Average-case Analysis on Simple Families of Trees Using a Balanced Probability Model*

R. Casas J. Díaz C. Martínez
Dept. Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
Pau Gargallo 5, 08028-Barcelona

Abstract

In this paper we study a balanced probability model defined over simple families of trees, which from the point of doing statistics on trees, behave in quite a different way from the uniform distribution.

Using the new model, the analysis itself is more complex, but feasible. We illustrate our point by working out some particular simple cases of study : the computation of occupancy (a measure of the degree of balancing in trees), and the average size of the intersection of two m-ary trees. The development of the last analysis involves the use of sophisticated tools, such as Riemman's method to solve partial differential equations.

These average complexities are then compared with the ones obtained if the uniform model is assumed.

Introduction

e uniform probability model for the input is a current assumption in many average-case analysis, : to several reasons. Among others there are many applications where this model accurately ›ugh approximates the actual probability distribution of the input. Moreover, the assump- ı has been historically used to cope with uncertainty situations, such as the one that arises if actual distribution is unknown. On the other hand, because of the limitations of our tech- ues [Kar86, Sed83], only in a few cases we are able to carry on the analysis for probabilistic dels other than the simplest : the uniform probability model. Finally, even if the derived results ng the uniform model might not always be considered of relevance in absolute terms, they can of much value in order to compare the performance of different algorithms, which solve the en problem.

Nevertheless, it would be interesting to do these analysis for other reasonable probability dels, and compare the differences, if any, existing between the average complexities deduced the different models. Also, it would be interesting to compare "discriminative" powers of se models, that is, how different are the relative comparisons of the complexities of distinct orithms.

*This research was partially supported by the ESPRIT BRA Program of the EC under contract no. 3075, ject ALCOM.

In this paper, we propose what we call the ***balanced probability model*** for various families of trees. In this model, probabilities are high for well balanced trees and low for those which are poorly balanced.

This model is introduced in Section 2, where we also show the particular models for binary trees, m-ary trees, etc. and how can be defined the probability of a r-tuple of trees drawn from a given family of trees.

It turns out that the balanced probability model corresponding to binary trees coincides with that of the binary search trees, when the permutations from which they are built are assumed to be equally likely.

Sections 3 and 4 are devoted to show the kind of problems which arise in the average-case analysis of some algorithms and characteristics of the trees under the proposed model, as well as the ready tools to try to solve them. The first one covers the analysis of occupancy under the new assumption. Occupancy is a complexity measure of the "completeness" of trees, being maximum for complete trees and minimum for linear-list like trees. Our second case of study deals with the average size of the intersection of m-ary trees. While these analysis lead to algebraic functional relations for the generating functions defined if the uniform model is adopted, differential (partial and ordinary) equations arise for the balanced model.

Finally, in Section 5 main conclusions and further research are discussed.

2 The Balanced Probability Model

Let S be a set of symbols and $\nu : S \longrightarrow \mathbb{N}$ an arity application defined over S. Let $s(T_1, \ldots, T_{\nu(s)})$ be the tree whose root is labelled s and its subtrees are left-to-right : $T_1, \ldots, T_{\nu(s)}$.

Then the set of trees $\mathcal{F}$ recursively defined by

$$\begin{array}{rl} \forall s \in S & \nu(s) = 0 \Longrightarrow s \in \mathcal{F} \\ \forall s \in S \quad \forall T_1, \ldots, T_{\nu(s)} \in \mathcal{F} & s(T_1, \ldots, T_{\nu(s)}) \in \mathcal{F} \end{array}$$

is a *simple family of trees*, generated by S, if and only if

$$\exists M \in \mathbb{N} \quad \forall n \quad \#\nu^{-1}(n) \leq M$$

where $\nu^{-1}(n)$ denotes the set of symbols of arity n.

Simple families of trees were first defined in [MM78], but the definition and notation used here is that of [Ste84].

Furthermore, we shall consider probability distributions over the sets of symbols of the same arity. Let $p(s)$ be the probability of the symbol $s \in S$. We impose that,

$$\forall k \geq 0 \quad \sum_{s \in \nu^{-1}(k)} p(s) = 1, \qquad \text{if } \nu^{-1}(k) \neq \emptyset$$

If $\nu(s) = 0$ then we say that s is a leaf (0-ary symbol). The size of a tree T is the total number of nodes it contains, and will be denoted $|T|$. In the particular case of m-ary trees ($m \geq 1$) we define the size to be the number of internal nodes (those which are not leaves) since the total number of nodes depends only of the number of internal nodes.

Now, we are ready to define the balanced probability model. To start with, the weight measure w over $\mathcal{F}$ is recursively defined by

finition 2.1

$$w(T) \stackrel{\text{def}}{=} \begin{cases} p(s) & \text{if } T = s \text{ and } \nu(s) = 0 \\ p(s) \cdot w(T_1) \cdots w(T_k) \cdot \frac{1}{|T|} & \text{if } T = s(T_1, \ldots, T_k) \text{ and } \nu(s) = k \end{cases}$$

And in order to obtain a probability measure over $\mathcal{F}$, we define the probability of a tree T as

finition 2.2

$$p(T) \stackrel{\text{def}}{=} \frac{w(T)}{\sum_{|t|=|T|} w(t)}$$

cing it to verify Kolmogorov's axioms.

Note that $p(T)$ greatly varies depending on how balanced the tree is, being maximum for nplete trees and minimum for linear list-like trees (see Figure 1).

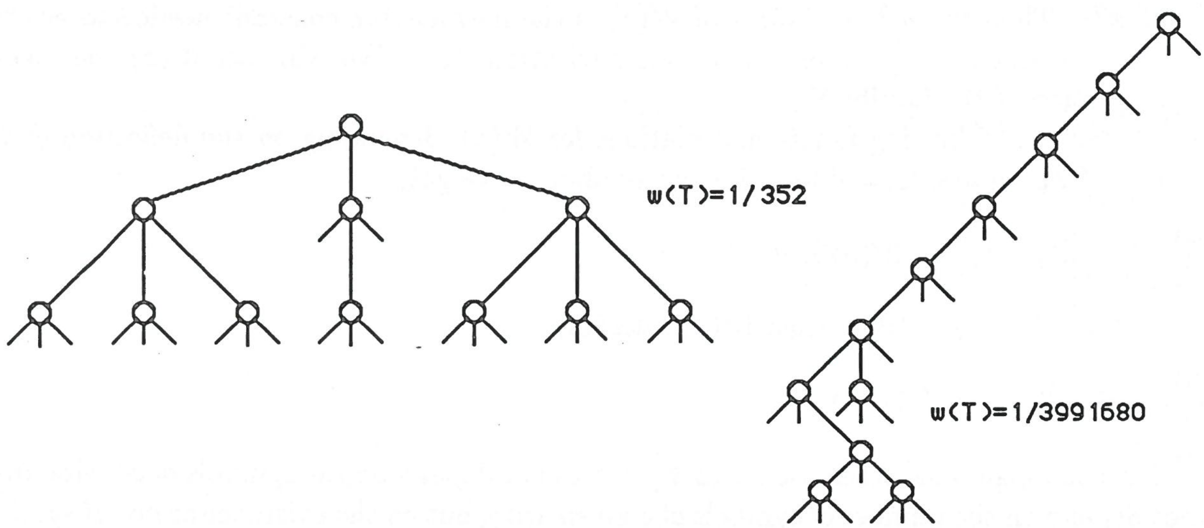

Figure 1: The Balanced Probability Model : An Example.

The above definitions can be extended to k-tuples of trees from a simple family by

finition 2.3

$$w(T_1, \ldots, T_k) \stackrel{\text{def}}{=} w(T_1) \cdots w(T_k)$$

l then define a probability measure over k-tuples $(T_1, \ldots, T_k)$ of size $|T_1| + \cdots + |T_k| = n$ as it been done previously,

finition 2.4

$$p(T_1, \ldots, T_k) \stackrel{\text{def}}{=} \frac{w(T_1, \ldots, T_k)}{\sum_{|t_1|+\cdots+|t_k|=n} w(t_1, \ldots, t_k)}$$

Family	$\Phi(u)$	$W(z)$
unary trees	$1+u$	e^z
binary trees ($\mathcal{B}$)	$1+u^2$	$\frac{1}{1-z}$
m-ary trees, $m>1$ ($\mathcal{T}_m$)	$1+u^m$	$(1-(m-1)z)^{-1/(m-1)}$
general trees	$\frac{1}{1-u}$	$1-\sqrt{1-2z}$

Table 1: Weight Characteristic Series and Generating Functions.

Given a family $\mathcal{F}$ let its *weight characteristic series* be

$$\Phi(u) = \sum_{s\in S} p(s)\cdot u^{\nu(s)} = \sum_{n\in \mathrm{Im}\nu} u^n$$

where $\mathrm{Im}\nu$ denotes the set of arities for which there are symbols in S having them. Let $W(z) = \sum_{T\in\mathcal{F}} w(T)\, z^{|T|}$. Then the n^{th} coefficient of $W(z)$ is the normalizing constant needed to get the probability distribution for trees of size n (see Definition 2.2). We will call $W(z)$ the ***weight*** *generating function* of the family $\mathcal{F}$.

Now, we have the following functional relations for $W(z)$, depending on the definition of the size of a leaf. If we impose $|s|=0$ for all 0-ary symbols s, we get,

$$\frac{dW}{dz} = \Phi(W) - 1, \qquad W(0) = 0$$

whereas if $|s| = 1$ for $s\in\nu^{-1}(0)$, then $W(z)$ satisfies,

$$\frac{dW}{dz} = \Phi(W), \qquad W(0) = 1$$

Note that the weight characteristic series $\Phi(u)$ does not depend on the symbols of S. Moreover, it does not depend on the number of symbols of a given arity, but on the existence or not of symbols of the given arity, and consequently, $W(z)$ has the same properties.

Some interesting characteristic series and weight generating functions are given in Table 1, where in the three first ones we have assumed that the size of a tree is the number of internal nodes (leaves have null size) and in the last family each node, of whatever arity, contributes to the total size.

In the particular case of binary trees,where only symbols of arity 0 and 2 exist, say $\square$ and $\circ$, it is obvious that $p(T) = w(T)$, since $[z^n]W(z) = 1$ for all $n\geq 0$ ($[z^n]W(z)$ stands for the n^{th} coefficient of $W(z)$). Therefore, $p(T)$ adopts a nice simple form :

$$p(T) = \begin{cases} 1 & \text{if } T=\square \\ \frac{p(T_1)\cdot p(T_2)}{|T|} & \text{if } T = \circ(T_1,T_2) = {}_{T_1}\!\wedge_{T_2} \end{cases} \tag{2.1}$$

For this case, the balanced model coincides with that of binary search trees built from random equiprobable permutations (see [BCDM89, Fla88b, Knu73]). This probability distribution also applies to heap-ordered trees [Fla88b] and k-d-trees [Ben75, FP86]. Nevertheless, our balanced model for m-ary trees is not that of m-ary search trees. In fact, these families are not isomorphic, except for $m=2$.

A useful way to view this random tree model for binary trees is the underlying splitting process [Fla88b] : Suppose we have defined the balanced distribution for binary trees of sizes 0 to $n-1$. In order to construct a binary tree T with n internal nodes, select randomly the size of its left subtree, say i, from $0,\ldots,n-1$. Then pick a tree T_1 of size i with probability $p(T_1)$, pick another tree T_2 of size $n-1-i$ with probability $p(T_2)$ and set T_1 and T_2 as the left and right subtrees respectively of the tree T. This protocol defines recursively the balanced distribution as can be found in Definition 2.2, for binary trees, and reformulates the classical characterization of random binary search trees, which states that the size of the left subtree (or of the right subtree) of a tree of size n is a random discrete variable X taking value in the range $[0,\ldots,n-1]$ such that $\Pr(X=i)=1/n, \quad i=0,\ldots,n-1$.

Notice also that for pairs of binary trees, Definition 2.3 implies

$$p(T_1,T_2)=\frac{p(T_1)\cdot p(T_2)}{|T_1|+|T_2|+1}=p({}_{T_1}\!\bigwedge\!{}_{T_2}) \tag{2.2}$$

3 Average Occupancy of m-ary Trees

By occupancy of a tree we mean the sum of the ratios between the number of internal nodes and the maximum number of nodes at a given level. This characteristic can be defined for any simple family of trees with bounded arity (i.e. there exists some K such that $\nu^{-1}(k)=\emptyset$ for all $k>K$) and gives indication of the degree of balancing of the tree. As far as we know, this measure has not appeared in previous literature. Without loss of generality we will examine the behavior of occupancy for m-ary trees. From now on this family will be denoted $\mathcal{T}_m$ and we shall assume that it is generated by $S=\{\square,\circ\}$, where $\nu(\square)=0$ and $\nu(\circ)=m$. Moreover, we assume that leaves do not contribute to the size of an m-ary tree.

The occupancy for an m-ary tree T is,

$$OCC(T)=\sum_{k\geq 0}\frac{N_k(T)}{m^k}$$

where $N_k(T)$ is the *width* of T at level k, i.e. the number of internal nodes (not leaves) of T at level k, and is recursively defined as

$$N_k(T)=\begin{cases}0 & \text{if } T=\square\\ N_{k-1}(T_1)+\cdots+N_{k-1}(T_m) & \text{if } T=\circ(T_1,\ldots,T_m)\end{cases}$$

and

$$N_0(T)=\begin{cases}0 & \text{if } T=\square\\ 1 & \text{otherwise}\end{cases}$$

Let $OCC_{MIN}(n)$ and $OCC_{MAX}(n)$ denote the minimum occupancy and maximum occupancy achieved by m-ary trees of size n. No much effort is needed to get,

$$\begin{aligned}OCC_{MIN}(n) &= \frac{m}{m-1}-\frac{m^{-n}}{m-1}\approx\frac{m}{m-1}\\ OCC_{MAX}(n) &= d+\frac{\ell}{m^d}\leq\lceil\log_m(n+1)\rceil\end{aligned}$$

where $n=m^d-1+\ell$, $0<\ell<m^d$.

3.1 Average Occupancy for the Uniform Probability Model

We start computing the average occupancy $\overline{OCC(n)}$, for m-ary trees of size n supposing that al them are equiprobable. Let

$$OCC(z) = \sum_{T \in \mathcal{T}_m} OCC(T)\, z^{|T|}$$

The average occupancy is therefore,

$$\overline{OCC(n)} = \frac{[z^n]OCC(z)}{t_{n,m}}$$

where $t_{n,m}$ stands for the number of m-ary trees of n nodes.

From the definition of $OCC(T)$ is not difficult to verify that,

$$OCC(z) = \frac{zF^m(z)}{1 - zF^{m-1}(z)} = F^2(z) - F(z)$$

where $F(z) = \sum_{n \geq 0} t_{n,m}\, z^n = 1 + zF^m(z)$.

Applying Lagrange-Bürmann inversion formula [Hen77], we have

$$\begin{aligned} OCC(z) &= \sum_{n \geq 1} \frac{z^n}{n!} \left[\frac{d^{n-1}}{dy^{n-1}} \left((2y-1)y^{mn}\right)_{y=1} \right] \\ [z^n]OCC(z) &= \binom{mn}{n} \cdot \left(2\frac{mn+1}{(mn-n+2)(mn-n+1)} - \frac{1}{mn-n+1} \right) \end{aligned}$$

And using the same inversion formula, it turns out that,

$$[z^n]F(z) = t_{n,m} = \binom{mn}{n} \cdot \frac{1}{mn-n+1}$$

Hence,

Theorem 3.1 *The average occupancy of m-ary trees of size n, if the uniform probability mod is assumed is*

$$\overline{OCC(n)} = 2\frac{mn+1}{(m-1)n+2} - 1 \approx \frac{m+1}{m-1} + O(\frac{1}{n})$$

From the above theorem is easy to see that

$$\overline{OCC(n)} \approx 2 \cdot OCC_{MIN}(n)$$

3.2 Average Occupancy for the Balanced Probability Model

If, on the other hand, we use the balanced probability model defined in Section 2 for m-ary tree the average occupancy is now given by

$$\overline{OCC(n)} = \sum_{|T|=n} OCC(T)\, p(T) = \frac{[z^n] \sum_{T \in \mathcal{T}_m} OCC(T)\, w(T)\, z^{|T|}}{[z^n]W(z)}$$

Denoting by $OCC(z)$ the generating function in the numerator, the relationships for the occupanc translate into a first-order ordinary differential equation,

$$\frac{d}{dz}OCC - \frac{1}{1-(m-1)z}OCC = \frac{1}{(1-(m-1)z)^{m/(m-1)}}, \quad OCC(0) = 0$$

whose solution is,

$$OCC(z) = \frac{1}{m-1}W(z)\ln\left(\frac{1}{1-(m-1)z}\right)$$

Therefore, using the appropriate transfer lemmas [Fla88a], $[z^n]OCC(z)$ is asymptotically,

$$[z^n]OCC(z) \approx \frac{1}{(m-1)^{n+1}} \cdot n^{-\frac{m-2}{m-1}} \cdot \ln n \cdot \frac{1}{\Gamma(1/(m-1))}\left(1 + O\left(\frac{1}{n}\right)\right)$$

Theorem 3.2 *The average occupancy for m-ary trees, under the balanced probability model is*

$$\overline{OCC(n)} \approx \frac{1}{m-1}\ln n$$

The relationship of the average occupancy to the extremal values is now

$$\overline{OCC(n)} \approx \frac{1}{m-1}\ln m \cdot OCC_{MAX}(n)$$

4 Average Size of the Intersection of Two m-ary Trees

In this section we examine a clear example, the intersection of a pair of trees, where the balanced model marks an important difference with respect to the uniform model.

Intersection of trees appears in a natural way in a certain number of algorithms, such as tree matching [SF83] or unification [CDS89]. It turns out that, if the uniform model is assumed, the average size of the intersection is $O(1)$, independently of the size of the trees and its arity. The average size of intersection of binary trees using a balanced model has been covered in [BCDM89].

Now, given trees $T', T'' \in \mathcal{T}_m$ we wish to compute the average size of the intersection of the two trees, where the intersection of T' and T'' is a m-ary tree given by :

$$T' \cap T'' = \begin{cases} \square & \text{if } T' = \square \text{ or } T'' = \square \\ \circ((T_1' \cap T_1''), \ldots, (T_m' \cap T_m'')) & \text{if } T' = \circ(T_1', \ldots, T_m') \\ & \text{and } T'' = \circ(T_1'', \ldots, T_m'') \end{cases}$$

The size of the intersection of two trees T' and T'', is thus null if any of the trees is a leaf, and the sum of the intersections of the subtrees, otherwise, and it will denoted by $s(T', T'')$.

Let $\overline{S(n)}$ be the average value of $s(T', T'')$ over all the pairs $(T', T'') \in \mathcal{T}_m^2$ with $|T'| + |T''| = n$. We have, by application of standard generating function techniques[GJ83, VF90],

$$\begin{aligned} \overline{S(n)} &= \sum_{|T'|+|T''|=n} s(T', T'') \cdot p(T', T'') = \\ &= \frac{\sum_{|T'|+|T''|=n} s(T', T'') \cdot w(T', T'')}{\sum_{\substack{T_1, T_2 \in \mathcal{T}_m \\ |T_1|+|T_2|=n}} w(T_1) \cdot w(T_2)} = \frac{[z^n]S(z,z)}{[z^n]W^2(z)} \end{aligned} \tag{4.1}$$

where

$$S(x,y) = \sum_{(T',T'') \in \mathcal{T}_m^2} s(T', T'') w(T') w(T'') x^{|T'|} y^{|T''|} \tag{4.2}$$

and

$$W(z) = \sum_{T \in \mathcal{T}_m} w(T) z^{|T|} = (1-(m-1)z)^{-1/(m-1)} \tag{4.3}$$

The asymptotic behavior of the n^{th} coefficient ($n \to \infty$) of $W^2(z)$ can be easily derived by means of Darboux's theorem [Hen77] which yields :

$$[z^n]W^2(z) \approx (m-1)^n \cdot n^{-\frac{m-3}{m-1}} \cdot \frac{1}{\Gamma(2/(m-1))} \cdot (1 + O(1/n)) \tag{4.4}$$

where $\Gamma(\cdot)$ denotes Euler's gamma function.

Differentiating Eq. (4.2) and using the previous definitions of $w(T)$ and $s(T', T'')$ together with Eq. (4.3) we get the following hyperbolic partial differential equation

$$\frac{\partial^2 S(x,y)}{\partial x \partial y} = [W(x) \cdot W(y)]^m + [W(x) \cdot W(y)]^{m-1} \cdot m \cdot S(x,y) \tag{4.5}$$

subject to the boundary conditions: for all x and y, $S(x,0) = 0$ and $S(0,y) = 0$.

A particular solution to Eq. (4.5) is $-\frac{1}{m-1} W(x)W(y)$. Therefore, it can be rewritten as

$$S(x,y) = \Psi(x,y) - \frac{1}{m-1} W(x)W(y) \tag{4.6}$$

where $\Psi(x,y)$ satisfies the homogeneous equation

$$\frac{\partial^2 \Psi(x,y)}{\partial x \partial y} = \frac{m\Psi(x,y)}{(1-(m-1)x)(1-(m-1)y)}$$

with boundary conditions $\Psi(x,0) = \frac{1}{m-1} W(x)$ and $\Psi(0,y) = \frac{1}{m-1} W(y)$. This last differential equation can be transformed into a simpler form, making the change of variables

$$\begin{cases} X = -\frac{\sqrt{m}}{m-1} \ln(1-(m-1)x) \\ Y = -\frac{\sqrt{m}}{m-1} \ln(1-(m-1)y) \end{cases}$$

and making $H(X,Y)$ be $\Psi\left(\frac{1}{m-1}(1-e^{-X\frac{m-1}{\sqrt{m}}}), \frac{1}{m-1}(1-e^{-Y\frac{m-1}{\sqrt{m}}}\right)$, we finally obtain the hyperbolic differential equation

$$\frac{\partial^2 H}{\partial X \partial Y} = H$$

subject to boundary conditions $H(X,0) = \frac{1}{m-1} e^{X/\sqrt{m}}$, $H(0,Y) = \frac{1}{m-1} e^{Y/\sqrt{m}}$.

This system can be solved by the method of Riemann [Cop75] thus yielding

$$\begin{aligned} H(X,Y) &= \frac{1}{(m-1)\sqrt{m}} \int_0^X e^{t/\sqrt{m}} J_0\left(2i\sqrt{(X-t)Y}\right) dt + \\ &\quad \frac{1}{(m-1)\sqrt{m}} \int_0^Y e^{t/\sqrt{m}} J_0\left(2i\sqrt{(Y-t)X}\right) dt + \\ &\quad \frac{1}{m-1} J_0(2i\sqrt{XY}) \end{aligned}$$

where J_0 denotes the Bessel function of the first kind and order 0.

Since we are interested in obtaining asymptotics for $\overline{S(n)}$, we shall deduce an asymptotic valu for the n^{th} coefficient in the Taylor series expansion of $\Psi(z,z)$:

$$\overline{S(n)} = \frac{[z^n]\Psi(z,z)}{[z^n]W^2(z)} - \frac{1}{m-1} \tag{4.7}$$

If we denote $H(Z,Z) = A(Z) + \frac{1}{m-1}J_0(2iZ)$, from the series expansion of the Bessel function we get

$$A(Z) = \frac{2}{m-1}\sum_{i\geq 0}\frac{(Z\sqrt{m})^i}{i!}\left(\sum_{j>i}\frac{(Z/\sqrt{m})^j}{j!}\right)$$

And straightforward manipulation of the coefficients of $[Z^n]A(Z)$ yields,

$$H(Z,Z) \approx \frac{m+1}{(m-1)^2}J_0(2iZ) + \frac{\sqrt{m}}{(m-1)^2}\frac{d}{dZ}J_0(2iZ)$$

so we can conclude

Lemma 4.1

$$[z^n]\Psi(z,z) \approx c_1 \cdot [z^n]J_0\left(-2\frac{\sqrt{m}}{m-1}\cdot i\cdot \ln(1-(m-1)z)\right)$$

where $\approx$ stands for asymptotical equivalence, and $c_1 = \frac{m+2\sqrt{m}+1}{(m-1)^2}$

On the other hand, using singularity analysis (when $z \to 1$, the modulus of the argument of J_0 becomes infinite) and the appropriate transfer lemmas [AS64, Fla88a], we derive the asymptotics for the n^{th} coefficient in the expansion of the Bessel function,

Lemma 4.2

$$[z^n]J_0(-2\frac{\sqrt{m}}{m-1}\cdot i\cdot \ln(1-(m-1)z)) \approx c_2\cdot(m-1)^n\frac{n^{\frac{2\sqrt{m}}{m-1}-1}}{\sqrt{\ln n}}\cdot\left(1+O(\frac{1}{\ln n})\right)$$

where the value of constant c_2 is given by

$$c_2 = \frac{\sqrt{m-1}}{2m^{1/4}\sqrt{\pi}\cdot\Gamma(\frac{2\sqrt{m}}{m-1})}$$

Lemma 4.1 together with Lemma 4.2 give the following result,

$$[z^n]\Psi(z,z) \approx c_1\cdot c_2\cdot(m-1)^n\cdot\frac{n^{2\frac{\sqrt{m}}{m-1}-1}}{\sqrt{\ln n}}$$

In the light of Eq. (4.7), and using Eq. (4.4) we can conclude

Theorem 4.1 *Under the distribution defined in Section 2, the average size of the intersection of two m-ary trees behaves asymptotically as*

$$\overline{S(n)} \approx c\cdot\frac{n^{\frac{2\sqrt{m}-2}{m-1}}}{\sqrt{\ln n}}\cdot\left(1+O(\frac{1}{\ln n})\right)$$

where the constant c has value

$$c = \frac{m+2\sqrt{m}+1}{2\sqrt{\pi}m^{1/4}(m-1)^{3/2}}\cdot\frac{\Gamma(\frac{2}{m-1})}{\Gamma(\frac{2\sqrt{m}}{m-1})}$$

5 Conclusions. Further Research.

The cases of study can be extended to other families of trees, by using the adequate weight generating function, as presented in Section 2.

The present analysis suggests that the apparition of hyperbolic partial differential equations in analysis involving pairs of trees, is rather independent of the nature of the problem and that relies on the underlying probability distribution, in the same way as ordinary differential equations arise everywhere when dealing with an algorithm which has a unique tree as input.

Further continuation of the present work will explore the uses of the balanced probability model to analyze other well known algorithms on trees, and the effects of the probability models with respect to average measures.

We are also interested in the development of tools which simplify the analysis of the problems to be faced.

References

[AS64] M. Abramowitz and I.A. Stegun, editors. *Handbook of Mathematical Functions.* Dover Publications, Inc., New York, 1964.

[BCDM89] R. Baeza-Yates, R. Casas, J. Díaz, and C. Martínez. On the average size of the intersection of binary trees. Technical Report 23, LSI-UPC, 1989.

[Ben75] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Comm. of the ACM*, 18(9):509–517, Sept 1975.

[CDS89] R. Casas, J. Díaz, and J. M. Steyaert. Average-case analysis of Robinson's unification algorithm with two different variables. *Information Processing Letters*, 31:227–232, June 1989.

[Cop75] E.T. Copson. *Partial Differential Equations.* Cambridge University Press, Cambridge, 1975.

[Fla88a] P. Flajolet. Mathematical methods in the analysis of algorithms and data structures. In Egon Börger, editor, *Trends in Theoretical Computer Science*, chapter 6, pages 225–304. Computer Science Press, 1988.

[Fla88b] P. Flajolet. Random tree models in the analysis of algorithms. In *Performance'87*, pages 171–187, 1988.

[FP86] P. Flajolet and C. Puech. Partial match retrieval of multidimensional data. *J. ACM*, 33(2):371–407, Apr 1986.

[GJ83] I. Goulden and D. Jackson. *Combinatorial Enumerations.* Wiley, New York, 1983.

[Hen77] P. Henrici. *Applied and Computational Complex Analysis (volume 2).* Wiley Interscience, New York, 1977.

[Kar86] R. Karp. Combinatorics, complexity and randomness. *Comm. of the ACM*, 29(2):98–119, Feb 1986.

[Knu73] D.E. Knuth. *The Art of Computer Programming: Sorting and Searching*, volume 3. Addison-Wesley, Reading, Mass., 1973.

[MM78] A. Meir and J.W. Moon. On the altitude of nodes in random trees. *Canad J Math*, 30(5):997–1015, 1978.

[Sed83] R. Sedgewick. Mathematical Analysis of Combinatorial Algorithms. In G. Louchard and G. Latouche, editors, *Probability Theory and Computer Science*, part III, pages 124–205. Academic Press, 1983.

[SF83] J. M. Steyaert and P. Flajolet. Patterns and pattern-matching in trees: an analysis. *Information and Control*, 58(1-3):19–58, 1983.

[Ste84] J.M. Steyaert. *Structure et Complexité des Algorithmes.* PhD thesis, Université Paris 7, April 1984.

[VF90] J.S. Vitter and P. Flajolet. Average-case Analysis of Algorithms and Data Structures. In J. van Leeuwen, editor, *Algorithms and Complexity*, volume A, chapter 9, pages 433–524. Elsevier, 1990.