Vertices of Degree k in Random Unlabeled Trees

Konstantinos Panagiotou *　　Makrand Sinha †‡

Abstract

Let $\mathcal{H}_n$ be the class of unlabeled trees with n vertices, and denote by H_n a tree that is drawn uniformly at random from this set. The asymptotic behavior of the random variable $\deg_k(\mathsf{H}_n)$ that counts vertices of degree k in H_n was studied, among others, by Drmota and Gittenberger in [3], who showed that this quantity satisfies a central limit theorem. This result provides a very precise characterization of the "central region" of the distribution, but does not give any non-trivial information about its tails.

In this work we study further the number of vertices of degree k in H_n. In particular, for $k = \mathcal{O}((\frac{\log n}{\log\log n})^{1/2})$ we show exponential-type bounds for the probability that $\deg_k(\mathsf{H}_n)$ deviates from its expectation. On the technical side, our proofs are based on the analysis of a randomized algorithm that generates unlabeled trees in the so-called *Boltzmann model*. The analysis of such algorithms is quite well-understood for classes of *labeled* graphs, see e.g. the work [1, 2] by Bernasconi, the first author, and Steger. Comparable algorithms for unlabeled classes are unfortunately much more complex. We demonstrate in this work that they can be analyzed very precisely for classes of *unlabeled* graphs as well.

1 Introduction

Let $\mathcal{H}_n$ denote the class of vertex-rooted unlabeled trees with n vertices. Pólya [11] and Otter [9] first studied the number of such trees, and showed that there is a constant $C > 0$ such that for large n

$$|\mathcal{H}_n| = (C + o(1))n^{-3/2}\rho^{-n},$$

where $\rho \approx 0.3383$ is the unique positive singularity of the ordinary generating function $H(z) = \sum_{n\ge1} |\mathcal{H}_n|z^n$ enumerating vertex-rooted unlabeled trees.

Robinson and Schwenk [12] showed by extending Otter's methods that the mean number of vertices of degree k in a random unlabeled tree is approximately $d(k)n$ for sufficiently large n, where the behavior of $d(k)$ is asymptotically given by $d(k) \sim c\rho^k$, and $c \approx 6.380045$ was estimated in [13]. However, they gave no further information about the number $\mathsf{deg}_k(\mathsf{H}_n)$ of vertices of degree k in a tree H_n drawn uniformly at random from $\mathcal{H}_n$.

Drmota and Gittenberger determined in [3], among other results, the limiting distribution of $\mathsf{deg}_k(\mathsf{H}_n)$ for any constant $k \ge 1$. Later, Gittenberger [7] studied systematically the distribution of $\mathsf{deg}_k(\mathsf{H}_n)$, when n *and* k both tend to infinity. He showed that $\mathsf{deg}_k(\mathsf{H}_n)$ is asymptotically normally distributed if $\mathbb{E}[\mathsf{deg}_k(\mathsf{H}_n)] \to \infty$, asymptotically Poisson distributed if $\mathbb{E}[\mathsf{deg}_k(\mathsf{H}_n)] \to C > 0$, and that if $\mathbb{E}[\mathsf{deg}_k(\mathsf{H}_n)] \to 0$, then the distribution degenerates. In other results, Goh and Schmutz [8] described the distribution of the maximum degree.

Although all the above results provide a precise characterization of the "central region" of the distribution, they do not give any *very sharp* concentration results, i.e., they do not provide any

*Department of Algorithms and Complexity, Max-Planck-Institut für Informatik, 66123 Saarbrücken, Germany. Email: kpanagio@mpi-inf.mpg.de

†Department of Computer Science, ETH Zürich, 8092 Zürich, Switzerland. Email: makrand@student.ethz.ch

‡Part of this work was done while visiting ETH Zürich in May-July '08 when the author was an undergraduate student at the Indian Institute of Technology Kanpur.

explicit description of the tails of the distribution. This question is one main topic of our work. In particular, for $k = \mathcal{O}((\frac{\log n}{\log\log n})^{1/2})$ we show exponential-type bounds for the probability that $\mathsf{deg}_k(\mathsf{H}_n)$ deviates from its expectation. More precisely, we prove the following statement. For brevity we write "$(1 \pm \varepsilon)X$" to denote the interval $((1-\varepsilon)X, (1+\varepsilon)X)$.

Theorem 1.1. *Let H_n be a tree drawn uniformly at random from $\mathcal{H}_n$, and let $k_0 = \frac{1}{5}\sqrt{\frac{\log n}{\log\log n}}$. There exists a $C > 0$ such that the following is true. For any $k \leq k_0$ there is a function $d(k)$ such that for any $0 < \varepsilon < 1$ and sufficiently large n*

$$\mathbb{P}[\mathsf{deg}_k(\mathsf{H}_n) \notin (1 \pm \varepsilon)d(k)n] \leq e^{-C\varepsilon^2 k^{-4k^2} n^{1/4}}.$$

Before we proceed with a more detailed exposition of the main proof ideas, let us collect a few general facts about our approach. In recent work, Bernasconi, the first author, and Steger [1, 2] and the first author and Weißl [10], investigated the properties of random graphs from certain classes of *labeled* graphs by exploiting the so-called *Boltzmann sampling* framework, which was introduced by Duchon, Flajolet, Louchard and Schaeffer [4]. More precisely, they defined an algorithm that generates graphs from the desired class according to some well-defined distribution, and then analyzed the characteristics of the returned object. This approach has the benefit that it reduces properties of a random graph from the class under consideration to properties of *independent* random variables.

The basic Boltzmann sampling framework for labeled structures was extended in [5] by Flajolet, Fusy, and Pivoteau to *unlabeled* structures. Due to inherent internal symmetries of unlabeled graphs, the corresponding sampling algorithms have a considerably more complex structure than their counterparts for labeled graphs. However, in this work we demonstrate that such algorithms also can be analyzed very precisely.

Outline In Section 2 we collect some basic facts that we will use throughout, and introduce briefly the Boltzmann model, tailored to our specific application. Section 3.1 describes the actual version of the sampler that we will exploit, and Section 3.2 states its most relevant combinatorial properties. Finally, the proof of the main theorem is described in Section 3.3, and Section 4 contains the proof of technical lemmas omitted in Section 3.

2 Preliminaries

2.1 Notation

Let $\mathcal{G}$ be a class of unlabeled graphs. Moreover, let $\mathcal{G}_n$ be the subset of graphs in $\mathcal{G}$ that have precisely n vertices, and set $G_n = |\mathcal{G}_n|$. We will write $G(x) = \sum_{n\geq 0} G_n x^n$ for the *ordinary generating function (OGF)* enumerating $\mathcal{G}$. We use $\mathsf{deg}_k(\gamma)$ to denote the number of vertices of degree k in a graph γ and $\mathsf{rdeg}(\gamma)$ to denote the degree of the root vertex in γ. Furthermore, let $\mathsf{deg}'_k(\gamma)$ be the number of vertices different from the root that have degree k in γ. Finally, the notation $\mathbf{I}[\mathfrak{A}]$ is used to represent an indicator variable that assumes value 1 when the statement $\mathfrak{A}$ is true, and 0 otherwise.

Basic Combinatorial Constructions Let $\mathcal{A}$ and $\mathcal{B}$ be two classes of unlabeled combinatorial objects. As described very detailed in [6], we will use the notation "$\mathcal{A} = \mathsf{MSET}(\mathcal{B})$" to denote that $\mathcal{A}$ is obtained by forming all finite multisets of elements from $\mathcal{B}$. Here, following common mathematical terminology, in a multiset the order between the elements does not count, but arbitrary repetitions of elements are allowed. Moreover, we will denote by $\mathcal{A} \times \mathcal{B}$ the class that contains all objects that are *ordered* pairs of objects from $\mathcal{A}$ and $\mathcal{B}$. For a very extensive treatment and many examples of those and several other constructions we refer the reader to [6].

2.2 The Boltzmann Model

Let $\mathcal{A}$ be an unlabeled combinatorial class. In the Boltzmann model of parameter x we assign to any $\gamma \in \mathcal{A}$ the probability

$$\mathbb{P}_x[\gamma] = \frac{x^{|\gamma|}}{A(x)}, \tag{2.1}$$

if the expression above is well-defined.

A *Boltzmann sampler* $\Gamma\mathcal{A}(x)$ for $\mathcal{A}$ is an algorithm that generates objects from $\mathcal{A}$ according to the probability distribution (2.1). Note that the sampler generates two objects of the same size with the same probability, since this probability only depends on the parameter x and the size n of the object. So, if we condition on the output being of a particular size n, then $\Gamma\mathcal{A}(x)$ is a uniform sampler for the class $\mathcal{A}_n$. The parameter x can be chosen accordingly to control the expected size of a random object from $\mathcal{A}$; for larger x the expected size becomes also bigger.

The Boltzmann sampling framework was introduced in [4, 5], where also numerous applications and examples can be found. In particular, in [5] the authors presented several rules that can be used to translate common combinatorial constructions like multiset or cartesian product in algorithms that sample objects from the underlying combinatorial classes according to the distribution (2.1). We will exploit these results later.

2.3 Tail Bounds

In our proofs we will often need to bound the probability that certain binomially or Poisson distributed random variables assume values far away from their expectations. The next two statements provide sharp bounds for this purpose.

Lemma 2.1 (Chernoff's Bounds). *Let X be distributed as $\mathsf{Bin}(n,p)$ and set $\mu = \mathbb{E}[X] = np$. There exists a $C > 0$ such that for every $0 < \varepsilon < 1$*

$$\mathbb{P}[X \in (1 \pm \varepsilon)\mu] \geq 1 - e^{-C\varepsilon^2\mu}.$$

Lemma 2.2 (Concentration of Poisson Variables). *Let X be distributed as $\mathsf{Po}(\mu)$, where $\mu > 0$. There exists a $C > 0$ such that for every $0 < \varepsilon < 1$*

$$\mathbb{P}[X \in (1 \pm \varepsilon)\mu] \geq 1 - e^{-C\varepsilon^2\mu}.$$

We will also need the following elementary statement for Poisson variables with small expectation. It is not best possible, but will suffice for our purposes.

Lemma 2.3. *Let X be distributed as $\mathsf{Po}(\mu)$, where $\mu > 0$. Then, for any $\delta > 0$,*

$$\mathbb{P}[X > 9\mu + \delta] \leq e^{-\delta}.$$

Proof. We will use the well-known fact $x! \geq (\frac{x}{e})^x$, which is true for any $x \geq 1$:

$$\mathbb{P}[X > 9\mu + \delta] = \sum_{a \geq 0} e^{-\mu} \frac{\mu^{\lceil 9\mu+\delta \rceil + a}}{(\lceil 9\mu + \delta \rceil + a)!} \ \leq e^{-\mu} \frac{\mu^{\lceil 9\mu+\delta \rceil}}{(\lceil 9\mu + \delta \rceil)!} \sum_{a \geq 0} \frac{\mu^a}{a!} \ \leq \frac{\mu^{\lceil 9\mu+\delta \rceil}}{\left(\frac{\lceil 9\mu+\delta \rceil}{e}\right)^{\lceil 9\mu+\delta \rceil}}.$$

Note that $\frac{e\mu}{\lceil 9\mu+\delta \rceil} \leq \frac{e}{9} \leq e^{-1}$, which completes the proof. ∎

3 Unlabelled non-plane trees

In this section we will use the framework of Boltzmann samplers to determine tail bounds for the degree sequence of random rooted unlabeled non-plane trees with n vertices, i.e., we will present the proof of Theorem 1.1. Note that these trees are called non-plane because the subtrees attached to a node are not ordered. This description suggests a natural recursive decomposition by viewing a tree as a *multiset* of trees attached to a special vertex, which is the root vertex of the whole tree. The class $\mathcal{H}$ of all such trees is definable by the following symbolic equation, which also provides an implicit equation satisfied by the OGF $H(z)$:

$$\mathcal{H} = \mathcal{Z} \times \mathsf{MSET}(\mathcal{H}) \implies H(z) = z \exp\left(\sum_{j\geq 1} \frac{H(z^j)}{j}\right). \tag{3.1}$$

where $\mathcal{Z}$ represents the class consisting of a single (unlabeled) node.

The above equation does not permit a closed form specification for the OGF $H(z)$, though all coefficients H_n can be determined recursively. However, it can be analyzed rigorously to obtain the following statement.

Lemma 3.1 (Polya, Otter). *$H(z)$ has a unique dominant singularity at $\rho \approx 0.3383219$ and satisfies locally around ρ*

$$H(z) = 1 - \ \eta\left(1 - \frac{z}{\rho}\right)^{1/2} + O\left(\left(1 - \frac{z}{\rho}\right)\right), \tag{3.2}$$

where $\eta = \alpha\sqrt{\rho}$ and $\alpha \approx 2.6811266$. Furthermore, for large n we have $H_n \sim \frac{\eta}{2\sqrt{\pi n^3}}\rho^{-n}$.

We will need very often the value of $H(z)$ and $H'(z)$ at some (large) power of its singularity. The following lemma gives a useful bound for this purpose.

Proposition 3.2. *Let $\rho < 1$ be the singularity of the generating function $H(z)$ for the class $\mathcal{H}$. Then, uniformly for all $j \geq 2$,*

a) $H(\rho^j) = \rho^j + \Theta(\rho^{2j})$.

b) $H'(\rho^j) = 1 + \Theta(\rho^j)$.

3.1 The Boltzmann Sampler

Based on the recursive description (3.1) and the rules for the construction of Boltzmann samplers for unlabeled structures in [5] we can construct a sampler $\Gamma\mathcal{H}(x)$ for $\mathcal{H}$ (Algorithm 1). The sampler starts by generating a single vertex, which will become the root vertex of the final tree, then generates a multiset of new nodes, and finally substitutes each of them by a random rooted tree in a recursive fashion.

We wish to remark that the Boltzmann sampler construction for multisets presented in [5] yields a sampler that runs in finite time. However, since here we are only interested in the distribution of the number of vertices of degree k, the only property of sampler which is of relevance is the distribution generated by it and not its running time. Hence, to simplify the analysis we use an equivalent sampler that requires *infinite* time; this sampler is in fact used in [5] as an intermediate step in the construction of the final finite sampler.

Lemma 3.3. *Let $\gamma \in \mathcal{H}$ and $0 < x \leq \rho$. Then $\mathbb{P}[\Gamma\mathcal{H}(x) = \gamma] = \frac{x^{|\gamma|}}{H(x)}$, and hence $\mathbb{P}[|\Gamma\mathcal{H}(x)| = n] = \frac{H_n x^n}{H(x)}$.*

Algorithm 1 Boltzmann Sampler $\Gamma\mathcal{H}(x)$ for class $\mathcal{H}$:

$\gamma \leftarrow$ a single node v
for $j = 1$ to ∞ **do**
  $\ell_j \leftarrow \mathsf{Po}\left(\frac{H(x^j)}{j}\right)$
  for $i = 1$ to ℓ_j **do**
    $\alpha_{j,i} \leftarrow \Gamma\mathcal{H}(x^j)$
    make j copies of $\alpha_{j,i}$
    connect the root of each one of them through an edge to v
  end for
end for
return γ, rooted at v

By exploiting this statement and Lemma 3.1 we obtain easily an asymptotic estimate for the probability that $\Gamma\mathcal{H}(x)$ outputs an object of a given size n. The proof is immediate and hence omitted.

Lemma 3.4. *Let $0 < x \le \rho$. There is a constant $C = C(x) > 0$ such that*

$$\mathbb{P}\left[|\Gamma\mathcal{H}(x)| = n\right] \sim \frac{C}{H(x)} n^{-3/2} \left(\frac{x}{\rho}\right)^n.$$

3.2 Properties of the Sampler

To inspect the properties of trees generated by Algorithm 1 we will now describe $\Gamma\mathcal{H}(x)$ in terms of an equivalent deterministic algorithm $\Gamma\mathcal{H}(x, L, T_2, T_3, \dots)$. We will do this by making the random choices necessary for the sampler in advance, and providing them to $\Gamma\mathcal{H}(x, L, T_2, T_3, \dots)$ as part of the input. More precisely, let $L = (B_1, B_2, \dots)$ be an infinite sequence of *blocks* $B_1, B_2, \dots$, each of which is again an infinite sequence of variables. The jth element of the mth block, denoted by $B_m[j]$, is a random value distributed as $\mathsf{Po}\left(H(x^j)/j\right)$. Moreover, for $j \ge 2$, let T_j be an infinite list of trees drawn from $\mathcal{H}$ according to the Boltzmann Model with parameter x^j.

$\Gamma\mathcal{H}(x, L, T_2, T_3, \dots)$ works as follows: it starts by generating a single vertex (the root), and by reading the first block of L. For $j \ge 2$, it takes the first $B_1[j]$ trees from T_j, makes j copies of each one of them, and attaches them by an edge to the root vertex. Then it removes the first block from L, and also the first $B_1[j]$ trees from T_j for all $j \ge 2$. Finally, it calls itself recursively $B_1[1]$ times (consequently, these runs of the sampler will use the next blocks from L, and the next trees from the lists $T_2, T_3, \dots$), and attaches the $B_1[1]$ outputs to the root vertex.

Clearly, $\Gamma\mathcal{H}(x, L, T_2, T_3, \dots)$ simulates the execution of $\Gamma\mathcal{H}(x)$ – in particular, all random choices made in the samplers are made according to exactly the same probability distributions. This implies that $\Gamma\mathcal{H}(x, L, T_2, T_3, \dots)$ and $\Gamma\mathcal{H}(x)$ have precisely the same output distribution. Hence, in the sequel we shall consider only $\Gamma\mathcal{H}(x, L, T_2, T_3, \dots)$.

Based on this modification, we can state the following structural lemma, which relates the number of vertices of degree k in the output of the sampler to the random choices performed during its execution. The statement of the lemma follows immediately from the model, but for completeness we give a formal proof in Section 4. For notational convenience we will sometimes omit the input lists from the description of $\Gamma\mathcal{H}(x, L, T_1, T_2, \dots)$.

Lemma 3.5. *Consider one run of $\Gamma\mathcal{H}(x, L, T_2, T_3, \dots)$ that returned a tree $\gamma \in \mathcal{H}$. Let N_j, $j \ge 2$, be the number of trees used by the sampler from the list T_j, and B the number of blocks from L read by the sampler. Then the following holds.*

a) $N_j = \sum_{m=1}^{B} B_m[j]$.

b) $|\gamma| = B + \sum_{j\geq 2} j\left[\sum_{i=1}^{N_j} |T_j[i]|\right]$.

c) $\mathsf{deg}'_k(\gamma) = \sum_{j\geq 2} j\left[\sum_{i=1}^{N_j} \mathsf{deg}'_k(T_j[i]) + \mathbf{I}[\mathsf{rdeg}(T_j[i]) = k-1]\right] + \sum_{m=2}^{B} \mathbf{I}\left[\sum_{i\geq 1} iB_m[i] = k-1\right]$.

In the remainder of this section we collect some additional properties of trees generated by $\Gamma\mathcal{H}(x)$. First, recall that Lemma 3.4 states that the probability that $\Gamma\mathcal{H}(x)$ (and thus also $\Gamma\mathcal{H}(x, L, T_2, T_3, \dots)$) produces a tree of size precisely n is asymptotically $\frac{C}{H(x)} n^{-3/2}\left(\frac{x}{\rho}\right)^n$, for some absolute $C > 0$. If we set $x = \rho^j$, the probability of drawing a tree of size n is exponentially small (as $\rho < 1$), while if $x = \rho$, then this probability is proportional to $n^{-3/2}$. So, if $0 < x < \rho$, then many independently generated trees should behave as they are expected to behave, as their sizes are concentrated among small values. The next lemma formalizes this statement, with respect to the total size of those trees.

Lemma 3.6. *There is a constant $C > 0$ such that the following holds uniformly for all $x \in (0, \rho^2]$. Let $0 < \varepsilon < 1$ and let $\gamma_1, \dots, \gamma_N$ be random trees from $\mathcal{H}$ drawn independently according to the Boltzmann model with parameter x. Then*

$$\mathbb{P}\left[\sum_{i=1}^{N} |\gamma_i| \notin (1\pm\varepsilon)\frac{xH'(x)}{H(x)}N\right] \leq e^{-C\varepsilon^2 \frac{xH'(x)}{H(x)}N + O(\log \varepsilon^{-1})}.$$

Proof. Let $\delta \geq 0$ be such that $x + \delta \leq \frac{\rho^2+\rho}{2}$. Then, as $H(z)$ has only non-negative coefficients and is analytic in its disc of convergence, there is an absolute constant $c > 0$ such that

$$H(x+\delta) \leq H(x) + \delta H'(x) + \delta^2 c.$$

Here, one might for example choose $c = \frac{1}{2}H''(\frac{\rho^2+\rho}{2})$. A straightforward induction argument over N shows that for any s

$$\mathbb{P}\left[\sum_{i=1}^{N} |\gamma_i| = s\right] = \frac{x^s [z^s] H(z)^N}{H(x)^N},$$

where, as usual, $[z^n]G(z)$ denotes the nth coefficient in the Taylor expansion of a function G around $z = 0$. Note that if $G(z)$ has only non-negative coefficients, then for any $r > 0$ that is strictly smaller than the radius of convergence of G we have that $[z^n]G(z) \leq G(r)r^{-n}$.

Let us consider first the upper tail. For this let $s = (1+\varepsilon)\frac{xH'(x)}{H(x)}N$. Using the above facts, we obtain for any $x < r \leq (\rho^2+\rho)/2$

$$p_s := \mathbb{P}\left[\sum_{i=1}^{N} |\gamma_i| \geq s\right] \leq \left(\frac{x}{r}\right)^s \left(\frac{H(r)}{H(x)}\right)^N \frac{r}{r-x} \tag{3.3}$$

Write $r = x + \delta$. Then by using the fact that $1 + z \leq e^z$ for all z we may estimate

$$\left(\frac{x}{r}\right)^s \leq \exp\left\{-\frac{\delta s}{x+\delta}\right\} \leq \exp\left\{-\frac{\delta s}{x} + \frac{\delta^2 s}{x^2}\right\}.$$

Moreover, by exploiting the Taylor expansion of H around x we obtain

$$\left(\frac{H(r)}{H(x)}\right)^N \leq \left(1 + \delta\frac{H'(x)}{H(x)} + \delta^2\frac{c}{H(x)}\right)^N \leq \exp\left\{\left(\delta\frac{H'(x)}{H(x)} + \delta^2\frac{c}{H(x)}\right)N\right\}.$$

All the above bounds are true for any $\delta \geq 0$ such that $x+\delta \leq (\rho^2+\rho)/2$. For the remaining calculation we shall assume that

$$\delta = \min\left\{\frac{\varepsilon}{2(\frac{1+\varepsilon}{x} + c)}, \frac{\rho^2+\rho}{2} - x\right\}.$$

Using the above, we can get an upper bound on $\frac{r}{r-x} = \frac{x+\delta}{\delta}$ as follows. If $\delta = \frac{\varepsilon}{2(\frac{1+\varepsilon}{x}+c)}$, then $\frac{r}{r-x} = \mathcal{O}(\varepsilon^{-1})$. On the other hand, if $\delta = \frac{\rho^2+\rho}{2} - x$, then $\frac{r}{r-x} = (\frac{\rho^2+\rho}{2})/(\frac{\rho^2+\rho}{2} - x) = \mathcal{O}(1)$ as $x \le \rho^2$. Altogether, we have that $\frac{r}{r-x} = \mathcal{O}(\varepsilon^{-1})$. By plugging all these bounds into (3.3) we obtain due to $H'(x) \ge 1$

$$\log p_s \le \delta \frac{H'(x)}{H(x)} \left(-\varepsilon + \delta \left(\frac{1+\varepsilon}{x} + c \right) \right) N + O(\log \varepsilon^{-1}).$$

By the choice of δ, this is at most

$$-\delta \frac{H'(x)}{H(x)} \frac{\varepsilon}{2} N + O(\log \varepsilon^{-1}) = -\frac{\varepsilon^2}{4(2+c)} \frac{xH'(x)}{H(x)} N + O(\log \varepsilon^{-1}).$$

A completely analogous argument can be performed to bound also the lower tail, i.e., the probability that $\sum_{i=1}^{N} |\gamma_i|$ is less than $(1-\varepsilon)\frac{xH'(x)}{H(x)}N$. This completes the proof. □

The next statement says that in a sequence of trees drawn from $\mathcal{H}$ independently according to the Boltzmann distribution with parameter $0 < x < \rho$ the total number of vertices of degree k is sharply concentrated around its expected value. This shall become useful later on, when we will count the number of vertices of degree k in large uniformly drawn random trees.

Lemma 3.7. *Let $\gamma_1, \dots, \gamma_N$ be trees from $\mathcal{H}$ drawn independently according to the Boltzmann model with parameter $0 < x < \rho$. Let $D_k = D_k(\gamma_1, \dots, \gamma_N) := \sum_{i=1}^{N} \mathsf{deg}'_k(\gamma_i)$. Let $k = k(N)$ be such that $\mathbb{E}[\mathsf{deg}'_k(\gamma_i)] \ge \log^3 N/\sqrt{N}$. Then there is a $C > 0$ such that for every $0 < \varepsilon < 1$ and sufficiently large N*

$$\mathbb{P}\left[D_k \notin (1 \pm \varepsilon)\mathbb{E}[D_k]\right] \le e^{-C\varepsilon^2 \mathbb{E}[\mathsf{deg}'_k(\gamma_1)]^{2/3} N^{1/3}}.$$

Proof. Since $0 < x < \rho$, by applying Lemma 3.4 we observe that $\mathbb{P}[|\gamma_i| = n] \le C^{-n}$ for some $C > 1$. The proof then follows by a straightforward application of Lemma 5.7 from [2] to $\gamma_1, \dots, \gamma_N$ and by setting $X(\gamma_i) = \mathsf{deg}'_k(\gamma_i)$ in that lemma. □

3.3 Distribution of Degree

In this subsection we prove our main result, namely that the number of vertices of degree k in large uniformly drawn unlabeled trees is concentrated around its expected value. More precisely, we have the following result.

Lemma 3.8. *Let $\gamma = \Gamma\mathcal{H}(\rho, L, T_2, T_3, \dots)$, n be a positive integer and set $k_0 = \frac{1}{5}\sqrt{\frac{\log n}{\log\log n}}$. There is a $C > 0$ such that for every $k \le k_0$ and $0 < \varepsilon < 1$ the following is true for sufficiently large n. There is a $d_k > 0$ such that*

$$\mathbb{P}\big[\mathsf{deg}_k(\gamma) \notin (1 \pm \varepsilon)d_k n \bigm| |\gamma| = n\big] \le e^{-C\varepsilon^2 k^{-4k^2} n^{1/4}}.$$

Using this statement, by applying Lemma 3.3 we readily obtain Theorem 1.1. In the sequel we present the proof of Lemma 3.8, which proceeds as follows. Recall that by Lemma 3.5 we have

$$\mathsf{deg}'_k(\gamma) = \sum_{j \ge 2} j \left[\sum_{i=1}^{N_j} \mathsf{deg}'_k(T_j[i]) + \mathbf{I}[\mathsf{rdeg}(T_j[i]) = k-1] \right] + \sum_{m=2}^{B} \mathbf{I}\left[\sum_{i \ge 1} iB_m[i] = k-1 \right], \qquad (3.4)$$

where N_j denotes the number of trees that were used from the sequence T_j, and B the number of blocks read from L. Our strategy will be, conditioned on the event "$|\gamma| = n$", to determine tight bounds for the random quantities N_j, where $j \ge 2$, and B. Then, by using Lemmas 3.6, 3.7 and standard tail bounds we will bound the sums of random variables $\mathsf{deg}'_k(T_j[i])$ and indicator variables in (3.4). For a clear presentation of the proof of Lemma 3.8, we shall defer the proofs of all the intermediate technical lemmas to Section 4.

To start with, we consider trees that are drawn from T_j, where j is "large".

Proposition 3.9. *Let $\gamma = \Gamma\mathcal{H}(\rho, L, T_2, T_3, \dots)$. Moreover, let N_j, $j \geq 2$, be the random variable counting the number of trees used from T_j, and B be the random number of used blocks from L. Then if $b \leq n$, $j_0 \geq \log^2 n$ and n is sufficiently large the following holds,*

$$\mathbb{P}\left[\sum_{j>j_0} N_j > 0 \;\middle|\; B = b\right] \leq e^{-j_0/2}.$$

Next we turn to the case $2 < j \leq j_0$. We prove that the total number of vertices due to subtrees read from T_j is "small", if j lies in the mentioned range.

Proposition 3.10. *Let $\gamma = \Gamma\mathcal{H}(\rho, L, T_2, T_3, \dots)$. Moreover, let N_j, $j \geq 2$, be the random variable counting the number of trees from T_j used by the sampler, and B the random number of used blocks from L. Furthermore, for $j \geq 2$, set $V_j := j \sum_{i=1}^{N_j} |T_j[i]|$. Then, for any constant $c > 0$, $n \geq b$, $y \geq \log\log n$, $t \geq \log^2 n$, any $y \leq j_0 \leq n^c$ and sufficiently large b*

$$\mathbb{P}\left[\sum_{j=y}^{j_0} V_j \leq 22b \cdot \rho^y + t j_0^2 \;\middle|\; B = b\right] \geq 1 - e^{-\Omega(t)}.$$

Equipped with the above lemma, part *b)* of Lemma 3.5 and Proposition 3.2 we obtain the following milestone relating B with the size of the tree, which essentially establishes that generating a large tree also requires a large number of blocks from the list L.

Lemma 3.11. *Let $\gamma = \Gamma\mathcal{H}(\rho, L, T_2, T_3, \dots)$. Moreover, let B be the random number of blocks used from L. Then for any $0 < \varepsilon < 1$ and sufficiently large n*

$$\mathbb{P}\big[B \notin (1 \pm \varepsilon)bn \mid |\gamma| = n\big] = e^{-\Omega(n^{1/4})},$$

where $b = 1/(1 + \eta^2/2)$ with η given by Lemma 3.1.

Equipped with all those tools we are ready to prove our main lemma.

Proof of Lemma 3.8. Recall that by Lemma 3.5 we have

$$\mathsf{deg}'_k(\gamma) = \sum_{j\geq 2} j \left[\sum_{i=1}^{N_j} \mathsf{deg}'_k(T_j[i]) + \mathbf{I}[\mathsf{rdeg}(T_j[i]) = k-1]\right] + \sum_{m=2}^{B} \mathbf{I}\left[\sum_{i\geq 1} iB_m[i] = k-1\right], \tag{3.5}$$

where N_j is the number of trees that were used from the list T_j and B the number of blocks used from L. Let $t_k = 3k \log_{1/\rho} k + \log\log n$. We shall split the above expression as follows.

$$S_1 = \sum_{m=2}^{B} \mathbf{I}\left[\sum_{i\geq 1} iB_m[i] = k-1\right],$$

$$S_2 = \sum_{j=2}^{t_k} j \left[\sum_{i=1}^{N_j} \mathsf{deg}'_k(T_j[i])\right],$$

$$S_3 = \sum_{j=2}^{t_k} j \left[\sum_{i=1}^{N_j} \mathbf{I}[\mathsf{rdeg}(T_j[i]) = k-1]\right].$$

Moreover, let

$$S_4 = \sum_{j=t_k+1}^{n^{1/4}} j \left[\sum_{i=1}^{N_j} \mathsf{deg}'_k \left(T_j[i] \right) + \mathbf{I}[\mathsf{rdeg}\left(T_j[i] \right) = k-1] \right],$$

$$S_5 = \sum_{j>n^{1/4}} j \left[\sum_{i=1}^{N_j} \mathsf{deg}'_k \left(T_j[i] \right) + \mathbf{I}[\mathsf{rdeg}\left(T_j[i] \right) = k-1] \right].$$

Denote the event "$B \in (1 \pm \varepsilon/2)bn$", where b is given in Lemma 3.11, by $\mathfrak{B}$. Conditioned on $\mathfrak{B}$ we claim the following about the random variables $S_1, \dots, S_5$. For S_1 we claim that there is a quantity $p_k \geq n^{-o(1)}$ and $C_1 > 0$ such that

$$(\mathcal{S}_1) \; : \; \mathbb{P}[S_1 \notin (1 \pm \varepsilon) p_k bn \bigm| \mathfrak{B}] \leq e^{-C_1 \varepsilon^2 p_k n}.$$

For the random variables S_2 and S_3 we make statements similar to $\mathcal{S}_1$, although the concentration is not as sharp. In particular, we show that there are $s_2 = s_2(k) > 0$ and $s_3 = s_3(k) > 0$ such that

$$(\mathcal{S}_2) \; : \; \mathbb{P}[S_2 \notin (1 \pm \varepsilon) s_2 bn \bigm| \mathfrak{B}] \leq e^{-C_2 \varepsilon^2 k^{-3k^2} n^{1/4}}.$$

$$(\mathcal{S}_3) \; : \; \mathbb{P}[S_3 \notin (1 \pm \varepsilon) s_3 bn \bigm| \mathfrak{B}] \leq e^{-C_3 \varepsilon^2 k^{-4k^2} n^{1/4}}.$$

For S_4 and S_5 we cannot make such concentration statements. However, we show that with sufficiently high probability the value that S_4 assumes is negligible and that $S_5 = 0$.

$$(\mathcal{S}_4) \; : \; \mathbb{P}[S_4 = o(p_k n) \bigm| \mathfrak{B}] \geq 1 - e^{-\Omega(n^{1/4})}.$$

$$(\mathcal{S}_5) \; : \; \mathbb{P}[S_5 = 0 \bigm| \mathfrak{B}] \geq 1 - e^{-n^{1/4}/2}.$$

By taking union of the statements $\mathcal{S}_1, \dots, \mathcal{S}_5$ it follows that for sufficiently large n and a positive constant c, there exists a $d_k > 0$ such that

$$\mathbb{P}\big[\mathsf{deg}_k(\gamma) \notin (1 \pm \varepsilon) d_k n \bigm| \mathfrak{B}\big] \leq e^{-c\varepsilon^2 k^{-3k^2} n^{1/4}}.$$

Lemma 3.11 says that $\mathbb{P}[\overline{\mathfrak{B}} \mid |\gamma| = n] = e^{-\Omega(n^{1/4})}$. The proof of the lemma then routinely completes by a simple averaging argument.

Proof of claims $\mathcal{S}_1, \dots, \mathcal{S}_5$.

($\mathcal{S}_1$): Note that, S_1 is distributed like $\mathsf{Bin}(B, p_k)$, where $p_k = \mathbb{P}[\sum_{i\geq 1} i B_1[i] = k-1]$. By conditioning on $\mathfrak{B}$ and applying the Chernoff's Bounds, we deduce that for some $C_1 > 0$ and sufficiently large n

$$\mathbb{P}[S_1 \notin (1 \pm \varepsilon) p_k bn \bigm| \mathfrak{B}] \leq e^{-C_1 \varepsilon^2 p_k n}.$$

As $k \leq \frac{1}{5}\sqrt{\frac{\log n}{\log\log n}}$, for further reference let us note that with plenty of room to spare that p_k is not too small :

$$p_k \geq \mathbb{P}[\mathsf{Po}\left(H(\rho)\right) = k-1] \cdot \prod_{j\geq 2} \mathbb{P}\left[\mathsf{Po}\left(\frac{H(\rho^j)}{j}\right) = 0\right] = \frac{\rho H(\rho)^{k-2}}{(k-1)!} = n^{-o(1)}. \tag{3.6}$$

($\mathcal{S}_2$): Let $2 \leq j \leq t_k$ and $\mathfrak{N}_j$ denote the event "$N_j \in (1 \pm \frac{2\varepsilon}{3}) bn \frac{H(\rho^j)}{j}$". Since N_j is distributed like $\mathsf{Po}\left(B\frac{H(\rho^j)}{j}\right)$, conditioning on $\mathfrak{B}$ and by applying Lemma 2.2 and Proposition 3.2, part *a)* we obtain that there is a $c > 0$ such that

$$\mathbb{P}\left[\neg\mathfrak{N}_j \bigm| \mathfrak{B}\right] \leq e^{-c\varepsilon^2 n\rho^j/j}. \tag{3.7}$$

To simplify the notation set $D'_{k,j} = \sum_{i=1}^{N_j} \mathsf{deg}'_k\,(T_j[i])$. Note that $\mathbb{E}[\mathsf{deg}'_k\,(T_j[i])] \geq \frac{\rho^{j(k+1)}}{H(\rho^j)}$, as there is a tree with $k+1$ vertices such that one vertex is connected to all other vertices one of which is the root and its probability in the Boltzmann model is exactly $\frac{\rho^{j(k+1)}}{H(\rho^j)}$. Therefore, with Proposition 3.2, part *a)* we obtain for a suitable $c' > 0$ and large n

$$\mathbb{E}[\mathsf{deg}'_k\,(T_j[i])] \geq \frac{\rho^{j(k+1)}}{\rho^j + \Theta(\rho^{2j})} \geq c'\rho^{jk} \geq c'\rho^{3k^2 \log_{1/\rho} k} \geq c' k^{-3k^2} \geq n^{-1/5}.$$

On the other hand, note that with plenty of room to spare $bn\frac{H(\rho^j)}{j} = \Omega(n\rho^j/j) \geq n^{1-o(1)}$. Hence, conditioned on $\mathfrak{N}_j$, we have for large n that $\mathbb{E}[\mathsf{deg}'_k\,(T_j[i])] \geq \frac{\log^3 N_j}{N_j^{1/2}}$. By applying Lemma 3.7 we may thus infer with plenty of room to spare that, say,

$$\mathbb{P}\left[jD'_{k,j} \notin (1 \pm \varepsilon)j\mathbb{E}[D'_{k,j} \mid \mathfrak{N}_j] \;\middle|\; \mathfrak{N}_j\right] \leq e^{-\Omega(\varepsilon^2 \mathbb{E}[\mathsf{deg}'_k(T_j[1])]^{2/3} N_j^{1/3})} \leq e^{-\Omega(\varepsilon^2 k^{-3k^2} n^{1/4})},$$

where

$$\mathbb{E}[D'_{k,j} \mid \mathfrak{N}_j] \sim \mathbb{E}[D'_{k,j} \mid \mathfrak{B}] \sim bn\frac{H(\rho^j)}{j}\mathbb{E}[\mathsf{deg}'_k\left(\Gamma\mathcal{H}(\rho^j)\right)].$$

By a simple averaging argument it then follows that

$$\mathbb{P}\left[jD'_{k,j} \notin (1 \pm \varepsilon)j\mathbb{E}[D'_{k,j} \;\middle|\; \mathfrak{B}] \;\middle|\; \mathfrak{B}\right] \leq e^{-\Omega(\varepsilon^2 k^{-3k^2} n^{1/4})}.$$

Now, by summing over all values of $2 \leq j \leq t_k = 3k\log_{1/\rho} k + \log\log n$ we obtain by using the union bound

$$\mathbb{P}[S_2 \notin (1 \pm \varepsilon)s_2(k,n) \;|\; \mathfrak{B}] \leq t_k e^{-\Omega(\varepsilon^2 k^{-3k^2} n^{1/4})} \leq e^{-\Omega(\varepsilon^2 k^{-3k^2} n^{1/4})},$$

where $s_2(k,n) = \sum_{j=2}^{t_k} j\mathbb{E}[D'_{k,j} \;\big|\; \mathfrak{B}] \sim bn\sum_{j\geq 2} H(\rho^j)\mathbb{E}[\mathsf{deg}'_k\left(\Gamma\mathcal{H}(\rho^j)\right)]$. The claim follows by setting $s_2 = s_2(k,n)/bn$.

($\mathcal{S}_3$): Similar to the case where we estimated S_1, the sum $\mathsf{Rdeg}_{k,j} := \sum_{i=1}^{N_j} \mathbf{I}[\mathsf{rdeg}\,(T_j[i]) = k-1]$ is distributed as $\mathsf{Bin}(N_j, p_{k,j})$ where $p_{k,j} = \mathbb{P}[\mathsf{rdeg}\,(T_j[i]) = k-1]$. Note that for sufficiently large k

$$p_{k,j} \geq \mathbb{P}[\mathsf{Po}\left(H(\rho^j)\right) = k-1] \cdot \prod_{m\geq 2} \mathbb{P}\left[\mathsf{Po}\left(\frac{H(\rho^{jm})}{j}\right) = 0\right] = \frac{\rho^j H(\rho^j)^{k-2}}{(k-1)!} \geq k^{-4k^2},$$

where in the last step we used Proposition 3.2. Conditioned on the event $\mathfrak{N}_j$ we obtain by Chernoff's Bounds

$$\mathbb{P}\left[\mathsf{Rdeg}_{k,j} \notin (1 \pm \varepsilon)p_{k,j} bn\frac{H(\rho^j)}{j} \;\middle|\; \mathfrak{N}_j\right] \leq e^{-\Omega(\varepsilon^2 p_{k,j} N_j)} \leq e^{-\Omega(\varepsilon^2 k^{-4k^2} n^{1-o(1)})}.$$

Using (3.7) we obtain for some $C > 0$

$$\mathbb{P}\left[j\mathsf{Rdeg}_{k,j} \notin (1 \pm \varepsilon)p_{k,j} bnH(\rho^j) \;\middle|\; \mathfrak{B}\right] \leq e^{-C\varepsilon^2 k^{-4k^2} n^{1/4}},$$

and using the union bound for all $2 \leq j \leq t_k$ yields the claimed statement, for $s_3(k,n) = bn\sum_{j=2}^{t_k} p_{k,j} H(\rho^j)$, and $s_3 = s_3(k,n)/bn$.

($\mathcal{S}_4$): By applying Proposition 3.10 for $B \in (1 \pm \frac{\varepsilon}{2})bn$ and setting $y = t_k + 1$, $j_0 = n^{1/4}$ and $t = n^{1/4}$ we obtain for large n that

$$\mathbb{P}\left[\sum_{j=t_k+1}^{n^{1/4}} V_j \leq 33bn\rho^{t_k+1} + n^{3/4} \;\middle|\; \mathfrak{B}\right] \geq 1 - e^{-\Omega(n^{1/4})}$$

where $V_j = j\sum_{i=1}^{N_j}|T_j[i]|$. As $k \le k_0$, it follows that

$$33bn\rho^{t_k+1} + n^{3/4} = O\left(\frac{k^{-3k}n}{\log\log n} + n^{3/4}\right) \overset{(3.6)}{=} o(p_k n).$$

Since $\mathsf{deg}'_k(T_j[i]) + \mathbf{I}[\mathsf{rdeg}(T_j[i]) = k-1] \le |T_j[i]|$ we obtain that $S_4 = o(p_k n)$ with probability at least $1 - e^{-\Omega(n^{1/4})}$, as claimed.

($\mathcal{S}_5$): By applying Lemma 3.9 conditioned on $\mathfrak{B}$ we obtain

$$\mathbb{P}[S_5 \neq 0 \mid \mathfrak{B}] = \mathbb{P}\left[\sum_{j>n^{1/4}} N_j > 0 \;\middle|\; \mathfrak{B}\right] \le e^{-n^{1/4}/2}. \qquad \square$$

4 Proofs of Technical Lemmas

4.1 Proofs of Section 3.2

Proof of Proposition 3.2. By applying Lemma 3.1 we obtain that there exists a $c > 0$ such that $H_n \le cn^{-3/2}\rho^{-n} \le c\rho^{-n}$. It is also easily established that $H_1 = 1$ and $H_2 = 1$. Thus,

$$H(\rho^j) = \sum_{n\ge1} H_n\rho^{jn} \le \rho^j + \rho^{2j} + \sum_{n\ge3} c\rho^{-n}\rho^{jn} = \rho^j + \rho^{2j} + o(\rho^{2j}).$$

To see the lower bound on $H(\rho^j)$, note that trivially $H(\rho^j) = \sum_{n\ge1} H_n\rho^{jn} \ge \rho^j + \rho^{2j}$. This proves part *a)*. Part *b)* follows from similar analysis. $\square$

Proof of Proposition 3.5. We shall prove the statements *a)*–*c)* by induction over the structure of γ. We first sketch our general strategy, and then implement the specific details for each case. First of all, note that if γ is a single node, then all statements are trivially true. In all other cases γ looks as follows. There are

- $B_1[1]$ subtrees $\gamma_1, \ldots, \gamma_{B_1[1]}$ generated by recursive calls to $\Gamma\mathcal{H}(x, L, T_2, T_3, \ldots)$, and
- for $j \ge 2$, j copies of each of $T_j[1], \ldots, T_j[B_1[j]]$

adjacent to the root of γ. Note that the trees $\gamma_1, \ldots, \gamma_{B_1[1]}$ are generated recursively and sequentially. So, there are indexes $2 = i_1 < i_2 \cdots < i_{B_1[1]+1} = B + 1$ such that γ_ℓ was generated by using the list $L_\ell = (B_{i_\ell}, \ldots, B_{i_{\ell+1}-1})$, and for $j \ge 2$ there are indexes $B_1[j]+1 = i_{j,1} \le \cdots \le i_{j,B_1[1]+1} = N_j + 1$ such that γ_ℓ was generated by using from T_j the trees in $T'_{j,\ell} = (T_j[i_{j,\ell}], \ldots, T_j[i_{j,\ell+1} - 1])$. We now may apply the induction hypothesis to γ_ℓ with the lists L_ℓ and $T'_{j,\ell}$, for all $j \ge 2$ and $1 \le \ell \le B_1[1]$.

To see *a)*, note that by applying the induction hypothesis we can assume that the total number of blocks from L needed to generate $\gamma_1, \ldots, \gamma_{B_1[1]}$ is precisely $\sum_{m=2}^{B} B_m[j]$. But then,

$$N_j = B_1[j] + \sum_{m=2}^{B} B_m[j] = \sum_{m=1}^{B} B_m[j].$$

Next we show *b)*. By applying the induction hypothesis we readily obtain

$$|\gamma_1| + \cdots + |\gamma_{B_1[1]}| = (B-1) + \sum_{j\ge2} j\left[\sum_{i=B_1[j]+1}^{N_j} |T_j[i]|\right].$$

But then, as the trees from T_j are copied j times, we deduce that the number of vertices in γ is

$$|\gamma| = 1 + |\gamma_1| + \cdots + |\gamma_{B_1[1]}| + \sum_{j\geq 2} j \left[\sum_{i=1}^{B_1[j]} |T_j[i]|\right] = B + \sum_{j\geq 2} j \left[\sum_{i=1}^{N_j} |T_j[i]|\right],$$

as claimed. In order to prove *c)*, note that by the induction hypothesis we obtain for $1 \leq \ell \leq B_1[1]$

$$\mathsf{deg}'_k(\gamma_\ell) = \sum_{j\geq 2} j \left[\sum_{i=i_{j,\ell}}^{i_{j,\ell+1}-1} \mathsf{deg}'_k(T_j[i]) + \mathbf{I}[\mathsf{rdeg}(T_j[i]) = k-1]\right] + \sum_{m=i_\ell}^{i_{\ell+1}-1} \mathbf{I}\left[\sum_{i\geq 1} iB_m[i] = k-1\right].$$

Also observe that $\mathsf{rdeg}(\gamma_\ell) = \sum_{m\geq 1} mB_{i_\ell}[m]$, as for every $m \geq 1$ there are m copies of $B_{i_\ell}[m]$ trees read from $T'_{j,\ell}$ that are attached to the root of γ_ℓ by an edge. So, by using *a)* and $i_{j,1} = B_1[j] + 1$ for $j \geq 2$ we obtain

$$\begin{aligned}\mathsf{deg}'_k(\gamma) &= \sum_{\ell=1}^{B_1[1]} \big(\mathsf{deg}'_k(\gamma_\ell) + \mathbf{I}[\mathsf{rdeg}(\gamma_\ell) = k-1]\big) + \sum_{j\geq 2} j \sum_{i=1}^{B_1[j]} \big(\mathsf{deg}'_k(T_j[i]) + \mathbf{I}[\mathsf{rdeg}(T_j[i]) = k-1]\big) \\ &= \sum_{j\geq 2} j \left[\sum_{i=1}^{N_j} \mathsf{deg}'_k(T_j[i]) + \mathbf{I}[\mathsf{rdeg}(T_j[i]) = k-1]\right] + \sum_{m=2}^{B} \mathbf{I}\left[\sum_{i\geq 1} iB_m[i] = k-1\right].\end{aligned}$$

□

4.2 Proofs of Section 3.3

Proof of Proposition 3.9. By applying Lemma 3.5 we obtain that $N_j = \sum_{m=1}^{B} B_m[j]$. As we have that for $j \geq 2$, $B_m[j]$ is distributed as $\mathsf{Po}\left(H(\rho^j)/j\right)$. Hence, conditioned on the event "$B = b$" we have that N_j is distributed as $\mathsf{Po}\left(bH(\rho^j)/j\right)$, and the probability that $\sum_{j>j_0} N_j > 0$ can then be bounded by

$$\mathbb{P}\left[\mathsf{Po}\left(\sum_{j>j_0} b\frac{H(\rho^j)}{j}\right) > 0\right] = 1 - \exp\left(-b\sum_{j>j_0} \frac{H(\rho^j)}{j}\right) \leq b\sum_{j>j_0} \frac{H(\rho^j)}{j}.$$

Using Proposition 3.2, we see that this expression is $bO(\rho^{j_0})$, and as $\rho < e^{-1}$ and $j_0 \geq \log^2 n$, the proof is finished for large n. □

Proof of Proposition 3.10. Set $\lambda_j = 9b\frac{H(\rho^j)}{j} + t$ and $V'_j = j\sum_{i=1}^{\lambda_j} |T_j[i]|$. In the sequel we will show that $\mathbb{P}\big[V_j > \frac{3}{2}j\lambda_j \;\big|\; B = b\big] \leq e^{-\Omega(t)}$. The statement then follows by summing over all values of j and by applying Proposition 3.2, part *a)*.

Let $y \leq j \leq j_0$. First, note that by applying Lemma 3.5 we obtain that N_j is distributed as $\mathsf{Po}\left(BH(\rho^j)/j\right)$. Lemma 2.2 then yields that $\mathbb{P}[N_j > \lambda_j \mid B = b] \leq e^{-t}$. Furthermore, by setting $N = \lambda_j$ and $\varepsilon = \frac{1}{3}$ in Lemma 3.6 we obtain that there is a $c > 0$ such that for sufficiently large b

$$\mathbb{P}\left[\frac{V'_j}{j} > \frac{4}{3}\frac{\rho^j H'(\rho^j)}{H(\rho^j)}\lambda_j \;\Big|\; B = b\right] \leq e^{-c\frac{\rho^j H'(\rho^j)}{H(\rho^j)}\lambda_j}. \tag{4.1}$$

Note that from Proposition 3.2, part *b)* we have $H'(\rho^j) = 1 + \Theta(\rho^j)$. Hence, for $y \le j \le j_0$ and large enough n, it follows that $\frac{\rho^j H'(\rho^j)}{H(\rho^j)} = 1 + o(1)$, and we deduce with plenty of room to spare that

$$\mathbb{P}\left[\frac{V_j'}{j} > \frac{3}{2}\lambda_j \;\middle|\; B = b\right] \le e^{-\Omega(t)}.$$

With these facts at hand the proof completes routinely as follows. We have

$$\begin{aligned}\mathbb{P}\left[V_j > \frac{3}{2}j\lambda_j \;\middle|\; B = b\right] &\le \mathbb{P}\left[V_j > \frac{3}{2}j\lambda_j \;\middle|\; B = b, N_j \le \lambda_j\right]\mathbb{P}[N_j \le \lambda_j \;|B = b] \;+ e^{-t} \\ &\le \mathbb{P}\left[V_j' > \frac{3}{2}j\lambda_j \;\middle|\; B = b, N_j \le \lambda_j\right] + e^{-t} = e^{-\Omega(t)}.\end{aligned}$$

□

Proof of Lemma 3.11. Let N_j, where $j \ge 2$, be the random variable counting the number of trees from T_j used by $\Gamma\mathcal{H}(\rho, L, T_2, T_3, \dots)$. Furthermore, set $V_j := j\sum_{i=1}^{N_j} |T_j[i]|$. First, by applying Proposition 3.9 with $j_0 = n^{1/4}$ we obtain

$$\mathbb{P}\left[\sum_{j \ge n^{1/4}} V_j > 0 \;\middle|\; B \le n\right] \le e^{-\Omega(n^{1/4})}. \tag{4.2}$$

Moreover, by applying Proposition 3.10 with $y = \log\log n$ and $j_0 = t = n^{1/4}$ we obtain

$$\mathbb{P}\left[\sum_{j=\log\log n}^{n^{1/4}} V_j \le 22B \cdot \rho^y + n^{3/4} \;\middle|\; B \le n\right] \ge 1 - e^{-\Omega(n^{1/4})}. \tag{4.3}$$

Note that $22B \cdot \rho^y + n^{3/4} = o(B + n)$. What remains to handle are the cases $2 \le j \le \log\log n$. By using Lemma 2.2 and Proposition 3.2 we infer that for any $0 < \delta < 1$ it holds that

$$\mathbb{P}\left[\left|N_j - \frac{H(\rho^j)B}{j}\right| \ge \delta\frac{H(\rho^j)B}{j} \;\middle|\; n^{2/3} \le B \le n\right] \le e^{-\Omega(n^{1/2})}.$$

Using the above fact, along with an application of Lemma 3.6 with N satisfying

$$\left(1 - \frac{\varepsilon}{2}\right)\frac{H(\rho^j)B}{j} \le N \le \left(1 + \frac{\varepsilon}{2}\right)\frac{H(\rho^j)B}{j}$$

where $0 < \varepsilon < 1$, we infer again with plenty of room to spare that

$$\mathbb{P}\left[\left|V_j - \rho^j H'(\rho^j)B\right| \ge \varepsilon\rho^j H'(\rho^j)B \;\middle|\; n^{2/3} \le B \le n\right] \le e^{-\Omega(n^{1/2})}. \tag{4.4}$$

Moreover, if $B \le n^{2/3}$ we readily obtain that with probability at most $e^{-\Omega(n^{1/2})}$ that $N_j \ge 2n^{2/3}$, and by again applying Lemma 3.6 with $N = 2n^{2/3}$ we obtain

$$\mathbb{P}\left[V_j \ge 10jn^{2/3} \;\middle|\; B \le n^{2/3}\right] \le e^{-\Omega(n^{1/2})}. \tag{4.5}$$

Since $j \le j_0 = n^{1/4}$, conditioned on the event "$B \le n^{2/3}$" we have that $V_j = o(n^{3/4})$ with probability at least $1 - e^{-\Omega(n^{1/2})}$.

With all those facts at hand we conclude the proof of the statement as follows. Lemma 3.5, statement *b)* yields that

$$|\gamma| = B + \sum_{j\geq 2} V_j.$$

So, if $B \leq n^{2/3}$, we obtain with (4.2), (4.3), and (4.5)

$$\mathbb{P}[|\gamma| = o(n) \mid B \leq n^{2/3}] \geq 1 - e^{-\Omega(n^{1/4})}. \tag{4.6}$$

On the other hand, let $n^{2/3} \leq B \leq n$. Then, by using (4.4) instead of (4.5), we infer that with probability at least $1 - e^{-\Omega(n^{1/4})}$ we have

$$|\gamma| = B + B\sum_{j\geq 2} \rho^j H'(\rho^j) + o(n).$$

As we shall derive later that the sum $\sum_{j\geq 2} \rho^j H'(\rho^j)$ evaluates to $\eta^2/2$. Thus, setting $b = 1/(1+\eta^2/2)$ in the above equation for $|\gamma|$ we readily obtain

$$\mathbb{P}[|\gamma| = b^{-1}B + o(n) \mid n^{2/3} \leq B \leq n] \geq 1 - e^{-\Omega(n^{1/4})}.$$

Note that the observation that $|\gamma| = n$ implies $B \leq n$, which follows from Lemma 3.5, part *b)*. Using this and (4.6) yields the statement of the lemma as

$$\mathbb{P}[B \notin (1 \pm \varepsilon)bn \mid |\gamma| = n] = \mathbb{P}[|\gamma| = n \mid B \notin (1 \pm \varepsilon)bn, B \leq n] \cdot \frac{\mathbb{P}[B \notin (1 \pm \varepsilon)bn, B \leq n]}{\mathbb{P}[|\gamma| = n]} = e^{-\Omega(n^{1/4})},$$

where in the last step we used the trivial estimate $\mathbb{P}[B \notin (1 \pm \varepsilon)bn, B \leq n] \leq 1$ and Lemma 3.4.

Now to complete the proof we will show that the sum $\sum_{j\geq 2} \rho^j H'(\rho^j) = \eta^2/2$. We proceed as follows. First note that by differentiating Equation (3.1) we obtain that $\sum_{j\geq 2} z^j H'(z^j) = zH'(z)[1/H(z)-1]$. From Lemma 3.1 we have

$$H'(z) \overset{z\to\rho}{=} \frac{\eta}{2\rho}\left(1 - \frac{z}{\rho}\right)^{-1/2} + O(1) \quad \text{and} \qquad \frac{1}{H(z)} \overset{z\to\rho}{=} 1 + \eta\left(1 - \frac{z}{\rho}\right)^{1/2} + O\left(1 - \frac{z}{\rho}\right).$$

Hence for $z \to \rho$ the sum $\sum_{j\geq 2} z^j H'(z^j)$ equals $\eta^2/2 + O((1 - z/\rho)^{1/2})$, from which we derive that $\sum_{j\geq 2} \rho^j H'(\rho^j) = \eta^2/2$.

□

Acknowledgement The authors thank Nicla Bernasconi for carefully reading earlier versions of this manuscript.

References

[1] N. Bernasconi, K. Panagiotou, and A. Steger. On properties of random dissections and triangulations. In *SODA '08: Proceedings of the 19th annual ACM-SIAM symposium on Discrete algorithms*, pages 132–141. Society for Industrial and Applied Mathematics, 2008.

[2] N. Bernasconi, K. Panagiotou, and A. Steger. On the Degree Sequences of Random Outerplanar and Series-Parallel Graphs. In *Proceedings of the 11th international workshop, APPROX 2008, and 12th international workshop, RANDOM 2008 on Approximation, Randomization and Combinatorial Optimization*, pages 303–316. Springer-Verlag, 2008.

[3] M. Drmota and B. Gittenberger. The distribution of nodes of given degree in random trees. *Journal of Graph Theory*, 31(3):227–253, 1999.

[4] P. Duchon, P. Flajolet, G. Louchard, and G. Schaeffer. Boltzmann Samplers for the Random Generation of Combinatorial Structures. *Combinatorics Probability and Computing*, 13(4-5):577–625, 2004.

[5] P. Flajolet, É. Fusy, and C. Pivoteau. Boltzmann sampling of Unlabelled Structures. In *Proceedings of the 9th Workshop on Algorithm Engineering and Experiments and the 4th Workshop on Analytic Algorithmics and Combinatorics*, pages 201–211. SIAM Press, 2007.

[6] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[7] B. Gittenberger. Nodes of large degree in random trees and forests. *Random Structures and Algorithms*, 28(3):374–385, 2006.

[8] W. M. Y. Goh and E. Schmutz. Unlabeled trees: Distribution of the Maximum Degree. *Random Structures and Algorithms*, 5(3):411–440, 1994.

[9] R. Otter. The number of trees. *Annals of Mathematics*, 49:583–599, 1948.

[10] K. Panagiotou and A. Weißl. Properties of Random Graphs via Boltzmann Samplers. In *2007 Conference on Analysis of Algorithms, AofA'07*. Springer DMTCS, 2007.

[11] G. Pólya. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math*, 68:145–254, 1937.

[12] R. W. Robinson and A. J. Schwenk. The distribution of degrees in a large random tree. *Discrete Mathematics*, 12:359–372, 1975.

[13] A. J. Schwenk. An asymptotic evaluation of the cycle index of a symmetric group. *Discrete Mathematics*, 18:71–78, 1977.