $\mathcal{C}_1$ is contiguous to the model $\tilde{\mathcal{C}}_1$, constructed in 3 steps as follows:*

1. Let $Z \sim \mathcal{N}\left(\frac{2}{3}\varepsilon^3 n, \varepsilon^3 n\right)$, and select a random 3-regular graph $\mathcal{K}$ on $N = 2\lfloor Z \rfloor$ vertices.

2. Replace each edge of $\mathcal{K}$ by a path, where the path lengths are i.i.d. $\text{Geom}(\varepsilon)$.

3. Attach an independent $\text{Poisson}(1 - \varepsilon)$-Galton-Watson tree to each vertex.

That is, $\mathbb{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A}) \to 0$ implies $\mathbb{P}(\mathcal{C}_1 \in \mathcal{A}) \to 0$ for any set of graphs $\mathcal{A}$.

In the above, a $\text{Poisson}(\mu)$-Galton-Watson tree is the family tree of a Galton-Watson branching process with offspring distribution $\text{Poisson}(\mu)$.

Two well-known objects relevant to the study of the giant component are its 2-core $\mathcal{C}_1^{(2)}$ and its kernel $\mathcal{K}$. The 2-core of a graph is its maximum subgraph where all degrees are at least 2. The kernel is obtained from the

2-core by replacing every maximal 2-path by an edge (where a 2-path is a path where all internal vertices have degree 2). Note that our description of $\tilde{\mathcal{C}}_1$ constructs the kernel in Step 1, the 2-core in Step 2 and the entire component $\tilde{\mathcal{C}}_1$ in Step 3.

The above theorem not only states that the kernel of $\mathcal{C}_1$ in this regime is an expander, but it is in fact contiguous to a random 3-regular graph, an object whose expansion properties are well understood (cf., e.g., [18]). Furthermore, the 2-core is obtained from the kernel by a simple operation ("stretching" the edges into paths of lengths i.i.d. geometric with mean $1/\varepsilon$). This allows us to pinpoint the expansion properties of the 2-core and their dependence on ε as it tends to 0.

A few known (yet nontrivial) properties of the 2-core of $\mathcal{C}_1$ can be immediately read off from Theorem 1. For instance, w.h.p. the 2-core contains $(2+o(1))\varepsilon^2 n$ vertices while the kernel has $(\frac{4}{3}+o(1))\varepsilon^3 n$ vertices (see [25, 31]). As there are w.h.p. $(2 + o(1))\varepsilon^3 n$ edges in the kernel, a simple estimate of the maximum of i.i.d. geometric variables gives the following corollary.

Corollary 1. *Let $\mathcal{C}_1^{(2)}$ be the 2-core of the largest component of $\mathcal{G}(n,p)$ for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon^3 n \to \infty$ and $\varepsilon = o(n^{-1/4})$. The maximal 2-path in $\mathcal{C}_1^{(2)}$ has length $(1/\varepsilon)\log(\varepsilon^3 n) + O_P(1/\varepsilon)$.*

Moreover, Theorem 1 enables us to interpret distances in the 2-core as passage times in first-passage percolation (for further information on this thoroughly studied topic, see, e.g., [21]). As we state in Theorem 3 below, this connection (used in a companion paper [12]) gives the asymptotic behavior of the diameter throughout the regime $\varepsilon^3 n \to \infty$ and $\varepsilon = o(1)$.

1.1. **Main results.** We now state the extension of Theorem 1 to all $\varepsilon = o(1)$ outside the critical window.

Theorem 2. *Let $\mathcal{C}_1$ be the largest component of $\mathcal{G}(n,p)$ for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon^3 n \to \infty$ and $\varepsilon \to 0$. Let $\mu < 1$ denote the conjugate of $1 + \varepsilon$, that is, $\mu e^{-\mu} = (1 + \varepsilon)e^{-(1+\varepsilon)}$. Then $\mathcal{C}_1$ is contiguous to the following model $\tilde{\mathcal{C}}_1$:*

1. *Let $\Lambda \sim \mathcal{N}\left(1 + \varepsilon - \mu, \frac{1}{\varepsilon n}\right)$ and assign i.i.d. variables $D_u \sim \mathrm{Poisson}(\Lambda)$ ($u \in [n]$) to the vertices, conditioned that $\sum D_u \mathbf{1}_{D_u \geq 3}$ is even.*
   Let $N_k = \#\{u : D_u = k\}$ and $N = \sum_{k \geq 3} N_k$. Select a random graph $\mathcal{K}$ on N vertices, uniformly among all graphs with N_k vertices of degree k for $k \geq 3$.
2. *Replace the edges of $\mathcal{K}$ by paths of lengths i.i.d. $\mathrm{Geom}(1 - \mu)$.*
3. *Attach an independent $\mathrm{Poisson}(\mu)$-Galton-Watson tree to each vertex.*

That is, $\mathbb{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A}) \to 0$ implies $\mathbb{P}(\mathcal{C}_1 \in \mathcal{A}) \to 0$ for any set of graphs $\mathcal{A}$.

We note that conditioning that the sum of degrees is even can easily be realized by rejection sampling. The differences between the two theorems are the approximation of $1 - \mu \approx \varepsilon$ in Steps 2,3, and a richer degree distribution of the random graph $\mathcal{K}$ in Step 1.

Combining Theorem 2 with some known results on first-passage percolation from [7] gives an immediate corollary on the typical distances between vertices of degree at least 3 in the 2-core.

Corollary 2. *Let $\mathcal{C}_1^{(2)}$ be the 2-core of the largest component of $\mathcal{G}(n,p)$ for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon^3 n \to \infty$ and $\varepsilon = o(1)$. Let u, v be two vertices of degree at least 3 in $\mathcal{C}_1^{(2)}$, chosen u.a.r. among all such vertices. The distance between u, v is w.h.p. $(1/\varepsilon + O(1)) \log(\varepsilon^3 n)$.*

However, maximal distances in the 2-core can differ from typical distances; compare the above result to (1.3) in the next theorem, which we prove in a companion paper.

Theorem 3 ([12])**.** *Consider the random graph $\mathcal{G}(n,p)$ for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon^3 n \to \infty$ and $\varepsilon = o(1)$. Let $\mathcal{C}_1$ be the largest component G, let $\mathcal{C}_1^{(2)}$ be its 2-core and let $\mathcal{K}$ denote its kernel. Then w.h.p.,*

$$\operatorname{diam}(\mathcal{C}_1) = \big(3 + o(1)\big)(1/\varepsilon) \log(\varepsilon^3 n), \qquad (1.1)$$

$$\operatorname{diam}(\mathcal{C}_1^{(2)}) = \big(2 + o(1)\big)(1/\varepsilon) \log(\varepsilon^3 n), \qquad (1.2)$$

$$\max_{u,v \in \mathcal{K}} \operatorname{dist}_{\mathcal{C}_1^{(2)}}(u,v) = \big(\tfrac{5}{3} + o(1)\big)(1/\varepsilon) \log(\varepsilon^3 n). \qquad (1.3)$$

To prove the above theorem, we need to go beyond typical distances and obtain new large deviation estimates for the relevant parameters (see [12] for further details). The result (1.1) on the diameter of the giant component concludes a long list of studies of this parameter in the supercritical random graph (e.g., [11, 16, 27, 32]). First results for the challenging regime where $\varepsilon = o(1)$ appeared only recently: Riordan and Wormald [32] obtained very accurate estimates of the diameter for most of this regime, but did not cover the range where the random graph emerges from the critical window (i.e., $\varepsilon^3 n$ tends to ∞ arbitrarily slowly). Łuczak and Seierstad [27] then gave estimates for the diameter that do apply to the entire supercritical regime, yet their upper and lower bounds differ by a factor of $\frac{1000}{7}$.

Controlling typical and maximal distances between vertices in the giant component is but one of several prerequisites for estimating the mixing time of the (lazy) random walk on $\mathcal{C}_1$. For instance, as this parameter is highly sensitive to bottlenecks in $\mathcal{C}_1$, one also needs to fully understand the isoperimetric profile of the 2-core and the structure of the trees attached to it.

In the fully supercritical case, Fountoulakis and Reed [17] and Benjamini, Kozma and Wormald [6] independently proved that the mixing time on

$\mathcal{C}_1$ is of order $\log^2 n$. However, as evident from the structure description in Theorem 2, methods for the fully-supercritical case that depend on large sets in the 2-core having edge expansion bounded away from 0 will break down as $\varepsilon \to 0$. Within the critical window, it was shown in [28] that the mixing time on $\mathcal{C}_1$ has order n. For $\varepsilon = o(1)$ outside the critical window, the problem of estimating the mixing-time on $\mathcal{C}_1$ remained open, and furthermore, it was unclear what the answer should be, as one would expect some interpolation between $\log^2 n$ for fixed $\varepsilon > 0$ and order n at criticality.

The following theorem, proved in a companion paper, settles this problem by exploiting the geometric understanding of $\mathcal{C}_1$ provided by Theorem 2. This completes the picture of the supercritical mixing time.

Theorem 4 ([13])**.** *Let $\mathcal{C}_1$ be the largest component of $\mathcal{G}(n,p)$ for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon^3 n \to \infty$ and $\varepsilon = o(1)$. With high probability, the mixing time of the lazy random walk on $\mathcal{C}_1$ is of order $(1/\varepsilon^3) \log^2(\varepsilon^3 n)$.*

Indeed, the mixing time exhibits a smooth evolution from the critical regime $\varepsilon = O(n^{-1/3})$ to the fully-supercritical regime of $\varepsilon > 0$ fixed.

1.2. **Main techniques.** A key ingredient in the proofs is the Poisson cloning model $\mathcal{G}_{\mathrm{PC}}(n,p)$, introduced in [23] and shown to be contiguous to $\mathcal{G}(n,p)$ (see Section 2). It thus suffices to establish the contiguity of our model $\tilde{\mathcal{C}}_1$ to the giant component of Poisson cloning, a fact we establish in several stages.

We first show the contiguity of the 2-cores in the models through a careful analysis of $\mathcal{G}_{\mathrm{PC}}(n,p)$. We then perform a series of contiguous translations of the model, in order to remove dependencies between maximal 2-paths in the 2-core, as well as incorporate the trees attached to the 2-core in $\mathcal{C}_1$. To establish these, we use local central limit results for various parameters, including a powerful local CLT of Pittel and Wormald [31].

1.3. **Organization.** Section 2 contains several preliminary facts needed for the proofs. In Section 3 we reduce the 2-core of the Poisson cloning model to an intermediate simplified model (the proof of a technical lemma on Poisson cloning used here is postponed to Section 7). This model is subsequently reduced in Section 4 to one that is essentially the 2-core of our model $\tilde{\mathcal{C}}_1$. The complete structure of the giant component is thereafter analyzed in Section 5, which concludes the proof of Theorem 2. In Section 6 we prove Theorem 1, addressing the special case of the early giant component.

2. Preliminaries

2.1. **Cores and kernels.** The *k-core* of a graph G, denoted by $G^{(k)}$, is its maximum subgraph $H \subset G$ where every vertex has degree at least k. It is well known (and easy to verify) that this subgraph is unique, and can be

obtained by repeatedly deleting any vertex whose degree is smaller than k (at an arbitrary order).

We call a path $\mathcal{P} = v_0, v_1, \ldots, v_k$ for $k > 1$ (i.e., a sequence of vertices with $v_i v_{i+1}$ an edge for each i) a *2-path* if and only if v_i has degree 2 for all $i = 1, \ldots, k-1$ (while the endpoints v_0, v_k may have degree larger than 2, and possibly $v_0 = v_k$).

The *kernel* $\mathcal{K}$ of G is obtained by taking its 2-core $G^{(2)}$ minus its disjoint cycles, then repeatedly contracting all 2-paths (replacing each by a single edge). Notice that, by definition, the degree of every vertex in $\mathcal{K}$ is at least 3. At certain times the notation $\ker(G)$ will be useful to denote a kernel with respect to some specific graph G.

2.2. **Configuration model.** This model, introduced by Bollobás [8], provides a remarkable method for constructing random graphs with a given degree distribution, which is highly useful to their analysis. We describe this for the case of random d-regular graphs for d fixed (the model is similar for other degree distributions); see [9, 19, 34] for additional information.

Associate each of the n vertices with d distinct points (also referred to as "half-edges"), and consider a uniform perfect matching on these points. The random d-regular graph is obtained by contracting each cluster of the d points corresponding to a vertex, possibly introducing multiple edges and self-loops. Clearly, on the event that the obtained graph is simple, it is uniformly distributed among all d-regular graphs, and furthermore, one can show that this event occurs with probability bounded away from 0 (namely, with probability about $\exp(\frac{1-d^2}{4})$). Hence, every event that occurs w.h.p. for this model, also occurs w.h.p. for a random d-regular graph.

One particularly useful property of the above model is that it allows one to construct the graph gradually, exposing the edges of the matching one by one. This way, having exposed part of the graph, the edges on the remaining unmatched points are still distributed as a uniform perfect matching.

2.3. **Poisson cloning model.** In order to analyze the delicate structure of the near-critical giant component, we need to use Poisson cloning model $\mathcal{G}_{\mathrm{PC}}(n, p)$, which was introduced in [23]. We incorporate a brief account on Poisson cloning model as follows, and one can see [22] and [23] for more.

Let V be the set of n vertices, and $\mathrm{Po}(\lambda)$ denote a Poisson random variable with mean λ. Let $\{d_v\}_{v \in V}$ be a sequence of i.i.d. $\mathrm{Po}(\lambda)$ variables with $\lambda = (n-1)p$. Then, take $d(v)$ copies of each vertex $v \in V$ and the copies of v are called *clones* of v or simply *v-clones*. Define $N_\lambda \triangleq \sum_{v \in V} d(v)$.

If N_λ is even, the multi-graph $\mathcal{G}_{\mathrm{PC}}(n, p)$ is obtained by generating a uniform random perfect matching of those N_λ clones (e.g., via the configuration model, where every clone is considered to be a half-edge) and contracting

clones of the same vertex. That is to say, each matching of a v-clone and a w-clone is translated into the edge (v, w) with multiplicity. In the case that $v = w$, it contributes a self-loop with degree 2. On the other hand, if N_λ is odd, we first pick a uniform clone and translate it to a special self-loop contributing degree 1 of the corresponding vertex. For the remaining clones, we generate a perfect matching and contract them as in the N_λ even case.

The following theorem of [23] states that the Poisson cloning model is *contiguous* with Erdős-Rényi model. Hence, it suffices to study Poisson cloning model in order to establish properties of Erdős-Rényi model.

Theorem 2.1 ([23, Theorem 1.1]). *Suppose $p = \Theta(n^{-1})$. Then there exist constants $c_1, c_2 > 0$ such that for any collection $\mathcal{F}$ of simple graphs, we have*

$$c_1 \mathbb{P}(\mathcal{G}_{\mathrm{PC}}(n, p) \in \mathcal{F}) \leq \mathbb{P}(\mathcal{G}(n, p) \in \mathcal{F}) \leq c_2 \big(\mathbb{P}(\mathcal{G}_{\mathrm{PC}}(n, p) \in \mathcal{F}) \big)^{1/2} + \mathrm{e}^{-n} \big) .$$

Note that in our regime ($p = \frac{1+\varepsilon}{n}$ for $\varepsilon = o(1)$ and $\varepsilon^3 n \to \infty$) we may replace the rate $\lambda = (n-1)p$ in the Poisson-cloning model definition simply by $\lambda = np$, for convenience.

3. The 2-core of Poisson cloning

By the results of [23], the random graph $\mathcal{G}(n, p)$ in our range of parameters is contiguous to the Poisson cloning model, where every vertex gets an i.i.d. $\mathrm{Po}(np)$ number of half-edges (clones), and the final (multi)graph is obtained thereafter via the configuration model. As opposed to $\mathcal{G}(n, p)$, the Poisson cloning model features vertex degrees that are independently distributed, often contributing to an easier analysis. Nevertheless, the structure of the 2-core in this model just beyond criticality is still highly nontrivial.

The main goal in this section is to reduce the 2-core of the supercritical Poisson cloning model to the following tractable model, which is simply a random graph uniformly chosen over all graphs with a given degree sequence.

Definition 3.1 (Poisson-configuration model for n and $p = \frac{1+\varepsilon}{n}$).
(1) *Let $\Lambda \sim \mathcal{N} \left(1 + \varepsilon - \mu, \frac{1}{\varepsilon n} \right)$ and assign an independent variable $D_u \sim \mathrm{Po}(\Lambda)$ to each vertex u. Let $N_k = \#\{u : D_u = k\}$ and $N = \sum_{k \geq 2} N_k$.*
(2) *Construct a random graph on N vertices, uniformly chosen over all graphs with N_k degree-k vertices for $k \geq 2$ (if N is odd, choose a vertex u with $D_u = k \geq 2$ with probability proportional to k, and give it $k - 1$ half-edges and a self-loop).*

Theorem 3.2. *Let $G \sim \mathcal{G}_{\mathrm{PC}}(n, p)$ be generated by the Poisson cloning model for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon \to 0$ and $\varepsilon^3 n \to \infty$. Let $G^{(2)}$ be its 2-core, and H be generated by the Poisson-configuration model corresponding to n, p. Then for any set of graphs $\mathcal{A}$ such that $\mathbb{P}(H \in \mathcal{A}) \to 0$, we have $\mathbb{P}(G^{(2)} \in \mathcal{A}) \to 0$.*

In order to prove the above Theorem 3.2, in what follows we review a specific way to generate $\mathcal{G}_{\mathrm{PC}}(n,p)$, introduced in [23]. Let V be a set of n vertices and let

$$\lambda \overset{\triangle}{=} np = 1 + \varepsilon \;,$$

be the mean of the degree. Consider n horizontal line segments ranging from $(0,j)$ to (λ, j), for $j = 1, \ldots, n$ in $\mathbb{R}^2$. Assign a Poisson point process with rate 1 on each line segment independently. Each point (x,v) in these processes is referred to as a v-clone with the assigned number x. The entire set of Poisson point processes is called a Poisson λ-cell.

Given the Poisson λ-cell, there are various schemes to generate a perfect matching on all points (thus yielding a random graph). One such way is the "Cut-Off Line Algorithm" (COLA), defined in [22], which is useful in finding the 2-core $G^{(2)}$. We next describe this algorithm in detail.

First define θ_λ to be the unique positive solution to the following equation:

$$\theta = 1 - \mathrm{e}^{-\theta\lambda} \;. \tag{3.1}$$

It is straightforward to verify that

$$\theta_\lambda = (2 + o(1))\varepsilon \;. \tag{3.2}$$

Let β be some real, to be specified later, satisfying

$$\frac{1 - \theta_\lambda}{3} \leq \beta \leq \frac{1 - \theta_\lambda}{2} \;. \tag{3.3}$$

Next, construct a Poisson λ-cell as follows. The COLA procedure consists of multiple phases, formally defined in Algorithm 1 below. Throughout these phases, the algorithm maintains the position of a "cut-off line", a vertical line in $\mathbb{R}^2$ whose initial x-coordinate equals λ, and gradually moves leftwards. The j-th phase $(j \geq 1)$ begins when the line is at $(1-\beta)^{j-1}\lambda$ and ends once it reaches $(1-\beta)^j\lambda$.

The result of each phase is a matching on (previously unmatched) clones. In order to describe the rule of constructing this matching, we need the following definitions. At any given point, we call a vertex $v \in V$ (and its unmatched clones) *light* if it has at most one unmatched clone and *heavy* otherwise. Furthermore, for each j, we label each vertex (and its unmatched clones) at the beginning of phase j as either j-*active* or j-*passive*, as follows. A vertex $v \in V$ (and its clones) is j-passive if it has precisely 2 clones to the left of the cut-off line, and both are unmatched. This partition of the unmatched clones into j-active and j-passive ones remains fixed throughout phase j.

At the beginning of the process, all the light clones are placed in a stack (whose state is maintained without being re-initialized after each phase).

The order by which these clones are inserted into the stack can be arbitrary, as long as it is oblivious of the values assigned to the clones.

Algorithm 1 CUT-OFF LINE ALGORITHM: PHASE j DESCRIPTION

---

1. As long as the stack is nonempty, repeat the following:
   - Let (u, i) be the first clone in the stack.
   - Move the cut-off line leftwards until one of the following occurs:
     - (a) If the line hits $(1 - \beta)^j$, the phase is conlcuded (quit).
     - (b) The line hits an unmatched clone $(v, j) \neq (u, i)$.
   - Remove (u, i) from the stack, as well as (v, j) (if it is there).
   - Match (u, i) and (v, j), and re-evaluate u and v as light/heavy.
   - Add any clone that just became *light* into the stack.
2. If there are *active* unmatched clones:
   - (a) Choose such a clone uniformly at random and put it in the stack.
   - (b) Return to Step (1).

   Otherwise, the algorithm is concluded (no additional phases).

Define Λ_C to be the x-coordinate of the cut-off line once Step 2 is reached for the first time in the course of the algorithm, i.e., at the first time when there are no light clones. The next lemma states that Λ_C is concentrated about $\theta_\lambda \lambda$ with a standard deviation of $1/\sqrt{\theta_\lambda n}$.

Before giving the explicit statement on the concentration of Λ_C, we elaborate on its important role in understanding the structure of the 2-core of the graph. Until reaching Step 2 for the first time, the above algorithm repeatedly matches light clones until all of them are exhausted — precisely as the cut-off line reaches Λ_C. As stated in Section 2, the k-core of a graph can be obtained by repeatedly removing vertices of degree at most $k - 1$ (at any arbitrary order). Therefore, the 2-core is precisely comprised of all the unmatched clones at the moment we reach Λ_C. Crucially, continuing the algorithm will further reveal the inner structure of the 2-core, and these further steps are equivalent to running the configuration model on the clones to the left of Λ_C.

The following theorem gives tight concentration bounds for Λ_C. Its proof relies on a delicate analysis of the above mentioned Algorithm 1, and we postpone it to Section 7.

Theorem 3.3. *[Upper bound on the window of Λ_C] There exist some constant $c > 0$ so that for all $\gamma > 0$ with $\gamma = o\big(\sqrt{\theta_\lambda^3 n}\big)$, the following holds:*

$$\mathbb{P}\left(|\Lambda_C - \theta_\lambda \lambda| \geq \tfrac{\gamma}{\sqrt{\theta_\lambda n}}\right) \leq \mathrm{e}^{-c\gamma^2}. \tag{3.4}$$

3.1. **Size of the 2-core and its disjoint cycles.** Using the above theorem, we will now be able to characterize the structure of $G^{(2)}$. Indeed, by the discussion preceding the theorem, the 2-core of Poisson-cloning given that $\Lambda_C = \ell$ has the same distribution as the graph generated by the Poisson-configuration model given that $\Lambda = \ell$. The above theorem implies that w.h.p. we only need to consider

$$\ell = (1 + o(1))\theta_\lambda\lambda = (2 + o(1))\varepsilon\,,$$

and our next step is to estimate the basic properties of the 2-core (size, the number of vertices that comprise disjoint cycles) and its kernel on this event.

The next proposition thus applies not only to the Poisson-configuration model but also to the 2-core of Poisson-cloning. The term *expander* used here refers (informally) to a graph where the ratio between the boundary and volume of each set is bounded from below by some constant $c > 0$ (a precise definition appears below).

Proposition 3.4. *Let H be generated by the Poisson-configuration model given $\Lambda = \ell$, where $\ell = (2 + o(1))\varepsilon$. Define H' as the graph obtained by deleting every disjoint cycle from H. Let N_2 be the number of vertices with degree 2 in H, and N_2' be the corresponding quantity for H'. Then w.h.p.*

$$N_2 = (2 + o(1))\varepsilon^2 n, \qquad\qquad N_2' = (1 + o(1))N_2\,.$$

In addition, w.h.p. the kernel $\mathcal{K}$ of H is an expander graph with

$$|\mathcal{K}| = \left(\tfrac{4}{3} + o(1)\right)\varepsilon^3 n\,, \qquad\qquad |E(\mathcal{K})| = (2 + o(1))\varepsilon^3 n\,.$$

The first step in the proof is to establish the size of the kernel $\mathcal{K} = \ker(H)$, as well as show that it is an expander. This latter fact is of independent interest and will have important applications, e.g., for the mixing time of the random walk on $\mathcal{C}_1$. In what follows, for a subset S of the vertices of a graph G, we let

$$d_G(S) \triangleq \sum_{v \in S} d_G(v)$$

denote the sum of the degrees of its vertices (also referred to as the *volume* of S in G). Further define the *isoperimetric number* of a graph G, denoted by $i(G)$, as

$$i(G) \triangleq \min\left\{\frac{e(S, S^c)}{d_G(S)} \ : \ S \subset V(G)\,, d_G(S) \leq |E(G)|\right\}\,.$$

We say that G is a *c-edge-expander* for some fixed $c > 0$ iff $i(G) > c$.

Lemma 3.5. *For $\mathcal{K}$ the kernel of H as defined in Proposition 3.4, w.h.p.*

$$|\mathcal{K}| = \left(\tfrac{4}{3} + o(1)\right)\varepsilon^3 n\,, \quad |E(\mathcal{K})| = (2 + o(1))\varepsilon^3 n\,,$$

and $\mathcal{K}$ is an α-edge-expander for some constant $\alpha > 0$.

Proof. By Definition 3.1, the kernel consists of exactly those vertices $u \in V$ that have $D_u \geq 3$. Combining this with the assumption that $\Lambda = (2+o(1))\varepsilon$, it follows that $|\mathcal{K}| \sim \text{Bin}(n, p_3^+(\Lambda))$, where

$$p_3^+(\Lambda) \triangleq \sum_{k \geq 3} e^{-\Lambda} \frac{\Lambda^k}{k!} = e^{-\Lambda}(1 + O(\Lambda)) \frac{\Lambda^3}{3!} = \left(\tfrac{4}{3} + o(1)\right) \varepsilon^3 .$$

Since $\varepsilon^3 n \to \infty$, we get that w.h.p.

$$|\mathcal{K}| = \left(\tfrac{4}{3} + o(1)\right) \varepsilon^3 n . \tag{3.5}$$

Similarly, the total sum of degrees in $\mathcal{K}$ is simply $\sum_{u \in \mathcal{K}} D_u$, and therefore, $|E(\mathcal{K})|$ is the sum of n i.i.d. variables distributed as $Y \sim \text{Po}(\Lambda)\mathbf{1}_{[3,\infty)}$. A similar calculation to the one above now gives that $\mathbb{E}Y = (4 + o(1))\varepsilon^3$, and so (by CLT) w.h.p.

$$|E(\mathcal{K})| = (2 + o(1))\varepsilon^3 n . \tag{3.6}$$

To show that the isoperimetric number is bounded away from 0, we apply standard techniques used to analyze the configuration model (for definitions, see Subsection 2.2), while assuming (3.5) and (3.6). Let $0 < \alpha < \tfrac{1}{4}$ be specified later, and let

$$D = 2|E(\mathcal{K})| = (4 + o(1))\varepsilon^3 n$$

be the total number of points to be matched in the configuration model. We will next prove a lower bound $i(\mathcal{K}) > \alpha > 0$ for the case where D is even, and under the relaxed condition that perhaps one of the vertices of $\mathcal{K}$ has degree 2 (all others have degree 3 or more).

To see that this gives a bound on $i(\mathcal{K})$ for D odd, recall that in that case precisely one of the D points will have a self-loop (by Definition 3.1). Clearly, omitting one point produces a kernel as handled above (with an isoperimetric number at least α), and reintroducing it (to a vertex with at least 2 other points) would give $i(\mathcal{K}) \geq \tfrac{2}{3}\alpha$.

Consider the probability that $\{e(S, S^c) \leq \alpha d_{\mathcal{K}}(S)\}$, where S is a subset of $\mathcal{K}$ with $d_{\mathcal{K}}(S) = s$. This is precisely the probability that $k \leq \alpha s$ points out of the s that comprise S are matched with points in S^c, whereas the remaining $s - k$ points form a perfect matching. Thus,

$$\mathbb{P}(e(S, S^c) \leq \alpha s) = \sum_{k=0}^{\alpha s} \mathbb{P}(e(S, S^c) = k)\mathbf{1}_{\{s \equiv k \pmod 2\}}$$

$$\leq \frac{1}{(D-1)!!} \sum_{k=0}^{\alpha s} \binom{s}{k}(s - k - 1)!! \binom{D-s}{k} k!(D - s - k - 1)!!$$

$$\leq \frac{4}{D!!} \sum_{k=0}^{\alpha s} \binom{s}{k}(s - k)!! \binom{D-s}{k} k!(D - s - k)!! ,$$

where we used the fact that $\frac{m!!}{(m-1)!!} < 2\sqrt{m}$ for sufficiently large m, and that $\sqrt{(D-s-k)(s-k)} \geq \sqrt{D-2k} \geq \sqrt{D/2}$, as $k \leq \alpha s \leq s/4$ and $s \leq D/2$.

A standard application of Stirling's formula gives $n!! = \Theta((n!)^{1/2} n^{1/4})$. Hence, for some constant $c_1 > 0$,

$$\mathbb{P}(e(S, S^c) \leq \alpha s) \leq c_1 \sum_{k=0}^{\alpha s} \frac{(D-s-k)^{1/4}}{(D!)^{1/2} D^{1/4}} \frac{s!(s-k)^{1/4}}{k!((s-k)!)^{1/2}} \frac{(D-s)!}{((D-s-k)!)^{1/2}}$$

$$= c_1 \sum_{k=0}^{\alpha s} \left(\frac{(D-s-k)(s-k)}{D} \right)^{1/4} \left(\frac{\binom{s}{k}\binom{D-s}{k}}{\binom{D}{s}} \right)^{1/2}$$

$$\leq c_1 s^{1/4} \sum_{k=0}^{\alpha s} \left(\frac{\binom{s}{k}\binom{D-s}{k}}{\binom{D}{s}} \right)^{1/2}.$$

It is well known that, by Stirling's formula, $\binom{n}{k} \asymp \sqrt{\frac{n}{k(n-k)}} \exp\left[-H(\frac{k}{n})n\right]$, where $H(x)$ is the entropy function $H(x) \triangleq -x \log x - (1-x)\log(1-x)$. Thus, for some constant $c_2 > 0$,

$$\mathbb{P}(e(S, S^c) \leq \alpha s) \leq c_2 \sqrt{s} \sum_{k=0}^{\alpha s} e^{\frac{1}{2}\left[H\left(\frac{k}{s}\right)s + H\left(\frac{k}{D-s}\right)(D-s) - H\left(\frac{s}{D}\right)D\right]}$$

$$\leq c_2 s^{3/2} e^{\frac{1}{2}\left[H(\alpha)s + H\left(\frac{2\alpha s}{D}\right)D - H\left(\frac{s}{D}\right)D\right]},$$

where we applied the fact that $\frac{k}{D-s} \leq \frac{\alpha s}{D/2}$. Recalling that each vertex of $\mathcal{K}$ has degree at least 3, except for possibly one vertex of degree 2, we have

$$s = d_{\mathcal{K}}(S) \geq 3|S| - 1, \quad D \geq 3|\mathcal{K}| - 1,$$

and it follows that $|S| \leq \frac{s+1}{3}$. Thus,

$$\sum_{S:d_{\mathcal{K}}(S)=s} \mathbb{P}(e(S, S^c) \leq \alpha s) \leq \sum_{l \leq \frac{s+1}{3}} \sum_{\substack{S:d_{\mathcal{K}}(S)=s \\ |S|=l}} \mathbb{P}(e(S, S^c) < \alpha s)$$

$$\leq \sum_{l \leq \frac{s+1}{3}} c_2 s^{3/2} \binom{|\mathcal{K}|}{l} e^{\frac{1}{2}\left[H(\alpha)s + H\left(\frac{2\alpha s}{D}\right)D - H\left(\frac{s}{D}\right)D\right]}.$$

Since $D = (3+o(1))|\mathcal{K}|$, any $s \leq D/2$ satisfies $s/3 \leq (\frac{1}{2}+o(1))|\mathcal{K}|$. Therefore, another application of the above estimate of the binomial coefficient gives that for some constant $c_3 > 0$,

$$\sum_{S:d_{\mathcal{K}}(S)=s} \mathbb{P}(e(S, S^c) \leq \alpha s) \leq c_3 s^{5/2} e^{H\left(\frac{s}{3|\mathcal{K}|}+o(1)\right)|\mathcal{K}| + \frac{1}{2}\left[H(\alpha)s + H\left(\frac{2\alpha s}{D}\right)D - H\left(\frac{s}{D}\right)D\right]}$$

$$= c_3 s^{5/2} e^{(\frac{1}{3}+o(1))H\left(\frac{s}{D}\right)D + \frac{1}{2}\left[H(\alpha)\frac{s}{D} + H\left(\frac{2\alpha s}{D}\right) - H\left(\frac{s}{D}\right)\right]D}.$$

It is then clear that we can choose a sufficiently small $\alpha > 0$ such that

$$\mathbb{P}(e(S, S^c) \leq \alpha s) \leq c_3 s^{5/2} e^{-\frac{s}{10}H\left(\frac{s}{D}\right)D} \leq c_3 s^{5/2} e^{-\frac{s}{10}\log(D/s)}.$$

Combined with the fact that $D = (4+o(1))\varepsilon^3 n \to \infty$ by (3.5), and summing over the possible values of s, we deduce that

$$\mathbb{P}(\exists S \subset \mathcal{K}, \ d_{\mathcal{K}}(S) \leq |E(\mathcal{K})|, \ e(S, S^c) < \alpha d_{\mathcal{K}}(S)) = o(1) \ ,$$

as required. ∎

Proof of Proposition 3.4. The required statement on the typical number of vertices and edges in the kernel, $|\mathcal{K}|$ and $|E(\mathcal{K})|$ resp., has already been established in Lemma 3.5 (along with the expansion properties of the kernel). It remains to show that, w.h.p., N_2 and N_2' are both $(2 + o(1))\varepsilon^2 n$, where N_2 is the number of degree-2 vertices in the 2-core, and N_2' equals N_2 minus the number of vertices that belong to disjoint cycles in the 2-core.

The main issue left is to distinguish between H and H' (the graph before and after removing its disjoint cycles).

Let $p_2(x)$ be the probability that a $\mathrm{Po}(x)$ variable equals 2:

$$p_2(x) = \mathrm{e}^{-x} \frac{x^2}{2} \ .$$

By Definition 3.1 and our assumption that $\Lambda = (2 + o(1))\varepsilon$,

$$N_2 \sim \mathrm{Bin}(n, p_2(\Lambda)) \ , \ \text{and} \ \ p_2(\Lambda_C) = (2 + o(1))\varepsilon^2 \ .$$

Hence, by a standard concentration argument (using the fact that $\varepsilon^2 n \to \infty$) we deduce that $N_2 = (2 + o(1))\varepsilon^2 n$ w.h.p. Assume therefore that this is the case (i.e., there are $(2 + o(1))\varepsilon^2 n$ vertices of degree 2 in H), and that $|E(\mathcal{K})| = (2 + o(1))\varepsilon^3 n$.

We next consider the disjoint cycles in H. For a given degree-2 vertex v in H, let $A_{v,k}$ denote the event that v belongs to such a disjoint cycle whose length is k, and further let $A_v = \cup_{k=1}^{N_2} A_{v,k}$. To form a disjoint cycle of length k via the configuration model, we must repeatedly match points of degree-2 vertices, that is (noticing that $2|E(\mathcal{K})|$ counts the total number of points to be matched via the configuration model):

$$\mathbb{P}(A_{v,k}) = \frac{1}{2N_2 + 2|E(\mathcal{K})| - 2k + 1} \prod_{j=0}^{k-2} \frac{2N_2 - 2j - 1}{2N_2 + 2|E(\mathcal{K})| - 2j - 1}$$

$$= \frac{1}{2N_2 + 2|E(\mathcal{K})| - 1} \prod_{j=0}^{k-2} \frac{2N_2 - 2j - 1}{2N_2 + 2|E(\mathcal{K})| - 2j - 3}$$

$$\leq \frac{1 + o(1)}{4\varepsilon^2 n} \left(1 - (1 + o(1))\varepsilon\right)^{k-1} \ ,$$

since the terms in the above product over j formed a decreasing sequence. Summing over the values of k,

$$\mathbb{P}(A_v) = \sum_{k=1}^{N_2} \mathbb{P}(A_{v,k}) \leq \frac{1+o(1)}{4\varepsilon^2 n} \sum_{k=1}^{N_2} (1-(1+o(1))\varepsilon)^{k-1}$$

$$\leq \frac{1+o(1)}{4\varepsilon^2 n} \cdot \frac{1+o(1)}{\varepsilon} = \frac{1+o(1)}{4\varepsilon^3 n} = o(1) \ . \tag{3.7}$$

It then follows that w.h.p. $N_2' = (1-o(1))N_2 = (2+o(1))\varepsilon^2 n$, as required. ■

3.2. **Contiguity of Poisson-cloning and Poisson-configuration.** A key part of showing the contiguity result is a counterpart for Theorem 3.3, which together implies that Λ_C has a tight concentration window of order $1/\sqrt{\varepsilon n}$.

Theorem 3.6. *[Lower bound on the window of Λ_C] There exist some constant $c > 0$ such that for any $t = t(n) > 0$ and fixed $\delta > 0$,*

$$\mathbb{P}\left(t \leq \Lambda_C \leq t + \frac{\delta}{\sqrt{\varepsilon n}}\right) \leq c\delta \ . \tag{3.8}$$

Proof. The results of the previous subsection imply that, for some suitably large constant $M = M(\delta) > 0$,

$$\mathbb{P}\left(|\Lambda_C - \lambda\theta_\lambda| \geq \frac{M}{\sqrt{\varepsilon n}}\right) \leq \delta \ .$$

It follows that (3.8) holds trivially for any $c \geq 1$ when $|t - \lambda\theta_\lambda| \geq \frac{M+1}{\sqrt{\varepsilon n}}$. Therefore, we assume in what follows $|t - \lambda\theta_\lambda| < \frac{M+1}{\sqrt{\varepsilon n}}$. Denote by A the event $\{t \leq \Lambda_C \leq t + \delta/\sqrt{\varepsilon n}\}$. Recall the fact that $|G^{(2)}| \sim \mathrm{Bin}(n, p_2^+(\Lambda_C))$, where $G^{(2)}$ is the 2-core of the Poisson-cloning model and $p_2^+(x)$ stands for the probability for a $\mathrm{Po}(x)$ variable to be at least 2. Standard analytical arguments give that

$$p_2^+\left(t + \frac{\delta}{\sqrt{\varepsilon M}}\right) - p_2^+(t) \leq \frac{3\delta}{\sqrt{n/\varepsilon}} \ .$$

Now, an application of CLT implies that for some interval B of length $4\delta\sqrt{\varepsilon n}$,

$$\mathbb{P}\left(|G^{(2)}| \in B \,\big|\, A\right) = 1 - o(1) \ . \tag{3.9}$$

Consider $\mathcal{C}_1^{(2)}$, the 2-core of the giant component in the Poisson-cloning. Recalling (3.7) (which, as the discussion before Proposition 3.4, also applies to the Poisson-cloning model), we know that w.h.p. only an $\frac{w}{\varepsilon^3 n}$ fraction of vertices in $G^{(2)}$ will appear in disjoint cycles, where $w = (\varepsilon^3 n)^{1/4}$ ($\to \infty$). Therefore, we have

$$|\mathcal{C}_1^{(2)}| = |G^{(2)}|\left(1 - \frac{w}{\varepsilon^3 n}\right) = |G^{(2)}| - O\left(\frac{w}{\varepsilon}\right) = |G^{(2)}| - o(\sqrt{\varepsilon n}) \ .$$

Together with (3.9), we conclude that there exists an interval B' with length $5\delta\sqrt{\varepsilon n}$ such that

$$\mathbb{P}\left(|\mathcal{C}_1^{(2)}| \in B' \,\big|\, A\right) = 1 - o(1) \,. \tag{3.10}$$

Now, for the 2-core H of the giant component of the Erdős-Rényi graph $\mathcal{G}(n, p)$, it is known (see [31, Theorem 6], reformulated here in Theorem 5.1) that $|H|$ is in the limit Gaussian with variance $(12 + o(1))\varepsilon n$. Therefore,

$$\mathbb{P}(|H| \in B') \leq 5\delta \,.$$

Combining this with contiguity of Poisson-cloning and $\mathcal{G}(n, p)$ (as stated in Theorem 2.1), we obtain that for some constant $c_0 > 0$

$$\mathbb{P}(|\mathcal{C}_1^{(2)}| \in B') \leq 5c_0\delta \,.$$

The proof is completed by choosing $c = 5c_0 + 1$ and applying (3.10). ∎

Using the above estimate for Λ_C, we are now able to conclude the main result of this section, which reduces the 2-core of Poisson-cloning to the graph generated by the Poisson-configuration model.

Proof of Theorem 3.2. Recall that H is the random graph generated by the Poisson-configuration model, and $G^{(2)}$ is the 2-core of Poisson-cloning.

Let $\delta > 0$, and set

$$B = \left(\lambda\theta_\lambda - M\sqrt{\varepsilon n}, \ \lambda\theta_\lambda + M\sqrt{\varepsilon n}\right),$$

where $M = M(\delta)$ is a sufficiently large constant such that $\mathbb{P}(\Lambda_C \in B) \geq 1 - \delta$. Further define

$$D \triangleq \{x : \mathbb{P}(H \in \mathcal{A} \mid \Lambda = x) \geq \delta\} \,.$$

Since $\mathbb{P}(H \in \mathcal{A}) = o(1)$, we obtain that $\mathbb{P}(\Lambda \in D) \to 0$, and consequently

$$\mathbb{P}(\Lambda \in B \cap D) = o(1) \,.$$

Recalling that $\Lambda \sim \mathcal{N}\left(\lambda\theta_\lambda, 1/(\varepsilon n)\right)$, we deduce that $\mathcal{L}(B \cap D)\sqrt{\varepsilon n} \to 0$, where $\mathcal{L}(\cdot)$ stands for the Lebesgue measure on $\mathbb{R}$. At this point, Theorem 3.6 gives that $\mathbb{P}(\Lambda_C \in B \cap D) \to 0$. Recalling that $G^{(2)}$ and H are generated by the same scheme (and hence have the same distribution) given the event $\Lambda_C = \Lambda$, we obtain that

$$\mathbb{P}(G^{(2)} \in \mathcal{A}) \leq 2\delta + o(1) \,,$$

as required. ∎

4. Constructing the 2-core of the random graph

In the previous section, we have shown that the 2-core of Poisson-cloning is contiguous to a simpler model, which we called the Poisson-configuration model (see Definition 3.1). The goal of this section is to reduce the Poisson-configuration model to the following, where here and in what follows, μ is defined to be the conjugate of $\lambda = 1 + \varepsilon$. That is to say, $\mu < 1$ and

$$\mu \mathrm{e}^{-\mu} = \lambda \mathrm{e}^{-\lambda} \ . \tag{4.1}$$

Definition 4.1 (Poisson-geometric model for n and $p = \frac{1+\varepsilon}{n}$)**.**

(1) Let $\Lambda \sim \mathcal{N}\left(1 + \varepsilon - \mu, \frac{1}{\varepsilon n}\right)$ and assign an independent $\mathrm{Po}(\Lambda)$ variable D_u to each vertex u. Let $N_k = \#\{u : D_u = k\}$ and $N = \sum_{k \geq 3} N_k$.
(2) Construct a random graph $\mathcal{K}$ on N vertices, uniformly chosen over all graphs with N_k degree-k vertices for $k \geq 3$ (if $\sum_{k \geq 3} k N_k$ is odd, choose a vertex u with $D_u = k \geq 3$ with probability proportional to k, and give it $k - 1$ half-edges and a self-loop).
(3) Replace the edges of $\mathcal{K}$ by paths of length i.i.d. $\mathrm{Geom}(1 - \mu)$.

Theorem 4.2. *Let H be generated by the Poisson-configuration model w.r.t. n and $p = \frac{1+\varepsilon}{n}$, where $\varepsilon \to 0$ and $\varepsilon^3 n \to \infty$. Let $\tilde{H}$ be generated by the Poisson-geometric model corresponding to n, p. Then for any set of graphs $\mathcal{A}$ such that $\mathbb{P}(\tilde{H} \in \mathcal{A}) \to 0$, we have $\mathbb{P}(H \in \mathcal{A}) \to 0$.*

Clearly, both models have the same kernel, and they only differ in the way this kernel is thereafter expanded to form the entire graph (replacing edges by paths). To prove the above statement, we need to estimate the distribution of the total number of edges in each of the models; we will show that they are in fact contiguous.

4.1. Edge distribution in the Poisson-configuration model.

Lemma 4.3. *Let N_k denote the number of degree-k vertices in the Poisson-configuration model, and set $\Lambda_0 = \lambda - \mu$. For any fixed $M > 0$ there exist some $c_1, c_2 > 0$ such that the following holds: If $n_3, n_4, \ldots$ satisfy*

$$\begin{aligned} \left| n \left(1 - \mathrm{e}^{-\Lambda_0}(1 + \Lambda_0 + \tfrac{\Lambda_0^2}{2})\right) - \sum_{k \geq 3} n_k \right| &\leq M\sqrt{\varepsilon^3 n} \,, \\ \left| n \Lambda_0 \left(1 - \mathrm{e}^{-\Lambda_0}(1 + \Lambda_0)\right) - \sum_{k \geq 3} k n_k \right| &\leq M\sqrt{\varepsilon^3 n} \end{aligned}$$

and x satisfies $|x - \Lambda_0| \leq \frac{M}{\sqrt{\varepsilon n}}$ then

$$c_1 \leq \frac{\mathbb{P}\left(N_k = n_k \ \ for \ all \ k \geq 3 \,\big|\, \Lambda = x\right)}{\mathbb{P}\left(N_k = n_k \ \ for \ all \ k \geq 3 \,\big|\, \Lambda = \Lambda_0\right)} \leq c_2 \,.$$

Proof. Throughout the proof of the lemma, the implicit constants in the $O(\cdot)$ notation depend on M.

Write $m = \sum_{k \geq 3} n_k$ and $r = \sum_{k \geq 3} k n_k$, and let $A = A(n_3, n_4, \ldots)$ denote the event $\{N_k = n_k \text{ for all } k \geq 3\}$. As usual, we use the abbreviations

$$p_k(x) = \mathbb{P}(\mathrm{Po}(x) = k) = \mathrm{e}^{-x} x^k / k! \,, \quad \text{and} \quad p_k^- = \mathbb{P}(\mathrm{Po}(x) \leq k) \,.$$

It follows that

$$\frac{\mathbb{P}\left(A \mid \Lambda = x\right)}{\mathbb{P}\left(A \mid \Lambda = \Lambda_0\right)} = \left(\frac{p_2^-(x)}{p_2^-(\Lambda_0)}\right)^{n-m} \prod_k \left(\frac{p_k(x)}{p_k(\Lambda_0)}\right)^{n_k}$$

$$= \mathrm{e}^{-n(x-\Lambda_0)} \left(\frac{1 + x + \frac{x^2}{2}}{1 + \Lambda_0 + \frac{\Lambda_0^2}{2}}\right)^{n-m} \left(\frac{x}{\Lambda_0}\right)^r \,,$$

and so

$$\log \frac{\mathbb{P}\left(A \mid \Lambda = x\right)}{\mathbb{P}\left(A \mid \Lambda = \Lambda_0\right)} = n(\Lambda_0 - x) + (n - m) \log \left(\frac{1 + x + \frac{x^2}{2}}{1 + \Lambda_0 + \frac{\Lambda_0^2}{2}}\right) + r \log \frac{x}{\Lambda_0} \,.$$

Using Taylor's expansion and recalling that $x - \Lambda_0 = O(1/\sqrt{\varepsilon n}) = o(\Lambda_0)$,

$$\log \left(\frac{1 + x + \frac{x^2}{2}}{1 + \Lambda_0 + \frac{\Lambda_0^2}{2}}\right) = \frac{1 + \Lambda_0}{1 + \Lambda_0 + \frac{\Lambda_0^2}{2}} (x - \Lambda_0) - (\Lambda_0 - o(1))(x - \Lambda_0)^2$$

$$= \frac{1 + \Lambda_0}{1 + \Lambda_0 + \frac{\Lambda_0^2}{2}} (x - \Lambda_0) + O(1/n) \,,$$

and we deduce that

$$\log \frac{\mathbb{P}\left(A \mid \Lambda = x\right)}{\mathbb{P}\left(A \mid \Lambda = \Lambda_0\right)} = n(\Lambda_0 - x) + (n - m) \frac{1 + \Lambda_0}{1 + \Lambda_0 + \frac{\Lambda_0^2}{2}} (x - \Lambda_0) - O(1)$$

$$+ r \frac{x - \Lambda_0}{\Lambda_0} - O\left(r/\varepsilon^3 n\right) \,.$$

Our assumptions on m, r now yield that

$$\log \frac{\mathbb{P}\left(A \mid \Lambda = x\right)}{\mathbb{P}\left(A \mid \Lambda = \Lambda_0\right)} = n(\Lambda_0 - x) + n \mathrm{e}^{-\Lambda_0}(1 + \Lambda_0)(x - \Lambda_0)$$

$$+ n\left(1 - \mathrm{e}^{-\Lambda_0}(1 + \Lambda_0)\right)(x - \Lambda_0) + O(1) = O(1) \,,$$

completing the proof. ∎

Fix $M > 0$, and let $\mathcal{B}_M$ denote the following set of "good" kernels:

$$\mathcal{B}_M \triangleq \left\{ \mathcal{K} : \begin{array}{l} \left| |\mathcal{K}| - n\left(1 - \mathrm{e}^{-\Lambda_0}(1 + \Lambda_0 + \frac{\Lambda_0^2}{2})\right) \right| \leq M\sqrt{\varepsilon^3 n} \\ \left| |E(\mathcal{K})| - \frac{1}{2} n \Lambda_0 \left(1 - \mathrm{e}^{-\Lambda_0}(1 + \Lambda_0)\right) \right| \leq M\sqrt{\varepsilon^3 n} \end{array} \right\} . \quad (4.2)$$

Let $f_\Lambda(\cdot \mid \cdot)$ denote the density function of Λ given the kernel $\mathcal{K}$ (or equivalently, given its degree sequence). By applying Bayes' formula, the above lemma gives that

$$\frac{f_\Lambda(x \mid \mathcal{K})}{f_\Lambda(\Lambda_0 \mid \mathcal{K})} = \Theta(1)$$

for all $\mathcal{K} \in \mathcal{B}_M$ and x in the interval $I_M = [\Lambda_0 - \frac{M}{\sqrt{\varepsilon n}}, \Lambda_0 + \frac{M}{\sqrt{\varepsilon n}}]$. Clearly, by volume considerations, this implies that for some $c = c(M) > 0$ we have

$$f_\Lambda(x \mid \mathcal{K}) \leq c\sqrt{\varepsilon n} \quad \text{for all } x \in I_M \text{ and } \mathcal{K} \in \mathcal{B}_M. \tag{4.3}$$

Lemma 4.4. *Define $M > 0$, I_M and $\mathcal{B}_M$ as above. Let H be generated by the Poisson-configuration model. There exists some constant $c = c(M) > 0$ so that for any $\mathcal{K} \in \mathcal{B}_M$ and s with $\left| s - \frac{n}{2}\left(\Lambda_0 - e^{-\Lambda_0}\Lambda_0\right) \right| \leq M\sqrt{\varepsilon n}$,*

$$\mathbb{P}\left(|E(H)| = s\,,\ \Lambda \in I_M \,\big|\, \ker(H) = \mathcal{K}\right) \leq \frac{c}{\sqrt{\varepsilon n}}\,.$$

Proof. Let $x \in I_M$ and $\mathcal{K} \in \mathcal{B}_M$, and write $m = |\mathcal{K}|$ and $r = |E(\mathcal{K})|$ for the number of vertices and the edges in the kernel respectively. We will first estimate $\mathbb{P}(|E(H)| = s \mid \Lambda = x\,,\ \ker(H) = \mathcal{K})$, and the required inequality will then readily follow from an integration over $x \in I_M$.

Note that, given $\Lambda = x$ and $\ker(H) = \mathcal{K}$, the number of edges in H is the r edges of $\mathcal{K}$ plus an added edge for each degree 2 variable out of the $n - m$ variables (i.i.d. $\mathrm{Po}(x)$) that have $\{u : D_u \leq 2\}$. That is, in this case

$$|E(H)| \sim r + \mathrm{Bin}\left(n - m,\ \frac{x^2/2}{1 + x + x^2/2}\right),$$

and therefore,

$$\mathbb{P}\left(|E(H)| = s \,\big|\, \Lambda = x\,,\ \ker(H) = \mathcal{K}\right)$$
$$= \binom{n - m}{s - r}\left(\frac{x^2/2}{1 + x + x^2/2}\right)^{s-r}\left(\frac{1 + x}{1 + x + x^2/2}\right)^{n-m-(s-r)}.$$

Write

$$q_0 = \frac{\Lambda_0^2/2}{1 + \Lambda_0 + \Lambda_0^2}$$

and define

$$q = q(x) = \frac{x^2/2}{1 + x + x^2/2} \qquad \text{and} \qquad t = \frac{s - r}{n - m}\,.$$

Since $x \in I_M$ and $\mathcal{K} \in \mathcal{B}_M$, we have

$$q = q_0 + O(\sqrt{\varepsilon/n})\,, \quad \text{and } t = q_0 + O(\sqrt{\varepsilon/n})\,. \tag{4.4}$$

Using Stirling's formula, we obtain that

$$\mathbb{P}\left(|E(H)| = s \mid \Lambda = x\,,\ \ker(H) = \mathcal{K}\right)$$

$$\leq \left(\frac{(1+o(1))(n-m)}{2\pi(s-r)\,(n-m-(s-r))}\right)^{1/2} \left(\frac{q}{t}\right)^{s-r} \left(\frac{1-q}{1-t}\right)^{n-m-s+r}$$

$$\leq \frac{1}{\sqrt{s-r}}e^{(n-m)g_t(x)}\,,$$

where $g_t(q)$ is given by

$$g_t(q) \stackrel{\triangle}{=} -t\log t - (1-t)\log(1-t) + t\log q + (1-t)\log(1-q)\,.$$

It is easy to verify that

$$g_t'(q) = -\frac{1-t}{1-q} + \frac{t}{q}\,,$$

and so for any q, t satisfying (4.4) we have

$$g_t(t) = 0\,, \qquad g_t'(t) = 0\,,$$

$$g_t''(q) = -\frac{t}{q^2} - \frac{1-t}{(1-q)^2} = -\frac{1+o(1)}{q_0}\,.$$

Thus, for any large n (absorbing the $o(1)$-term in the constant) we have

$$g_t(q) \leq -\frac{1}{3q_0}(q-t)^2\,. \tag{4.5}$$

Clearly, the function $q(x) = \frac{x^2/2}{1+x+x^2/2}$ satisfies $q'(x) = \frac{x(2+x)}{2(1+x+x^2/2)^2}$, and in particular q is strictly monotone increasing from 0 to 1 for $x \in [0, \infty)$. Thus, there exists a unique $x_t > 0$ such that $q(x_t) = t$. Noticing that for all $x = (1+o(1))\Lambda_0$ we have $q'(x) = (1+o(1))\Lambda_0 = (2+o(1))\varepsilon$, it follows that for any $M_1 > 0$ one can choose $M_2 > 0$ such that

$$\left[\Lambda_0 - M_1\sqrt{\varepsilon/n},\, \Lambda_0 + M_1\sqrt{\varepsilon/n}\right] \subset q\left(\left[\Lambda_0 - \tfrac{M_2}{\sqrt{\varepsilon n}},\, \Lambda_0 + \tfrac{M_2}{\sqrt{\varepsilon n}}\right]\right)\,,$$

and in particular, $x_t = \Lambda_0 + O(1/\sqrt{\varepsilon n})$. We can now apply the Mean Value Theorem to q in (4.5) and obtain that

$$g_t(q) \leq -\frac{1}{3q_0}(q(x) - q(x_t))^2 = -\frac{((1+o(1))\Lambda_0)^2}{3q_0}(x - x_t)^2 \leq -\frac{3}{5}(x - x_t)^2\,,$$

where the last inequality holds for any sufficiently large n. Altogether, absorbing the change from $(n-m)$ to n in the constant, we conclude that

$$\mathbb{P}(|E(H)| = s \mid \Lambda = x,\ \ker(H) = \mathcal{K})$$

$$\leq \frac{1}{\sqrt{s-r}}e^{-\frac{1}{2}n(x-x_t)^2} \leq \frac{1}{\sqrt{\varepsilon^2 n}}e^{-\frac{1}{2}n(x-x_t)^2}\,. \tag{4.6}$$

It remains to integrate the above conditional probability over $x \in I_M$. Combining (4.3) and (4.6), we obtain that for some constant $c > 0$

$$
\mathbb{P}\left(|E(H)| = s \ , \ \Lambda \in I_M \,\big|\, \ker(H) = \mathcal{K}\right)
$$

$$
\leq \int_{I_M} \mathbb{P}\left(|E(H)| = s \,\big|\, \Lambda = x \ , \ \ker(H) = \mathcal{K}\right) f_\Lambda(x|\mathcal{K})\, dx
$$

$$
\leq \int_{I_M} \frac{1}{\sqrt{\varepsilon^2 n}} \mathrm{e}^{-\frac{1}{2}n(x-x_t)^2} c\sqrt{\varepsilon n} \leq \frac{c}{\sqrt{\varepsilon n}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}y^2}\, dy = \frac{c\sqrt{2\pi}}{\sqrt{\varepsilon n}},
$$

as required. ∎

4.2. Edge distribution in the Poisson-geometric model.

Lemma 4.5. *Let $M > 0$ and $\mathcal{B}_M$ be as in (4.2). Let $\tilde{H}$ be generated by the Poisson-geometric model. There exists some constant $c = c(M) > 0$ so that for any $\mathcal{K} \in \mathcal{B}_M$ and s with $\left|s - \frac{n}{2}(\lambda - \mu)\left(1 - \frac{\mu}{\lambda}\right)\right| \leq M\sqrt{\varepsilon n}$,*

$$
\mathbb{P}\left(|E(\tilde{H})| = s \,\big|\, \ker(\tilde{H}) = \mathcal{K}\right) \geq \frac{c}{\sqrt{\varepsilon n}}\,.
$$

Proof. By definition, given that $\ker(\tilde{H}) = \mathcal{K}$, the variable $|E(\tilde{H})$ is the sum of $|E(\mathcal{K})|$ i.i.d. geometric variables with mean $1/(1-\mu)$.

Denote by r the number of edges in the kernel $\mathcal{K}$, and let s be a candidate for the number of edges in the expanded 2-core $\tilde{H}$. As stated in the lemma (recall definition (4.2)), we are interested in the following range for r, s:

$$
r = \frac{n}{2}(\lambda - \mu)(1 - \tfrac{\mu}{\lambda})(1 - \mu) + c_1\sqrt{\varepsilon^3 n}, \qquad (|c_1| \leq M)\,,
$$

$$
s = \frac{n}{2}(\lambda - \mu)(1 - \tfrac{\mu}{\lambda}) + c_2\sqrt{\varepsilon n}, \qquad (|c_2| \leq M)\,.
$$

In this case, we have that

$$
\frac{s-1}{r-1} = \frac{1}{1-\mu} + \frac{c_2\sqrt{\varepsilon n} - c_1\sqrt{\varepsilon^3 n}/(1-\mu) - 1}{\frac{n}{2}(\lambda - \mu)(1 - \tfrac{\mu}{\lambda})(1-\mu) + c_1\sqrt{\varepsilon^3 n} - 1} = \frac{1 + \xi}{1 - \mu}\,, \quad (4.7)
$$

where $\xi \triangleq \frac{1 + o(1)}{2}\frac{c_2 - c_1}{\sqrt{\varepsilon^3 n}}$. Let X_i be independent geometric random variables with mean $\frac{1}{1-\mu}$, i.e., $\mathbb{P}(X_i = k) = \mu^{k-1}(1-\mu)$ for $k = 1, 2, \ldots$; further set $S_k = \sum_{i=1}^{k}(X_i - 1)$. Since S_k follows a negative binomial distribution,

$$
\mathbb{P}(S_r = s - r) = \binom{s-1}{r-1}(1-\mu)^r \mu^{s-r}\,.
$$

Using Stirling's formula, we get that for some constant $c_3 > 0$

$$
\mathbb{P}(S_r = s - r) \geq \frac{c_3}{\sqrt{r}}\left(\frac{s-1}{r-1}\right)^{r-1}\left(\frac{s-1}{s-r}\right)^{s-r}(1-\mu)^r \mu^{s-r}\,.
$$

Substituting (4.7) in the above, and using the fact that

$$\frac{s-1}{s-r} = \frac{(1+\xi)/(1-\mu)}{(1+\xi)/(1-\mu)-1} = \frac{1+\xi}{\xi+\mu} \ ,$$

we obtain that

$$\mathbb{P}(S_r = s - r) \geq \frac{c_3}{\sqrt{r}} \Big(\frac{1+\xi}{1-\mu}\Big)^{r-1} \Big(\frac{1+\xi}{\xi+\mu}\Big)^{s-r} (1-\mu)^r \mu^{s-r}$$

$$= \frac{c_3(1-\mu)}{\sqrt{r}} (1+\xi)^{s-1} \Big(\frac{\mu}{\xi+\mu}\Big)^{s-r}$$

$$= \frac{c_3(1-\mu)}{\sqrt{r}} \exp\Big(g(\xi)\frac{r-1}{1-\mu}\Big) \ ,$$

where

$$g(x) \triangleq (1+x)\log(1+x) + (x+\mu)\log\Big(\frac{\mu}{x+\mu}\Big) \ .$$

Clearly, we have that $g(0) = 0$ and a standard calculation yields that

$$g'(x) = \log(1+x) + \log\Big(\frac{\mu}{x+\mu}\Big) \text{ and } g''(x) = \frac{\mu-1}{(x+\mu)(1+x)} \ .$$

In particular, we have that $g'(0) = 0$ and $|g''(x)| \leq 2(1-\mu)$ when $|x| \leq |\xi|$, where ξ is defined as above (recall that $\xi = o(1)$). Therefore, for all such x we have $|g(x)| \leq 2(1-\mu)x^2$, and altogether,

$$\mathbb{P}(S_r = s - r) \geq \frac{c_3(1-\mu)}{\sqrt{r}} e^{-2(r-1)|\xi|^2} \ .$$

Since

$$r|\xi|^2 = \frac{1+o(1)}{2}(c_2 - c_1)^2 \leq (2+o(1))M^2 \ ,$$

we deduce that for some $c_3' > 0$,

$$\mathbb{P}(S_r = s - r) \geq c_3'/\sqrt{\varepsilon n} \ ,$$

completing the proof. ∎

4.3. Contiguity of the two models.
We are now ready to prove the main result of this section, Theorem 4.2, which reduces the Poisson-configuration model to the Poisson-geometric model.

Proof. For some constant $M > 0$ to be specified later, define the event

$$A_M \triangleq \Big\{ \ \Lambda \in I_M, \ \ker(H) \in \mathcal{B}_M, \ \big||E(H)| - \tfrac{n}{2}(\lambda - \mu)(1 - \tfrac{\mu}{\lambda})\big| \leq M\sqrt{\varepsilon n} \ \Big\} \ .$$

Fix $\delta > 0$. We claim that for a sufficiently large $M = M(\delta)$ we have $\mathbb{P}(A_M) \geq 1 - \delta$. To see this, note the following:

1. In the Poisson-configuration model, $\Lambda \sim \mathcal{N}\big(\Lambda_0, \frac{1}{\varepsilon n}\big)$, and I_M includes at least M standard deviations about its mean.

2. Each of the variables $|\mathcal{K}|$ and $E(\mathcal{K})$ is a sum of i.i.d. random variables with variance $O(\varepsilon^3 n)$ and mean as specified in the definition of $\mathcal{B}_M$, hence their concentration follows from CLT.

3. Finally, $E(H)$ is again a sum of i.i.d. variables and has variance $O(\varepsilon n)$, only here we must subtract the vertices that comprise disjoint cycles. By (3.7) and the estimate in Proposition 3.4 on the size of the 2-core in the Poisson-configuration model, the number of such vertices is $O(1/\varepsilon)$ w.h.p. Compared to the standard deviation of $O(\sqrt{\varepsilon n})$, this amounts to a negligible error, as $\varepsilon^3 n \to \infty$.

Given an integer s and a kernel $\mathcal{K}$, let $\mathcal{D}_{s,\mathcal{K}}$ denote every possible 2-core with s edges and kernel $\mathcal{K}$. Crucially, the distribution of the Poisson-configuration model given $E(H) = s$ and $\ker(H) = \mathcal{K}$ is uniform over $\mathcal{D}_{s,\mathcal{K}}$, and so is the Poisson-geometric model given $E(\tilde{H}) = s$ and $\ker(\tilde{H}) = \mathcal{K}$. Therefore, for any graph $D \in \mathcal{D}_{s,\mathcal{K}}$,

$$\frac{\mathbb{P}(H = D \mid \ker(H) = \mathcal{K})}{\mathbb{P}(\tilde{H} = D \mid \ker(\tilde{H}) = \mathcal{K})} = \frac{\mathbb{P}(|E(H)| = s \mid \ker(H) = \mathcal{K})}{\mathbb{P}(|E(\tilde{H})| = s \mid \ker(\tilde{H}) = \mathcal{K})} .$$

Combining Lemmas 4.4, 4.5 we get that for some $c = c(M) > 0$,

$$\frac{\mathbb{P}(|E(H)| = s , A_M \mid \ker(H) = \mathcal{K})}{\mathbb{P}(|E(\tilde{H})| = s \mid \ker(\tilde{H}) = \mathcal{K})} \leq c .$$

Recalling that $\mathbb{P}(A_M) \geq 1 - \delta$ and letting $\delta \to 0$, we deduce that for any family of graphs $\mathcal{A}$, if $\mathbb{P}(\tilde{H} \in \mathcal{A}) \to 0$ then also $\mathbb{P}(H \in \mathcal{A}) \to 0$.   ∎

5. Constructing the giant component

Throughout the section, let $p = (1 + \varepsilon)/n$, where $\varepsilon \to 0$ and $\varepsilon^3 n \to \infty$ with n, and let G be a random graph $G \sim \mathcal{G}(n, p)$. We begin by analyzing the "bushes", i.e., the trees that are attached to $G^{(2)}$, the 2-core of G.

As before, $\mu < 1$ is defined to be the conjugate of $\lambda = 1 + \varepsilon$ (see (4.1)). Since $\varepsilon \to 0$, we can infer from a standard Taylor expansion that

$$\mu = 1 - \varepsilon + \tfrac{2}{3}\varepsilon^2 + O(\varepsilon^3) . \tag{5.1}$$

Proof of Theorem 2. In what follows, we use the abbreviation PGW(μ)-tree for a Poisson(μ)-Galton-Watson tree. Let $\widehat{\mathcal{C}}_1$ denote the graph obtained as follows:

- Let H be a copy of $\mathcal{C}_1^{(2)}$ (the 2-core of the giant component of G).
- For each $v \in H$, attach an independent PGW(μ) tree rooted at v.

By this definition, $\mathcal{C}_1$ and $\widehat{\mathcal{C}}_1$ have exactly the same 2-core H. For simplicity, we will refer directly to H as the 2-core of the model, whenever the context of either $\mathcal{C}_1$ or $\widehat{\mathcal{C}}_1$ is clear. We first establish the contiguity of $\mathcal{C}_1$ and $\widehat{\mathcal{C}}_1$.

Define the bushes of $\mathcal{C}_1$ as follows:

$$T_u \triangleq \{v \in \mathcal{C}_1 : v \text{ is connected to } u \text{ in } \mathcal{C}_1 \setminus H\} \quad \text{for } u \in H \,.$$

Clearly, each T_u is a tree as it is connected and does not contain any cycles (its vertices were not included in the 2-core). To conclude, we go from H to $\mathcal{C}_1$ by attaching a tree T_u to each vertex $u \in H$ (while identifying the root of T_u with u). Analogously, let $\{\tilde{T}_u\}_{u \in H}$ be the corresponding bushes in $\widehat{\mathcal{C}}_1$.

We next introduce notations for the labeled and unlabeled trees as well as their distributions. For $t \in \mathbb{N}$, let $\mathcal{R}_t$ be the set of all labeled rooted trees on the vertex set $[t]$, and let U_t be chosen uniformly at random from $\mathcal{R}_t$. For $T \in \mathcal{R}_t$ and a bijection ϕ on $[t]$, let $\phi(T)$ be the tree obtained by relabeling the vertices in T according to ϕ. Furthermore, define

$$T' \triangleq \{\phi(T) : \phi \text{ is a bijection on } [t]\}$$

to be the corresponding rooted unlabeled tree.

Let $\{t_u : u \in H\}$ be some integers. Conditioning on the event

$$\{\, |T_u| = t_u \text{ for all } u \in H \,\}\,,$$

we know from the definition of $\mathcal{G}(n,p)$ that T_u is distributed independently and uniformly among all labeled trees of size t_u rooted at u. In particular, in that case each T'_u is independently distributed as U'_{t_u} (the unlabeled counterparts of T_u and U_{t_u}).

On the other hand, Aldous [2] (see also, e.g., [3]) observed that, if T is a PGW-tree then T' has the same distribution as U'_t on the event $\{|T| = t\}$. Therefore, conditioning on the event

$$\{\, |\tilde{T}_u| = t_u \text{ for all } u \in H \,\}\,,$$

we also get that $\tilde{T}'_k$ has the same distribution as U'_{t_k}.

We therefore turn to study the sizes of the bushes in $\mathcal{C}_1$ and $\widehat{\mathcal{C}}_1$. Letting $\{t_u : u \in H\}$ be some integers and writing

$$N = \sum_{u \in H} t_u \,,$$

we claim that by definition of $\mathcal{G}(n,p)$ every extension of the 2-core H to the component $\mathcal{C}_1$, using trees whose sizes sum up to N, has the same probability. To see this, fix H, and notice that the probability of obtaining a component with a 2-core is H and an extension X connecting it to $N - |H|$ additional vertices only depends on the number of edges in H and X (and the fact that this is a legal configuration, i.e., H is a valid 2-core and X is comprised of trees). Therefore, upon conditioning on H the probabilities of the various

extensions X remain all equal. Cayley's formula gives that there are m^{m-1} labeled rooted trees on m vertices, and so,

$$
\begin{aligned}
&\mathbb{P}\left(|T_u| = t_u \text{ for all } u \in H \mid H\right) \\
&\quad = \mathbb{P}\left(|\mathcal{C}_1| = N \mid H\right) \mathbb{P}\left(|T_u| = t_u \text{ for all } u \in H \mid H, |\mathcal{C}_1| = N\right) \\
&\quad = \mathbb{P}\left(|\mathcal{C}_1| = N \mid H\right) \frac{1}{Z(N)} \frac{N!}{\prod_{u \in H} t_u!} \prod_{u \in H} t_u^{t_u - 1} \\
&\quad = \mathbb{P}(|\mathcal{C}_1| = N \mid H) \frac{1}{Z'(N)} \prod_{u \in H} \left[\frac{t_u^{t_u-1}}{\mu t_u!}(\mu \mathrm{e}^{-\mu})^{t_u}\right],
\end{aligned}
$$
(5.2)

where $Z(N)$ and $Z'(N)$ are the following normalizing constants

$$
Z'(N) = \sum_{\{r_u\}: \sum_{u \in H} r_u = N} \prod_{u \in H} \left[\frac{r_u^{r_u-1}}{\mu r_u!}(\mu \mathrm{e}^{-\mu})^{r_u}\right] ,
$$

$$
Z(N) = Z'(N)\mu^{N-|H|}\mathrm{e}^{-\mu N} .
$$

Notice that the size of a Poisson(γ)-Galton-Watson tree T follows a Borel(γ) distribution (see, e.g., [30]), namely,

$$
\mathbb{P}(|T| = t) = \frac{t^{t-1}}{\gamma t!}(\gamma \mathrm{e}^{-\gamma})^t .
$$
(5.3)

Recalling that $\tilde{T}_u$ are independent PGW(μ)-trees, it follows that

$$
Z'(N) = \sum_{\{r_u\}: \sum_{u \in H} r_u = N} \left[\prod_{u \in H} \mathbb{P}(|\tilde{T}_u| = r_u)\right] = \mathbb{P}\left(|\widehat{\mathcal{C}}_1| = N \mid H\right) .
$$

Combining this with (5.2) and (5.3), we obtain that

$$
\frac{\mathbb{P}(|T_u| = t_u \text{ for all } u \in H \mid H)}{\mathbb{P}(|\tilde{T}_u| = t_u \text{ for all } u \in H \mid H)} = \frac{\mathbb{P}(|\mathcal{C}_1| = N \mid H)}{\mathbb{P}(|\widehat{\mathcal{C}}_1| = N \mid H)} .
$$
(5.4)

At this point, we wish to estimate the ratio in the right hand side above. To this end, we need the following result of [31], which we restate in our setting of the near-critical regime.

Theorem 5.1 ([31, Theorem 6], reformulated). *Let* $\mathbf{b}(\lambda) = \begin{pmatrix} b_1(\lambda) \\ b_2(\lambda) \\ b_3(\lambda) \end{pmatrix}$ *where*

$b_1(\lambda) = (1 - \mu)\left(1 - \frac{\mu}{\lambda}\right)$, $b_2(\lambda) = \mu\left(1 - \frac{\mu}{\lambda}\right)$, $b_3(\lambda) = \frac{1}{2}\left(1 - \frac{\mu}{\lambda}\right)(\lambda + \mu - 2)$.

There exist positive definite matrices K_p, K_m *satisfying*

$$
K_p = \begin{pmatrix}
(12 + o(1))\varepsilon & 4 + o(1) & (6 + o(1))\varepsilon^2 \\
4 + o(1) & (2 + o(1))/\varepsilon & (2 + o(1))\varepsilon \\
(6 + o(1))\varepsilon^2 & (2 + o(1))\varepsilon & (\frac{10}{3} + o(1))\varepsilon^3
\end{pmatrix} ,
$$

$$
K_m = K_p - 2\lambda \frac{d\mathbf{b}(\lambda)}{d\lambda} \cdot \frac{d\mathbf{b}(\lambda)^T}{d\lambda} , \quad \det(K_m) = (\tfrac{8}{3} + o(1))\varepsilon^3 ,
$$

and such that

(i) $(|H|, |\mathcal{C}_1| - |H|, |E(H)| - |H|)$ *is in the limit Gaussian with a mean vector $n\mathbf{b}$ and a covariance matrix nK_p.*

(ii) *If $A_m \triangleq K_m^{-1}$ and B denotes the event that $|E(G)| = m$ for some $m = (1 + (1 + o(1))\varepsilon)\frac{n}{2}$, and there is a unique component of size between εn and $4\varepsilon n$ and none larger, then*

$$\mathbb{P}\left(|H| = n_1, |\mathcal{C}_1| - |H| = n_2, |E(H)| - |H| = n_3 \,\big|\, B\right)$$
$$= \frac{\sqrt{3} + o(1)}{8(\pi n \varepsilon)^{3/2}} \exp\left(-\tfrac{1}{2}\mathbf{x}^T A_m \mathbf{x}\right) , \tag{5.5}$$

uniformly for all $(n_1, n_2, n_3) \in \mathbb{N}^3$ such that

$$(K_p(1,1)^{-1/2} x_1, K_p(2,2)^{-1/2} x_2, K_p(3,3)^{-1/2} x_3)$$

is bounded, where $\mathbf{x}^T = (x_1, x_2, x_3)$ is defined by

$$\mathbf{x}^T = \frac{1}{\sqrt{n}}(n_1 - b_1 n, n_2 - b_2 n, n_3 - b_3 n) .$$

Since $\varepsilon^2 n \to \infty$, it is clear by CLT that w.h.p. the total number of edges in $G \sim \mathcal{G}(n, p)$ is $(1 + (1 + o(1))\varepsilon)\frac{n}{2}$. Furthermore, by the results of [10] and [24] (see also [19]), w.h.p. our graph G has a unique giant component of size

$$(1 + o(1))(1 - \mu/\lambda)n = (2 + o(1))\varepsilon n .$$

Altogether, we deduce that the event B happens w.h.p.; assume therefore that B indeed occurs. Define the event Q by

$$\mathcal{Q}_M \triangleq \left\{(n_1, n_2, n_3) \in \mathbb{N}^3 : |x_1| \leq \sqrt{\varepsilon}M \ , \ |x_2| \leq \frac{M}{\sqrt{\varepsilon}} \ , \ |x_3| \leq \varepsilon^{3/2}M\right\} ,$$
$$Q \triangleq \left\{(|H|, |\mathcal{C}_1| - |H|, |E(H)| - |H|) \in \mathcal{Q}_M\right\} .$$

By part (i) of Theorem 5.1, for any fixed $\delta > 0$ there exists some $M > 0$ such that $\mathbb{P}(Q^c) < \delta$ for a sufficiently large n. Next, define

$$P_{\max} = \max_{(n_1, n_2, n_3) \in \mathcal{Q}_M} \mathbb{P}\left(|H| = n_1, \ |\mathcal{C}_1| - |H| = n_2, \ |E(H)| - |H| = n_3\right) ,$$
$$P_{\min} = \min_{(n_1, n_2, n_3) \in \mathcal{Q}_M} \mathbb{P}\left(|H| = n_1, \ |\mathcal{C}_1| - |H| = n_2, \ |E(H)| - |H| = n_3\right) .$$

It follows from part (ii) of Theorem 5.1 that there exists some $c = c(M) > 0$ such that

$$P_{\max} \leq c \cdot P_{\min} , \tag{5.6}$$

when n is sufficiently large. Notice that by definition of $\mathbf{x}$,

$$\#\{n_2 \in \mathbb{N} : |x_2| \leq M/\sqrt{\varepsilon}\} \geq M\sqrt{n/\varepsilon} .$$

Combined with (5.6), it follows that for any $(n_1, n_2, n_3) \in \mathcal{Q}_M$ we have

$$\mathbb{P}\left(|\mathcal{C}_1| = n_1 + n_2 \ , \ Q \,\big|\, |H| = n_1\right) \leq \frac{c}{M\sqrt{n/\varepsilon}} \ . \tag{5.7}$$

With this estimate for $\mathbb{P}(|\mathcal{C}_1| = N \,\big|\, H)$, the numerator in the right-hand-side of (5.4), it remains to estimate the denominator, $\mathbb{P}(|\widehat{\mathcal{C}}_1| = N \,\big|\, H)$.

Recall that, given H, the quantity $|\widehat{\mathcal{C}}_1|$ is a sum of $|H|$ i.i.d. Borel(μ) random variables (each such variable is the size of a PGW(μ)-tree). We would now like to derive a local central limit theorem for $|\widehat{\mathcal{C}}_1|$. Unfortunately, each Borel(μ) variable $|T_u|$ has $\operatorname{Var}|T_u| \asymp 1/\varepsilon^3 \to \infty$, and standard versions of local CLT do not apply here. To bypass this obstacle, we use a different characterization of the tree-sizes $\{|T_u| : u \in H\}$.

It is well known that the total progeny in a branching process with off-spring distribution Z has the same law as the hitting time from 1 to 0 of a one-dimensional random walk whose increments are i.i.d. variables distributed as $Z - 1$ (see, e.g., [33, page 234]). Hence, the total size of k i.i.d. such branching processes is exactly the hitting time of this walk from k to 0. The following theorem of Otter [29] characterizes this quantity:

Theorem 5.2 ([29], see also [20]). *Let W_t be a random walk, whose steps Y_i are i.i.d. random variables satisfying $Y_i \geq -1$. Then*

$$\mathbb{P}_k(\tau_0 = t) = \frac{k}{t}\mathbb{P}_k(W_t = 0) \ .$$

In our setting, $Y_i \sim \operatorname{Po}(\mu) - 1$, hence $\mathbb{P}_k(W_t = 0) = \mathbb{P}(S_t = t - k)$ where $S_t = \sum_{i=1}^{t} X_i$ and the X_i's are i.i.d. $\operatorname{Po}(\mu)$ variables. In light of the above theorem, it follows that for any integer n_2,

$$\mathbb{P}(|\widehat{\mathcal{C}}_1| = n_1 + n_2 \,\big|\, |H| = n_1) = \frac{n_1}{n_1 + n_2}\mathbb{P}(S_{n_1+n_2} = n_2) \ . \tag{5.8}$$

Since $S_m \sim \operatorname{Po}(m\mu)$ for any m, we have that

$$\mathbb{P}(S_{n_1+n_2} = n_2) = \mathrm{e}^{-(n_1+n_2)\mu}\frac{((n_1 + n_2)\mu)^{n_2}}{n_2!} \ . \tag{5.9}$$

Recalling the definition of $\mathcal{Q}_m$, we are interested in the following range for n_1 and n_2:

$$n_1 = (1 - \mu)(1 - \tfrac{\mu}{\lambda})n + c_1\sqrt{\varepsilon n} \qquad\qquad (|c_1| \leq M) \ ,$$
$$n_2 = \mu(1 - \tfrac{\mu}{\lambda})n + c_2\sqrt{n/\varepsilon} \qquad\qquad (|c_2| \leq M) \ .$$

In this case, we have

$$\frac{n_1 + n_2}{n_2} = \frac{(1 - \frac{\mu}{\lambda})n + c_1\sqrt{\varepsilon n} + c_2\sqrt{n/\varepsilon}}{\mu(1 - \frac{\mu}{\lambda})n + c_2\sqrt{n/\varepsilon}} = \frac{1}{\mu} + \frac{c_1\sqrt{\varepsilon n} + c_2(1 - \frac{1}{\mu})\sqrt{n/\varepsilon}}{\mu(1 - \frac{\mu}{\lambda})n + c_2\sqrt{n/\varepsilon}}$$

$$\triangleq \frac{1}{\mu} + \xi \ ,$$

where $\xi = \xi(n) = (\frac{1}{2} + o(1))(c_1 - c_2)/\sqrt{\varepsilon n}$. Applying Stirling's formula to (5.9) and using the fact that $1 + x \geq \exp(x - x^2)$ for $x \geq 0$ gives

$$\mathbb{P}(S_{n_1+n_2} = n_2) = \exp\left[\left(1 - \frac{n_1 + n_2}{n_2}\mu\right)n_2\right]\frac{1}{\sqrt{2\pi n_2}}\left(\frac{(n_1 + n_2)\mu}{n_2}\right)^{n_2}$$

$$= \exp(-\xi\mu n_2)\frac{1}{\sqrt{2\pi n_2}}(1 + \xi\mu)^{n_2} \geq \frac{1}{\sqrt{2\pi n_2}}e^{-\xi^2\mu^2 n_2} \ .$$

Now, since $n_2 = (2 + o(1))\varepsilon n$ and

$$\xi^2\mu^2 n_2 = (1 + o(1))\frac{(c_1 - c_2)^2}{4\varepsilon n}2\varepsilon n \leq (2 + o(1))M^2 \ ,$$

we conclude that for some constant $\delta' = \delta'(M)$,

$$\mathbb{P}(S_{n_1+n_2} = n_2) \geq \frac{\delta'}{\sqrt{n\varepsilon}} \ .$$

Recalling that $\frac{n_1}{n_1+n_2} = (1 + o(1))\varepsilon$, we can decrease δ' to absorb this $o(1)$ error-term for a sufficiently large n, and together with (5.8) get

$$\mathbb{P}(|\widehat{\mathcal{C}}_1| = n_1 + n_2 \,|\, |H| = n_1) \geq \delta'\frac{\varepsilon}{\sqrt{n\varepsilon}} = \frac{\delta'}{\sqrt{n/\varepsilon}} \ . \qquad (5.10)$$

Combining (5.7) and (5.10), we obtain that when n is sufficiently large,

$$\frac{\mathbb{P}\left(|\mathcal{C}_1| = N \ , \ Q \,\big|\, |H|\right)}{\mathbb{P}\left(|\widehat{\mathcal{C}}_1| = N \,\big|\, |H|\right)} \leq \frac{c}{M\delta'} \ .$$

By (5.4) (and recalling the fact that conditioned on $|T_i|$, the tree T_i is uniformly distributed among all unlabeled trees of this size, and a similar statement holds for $\tilde{T}_i$), we conclude that for some $c' = c'(M) > 0$ and any unlabeled graph A

$$\mathbb{P}(\mathcal{C}_1 = A, Q, B \mid H) \leq c'\,\mathbb{P}(\widehat{\mathcal{C}}_1 = A \mid H) \ . \qquad (5.11)$$

We are now ready to conclude the proof of the main theorem. Let $\tilde{\mathcal{C}}_1$ be defined as in Theorem 2. For any set of simple graphs $\mathcal{A}$, define

$$\mathcal{H} = \left\{H : \mathbb{P}(\mathcal{C}_1 \in \mathcal{A}, Q, B \mid \mathcal{C}_1^{(2)} = H) \geq (\mathbb{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A}))^{1/2}\right\} \ . \qquad (5.12)$$

Recall that by definition, $\tilde{\mathcal{C}}_1$ is produced by first constructing its 2-core (first two steps of the description), then attaching to each of its vertices independent PGW(μ)-trees. Hence, for any H, the graphs $\widehat{\mathcal{C}}_1$ and $\tilde{\mathcal{C}}_1$ have

the same conditional distribution given $\widehat{\mathcal{C}}_1^{(2)} = \tilde{\mathcal{C}}_1^{(2)} = H$. It then follows from (5.11),(5.12) that for some constant $c'' > 0$ and any $H \in \mathcal{H}$,

$$\mathbb{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A} \mid \tilde{\mathcal{C}}_1^{(2)} = H) \geq c''(\mathbb{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A}))^{1/2}.$$

Since

$$\mathbb{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A}) \geq c''(\mathbb{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A}))^{1/2}\mathbb{P}(\tilde{\mathcal{C}}_1^{(2)} \in \mathcal{H}),$$

the assumption that $\mathbb{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A}) \to 0$ now gives that $\mathbb{P}(\tilde{\mathcal{C}}_1^{(2)} \in \mathcal{H}) \to 0$.

At this point, we combine all the contiguity results thus far to claim that, for any family of simple graphs $\mathcal{F}$,

$$\mathbb{P}(\tilde{\mathcal{C}}_1^{(2)} \in \mathcal{F}) = o(1) \text{ implies that } \mathbb{P}(\mathcal{C}_1^{(2)} \in \mathcal{F}) = o(1).$$

Indeed, by definition, the 2-core of $\tilde{\mathcal{C}}_1$ is precisely the Poisson-geometric model, conditioned on the sum of the degrees $(\sum_u D_u \mathbf{1}_{D_u \geq 3})$ being even. Therefore, as $\mathcal{F}$ consists only of simple graphs, clearly we may consider this model condition on the graph produced being simple, and in particular, that $\sum_u D_u \mathbf{1}_{D_u \geq 3}$ is even. Applying Theorem 4.2 (contiguity with Poisson-configuration), Theorem 3.2 (contiguity with Poisson-cloning) and Theorem 2.1 (contiguity with Erdős-Rényi graphs), in that order, now gives the above statement.

This fact and the arguments above now give that $\mathbb{P}(\mathcal{C}_1^{(2)} \in \mathcal{H}) \to 0$. By the definition of $\mathcal{H}$, we now conclude that

$$\mathbb{P}(\mathcal{C}_1 \in \mathcal{A}) \leq \mathbb{P}(B^c) + \mathbb{P}(Q^c) + \mathbb{P}(\mathcal{C}_1^{(2)} \in \mathcal{H}) + (\mathbb{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A}))^{1/2}.$$

Taking a limit, we get that $\limsup_{n\to\infty} \mathbb{P}(\mathcal{C}_1 \in \mathcal{A}) \leq \delta$ and the proof is completed by letting $\delta \to 0$. ∎

6. The early giant component

In this section, we consider the special case of Theorem 2 for $\varepsilon = o(n^{-1/4})$, and namely prove Theorem 1. We show how each of the three steps described in Theorem 2 reduces to the corresponding steps in Theorem 1.

6.1. **Step 1: The kernel.** Let $\Lambda \sim \mathcal{N} \sim (1+\varepsilon-\mu, \frac{1}{\varepsilon n})$. By (5.1), we have $\mu = 1 - \varepsilon + O(\varepsilon^2)$, and so

$$\mathbb{E}\Lambda = 1 + \varepsilon - \mu = 2\varepsilon + O(\varepsilon^2), \quad \mathrm{Var}(\Lambda) = \frac{1}{\sqrt{\varepsilon n}},$$

giving that $\Lambda = (2+o(1))\varepsilon$ w.h.p. In particular, the probability that $D_u \geq 4$ for some vertex u is

$$\mathbb{P}(\mathrm{Po}(\Lambda) \geq 4) = O(\varepsilon^4) = o(n^{-1}),$$

and a union bound thus implies that the kernel is 3-regular w.h.p. In other words, we have $N_k = 0$ for all $k \geq 4$, and

$$N = N_3 \sim \mathrm{Bin}(n, \mathrm{e}^{-\Lambda}\Lambda^3/6) \text{ conditioned to be even.}$$

It remains to compare the distributions of N and $2\lfloor Z \rfloor$, where

$$Z \sim \mathcal{N}(\tfrac{2}{3}\varepsilon^3 n, \varepsilon^3 n)\,.$$

The first step in this direction is to approximate the binomial variable by a Poisson variable. A well-known and straightforward application of the Stein-Chen method (see, e.g., [5]) is that for any n and q,

$$\|\mathrm{Bin}(n, q) - \mathrm{Po}(nq)\|_{\mathrm{TV}} \leq q \,\wedge\, nq^2\,,$$

where the total-variation distance $\|\cdot\|_{\mathrm{TV}}$ between two distributions σ, π on a finite space Ω is given by

$$\|\sigma - \pi\|_{\mathrm{TV}} \triangleq \sup_{A \subset \Omega} |\sigma(A) - \pi(A)| = \frac{1}{2} \sum_{x \in \Omega} |\sigma(x) - \pi(x)|\,. \qquad (6.1)$$

Therefore, given that $\Lambda = (2 + o(1))\varepsilon$ (again, this holds w.h.p.) we have

$$\|\mathrm{Bin}(n, \mathrm{e}^{-\Lambda}\Lambda^3/6) - \mathrm{Po}(n\mathrm{e}^{-\Lambda}\Lambda^3/6)\|_{\mathrm{TV}} \leq O(n\varepsilon^6) = o(n^{-1/2})\,.$$

Clearly, by definition (6.1), a negligible total-variation distance between two distributions already implies they are contiguous (in both directions), hence it suffices to compare $2\lfloor Z \rfloor$ to the variable Y, distributed as $\mathrm{Po}(n\mathrm{e}^{-\Lambda}\Lambda^3/6)$ conditioned to be even. We will show that for some region $\mathcal{Q}$ such that $\mathbb{P}(Y \in \mathcal{Q}) \to 1$ and some $c = c(\mathcal{Q}) > 0$,

$$\mathbb{P}(Y = t) \leq c \cdot \mathbb{P}(2\lfloor Z \rfloor = t) \quad \text{ for all even } t \in \mathcal{Q}\,. \qquad (6.2)$$

Let $\delta > 0$. By the above properties of Λ, there exists some $M > 0$ such that

$$\mathbb{P}\left(|\Lambda - 2\varepsilon| > M/\sqrt{\varepsilon n}\right) < \delta\,.$$

The following result is a special case of a theorem of [14].

Theorem 6.1 ([14, Ch. 2, Theorem 5.2], reformulated). *Let X be a random variable on $\mathbb{N}$ with $\mathbb{P}(X = k) > 0$ for all $k \in \mathbb{N}$. Suppose that $\mathbb{E}X = \nu < \infty$ and $\mathrm{Var}\, X = \sigma^2 < \infty$. Let X_i be i.i.d. distributed as X and $S_m = \sum_{i=1}^{m} X_i$. Then as $m \to \infty$, we have*

$$\sup_{x \in \mathcal{L}_m} \left| \sqrt{m}\, \mathbb{P}\left(\frac{S_m - m\nu}{\sqrt{m}} = x\right) - \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-x^2/\sigma^2} \right| \to 0\,,$$

where $\mathcal{L}_m = \{(z - m\nu)/\sqrt{m} : z \in \mathbb{Z}\}$.

In our setting, given that $|\Lambda - 2\varepsilon| \leq M/\sqrt{\varepsilon n}$ we have a Poisson distribution with parameter

$$m = \tfrac{4}{3}\varepsilon^3 n + O(\sqrt{\varepsilon^3 n}),$$

and clearly the effect of conditioning that it is even, as well as rounding m to its integer part, only affect the density function by a constant factor. This translates into a sum of $\lfloor m \rfloor$ i.i.d. Po(1) variables, and by the above theorem we conclude that, as long as $|\Lambda - 2\varepsilon| < M/\sqrt{\varepsilon n}$, there exists some $c = c(M) > 0$ such that

$$\mathbb{P}(Y = t) \leq \frac{c}{\sqrt{\varepsilon^3 n}} \qquad \text{for any integer } t.$$

Furthermore, we can choose $\mathcal{Q} = [m - M'\sqrt{\varepsilon^3 n}, m + M'\sqrt{\varepsilon^3 n}]$ for a suitably large $M' > 0$ so that

$$\mathbb{P}(Y \notin \mathcal{Q}) < \delta.$$

By the definition of Z (note that $2\lfloor Z \rfloor$ has mean $\tfrac{4}{3}\varepsilon^3 n + O(1)$ and variance of order $\varepsilon^3 n$), there exists some $c' = c'(M')$ such that

$$\mathbb{P}(2\lfloor Z \rfloor = t) \geq \frac{c'}{\sqrt{\varepsilon^3 n}} \qquad \text{for all even } t \in \mathcal{Q}.$$

Altogether, it follows that for any sequence of subsets of integers $\mathcal{S} = \mathcal{S}(n)$, if $\mathbb{P}(2\lfloor Z \rfloor \in \mathcal{S}) = o(1)$ then $\mathbb{P}(N \in \mathcal{S}) \leq 2\delta + o(1)$. We now let $\delta \to 0$ to complete the contiguity of N and $2\lfloor Z \rfloor$.

6.2. **Step 2: The 2-core.** Here we need to compare the effect of replacing the 2-paths by i.i.d. Geom(ε) variables rather than Geom$(1 - \mu)$. Rather than just showing contiguity between the two models, we will show a stronger statement, namely that the total-variation between the joint distributions of the path lengths are negligible. The total-variation distance $\|\cdot\|_{\mathrm{TV}}$ between two distributions σ, π on a finite space Ω is given by

$$\|\sigma - \pi\|_{\mathrm{TV}} \triangleq \sup_{A \subset \Omega} |\sigma(A) - \pi(A)| = \frac{1}{2} \sum_{x \in \Omega} |\sigma(x) - \pi(x)|.$$

In our case, since the ratio of $\frac{p(1-p)^k}{q(1-q)^k}$ for $0 < p < q < 1$ is monotone increasing in k, we can clearly consider sets of the $A = \{k, k+1, \ldots\}$ in the supremum above. Hence,

$$\| \operatorname{Geom}(1 - \mu) - \operatorname{Geom}(\varepsilon) \|_{\mathrm{TV}} = \sup_k |\mu^k - (1 - \varepsilon)^k|$$

$$\leq \sup_k |\mu - (1 - \varepsilon)| \cdot k[\mu \vee (1 - \varepsilon)]^{k-1} = O(\varepsilon),$$

where, as $|\mu - (1 - \varepsilon)| = \Theta(\varepsilon^2)$, the value of k optimizing the above has order $1/\varepsilon$.

Recalling that $|E(\mathcal{K})| = O(\varepsilon^3 n)$ w.h.p., we infer that the total-variation distance between the joint distribution of the 2-path lengths in the two models (i.i.d. $\mathrm{Geom}(1 - \mu)$ variables and i.i.d. $\mathrm{Geom}(\varepsilon)$ variables) is $O(\varepsilon^4 n)$. Our assumption that $\varepsilon = o(n^{-1/4})$ now gives that this is $o(1)$.

6.3. **Step 3: The attached trees.** We now wish to compare the distributions of i.i.d. $\mathrm{PGW}(\mu)$-trees to i.i.d. $\mathrm{PGW}(1 - \varepsilon)$-trees. Recall that the size of a $\mathrm{PGW}(\gamma)$-tree follows a $\mathrm{Borel}(\gamma)$ distribution, as given in (5.3). Thus,

$$\|\mathrm{Borel}(\mu) - \mathrm{Borel}(1 - \varepsilon)\|_{\mathrm{TV}} = \frac{1}{2} \sum_t \frac{t^{t-1}}{t!} |\mu^{t-1} e^{-\mu t} - (1 - \varepsilon)^{t-1} e^{-(1-\varepsilon)t}|$$

$$\leq O(\varepsilon^2) \sum_t \frac{t^{t-1}}{t!} |x_t^{t-2} e^{-x_t t} (t - 1 - t x_t)|,$$

Noticing that $x_t = 1 - \varepsilon + O(\varepsilon^2)$ and applying Stirling's inequality, we get

$$\frac{t^{t-1}}{t!} |x_t^{t-2} e^{-x_t t}(t - 1 - t x_t)| = \Theta(1) t^{-3/2} (x_t e^{1-x_t})^t |-1 + t\varepsilon + O(t\varepsilon^2)|$$

$$\leq O(t^{-3/2}) e^{-(1-x_t)^2/2}(1 + t\varepsilon) \leq O(t^{-3/2}) e^{-\varepsilon^2 t/3}(1 + t\varepsilon),$$

where the first inequality is by the fact that $1 - y \leq e^{-y - y^2/2}$ for all $y > 0$, and the second one holds for any large n by the definition of x_t. Therefore,

$$\|\mathrm{Borel}(\mu) - \mathrm{Borel}(1 - \varepsilon)\|_{\mathrm{TV}} = O(\varepsilon^2) + O\left(\varepsilon^2 \sum_t \frac{\varepsilon e^{-\varepsilon^2 t/3}}{\sqrt{t}}\right)$$

$$\leq O(\varepsilon^2) + O(\varepsilon^2) \int_0^\infty \frac{1}{\sqrt{\varepsilon^2 x/3}} e^{-\varepsilon^2 x/3} d(\varepsilon^2 x/3) = O(\varepsilon^2),$$

where we used the fact that $\int_0^\infty \frac{1}{\sqrt{y}} e^{-y}$ converges (to $\sqrt{\pi}$).

With high probability, the size of the 2-core (that is, the number of PGW-trees we attach) is $O(\varepsilon^2 n)$, and so the total-variation distance between the two joint distributions is at most $O(\varepsilon^4 n) = o(1)$, as required.

7. Analysis of the Cut-Off Line Algorithm

In this section, we analyze Algorithm 1 for generating the Poisson cloning model, and establish a tight concentration result for Λ_C (the location of the cut-off line when all light clones are exhausted), as stated in Theorem 3.3.

Proof of Theorem 3.3. We wish to prove inequality (3.4), i.e., that for some fixed $c > 0$, the probability that $|\Lambda_C - \theta_\lambda \lambda| \geq \frac{\gamma}{\sqrt{\theta_\lambda n}}$ is at most $\exp(-c\gamma^2)$.

Notice that, prior to the first time the algorithm reaches Step 2, the notion of active/passive vertices does not play a role in its decisions. Since this is the only change between subsequent phases, it follows that Λ_C is precisely

the same regardless of the choice of phase boundaries. In particular, we may choose β as follows: Take $\frac{1-\theta_\lambda}{3} \leq \beta \leq \frac{1-\theta_\lambda}{2}$ and an integer m such that

$$(1-\beta)^{m-1} = \theta_\lambda + \frac{\gamma}{\sqrt{\theta_\lambda n}} \,, \tag{7.1}$$

where γ as given in the lemma, i.e., $\gamma = o\big(\sqrt{\theta_\lambda^3 n}\big)$.

In order to prove the lemma, we first estimate the number of j-active clones for each j, denoted by N_j. Let M_j be the number of j-active clones that are *matched* during phase j. We need the following lemma to estimate M_j given N_j.

Lemma 7.1 ([22, Lemma 2.2]). *Consider a Poisson μ-cell for $\mu > 0$, and let N be its total number of clones. For $0 < \theta \leq 1$, let $N(\theta)$ be the number of matched clones once the cut-off line reaches $\theta\mu$. Then there exists some $c > 0$ so that the following holds: For any $0 < \theta_0 < 1$ and $l, \Delta > 0$,*

$$\mathbb{P}\left(\max_{\theta_0 \leq \theta \leq 1} |N(\theta) - (1-\theta^2)k| \geq \Delta \,\Big|\, N = k\right) \leq 2\exp\big[-c\big(\Delta \wedge \tfrac{\Delta^2}{(1-\theta_0)k}\big)\big] \,.$$

By definition of the Cut-Off Line Algorithm, if either one of the two unmatched clones of a passive vertex was matched in a given step, then the other clone is guaranteed to be matched in the next step (either in this phase or in a later one), as it is inserted to the top of the stack. This means that, for the purpose of determining the number of matched active clones throughout the phase, M_j, applying the algorithm with or without passive vertices is effectively the same (since one can always identify the two clones of each passive vertex, then contract the 2-paths into edges between active clones).

That said, one must consider the following delicate point. If the end-of-phase boundary is reached while the top of the stack contains a passive clone, this corresponds to a path whose one endpoint is an active clone (u, i), yet its other endpoint is a passive clone. In this case, the active clone should not be considered as matched when disregarding all passive clones. Let A_j denote this event for phase j, and define

$$M_j' \triangleq M_j - \mathbf{1}_{A_j}$$

to be the number of j-active clones that are matched during phase j while disregarding (contracting) the passive clones.

Combining the above observation with the fact that phase j began with N_j active clones and a cut-off line at $\mu = (1-\beta)^{j-1}\lambda$ and ended as soon as the cut-off line reached $(1-\beta)\mu$, we apply Lemma 7.1 for $\theta = \theta_0 = 1-\beta$ and conclude that for some constant $c > 0$,

$$\mathbb{P}\big(|M_j' - (1-(1-\beta)^2)N_j| \geq \Delta \,\big|\, N_j\big) \leq 2\exp\big[-c\big(\Delta \wedge \tfrac{\Delta^2}{\beta N_j}\big)\big] \,. \tag{7.2}$$

Let B_j denote the number of vertices which have precisely 2 clones to the left of the end-of-phase boundary of phase j, and at least 1 more clone in the interval of phase j. Note that, for such a vertex v, it is clearly j-active, and it would become $(j+1)$-passive if and only if the formerly mentioned 2 clones are unmatched by the end of phase j. In this case, two formerly active clones will be relabeled as passive. In particular, the number of clones that transition from being j-active to being $(j+1)$-passive is at most $2B_j$.

On the other hand, a clone can transition from being j-passive to being $(j+1)$-active if and only if it happened to be at the top of the stack when phase j ended, and in particular, the event A_j occurred.

Adding these two, along with the number of j-active clones matched in this phase M'_j, we conclude that the number of $(j+1)$-active clones satisfies

$$N_{j+1} \geq N_j - M_j + \mathbf{1}_{A_j} - 2B_j = N_j - M'_j - 2B_j \ . \qquad (7.3)$$

Note that, as long as Step 2 of the algorithm is not reached, the stack always consists of light clones exclusively. Therefore, up till that point, if a vertex has 2 unmatched clones, both will remain unmatched until the cut-off line reaches one of them. Suppose that phase j_0 is the first one where the algorithm invoked Step 2. In that case, for any $j < j_0$, the vertices counted in B_j are precisely those that were j-active yet became $(j+1)$-passive. We deduce that (7.3) is in fact an equality for all $j < j_0$.

In order to analyze B_j, for each $v \in V$ and $0 \leq \theta < \theta' \leq 1$ let $d_v(\theta, \theta')$ denote the number of v-clones whose assigned value belongs to $[\theta\lambda, \theta'\lambda)$. Further let $d_v(\theta) \triangleq d_v(0, \theta)$. Recall that phase j begins with the line at $(1-\beta)^{j-1}\lambda$ and ends with the line at $(1-\beta)^j\lambda$. Hence, for $\theta_j = (1-\beta)^{j-1}$, we have by the definition of B_j that

$$B_j = \sum_{v \in V} \mathbf{1}_{\{d_v(\theta_{j+1})=2\}} \ \mathbf{1}_{\{d_v(\theta_{j+1}, \theta_j) \geq 1\}} \ .$$

Observe that $(d_v(\theta_{j+1}), d_v(\theta_{j+1}, \theta_j))$ for $v \in V$ are i.i.d. pairs of independent Poisson random variables with means $\theta_{j+1}\lambda$ and $(\theta_j - \theta_{j+1})\lambda$ respectively. Applying Chernoff's bound (cf., e.g., [4]) we have that for some $c_1 > 0$,

$$\mathbb{P}\left(\left| B_j - \frac{(\theta_{j+1}\lambda)^2}{2} \mathrm{e}^{-\theta_{j+1}\lambda}(1 - \mathrm{e}^{-\beta\theta_j\lambda})n \right| \geq \Delta \right) \leq 2 \exp\left(-c_1 \frac{\Delta^2}{\theta_j^3 n} \right) \ . \quad (7.4)$$

Combined with (7.2) and (7.3), we arrive at the following estimated lower bound for N_{j+1}:

$$(1-\beta)^2 N_j - (\theta_{j+1}\lambda)^2 \mathrm{e}^{-\theta_{j+1}\lambda}(1 - \mathrm{e}^{-\beta\theta_j\lambda})n \ .$$

Applying this inductively, we expect that the following would be a lower bound for N_j:

$$\theta_j^2 \lambda (1 - \lambda \mathrm{e}^{-\theta_j\lambda})n \qquad (7.5)$$

(indeed, this is later shown in Lemma 7.3).

Suppose that the algorithm is at the beginning of phase j, and Step 2 has not been reached yet (in any of the phases thus far). By the discussion above, any j-active vertex has either 1 or strictly more than 2 clones with values in $(0, \theta_j \lambda)$. In particular, the number of j-active clones that are *heavy* at the beginning of phase j can then be written as

$$H_j = \sum_{v \in V} d_v(\theta_j) \mathbf{1}_{\{d_v(\theta_j) > 2\}} ,$$

and the number of light clones at the start of phase j is then precisely

$$L_j \triangleq N_j - H_j .$$

In general (once Step 2 is invoked), H_j is an upper bound for the number of j-active clones that are heavy at this point. (Note that the only reason for this bound not to be tight is on account of clones that are already matched. That is, $\sum_{v \in V} \mathbf{1}_{\{d_v(\theta_j) > 2\}}$ counts all j-active heavy *vertices*, in addition to perhaps some whose clones are all matched by phase j.) Hence, L_j is always a lower bound for the number of light clones at the start of phase j.

We need the following large deviation inequality:

Lemma 7.2 ([23, Corollary 4.2]). *Let $X_1, \ldots, X_m$ be a sequence of independent random variables. Suppose $\mathbb{E}[X_i] = \mu_i$ and there are b_i, d_i and ξ_0 such that $\mathbb{E}[(X_i - \mu_i)^2] \leq b_i$, and*

$$\left| \mathbb{E}\left[(X_i - \mu_i)^3 e^{\xi(X_i - \mu_i)} \right] \right| \leq d_i \qquad \text{for all } 0 \leq |\xi| \leq \xi_0 .$$

If $\delta \xi_0 \sum_{i=1}^m d_i \leq \sum_{i=1}^m b_i$ for some $0 < \delta \leq 1$, then

$$\mathbb{P}\left(\left| \sum_{i=1}^m X_i - \sum_{i=1}^m \mu_i \right| \geq \Delta \right) \leq e^{-\frac{1}{3} \min\{\delta \xi_0 \Delta, \frac{\Delta^2}{\sum_{i=1}^m b_i}\}} ,$$

for all $\Delta > 0$.

Since $d_v(\theta_j)$ are i.i.d. $\mathrm{Po}(\theta_j \lambda)$ variables, an application of the above lemma gives that for some constant $c_2 > 0$,

$$\mathbb{P}\left(|H_j - \theta_j \lambda (1 - e^{-\theta_j \lambda} - \theta_j \lambda e^{-\theta_j \lambda}) n| \geq \Delta \right) \leq 2 \exp\left[-c_2 \left(\Delta \wedge \tfrac{\Delta^2}{\theta_j^3 n} \right) \right] . \quad (7.6)$$

Recalling that m is an integer with $(1 - \beta)^{m-1} = \theta_\lambda + \frac{\gamma}{\sqrt{\theta_\lambda n}}$ (see (7.1)), set

$$\Delta_j \triangleq \frac{\gamma}{100} \sqrt{\theta_j^3 n} \sum_{i=1}^j (1 - \beta)^{(2j - i - m)/4} . \quad (7.7)$$

Observe that the following sequence is increasing in j:

$$(1 - \beta)^{\frac{j}{4}} \theta_j^{-3/2} = (1 - \beta)^{3/2} (1 - \beta)^{-5j/4} .$$

We then get that for all $j \in [m]$,

$$\gamma(1-\beta)^{(j-m)/4}(\theta_j^3 n)^{-1/2} \leq \frac{\gamma}{\sqrt{\theta_m^3 n}} = o(1) , \tag{7.8}$$

where the last equality used the facts $\gamma = o\big(\sqrt{\theta_\lambda^3 n}\big)$ and

$$\theta_m = (1-\beta)^{m-1} = \theta_\lambda + \frac{\gamma}{\sqrt{\theta_\lambda n}} = (1+o(1))\theta_\lambda . \tag{7.9}$$

It then follows from (7.8) that for any $j \in [m]$,

$$
\begin{aligned}
\Delta_j &= \frac{\gamma}{100}\sqrt{\theta_j^3 n}\sum_{i=1}^{j}(1-\beta)^{(2j-i-m)/4} \\
&= \frac{\gamma}{100}\sqrt{\theta_j^3 n}(1-\beta)^{(j-m)/4}\sum_{i=1}^{j}(1-\beta)^{(j-i)/4} = o(\theta_j^3 n) .
\end{aligned} \tag{7.10}
$$

Recalling (7.5) and (7.6), we will next establish lower bounds for the N_j's and L_j's in terms of the following parameters:

$$
\begin{aligned}
n_j &\triangleq \theta_j^2\lambda(1-\lambda e^{-\theta_j\lambda})n \\
l_j &\triangleq \theta_j\lambda(\theta_j - 1 + e^{-\theta_j\lambda})n
\end{aligned}
\qquad \text{for } j \in [m]. \tag{7.11}
$$

Lemma 7.3. *There exists a constant $c > 0$ such that the following holds:*

$$\mathbb{P}\left(\exists j \in [m] : N_j < n_j - \Delta_j\right) \leq e^{-c\gamma^2} ,$$

$$\mathbb{P}\left(\exists j \in [m] : L_j < l_j - 2\Delta_j\right) \leq e^{-c\gamma^2} .$$

Proof. With (7.4) in mind, define the following for each $j \in [m]$:

$$
\begin{aligned}
b_j &\triangleq \frac{(\theta_{j+1}\lambda)^2}{2}e^{-\theta_{j+1}\lambda}(1-e^{-\beta\theta_j\lambda})n , \\
\gamma_j &\triangleq (1-\beta)^{(j-m)/4}\gamma .
\end{aligned} \tag{7.12}
$$

It is clear (see definition (7.11)) that

$$n_{j+1} = (1-\beta)^2 n_j - 2b_j , \tag{7.13}$$

and furthermore, by (7.8), we have that

$$\gamma_j = o\big(\sqrt{\theta_j^3 n}\big) . \tag{7.14}$$

For $\ell \in [m]$, decomposing the events in the required lower bound on N_j gives

$$
\begin{aligned}
\mathbb{P}\left(\exists j \in [\ell] : N_j < n_j - \Delta_j\right) &= \mathbb{P}\big(\exists j \in [\ell-1] : N_j < n_j - \Delta_j\big) \\
&\quad + \mathbb{P}\big(\cap_{j\in[\ell-1]}\left\{N_j \geq n_j - \Delta_j\right\} \cap \left\{N_\ell < n_\ell - \Delta_\ell\right\}\big) ,
\end{aligned}
$$

as well as

$$\mathbb{P}\big(\cap_{j\in[\ell-1]}\{N_j \geq n_j - \Delta_j\} \cap \{N_\ell < n_\ell - \Delta_\ell\}\big)$$
$$\leq \mathbb{P}\,(N_{\ell-1} \geq n_{\ell-1} - \Delta_{\ell-1}\,,\ N_\ell < n_\ell - \Delta_\ell) \triangleq P_\ell\,.$$

Recall that γ_j is decreasing in j, thereby for any constant $c_1' > 0$ there exists a constant $c_2' > 0$ such that

$$\sum_{j=1}^{\ell} \mathrm{e}^{-c_1'\gamma_j^2} \leq \mathrm{e}^{-c_2'\gamma_\ell^2}\,.$$

It will thus suffice to show that $P_j \leq \mathrm{e}^{-c'\gamma_j^2}$ for some constant $c' > 0$ and every $j \in [m]$.

For $j = 1$, recall that N_1 is the number of active clones at the beginning of the algorithm (since all clones are initially unmatched, the passive vertices are those with precisely 2 clones, and all other vertices are active) and $n_1 = \lambda(1 - \lambda \mathrm{e}^{-\lambda})n = \mathbb{E}N_1$. In addition, $\Delta_1 = \frac{1}{100}\gamma_1\sqrt{n}$ and $\gamma_1 \to \infty$ with n (by the fact that $(1 - \beta)^{-m} \asymp 1/\varepsilon \to \infty$). Hence, by a standard application of the Central Limit Theorem to the i.i.d. random variables defined by the number of active clones that each vertex contributes, we deduce that $P_1 \leq \mathrm{e}^{-c_0\gamma_1^2}$ for some $c_0 > 0$ fixed.

Next, consider P_{j+1} for $j \in \{1,\ldots, m-1\}$. Combining (7.3) with (7.13) we get that for each such j

$$N_{j+1} - n_{j+1} \geq N_j - (1 - \beta)^2 n_j - M_j' - 2(B_j - b_j)$$
$$= (1 - (1 - \beta)^2)N_j - M_j' + (1 - \beta)^2(N_j - n_j) - 2(B_j - b_j)\,.$$

Therefore, the event addressed in P_{j+1} implies that

$$\Delta_{j+1} < n_{j+1} - N_{j+1} \leq M_j' - (1 - (1 - \beta)^2)N_j + (1 - \beta)^2\Delta_j + 2(B_j - b_j)\,,$$

and also $N_j \geq n_j - \Delta_j$. In particular, P_{j+1} is at most the probability that

$$M_j' - (1 - (1 - \beta)^2)N_j + 2(B_j - b_j) \geq \Delta_{j+1} - (1 - \beta)^2\Delta_j\,,\ \text{and}$$
$$N_j \geq n_j - \Delta_j\,.$$

Adding the fact that, by definitions (7.7) and (7.12),

$$\Delta_{j+1} = (1 - \beta)^2\Delta_j + \frac{\gamma_{j+1}}{100}\sqrt{\theta_{j+1}^3 n}\,,$$

and we deduce that

$$P_{j+1} \leq \mathbb{P}\left(B_j - b_j > \frac{\gamma_{j+1}}{400}\sqrt{\theta_{j+1}^3 n}\right)$$
$$+ \mathbb{P}\left(M_j' - (1 - (1 - \beta)^2)N_j > \frac{\gamma_{j+1}}{200}\sqrt{\theta_{\ell+1}^3 n}\,\big|\,N_j \geq n_j - \Delta_j\right)\,.$$

Combining (7.4) and (7.14), we can obtain an upper bound $e^{-c_1 \gamma_{\ell+1}^2}$ on the first term, for some constant $c_1 > 0$. At the same time, (7.2) provides an upper bound of $e^{-c_2 \gamma_{\ell+1}^2}$ on the second term, for some $c_2 > 0$ fixed.

Altogether, we have shown the desired upper bound on P_j for all $j \in [m]$, implying the first inequality in the lemma.

Recall now that inequality (7.6) gives that for some constant $c_3, c_4 > 0$,

$$\mathbb{P}\left(\exists j \in [\ell] : |H_j - \theta_j \lambda(1 - e^{-\theta_j \lambda} - \theta_j \lambda e^{-\theta_j \lambda})n| \geq \Delta_j\right)$$

$$\leq 2 \sum_{j=1}^{\ell} e^{-c_3 \gamma_j^2} \leq e^{-c_4 \gamma_\ell^2} .$$

Combining this with the fact $L_j \geq N_j - H_j$, as well as the above lower bound on N_j, yields the second statement of the lemma, as required. ∎

We can now derive a lower bound on the first time that Step 2 is applied (that is, the first time at which there are no light clones). Equivalently, this gives an upper bound on Λ_C (the x-coordinate of the cut-off line at that point).

By the definition, the number of light clones throughout the algorithm has the following property:

- As long as there light clones in the stack, in each step one of them will be popped and matched, and as a result, at most one new light clone will be created.
- If there are no light clones in the stack (and the algorithm is not concluded) then following Step 2 the stack will necessarily be comprised of a single heavy clone. This clone will then be popped in the next iteration of Step 1, while creating at most two new light clones.

That is to say, once the number of light clones drops to 0, it can never again exceed 2. In particular, if all the light clones disappear for the first time during phase j for some $j = 1, \ldots, m-1$, we must have that $L_m \leq 2$ (since L_j is a lower bound on the number of light clones at the start of phase j).

By (7.11) we have that

$$l_m = \theta_m \lambda(\theta_m - 1 + e^{-\theta_m \lambda})n .$$

By the definition of θ_λ and its asymptotic behavior (see (3.1),(3.2)), as well as the fact that $\theta_m = (1 + o(1))\theta_\lambda$, we have that for all $\theta_\lambda \leq x \leq \theta_m$ the function $f(x) = x - 1 + e^{-\lambda x}$ satisfies

$$f'(x) = 1 - \lambda e^{-\lambda x} = 1 - \lambda e^{-(1+o(1))\theta_\lambda \lambda} = 1 - \lambda(1 - \theta_\lambda)(1 - o(\theta_\lambda)) = (1 - o(1))\varepsilon .$$

Since $f(\theta_\lambda) = 0$ and $\theta_m = \theta_\lambda + \frac{\gamma}{\sqrt{\theta_m n}} = (2 + o(1))\varepsilon$ (see (7.9)), we can apply the Mean Value Theorem and get

$$l_m = \theta_m \lambda f(\theta_m) n = \theta_m \lambda \cdot (1 - o(1))\varepsilon \frac{\gamma}{\sqrt{\theta_m n}} n = (\tfrac{1}{2} - o(1))\gamma \sqrt{\theta_m^3 n} \ . \quad (7.15)$$

On the other hand, by (7.7) (and recalling requirement (3.3) from β)

$$\Delta_m = \frac{\gamma}{100} \sqrt{\theta_m^3 n} \sum_{i=1}^m (1 - \beta)^{(m-i)/4} \le \frac{\gamma}{100} \sqrt{\theta_m^3 n} \frac{1}{1 - (1 - \beta)^{1/4}}$$

$$\le \frac{\gamma}{100} \sqrt{\theta_m^3 n} \frac{1}{1 - (\tfrac{2}{3} - o(1))^{1/4}} \le \frac{\gamma}{8} \sqrt{\theta_m^3 n} \le \tfrac{1}{3} l_m \ , \quad (7.16)$$

where the last two inequalities hold for any large n. As $\theta_m = (1 + o(1))\theta_\lambda$ and $\theta_\lambda^3 n \to \infty$ with n, we immediately have that $l_m \to \infty$ as well. However, Lemma 7.3 gives that

$$\mathbb{P}(L_m \le 2) \le \mathbb{P}(L_m \le l_m - 2\Delta_m) \le e^{-c\gamma^2}$$

for some fixed $c > 0$. By the above discussion, this translates into an upper bound of Λ_C:

$$\mathbb{P}\left(\Lambda_C \ge \theta_\lambda \lambda + \tfrac{\gamma}{\sqrt{\theta_\lambda n}}\right) \le e^{-c\gamma^2} \ . \quad (7.17)$$

The above upper bound on Λ_C ensures that Step 2 is not applied in the first $m - 1$ phases, except with probability $\exp(-c\gamma^2)$. Recall that, if Step 2 has not yet been applied in phases $1, \ldots, j - 1$ then our lower bound (7.3) for N_j is in fact an equality, and similarly, L_j is precisely the number of light clones at the beginning of phase j. Therefore, assuming that indeed Step 2 was not applied in phases $1, \ldots, m - 1$ (we account for the above error probability), we may now apply the same proof of Lemma 7.3, this time with respect to the events $\{N_j > n_j + \Delta_j\}$ and $\{L_j > l_j + 2\Delta_j\}$. This gives the following matching upper bounds on the N_j's and L_j's:

Lemma 7.4. *There exists a constant $c > 0$ such that the following holds:*

$$\mathbb{P}\left(\exists j \in [m] : N_j > n_j + \Delta_j\right) \le e^{-c\gamma^2} \ ,$$

$$\mathbb{P}\left(\exists j \in [m] : L_j > l_j + 2\Delta_j\right) \le e^{-c\gamma^2} \ .$$

To obtain the required upper bound on Λ_C, assume that Step 2 was not applied in phases $1, \ldots, m - 1$ (this happens except with probability $\exp(-c\gamma^2)$). We now wish to show that all light clones will disappear shortly after commencing phase m with probability at least $1 - \exp(-c\gamma^2)$.

Since we did not apply Step 2 yet, the stack exclusively contains light clones, and a clone is active (more precisely, m-active) iff it has 3 unmatched clones or more. Now, if we ignore the passive clones, the algorithm must

remove at least 2 light clones from the stack in order to create a new light clone.

Suppose that at the beginning of phase m, the stack contains k light clones. By the above discussion, after matching k light clones, the stack will be of size at most $k/2$. Iterating, it follows that after matching at most $2k$ active clones, every light clone will disappear (the stack will be exhausted).

By Lemma 7.4 and the fact that $l_m + 2\Delta_m \leq 2\gamma\sqrt{\theta_m^3 n}$ for any large n with room to spare (as established in (7.15),(7.16)), there are $L_m \leq 2\gamma\sqrt{\theta_\lambda^3 n}$ light clones at the beginning of phase m, except with probability $1 - \mathrm{e}^{-c\gamma^2}$. Combined with the above argument, we conclude that the stack of light clones will be exhausted after matching at most $4\gamma\sqrt{\theta_\lambda^3 n}$ active clones, except with the above error probability.

We will next show that at least this many active clones will be matched by the time the cut-off line reaches the point $\theta_m\lambda - 10\frac{\gamma}{\sqrt{\theta_\lambda n}}$. Since $\theta_m \geq \theta_\lambda$, by definition (7.11) we have

$$n_m = \theta_m^2\lambda(1 - \lambda\mathrm{e}^{-\theta_m\lambda})n \geq \theta_m^2\lambda(1 - \lambda\mathrm{e}^{-\theta_\lambda\lambda})n = \theta_m^2(1 - \lambda(1 - \theta_\lambda))n$$
$$= \theta_m^2(-\varepsilon + (2 + o(1))\varepsilon)n = (\tfrac{1}{2} + o(1))\theta_m^3 n \ .$$

Together with Lemma 7.3, we deduce that for a sufficiently large n, there are $N_m \geq \frac{1}{3}\theta_\lambda^3 n$ unmatched active clones at the beginning of phase m, except with probability $1 - \mathrm{e}^{-c\gamma^2}$. Note that, by the assumption on γ,

$$\theta_m\lambda - 10\frac{\gamma}{\sqrt{\theta_\lambda n}} = \left(1 - 10\frac{\gamma}{(1 + o(1))\sqrt{\theta_\lambda^3 n}}\right)\theta_m = (1 - o(1))\theta_m \ .$$

Since the boundary marking the end of phase m is at $(1 - \beta)\theta_m$ (and β is bounded away from 0), the cut-off line moves through the entire interval between $\theta_m\lambda$ and the above point as part of phase m. Hence, we can use the original version of the Cut-Off Line Algorithm in order to analyze the number of active clones that are matched along this interval (without considering a potential change of phase).

Applying Lemma 7.1 with $\theta = 1 - 10\frac{\gamma}{\sqrt{\theta_\lambda^3 n}}$ and $k = \frac{1}{3}\theta_\lambda^3 n$, we can now deduce that, except with probability $\exp(-c\gamma^2)$ we match at least

$$\tfrac{5}{6}(1 - \theta^2)k = \tfrac{5}{6}(20 - o(1))\frac{\gamma}{\sqrt{\theta_\lambda^3 n}}k \geq (5 - o(1))\gamma\sqrt{\theta_\lambda^3 n}$$

active clones. Therefore,

$$\mathbb{P}\left(\Lambda_C \leq \theta_\lambda\lambda - \tfrac{\gamma}{\sqrt{\theta_\lambda n}}\right) \leq \mathrm{e}^{-c\gamma^2} \ .$$

Combining this bound with (7.17) completes the proof of (3.4). ∎

Acknowledgments

We wish to thank Asaf Nachmias for helpful discussions at an early stage of this project.

References

[1] D. Aldous, *Brownian excursions, critical random graphs and the multiplicative coalescent*, Ann. Probab. **25** (1997), no. 2, 812–854.

[2] D. Aldous, *The continuum random tree. III*, Ann. Probab. **21** (1993), no. 1, 248–289.

[3] D. Aldous and J. Pitman, *Tree-valued Markov chains derived from Galton-Watson processes*, Ann. Inst. H. Poincaré Probab. Statist. **34** (1998), no. 5, 637–686 (English, with English and French summaries).

[4] N. Alon and J. H. Spencer, *The probabilistic method*, 3rd ed., John Wiley & Sons Inc., Hoboken, NJ, 2008.

[5] A. D. Barbour and L. H. Y. Chen (eds.), *Stein's method and applications*, Lecture Notes Series. Institute for Mathematical Sciences. National University of Singapore, vol. 5, Published jointly by Singapore University Press, Singapore, 2005.

[6] I. Benjamini, G. Kozma, and N. C. Wormald, *The mixing time of the giant component of a random graph*. preprint, available at `http://arxiv.org/abs/math/0610459`.

[7] S. Bhamidi, G. Hooghiemstra, and R. van der Hofstad, *First passage percolation on random graphs with finite mean degrees*. preprint.

[8] B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin. **1** (1980), no. 4, 311–316.

[9] B. Bollobás, *Random graphs*, 2nd ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.

[10] B. Bollobás, *The evolution of random graphs*, Trans. Amer. Math. Soc. **286** (1984), no. 1, 257–274.

[11] F. Chung and L. Lu, *The diameter of sparse random graphs*, Adv. in Appl. Math. **26** (2001), no. 4, 257–279.

[12] J. Ding, J. H. Kim, E. Lubetzky, and Y. Peres, *Diameters in supercritical random graphs via first-passage percolation*. preprint.

[13] J. Ding, E. Lubetzky, and Y. Peres, *The mixing time of the near-critical random graph*. preprint.

[14] R. Durrett, *Probability: theory and examples*, 3rd ed., Duxbury Press, 2004.

[15] P. Erdős and A. Rényi, *On random graphs. I*, Publ. Math. Debrecen **6** (1959), 290–297.

[16] D. Fernholz and V. Ramachandran, *The diameter of sparse random graphs*, Random Structures Algorithms **31** (2007), no. 4, 482–516.

[17] N. Fountoulakis and B. A. Reed, *The evolution of the mixing rate of a simple random walk on the giant component of a random graph*, Random Structures Algorithms **33** (2008), no. 1, 68–86.

[18] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc. (N.S.) **43** (2006), no. 4, 439–561 (electronic).

[19] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[20] R. van der Hofstad and M. Keane, *An elementary proof of the hitting time theorem*, Amer. Math. Monthly **115** (2008), no. 8, 753–756.

[21] H. Kesten, *Aspects of first passage percolation*, École d'été de probabilités de Saint-Flour, XIV—1984, Lecture Notes in Math., vol. 1180, Springer, Berlin, 1986, pp. 125–264.

[22] J. H. Kim, *Finding cores of random 2-SAT formulae via Poisson cloning.* Preprint, `http://arxiv.org/abs/0808.1599`.

[23] J. H. Kim, *Poisson cloning model for random graphs*, International Congress of Mathematicians (ICM) (Madrid, Spain, August 2006). Preprint, available at `http://arxiv.org/abs/0805.4133`.

[24] T. Łuczak, *Component behavior near the critical point of the random graph process*, Random Structures Algorithms **1** (1990), no. 3, 287–310.

[25] T. Łuczak, *Cycles in a random graph near the critical point*, Random Structures Algorithms **2** (1991), no. 4, 421–439.

[26] T. Łuczak, B. Pittel, and J. C. Wierman, *The structure of a random graph at the point of the phase transition*, Trans. Amer. Math. Soc. **341** (1994), no. 2, 721–748.

[27] T. Łuczak and T. G. Seierstad, *The diameter behavior in the random graph process.* preprint.

[28] A. Nachmias and Y. Peres, *Critical random graphs: diameter and mixing time*, Ann. Probab. **36** (2008), no. 4, 1267–1286.

[29] R. Otter, *The multiplicative process*, Ann. Math. Statistics **20** (1949), 206–224.

[30] J. Pitman, *Enumerations of trees and forests related to branching processes and random walks*, Microsurveys in discrete probability (Princeton, NJ, 1997), DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 41, Amer. Math. Soc., Providence, RI, 1998, pp. 163–180.

[31] B. Pittel and N. C. Wormald, *Counting connected graphs inside-out*, J. Combin. Theory Ser. B **93** (2005), no. 2, 127–172.

[32] O. Riordan and N. C. Wormald, *The diameter of sparse random graphs.* preprint, available at `http://arxiv.org/abs/0808.4067`.

[33] F. Spitzer, *Principles of random walks*, 2nd ed., Springer-Verlag, New York, 1976. Graduate Texts in Mathematics, Vol. 34.

[34] N. C. Wormald, *Models of random regular graphs*, Surveys in combinatorics, 1999 (Canterbury), London Math. Soc. Lecture Note Ser., vol. 267, Cambridge Univ. Press, Cambridge, 1999, pp. 239–298.

Jian Ding
Department of Statistics, UC Berkeley, Berkeley, CA 94720, USA.
 E-mail address: `jding@stat.berkeley.edu`

Jeong Han Kim
Department of Mathematics, Yonsei University, Seoul 120-749 Korea, and
National Institute for Mathematical Sciences, Daejeon 305-340, Korea.
 E-mail address: `jehkim@yonsei.ac.kr`

Eyal Lubetzky
Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399, USA.
 E-mail address: `eyal@microsoft.com`

Yuval Peres
Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399, USA.
 E-mail address: `peres@microsoft.com`