A combinatorial link between labelled graphs and increasingly labelled Schröder trees

Olivier Bodini, Antoine Genitrini, Mehdi Naima

A combinatorial link between Erdös-Renyi graphs and increasingly labelled Schröder trees

Olivier Bodini[1][0000−0002−1867−667X], Antoine Genitrini[2][0000−0002−5480−0236], and Mehdi Naima[2][0000−0003−2735−2420]

[1] Université Sorbonne Paris Nord – LIPN – CNRS, UMR 7030
`Olivier.Bodini@lipn.univ-paris13.fr`
[2] Sorbonne Université – LIP6 – CNRS UMR 7606 {`Antoine.Genitrini`, `Mehdi.Naima`}`@lip6.fr`

Abstract. In this paper we study a model of Schröder trees whose labelling is increasing along the branches. Such tree family is useful in the context of phylogenetic. The tree nodes are of arbitrary arity (i.e. out-degree) and the node labels can be repeated throughout different branches of the tree. Once a formal construction of the trees is formalized, we then turn to the enumeration of the trees inspired by a renormalisation due to Stanley on acyclic orientations of graphs. We thus exhibit links between our tree model and labelled graphs and prove a one-to-one correspondence between a subclass of our trees and labelled graphs. As a by-product we obtain a new natural combinatorial interpretation of Stanley's renormalising factor. We then study different combinatorial characteristics of our tree model and finally, we design an efficient uniform random sampler for our tree model which allows to generate uniformly Erdös-Renyi graph with a constant number of rejections on average.

Keywords: Evolution process · Schröder trees · Increasing trees · Monotonic trees · Erdös-Rényi graphs · Combinatorics · Uniform sampling.

1 Introduction

Increasing trees are ubiquitous in combinatorics especially because they aim at modelling various classical phenomena: phylogenetics, the frequencies of family names or the graph of the Internet [24] for example. Meir and Moon [19] studied the distance between nodes in their now classical model of recursive trees. Bergeron *et al.* [2] studied several families of increasingly-labelled trees for a wide range of models embedded in the simple families of trees. We also refer to [8] where recent results on various families of increasing trees and the methods to study them, from a quantitative point of view, are surveyed.

Increasing trees can often be described as the result of a dynamical construction: the nodes are added one by one at successive integer-times in the tree (their labels being the time when they are added). This dynamical process allows sometimes to apply probabilistic methods to show results about different characteristics on the trees and often gives an efficient way to uniformly sample large trees using simple, iterative and local rules.

In the recent years, many links were found between evolution processes in the form of increasing trees and classical combinatorial structures, for instance permutations are known to be in bijection with increasing binary trees [11], increasing even trees and alternating permutations are put in bijection in [17,6], plane recursive trees are related to Stirling permutations [16] and more recently increasing Schröder trees have been proved in one-to-one correspondence with even permutations and with weak orderings on sets of n elements (counted by ordered Bell numbers) in [3,4]. By adding some constraint in the increasing labelling of the latter model, Zhicong *et al.* [26] exhibited closed relationships between various families of polynomials (especially Eulerian, Narayana and Savage and Schuster polynomials).

The theory of analytic combinatorics developed in [11] gives a framework to study many classes of discrete structures by applying principles based on the now classical *symbolic method*. In various situations we get direct answers to questions concerning the count of the number of objects, the study of typical shapes and the development of methods for the uniform sampling of objects. Using this approach we explore links between labelled directed graphs and an evolution process that generates increasing trees seen as enriched Schröder trees. Schröder in [23] studied trees with possible multifurcations to model phylogenetic. The trees he studied where counted by their number of leaves which represent the number of species. We pursue enriching Schröder trees in the same vein as [3,4] but with a more general model.

Our evolution process can be reinterpreted as a builder for phylogenetic tree that represents the evolutionary relationship among species. At each branching node of the tree, the descendant species from distinct branches have distinguished themselves in some manner and are no more dependent: the past is shared but the futures are independent. For more information on the phylogenetic links the reader may refer to the thesis [20].

The study of this evolution process leads to exhibit unexpected links between our trees and labelled graphs: we then prove a bijection between both families of structures. The links we find also give a new combinatorial interpretation of the renormalisation factor that Stanley used in [25] based on ideas of [7] and more recently for *graphic generating functions* by de Panafieu and Dovgal in [22].

Our main contributions: A study of an evolution process that produces increasing trees with label repetitions. The study of this evolution process using tools of analytic combinatorics produces functional equations for generating functions that are divergent. Next, using a renormalisation we provide a quantitative study for the enumeration problem and the asymptotic analysis of several parameters. After that, we introduce a one-to-one correspondence between a sub-family of our increasing trees and directed labelled graphs. Finally, we design a uniform random sampler for the increasing trees which easily translates to a uniform random sampler for labelled directed graphs with constant time rejection.

This work is part of a long term project, in which we aim at relaxing the classical rules of increasing labelling (described in, e.g., [2]), by allowing label repetitions in the tree.

In Table 1 we provide the main statistics of our enriched Schröder tree model that we will call *strict monotonic general tree model*. Due to its relationship with Schröder trees the size of a tree is given by its number of leaves, independently of its number of internal nodes.

	Number of trees	Average number of distinct labels	Average number of internal nodes	Average height
Strict monotonic general trees	$c\ (n-1)! \cdot 2^{(n-1)(n-2)/2}$	$\Theta(n)$	$\Theta(n^2)$	$\Theta(n)$

Table 1: Main analytic results for the characteristics of a large typical tree. n stands for the size of the trees and the results are asymptotic when $n \to +\infty$.

Plan of the paper: The paper is organized as follows. First, in Section 2 we present our evolution process and then extract from it a general recursive formula to count the number of trees of a given size. We end this section by giving the statement of Theorem 2 on the asymptotic enumeration of the trees. Next, in Section 3 we count the trees according to their sizes, we also study the distribution of the number of iteration steps to construct the trees of a given as size. As a result, we can simply prove Theorem 2 and give detailed characteristics of the shape of large trees. Based on the previous results, in Section 4.1 we make quantitative studies of several tree parameters: the number of internal labels and the height of the trees. We then turn to show the relationship between our trees and labelled graphs. We exhibit a bijection in Section 5. Finally, in Section 6 we design an unranking method for the sampling of strict monotonic general trees. It naturally can be tuned to define a uniform random sampler. We conclude the paper by showing of this sampler can be used to generate Erdös-Renyi graphs with a constant number of rejections.

2 The model and its enumeration

Definition 1 *A* strict monotonic general tree *is a labelled tree constructed by the following evolution process:*

- *Start with a single (unlabelled) leaf.*
- *At every step $\ell \geq 1$, select a non-empty subset of leaves, replace all of them by internal nodes labelled by ℓ, attach to at least one of them a sequence of two leaves or more, and attach to all others a unique leaf.*

The two trees in Fig. 1 are sampled uniformly among all strict monotonic general trees of respective sizes (i.e. number of leaves) 15 and 500. The left-hand-side

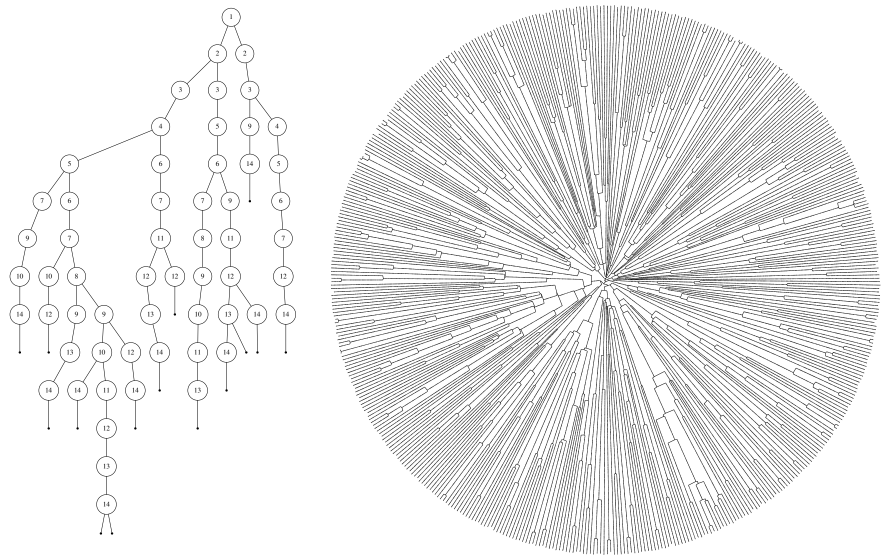

Fig. 1: Two strict monotonic general trees, with respective sizes 15 and 500

tree has 14 distinct node-labels, i.e. it can be built in 14 steps using Definition 1. The right-hand-side tree is represented as a circular tree with stretched edges: the length of an edge is proportional to the label difference of the two nodes it connects. Here the tree contains 500 leaves built with 499 iterations of the growth process. Thus the maximal arity is 2. This tree contains 62494 internal nodes almost all of them (unless 499) being unary nodes.

We can specify strict monotonic general trees using the symbolic method [11]; the internal node labelling is transparent and does not appear in the specification in consequence, we use ordinary generating functions. We denote by $F(z)$ the generating function of strict monotonic general trees and by $\mathcal{F}_n$ the set of all strict monotonic general trees of size n; from Definition 1, we get

$$F(z) = z + F\left(z + \frac{z}{1-z}\right) - F(2z). \qquad (1)$$

The combinatorial meaning of this specification is the following: A tree of is either a single leaf, or it is obtained by taking an already constructed tree, and replace each leaf by either a leaf (i.e. no change) or an internal node attached to a sequence of at least one leaf. Furthermore we omit the case where no leaf is replaced by an internal node with at least two children (this is encoded in the subtracting $F(2z)$).

From this equation we extract the recurrence for the number f_n of strict monotonic general trees with n leaves. In fact we get

$$\begin{aligned} f_n = [z^n]F(z) &= [z^n]\left(z + F\left(z + \frac{z}{1-z}\right) - F(2z)\right) \\ &= \delta_{n,1} - 2^n f_n + [z^n]\sum_{\ell\geq 1} f_\ell \left(z + \frac{z}{1-z}\right)^\ell \\ &= \delta_{n,1} - 2^n f_n + \sum_{\ell\geq 1} f_\ell \, [z^{n-\ell}] \sum_{i=0}^{\ell} \binom{\ell}{i} \left(\frac{1}{1-z}\right)^i , \end{aligned}$$

which implies that

$$f_n = \begin{cases} 1 & \text{if } n = 1, \\ \sum_{\ell=1}^{n-1} \sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i} \, 2^{\ell-i} \, \binom{n-\ell-1}{i-1} \, f_\ell & \text{for all } n \geq 2. \end{cases} \tag{2}$$

The inner sum can be explained combinatorially: starting with a tree of size ℓ we reach a tree of size n in one iteration by adding $n-\ell$ leaves. The index i in the inner sum stands for the number of leaves that are replaced by internal nodes or arity at least 2, by definition of the model we have $1 \leq i \leq \min(n-\ell, \ell)$. There are $\binom{\ell}{i}$ possible choices for the i leaves that are replaced by nodes of arity at least 2. Each of the remaining $\ell - i$ leaves is either kept unchanged or replaced by a unary node, which gives $2^{\ell-i}$ possible choices. And finally, there are $\binom{n-\ell-1}{\ell-1}$ possible ways to distribute the (indistinguishable) $n-\ell$ additional leaves among the i new internal nodes so that each of the i nodes is given at least one additional leaf (it already has one leaf, which is the leaf that was replaced by an internal node). The first terms of the sequence are the following:

$$(f_n)_{n\geq 0} = (0, 1, 1, 5, 66, 2209, 180549, 35024830, 15769748262, \dots) \, .$$

Theorem 2. *There exists a constant c such that the number f_n of strict monotonic general trees of size n satisfies, asymptotically when n tends to infinity,*

$$f_n \underset{n\to\infty}{\sim} c \, (n-1)! \, 2^{\frac{(n-1)(n-2)}{2}} .$$

In the proof of the theorem we exhibit the following bounds: $1.4991 < c < 1.8932$. But through several experimentations we see that $c < 3/2$ but it is close to $3/2$. For instance when $n = 1000$, we get $c \approx 1.49913911$. We postpone the proof to the next section to make use of the number of iteration steps.

3 Iteration steps and asymptotic enumeration

In this section, we look at the number of distinct internal-node labels that occur in a typical strict monotonic general tree, i.e. the number of iterations needed to build it.

Proposition 1. *Let $f_{n,k}$ denotes the number of strict monotonic general trees of size n with k distinct node-labels, then, for all $n \geq 1$,*

$$f_{n,n-1} = (n-1)!\; 2^{\frac{(n-1)(n-2)}{2}}.$$

Note that the first terms are

$$(f_{n,n-1})_{n\geq 0} = (0, 1, 1, 4, 48, 1536, 122880, 23592960, 10569646080, \dots)\,.$$

This is a shifted version of the sequence OEIS A011266 used by Stanley in [25] that is in relation with acyclic orientations of graphs.

Proof. We use a new variable u to mark the number of iterations (i.e. the number of distinct node-labels) in the iterative Equation (1). We get

$$F(z,u) = z + u\; F\left(z + \frac{z}{1-z}, u\right) - u\; F(2z,u). \tag{3}$$

Using either Equation (3) or a direct combinatorial argument, we get that, for all $k \geq n$, $f_{n,k} = 0$ and

$$f_{n,k} = \begin{cases} 1 & \text{if } n=1 \text{ and } k=0,\\ \sum_{\ell=k}^{n-1} \sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i}\; 2^{\ell-i}\; \binom{n-\ell-1}{i-1}\; f_{\ell,k-1} & \text{if } 1 \leq k < n. \end{cases}$$

In particular, for $k = n-1$, we get $f_{n,n-1} = (n-1)\; 2^{n-2}\; f_{n-1,n-2}$. Solving the recurrence we get

$$f_{n,n-1} = f_{1,0} \prod_{j=1}^{n-1} j\; 2^{j-1} = (n-1)!\; 2^{\sum_{j=0}^{n-2} j} = (n-1)!\; 2^{\frac{(n-1)(n-2)}{2}},$$

because $f_{1,0} = 1$. This concludes the proof. □

Alternatively the recurrence of $f_{n,n-1}$ can be obtained by extracting the coefficient $[z^n]$ in the following functional equation $T(z) = z + z^2\; T'(2z)$.

Lemma 1. *Both sequences (f_n) and $(f_{n,n-1})$ have the same asymptotic behaviour up to a multiplicative constant.*

Proof. Let us start with the definition of a new sequence

$$g_n = \begin{cases} 1 & \text{if } n=1,\\ f_n / f_{n,n-1} & \text{otherwise.} \end{cases}$$

This sequence g_n satisfies the following recurrence:

$$g_n = \begin{cases} 1 & \text{if } n=1,\\ \sum_{\ell=1}^{n-1} \sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i}\; 2^{\ell-i}\; \binom{n-\ell-1}{i-1}\; g_\ell \frac{(\ell-1)!\; 2^{(\ell-1)(\ell-2)/2}}{(n-1)!\; 2^{(n-1)(n-2)/2}} & \text{otherwise.} \end{cases}$$

When $n > 1$, extracting the term g_{n-1} from the sum we get

$$g_n = g_{n-1} + \sum_{\ell=1}^{n-2} \sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i} 2^{\ell-i} \binom{n-\ell-1}{i-1} g_\ell \frac{(\ell-1)!\, 2^{(\ell-1)(\ell-2)/2}}{(n-1)!\, 2^{(n-1)(n-2)/2}}.$$

Since all summands are non-negative, this implies that $g_n \geq g_{n-1}$, and thus that this sequence is non-decreasing. To prove that this sequence converges, it only remains to prove that it is (upper-)bounded.

Equation (2) implies that, for $n \geq 2$,

$$f_n \leq \sum_{\ell=1}^{n-1} 2^{\ell-1} \sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i} \binom{n-\ell-1}{i-1} f_\ell.$$

Chu-Vandermonde's identity states that, for all $\ell \leq n$,

$$\sum_{i=1}^{\min(n-\ell,\ell)} \binom{\ell}{i} \binom{n-\ell-1}{i-1} = \binom{n-1}{\ell-1}.$$

This implies the following upper-bound for f_n:

$$f_n \leq \sum_{\ell=1}^{n-1} 2^{\ell-1} \binom{n-1}{\ell-1} f_\ell = \sum_{\ell=1}^{n-1} 2^{n-\ell-1} \binom{n-1}{\ell} f_{n-\ell}.$$

Using the same argument for g_n we get

$$g_n \leq g_{n-1} + \sum_{\ell=2}^{n-1} \frac{2^{(\ell-1)(\ell-2n+2)/2}}{\ell!} \; g_{n-\ell}.$$

We look at the exponent of 1 in the sum: For all $\ell \geq 2$ (as in the sum), we have $2\ell \geq \ell + 2$, and thus $2n - \ell - 2 \geq 2(n - \ell)$. This implies that for all $\ell \geq 2$, $(\ell - 1)(\ell - 2n + 2)/2 \leq -(n - \ell)$, and thus that

$$g_n \leq g_{n-1} + \sum_{\ell=2}^{n-1} \frac{1}{\ell!\, 2^{n-\ell}} \; g_{n-\ell}.$$

Since the sequence $(g_n)_n$ is non-decreasing, we obtain

$$g_n \leq g_{n-1} + \frac{g_{n-1}}{2^n} \sum_{\ell=2}^{n-1} \frac{2^\ell}{\ell!} \leq g_{n-1} + g_{n-1} \frac{\mathrm{e}^2 - 3}{2^n}.$$

We set $\alpha = \mathrm{e}^2 - 3$. Iterating the last inequality, we get that

$$g_n \leq g_{n-1} \left(1 + \frac{\alpha}{2^n}\right) \leq g_1 \prod_{i=2}^{n} \left(1 + \frac{\alpha}{2^i}\right) = \exp\left(\sum_{i=2}^{n} \ln\left(1 + \frac{\alpha}{2^i}\right)\right),$$

because $g_1 = 1$. Note that, when $i \to +\infty$, we have $\ln(1+\alpha 2^{-i}) \leq \alpha 2^{-i}$ (because $\ln(1+x) \leq x$ for all $x \geq 0$). This implies that, for all $n \geq 1$,

$$g_n \leq \exp\left(\alpha \sum_{i=2}^{\infty} 2^{-i}\right) = \exp(\alpha/2).$$

In other words, the sequence $(g_n)_n$ is bounded. Since it is also non-decreasing, it converges to a finite limit c, which is also non-zero since $g_n \geq g_1 \neq 0$ for all $n \geq 1$. This is equivalent to $f_n \sim c f_{n,n-1}$ when $n \to +\infty$ as claimed. To get a lower wound on c, note that, for all $n \geq 1$, $c \geq g_n \geq g_{1000} = f_{1000}/f_{1000,999} \approx 1.49913911$. □

Lemma 1 gives first step for a proof of Theorem 2. But in order to get an upper bound for the constant c, we have to introduce a proof slightly more technical. It is given in the Appendix A.

This result means that asymptotically a constant fraction of the strict monotonic general trees of size n are built in $(n-1)$ steps. For these trees, at each step of construction only one single leaf expands into a binary node. All other leaves either become a unary node or stay unchanged, meaning that on average half of the leaves will expand into unary node with one leaf expanding into a binary node. The number of internal nodes of these trees then grows like $n^2/4$.

4 Analysis of typical parameters

4.1 Quantitative analysis of the number of internal nodes

Theorem 3. *Let $I_n^{\mathcal{F}}$ be the number of internal nodes in a tree taken uniformly at random among all strict monotonic general trees of size n. Then for all $n \geq 1$, we have*

$$\frac{(n-1)(n+2)}{6} \leq \mathbb{E}[I_n^{\mathcal{F}}] \leq \frac{(n-1)n}{2}.$$

To prove this theorem, we use the following proposition.

Proposition 2. *Let us denote by $s_{n,k}$ the number of strict monotonic general trees of size n that have $n-1$ distinct node-labels and k internal nodes. For all $n \geq 1$ and $k \geq 0$,*

$$s_{n,k} = (n-1)! \binom{(n-1)(n-2)/2}{k-(n-1)},$$

and thus, if $I_n^{\mathcal{S}}$ is the number of internal nodes in a tree taken uniformly at random among all strict monotonic general trees of size n that have $n-1$ distinct label nodes, then, for all $n \geq 1$,

$$\mathbb{E}[I_n^{\mathcal{S}}] = \frac{(n-1)(n+2)}{4}.$$

The proof is provided in the Appendix B. We are now ready to prove the main theorem of this section.

Proof (of Theorem 3). Note that the number of internal nodes of a strict monotonic general tree of size n belongs to $\{1,\dots,n(n-1)/2\}$. The upper bound follows from the fact that, at the ℓ-th iteration in Definition 1, a maximum of ℓ internal nodes is added to the tree, and $\sum_{\ell=1}^{n}\ell = n(n-1)/2$. In particular, we thus have that, almost surely for all $n\geq 1$, $I_n^{\mathcal{F}} \leq n(n-1)/2$, and thus $\mathbb{E}[I_n^{\mathcal{F}}] = \mathcal{O}(n^2)$.

For the lower bound, we denote by $\mathcal{S}_n$ the set of strict monotonic general trees of size n that have $n-1$ distinct node-labels. Moreover, we denote by t_n a tree taken uniformly at random in $\mathcal{F}_n$, and by $I_n^{\mathcal{F}}$ its number of internal nodes. We have, for all $n\geq 1$,

$$\begin{aligned}\mathbb{E}[I_n^{\mathcal{F}}] &= \mathbb{E}[I_n^{\mathcal{F}} \mid t_n\in\mathcal{S}_n]\cdot\mathbb{P}(t_n\in\mathcal{S}_n) + \mathbb{E}[I_n^{\mathcal{F}} \mid t_n\notin\mathcal{S}_n]\cdot\mathbb{P}(t_n\notin\mathcal{S}_n)\\ &\geq \mathbb{E}[I_n^{\mathcal{F}} \mid t_n\in\mathcal{S}_n]\cdot\mathbb{P}(t_n\in\mathcal{S}_n) = \mathbb{E}[I_n^{\mathcal{S}}]\ \cdot \frac{f_{n,n-1}}{f_n},\end{aligned}$$

where we have used conditional expectations and the fact that conditionally on being in $\mathcal{S}_n$, t_n is uniformly distributed in this set, and, in particular, $\mathbb{E}[I_n^{\mathcal{F}} \mid t_n \in \mathcal{S}_n] = \mathbb{E}I_n^{\mathcal{S}}$. Using Proposition 2 and the upper bound of Proposition 1, we thus get

$$\mathbb{E}[I_n^{\mathcal{F}}] \geq \frac{2}{3}\ \frac{(n-1)(n+2)}{4},$$

which concludes the proof. □

4.2 Quantitative analysis of the number of distinct labels

Theorem 4. *Let $X_n^{\mathcal{F}}$ denotes the number of distinct internal-node labels (or construction steps) is a tree taken uniformly at random among all strict monotonic general trees of size n, then for all $n\geq 1$,*

$$\frac{2}{3}\,(n-1) \leq \mathbb{E}[X_n^{\mathcal{F}}] \leq n-1.$$

Proof. First note that since at every construction step in Definition 1 we add at least one leaf in the tree, then after ℓ construction steps, there are exactly ℓ distinct labels and at least $\ell+1$ leaves in the tree. Therefore, $n\geq X_n^{\mathcal{F}}+1$ almost surely for all $n\geq 1$, which implies in particular that $\mathbb{E}[X_n]\leq n-1$, as claimed.

For the lower bound, we reason as in the proof of Theorem 3, and using the same notations:

$$\mathbb{E}[X_n^{\mathcal{F}}] \geq \mathbb{E}[X_n^{\mathcal{F}} \mid t_n\in\mathcal{S}_n]\cdot\mathbb{P}(t_n\in\mathcal{S}_n) = (n-1)\ \frac{f_{n,n-1}}{f_n},$$

because $\mathbb{E}[X_n^{\mathcal{F}} \mid t_n\in\mathcal{S}_n] = n-1$ by definition of $\mathcal{S}_n$ (being the set of all strict monotonic general trees of size n that have $n-1$ distinct node-labels). Using the upper bound of Proposition 1 gives that $\mathbb{E}[X_n^{\mathcal{F}}]\geq 2(n-1)/3$, which concludes the proof. □

4.3 Quantitative analysis of the height of the trees

Theorem 5. *Let $H_n^{\mathcal{F}}$ denotes the height of a tree taken uniformly at random in $\mathcal{F}_n$, the set of all strict monotonic general trees of size n. Then we have, for all $n \geq 0$,*

$$\frac{n}{3} \leq \mathbb{E}[H_n^{\mathcal{F}}] \leq n-1.$$

To prove this theorem, we first prove the following:

Proposition 3. *Let us denote by $H_n^{\mathcal{S}}$ the height of a tree taken uniformly at random in $\mathcal{S}_n$, the set of all strict monotonic general trees of size n that have $n-1$ distinct labels. Then we have, for all $n \geq 0$,*

$$\frac{n}{2} \leq \mathbb{E}[H_n^{\mathcal{S}}] \leq n-1.$$

Proof. Define the sequence of random trees $(t_n)_{n\geq 0}$ recursively as: t_1 is a single leaf; and given t_{n-1}, we define t_n as the tree obtained by choosing a leaf uniformly at random among all leaves of t_{n-1}, replacing it by an internal nodes to which two leaves are attached, and, for each of the other leaves of t_{n-1}, choose with probability $1/2$ (independently from the rest) whether to leave it unchanged or to replace it by a unary node to which one leaf is attached.

One can prove by induction on n that for all $n \geq 1$, t_n is uniformly distributed in $\mathcal{S}_n$. We denote by $H_n^{\mathcal{F}}$ the height of t_n. Since the height of t_n is at most the height of t_{n-1} plus 1 for all $n \geq 2$, we get that $H_n^{\mathcal{S}} \leq n-1$ almost surely.

For the upper bound, we note that, for the height of t_n to be larger than the height of t_{n-1}, we need to have replaced at least one of the maximal-height leaves in t_{n-1}. There is at least one leaf of t_{n-1} which is at height $H_{n-1}^{\mathcal{S}}$ and this leaf is replaced by an internal node with probability

$$\frac{1}{2}\left(1-\frac{1}{n-1}\right)+\frac{1}{n-1} \geq \frac{1}{2}.$$

Therefore, for all $n \geq 1$, we have

$$\mathbb{P}(H_n^{\mathcal{S}} = H_{n-1}^{\mathcal{S}}+1) \geq \frac{1}{2},$$

which implies, since $H_n^{\mathcal{S}} \in \{H_{n-1}^{\mathcal{S}}, H_{n-1}^{\mathcal{S}}+1\}$ almost surely,

$$\mathbb{E}[H_n^{\mathcal{S}}] = \mathbb{E}[H_{n-1}^{\mathcal{S}}] + \mathbb{P}(H_n^{\mathcal{S}} = H_{n-1}^{\mathcal{S}}+1) \geq \mathbb{E}[H_{n-1}^{\mathcal{S}}] + \frac{1}{2}.$$

Therefore, for all $n \geq 1$, we have $\mathbb{E}[H_n^{\mathcal{S}}] \geq \mathbb{E}[H_0^{\mathcal{S}}] + n/2 = n/2$, as claimed. □

Proof (of Theorem 5). By Definition 1, it is straightforward to see that the height of a tree built in ℓ steps is at most ℓ since the height increases by at most one per construction step. Since a tree of size n is built in at most $n-1$ steps, we get that $H_n^{\mathcal{F}} \leq n-1$ almost surely, which implies, in particular, that $\mathbb{E}[H_n^{\mathcal{F}}] \leq n-1$.

For the lower bound, note that, if t_n is a tree taken uniformly at random in $\mathcal{F}_n$ and $H_n^{\mathcal{F}}$ is its height, then

$$\mathbb{E}[H_n^{\mathcal{F}}] \geq \mathbb{E}[H_n^{\mathcal{F}} | t_n \in \mathcal{S}_n] \cdot \mathbb{P}(X \in \mathcal{S}_n) \geq \frac{2}{3}\, \mathbb{E}[H_n^{\mathcal{S}}],$$

where we have used Proposition 1 and the fact that t_n conditioned on being in $\mathcal{S}_n$ is uniformly distributed in this set and thus $E[H_n^{\mathcal{F}} \mid t_n \in \mathcal{S}_n] = \mathbb{E}H_n^{\mathcal{S}}$. By Proposition 3, we thus get $\mathbb{E}[H_n^{\mathcal{F}}] \geq n/3$, as claimed. □

4.4 Quantitative analysis of the depth of the leftmost leaf

Theorem 6. *Let us denote by $D_n^{\mathcal{F}}$ the depth of the leftmost leaf of a tree taken uniformly at random in $\mathcal{F}_n$, the set of all strict monotonic general trees of size n. Then we have, for all $n \geq 0$,*

$$\frac{n}{3} \leq \mathbb{E}[H_n^{\mathcal{F}}] \leq n - 1.$$

Proposition 4. *Let us denote by $D_n^{\mathcal{S}}$ the depth of the leftmost leaf of a tree taken uniformly at random in $\mathcal{S}_n$, the set of all strict monotonic general trees of size n that have $n-1$ distinct labels. Then we have, for all $n \geq 0$,*

$$\frac{n}{2} \leq \mathbb{E}[D_n^{\mathcal{S}}] \leq n - 1.$$

Proof. Given the uniform process of trees t_n presented in Proposition 3. The depth of the leftmost leaf is always smaller than $n-1$. Let X_n be a Bernoulli variable taking value 1 if the leftmost leaf of t_n has been expanded at iteration n and the value 0 otherwise. Then for $n \geq 1$,

$$\mathbb{P}\left(X_n = 1\right) = \frac{1}{n} + \frac{(n-1)}{n}\frac{1}{2} = \frac{n+1}{2n} \geq \frac{1}{2}.$$

Since at each iteration step either the leftmost leaf expand to make a binary node which gives $\frac{1}{n}$ or it has not created a binary and then it has $\frac{1}{2}$ probability to make a unary node. The depth of the leftmost leaf is $D_n^{\mathcal{S}} = \sum_{k=1}^{n} X_k$. Therefore for $n \geq 1$,

$$\mathbb{E}[D_n^{\mathcal{S}}] \geq \frac{n}{2}.$$

Which concludes the proof. □

Proof (of Theorem 6). By the same arguments as in Theorem 5 the result follows directly since we have the same bounds on the depth of leftmost leaf as we had in the height of the tree. □

5 Correspondence with labelled graphs

In Section 3 we defined $f_{n,k}$ the number of strict monotonic general trees of size n with exactly k distinct node-labels. Then we have shown that, for all $n \geq 1$,

$$f_{n,n-1} = (n-1)!\, 2^{\frac{(n-1)(n-2)}{2}}.$$

The factor $2^{(n-1)(n-2)/2} = 2^{\binom{n-1}{2}}$ in the context of graphs with $n-1$ vertices counts the different combinations of undirected edges between vertices. The factor $(n-1)!$ accounts for all possible permutations of vertices. We will denote $\mathcal{S}_n$ to be the trees that $f_{n,n-1}$ counts and exhibit a bijection between strict monotonic general trees of $\mathcal{S} = \cup_{n\geq 1}\mathcal{S}_n$ with a class of labelled graphs with $n-1$ vertices defined in the following.

For all $n \geq 1$, we denote by $\mathcal{G}_n$ the set of all labelled graphs (V, ℓ, E) such that $V = \{1, \dots, n\}$, $E \subseteq \{\{i,j\} \colon i \neq j \in V\}$ and $\ell = (\ell_1, \dots, \ell_n)$ is a permutation of V (see Fig. 2 for an example). We set $\mathcal{G} = \cup_{n=0}^{\infty}\mathcal{G}_n$. Choosing a graph in $\mathcal{G}_n$ is equivalent to (1) choosing ℓ (there are $n!$ choices) and (2) for each of the $\binom{n}{2}$ possible edges, choose whether it belongs to E or not (there are $2^{\binom{n}{2}}$ choices in total). In total, we thus get that $|\mathcal{G}_n| = n!\, 2^{\binom{n}{2}}$.

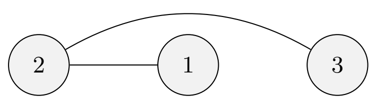

Fig. 2: The graph $\mathcal{G}_3$. In this representation, the vertices $V = \{1, \dots, n\}$ are drawn from left to right (node 1 is the leftmost one, node n is the rightmost one), and their label is their image by ℓ: in this example $\ell = (2, 1, 3)$.

A size-n permutation σ is denoted by $(\sigma_1, \dots, \sigma_n)$, and σ_i is its i-th element (the image of i), while $\sigma^{-1}(k)$ is the preimage of k (the position of k in the permutation).

Another important bijection that we will use is the bijection between binary increasing trees and permutations, see [11, p. 143].

We define $\mathcal{M}'' : \mathcal{S} \to \mathcal{G}$ recursively on the size of the tree it takes as an input: first, if t is the tree of size 1 (which contains only one leaf) or the tree of size 2 (one internal node attached to two leaves), then we set $\mathcal{M}''(t)$ to be the graph $(\{1\}, (1),)$ (the graph with one vertex labelled 1 and no edge). Now assume we have defined $\mathcal{M}''$ on $\cup_{\ell=1}^{n-1}\mathcal{S}_\ell$, and consider a tree $t \in \mathcal{S}_n$. By Definition 1 and since $t \in \mathcal{S}_n$, then there exists a unique binary node in t labelled by $n-1$, and this node is attached to two leaves. Consider $\hat{t}$ the tree obtained when removing all internal nodes labelled by $n-1$ (and all the leaves attached to them) from t and replacing them by leaves. Denote by v_n the position (in, e.g., depth-first order) of the leaf of $\hat{t}$ that previously contained the binary node labelled by $n-1$ in t.

Denote by $u_1, \ldots, u_m$ the positions or the leaves of $\hat{t}$ that previously contained unary nodes labelled by $n-1$ in t. We set $\mathcal{M}^{''}(\hat{t}) = (\{1, \ldots, n-1\}, \hat{\ell}, \hat{E})$ and define $\mathcal{M}^{''}(t) = (\{1, \ldots, n\}, \ell, E)$ where

$$\ell_i = \begin{cases} v_n & \text{if } i = n \\ \hat{\ell}_i & \text{if } \hat{\ell}_i < v_n \\ \hat{\ell}_i + 1 & \text{if } \hat{\ell}_i \geq v_n, \end{cases}$$

$E = \hat{E} \cup \{\{\hat{\ell}^{-1}(u_j), n\} : 1 \leq j \leq m\}$. An example of the bijection is depicted in Fig. 3.

Theorem 7. *The mapping $\mathcal{M}^{''}$ is bijective, and $\mathcal{M}^{''}(\mathcal{S}_n) = \mathcal{G}_{n-1}$.*

Proof. From the definition, it si clear that two different trees have two distinct images by $\mathcal{M}^{''}$, thus implying that $\mathcal{M}^{''}$ is injective; this is enough to conclude since $|\mathcal{G}_{n-1}| = |\mathcal{S}_n|$ (see Theorem 1 for the cardinality of $\mathcal{S}_n$).

*R*emark: It is interesting to note that this graph model is a labelled version of the binomial random graph $\mathcal{G}_n(1/2) = (V, E)$ defined as follows: $V = \{1, \ldots, n\}$ and each edge belong to E with probability $1/2$, independently from the other edges. This model, also called the Erdös-Renyi random graph was originally introduced by Erdös and Renyi [10], and simultaneously by Gilbert [14], and has been since then extensively studied in the probability and combinatorics literature (see, for example, the books [5] and [9] for introductory surveys).

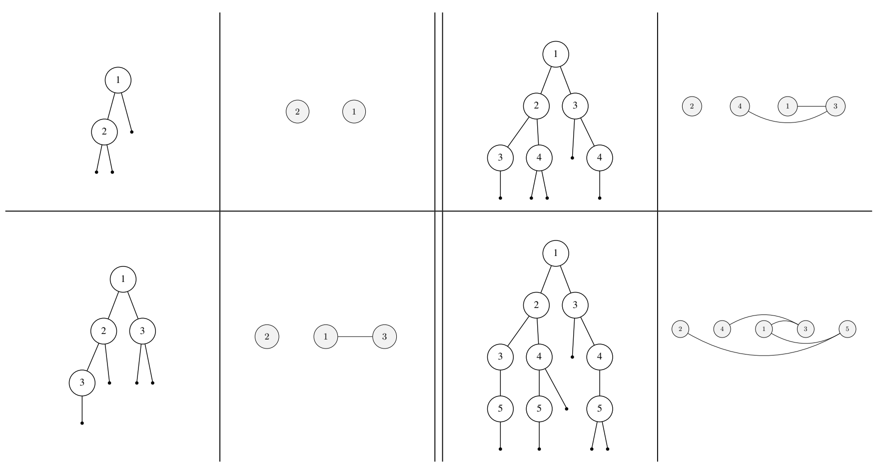

Fig. 3: Bijection between an evolving tree in $\mathcal{S}$ from size 3 to 6 and its corresponding graph in $\mathcal{G}$

6 Uniform random sampling

In this section we exhibit a very efficient way for the uniform sampling of strict monotonic general trees using the described evolution process. We finally explain how this same uniform sampler can be used to generate Erdös-Renyi graphs.

The global approach for our algorithmic framework deals with the *recursive generation method* adapted to the analytic combinatorics point of view in [12]. But in our context we note that we can obtain for free (from a complexity view) an *unranking algorithm*. This fact is sufficiently rare to mention it: usually unranking algorithm are less efficient than recursive generation ones. Unranking algorithmic has been developed in the 70's by Nijenhuis and Wilf [21] and then has been introduced to the context of analytic combinatorics by Martínez and Molinero [18]. We use the same method as the one described in [4].

In our recurrence when r grows, the sequence $(f_{n-r})_r$ decreases extremely fast. Thus for the uniform random sampling, it will appear more efficient to read Equation (2) in the following way:

$$\begin{aligned} f_n =& \binom{n-1}{1} 2^{n-2} f_{n-1} + \sum_{i=1}^{2} \binom{n-2}{i} 2^{n-2-i} f_{n-2} \\ &+ \sum_{i=1}^{3} \binom{n-3}{i} 2^{n-3-i} \binom{2}{i-1} f_{n-3} + \cdots + f_1. \end{aligned} \tag{4}$$

Using the latter decomposition the algorithm can now be described as Algorithm 1.

In Algorithm 1 note that the `While` loop allows to determine the values for ℓ, i and r (see Equation (2) to identify the variables). Then the recursive call is done using the adequate rank $r \mod f_{n-\ell}$. The last lines of the algorithm (for 21 to 27) are necessary to modify the tree T of size $n-\ell$ that has just been built. In line 22 we determine which leaves of T will be substituted by internal nodes (of arity at most 2) with new leaves. It is based on the unranking of combinations, see [13] for a survey in this context. Then for the other leaves that are either kept as they are of replaced by unary internal nodes attached to a leaf we use the integer F seen as a $n-\ell-i$-bit integer: if the bit $\#s$ is 0 then the corresponding leaf is kept, and if it is 1 then the leaf is substituted. And finally the composition unranking allows to determine how many leaves are attached to the nodes selected with B.

Theorem 8. *The function* UNRANKTREE *is an unranking algorithm and calling it with the parameters n and a uniformly-sampled integer s in $\{0, \ldots, f_n - 1\}$ gives as output a uniform strict monotonic general tree of size n .*

The correctness of the algorithm follows directly from the total order over the trees deduced from the decomposition Equation (4).

Theorem 9. *Once the pre-computations have been done, the function* UNRANKTREE *needs in average $\Theta(n)$ arithmetic operations to construct a tree of size n.*

Algorithm 1 Strict Monotonic General Tree Unranking

```
function UnrankTree(n, s)
    if n = 1 then
        return the tree reduced to a single leaf
    ℓ := 1
    r := s
    i := 1
    while r >= 0 do
        t := binom(n−ℓ, i) 2^(n−ℓ−i) binom(ℓ−1, i−1)
        r := r − t · f_{n−ℓ}
        i := i + 1
        if i > min(ℓ, n − ℓ) then
            i := 1
            ℓ := ℓ + 1
    if i > 1 then
        i := i − 1
    else
        ℓ := ℓ − 1
        i := min(ℓ, n − ℓ)
    r := r + t · f_{n−ℓ}
    T := UnrankTree(n − ℓ, r mod f_{n−ℓ})
    r := r // f_{n−ℓ}                              ▷ // stands for the integer division
    B := UnrankBinomial(n − ℓ, i, r // binom(n−ℓ, i))   ▷ see Algorithm 2 in [4]
    r := r mod binom(n−ℓ, i)
    F := r // binom(ℓ−1, i−1)
    C := UnrankComposition(ℓ, i, r mod binom(ℓ−1, i−1)) ▷ see Algorithm 2 in [4]
    Using F, substitute in T, using any traversal, the leaves selected with B with
    internal nodes and new leaves according to C; the other leaves are changed as
    unary nodes with a leaf or not
    return the tree T
```

The sequences $(f_\ell)_{\ell \leq n}$ and $(\ell!)_{\ell \in \{1,\ldots,n\}}$ have been pre-computed and stored.

Proof. The proof for this theorem is analogous to the one for Theorem 3.6.5 in [4] after showing that both UnrankBinomial and UnrankComposition run in $\Theta(n)$ in the number of arithmetic operations.

Finally, using the previous bijection between trees and Erdös-Renyi graphs we remark that Algorithm 1 generates uniformly an Erdös-Renyi graphs $\mathcal{G}_n(1/2)$ with probability $p > 1/2$. In fact, from the results of Theorem 2 we know that Algorithm 1 generates a tree of $\mathcal{S}_n$ in more than $1/1.8932 \approx 0.5282$ of the time, and from the correspondence shown in Section 5 we obtain an Erdös-Renyi graphs $\mathcal{G}_n(1/2)$.

As a last remark, we deduce, with an average of less than 1 rejection, Algorithm 1 generates uniformly an Erdös-Renyi graph of $\mathcal{G}_n(1/2)$.

Acknowledgement. The authors thank Cécile Mailler for fruitful discussions to relate this general model of Schröder trees to the models presented in [4].

References

1. Bender, C., Orszag, S.: Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory, vol. 1 (1999)
2. Bergeron, F., Flajolet, P., Salvy, B.: Varieties of increasing trees. In: CAAP. pp. 24–48 (1992)
3. Bodini, O., Genitrini, A., Naima, M.: Ranked Schröder Trees. In: Proceedings of the Sixteenth Workshop on Analytic Algorithmics and Combinatorics, ANALCO 2019. pp. 13–26 (2019)
4. Bodini, O., Genitrini, A., Mailler, C., Naima, M.: Strict monotonic trees arising from evolutionary processes: combinatorial and probabilistic study. Advances in Applied Mathematics **133**, 102284 (2022)
5. Bollobás, B.: Random Graphs. Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2 edn. (2001)
6. Donaghey, R.: Alternating permutations and binary increasing trees. Journal of Combinatorial Theory, Series A **18**(2), 141–148 (1975)
7. Doubilet, P., Rota, G.C., Stanley, R.: On the foundations of combinatorial theory. vi. the idea of generating function. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory. pp. 267–318. University of California Press (1972)
8. Drmota, M.: Random trees. Springer, Vienna-New York (2009)
9. Durrett, R.: Random Graph Dynamics. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press (2006)
10. Erdös, P., Rényi, A.: On random graphs. Publicationes mathematicae **6**(26), 290–297 (1959)
11. Flajolet, P., Sedgewick, R.: Analytic combinatorics. Cambridge University Press, Cambridge (2009)
12. Flajolet, P., Zimmermann, P., Van Cutsem, B.: A calculus for the random generation of labelled combinatorial structures. Theoretical Computer Science **132**(1-2), 1–35 (1994)
13. Genitrini, A., Pépin, M.: Lexicographic Unranking of Combinations Revisited. Algorithms **14**(3) (2021)
14. Gilbert, E.N.: Random graphs. The Annals of Mathematical Statistics **30**(4), 1141–1144 (1959)
15. Janson, S.: A divertent generating function that can be summed and analysed analytically. Discret. Math. Theor. Comput. Sci. **12**(2), 1–22 (2010)
16. Janson, S.: Plane recursive trees, stirling permutations and an urn model. Discrete Mathematics & Theoretical Computer Science (2008)
17. Kuznetsov, A.G., Pak, I.M., Postnikov, A.E.: Increasing trees and alternating permutations. Russian Mathematical Surveys **49**(6), 79 (1994)
18. Martínez, C., Molinero, X.: Generic algorithms for the generation of combinatorial objects. In: 28th International Symposium on Mathematical Foundations of Computer Science (MFCS). pp. 572–581. Springer Berlin Heidelberg (2003)
19. Meir, A., Moon, J.W.: On the altitude of nodes in random trees. Canad. J. Math. **30**(5), 997–1015 (1978)
20. Naima, M.: Combinatorics of trees under increasing labellings: Asymptotics, bijections and algorithms. Ph.D. thesis, Université Sorbonne Paris Nord (2020)
21. Nijenhuis, A., Wilf, H.S.: Combinatorial algorithms. Computer science and applied mathematics, Academic Press, New York, NY (1975)

22. de Panafieu, E., Dovgal, S.: Symbolic method and directed graph enumeration. Acta Mathematica Universitatis Comenianae **88**(3) (2019)
23. Schröder, E.: Vier Combinatorische Probleme. Z. Math. Phys. **15**, 361–376 (1870)
24. Sornette, D.: Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools. Springer (2006)
25. Stanley, R.P.: Acyclic orientations of graphs. Discrete Mathematics **5**(2), 171 – 178 (1973)
26. Zhicong, L., Jun, M., Shi-Mei, M., Yanghongbo, Z.: Weakly increasing trees on a multiset. Advances in Applied Mathematics **129**, 102206 (2021)

A Proof details for Section 3

Proof (of Theorem 2). In the proof of Lemma 1, by denoting $g_n = f_n/f_{n,n-1}$ when $n > 1$ and $g_1 = 1$, we have proved

$$g_n \leq g_{n-1} + g_{n-1}\frac{\mathrm{e}^2 - 3}{2^n}.$$

We set $\alpha = \mathrm{e}^2 - 3$ and define two other sequences as

$$\bar{g}_n = \begin{cases} 1 & \text{if } n = 1 \text{ or } n = 2, \\ \bar{g}_{n-1} + \frac{\alpha}{2^n}\, \bar{g}_{n-2} & \text{otherwise,} \end{cases}$$

and

$$\bar{\bar{g}}_n = \begin{cases} 1 & \text{if } n = 1 \text{ or } n = 2, \\ \bar{\bar{g}}_{n-1} + \frac{1}{n(n+1)}\, \bar{\bar{g}}_{n-2} & \text{otherwise.} \end{cases}$$

Due to the two first terms and the recurrence equation we have for all positive n, $g_n \leq \bar{g}_n \leq \bar{\bar{g}}_n$. By induction we prove a new expression for $\bar{\bar{g}}_n$:

$$\bar{\bar{g}}_n = \begin{cases} \bar{g}_n & \text{if } n \leq 3, \\ \bar{\bar{g}}_{n-1} + \frac{2}{(n+1)!}\, a_{n-1} & \text{otherwise,} \end{cases}$$

with the sequence $(a_n)_n$ such that $a_1 = 0, a_2 = 1$ and for $n \geq 3$, $a_n = na_{n-1} + a_{n-2}$. This sequence (a_n) is a shifted version of OEIS A058307. We can either follow the work of Janson [15] to study it, but we need less details than him so we describe here an easier approach. We define a new sequence as $b_n = a_n/n!$. We easily prove that $b_n = b_{n-1} + b_{n-2}/(n(n-1))$ with $b_1 = 0$ and $b_2 = 1/2$. Using the later recurrence, we obtain an equation satisfied by its generating function $B(z) = \sum_{n>0} b_n z^n$:

$$B(z) = \frac{z^2}{2} + zB(z) + \int_{t=0}^{u}\int_{t=0}^{z} B(u)\mathrm{d}u.$$

we thus obtain

$$(z-1)B''(z) + 2B'(z) + B(z) + 1 = 0,$$

with $B(0) = 0$ and $B'(0) = 0$. By dividing the equation by $\mathrm{i}\sqrt{1-z}$ and then by a change of variable: $u = 2\mathrm{i}\sqrt{1-z}$, we recognize the classical differential equation satisfied by Bessel functions [1]. We thus derive

$$B(z) = -1 + \frac{1}{\sqrt{1-z}}\left(\alpha\ J_1(2\mathrm{i}\sqrt{1-z}) + \beta\ Y_1(2\mathrm{i}\sqrt{1-z})\right),$$

where $J_\cdot(\cdot)$ and $Y_\cdot(\cdot)$ are the Bessel functions and α and β are two complex constants determined with the initial conditions:

$$\alpha = \frac{Y_1(2\mathrm{i}) - \mathrm{i}Y_0(2\mathrm{i})}{J_1(2)Y_0(2\mathrm{i}) + \mathrm{i}J_0(2)Y_1(2\mathrm{i})}, \qquad \beta = -\frac{J_1(2) - \mathrm{i}J_0(2)}{J_1(2)Y_0(2\mathrm{i}) + \mathrm{i}J_0(2)Y_1(2\mathrm{i})}.$$

We are interested in the asymptotic behaviour of b_n. The dominant singularity of $B(z)$ is at $z = 1$ and there

$$B(z) \underset{z\to 1}{\sim} -\frac{\beta}{\mathrm{i}\pi}\,\frac{1}{1-z}.$$

We thus deduce that b_n tends to $-\beta/(\mathrm{i}\pi) \approx 0.68894$. Since the sequence $\bar{\bar{g}}_n$ satisfies $\bar{\bar{g}}_n = \bar{\bar{g}}_{n-1} + \frac{2}{n(n+1)}\, b_{n-1}$. We deduce that the increasing sequence $(\bar{\bar{g}}_n)$ admits a finite limit. Hence it is also the case for the increasing sequence (g_n). Finally, Proposition 1 allows to conclude for the existence of the constant c. Furthermore we get

$$c < \bar{\bar{g}}_3 + \sum_{\ell\geq 4} \frac{2}{\ell(\ell+1)} \cdot \lim_{n\to\infty} b_n \approx 1.8932.$$

The stated result is thus proved. □

B Proof details for Section 4.1

Proof (of Proposition 2). Let us prove the formula for $s_{n,k}$ by induction. For $n = 1$, k can only be 0 thus $s_{1,0} = 1 = 0!\binom{0}{0}$.

We suppose that $s_{m,k} = (m-1)!\binom{(m-1)(m-2)/2}{k-(m-1)}$ holds for $m = n-1$ and $k \in \{n-1, \ldots, (n-2)(n-3)/2\}$.

Then, we are interested in the value of $s_{n,k}$:

$$s_{n,k} = \sum_{s=0}^{k-(n-1)} (n-2)!\binom{(n-2)(n-3)/2}{s-(n-2)}\binom{n-1}{k-s-1}(k-s-1).$$

Let $k' = k-(n-1)$ and $s' = s-(n-2)$. Replacing k' and s' in the equation gives,

$$\begin{aligned}\tilde{s}_{n,k'} &= \sum_{s'=0}^{k'} (n-2)!\binom{(n-2)(n-3)/2}{s'}\binom{n-1}{k'-s'+1}(k'-s'+1)\\ &= (n-1)!\sum_{s'=0}^{k'}\binom{(n-2)(n-3)/2}{s'}\binom{n-2}{k'-s'}.\end{aligned}$$

Using Chu-Vandermonde identity, we finally obtain

$$s_{n,k} = (n-1)!\binom{(n-1)(n-2)/2}{k-(n-1)}.$$

We now can compute the average number of internal nodes of $\mathcal{S}_n$:

$$\mathbb{E}_n[I_n^{\mathcal{S}}] = \frac{\sum_{k=n-1}^{n(n-1)/2} k(n-1)!\binom{(n-1)(n-2)/2}{k-(n-1)}}{(n-1)!\,2^{(n-1)(n-2)/2}}.$$

Again we reverse the sum: $k' = k - (n-1)$,

$$\mathbb{E}[I_n^{\mathcal{S}}] = \frac{\sum_{k'=0}^{(n-1)(n-2)/2} (k' + (n-1))(n-1)!\binom{(n-1)(n-2)/2}{k'}}{(n-1)!\, 2^{(n-1)(n-2)/2}}$$
$$= \frac{\sum_{k'=0}^{(n-1)(n-2)/2} k\,'\binom{(n-1)(n-2)/2}{k'} + (n-1) \sum_{k'=0}^{(n-1)(n-2)/2} \binom{(n-1)(n-2)/2}{k'}}{2^{(n-1)(n-2)/2}}$$
$$= \frac{(n-1)(n-2)}{4} + (n-1) = \frac{(n-1)(n+2)}{4}.$$

□