REPORTS@SCM

AN ELECTRONIC JOURNAL OF THE
SOCIETAT CATALANA DE MATEMÀTIQUES

# Generating uniform spanning trees from conditioned Bienaymé–Galton–Watson trees

**Albert Vives**
Universitat Politècnica de Catalunya
*avivesbatalla@gmail.com*

**Resum** *(CAT)*
Aquest article explora la generació d'arbres generadors uniformes (UST), fonamentals en combinatòria i probabilitat, amb aplicacions en teoria de xarxes i física. Utilitzant processos de Bienaymé–Galton–Watson (BGW) condicionats a un nombre fix de vèrtexs, s'introdueix un mètode per generar arbres generadors uniformes i s'examinen propietats estructurals com l'alçada i l'amplada.

**Abstract** *(ENG)*
This report explores uniform spanning tree (UST) generation, essential in combinatorics and probability with applications in network theory and physics. Using conditioned Bienaymé–Galton–Watson (BGW) processes, it introduces a method to generate USTs. Rigorous proofs show that conditioning on a fixed number of vertices ensures uniform distribution and let us examine structural properties like height and width.

Societat Catalana de Matemàtiques

Institut d'Estudis Catalans

**Keywords:** *random trees, Bienaymé–Galton–Watson trees.*
**MSC (2020):** *Primary 05C05. Secondary 60J80.*

**Received:** *February 17, 2025.*
**Accepted:** *October 2, 2025.*

# 1. Introduction

The generation of uniform spanning trees (USTs) is a crucial problem in combinatorics and stochastic processes, with applications in network analysis, random structures, and statistical physics. A spanning tree is a connected subgraph of a graph that includes all vertices without forming cycles. When sampled uniformly at random, each spanning tree has an equal probability of being chosen, presenting mathematical challenges in defining an appropriate random model susceptible of being analyzed. This drives the search for effective approaches, and this report proposes novel methods using conditioned Bienaymé–Galton–Watson (BGW) processes. These processes, traditionally used to model population growth, are adapted to generate spanning trees under uniform distributions, bridging combinatorics with stochastic processes.

This work provides detailed mathematical formulations and rigorous proofs demonstrating that conditioning BGW processes on a fixed number of vertices results in uniform spanning trees. It explores classical distributions such as Poisson, Geometric, and Bernoulli branching to illustrate the method's versatility and correctness. It also delves into structural characteristics of random trees, analyzing height and width as primary metrics for complexity. Results on these parameters offer insights into the asymptotic behavior of USTs, supported by probabilistic bounds and examples. These findings contribute to a deeper understanding of USTs and their generation through probabilistic techniques.

# 2. Obtaining uniform spanning trees from BGW trees

Depending on the probability distribution governing the reproduction of some individuals, one may obtain different types of trees. Specifically, we aim to prove that if we condition a Bienaymé–Galton–Watson tree to have $n$ vertices and fix certain known distributions, we obtain various types of uniform random trees. The first description of random trees from conditioning a BGW process by its total progeny can be traced back to Kolchin [3] and Aldous [2].

**Lemma 2.1.** *Let $Z$ be a nonnegative integer-valued random variable and let $X$ be a* BGW($Z$) *process. Let $\mathcal{T}_n$ denote the class of trees with $n$ vertices which can be generated by the process and let $T \in \mathcal{T}_n$ be a tree in the class. If the probability that $T$ is generated by the process depends only on the number of vertices $n$, then*

$$\mathbb{P}(X = T||X| = n) = \frac{1}{|\mathcal{T}_n|}.$$

*Proof.* Let us denote $\mathbb{P}(X = T)$ by $f(n)$, where $n$ is the number of vertices of $T$. Then,

$$\mathbb{P}(|X| = n) = |\mathcal{T}_n| f(n),$$

since the last probability is the sum of the probabilities of obtaining each of the trees with $n$ vertices. Now, when conditioned on having $n$ vertices, all possible trees are equally likely to occur:

$$\mathbb{P}(X = T \mid |X| = n) = \frac{\mathbb{P}(\{X = T\} \cap \{|X| = n\})}{\mathbb{P}(|X| = n)} = \frac{f(n)}{|\mathcal{T}_n| f(n)} = \frac{1}{|\mathcal{T}_n|}.$$ □

We now aim to prove that, by selecting an appropriate offspring distribution, it is possible to generate several well-known types of trees.

The class of labeled trees with $n$ nodes is also called the class of Cayley trees, due to the Cayley formula enumerating them, its number being $n^{n-2}$. While Bienaymé–Galton–Watson trees are naturally rooted, Cayley trees are not; in this context we consider Cayley trees to be rooted by fixing one distinguished vertex as the root.

**Theorem 2.2.** *Conditioning a Bienaymé–Galton–Watson tree with offspring distribution* Poisson(1) *on having $n$ vertices results in a Cayley tree with $n$ vertices generated uniformly at random.*

*Proof.* Every rooted Cayley tree can be generated from a Poisson(1) BGW process by a labeling of its vertices. Let $T$ be a specific Cayley tree with $n$ vertices. To prove that each tree can be obtained uniformly, we first need to order all sibling sets in $T$ by increasing vertex labels. Let $\chi_1, \ldots, \chi_n$ represent the number of children of each node, where the vertices are indexed starting from the root and then recursively visiting the children from left to right. The first requirement for generating $T$ is ensuring the correct number of descendants for each vertex. Since these random variables are mutually independent, the probability of obtaining a specific number of children for all vertices is the product of their individual probabilities.

The second requirement is assigning the correct labeling, as we are considering labeled trees. With $n$ vertices, there are $n!$ possible ways to label them. Moreover, the children of the $i$-th vertex can be permuted in $\chi_i!$ distinct ways for each $i = 1, \ldots, n$, resulting in the same tree. Thus, the final calculation can be expressed clearly as follows:

$$\mathbb{P}(X = T) = \left( \prod_{i=1}^{n} \frac{1}{\chi_i!} \cdot e^{-1} \right) \cdot \frac{\prod_{i=1}^{n} \chi_i!}{n!} = e^{-n} \frac{1}{n!}.$$

Since the last probability depends only on the number of vertices, and it is known that there exist $n^{n-2}$ labeled trees, by Lemma 2.1 the theorem is proved. $\square$

The class $\mathcal{B}_n$ of full binary trees of $n$ vertices is the family of unlabelled rooted plane trees where every node has two or zero children. Being plane means that trees have distinguished left and right subtrees. The trees depicted in Figure 1 are considered to be distinct.

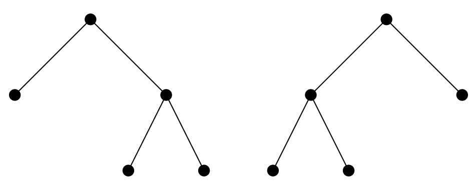

Figure 1: Two distinct binary trees.

A full binary tree with $n$ nodes has an odd number $n$ of vertices and $m = (n+1)/2$ leaves. The number of such trees is given by the Catalan number $C_{m-1}$.

**Theorem 2.3.** *Conditioning a* $2\,$Bernoulli$(1/2)$ *Bienaymé–Galton–Watson tree on having $n$ vertices results in a binary tree with $n$ vertices generated uniformly at random.*

*Proof.* It is clear that one can obtain every binary tree from a $2\,\mathrm{Be}(1/2)$ BGW tree and vice versa. Let $T$ be a particular binary tree with $n$ vertices. It is known that in a binary tree there are $(n-1)/2$ internal nodes, and $(n+1)/2$ leafs. To apply Lemma 2.1, we need to calculate $\mathbb{P}(X = T)$. Let $\chi \sim 2\,\mathrm{Be}(1/2)$.

$$\mathbb{P}(X = T) = \mathbb{P}(\chi = 0)^{\frac{n+1}{2}}\mathbb{P}(\chi = 2)^{\frac{n-1}{2}} = \left(\frac{1}{2}\right)^n.$$

Since the last probability depends just on the number of vertices, by Lemma 2.1 we have proved the theorem. □

The class $\mathcal{P}_n$ of ordered plane trees with $n$ vertices is the family of unlabelled rooted trees where the children of every node are ordered from left to right in the plane.

**Theorem 2.4.** *Conditioning a* $\mathrm{Geom}(1/2)$ *Bienaymé–Galton–Watson tree on having n vertices results in an ordered plane tree with n vertices generated uniformly at random, where* $\mathrm{Geom}(1/2)$ *denotes the geometric distribution on* $\{0, 1, 2, \dots\}$.

*Proof.* It can be observed that every ordered plane tree can be obtained from a $\mathrm{Geom}(1/2)$ BGW tree, and conversely, every $\mathrm{Geom}(1/2)$ BGW tree corresponds to an ordered plane tree. Let $T$ be a specific ordered plane tree with $n$ vertices. To generate such a tree, it is necessary to account for both internal nodes and leaves.

The probability of an internal node having exactly $k$ children is $(1/2)^k(1/2)$, where the first factor represents the probability of successfully having $k$ children, and the second factor accounts for the probability of no additional children. For a leaf, the requirement is simply to have no children, which occurs with probability $1/2$.

Consequently, the probability of achieving the correct number of children for each vertex is independent of whether the vertex is an internal node or a leaf. This probability is given by $(1/2)^{\chi_i+1}$, where $\chi_i$ denotes the number of children of vertex $i$. Thus, the following conclusion naturally arises:

$$\mathbb{P}(\mathcal{T} = T) = \prod_{i=1}^{n}\left(\frac{1}{2}\right)^{\chi_i+1} = \left(\frac{1}{2}\right)^{\sum_{i=1}^{n}\chi_i+1} = \left(\frac{1}{2}\right)^{n-1+n} = \left(\frac{1}{2}\right)^{2n-1}.$$

Furthermore, it is usually known that the number of ordered plane trees with $n$ vertices is $C_{n-1}$, where $C_n$ is the $n$-th Catalan number. Hence, by Lemma 2.1, the theorem has been proved. □

# 3. Studying some parameters of random trees

This chapter is focused on the study of key parameters of random trees: height, width, and the number of leaves. All results presented in this section are derived from the study L. Addario-Berry, L. Devroye, and S. Janson did in [1]. The approach in [1] applies not only to random Cayley trees, but to any family of trees arising from a $\mathrm{BGW}(\chi)$ tree as long $\chi$ has expectation 1 and finite variance (critical BGW trees). Furthermore, explicit proofs are provided for certain concepts that are often assumed without further justification in the literature. The exploration of these parameters offers a deeper understanding of random trees, contributing new insights into established results in this field.

## 3.1. Some preliminaries

The Breadth-First Search (BFS) on a BGW tree is an algorithm used to explore the tree level by level, starting from the root. BFS explores the tree by visiting all nodes at one level before moving to the next one, ensuring that nodes are processed in increasing distance from the root. Hence, this search keeps a queue $Q$ with $Q_i$ nodes at the $i$-th step, with $Q_0 = 1$. During the exploration of a vertex, its offspring are added to the back of the queue. Then, one can easily obtain the following recursion:

$$Q_i = Q_{i-1} - 1 + \chi_i,$$

where $\chi_i$ are independent and identically distributed copies of the offspring distribution $\chi$. Hence, by this recursion, $Q_j = 1 + \tilde{S}_j$, where $\tilde{S}_j := \sum_{i=1}^{j}(\chi_i - 1) = S_j - j$. The tree is completely explored when $Q_j = 0$. In this case, $\tilde{S}_n = -1$.

**Definition 3.1.** Let $T_n$ be a random tree with $n$ nodes. The width of a tree $T_n$, denoted by $W(T_n)$, is the maximum number of nodes at any depth level. Generally, let $d(v)$ denote the depth of a node $v$ in a rooted tree $T$. Then

$$\mathrm{width}(T_n) = \max_{k\geq 0} |\{v \in V(T_n) : d(v) = k\}|.$$

We now present key lemmas that are necessary for studying the expected width of a random tree.

**Lemma 3.2** (Raney's Lemma)**.** *Let $a_1, a_2, \dots, a_n$ be a sequence of integers such that $\sum_{i=1}^{n} a_i = -1$.*

*Then there exists a unique index $s$ such that the cyclic partial sums $S_k = \sum_{j=0}^{k-1} a_{(s+j) \bmod n}$ for $k = 1, 2, \dots, n$, satisfy:*

*(i) $S_k > 0$ for $1 \leq k < n$.*

*(ii) $S_n = -1$.*

**Lemma 3.3.** *Suppose that the individuals in a BGW process reproduce according to a random variable $\chi$, with $\mathbb{E}[\chi] = 1$ and $\mathrm{Var}(\chi) < \infty$. Then, there is a constant $c_1 \in \mathbb{R}$ such that, for all $n$ sufficiently large,*

$$\mathbb{P}(\tilde{S}_n = -1) \geq c_1 n^{-1/2}.$$

**Lemma 3.4.** *Suppose that $\chi_i$ are i.i.d., non-negative and integer-valued random variables, with $\mathbb{E}[\chi_i] = 1$ and $\mathrm{Var}[\chi_i] < \infty$, and let $S_n = \sum_{i=1}^{n} \chi_i$. Then, for all $n \geq 1$ and $m \geq 0$,*

$$\mathbb{P}(S_n = n - m) \leq \frac{c_2}{\sqrt{n}} e^{-c_3 m^2/n},$$

*where $c_2 > 0$ and $c_3 > 0$ are real constants.*

**Lemma 3.5.** *Let $\chi$ be a discrete random variable taking values in nonnegative integers. Suppose that $\mathbb{E}(\chi) = 1$ and $0 < \mathrm{Var}(\chi) < \infty$. Let $\mathcal{T}$ be a $\mathrm{BGW}(\chi)$ tree. Then, $\mathbb{P}(|\mathcal{T}| = n) \geq n^{-3/2}$.*

**Lemma 3.6** (Bernstein inequality)**.** *Let $X_1, X_2, \dots, X_n$ be independent random variables such that $X_i - \mathbb{E}[X_i] \leq b$ for every $i$, where $b \in \mathbb{R}$. Let $V := \sum_{i=1}^{n} \mathrm{Var}(X_i)$. Then,*

$$\mathbb{P}\left(\sum_{i=1}^{n}(X_i - \mathbb{E}X_i) \geq t\right) \leq \exp\left(-\frac{t^2}{2V + \frac{2bt}{3}}\right).$$

## 3.2. The width

**Theorem 3.7.** *Let $\chi$ be a discrete random variable taking values in nonnegative integers. Suppose that $\mathbb{E}[\chi] = 1$ and $0 < \mathrm{Var}(\chi) < \infty$. Let $T_n$ be a $\mathrm{BGW}(\chi)$ tree with $n$ nodes. Then,*

$$\mathbb{P}(W(T_n) \geq x) \leq c_4 e^{-c_5 x^2/n},$$

*for all $x \geq 0$ and $n \geq 1$, where $c_4 > 0$ and $c_5 > 0$.*

*Proof.* We will follow the proof strategy outlined in [1]. Let $Z_k$ be the number of individuals in the $k$-th level of a BGW tree. It is clear that every $Z_k$ is some $Q_j$, where $Q_j$ is the number of nodes in the queue of BFS in the $j$-th iteration. Hence,

$$W := \max_{k \geq 0} Z_k \leq \max_{j \geq 0} Q_j.$$

As a result, for the conditioned BGW tree $T_n$,

$$\mathbb{P}(W \geq x+1) \leq \mathbb{P}(\max_j Q_j \geq x+1) = \mathbb{P}(\max_j \tilde{S}_j \geq x \mid \tilde{S}_j \geq 0,\ j < n,\ \tilde{S}_n = -1).$$

Now, we aim to simplify the last conditioned probability. Our first goal is to get rid of the conditioning on $\tilde{S}_j \geq 0$, where $j < n$. By Raney's Lemma 3.2, conditioning on $\{\tilde{S}_j \geq 0,\ j < n,\ \tilde{S}_n = -1\}$ is equivalent to conditioning only on $\{\tilde{S}_n = -1\}$. However, $\max_j \tilde{S}_j$ may be changed. By conditioning on $\tilde{S}_n = -1$, we can write

$$\max_{j \leq n} \tilde{S}_j = \max_{j \leq n} \tilde{S}_j - \min_{j \leq n} \tilde{S}_j + 1,$$

and the latter quantity is changed by at most 1 by a rotation of $\tilde{\chi}_i := \chi_i - 1\ \forall i = 1, \dots, n$. Then

$$\mathbb{P}(\max_j \tilde{S}_j \geq x \mid \tilde{S}_j \geq 0,\ j < n,\ \tilde{S}_n = -1) \leq \mathbb{P}(\max_{j \leq n} \tilde{S}_j - \min_{j \leq n} \tilde{S}_j \geq x \mid \tilde{S}_n = -1).$$

Therefore, we now have more tools to bound $\mathbb{P}(W \geq x+1)$.

$$\begin{aligned}\mathbb{P}(W \geq 2x+2) &\leq \mathbb{P}(\max_j Q_j \geq 2x+2) \leq \mathbb{P}(\max_{j \leq n} \tilde{S}_j - \min_{j \leq n} \tilde{S}_j \geq 2x+1 \mid \tilde{S}_n = -1) \\ &\leq \mathbb{P}(\max_{j \leq n} \tilde{S}_j \geq x \mid \tilde{S}_n = -1) + \mathbb{P}(\min_{j \leq n} \tilde{S}_j \leq -x-1 \mid \tilde{S}_n = -1).\end{aligned}$$

The last inequality is due to the fact that, if $y - z \geq 2x+1$, then either $y \geq x$ or $z \leq -x-1$, where $x$, $y$, $z$ are real numbers. Furthermore, the reflection $\chi_i \leftrightarrow \chi_{n+1-i}$, which swaps $\tilde{S}_j \leftrightarrow -\tilde{S}_n - \tilde{S_{n-j}}$, shows that the last probabilities are the same. Hence,

$$\mathbb{P}(\max_j Q_j \geq 2x+2) \leq 2\,\mathbb{P}(\max_{j \leq n} \tilde{S}_j \geq x \mid \tilde{S}_n = -1).$$

The last expression can be written in terms of the first index such that $\tilde{S}_j \geq x$. So that, let us define $\tau = \min\{j \geq 0 : \tilde{S}_j \geq x\}$. Then,

$$\begin{aligned}\mathbb{P}(\max_j Q_j \geq 2x+2) &\leq 2\,\mathbb{P}(\tau < n \mid \tilde{S}_n = -1) \\ &= 2\,\mathbb{P}(\tilde{S}_n = -1 \mid \tau < n)\frac{\mathbb{P}(\tau < n)}{\mathbb{P}(\tilde{S}_n = -1)},\end{aligned}$$

where the last equality is due to the definition of conditioned probability. By definition of $\tau$, $\tilde{S}_\tau \geq x$. Then, if we fix some $t < n$ and $y \geq x$, by Lemma 3.4 we have the following:

$$\begin{aligned}\mathbb{P}(\tilde{S}_n = -1 \mid \tau < n) &= \mathbb{P}(\tilde{S}_n = -1 \mid \tau = t, \tilde{S}_\tau = y)\\ &= \mathbb{P}(\tilde{S}_n - \tilde{S}_t = -y-1) = \mathbb{P}(\tilde{S}_{n-t} = -(y+1)) = \mathbb{P}(S_{n-t} - (n-t) = -(y+1))\\ &= \mathbb{P}(S_{n-t} = (n-t) - (y+1)) \leq c_2 n^{-1/2} e^{-c_3(y+1)^2/(n-t)}\\ &\leq c_2 n^{-1/2} e^{-c_3 x^2/n},\end{aligned} \tag{1}$$

where we have used $\tilde{S}_n - \tilde{S}_t = \tilde{S}_{n-t}$ due to the fact that these $n-t$ random variables are i.i.d. Then, by (1) and Lemma 3.3 it is clear that we can now prove what we seek:

$$\mathbb{P}(\max_j Q_j \geq 2x+2) \leq c_2 n^{-1/2} e^{-c_3 x^2/n} \cdot \frac{\mathbb{P}(\tau < n)}{\mathbb{P}(\tilde{S}_n = -1)} \leq \frac{c_5 n^{-1/2} e^{-c_7 x^2/n}}{\mathbb{P}(\tilde{S}_n = -1)} \leq c_4 e^{-c_5 x^2/n}. \qquad \square$$

**Theorem 3.8.** *Let $\chi$ be a discrete random variable taking values in nonnegative integers. Suppose that $\mathbb{E}[\chi] = 1$ and $0 < \mathrm{Var}(\chi) < \infty$. Let $T_n$ be a $\mathrm{BGW}(\chi)$ tree with $n$ nodes. Then,*

$$\mathbb{E}[W(T_n)] = \mathcal{O}(\sqrt{n}),$$

*for all $n \geq 1$.*

*Proof.* This proof is not provided in [1]; however, we consider it highly relevant to the topic at hand. Therefore, we present a rigorous proof of this result herein. We aim to prove that the expected value $\mathbb{E}[W(T_n)]$ grows asymptotically at most as $\sqrt{n}$, given the inequality

$$\mathbb{P}(W(T_n) \geq w) \leq c_4 e^{-c_5 w^2/n}.$$

For a discrete random variable $W$, the expectation can be written as

$$\mathbb{E}[W(T_n)] = \sum_{w=0}^{\infty} w \cdot \mathbb{P}(W(T_n) = w),$$

or equivalently

$$\mathbb{E}[W(T_n)] = \sum_{w=1}^{\infty} \mathbb{P}(W(T_n) \geq w).$$

This equivalence holds because

$$\mathbb{P}(W(T_n) \geq w) = \sum_{k=w}^{\infty} \mathbb{P}(W(T_n) = k),$$

and reorganizing terms yields the alternative representation.

Using the given inequality $\mathbb{P}(W(T_n) \geq w) \leq c_4 e^{-c_5 w^2/n}$, we can bound the expectation as

$$\mathbb{E}[W(T_n)] \leq \sum_{w=1}^{\infty} c_4 e^{-c_5 w^2/n} = c_4 \sum_{w=1}^{\infty} e^{-c_5 w^2/n}.$$

The sum $\sum_{w=1}^{\infty} e^{-c_5 w^2/n}$ resembles a Riemann sum and can be approximated by an integral for large $n$. We have

$$\sum_{w=1}^{\infty} e^{-c_5 w^2/n} \leq \int_0^{\infty} e^{-c_5 x^2/n}\, dx.$$

To compute this integral, we use the substitution $u = \sqrt{\frac{c_5}{n}}x$, which implies $x = \sqrt{\frac{n}{c_5}}u$ and $dx = \sqrt{\frac{n}{c_5}}\, du$. Substituting into the integral gives

$$\int_0^{\infty} e^{-c_5 x^2/n}\, dx = \sqrt{\frac{n}{c_5}} \int_0^{\infty} e^{-u^2}\, du = \sqrt{\frac{n}{c_5}} \cdot \frac{\sqrt{\pi}}{2}.$$

Therefore,

$$\int_0^{\infty} e^{-c_5 x^2/n}\, dx = \sqrt{\frac{n}{c_5}} \cdot \frac{\sqrt{\pi}}{2}.$$

Thus, the sum can be approximated as

$$\sum_{w=1}^{\infty} e^{-c_5 w^2/n} \sim \sqrt{\frac{n}{c_5}} \cdot \frac{\sqrt{\pi}}{2}.$$

Substituting this back into the bound for $\mathbb{E}[W(T_n)]$, we find

$$\mathbb{E}[W(T_n)] \leq c_4 \cdot \sqrt{\frac{n}{c_5}} \cdot \frac{\sqrt{\pi}}{2}.$$

This shows that

$$\mathbb{E}[W(T_n)] = \mathcal{O}(\sqrt{n}).$$

□

## 3.3. The height

**Definition 3.9.** The height of a tree $T$, denoted by $H(T)$, is the maximum depth of any node in the tree.

$$\mathrm{H}(T) = \max_{v \in V(T)} d(v).$$

A lexicographic depth-first search (DFS) is a linear time algorithm for ordering the vertices of a labelled graph. At each node, its children are visited in lexicographical order. The children of the first children are explored before going to its siblings. This idea is applied recursively for each vertex. Let $Q_i^d$ be the number of nodes in the stack of the DFS at the $i$-th step. We define $Q_0^d = 1$. At each step, we get rid of a vertex from the queue and add its children, which we read in the lexicographical order. Hence, since all individuals reproduce with the same probability distribution, the following recursion is clear:

$$Q_i^d = Q_{i-1}^d - 1 + \chi_i.$$

The reverse-lexicographic depth-first search is a variation of the standard depth-first search (DFS) where, instead of visiting the children of a node in lexicographical order (smallest first), we visit the children in reverse lexicographical order (largest first). We will denote by $Q_i^r$ the number of nodes in the queue of the algorithm at the $i$-th step.

One can easily observe that in a uniformly random tree $T_n$, the labels of the vertices do not affect the probability distribution of the labeled tree. While BFS explores nodes level by level and DFS goes deeper first, lexicographic and reverse-lexicographic DFS are merely different ways of ordering the visits to the nodes. Therefore, the resulting labelings observed along these procedures follow the same distribution.

Now, we will stand out several preliminary lemmas:

**Lemma 3.10.** *Let $P$ be the unique path in a tree from a vertex which has height $h$ to the root. The expected value of nodes in $P$ that have more than one child is $h \cdot q_1$, where $q_1 = 1 - \mathbb{P}(\chi = 1)$.*

By the previous lemma, the expected number of nodes in $P$ that have exactly one child is $h \cdot p_1$, where $p_1 = \mathbb{P}(\chi = 1)$.

## 3.4. A modified Bienaymé–Galton–Watson tree

To complete the study of the height of a random tree, we will need to introduce a new concept.

Let $\hat{\chi}$ be a random variable with the size-biased distribution

$$\mathbb{P}(\hat{\chi} = m) = m \cdot \mathbb{P}(\chi = m).$$

It is clear that this is a probability distribution since $\sum_m \mathbb{P}(\tilde{\chi} = m) = \mathbb{E}[\chi] = 1$, and that $\hat{\chi} \geq 1$.

Let, for $k \geq 1$, $T^{(k)}$ be the modified BGW defined as follows. It has two different types of nodes: normal and mutant. Normal constitute the offspring of $\chi$, while mutant nodes have offspring according to independent copies of $\hat{\chi}$. All children of normal nodes are also normal. Exactly one child of each mutant node is chosen at random and it is called its heir. If this heir has depth less than $k$, it is mutant. If not, it is normal. Hence, there are exactly $k$ mutant nodes, which together with the heir $v^*$ of the last mutant node, form a path from the root to $v^*$ at depth $k$. This path is what we call the spine $\gamma$ of $\hat{T}^{(k)}$.

An equivalent construction of the modified BGW tree can be described as follows:

1. **Construct the spine:**
   - The spine is a path consisting of $k$ nodes, starting from the root at depth 0 and ending at a node at depth $k$.
   - These nodes correspond to the mutant nodes in the original construction.
2. **Attach independent subtrees to the nodes of the spine:**
   - For each node along the spine, except the last one at depth $k$, attach a number of independent subtrees. The number of such subtrees is distributed as $\hat{\chi} - 1$, meaning it follows the distribution $\hat{\chi}$ reduced by 1.
   - For the node of the spine at depth $k$, attach independent subtrees according to the distribution $\chi$, which corresponds to the behavior of normal nodes.

Hence, the spine forms the central path from the root to depth $k$, while the subtrees, attached independently, reflect the probabilistic structure of the original model.

Now, we aim to study the probability of obtaining a particular tree $T$ from a modified BGW tree $\hat{T}^{(k)}$. To start with it, it is not difficult to find the probability that a given mutant node has $m$ children and one of them is selected as heir. This is:

$$\frac{1}{m}\mathbb{P}(\hat{\chi} = m) = \mathbb{P}(\chi = m).$$

Hence, since every normal node has distribution $\chi$ and the process consisting of taking a mutant node with $m$ children and selecting one of them to be mutant also distributes as $\chi$, the following is clear:

$$\mathbb{P}(\hat{T}^{(k)} = T \text{ with } \gamma \text{ as spine}) = \prod_v \mathbb{P}(\chi = d_v) = \mathbb{P}(X = T).$$

Then, the probability of getting a particular tree $T$ from a modified BGW with a fixed spine is the same of obtaining this $T$ from the BGW tree. Since for every node $v$ at depth $k$, there exists one unique path from $v$ to the root, there is a bijection between nodes at depth $k$ and spines. Therefore, summing for all nodes at depth $k$:

$$\mathbb{P}(\hat{\mathcal{T}}^{(k)} = T) = Z_k(T) \cdot \mathbb{P}(\mathcal{T} = T).$$

In conclusion, $\hat{\mathcal{T}}^{(k)}$ has the distribution of $\mathcal{T}$ biased by $Z_k$, which is the size of the $k$-th generation.

Returning to the broader discussion, with these preliminaries in place, we now have the necessary tools to prove the following.

**Theorem 3.11.** *Let $\chi$ be a discrete random variable taking values in nonnegative integers. Suppose that $\mathbb{E}[\chi] = 1$ and $0 < \mathrm{Var}(\chi) < \infty$. Let $T_n$ be a BGW($\chi$) tree conditional on having $n$ nodes. Then, for all $n \geq 1$ and $h \geq \sqrt{n}$,*

$$\mathbb{P}(H(T_n) \geq h) \leq C_2 e^{-c_2 h^2/n},$$

*where $C_2 > 0$ and $c_2 > 0$.*

*Proof.* We will follow the proof approach provided in [1]. Let $h$ be the height of $T_n$. We may assume that $h \in \mathbb{Z}$. We will base the proof on the next observation. Since we have proved that the width is expected to be $\sqrt{n}$, we can suppose there is a vertex $v \in V(T_n)$ with "large" height. Hence, there are two possible cases of obtaining a tree with height $h$: either there are many edges leaving the unique path $P$ from $v$ to the root, where $v$ is a vertex from the $h$-th level, or there are many of the ancestors of $v$ with just one child.

In the first case, the objective is to bound $\max(Q_j^d, Q_k^r)$ and connect this bound with the probabilistic structure of the tree. The quantities $Q_j^d$ and $Q_k^r$ measure how many vertices are simultaneously active during the lexicographic and reverse-lexicographic depth-first explorations, respectively. Whenever a node on the path from a vertex at height $h$ to the root has more than one child, all of its extra children are added to the exploration data structure and remain there until they are processed. This causes $Q_j^d$ or $Q_k^r$ to increase, and their maximum size thus reflects the cumulative effect of branching along the path. Therefore, the height $h$ of the tree can be directly related to $\max(Q_j^d, Q_k^r)$: if the tree has height $h$, then there must exist an index $j$ with $Q_j^d = h$ or an index $k$ with $Q_k^r = h$.

Let $p_1 = \mathbb{P}(\chi = 1)$ and let $q_1 = 1 - p_1$. Let $v \in V(T_n)$ such that $h(v) = h$. Let $j$ (resp. $k$) be the index of $v$ in lexicographic (resp. reverse-lexicographic) order. Let $X$ be the number of nodes which have more than one child in $P$. Each ancestor of $v$ with more than one child contributes to at least one unit to $Q_j^d$ or to $Q_k^r$. We distinguish two cases, either $\max(Q_j^d, Q_k^r) \geq \frac{q_1}{3}h$ or $\max(Q_j^d, Q_k^r) < \frac{q_1}{3}h$. In the second case, by the above remark, the number of ancestors of $v$ with exactly one child is at least $(1-2q_1/3)h = (p_1+q_1/3)h$.

Now, it is not useful to think about the queues because when the algorithm processes a node which has just one child, the size of the queue does not increase, so it is not a good representation for the height of the tree. Let $S(h)$ be the set of trees $T$ with $|T| = n$ and containing a node $v$ such that $h(v) = h$ that has at least $(p_1 + q_1/3)h$ ancestors in $P$ with exactly one child. Then, let $\alpha := \{T_n \in S\} = \bigcup_{T\in S}\{T_n = T\}$.

Then, we can apply these two cases described to the calculus of the following probability. The first two terms correspond to the first case, while the remaining term corresponds to the second case:

$$\begin{aligned}\mathbb{P}(H(T_n) \geq h) &\leq \mathbb{P}\left(\max_j Q_j^d \geq \frac{q_1}{3}h\right) + \mathbb{P}\left(\max_k Q_k^r \geq \frac{q_1}{3}h\right) + \mathbb{P}(\alpha)\\ &= 2\mathbb{P}\left(\max_i Q_i \geq \frac{q_1}{3}h\right) + \mathbb{P}(\alpha) \leq \; C_{11}e^{-c_{11}h^2/n} + \mathbb{P}(\alpha),\end{aligned}$$

where the last inequality has been seen in the proof of Theorem 3.7.

Then, we only need to bound $\mathbb{P}(\alpha)$:

$$\begin{aligned}\mathbb{P}(\mathcal{T} \in S) &= \sum_{T\in S}\mathbb{P}(\mathcal{T} = T) = \sum_{T\in S}\mathbb{P}(\hat{T}^{(h)} = T \text{ with } \gamma_T \text{ as spine})\\ &= \mathbb{P}\left(\bigcup_{T\in S}\{\hat{T}^{(h)} = T \text{ with } \gamma_T \text{ as spine}\}\right) \leq \mathbb{P}\left(\sum_{i=0}^{h-1}\mathbf{1}_{\hat{\chi}_i=1} \geq (p_1+q_1/3)h\right).\end{aligned}$$

To justify the last inequality, recall the construction of $\hat{T}^{(h)}$: the spine $\gamma$ consists of the $h$ mutant nodes from the root to depth $h$, and for each $i = 0, \ldots, h-1$ the number of offspring of the $i$-th mutant is distributed as $\hat{\chi}_i$ (the size-biased distribution), one of whose children is then chosen uniformly to continue the spine.

There is a one-to-one correspondence between the property "the $i$-th vertex on the spine has exactly one child in the realised tree $T$" and the event "$\hat{\chi}_i = 1$ in the construction of $\hat{T}^{(h)}$": if the $i$-th mutant has exactly one child, then necessarily that child is the heir (so $\hat{\chi}_i = 1$), and conversely, if $\hat{\chi}_i = 1$, then the $i$-th spine vertex has exactly one child in the realised tree.

Consequently, whenever $\hat{T}^{(h)}$ equals some $T \in S$ with spine $\gamma_T$, the number of spine vertices having exactly one child is at least $(p_1 + q_1/3)h$ by the definition of $S$. Therefore the event

$$\bigcup_{T\in\mathcal{S}}\{\hat{T}^{(h)} = T \text{ with } \gamma_T \text{ as spine}\}$$

is contained in the event

$$\left\{\sum_{i=0}^{h-1}\mathbf{1}_{\{\hat{\chi}_i=1\}} \geq (p_1+q_1/3)h\right\},$$

which yields the displayed inequality.

Since $\mathbf{1}_{\hat{\chi}_i=1}$ are Bernoulli($p_1$), by Lemma 3.6, and taking $t = q_1 h/3$, $V = p_1 q_1 h$, $b = q_1$, we have the following bound:

$$\mathbb{P}\left(\sum_{i=0}^{h-1} \mathbf{1}_{\hat{\chi}_i=1} \geq (p_1 + q_1/3)h\right) \leq \exp\left(-\frac{(q_1 h/3)^2}{2p_1 q_1 h + 2q_1^2 h/9}\right) = \exp\left(-\frac{h}{18p_1/q_1 + 2}\right).$$

Furthermore, the following equality is clear:

$$\mathbb{P}(\mathcal{T} \in S) = \mathbb{P}(\alpha) \cdot \mathbb{P}(|\mathcal{T}| = n).$$

By Lemma 3.5, $\mathbb{P}(|\mathcal{T}| = n) \geq n^{-3/2}$ and we can establish an upper bound for $\mathbb{P}(\delta)$.

$$\mathbb{P}(\delta) = \frac{P(\mathcal{T} \in S)}{P(|\mathcal{T}| = n)} \leq C_{12} n^{3/2} \exp\left(-\frac{h}{18p_1/q_1 + 2}\right) \leq C_{12} n^{3/2} \exp\left(\frac{h}{18p_1/q_1 + 2}\frac{h}{\sqrt{n}}\right) \leq C_{13} e^{-c_{12} h^2/n},$$

for all $h \geq \sqrt{n}$. Taking everything into account, what we have is the following bound for the height of a BGW tree:

$$\mathbb{P}(H(\mathcal{T}_n) \geq h) \leq C_{11} e^{-c_{11} h^2/n} + C_{13} e^{-c_{12} h^2/n}.$$

Now, taking

$$C_2 := \max\{C_{11}, C_{13}\}, \quad c_2 := \min\{c_{11}, c_{12}\} \Longrightarrow \mathbb{P}(H(\mathcal{T}_n) \geq h) \leq C_2 e^{-c_2 h^2/n}. \qquad \square$$

Having proved that, we left the following result without proof due to its analogy with Theorem 3.8.

**Theorem 3.12.** *Let $\chi$ be a discrete random variable taking values in nonnegative integers. Let $T_n$ be a* BGW($\chi$) *tree conditional on having $n$ nodes. Suppose that $\mathbb{E}[\chi] = 1$ and $0 < \mathrm{Var}(\chi) < \infty$. Then, for all $n \geq 1$,*

$$\mathbb{E}[H(T_n)] = \mathcal{O}(\sqrt{n}).$$

# References


[1] L. Addario-Berry, L. Devroye, S. Janson, Sub-Gaussian tail bounds for the width and height of conditioned Galton–Watson trees, *Ann. Probab.* **41(2)** (2013), 1072–1087.

[2] D. Aldous, The continuum random tree. I, *Ann. Probab.* **19(1)** (1991), 1–28.

[3] V.F. Kolchin, *Random Mappings*, Translated from the Russian, With a foreword by S. R. S. Varadhan, Transl. Ser. Math. Engrg., Optimization Software, Inc., Publications Division, New York, 1986.