Cuts in Increasing Trees*

Olivier Bodini†, and Antoine Genitrini‡§
Olivier.Bodini@lipn.univ-paris13.fr and Antoine.Genitrini@lip6.fr.

October 27, 2014

Abstract. Increasing trees have been extensively studied, since it is a simple model for many natural phenomena. Our paper focuses on sub-families of increasing trees. We measure the number of connected components obtained after having removed the nodes whose labels are smaller than a given value. This measure of cut-length allows, for example, to analyse in average an algorithm for tree-labelling.
It is noticeable that we give exact formulae for the distribution of trees according to their size and cut-lengths. Our approach is based on a construction using grafting processes.

1 Introduction

Scale free networks occur in a large varieties of phenomena, like in natural systems as well as in human activities. We mention for example the size of craters on the moon, the frequencies of family names or the graph of the Internet [Sor06]. A simple way to model such networks is to grow a graph using a preferential attachment dynamics, i.e. at each turn a new vertex is added and connected to a previous one, that is chosen by using a probability proportional to the numbers of links that the existing nodes already have. Barabási and Albert [BA99] used such an approach to model the World Wide Web, because the degree distribution of such a random graph follows a power law that seems realistic in this context.

By adding the constraint that the original graph contains a single node (the source of the diffusion) and by labelling the nodes by their arrival order, we are constructing well known trees, usually called *plane oriented recursive trees*. They belong to a wider family, the *recursive trees* that are under study since the 70's. The earliest works are due to Moon [Moo74] and Meir and Moon [MM74, MM78]. According to them, these combinatorial objects naturally describe the spread of epidemics for example. The first results dealt with the expectation and variance of the distance between two random nodes in a large recursive tree.

Besides the random tree model, increasingly labelled trees appear naturally as computer science data structures. In that context, they are named by *increasing trees*, or heap ordered trees (cf. e.g. [FS09]) and are directly related to search trees, and particularly to binary search trees (in the case of *binary increasing trees*). They appear also as the key notion in the analysis of Quicksort: the link between the data structure and the algorithm is well presented, for example, in [Drm09]. Using enumeration methods and *analytic combinatorics*, it has been proved that increasing trees are shorter and bushier than Catalan trees.

Our paper focuses on four families of increasing trees [MS95]. We measure the number of connected components obtained after having removed all the nodes which labels are smaller than a given value. Let us recall that this notion plays a fundamental rôle in many graph problems. For example, the graph-cut appears as a key tool in order to compute the minimum spanning

*This research was partially supported by the A.N.R. project *MAGNUM*, ANR 2010-BLAN-0204.

†Laboratoire d'Informatique de Paris-Nord, CNRS UMR 7030 - Institut Galilée - Université Paris-Nord, 99, avenue Jean-Baptiste Clément, 93430 Villetaneuse, France.
‡Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6, F-75005, Paris, France.
§CNRS, UMR 7606, LIP6, F-75005, Paris, France.

tree of a graph: Graham and Hell wrote a survey [GH85] about this problem. In some epidemics context (the initial motivation of Meir and Moon), the measure, we compute, corresponds to the average number of new sources, once the first nodes have been removed, because they are healthy again. Our approach uses a construction based on *tree-grafting*. We have developed similar ideas in [BGP12], in the easier case of non-labelled trees. In the context of increasing trees, an appropriate name for such an approach is *past-present-future method*. This name will be explained in the next section. During the study, we will emphasize an interesting relation to urn processes (e.g. [FDP06, Jan06]). This important link will be analysed in order to establish which urn processes could bring information in our context.

Let us finally highlight the fact that using our analytic combinatorics approach, we derive very *precise and exact quantitative results* in three out of four models, in particular, we give the exact distribution of trees according to their size and their cut-lengths. For the last model, we point out that our approach is very robust, and a much more involved proof gives the asymptotic of the average value for the parameter under consideration.

The papers is organized in the following way. Section 2 is devoted to detail the key notions in the context of increasing trees and then, it presents the main ideas and the fundamental method of the paper. Sections 3, 4 and 5 prove the main results about the three first tree models. In Section 6, we present an interesting link to urn theory. Furthermore, we explain the reasons why this approach is not sufficiently robust for our problem. In Section 7, we study the model of strict binary trees, that cannot be treated by urns. And finally, before the conclusion, we derive from our results the average complexity analysis of an algorithm of tree-labelling (Section 8).

2 Context

We are dealing with distinct tree models that are families of *increasing trees*. Let us recall that increasing trees are rooted labelled trees such that each sequence of labels from the root to a leaf is increasing. The reader will find many details about increasing trees in e.g. [Drm09, Section 1]. Our first goal is to derive a detailed analysis of the measure we are interested in and not only asymptotic results. The second goal is for our original application: the average complexity analysis of an algorithm of tree-labelling (Section 8).

The parameter, we analyse in this paper, appears naturally in *increasing trees*. Let us recall their definition. For a given tree-structure, either plane or not, the way of labelling it follows this constraint: the nodes are labelled by distinct integers in $\{1, \ldots, n\}$, where n corresponds to the *size* (i.e. the total number of nodes) of the tree, and the node-labelling is such that all label-sequences of branches from the root to a leaf are increasing.

We are now ready to define the parameter we are interested in:

Definition 1. *Let n be an integer and T an increasing tree of size n. Let $i \in \{1, \ldots, n\}$. The i-th* cut *of T is the subset of edges e_{j-k} from T, such that the associated nodes labelled j and k are satisfying $j \leq i < k$.*

Remark that the i-th cut (for $i \in \{1, \ldots, n-1\}$) defines a partition, of the tree, in two sets. The first one contains a single tree labelled by the integers from 1 to i and the second contains the largest subtrees of T whose node's labels are greater than i.

Definition 2. *Let n be an integer and T an increasing tree of size n. Let $i \in \{1, \ldots, n\}$. The* head *of the i-th cut corresponds to the tree with labels from 1 to i.*

In other contexts, like algebraic combinatorics (e.g. [CK98]) or concurrency theory [BGP13] the notion of head of a cut exists and is called there *admissible cut*. But in our context, this name would have been puzzling, in view of the natural notion of cut.

Definition 3. *Let n be an integer and T an increasing tree of size n. Let $i \in \{1, \ldots, n\}$. The* length *of the i-th cut is the number of edges in the cut.*

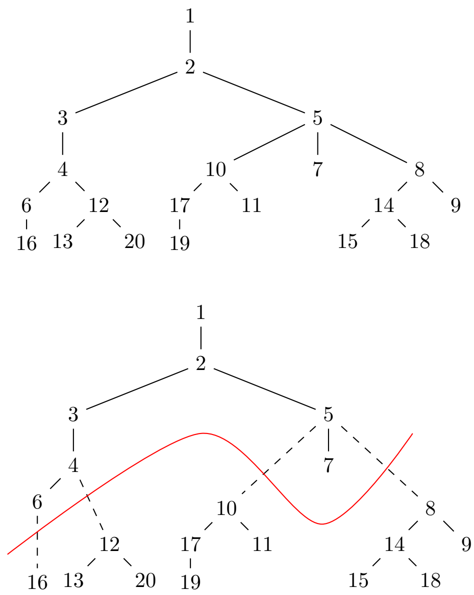

Figure 1: *An increasing tree (top) and the partition induced by its 7-th cut (bottom)*

See Figure 1 for an example of a cut. We have pictured the 7-th cut of T, the tree represented on the top. The 7-th cut corresponds to the subset $\{e_{6-16}, e_{4-12}, e_{5-10}, e_{5-8}\}$ of edges. It is of length 4.

The goal of our paper is the analysis of the cut-length in different models of increasing trees. We are studying four distinct models of trees. For the two first models, we are dealing with the classical binary and general plane trees. Then we study the increasing Cayley (non-plane) tree model and we conclude our study with strict binary plane tree model.

In order to analyse the cut-length in these contexts, we develop a combinatorial approach based on analytic combinatorics. With a precise specification of the trees, we are able to derive, in three cases, the distribution of the cut-length in the trees. These distributions allow us to compute the factorial moments of the cut-lengths: we compute the exact factorial moments formulae in one model. In the other cases we prove the mean value of the parameter (asymptotically for the worst case –strict binary tree–, or exactly for the others). From these results we can derive the asymptotic behaviour of the i-th cut-length when i is describing $\{1, \dots, n\}$. On Figure 2, we represent the limiting evolutions of the cut-lengths (in the central range), for our distinct models of increasing trees. In fact, we represent the ratio of the number of edges in the i-th cut according to the normalization i/n.

In order to conclude this section, let us give the common sketch for the proof of the cut-length behaviour.

First we define a specification for a trivariate generating function enumerating increasing trees according to their whole size, the sizes of their cut-heads and the length of each cut. The approach we have developed could be named by *past-present-future method*. In fact, for a tree, we characterize its cut-heads (*past*) then its cuts (*present*) and finally the subtrees grafted to the cut-heads (*future*). In order to define such a generating function, we decompose the complete objects according to some head of a cut and then we analyse the way to complete such a head in order to build the whole increasing tree. Since we are dealing with increasing trees, the *box operator* appears naturally and thus we will deal with differential equations satisfied by the generating functions. In all cases we are able to compute the closed form for the generating functions. Then a differentiation according to a specific variable (once or several times) and then by substituting it by 0, we compute the

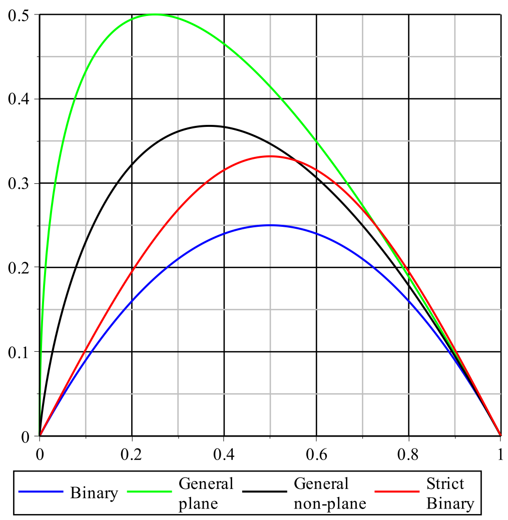

Figure 2: *Limiting evolution of the cut-lengths*

factorial moment of the cut-lengths. In order to obtain them explicitly, we are mainly using Lagrange inversion formula, and recurrences. In some specific cases, since our generating functions are holonomic, we are using some *Guess and Prove* strategy to obtain the results.

Almost all proofs need the use of a computer algebra system.

We are now ready to turn to the specific models.

3 Binary increasing trees

The underlying tree-structure of an increasing binary tree is an incomplete binary tree. Thus, the arity (or out-degree) of a node is either zero or one or two, and each child is either a left child or a right child of its parent. Thus, the combinatorial class of increasing binary trees is given by the following specification:

$$\mathcal{B}_{\mathcal{Z}} = \mathcal{Z}^{\square} \star (1 + 2 \cdot \mathcal{B}_{\mathcal{Z}} + \mathcal{B}_{\mathcal{Z}}^2).$$

The boxed product, denoted by $^{\square}\star$, consists in the labelled product with the constraint that all labels from the first factor (here $\mathcal{Z}$) are smaller than those of the second factor (here $(1+2\cdot\mathcal{B}_{\mathcal{Z}}+\mathcal{B}_{\mathcal{Z}}^2)$). The reader will find detailed information about this construction, for example in the book of Flajolet and Sedgewick [FS09].

Consequently, the corresponding exponential generating function $B(z)$ satisfies:

$$B(z) = \int_{v=0}^{z} 1 + 2B(v) + B(v)^2 \; dv.$$

Thus, $B(z) = z/(1-z)$ and the n-th coefficient $B_n = n!$, when $n > 0$.

Let us now introduce a trivariate generating function enumerating increasing binary trees.

Proposition 4. *Let $\mathcal{C}$ be the class enumerating the increasing binary trees. The class $\mathcal{C}$ follows the next specification:*

$$\mathcal{C} = \mathcal{Y}^{\square} \star \left(1 + 2 \cdot (\mathcal{C} + \mathcal{U} \times \mathcal{B}_{\mathcal{Z}}) + (\mathcal{C} + \mathcal{U} \times \mathcal{B}_{\mathcal{Z}})^2\right),$$

where $\mathcal{Y}$ is marking the nodes of the cut-heads, $\mathcal{Z}$ is marking the other nodes and $\mathcal{U}$ is marking the edges of the cuts.

Proof. Let us remark that the root of an object of $\mathcal{C}$ must be labelled with the smallest integer. If the root is not a leaf, then either it has a left child or a right one or two children. These children have either a root that belong to the head (and so are trees in $\mathcal{C}$) or do not belong to the head-cut: in that case they are nodes from a subtree in $\mathcal{B}_{\mathcal{Z}}$. Finally, note that the class $\mathcal{B}_{\mathcal{Z}}$ enumerates increasingly labelled trees and thus the class $\mathcal{C}$ is well-labelled. □□

The former specification allows to establish the exact distribution of binary increasing tree according to their cuts.

Theorem 5. *The number of binary increasing trees of size n such that the length of the i-th cut is k is exactly:*

$$\frac{(i+1)!\ i!\ (n-i)!\ (n-i-1)!}{k!\ (k-1)!\ (n-i-k)!\ (i+1-k)!}.$$

In Figure 3, the first numbers of binary increasing trees according to their size, to their cut and to their cut-length are presented.

n	i	k	nb. of trees
2	1	1	2
3	1	1	4
3	1	2	2
3	2	1	6
4	1	1	12
4	1	2	12
4	2	1	12
4	2	2	12
4	3	1	24
5	1	1	48
5	1	2	72
5	2	1	36
5	2	2	72
5	2	3	12
5	3	1	48
5	3	2	72
5	4	1	120

Figure 3: *The first numbers of binary increasing trees.*

Proof. Let $C(y,z,u)$ the generating function (exponential in y and z and ordinary in u) associated with $\mathcal{C}$, whose coefficient $C_{i,j,k}$ is the number of increasing binary trees of size $i+j$ whose i-th cut is of length k. From the specification, defined in Proposition 4, we derive the following partial differential equation satisfied by the generating function $C(y,z,u)$ of $\mathcal{C}$:

$$C(y,z,u) = \int_{v=0}^{y} \left(1 + C(v,z,u) + uB(z)\right)^2\ dv.$$

We easily deduce:

$$C(y,z,u) = \frac{y(1-2z+z^2+2uz-2uz^2+u^2z^2)}{1-uyz-y-2yz+uyz^2-yz^2-2z+z^2}.$$

In order to extract the coefficients of $C(y,z,u)$, a possible way consists in first extracting the coefficients (depending on z and u) of y^i. Then, using the Lagrange inversion [FS09, e.g. P. 127] on the result, we prove:

$$[y^i z^j u^k]C(y,z,u) = \binom{i+1}{k}\binom{j-1}{k-1}.$$

The theorem ensues by a change of variables. □□

Once the distribution of the trees is settled, we can derive information about the factorial moments.

Theorem 6. *Let n be a positive integer, and let T be a binary increasing tree of size n. Let $i \in \{1, \ldots, n\}$. The m-th factorial moment of the length of the i-th cut is denoted by $\mathbb{E}(N_{n,i}^{(m)})$ and the associated variance by $Var(N_{n,i})$. We get the exact formulae:*

$$\mathbb{E}(N_{n,i}^{(m)}) = \frac{(i+1)!\,(n-i)!\,(n-m)!}{(i-m+1)!\,(n-i-m)!\,n!},$$

$$Var(N_{n,i}) = \frac{(n-i)(n-i-1)(i+1)i}{n^2(n-1)}.$$

In particular, asymptotically, when n tends to infinity, in the central range $i = \alpha n$, with $0 < \alpha < 1$, the limiting distribution of N is Gaussian with parameters:

$$\mathbb{E}(N_{n,i}) := \mathbb{E}(N_{n,i}^{(1)}) \sim_{n\to\infty} \alpha(1-\alpha)n;$$

$$Var(N_{n,i}) \sim_{n\to\infty} \alpha^2(1-\alpha)^2 n.$$

Proof. The probability generating function of the length $N_{n,i}$ of the i-th cut in a tree of size n is:

$$\mathbb{E}(u^{N_{n,i}}) = \frac{\sum_{k=1}^{n} \binom{i+1}{k}\binom{n-i-1}{k-1}\ i!\ (n-i)!\ u^k}{n!}.$$

From this expression, we easily deduce all the factorial moments of $N_{n,i}$.
In order to compute the variance, since the factorial moment of order 2 of the length of the i-th cut is

$$\mathbb{E}(N_{n,i}(N_{n,i}-1)) = \frac{(n-i)\,(n-i-1)\,(i+1)\,i}{n\,(n-1)},$$

the variance can be extracted directly by using the next equation:

$$Var(N_{n,i}) = \mathbb{E}(N_{n,i}(N_{n,i}-1)) - \mathbb{E}(N_{n,i})(\mathbb{E}(N_{n,i})-1).$$

Thus, we get the stated result:

$$Var(N_{n,i}) = \frac{(n-i)\,(n-i-1)\,(i+1)\,i}{n^2\,(n-1)},$$

and the analysis in the central range follows immediately.
We can also estimate all the moments when n tends to infinity, we obtain:

$$\mathbb{E}(N_{n,i}^{(m)}) \sim \mathbb{E}((N_{n,i})^m) \sim \alpha^m(1-\alpha)^m n^m$$

This corresponds to the moments of a Gaussian law of parameters

$$\mathbb{E}(N_{n,i}) \sim_{n\to\infty} \alpha(1-\alpha)n;$$

$$Var(N_{n,i}) \sim_{n\to\infty} \alpha^2(1-\alpha)^2 n.$$

□□

In Figure 2, the evolution of the average lengths of the cuts is represented, more precisely, the curve corresponds to the limiting evolution when n tends to infinity. Let us remark that for a large part of the evolution, the average lengths of the cuts are equal to a positive ratio of the total size of the increasing trees.

Corollary 7. *Let n be a positive integer. The average length of the cuts issued of binary increasing trees of size n is:*

$$\frac{n^2+3n-4}{6n} =_{n\to\infty} \frac{n}{6}+\frac{1}{2}+O\left(\frac{1}{n}\right).$$

When n tends to infinity, the maximum of the average lengths of the cuts is equal to:

$$\frac{n}{4}+\frac{1}{2}+O\left(\frac{1}{n}\right),$$

and it is reached for the $n/2$-th cut.

Remark that we can compute the exact value of the maximum of the average lengths of the cuts. It can be derived by studying the parity of n and by noticing that $\mathbb{E}(N_{n,i}) = \mathbb{E}(N_{n,n-1-i})$, for $i \in \{1, \dots, n-2\}$.

4 General plane increasing trees

A general plane tree is a plane tree which nodes have an arity that belongs to $\mathbb{N}$. We organize the study of this model with an analogous approach as for the binary tree model.

Proposition 8. *Let $\mathcal{G}$ be the class enumerating general plane increasing trees. The class $\mathcal{G}$ follows the specification:*

$$\begin{aligned}&\mathcal{G} = \mathcal{Y}^{\square} \star \textsc{Seq}(\mathcal{G} + \mathcal{U} \times \mathcal{P}_{\mathcal{Z}}) \text{ and}\\ &\mathcal{P}_{\mathcal{Z}} = \mathcal{Z}^{\square} \star \textsc{Seq}(\mathcal{P}_{\mathcal{Z}}),\end{aligned}$$

where the class $\mathcal{P}_{\mathcal{Z}}$ denotes the class of plane increasing trees, $\mathcal{Y}$ is marking the nodes of the cut-heads, $\mathcal{Z}$ is marking all other nodes and $\mathcal{U}$ is marking the edges of the cuts.

The specification is directly obtained by adapting the binary case to the context of plane increasing trees. We thus get the following result.

Theorem 9. *The number of general plane increasing trees of size n such that the length of the i-th cut is k is exactly:*

$$\frac{2^{k+1-n}\,(2\,n-2\,i-k-1)!\;(2\,i+k-2)!}{(k-1)!\;(n-i-k)!\;(i-1)!}$$

In Figure 4, the first numbers of general plane increasing trees according to their size, to their cut and to their cut-length are presented.

Proof. Let $G(y,z,u)$ be the generating function (exponential in y and z and ordinary in u) associated to $\mathcal{G}$ (see Proposition 8), whose coefficient $G_{i,j,k}$ is the number of plane increasing trees of size $i+j$ whose i-th cut is of length k.

Since the class of plane increasing trees admits the following generating function $P_{\mathcal{Z}}(z) = 1-\sqrt{1-2z}$, an immediate integration of the following differential equation,

$$\frac{\partial G}{\partial y}(y,z,u) = \frac{1}{1-G(z,y,u)-u\left(1-\sqrt{1-2z}\right)},$$

gives:

$$\begin{aligned}G(y,z,u) = {} & 1-u(1-\sqrt{1-2z})\\ & -\sqrt{2u^2(1-z-\sqrt{1-2z})-2u(1-\sqrt{1-2z})+1-2y}.\end{aligned}$$

It remains to extract the coefficient we are interested in:

$$[y^i z^{n-i} u^k]G(y,z,u).$$

n	i	k	nb. of trees
2	1	1	1
3	1	1	1
3	1	2	2
3	2	1	3
4	1	1	3
4	1	2	6
4	1	3	6
4	2	1	3
4	2	2	12
4	3	1	15
5	1	1	15
5	1	2	30
5	1	3	36
5	1	4	24
5	2	1	9
5	2	2	36
5	2	3	60
5	3	1	15
5	3	2	90
5	4	1	105

Figure 4: *The first numbers of general plane increasing trees.*

We proceed in two steps. First, we guess the expression by using the fact that the factorizations into integers of the first coefficients contain only small factors. In the second step, we prove that the expression is valid. We proceed as follows:

- We compute an algebraic equation for $G(y, z, u)$:

$$\begin{aligned} G^4 + (-4 + 4\,z)\,G^3 + \left(4\,y + 8\,u^2 z - 8\,u + 4\right) G^2 \\ + (8\,yu - 8\,z)\,G + 4\,y^2 = 0. \end{aligned}$$

- As $G(y, z, u)$ is holonomic, we deduce recurrences for the coefficients. For instance the main one is:
$$G_{i,j+1,k} = \frac{2j-k}{j+1} G_{i,j,k} + \frac{2i-k-2}{j+1} G_{i,j,k-1}.$$

- We check that the closed form expression satisfies all recurrences. □

□

Theorem 10. *Let n be a positive integer, and let T be a plane increasing tree of size n. Let $i \in \{1, \ldots, n\}$. The average length of the i-th cut, denoted by $\mathbb{E}(N_{n,i})$, is equal to:*

$$\mathbb{E}(N_{n,i}) = \frac{2\,\Gamma\,(i+1/2)\,\Gamma\,(n)}{\Gamma\,(n-1/2)\,\Gamma\,(i)} - 2i + 1.$$

In particular, asymptotically, when n tends to infinity, in the central range $i = \alpha n$, with $0 < \alpha < 1$, we have

$$\mathbb{E}(N_{n,i}) \sim_{n\to\infty} 2\sqrt{\alpha}(1-\sqrt{\alpha})n.$$

Proof. Let us define

$$E(y,z) := \left(\frac{\partial G}{\partial u}(y,z,u)\right)_{|u=1},$$

and derive immediately:

$$E(y,z) = -1 + \sqrt{1-2z} - \frac{1-2z-\sqrt{1-2z}}{\sqrt{1-2z-2y}}.$$

Let us denote, by $e_{i,j}$, the coefficient of the monomial $y^i z^j$ in $E(y,z)$, then after a technical computation, we get:

$$\begin{aligned} e_{i,j} =& \frac{2^{i+j}\ \Gamma(i+1/2)\ \Gamma(i+j)}{\sqrt{\pi}\ \Gamma(i)\Gamma(i+1)\ \Gamma(j+1)} \\ &- \frac{2^{i+j}(2i-1)\ \Gamma(i+j-1/2)}{2\sqrt{\pi}\ \Gamma(i+1)\ \Gamma(j+1)}. \end{aligned}$$

But, the number, p_n, of plane increasing trees of size n is

$$[z^n](1-\sqrt{1-2z}) = \frac{2^{n-1}\ \Gamma(n-1/2)}{\sqrt{\pi}\ \Gamma(n+1)}.$$

Thus to obtain the curve of the evolution of the average lengths of the cuts, we just compute

$$\mathbb{E}(N_{n,i}) = i!\ (n-i)!\ \frac{e_{i,n-i}}{p_n}.$$

□□

Note that using the complete distribution of trees according to their cut-lengths (Theorem 9), we should be able to compute all the factorial moments of $N_{n,i}$. However it is a challenge to obtain the average value: we leave these computations out.

In order to describe the limiting curve (cf. Figure 2) of the evolution of the cut-lengths, we take $i = \alpha n$ and then analyse the limit of $\mathbb{E}(N_{n,i})$, when n tends to infinity.

Corollary 11. *Let n be a positive integer. Asymptotically, when n tends to infinity, the average length of the cuts issued of general plane increasing trees of size n is equal to:*

$$\frac{n}{3} + O\left(\frac{1}{n}\right).$$

When n tends to infinity, the maximum of the average length of the cuts is equal to:

$$\frac{n}{2} + \frac{1}{8} + O\left(\frac{1}{n}\right).$$

It is reached for the $n/4$-th cut.

5 General non-plane increasing trees

Usually, general non-plane labelled trees are called *Cayley trees.* So the following section is dealing with increasing Cayley trees.

Proposition 12. *Let $\mathcal{F}$ be the class enumerating general non-plane increasing trees. The class $\mathcal{F}$ follows the next specification:*

$$\begin{aligned} &\mathcal{F} = \mathcal{Y}^{\square} \star \textsc{Set}(\mathcal{F} + \mathcal{U} \times \mathcal{Q}_{\mathcal{Z}}) \text{ and} \\ &\mathcal{Q}_{\mathcal{Z}} = \mathcal{Z}^{\square} \star \textsc{Set}(\mathcal{Q}_{\mathcal{Z}}), \end{aligned}$$

where the class $\mathcal{Q}_{\mathcal{Z}}$ denotes the class of plane increasing trees, $\mathcal{Y}$ is marking the nodes of the cut-heads, $\mathcal{Z}$ is marking all other nodes and $\mathcal{U}$ is marking the edges of the cuts.

Theorem 13. *The number of general non-plane increasing trees of size n such that the length of the i-th cut is k is exactly:*

$$i!\, i^{k-1} \left| s\left(n-i,k\right) \right|,$$

where $|s(\cdot,\cdot)|$ denotes Stirling numbers of the first kind.

In Figure 5, the first numbers of general non-plane increasing trees according to their size, to their cut and to their cut-length are presented.

n	i	k	nb. of trees
2	1	1	1
3	1	1	1
3	1	2	1
3	2	1	2
4	1	1	2
4	1	2	3
4	1	3	1
4	2	1	2
4	2	2	4
4	3	1	6
5	1	1	6
5	1	2	11
5	1	3	6
5	1	4	1
5	2	1	4
5	2	2	12
5	2	3	8
5	3	1	6
5	3	2	18
5	4	1	24

Figure 5: *The first numbers of general non-plane increasing trees.*

Theorem 14. *Let n be a positive integer, and let T be a general non-plane increasing tree of size n. Let $i \in \{1,\dots,n\}$. The average length of the i-th cut, denoted by $\mathbb{E}(N_{n,i})$, is equal to:*

$$\mathbb{E}(N_{n,i}) = i\left(H\left(n-1\right) - H\left(i-1\right)\right),$$

where H is the harmonic function.
In particular, asymptotically, when n tends to infinity, in the central range $i = \alpha n$, with $0 < \alpha < 1$, we have

$$\mathbb{E}(N_{n,i}) \sim_{n\to\infty} (-\alpha \ln \alpha) n.$$

The limiting curve of the evolution of the cut-lengths is represented on Figure 2.

Corollary 15. *Let n be a positive integer. Asymptotically, when n tends to infinity, the average length of the cuts issued of general non-plane increasing trees of size n is equal to:*

$$\frac{n+1}{4} + O\left(\frac{1}{n}\right).$$

When n tends to infinity, the maximum of the average length of the cuts is equal to:

$$\frac{n}{\exp(1)} + \frac{1}{8} + O\left(\frac{1}{n}\right),$$

and it is reached for the $n/\exp(1)$-th cut.

6 Analysing cuts via urn processes

The three models we have detailed have an important common behaviour: *whatever the head structures (of a given size i), the numbers of grafts to complete them into full increasing trees (of size n) is the same.* Thus, it is a known result that the evolution of some families of increasing trees can be studied in the context of urn processes. Such a link is given, for example, in Morcrette's dissertation [Mor13, Chapter 7] about general plane increasing trees (named as PORT).

However, the cut-length parameter is not directly linked to the evolution of an increasing tree. We will give, in the following, the ideas to relate urn evolution to cut-length in increasing trees.

6.1 Binary increasing trees

The tree at the top of Figure 6 represents a head of size 7 and each of the bullet could contain a node of the complete increasing tree. In other terms, in each bullet, we could graft subtrees to complete the tree. In order to let the head evolve, we must choose one of the bullet that will be replace by 8. The new dashed edge between 6 and 8 counts for the cut-length. Then for the node 9, either one chooses a bullet or a square. In the first case, the new edge $\cdots - 9$ counts for the cut-length; in the second case it does not. For example, on the last tree of Figure 6, we have drawn an evolution (after three steps) from the first head. Both edges $6 - 8$ and $7 - 9$ count for the cut-length, but the edge $8 - 10$ does not.

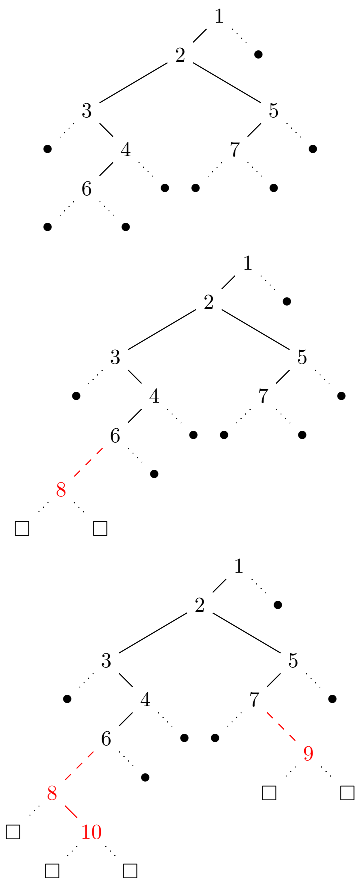

Figure 6: *A 7-th cut of a binary increasing tree, with the possible grafts for the next nodes (top), one of its evolutions 1 step later (middle) and one of its evolutions 3 steps later (bottom)*

Let us now make explicit the evolution rules in order to compute the cut-length. Let H be a binary head of size i. Whatever the tree-structure of H, it contains $i+1$ bullets. So the initial urn contains $i+1$ bullets and 0 square. Then we choose one ball from the urn, and we adapt our behaviour according to one of the following rule:

- if the ball represents a bullet, then it is replaced by two square-balls;
- if the ball represents a square, then it is replaced by two square-balls.

After $n-i$ repetitions of sampling, the initial composition of the head H is replaced by a composition of an increasing tree of size n. The difference between $i+1$ and the number of bullet-balls it contains at the end, corresponds to the cut-length of this tree. It remains to compute the average of this difference to obtain the average length of the i-th cut of binary trees of size n.

In the context of urn processes [FDP06], such an urn is encoded by the vector $[-1, 2, 0, 1]$ and is studied in the section about semi-sacrificials urns. In that paper, it is in particular shown that the moments (of the cut-length parameter) of such an urn are of hypergeometric form and that it tends to a Gaussian limit. Since we have studied, in some sense, the particular urn $[-1, 2, 0, 1]$ in Section 3 we have established precisely the values of the parameters we are interested in.

6.2 General plane increasing trees

Let us now explicit the evolution rules in order to compute the cut-length in general plane trees. Let H be a general head of size i. Whatever the tree-structure of H, it contains $2i-1$ bullets. So the initial urn contains $2i-1$ bullets and 0 square. Then we choose one ball from the urn, and we adapt our behaviour according to one of the following rule:

- if the ball represents a bullet, then it is replaced by two bullet-balls and one square-ball;
- if the ball represents a square, then it is replaced by three square-balls.

After $n-i$ steps, the initial composition of the head H is replaced by a composition of a general increasing tree of size n. The difference between the numbers of bullet-balls at the end the number of bullet-balls at the beginning $(2i-1)$ corresponds to the cut-length of this tree. It remains to compute the average of this difference to obtain the average length of the i-th cut of general plane trees of size n.

Such an urn is encoded by the vector $[1, 1, 0, 2]$ and is studied in independently in [FDP06] in the section about triangular urns and in the paper of Janson [Jan06]. There, it is in particular shown that the moments (of the cut-length parameter) of such an urn are of hypergeometric form and that it does not tend to a Gaussian limit but to a local limit law whose density is expressed in terms of Mittag-Leffler functions.

6.3 General non-plane increasing trees

Let us now explicit the evolution rules in order to compute the cut-length in general non-plane trees. Let H be a general head of size i. Whatever the tree-structure of H, it contains i bullets. So the initial urn contains i bullets and 0 square. Then we choose one ball from the urn, and we adapt our behaviour according to one of the following rule:

- if the ball represents a bullet, then it is replaced by one bullet-ball and one square-ball;
- if the ball represents a square, then it is replaced by two square-balls.

After $n-i$ steps, the initial composition of the head H is replaced by a composition of a general increasing tree of size n. However, in that context, we cannot determine the cut-length by the previous urn process. The 2-dimensional urn does not encode sufficiently information for our problem. This parameter requires a 3-dimensional urn such that a third type of balls represent the cut-length. The urn is encoded by $[0, 1, 0, 0, 0, 1, 0, 0, 1]$ and it is still balanced. Such an urn has neither been studied by [FDP06], nor by [Jan06]. To our knowledge, there is no paper about this model. However, the paper [FDP06] deals with some 3-dimensional triangular urns so we could maybe adapt their analysis to study this urn.

6.4 Limitations of urn models

In the last subsection, we have noticed that some models do not enter in the classical and widely studied of 2-dimensional urns. The reason is because it needs another parameter to encode the cut-length in the evolution process. However, there is another and a much more fundamental limitation of urn processes. In the three cases above, we highlight the fact that the tree-structure of the head does not matter for the initial state of the urn. But there are increasing tree models that does not follow this constraint: in particular strict d-ary models. In such models, different head-structures (of the same size i) do not give the same number of grafts for the new subtrees. In the next section, we study the cut-length in strict binary increasing trees.

7 Strict binary increasing trees

We define a *strict binary increasing tree* to be a plane binary increasing tree whose nodes have arity two or zero. Thus, the arity one is forbidden in such a tree. This model cannot be encoded properly in an urn process, as explained above, and this could be a reason for the jump in the mathematical difficulty of the analysis of the cut-length parameter in this context. In fact, we are not able to give an exact form for the factorial moments anymore. However an intricate analysis, based on our robust past-present-future approach, let us still derive the asymptotic of the average cut-length.

The approach is the same as the previous ones: the strict increasing binary trees satisfy the specification $\mathcal{T}_{\mathcal{Z}} = \mathcal{Z} + \mathcal{Z}^{\square} \star \mathcal{T}_{\mathcal{Z}}{}^2$. Thus the exponential generating function is solution of the differential equation $T'(z) = 1 + T(z)^2, T(0) = 0$. Thus, we get the classical solution $T(z) = \tan(z)$.

Proposition 16. *Let $\mathcal{E}$ be the class of strict binary increasing trees. It follows the specification:*

$$\mathcal{E} = \mathcal{Y}^{\square} \star \left(1 + (\mathcal{E} + \mathcal{U} \times \mathcal{T}_{\mathcal{Z}})^2\right),$$

where $\mathcal{Y}$ is marking the nodes of the cut-head, $\mathcal{Z}$ is marking the other nodes and $\mathcal{U}$ is marking the cut-length. Thus

$$E(y, z, u) = \frac{u \tan(z) + \tan(y)}{1 - u \tan(z) \tan(y)} - u \tan(z).$$

Thus the term $i!\ j!\ e_{i,j,k}$ that counts the number of trees of size $i + j$, whose i-th cut is exactly k, is equal to $[y^i z^j u^k]\ i!\ j!\ E(y, z, u)$. Nevertheless, it seems not to be realistic to expect a closed form (without any sum). Some intuition is given since the urn process cannot encode our model. In fact, some sum appears according to the different head-structures of size i.

In Figure 7, the first numbers of strict binary increasing trees according to their size, to their cut and to their cut-length are presented.

We concentrate our attention on the mean value of the length of the i-th cuts.

Theorem 17. *Let n be a positive integer, and let T be a strict binary increasing tree of size n. Let $i \in \{1, \ldots, n\}$. The average length of the i-th cut, denoted by $\mathbb{E}(N_{n,i})$, is in the central range $i = \alpha n$, with $0 < \alpha < 1$, asymptotically equal to:*

$$\mathbb{E}(N_{n,i}) \sim_{n\to\infty} \frac{1}{4}\, \frac{\mathrm{e}^2 \sin(\pi\,\alpha/2)\,(1 + \cos(\pi\,\alpha))}{\pi^{3/2} \cos(\pi\,\alpha/2)}\ n.$$

Proof. In order to prove the result, as previously, let us consider

$$M(y, z) := \frac{\partial C(y, z, u)}{\partial u}|_{u=1}.$$

We are going to apply multivariate saddle point analysis to evaluate $[y^i z^j] M(y, z)$. Precisely, we are going to deal with the generalisation of H-admissibility defined in [GM06]. Indeed, it is easy to show using the closure properties that the functions $\tilde{M}(y, z) := M(y, \sqrt{z}\,) - M(-y, \sqrt{z}\,)$ and

n	i	k	nb. of trees
3	1	2	2
3	2	1	2
5	1	2	16
5	2	1	4
5	2	3	12
5	3	2	16
5	4	1	16
7	1	2	272
7	2	1	32
7	2	3	240
7	3	2	128
7	3	4	144
7	4	1	32
7	4	3	240
7	5	2	272
7	6	1	272

Figure 7: *The first numbers of strict binary increasing trees.*

$\bar{M}(y,z) := (M(y,\sqrt{z}\,) + M(-y,\sqrt{z}\,))/\sqrt{z}$ (note that $M(y,z)$ is periodic of period 2 in z) are both H-admissible (as sum and product of H-admissible functions). So, we can conclude that

$$[y^i z^j]M(y,z) \sim 2 \times \frac{M(y_{i,j}, z_{i,j})}{2\pi |V| y_{i,j}^i z_{i,j}^j},$$

where $y_{i,j}$ and $z_{i,j}$ are the solution of the saddle point equations:

$$\frac{\dfrac{\partial M(y,z)}{\partial y}}{M(y,z)} = i, \qquad \frac{\dfrac{\partial M(y,z)}{\partial z}}{M(y,z)} = j$$

and $|V|$ is the following determinant:

$$|V| = \det\left(\left[\frac{\frac{\partial M(y,z)}{\partial y}}{M(y,z)}, \frac{\frac{\partial M(y,z)}{\partial z}}{M(y,z)}\right] \times \begin{pmatrix} y & 0 \\ 0 & z \end{pmatrix}\right).$$

Let us notice that the extra factor 2 in the beginning is just due to correct the 2-periodicity of $M(y,z)$. We start by solving asymptotically when i and j are large the saddle point equation, we get:

$$y_{i,j} = \frac{1}{2}\,\frac{\pi\, i}{i+j} + \frac{\beta}{i} + O\left(\frac{1}{(i+j)^2}\right)$$

and

$$z_{i,j} = \frac{1}{2}\,\frac{\pi\, j}{i+j} - \frac{\pi\, i + \beta\, i + \beta\, j}{(i+j)\, n} + O\left(\frac{1}{(i+j)^2}\right),$$

where β depends[1] from i, j and n. It just remains to reach an equivalent for $2 \times \dfrac{M(y_{i,j}, z_{i,j})}{2\pi |V| y_{i,j}^i z_{i,j}^j}$. Even, if it seems to be automatic, the size of the expression needs to carefully manage the problem. Finally, we get:

$$\begin{aligned}
[y^i z^j] M(y,z) \sim\ & 2^{-\frac{1}{2}+i+j} \left(1 + \cos\left(\frac{\pi\, j}{i+j}\right)\right) \\
& \times \sin\left(\frac{\pi\, j}{2(i+j)}\right) \left(\frac{i+j}{i}\right)^{i+j+3/2} \\
& \times \pi^{-i-j-3} \left(\frac{i}{j}\right)^{j+1/2} \mathrm{e}^2 i^{-3/2} \\
& \times \left(\cos\left(\frac{\pi\, j}{2(i+j)}\right)\right)^{-1}.
\end{aligned}$$

Thus the theorem follows directly after a change of variable and a renormalization by $[z^n]\tan(z)$. $\square\square$

Corollary 18. *Let n be a positive integer. Asymptotically, when n tends to infinity, the average length of the cuts issued of strict binary trees of size n is equal to:*

$$\frac{\mathrm{e}^2}{2\pi^{5/2}} n + O(1).$$

When n tends to infinity, the maximum of the average length of the cuts is equal to:

$$\frac{\mathrm{e}^2}{4\pi^{3/2}} n + O(1).$$

It is reached for the $n/2$-th cut.

8 Increasingly labelling of a tree-structure

A paper [BRS12], on Boltzmann samplers, is describing efficient algorithms in order to generate, in particular, increasing trees uniformly at random. However, their approach gives the tree-structure without the labelling. As it will be explained: the cut-length parameter is linked to the increasingly labelling of a tree-structure.

In the context of order theory, the labelling of a structure such that it satisfies the increasing constraint plays an important rôle because it corresponds to determine a linear extension. If the poset under consideration is a tree-like partial order, like in paper [Atk90], considering the way we can label the tree-structure is fundamental.

[1] The value of β satisfies:

$$\begin{aligned}
\beta = -\frac{\pi i^2}{4} \Bigg(& -\pi j + 4\sin\left(\frac{1}{2}\frac{\pi j}{i+j}\right)\cos\left(\frac{1}{2}\frac{\pi j}{i+j}\right) i + 2\pi j \sin\left(\frac{1}{2}\frac{\pi j}{i+j}\right) \\
& \cdot \cos\left(\frac{1}{2}\frac{\pi j}{i+j}\right)\sin\left(\frac{\pi j}{i+j}\right) + 4\sin\left(\frac{1}{2}\frac{\pi j}{i+j}\right)\cos\left(\frac{1}{2}\frac{\pi j}{i+j}\right) j \cos\left(\frac{\pi j}{i+j}\right) \\
& + 4\sin\left(\frac{1}{2}\frac{\pi j}{i+j}\right)\cos\left(\frac{1}{2}\frac{\pi j}{i+j}\right)\cos\left(\frac{\pi j}{i+j}\right) i - \cos\left(\frac{\pi j}{i+j}\right)\pi j \\
& + 4\sin\left(\frac{1}{2}\frac{\pi j}{i+j}\right)\cos\left(\frac{1}{2}\frac{\pi j}{i+j}\right) j \Bigg) \sin\left(\frac{1}{2}\frac{\pi j}{i+j}\right)^{-1} \cos\left(\frac{1}{2}\frac{\pi j}{i+j}\right)^{-1} \\
& \cdot \Bigg(i^3 + 3 j i^2 + \cos\left(\frac{\pi j}{i+j}\right) j^3 + 3 j^2 i + \cos\left(\frac{\pi j}{i+j}\right) i^3 + j^3 + 3\cos\left(\frac{\pi j}{i+j}\right) j i^2 \\
& + 3\cos\left(\frac{\pi j}{i+j}\right) j^2 n \Bigg)^{-1}.
\end{aligned}$$

In paper [BGP12][2] we have described an algorithm, based on the hook length formula in trees, in order to increasingly label the tree uniformly at random. The algorithm is based on a sampler in a dynamic multiset: the approach is completely described in the extended paper [BGP14]. This algorithm constructs a well balanced tree in order to represent the dynamic multiset. And now comes the link with cut lengths in increasing trees. In fact, the evolution of the average cut-lengths in increasing trees corresponds to the evolution of the average size of the data-structure that encodes the multiset. And we are able now to state that on average, whatever the increasing tree model among the four we have presented, the data structure contains a positive ratio of the total size of the tree that is being labelling.

The time-complexity of the algorithm that increasingly labels the tree is directly related to the size of the multiset. In [BGP12], we easily proved that the worst case of the algorithm is in $\Theta(n \log n)$. The logarithmic factor is given by the well balanced tree depth. Thus, in order to analyse the time-complexity in the average case, we should compute the arithmetic mean (over all $i \in \{1, \dots, n\}$) of the arithmetic mean (over all trees of size n) of the logarithm of their i-th cuts. This measure is equal to the arithmetic mean of the logarithm of the geometric mean over the lengths of the i-th cuts.

Corollary 19. *Let take one of the four tree models we have presented. The algorithm that increasingly labels a tree-structure of size n, described in [BGP12], has an average time-complexity equal to $\Theta(n \log n)$.*

Proof. We present a proof for the general plane increasing tree model. However the proof can be easily extended to the other three models.

Using Theorem 10, the average value $\mathbb{E}(N_{n,i})$ is decreasing for $i \in \{n/4, \dots, n/2\}$. The value $\mathbb{E}(N_{n,n/2}) = (\sqrt{2} - 1)n$. For $i \in \{n/4, \dots, n/2\}$, there exists two constants α and β, such that at least $n!/\alpha$ trees of size n (among $n!$ trees) have their i-th cut of length at least n/β.
By a *reductio ad absurdum* argument we obtain the absurd conclusion: $\mathbb{E}(N_{n,i})$ is equal to $o(n)$, and thus is smaller than $(\sqrt{2} - 1)n$. Consequently, the geometric mean (over the trees of size n) of their i-th cut, denoted by $G_{n,i}$ satisfies:

$$G_{n,i} \geq \left(1^{(1-\alpha)n/\alpha} \cdot \left(\frac{n}{\beta}\right)^{n/\alpha} \right)^{1/n} = \left(\frac{n}{\beta}\right)^{1/\alpha}.$$

We conclude that the arithmetic mean A_n of the logarithms of the $G_{n,i}$ satisfies:

$$A_n \geq \frac{1}{n} \cdot \frac{n}{4} \cdot \frac{1}{\alpha} \log \left(\frac{n}{\beta}\right) = \Theta(\log n).$$

Finally, using the worst case upper bound, we conclude that the average time-complexity of the algorithm is $\Theta(n \log n)$. □□

9 Conclusion

Using our robust past-present-future method, we are able to analyse a natural parameter in increasing trees. We establish Theorems 5, 9, 13 that define the distribution of trees according to their head-cuts and their cut-lengths. Since we get the distributions, now an interesting study would be to prove the results just using combinatorial arguments. Another perspective would consist in giving an expression for computing the distribution in the case of strict binary trees.

Finally, let us note that the past-present-future approach can be easily adapted to other context partially increasing trees (cf. e.g. [BGP13]), or unlabelled trees.

[2] A largely extended and detailed version of [BGP12] is currently under submission, a preprint is available online [BGP14].

References

[Atk90] M. D. Atkinson. On computing the number of linear extensions of a tree. *Order*, 7:23–25, 1990.

[BA99] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[BGP12] O. Bodini, A. Genitrini, and F. Peschanski. Enumeration and random generation of concurrent computations. In *DMTCS AofA'12 proceedings*, pages 83–96, 2012.

[BGP13] O. Bodini, A. Genitrini, and F. Peschanski. The Combinatorics of Non-determinism. In *FSTTCS*, volume 24, pages 425–436, 2013.

[BGP14] O. Bodini, A. Genitrini, and F. Peschanski. Enumeration and random generation of concurrent computations. In *arXiv/1407.1873*, page Under Submission, 2014.

[BRS12] O. Bodini, O. Roussel, and M. Soria. Boltzmann samplers for first-order differential specifications. *Discrete Applied Mathematics*, 160(18):2563–2572, 2012.

[CK98] A. Connes and D. Kreimer. Hopf algebras, renormalization and noncommutative geometry. *Comm. Math. Phys.*, 199(1):203–242, 1998.

[Drm09] M. Drmota. *Random trees.* Springer, Vienna-New York, 2009.

[FDP06] P. Flajolet, P. Dumas, and V. Puyhaubert. Some exactly solvable models of urn process theory. In *Fourth Colloquium on Mathematics and Computer Science*, pages 59–118, 2006.

[FS09] P. Flajolet and R. Sedgewick. *Analytic Combinatorics.* Cambridge University Press, 2009.

[GH85] R. L. Graham and Pavol Hell. On the history of the minimum spanning tree problem. *IEEE Ann. Hist. Comput.*, 7(1):43–57, January 1985.

[GM06] B. Gittenberger and J. Mandlbuger. Hayman Admissible Functions in Several Variables. *Electronic Journal of Combinatorics*, 13(1), 2006.

[Jan06] S. Janson. Limit theorems for triangular urn schemes. *Probability Theory and Related Fields*, 134:417–452, 2006.

[MM74] A. Meir and J.W. Moon. Cutting down recursive trees. *Mathematical Biosciences*, 21(3–4):173 – 181, 1974.

[MM78] A. Meir and J. Moon. On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, 30:997–1015, 1978.

[Moo74] J. Moon. The distance between nodes in recursive trees. In *London Math. Soc. Lecture Note Ser.*, volume 13, pages 125–132, 1974.

[Mor13] B. Morcrette. *Combinatoire Analytique et Modèles d'Urnes.* PhD thesis, UPMC, `http://perso.crans.org~morcrette`, 2013.

[MS95] H. M. Mahmoud and R. T. Smythe. A Survey of Recursive Trees. *Theo. Probability and Mathematical Statistics*, 51:1–37, 1995.

[Sor06] D. Sornette. *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools (Springer Series in Synergetics).* Springer, 2006.