Araucaria Trees: Construction and Grafting Theorems

Dominique Schmitt
Laboratoire LMIA, Equipe MAGE, Université de Haute Alsace
4, rue des Frères Lumière, 68093 Mulhouse, France
Dominique.Schmitt@uha.fr

Jean-Claude Spehner [C]
Laboratoire LMIA, Equipe MAGE, Université de Haute Alsace
4, rue des Frères Lumière, 68093 Mulhouse, France
Jean-Claude.Spehner@uha.fr

Abstract. Araucarias have been introduced by Schott and Spehner as trees which appear in the minimal automaton of the shuffle of words. We give here a new definition of araucarias which is more constructive and we prove that our definition of araucarias is equivalent to the original one. From the new definition we derive an optimal algorithm for the construction of araucarias and a new method for calculating their size. Moreover we characterize araucarias by properties of their maximal paths, by associating a capacity to every edge. We then show that every araucaria can be obtained by grafting and merging smaller araucarias. We prove also that every directed tree can be embedded in an araucaria. Moreover we define a capacity for every vertex of an araucaria, which leads to different new enumeration formulas for araucarias.

Keywords: Algorithms, trees, symmetric polynomials, combinatorics, shuffle of words

1. Introduction

If u and v are words of a free monoid, the shuffle of u and v (denoted $u \sqcup\!\sqcup v$) is the language whose words are of the form $u_1v_1u_2v_2 \dots u_mv_m$ where $u_1u_2 \dots u_m$ is a factorization of u, $v_1v_2 \dots v_m$ a factorization of v, and the factors u_1 and v_m are possibly empty. More generally, if I and J are two languages of a free monoid, the union of the sets $u \sqcup\!\sqcup v$ for $u \in I$ and $v \in J$ is called the shuffle of the languages I and J (denoted $I \sqcup\!\sqcup J$). The shuffle product $u_1 \sqcup\!\sqcup \dots \sqcup\!\sqcup u_k$ of k words $u_1, \dots, u_k$ can then be defined by induction on k.

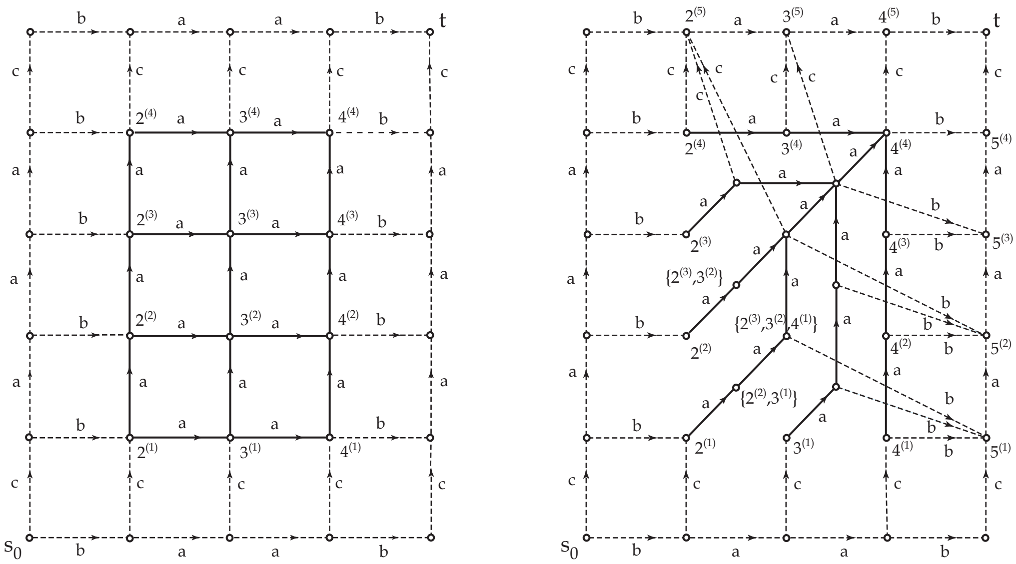
[C]Corresponding author

The shuffle product admits various applications, in particular in parallel processing [3, 8]. Theoretical results are given in [4, 5] and algorithmic results can be found in [12, 6, 1]. However, the construction of the minimal automaton which recognizes the shuffle product of k words is an old open problem.

The set of letters of a word is called its alphabet. If the alphabets of the words $u_1, \ldots, u_k$ are pairwise disjoint, the minimal automaton of the language $u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_k$ is simply its non-deterministic automaton and its size is equal to $\prod_{i=1}^{k}(|u_i| + 1) + 1$ where $|u_i|$ is the length of the word u_i. In the converse case, the minimal automaton contains, for every common letter of the words $u_1, \ldots, u_k$, some directed graphs having special properties, and its size is much bigger.

Recently, Biegler, Daley, and McQuillan [2] have shown that the size of the minimal automaton which recognizes the language $u \sqcup\!\sqcup v$ can be exponential relative to the lengths of the words u and v.

Schott and Spehner [10] (see also [9] and the erratum [11]) have studied the special case of k words of the form $b_1 a^{p_1} c_1, \ldots, b_k a^{p_k} c_k$ (where $p_1, \ldots, p_k$ are positive integers and $a, b_1, \ldots, b_k, c_1, \ldots, c_k$ are pairwise distinct letters up to the equalities $b_1 = c_1, \ldots, b_k = c_k$). They have proved that the minimal automaton of the language $b_1 a^{p_1} c_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup b_k a^{p_k} c_k$ contains a directed tree, called araucaria, which is characterized by the integer k, called its arity, and by the sequence of integers $(p_1, \ldots, p_k)$, called its type (see Fig. 1). The araucaria constitutes the most complicated part of the automaton.



(a) (b)

Figure 1. The grid (a) represents the part of the non-deterministic automaton which recognizes the language $L = caaac \sqcup\!\sqcup baab$ obtained by removing its absorbing state (the empty set); s_0 is the initial state and t the terminal state. (b) represents a part of the minimal automaton of L obtained by removing its absorbing state. If we change every edge (u, v) of the oriented subgraph of (b) with unbroken edges in its opposite edge (v, u), we obtain an araucaria of type $(3, 2)$.

This paper is a continuation of the first part of Schott and Spehner's paper by using a more algorithmic definition of araucarias and other graph-theoretic or combinatorial concepts. The new definition is independent of automata theory, but we also give some hints to interpret our results in the context of

the shuffle of words. The aim is to give more graphical and combinatorial properties of araucarias that should, in particular, help to prove the following conjecture :
"If the alphabets of the words $u_1, \ldots, u_k$ are pairwise disjoint up to a common letter a, then all subgraphs of the minimal automaton of $u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup u_k$ which are defined by the transitions relative to the common letter a are araucarias or homomorphic images of araucarias. The maximum size of the minimal automaton is a polynomial function of $|u_1|, \ldots, |u_k|$ whose coefficients are exponential in k".

The new definition of araucarias is given in Section 2. In the original definition, an araucaria is recursively defined as a union of non disjoint subaraucarias. This definition does not enable us to give an efficient algorithm to construct araucarias. We introduce here the notion of semi-araucaria and show that an araucaria is a union of semi-araucarias having only their roots in common.

This new definition leads, in Section 3, to an optimal algorithm to construct araucarias.

In Section 4, we characterize araucarias through their maximal paths (paths starting at the root and ending at a leaf). To this aim, we associate to every edge of an araucaria of type $(p_1, \ldots, p_k)$ a subset of $\{1, \ldots, k\}$, called the capacity of the edge. Then, a capacity chain corresponds to every maximal path of the araucaria. We show that these chains characterize the araucaria among all other directed trees.

In Section 5, we define a grafting operation from a directed tree C onto a directed tree B which consists in sticking at every vertex s of B a copy of C (with root s). We first prove that, if B and C are araucarias, then the grafting leads to a subtree of another araucaria. We then show that, for every positive integer $m < k$, any araucaria of arity k can be obtained by first grafting araucarias of arity m onto araucarias of arity $k - m$ and by then merging the resulting trees.

More generally, we show in Section 6 that every araucaria of arity k can be obtained by grafting and merging araucarias whose sum of arities is k. In the particular case where the grafted araucarias are reduced to paths, the grafting leads to a new family of directed trees having the property that every araucaria can be obtained by merging such trees.

In Section 7, we prove that every directed tree is isomorphic to some subtree of an araucaria.

In [10], Schott and Spehner proved that the size of an araucaria of type $(p_1, \ldots, p_k)$ is equal to the remarkable polynomial $\Upsilon_k(p_1, \ldots, p_k)$, where Υ_k is defined in the following way:

$$\Upsilon_k(X_1, \ldots, X_k) = \sum_{m=0}^{k} m!\Psi_m(X_1, \ldots, X_k)$$

where $\Psi_m(X_1, \ldots, X_k)$ is the elementary symmetric polynomial of degree m on variables $X_1, \ldots, X_k$ and $\Psi_0(X_1, \ldots, X_k) = 1$.

In the last section of this paper, we give some other enumeration formulas for araucarias. To this aim we first associate to every vertex of an araucaria of arity k a capacity, which is a subset I of $K = \{1, \ldots, k\}$. We then prove that the number of vertices of capacity I of an araucaria of type $(p_1, \ldots, p_k)$ is equal to $(k - |I|)! \prod_{i \in K \setminus I} p_i$. We also show that the size of a semi-araucaria of type $(p_1, \ldots, p_k)$ is equal to $k! \prod_{i \in K} p_i + 1$. This leads to two new and simple proofs that the size of an araucaria of type $(p_1, \ldots, p_k)$ is equal to $\Upsilon_k(p_1, \ldots, p_k)$.

An extended abstract of Sections 2, 3, and 4 has been presented in [7].

2. Semi-araucarias and araucarias

After basic definitions, we introduce a new kind of directed trees called semi-araucarias. Araucarias are then defined by using these semi-araucarias and we prove that this new definition of araucarias is equivalent to the original one given by Schott and Spehner [11]. We do not recall the original definition here because it needs several notions which are not introduced in this paper (see Remark 4.16). Moreover, the notations used in the original definition are incompatible with ours.

Definition 2.1. (i) Every pair $G = (V, E)$ where $E \subset V \times V$ is called a directed graph. Every $v \in V$ is called a vertex of G and every $(s, t) \in E$ is called an (oriented) edge of G. For every edge (s, t) of E, t is called a successor of s and s is called a predecessor of t.

(ii) Every sequence $\sigma = (s_0, \ldots, s_f)$ of vertices such that $(s_0, s_1), \ldots, (s_{f-1}, s_f)$ are edges of E is called a path of G from s_0 to s_f. The integer f is called the length of σ and is denoted by $|\sigma|$. If $\sigma = (s_0, \ldots, s_f)$ and $\tau = (s_f, \ldots, s_g)$ are paths of G, the path $\lambda = (s_0, \ldots, s_{f-1}, s_f, s_{f+1}, \ldots, s_g)$ is called the product of σ and τ and is denoted by $\lambda = \sigma\tau$.

(iii) A directed graph $G = (V, E)$ is called a directed tree if there exists one vertex $r \in V$ without predecessor and such that, for every vertex $s \in V \setminus \{r\}$, there exists a unique path from r to s. The vertex r is called the root of G. The length of the path from the root r to s is called the height of s. Every vertex s of G without successor is called a leaf of G. Each path starting at the root r and whose last vertex is a leaf of G is said to be maximal.

Definition 2.2. Throughout this paper, k will be a positive integer, K the set $\{1, \ldots, k\}$, and $(p_1, \ldots, p_k)$ a sequence of k positive integers.

A semi-araucaria $H(p_1, \ldots, p_k)$ of type $(p_1, \ldots, p_k)$ is a directed tree recursively defined in the following way (see Fig. 2(a) and Example 2.4):
If $k = 1$, $H(p_1)$ is a path of length p_1.
If $k > 1$, $H(p_1, \ldots, p_k)$ is the union of a path $\tau = (s_0, \ldots, s_p)$ of length $p = \sum_{i \in K} p_i$ and, for each non-empty proper subset $I = \{i_1, \ldots, i_m\}$ of K and each h in $\{1, \ldots, \sum_{j \in K \setminus I} p_j - 1\}$, of a semi-araucaria $H_{I,h}$ of type $(p_{i_1}, \ldots, p_{i_m})$ such that:
- the root of $H_{I,h}$ is the vertex s_h of τ and $H_{I,h} \cap \tau = \{s_h\}$,
- for each non-empty proper subset J of K ($J \neq I$) such that $h < \sum_{j \in K \setminus J} p_j$, $H_{I,h} \cap H_{J,h} = \{s_h\}$.

The integer k is called the arity of $H(p_1, \ldots, p_k)$ and the path τ is called its trunk.

Definition 2.3. An araucaria $A(p_1, \ldots, p_k)$ of type $(p_1, \ldots, p_k)$ is a directed tree which is the union of the semi-araucarias $H_{I,0}$ of root s_0 and of type $(p_{i_1}, \ldots, p_{i_m})$, where s_0 is a vertex, $I = \{i_1, \ldots, i_m\}$ is a non-empty subset of K and, for every non-empty subset J of K distinct from I, $H_{I,0} \cap H_{J,0} = \{s_0\}$ (see Fig. 2(b) and Example 2.4).

The integer k is called the arity of the araucaria and the trunk of $H_{K,0}$ is called its trunk.

Example 2.4. (i) The semi-araucaria H of type $(p_1, p_2, p_3) = (3, 2, 1)$ (Fig. 2(a)) is composed of a trunk $(s_0, \ldots, s_6)$ on which the twelve following subsemi-araucarias (i.e., subtrees which are semi-araucarias) are sticked:
- one subsemi-araucaria $H_{\{1,3\},1}$ of type $(p_1, p_3) = (3, 1)$ (Fig. 2(d)) on the vertex s_1,
- two subsemi-araucarias $H_{\{2,3\},1}$ and $H_{\{2,3\},2}$ of type $(p_2, p_3) = (2, 1)$ (Fig. 2(e)) on the vertices s_1

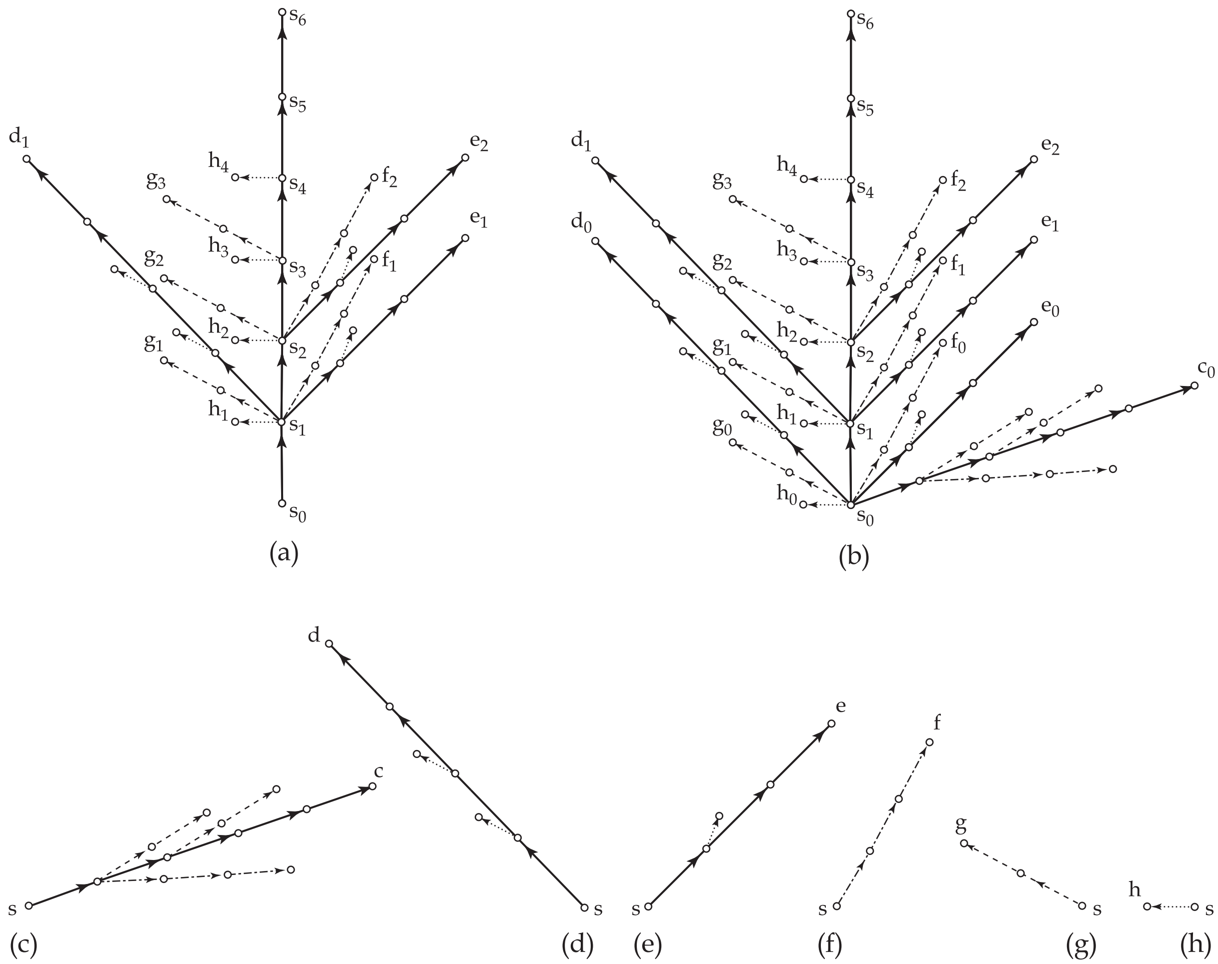


Figure 2. A semi-araucaria (a) and an araucaria (b) of type $(3, 2, 1)$. (c), (d), (e), (f), (g) and (h) are semi-araucarias of respective types $(3, 2)$, $(3, 1)$, $(2, 1)$, (3), (2), and (1).

and s_2,
- two subsemi-araucarias $H_{\{1\},1}$ and $H_{\{1\},2}$ of type $(p_1) = (3)$ (Fig. 2(f)) on the vertices s_1 and s_2,
- three subsemi-araucarias $H_{\{2\},1}$, $H_{\{2\},2}$, and $H_{\{2\},3}$ of type $(p_2) = (2)$ (Fig. 2(g)) on the vertices s_1, s_2, and s_3,
- four subsemi-araucarias $H_{\{3\},1}$, $H_{\{3\},2}$, $H_{\{3\},3}$, and $H_{\{3\},4}$ of type $(p_3) = (1)$ (Fig. 2(h)) on the vertices s_1, s_2, s_3, and s_4.

(ii) The araucaria A of type $(p_1, p_2, p_3) = (3, 2, 1)$ (Fig. 2(b)) is composed of a root s_0 on which are sticked the subsemi-araucarias $H_{\{1,2,3\},0}$ of type $(3, 2, 1)$, $H_{\{1,2\},0}$ of type $(3, 2)$ (Fig. 2(c)), $H_{\{1,3\},0}$ of type $(3, 1)$, $H_{\{2,3\},0}$ of type $(2, 1)$, $H_{\{1\},0}$ of type (3), $H_{\{2\},0}$ of type (2), and $H_{\{3\},0}$ of type (1).

We give now a characterization of araucarias similar to that in [11] which implies that our definition of araucarias is equivalent to the original one.

Theorem 2.5. A directed tree B is an araucaria of type $(p_1, \ldots, p_k)$ if and only if B has the following recursive properties:
If $k = 1$, B is a path of length p_1.
If $k > 1$, B is the union of a path τ_B of length $p_1 + \ldots + p_k$ and, for each non-empty proper subset $I = \{i_1, \ldots, i_m\}$ of K and each $h \in \{0, \ldots, \sum_{j\in K\setminus I} p_j - 1\}$, of a subtree $B_{I,h}$ such that:
- $B_{I,h}$ is an araucaria of type $(p_{i_1}, \ldots, p_{i_m})$,
- the root of $B_{I,h}$ is the vertex s_h of τ_B of height h and $B_{I,h} \cap \tau_B = \{s_h\}$,
- for every non-empty proper subset J of K such that $h < \sum_{i\in K\setminus J} p_i$, the common subtree of $B_{I,h}$ and $B_{J,h}$ is an araucaria of root s_h and type $(p_{j_1}, \ldots, p_{j_r})$ where $\{j_1, \ldots, j_r\} = I \cap J$.

Proof:
This theorem is trivial for $k = 1$ since an araucaria of type p_1 and a semi-araucaria of type p_1 are reduced to a single path of length p_1.

Assume the result for every araucaria of arity smaller than k ($k > 1$) and let A be an araucaria of type $(p_1, \ldots, p_k)$. Let B be a directed tree as given in the statement of the theorem.

Since the length of the trunk τ_A of $H_{K,0}$ is equal to the length of the path τ_B of B, there exists an isomorphism θ_0 from τ_A onto τ_B.

By Definitions 2.2 and 2.3, for each non-empty proper subset $I = \{i_1, \ldots, i_m\}$ of K and each h of $\{0, \ldots, \sum_{j\in K\setminus I} p_j - 1\}$, the vertex s_h of τ_A of height h is the root of the semi-araucaria $H_{I,h}$ of type $(p_{i_1}, \ldots, p_{i_m})$ and $H_{I,h} \cap \tau_A = \{s_h\}$. Since, for every non-empty proper subset $J = \{j_1, \ldots, j_l\}$ of I, $\sum_{j\in K\setminus J} p_j > \sum_{j\in K\setminus I} p_j$, s_h is also the root of the semi-araucaria $H_{J,h}$ of type $(p_{j_1}, \ldots, p_{j_l})$, and $H_{I,h} \cap H_{J,h} = \{s_h\}$. By induction, the union $A_{I,h}$ of $\{H_{J,h}; \emptyset \subset J \subseteq I\}$ is an araucaria of type $(p_{i_1}, \ldots, p_{i_m})$. Since, the vertex $\theta_0(s_h)$ is the root of $B_{I,h}$, θ_0 can be extended to an isomorphism θ such that $\theta(A_{I,h}) = B_{I,h}$.

Moreover, for every non-empty proper subset J of K such that $h < \sum_{i\in K\setminus J} p_i$, the common subtree of $A_{I,h}$ and $A_{J,h}$ is, by construction, the union of all $H_{L,h}$ such that $\emptyset \subset L \subseteq I \cap J$ which is, by induction, a subaraucaria of root s_h and type $(p_{j_1}, \ldots, p_{j_r})$, where $\{j_1, \ldots, j_r\} = I \cap J$. Hence, $\theta(A_{I,h} \cap A_{J,h}) = B_{I,h} \cap B_{J,h}$.

It follows that θ_0 can be extended to an isomorphism from A onto B. □

Corollary 2.6. Our definition of araucarias is equivalent to the original one given by Schott and Spehner.

Proof:
It follows from Theorem 2.5, by induction on the arity, that an araucaria of type $(p_1, \ldots, p_k)$ is unique up to an isomorphism.

In Theorem 1.1 of [11], Schott and Spehner characterize their araucarias by the same properties as those of Theorem 2.5, but with more restrictive conditions on the integer h. They also prove the unicity of their araucarias up to an isomorphism. Hence both definitions are equivalent. □

Remark 2.7. It is not difficult to prove by induction on the arity that, for each permutation φ of K, the araucaria $A(p_{\varphi(1)}, \ldots, p_{\varphi(k)})$ is isomorphic to $A(p_1, \ldots, p_k)$ [11]. Hence the type $(p_1, \ldots, p_k)$ of an araucaria can be given in a standard form such that $p_1 \geq p_2 \geq \ldots \geq p_k$.

Remark 2.8. The subaraucarias and the subsemi-araucarias appear in a natural way in the context of shuffle of words. The classical representation of the non-deterministic automaton that recognizes the language $b_1 a^{p_1} c_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup\, b_k a^{p_k} c_k$ contains a set of a-transitions that forms a k-dimensional hyper-parallelepiped of size $\prod_{i=1}^{k}(p_i + 1)$ (see Fig. 1(a)). The vertex r of this hyper-parallelepiped that has no outgoing a-transition corresponds to the root of the araucaria of type $(p_1, \ldots, p_k)$ in the minimal automaton (see Fig. 1(b)). For every integer $m \in \{1, \ldots, k\}$ and for every subset $\{i_1, \ldots, i_m\}$ of K, r is a vertex of an m-dimensional face of size $\prod_{j=1}^{m}(p_{i_j} + 1)$ of the hyper-parallelepiped and is the root of both a subaraucaria and a subsemi-araucaria of arity m and type $(p_{i_1}, \ldots, p_{i_m})$ in the minimal automaton. If the faces are considered as open (i.e. without their boundaries), they are pairwise disjoint like their corresponding subsemi-araucarias (up to the root r). Otherwise, two closed faces intersect in a subface if and only if their corresponding subaraucarias intersect in a subaraucaria.

3. Optimal implementation of araucarias

The aim of this section is to show that, with a simple data structure, Definitions 2.2 and 2.3 lead immediately to algorithms that construct araucarias in linear time (linear in the size of the constructed trees). In Section 4, it will be shown that the way araucarias are implemented with these algorithms also helps to easily report some characteristic values of araucarias (Remark 4.15). In Section 8, the complexity of the algorithms will be compared to that of other construction methods (Remark 8.11).

Araucarias and semi-araucarias can be stored in any data structure used for directed trees, such as first-child/next-sibling representation. Every vertex s in such a representation contains a link to its oldest child and another one to the immediately younger sibling of s. The data structure also contains direct access to the root of the tree. We suppose that we are also given a function $Copy(T)$, which returns a copy of the tree T, and a function $Path(n)$, which returns a path of length n in the form of a first-child/next-sibling representation (in this case no vertex has a sibling).

The algorithm to construct semi-araucarias is then a straightforward application of Definition 2.2. To this aim, we first write a recursive function $SemiAraucariaWithoutItsRoot(P, I)$ which, given a sequence $P = (p_1, \ldots, p_k)$ of positive integers and a non-empty subset $I = \{i_1, \ldots, i_m\}$ of $\{1, \ldots, k\}$ returns a semi-araucaria of type $(p_{i_1}, \ldots, p_{i_m})$ without its root (this is to avoid the creation of duplicate vertices that would have to be merged later).

$SemiAraucariaWithoutItsRoot(P, I)$
  $\tau = Path(\sum_{i \in I} p_i - 1)$ // the trunk of the semi-araucaria without its first vertex
  for every non-empty proper subset J of I
    $SA = SemiAraucariaWithoutItsRoot(P, J)$
    for every vertex s of τ of height strictly less than $\sum_{j \in I \setminus J} p_j - 1$
      $SA' = Copy(SA)$
      $next\_sibling(root(SA')) = next\_sibling(first\_child(s))$
      $next\_sibling(first\_child(s)) = root(SA')$
  return the constructed tree with root $root(\tau)$

And thus the function to construct a semi-araucaria of type P and arity k:

$SemiAraucaria(P, k)$
    create a vertex r // the root of the semi-araucaria
    $K = \{1, \ldots, k\}$
    $SA = SemiAraucariaWithoutItsRoot(P, K)$
    $first\_child(r) = root(SA)$
    return the constructed tree with root r

The function to construct an araucaria of type P and arity k is then an immediate consequence of Definition 2.3:

$Araucaria(P, k)$
    $SA = SemiAraucaria(P, k)$
    for every non-empty proper subset I of $\{1, \ldots, k\}$
        $SA' = SemiAraucariaWithoutItsRoot(P, I)$
        $next\_sibling(root(SA')) = next\_sibling(first\_child(root(SA)))$
        $next\_sibling(first\_child(root(SA))) = root(SA')$
    return the constructed tree with root $root(SA)$

Obviously, the given algorithms are linear in the size of the constructed araucarias or semi-araucarias. In Section 8, we give the explicit complexity of these algorithms.

4. Maximal paths and capacity chains

A maximal path in an araucaria corresponds to a maximal sequence of transitions with the same letter in the minimal automaton containing this araucaria. The right knowledge of maximal paths is thus important in the study of the minimal automaton of the shuffle of words. The notion of capacity of an edge introduced in this section, along with the notion of truncation given in [10], leads to a characterization of the set of maximal paths of an araucaria.

Definition 4.1. (i) For every vertex s of a directed tree B, let $\lambda(s)$ be the maximum length of the paths of B whose first vertex is s. If $\sigma = (s_0, \ldots, s_f)$ is a path of B, then every vertex s_j of $\sigma \setminus \{s_0, s_f\}$ such that $\lambda(s_j) > \lambda(s_{j+1}) + 1$ is called a breaking vertex of σ.

(ii) Let $\sigma = (s_0, \ldots, s_f)$ be a path of B and let $s_{j_1}, \ldots, s_{j_{t-1}}$ be the breaking vertices of σ such that $0 < j_1 < \ldots < j_{t-1} < f$.

The paths $\tau_1 = (s_0, \ldots, s_{j_1})$, $\tau_2 = (s_{j_1}, \ldots, s_{j_2})$, …, $\tau_t = (s_{j_{t-1}}, \ldots, s_f)$ are called the truncations of σ and the product $\tau_1 ... \tau_t$ is called the decomposition of σ into truncations.

A truncation τ of a maximal path of B is simply called a truncation of B. Its starting vertex is then either a breaking vertex of this path or the root of B.

(iii) A truncation τ of a path σ of B is called terminal if its last vertex is a leaf of B.

(iv) We say that a directed tree B satisfies the unicity condition if, for every vertex s of B which is not a leaf, there exists only one successor t of s such that $\lambda(s) = \lambda(t) + 1$ (i.e., there is a unique longest path starting at s).

Lemma 4.2. Let B be a semi-araucaria [resp. an araucaria] of type $(p_1, \ldots, p_k)$.

(i) The trunk of B is its unique longest path and does not contain any breaking vertex.

(ii) Every edge (s, t) of B belongs to a unique terminal truncation $\tau_{(s,t)}$ of B.

(iii) Every terminal truncation of the semi-araucaria B is the trunk of a unique subsemi-araucaria of B of type $(p_{i_1}, \ldots, p_{i_m})$ where $\{i_1, \ldots, i_m\} \subseteq K$.

(iv) B satisfies the unicity condition.

Proof:
First, we study the case where B is a semi-araucaria.

All these properties are trivially satisfied by semi-araucarias of arity 1. Assume now that they also hold true for every arity strictly smaller than k with $k > 1$.

(i) Let $\tau = (s_0, \ldots, s_p)$ be the trunk of B. By Definition 2.2, for every non-empty proper subset $I = \{i_1, \ldots, i_m\}$ of K and every $h \in \{1, \ldots, \sum_{i \in K \setminus I} p_i - 1\}$, the length of the trunk τ_I of the subsemi-araucaria $H_{I,h}$ of type $(p_{i_1}, \ldots, p_{i_m})$ whose root is the vertex s_h of τ is equal to $\sum_{i \in I} p_i$. By induction, τ_I is the longest path of $H_{I,h}$. Hence, for every path σ of B which contains the first edge of τ_I,

$$|\sigma| \leq h + \sum_{i \in I} p_i < \sum_{i \in K \setminus I} p_i + \sum_{i \in I} p_i = \sum_{i \in K} p_i = |\tau|.$$

It follows that τ is the longest path of B and does not contain any breaking vertex.

(ii) The property is trivially satisfied by every edge of the trunk of B. For every other edge (s, t) of B, there exist a unique non-empty proper subset I of K and a unique $h \in \{1, \ldots, \sum_{i \in K \setminus I} p_i - 1\}$ such that (s, t) is contained in the subsemi-araucaria $H_{I,h}$. By induction, (s, t) belongs to a unique terminal truncation $\tau_{(s,t)}$ of a maximal path σ of $H_{I,h}$. If $(r, \ldots, s_h)$ is the path of B from the root r of B to the root s_h of $H_{I,h}$, the path $\sigma' = (r, \ldots, s_h)\sigma$ is maximal in B and, by (i), s_h is a breaking vertex of σ'. It follows that $\tau_{(s,t)}$ is also the unique terminal truncation that contains (s, t) in B.

(iii) Let τ' be a terminal truncation of B. Since τ' does not contain any breaking vertex and is maximal with respect to this property, if τ' contains an edge of the trunk τ of B, then $\tau' = \tau$ by (i), and the property is true. The other case follows by induction, as in the proof of (ii).

(iv) The result is true for the root r of B since r has a unique successor. For any other vertex t, let s be the unique predecessor of t. By (ii), (s, t) belongs to a unique terminal truncation $\tau_{(s,t)}$ of B. If $\tau_{(s,t)}$ is the trunk of B, the result follows from (i). In the other cases, by (iii), $\tau_{(s,t)}$ is the trunk of a unique subsemi-araucaria of arity strictly smaller than k and the result holds by induction.

By Definition 2.3, all these properties hold also for araucarias. □

Definition 4.3. For every araucaria A of type $(p_1, \ldots, p_k)$, let cap be the mapping which associates to every edge (s, t) of A the subset $\{i_1, \ldots, i_m\}$ of K such that the terminal truncation $\tau_{(s,t)}$ of A that contains (s, t) is the trunk of a subsemi-araucaria of A of type $(p_{i_1}, \ldots, p_{i_m})$.

Example 4.4. In the araucaria A of type $(p_1, p_2, p_3) = (3, 2, 1)$ of Fig. 2(b), the decomposition into truncations of the maximal path $\sigma = (s_0, s_1, s_2, \ldots, e_2)$ is equal to $\tau_1\tau_2$ where $\tau_1 = (s_0, s_1, s_2)$ and $\tau_2 = (s_2, \ldots, e_2)$. τ_1 is a subpath of the trunk of the subsemi-araucaria of type $(3, 2, 1)$ and τ_2 is the trunk of a subsemi-araucaria of type $(p_2, p_3) = (2, 1)$. Thus, for every edge (s, t) of τ_1, $cap(s, t) = \{1, 2, 3\}$ and, for every edge (s, t) of τ_2, $cap(s, t) = \{2, 3\}$.

Definition 4.5. (i) Let $B = (V, E)$ be a directed tree and let cap_B be a mapping from the set of edges E into the set of non-empty subsets of the set $\mathbb{N}^*$ of positive integers. The mapping cap_B is called a capacity function of B if, for every path $\gamma = (s_1, s_2, s_3)$ of B, $cap_B(s_2, s_3) = cap_B(s_1, s_2)$ if s_2 is not a breaking vertex of γ, and $cap_B(s_2, s_3) \subsetneq cap_B(s_1, s_2)$ in the converse case.

Moreover, if, for every distinct successors t_1 and t_2 of each vertex s, $cap_B(s, t_1) \neq cap_B(s, t_2)$, then the capacity function cap_B is said to be discriminating.

(ii) If cap_B is a capacity function of B then, for every edge $(s, t) \in E$, $cap_B(s, t)$ is called the capacity of (s, t).

If τ is a path of B whose edges have the same capacity, then this capacity is called the capacity of τ and is denoted by $cap_B(\tau)$.

Lemma 4.6. The mapping cap is a discriminating capacity function of the araucaria A.

Proof:
(i) For every path $\gamma = (s_1, s_2, s_3)$ of A, if s_2 is not a breaking vertex of γ, the terminal truncations $\tau_{(s_2,s_3)}$ and $\tau_{(s_1,s_2)}$ are equal and hence $cap(s_2, s_3) = cap(s_1, s_2)$. In the converse case, $cap(s_2, s_3) \subsetneq cap(s_1, s_2)$ since $\tau_{(s_2,s_3)} \neq \tau_{(s_1,s_2)}$ and $\tau_{(s_2,s_3)}$ is a terminal truncation of the subsemi-araucaria which admits $\tau_{(s_1,s_2)}$ as trunk. Thus cap is a capacity function.

(ii) Moreover, we prove that cap is discriminating.

If $k = 1$, A is a path and the property is trivial. Hence, if $k > 1$, it is also the case for every subsemi-araucaria of A of arity 1. Assume that the property is satisfied for every subsemi-araucaria of arity strictly smaller than k and let t_1 and t_2 be two distinct successors of a vertex s in an araucaria A of arity $k > 1$. If s is not contained in the trunk τ of A, $cap(s, t_1) \neq cap(s, t_2)$ by induction. Suppose now that s is a vertex of τ of height h. If none of t_1 and t_2 is contained in τ, by Definitions 2.3 and 2.2, there exist two non-empty distinct proper subsets I_1 and I_2 of K such that the edges (s, t_1) and (s, t_2) are respectively contained in the trunks of the subsemi-araucarias $H_{I_1,h}$ and $H_{I_2,h}$. By Lemma 4.2, these trunks are the terminal truncations $\tau_{(s,t_1)}$ and $\tau_{(s,t_2)}$. Thus $cap(s, t_1) = I_1 \neq I_2 = cap(s, t_2)$. Otherwise, only one of t_1 and t_2 is contained in τ and either $cap(s, t_1) = K \neq I_2 = cap(s, t_2)$ or $cap(s, t_1) = I_1 \neq K = cap(s, t_2)$.

Hence cap is discriminating. □

Definition 4.7. Let B be a directed tree with capacity function cap_B and let $MP(B)$ be the set of maximal paths of B. If $\tau_1 ... \tau_f$ is the decomposition of a maximal path σ of B into truncations, let $cap_B(\sigma) = (cap_B(\tau_1), \ldots, cap_B(\tau_f))$ and $\pi(\sigma) = (|\tau_1|, \ldots, |\tau_f|)$.

The pair $(cap_B(\sigma), \pi(\sigma))$ is called the valued capacity chain of σ (relatively to cap_B).

The mapping κ such that, for every $\sigma \in MP(B)$, $\kappa(\sigma) = (cap_B(\sigma), \pi(\sigma))$, is said to be associated to the capacity function cap_B.

Definition 4.8. A pair $((I_1, \ldots, I_f), (q_1, \ldots, q_f))$ where $f > 0$, $\emptyset \neq I_f \subsetneq \ldots \subsetneq I_1 \subseteq K$ and $q_1, \ldots, q_f$ are positive integers, is said to be linked to $(p_1, \ldots, p_k)$ (relatively to K) if,

$$\text{for all } \ r \in \{1, \ldots, f-1\}, \ 0 < q_r < \sum_{j \in I_r \setminus I_{r+1}} p_j \ \text{ and } \ q_f = \sum_{j \in I_f} p_j.$$

Let $Link(p_1, \ldots, p_k)$ be the set of all pairs which are linked to $(p_1, \ldots, p_k)$.

Remark 4.9. Let $((I_1, \ldots, I_f), (q_1, \ldots, q_f))$ be a pair linked to $(p_1, \ldots, p_k)$ (relatively to K). For every non-empty proper subset $J = \{j_1, \ldots, j_m\}$ of K such that $I_1 \subseteq J$, $((I_1, \ldots, I_f), (q_1, \ldots, q_f))$ is also linked to $(p_{j_1}, \ldots, p_{j_m})$ (relatively to J).

It follows that $Link(p_{j_1}, \ldots, p_{j_m}) \subset Link(p_1, \ldots, p_k)$.

Lemma 4.10. The valued capacity chain of every maximal path of the araucaria A (relatively to cap) is linked to $(p_1, \ldots, p_k)$.

Proof:
For every subsemi-araucaria $H_{\{i\},h}$ of A of arity 1, the trunk τ of $H_{\{i\},h}$ is its unique longest path and the valued capacity chain $((\{i\}), (|\tau|))$ is linked to (p_i). Assume by induction that, for every subsemi-araucaria $H_{I,h}$ of A, where $I = \{i_1, \ldots, i_m\}$ is a non-empty proper subset of K, and for every maximal path σ of $H_{I,h}$, $(cap(\sigma), \pi(\sigma))$ is linked to $(p_{i_1}, \ldots, p_{i_m})$. Let $\sigma = (s_0, \ldots, s_t)$ be a maximal path of A, and τ the trunk of A.

If $(s_0, s_1) \notin \tau$, there exists a non-empty proper subset $I = \{i_1, \ldots, i_m\}$ of K such that (s_0, s_1) is contained in the trunk τ_I of the subsemi-araucaria $H_{I,0}$ of A of type $(p_{i_1}, \ldots, p_{i_m})$. By the induction hypothesis, $(cap(\sigma), \pi(\sigma))$ is linked to $(p_{i_1}, \ldots, p_{i_m})$ and, by Remark 4.9, to $(p_1, \ldots, p_k)$.

If $\sigma = \tau$, then $cap(\sigma) = (K)$ and $\pi(\sigma) = |\tau| = \sum_{i \in K} p_i$. Hence $(cap(\sigma), \pi(\sigma))$ is linked to $(p_1, \ldots, p_k)$.

If $(s_0, s_1) \in \tau$ and $\sigma \neq \tau$, there exist a non-empty proper subset $I = \{i_1, \ldots, i_m\}$ of K and an integer h of $\{1, \ldots, \sum_{j \in K \setminus I} p_j - 1\}$ such that the first truncation of σ is $\tau_1 = (s_0, \ldots, s_h)$, where s_h is the vertex of height h of τ, and $\sigma' = (s_h, \ldots, s_t)$ is a maximal path of the subsemi-araucaria $H_{I,h}$ of type $(p_{i_1}, \ldots, p_{i_m})$. By induction, $(cap(\sigma'), \pi(\sigma'))$ is linked to $(p_{i_1}, \ldots, p_{i_m})$. Since $\sigma = \tau_1 \sigma'$, we have $cap(\sigma) = (K, cap(\sigma'))$ and $\pi(\sigma) = (h, \pi(\sigma'))$. Thus $(cap(\sigma), \pi(\sigma))$ is linked to $(p_1, \ldots, p_k)$. □

Corollary 4.11. Let σ be a path from the root of A to a vertex which is not a leaf of A and $\tau_1 ... \tau_f$ the decomposition of σ into truncations. Then,

$$\text{for all } r \in \{1, \ldots, f-1\},\ 0 < |\tau_r| < \sum_{j \in cap(\tau_r) \setminus cap(\tau_{r+1})} p_j \text{ and } 0 < |\tau_f| < \sum_{j \in cap(\tau_f)} p_j.$$

Proof:
By Lemma 4.2, the first edge (u, v) of the last truncation τ_f belongs to a unique terminal truncation τ'_f. Since τ_f and τ'_f do not contain breaking vertices distinct from u and since A satisfies the unicity condition, τ_f is a subpath of τ'_f and, since the last vertex of τ'_f is a leaf, $\tau_f \neq \tau'_f$. Then $\sigma' = \tau_1 ... \tau_{f-1} \tau'_f$ is a maximal path of A and, by Lemma 4.10,

$$\text{for all } r \in \{1, \ldots, f-1\},\ 0 < |\tau_r| < \sum_{j \in cap(\tau_r) \setminus cap(\tau_{r+1})} p_j \text{ and } 0 < |\tau_f| < |\tau'_f| = \sum_{j \in cap(\tau_f)} p_j.$$

□

We prove now that the property of the valued capacity chain of the maximal paths given in Lemma 4.10 characterizes the araucarias.

Theorem 4.12. A directed tree B is isomorphic to an araucaria of type $(p_1, \dots, p_k)$ if and only if it admits a discriminating capacity function whose associated mapping is a bijection onto $Link(p_1, \dots, p_k)$.

Proof:
(i) Let A be an araucaria of type $(p_1, \dots, p_k)$. By Lemma 4.10, if κ is the mapping associated to cap, $\kappa(\sigma) \in Link(p_1, \dots, p_k)$ for every maximal path σ of A.

We prove now that κ is one-to-one.

Let σ and σ' two distinct maximal paths of $MP(A)$ and let $\tau_1 \dots \tau_f$ and $\tau'_1 \dots \tau'_{f'}$ be their decompositions into truncations. There exists a smallest $i \in \{1, \dots, \min(f, f')\}$ such that $\tau_i \neq \tau'_i$. If τ_i and τ'_i have only their first vertex in common, $cap(\tau_i) \neq cap(\tau'_i)$ since cap is discriminating. Otherwise, since A satisfies the unicity condition, either $\tau_i \subset \tau'_i$ or $\tau'_i \subset \tau_i$ (in the converse case one of τ_i and τ'_i would have a breaking vertex) and $|\tau_i| \neq |\tau'_i|$. Thus $\kappa(\sigma) \neq \kappa(\sigma')$.

We prove now, by induction on k and by using the results of Theorem 2.5, that κ is surjective.

This is trivial for $k = 1$. Assume that, for every subaraucaria $A_{J,h}$ of A of arity strictly smaller than k (as defined in Theorem 2.5), $\kappa(MP(A_{J,h})) = Link(p_{j_1}, \dots, p_{j_m})$ where $J = \{j_1, \dots, j_m\}$.

Let $((I_1, \dots, I_f), (q_1, \dots, q_f)) \in Link(p_1, \dots, p_k)$.

If $I_1 = \{i_1, \dots, i_m\} \neq K$, $((I_1, \dots, I_f), (q_1, \dots, q_f)) \in Link(p_{i_1}, \dots, p_{i_m})$. By the induction hypothesis, there exists a maximal path σ of the subaraucaria $A_{I_1,0}$ of A such that $cap(\sigma) = (I_1, \dots, I_f)$ and $\pi(\sigma) = (q_1, \dots, q_f)$. Moreover σ is also maximal in A.

If $I_1 = K$ and $f = 1$, by Definitions 4.8 and 2.3, $q_1 = \sum_{j \in K} p_j$ is the length of the trunk τ of A and $\kappa(\tau) = ((K), (|\tau|))$.

If $I_1 = K$ and $f > 1$, let τ_1 be the initial section of length q_1 of the trunk of A. Since $q_1 < \sum_{j \in K \setminus I_2} p_j$, it follows from Theorem 2.5 that, setting $I_2 = \{i_1, \dots, i_m\}$, A admits a subaraucaria A_{I_2,q_1} of type $(p_{i_1}, \dots, p_{i_m})$. By the induction hypothesis, there exists a maximal path σ' of A_{I_2,q_1} such that $cap(\sigma') = (I_2, \dots, I_f)$ and $\pi(\sigma') = (q_2, \dots, q_k)$. Then $\sigma = \tau_1\sigma'$ is a maximal path of A such that $cap(\sigma) = (K, cap(\sigma'))$ and $\pi(\sigma) = (q_1, \pi(\sigma')) = (q_1, \dots, q_f)$.

It follows that κ is a bijection.

(ii) Conversely, assume that B is a directed tree which admits a discriminating capacity function cap' whose associated mapping κ' is a bijection from $MP(B)$ onto $Link(p_1, \dots, p_k)$.

By (i), for every maximal path σ of the araucaria A, there exists a maximal path σ' of B such that $\kappa'(\sigma') = \kappa(\sigma) = ((I_1, \dots, I_f), (q_1, \dots, q_f))$. Hence $|\sigma'| = \sum_{g=1}^{f} q_g = |\sigma|$, and there exists an isomorphism φ from the path σ onto σ'.

Let now $\{\sigma_1, \dots, \sigma_l\}$ be a non-empty proper subset of paths of $MP(A)$, C the minimal subtree of A which contains $\{\sigma_1, \dots, \sigma_l\}$, and C' the minimal subtree of B which contains $\{\sigma'_1, \dots, \sigma'_l\}$ where $\sigma'_i = \kappa'^{-1} \circ \kappa(\sigma_i)$ for all $i \in \{1, \dots, l\}$. Assume, by induction, that there exists an isomorphism φ from C onto C' such that, for all $i \in \{1, \dots, l\}$, $\varphi(\sigma_i) = \sigma'_i$.

Let now $\sigma = (s_0, \dots, s_f)$ be a path of $MP(A) \setminus \{\sigma_1, \dots, \sigma_l\}$, $\sigma' = \kappa'^{-1} \circ \kappa(\sigma) = (s'_0, \dots, s'_f)$, and $s_i \in \{s_0, \dots, s_{f-1}\}$ be the highest vertex of σ belonging to C. Since the capacity function cap is discriminating, for every successor t of s_i in C, $cap(s_i, t) \neq cap(s_i, s_{i+1})$. Since cap' is also a discriminating capacity function and since the edges of σ' satisfy $cap'(s'_j, s'_{j+1}) = cap(s_j, s_{j+1})$, $(s'_0, \dots, s'_i)$ is the path $\varphi((s_0, \dots, s_i))$ of C'. For the same reason, (s'_i, s'_{i+1}) does not belong to C'. It follows that φ can be extended to an isomorphism φ' from $C \cup \sigma$ onto $C' \cup \sigma'$ such that $\varphi'(\sigma) = \sigma'$.

Hence, by induction, there exists a one-to-one morphism φ from A into B. Since $|MP(B)| = |MP(A)|$, φ is an isomorphism from A onto B. □

Remark 4.13. It follows from the proofs of Lemma 4.10 and Theorem 4.12 that, if $I = \{i_1, \ldots, i_m\}$ is a proper subset of K, a directed tree is an araucaria of type $(p_{i_1}, \ldots, p_{i_m})$ if and only if it admits a discriminating capacity function whose associated mapping is a bijection onto $Link(p_{i_1}, \ldots, p_{i_m})$.

Remark 4.14. Note that the "if" part of Theorem 4.12 does not hold if the capacity function is not discriminating.

Remark 4.15. (i) The algorithms of Section 3 construct the araucarias and semi-araucarias in such a way that, if s is a vertex which is not a leaf and if t is the first child of s, then the subsemi-araucaria trunk that contains (s, t) is the longest of all subsemi-araucaria trunks that contain s. Hence, s is a breaking vertex for all maximal paths that pass through a child of s other than the first one. It follows that the decomposition into truncations of any path in the tree can be reported while traversing the path.

(ii) If the capacities of the edges need to be stored in the data structure, it suffices to pass the parameter I of the function $SemiAraucariaWithoutItsRoot$ to the function $Path$ and to store I in all the vertices of the path generated by the function $Path$. At the end of the algorithm, the capacity of every edge (s, t) will then have been stored in the vertex t.

Remark 4.16. In [11], araucarias have been defined using attribution functions. These attribution functions and our capacity function are closely related. Indeed, if A is an araucaria of type $(p_1, \ldots, p_k)$ with capacity function cap, σ a maximal path of A, and $\tau_1 \ldots \tau_f$ the decomposition of σ into truncations then, for every $r \in \{1, \ldots, f-1\}$, the set $cap(\tau_r) \setminus cap(\tau_{r+1})$ can be attributed to the truncation τ_r, and $cap(\tau_f)$ can be attributed to τ_f. More precisely, τ_r [resp. τ_f] can be factorized into sections and each section can be attributed to an integer of $cap(\tau_r) \setminus cap(\tau_{r+1})$ [resp. $cap(\tau_f)$]. Such attribution functions are natural in the proof that every araucaria can be embedded in the minimal automaton of a shuffle of words, since every transition of this automaton is generated by a letter of one of these words and, hence, can be attributed to the index of this word. In the automaton of Fig. 1 for example, the a-transition from $4^{(1)}$ to $4^{(2)}$ can be attributed to the second letter of the word $caaac$. The a-transition from $2^{(2)}$ to $\{2^{(3)}, 3^{(2)}\}$ can be attributed either to the third letter of $caaac$ or to the second letter of $baab$.

Hence, a truncation in an araucaria can have several attribution sets while it admits a unique capacity set. This suggests that capacities are more interesting in the study of araucarias than attributions.

5. The first grafting theorem

In [10], it has been shown that an araucaria of arity k can be obtained by sticking araucarias of arity 1 on the vertices of araucarias of arity $k - 1$. This operation was called ramification. We introduce here the grafting of a directed tree onto another which generalizes the notion of ramification. If the two trees admit a capacity function, we define a capacity function for the resulting tree. In the particular case where the two trees are araucarias of respective arities k_1 and k_2, we prove that the resulting tree is a subtree of an araucaria of arity $k_1 + k_2$. Moreover, for every m such that $1 \leq m < k$, we show that every araucaria of arity k can be obtained by grafting araucarias of arity m onto araucarias of arity $k - m$ and by then merging the resulting trees.

Definition 5.1. Let $B = (S, U)$ and $C = (T, V)$ be directed trees and P a subset of S.

Let, for each vertex p of P, $C(p) = (T_p, V_p)$ be a directed tree isomorphic to C with root p, such that $S \cap T_p = \{p\}$ and, for each vertex q of P distinct from p, $T_p \cap T_q = \emptyset$.

The directed tree which is the union of B and of the trees $C(p)$, for all $p \in P$, is said to be obtained by grafting C onto B at P and is denoted $graft_P(C/B)$.

If $P = S$, this directed tree is said to be obtained by completely grafting C onto B and is denoted $graft(C/B)$ (see Fig. 3).

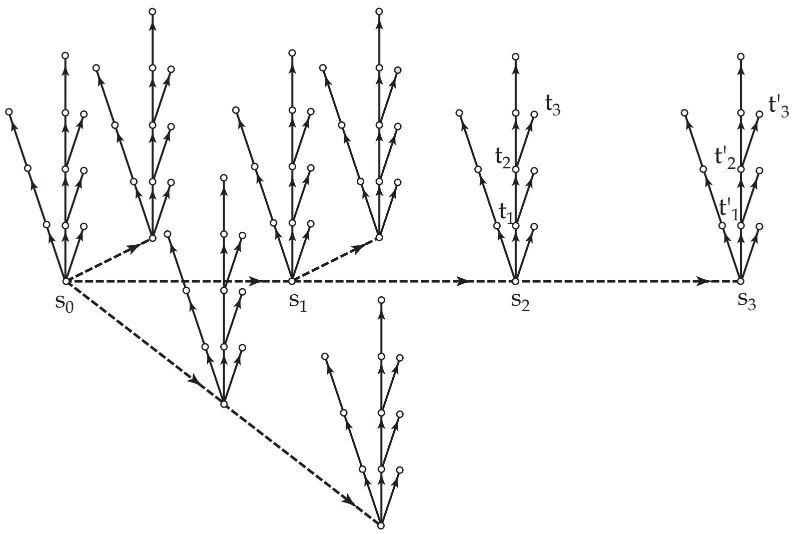


Figure 3. The tree obtained by completely grafting an araucaria of type $(3, 1)$ (full lines) onto an araucaria of type $(2, 1)$ (dotted lines). The decomposition into truncations of the maximal path $(s_0, \ldots, t_3)$ is $(s_0, s_1, s_2)(s_2, t_1, t_2)(t_2, t_3)$ and that of $(s_0, \ldots, t'_3)$ is $(s_0, \ldots, t'_2)(t'_2, t'_3)$.

Lemma 5.2. Let B and C be directed trees that satisfy the unicity condition and admit the respective discriminating capacity functions cap_B and cap_C such that, for every edge (s, t) of B and every edge (s', t') of C, $cap_B(s, t) \cap cap_C(s', t') = \emptyset$.

The tree $D = graft(C/B)$ satisfies the unicity condition and admits a discriminating capacity function cap_D.

Proof:

(i) Since B satisfies the unicity condition, for every vertex s of B, there exists a unique longest path $\sigma = (s, s_1, \ldots, s_h)$ of B starting at s. Let $\lambda_1(s)$ be the length of this path. In the same way, there exists a unique longest path $\tau_{(s_h)}$ in $C(s_h)$, and its length λ_2 is independent of s_h. $\sigma\tau_{(s_h)}$ is then the unique longest path of D starting at s. Hence, setting $\lambda(s)$ the length of this path, $\lambda(s) = \lambda_1(s) + \lambda_2$. Moreover, for every vertex s of B and every vertex u of $C(s)$, there exists a unique longest path of $C(s)$ starting at u and this path is also the longest path starting at u in D. Hence, setting respectively $\lambda_2(u)$ and $\lambda(u)$ the lengths of these paths, $\lambda_2(u) = \lambda(u)$. It follows that D satisfies the unicity condition.

(ii) Every maximal path σ of D is the product of a path σ_1 of B (possibly of length zero) and of a maximal path σ_2 of $C(s)$, where s is the last vertex of σ_1. The length of σ_1 is zero if and only if s is the root of B. Since D satisfies the unicity condition and since, for every vertex u of B, $\lambda(u) = \lambda_1(u) + \lambda_2$, the breaking vertices of σ_1 in B are also the breaking vertices of σ_1 in D. In the same way, since, for every vertex u of $C(s)$, $\lambda(u) = \lambda_2(u)$, the breaking vertices of σ_2 in $C(s)$ are also the breaking vertices of σ_2 in D. Moreover, since $\lambda(s) = \lambda_1(s) + \lambda_2$, s is a breaking vertex for σ if and only if, either s is not a leaf of B, or s is a leaf of B but the first edge of σ_2 is not contained in the longest path $\tau_{(s)}$ of $C(s)$.

Let $\tau_1 ... \tau_f$ be the decomposition of σ_1 into truncations in D and $\tau_{f+1} ... \tau_g$ be the decomposition of σ_2 into truncations in $C(s)$. Then, from above, the decomposition into truncations of σ in D is $\tau_1 ... \tau_f \tau_{f+1} ... \tau_g$ if, either s is not a leaf of B, or s is a leaf of B but the first edge of σ_2 is not contained in $\tau_{(s)}$. In the converse case, the decomposition is $\tau_1 ... \tau_{f-1} \tau'_f \tau_{f+2} ... \tau_g$, where $\tau'_f = \tau_f \tau_{f+1}$ is a truncation.

(iii) For every edge (u, v) of B, let $cap_D(u, v) = cap_B(u, v) \cup cap_C(\tau)$, where τ is the longest path of C. Since, for every $s \in B$, the tree $C(s)$ is isomorphic to C, it admits a discriminating capacity function corresponding to cap_C. For every edge (u, v) of $C(s)$, let $cap_D(u, v) = cap_C(u, v)$, except when s is a leaf of B and (u, v) is an edge of the longest path $\tau_{(s)}$ of $C(s)$. In this exception case, s admits a predecessor s' in B and we set $cap_D(u, v) = cap_D(s', s)$, for every edge (u, v) of $\tau_{(s)}$.

We prove now that cap_D is a capacity function. Let $\gamma = (s_1, s_2, s_3)$ be a path of D.

Case 1: γ is in B. From (ii), s_2 is a breaking vertex of γ in B if and only if it is also a breaking vertex of γ in D. Hence $cap_B(s_1, s_2) = cap_B(s_2, s_3)$ [resp. $cap_B(s_2, s_3) \subsetneq cap_B(s_1, s_2)$] if and only if $cap_D(s_1, s_2) = cap_D(s_2, s_3)$ [resp. $cap_D(s_2, s_3) \subsetneq cap_D(s_1, s_2)$].

Case 2: $s_1 \in B$ and $s_3 \in C(s_2)$. If s_2 is a leaf of B and $s_3 \in \tau_{(s_2)}$, s_2 is not a breaking vertex of γ in D and $cap_D(s_1, s_2) = cap_D(s_2, s_3)$. Otherwise, s_2 is a breaking vertex of γ in D and $cap_D(s_2, s_3) \subseteq cap_C(\tau_{(s_2)}) \subsetneq cap_D(s_1, s_2)$.

Case 3: γ is in some $C(s)$. The result is trivial when γ is in $\tau_{(s)}$ or when no edge of γ is in $\tau_{(s)}$. Otherwise, s_2 is a breaking vertex of γ in D and $cap_D(s_2, s_3) \subsetneq cap_C(\tau_{(s)}) \subsetneq cap_D(s_1, s_2)$.

Hence cap_D is a capacity function.

Moreover, since cap_B and cap_C are discriminating, the same easily follows for cap_D. □

Definition 5.3. (i) The capacity function cap_D of D defined in the preceding proof is said to be obtained by grafting cap_C onto cap_B and is denoted by $cap_D = graft(cap_C / cap_B)$.

(ii) For every non-empty proper subset I of K, let A_I be an araucaria isomorphic to the subaraucaria $A_{I,0}$ of A and whose capacity function cap_I is the restriction of cap to $A_{I,0}$.

(iii) Let B_I be the directed tree $graft(A_I / A_{K \setminus I})$ and $bcap_I = graft(cap_I / cap_{K \setminus I})$ its capacity function.

Example 5.4. Let A be the araucaria of type $(p_1, p_2, p_3, p_4) = (3, 2, 1, 1)$ and $I = \{1, 4\}$.

A_I is then the araucaria of type $(p_1, p_4) = (3, 1)$, $A_{K \setminus I}$ the araucaria of type $(p_2, p_3) = (2, 1)$, and $B_I = graft(A_{\{1,4\}} / A_{\{2,3\}})$ is the tree of Fig. 3.

In the araucaria $A_{K \setminus I}$, we have $cap_{K \setminus I}(s_0, s_1, s_2) = cap_{K \setminus I}(s_0, s_1, s_2, s_3) = K \setminus I = \{2, 3\}$. In $A_I(s_2)$, $cap_I(s_2, t_1, t_2) = I = \{1, 4\}$ and $cap_I(t_2, t_3) = \{4\}$. In $A_I(s_3)$, $cap_I(s_3, t'_1, t'_2) = \{1, 4\}$ and $cap_I(t'_2, t'_3) = \{4\}$. We then have in B_I, $bcap_I(s_0, s_1, s_2) = bcap_I(s_0, s_1, s_2, s_3, t'_1, t'_2) = \{1, 2, 3, 4\}$, $bcap_I(s_2, t_1, t_2) = \{1, 4\}$, and $bcap_I(t_2, t_3) = bcap_I(t'_2, t'_3) = \{4\}$.

Proposition 5.5. The mapping κ_I associated to the capacity function $bcap_I$ of B_I is a bijection onto the subset of pairs of $Link(p_1, \ldots, p_k)$ that are of the form $((I_1, \ldots, I_g), (q_1, \ldots, q_g))$ with either $I_h \subseteq I$ or $I \subseteq I_h$, for all $h \in \{1, \ldots, g\}$.

Proof:
(i) Since, by Lemma 5.2, $bcap_I$ is a discriminating capacity function and B_I satisfies the unicity condition, the mapping κ_I associated to $bcap_I$ is one-to-one, as in proof (i) of Theorem 4.12.

We show now that the valued capacity chain $\kappa_I(\sigma) = (bcap_I(\sigma), \pi(\sigma))$ of any maximal path σ of B_I is linked to $(p_1, \ldots, p_k)$.

Let σ_1 be the subpath of σ contained in $A_{K \setminus I}$, s the last vertex of σ_1 (or the unique vertex of σ_1 when $|\sigma_1| = 0$), and σ_2 the subpath of σ contained in $A_I(s)$. Let $\tau_1 ... \tau_f$ and $\tau_{f+1} ... \tau_g$ be the decompositions into truncations of σ_1 in $A_{K \setminus I}$ and of σ_2 in $A_I(s)$. Let $cap_{K \setminus I}(\sigma_1) = (I_1, \ldots, I_f)$ and $cap_I(\sigma_2) = (I_{f+1}, \ldots, I_g)$. From proof (ii) of Lemma 5.2, the decomposition into truncations of σ in B_I is either $\tau_1 ... \tau_f \tau_{f+1} ... \tau_g$ or $\tau_1 ... \tau_{f-1} \tau'_f \tau_{f+2} ... \tau_g$, where $\tau'_f = \tau_f \tau_{f+1}$ is a truncation.

By definition of $bcap_I$, in the first case

$$bcap_I(\sigma) = (I_1 \cup I, \ldots, I_f \cup I, I_{f+1}, \ldots, I_g)$$

and in the second case

$$bcap_I(\sigma) = (I_1 \cup I, \ldots, I_f \cup I, I_{f+2}, \ldots, I_g).$$

In both cases, $bcap_I(\sigma)$ is an inclusion chain of K.

Now, we consider $\pi(\sigma)$.

Case 1.- If s is not a leaf of $A_{K \setminus I}$, from proof of Lemma 5.2, the decomposition into truncations of σ is $\tau_1 ... \tau_f \tau_{f+1} ... \tau_g$. If $\sigma_1 \neq (s)$, by Corollary 4.11, for every $h \in \{1, \ldots, f-1\}$,

$$0 < |\tau_h| < \sum_{i \in I_h \setminus I_{h+1}} p_i = \sum_{i \in (I_h \cup I) \setminus (I_{h+1} \cup I)} p_i \text{ and } 0 < |\tau_f| < \sum_{i \in I_f} p_i \leq \sum_{i \in (I_f \cup I) \setminus I_{f+1}} p_i.$$

In the same way, by Lemma 4.10, whether σ_1 is reduced to (s) or not,

$$\text{for all } h \in \{f+1, \ldots, g-1\}, \; 0 < |\tau_h| < \sum_{i \in I_h \setminus I_{h+1}} p_i \text{ and } 0 < |\tau_g| = \sum_{i \in I_g} p_i.$$

Case 2.- If s is a leaf of $A_{K \setminus I}$ but τ_{f+1} does not belong to the trunk of $A_I(s)$, the decomposition into truncations of σ is $\tau_1 ... \tau_f \tau_{f+1} ... \tau_g$. Then, for all $h \in \{1, \ldots, g\} \setminus \{f\}$, $|\tau_h|$ satisfies the same relations as in case 1. Moreover, from Lemma 4.10,

$$0 < |\tau_f| = \sum_{i \in I_f} p_i < \sum_{i \in (I_f \cup I) \setminus I_{f+1}} p_i.$$

Case 3.- If s is a leaf of $A_{K \setminus I}$ and τ_{f+1} belongs to the trunk of $A_I(s)$, the decomposition into truncations of σ is $\tau_1 ... \tau_{f-1} \tau'_f \tau_{f+2} ... \tau_g$. For every $h \in \{1, \ldots, g\} \setminus \{f, f+1\}$, $|\tau_h|$ satisfies the same relations as in case 1. Moreover, by Lemma 4.10, if $f + 1 \neq g$, then

$$0 < |\tau'_f| = |\tau_f| + |\tau_{f+1}| < \sum_{i \in I_f} p_i + \sum_{i \in I_{f+1} \setminus I_{f+2}} p_i = \sum_{i \in I_f} p_i + \sum_{i \in I \setminus I_{f+2}} p_i = \sum_{i \in (I_f \cup I) \setminus I_{f+2}} p_i.$$

In the same way, if $f+1=g$,

$$0<|\tau'_f| = \sum_{i\in I_f} p_i + \sum_{i\in I_{f+1}} p_i = \sum_{i\in I_f\cup I} p_i.$$

It follows that, in all three cases, $(bcap_I(\sigma), \pi(\sigma)) \in Link(p_1,\dots,p_k)$.

Moreover, by definition of $bcap_I$, for every edge (u,v) of σ, either $I\subseteq bcap_I(u,v)$ or $bcap_I(u,v)\subseteq I$.

(ii) Let now $((I_1,\dots,I_g),(q_1,\dots,q_g)) \in Link(p_1,\dots,p_k)$ such that, for all $h\in\{1,\dots,g\}$, either $I\subseteq I_h$ or $I_h\subseteq I$. Set $\{i_1,\dots,i_m\}=I$ and $\{j_1,\dots,j_{k-m}\}=K\setminus I$. We show that there exists a maximal path σ of B_I such that $\kappa_I(\sigma)=((I_1,\dots,I_g),(q_1,\dots,q_g))$.

Case 1.- If $I_1\subseteq I$, by Remark 4.9, $((I_1,\dots,I_g),(q_1,\dots,q_g)) \in Link(p_{i_1},\dots,p_{i_m})$. Hence, if s is the root of B_I, there exists, by Remark 4.13, a maximal path σ of $A_I(s)$ such that $bcap_I(\sigma)=cap_I(\sigma)=(I_1,\dots,I_g)$ and $\pi(\sigma)=(q_1,\dots,q_g)$.

Case 2.- If there exists $f\in\{1,\dots,g-1\}$ such that $I_{f+1}\subseteq I\subsetneq I_f$, let, for every $h\in\{1,\dots,f\}$, $I'_h=I_h\setminus I$. Then $\emptyset\neq I'_f\subsetneq\dots\subsetneq I'_1\subseteq K\setminus I$ and, for all $h\in\{1,\dots,f-1\}$,

$$0<q_h<\sum_{i\in I_h\setminus I_{h+1}} p_i = \sum_{i\in I'_h\setminus I'_{h+1}} p_i.$$

Thus, setting $q'_f=\sum_{i\in I'_f} p_i$, $((I'_1,\dots,I'_f),(q_1,\dots,q_{f-1},q'_f)) \in Link(p_{j_1},\dots,p_{j_{k-m}})$ and, by Remark 4.13, there exists a maximal path σ_1 of $A_{K\setminus I}$ such that $cap_{K\setminus I}(\sigma_1)=(I'_1,\dots,I'_f)$ and $\pi(\sigma_1)=(q_1,\dots,q_{f-1},q'_f)$.

Subcase 2.1.- If $q_f\le q'_f$, let s be the vertex of σ_1 of height $\sum_{i=1}^f q_i$ and let σ'_1 be the subpath of σ_1 starting at the root of $A_{K\setminus I}$ and whose last vertex is s. Since, for all $h\in\{f+1,\dots,g\}$, $I_h\subseteq I$, $((I_{f+1},\dots,I_g),(q_{f+1},\dots,q_g)) \in Link(p_{i_1},\dots,p_{i_m})$ and, by Remark 4.13, there exists a maximal path σ_2 of $A_I(s)$ such that $cap_I(\sigma_2)=(I_{f+1},\dots,I_g)$ and $\pi(\sigma_2)=(q_{f+1},\dots,q_g)$. Obviously, $\sigma=\sigma'_1\sigma_2$ is a maximal path of B_I. Moreover, s is a leaf of $A_{K\setminus I}$ if and only if $q_f=q'_f$ and, in this case, $I_{f+1}\neq I$, since $q'_f=\sum_{i\in I'_f} p_i$ and $q_f<\sum_{i\in I_f\setminus I_{f+1}} p_i$. It follows that, if s is a leaf of $A_{K\setminus I}$, the first edge of σ_2 does not belong to the trunk of A_I. Then, by proof (ii) of Lemma 5.2, whether s is a leaf or not, the breaking vertices of σ are composed of the breaking vertices of σ_1, of those of σ_2, and of s. Thus, $\pi(\sigma)=(q_1,\dots,q_f,q_{f+1},\dots,q_g)$. Moreover, by definition of $bcap_I$, $bcap_I(\sigma)=(I'_1\cup I,\dots,I'_f\cup I,I_{f+1},\dots,I_g)=(I_1,\dots,I_g)$.

Subcase 2.2.- If $q'_f<q_f$, let s be the last vertex of σ_1. Since $q'_f=\sum_{i\in I_f\setminus I} p_i$ and $q_f<\sum_{i\in I_f\setminus I_{f+1}} p_i$, we have $I_{f+1}\neq I$ and $q_f-q'_f<\sum_{i\in I\setminus I_{f+1}} p_i$.

Thus $((I,I_{f+1},\dots,I_g),(q_f-q'_f,q_{f+1},\dots,q_g)) \in Link(p_{i_1},\dots,p_{i_m})$ and, by Remark 4.13, there exists a maximal path σ_2 of $A_I(s)$ such that

$$cap_I(\sigma_2)=(I,I_{f+1},\dots,I_g) \text{ and } \pi(\sigma_2)=(q_f-q'_f,q_{f+1},\dots,q_g).$$

Since the capacity of the first edge of σ_2 is equal to I, this edge belongs to the trunk of $A_I(s)$. s being a leaf of $A_{K\setminus I}$, it follows from proof of Lemma 5.2, that the breaking vertices of the maximal path $\sigma=\sigma_1\sigma_2$ of B_I are the breaking vertices of σ_1 and those of σ_2.

Thus $\pi(\sigma)=(q_1,\dots,q_{f-1},q'_f+(q_f-q'_f),q_{f+1},\dots,q_g)=(q_1,\dots,q_g)$. Moreover, by definition of $bcap_I$, $bcap_I(\sigma)=(I'_1\cup I,\dots,I'_f\cup I,I_{f+1},\dots,I_g)=(I_1,\dots,I_g)$.

Case 3.- If $I \subsetneq I_g$ then, for every $h \in \{1,\ldots,g\}$, $I \subsetneq I_h$. So let $I'_h = I_h \setminus I$. For every $h \neq g$,

$$0 < q_h < \sum_{i \in I_h \setminus I_{h+1}} p_i = \sum_{i \in I'_h \setminus I'_{h+1}} p_i \text{ and } q_g = \sum_{i \in I_g} p_i > \sum_{i \in I'_g} p_i.$$

Setting $q'_g = \sum_{i \in I'_g} p_i$, $((I'_1,\ldots,I'_g),(q_1,\ldots,q_{g-1},q'_g)) \in Link(p_{j_1},\ldots,p_{j_{k-m}})$. Thus there exists a maximal path σ_1 of $A_{K\setminus I}$ such that $cap_{K\setminus I}(\sigma_1) = (I'_1,\ldots,I'_g)$ and $\pi(\sigma_1) = (q_1,\ldots,q_{g-1},q'_g)$. Moreover, the last vertex s of σ_1 is a leaf of $A_{K\setminus I}$, and the trunk σ_2 of $A_I(s)$ is such that $cap_I(\sigma_2) = I$ and $\pi(\sigma_2) = (\sum_{i\in I} p_i) = (q_g - q'_g)$. It follows that the maximal path $\sigma = \sigma_1\sigma_2$ of B_I is such that $bcap_I(\sigma) = (I_1,\ldots,I_g)$ and $\pi(\sigma) = (q_1,\ldots,q_{g-1},q'_g + (q_g - q'_g)) = (q_1,\ldots,q_g)$. □

Lemma 5.6. There exists a one-to-one morphism φ_I from the directed tree B_I into the araucaria A, such that, for every edge (s,t) of B_I, $bcap_I(s,t) = cap(\varphi_I(s,t))$ and such that $\varphi_I(B_I)$ admits the same root as A, the same longest path as A, and every leaf of $\varphi_I(B_I)$ is a leaf of A.

Proof:
By Proposition 5.5, for every maximal path $\sigma = (s_0,\ldots,s_h)$ of B_I, $(bcap_I(\sigma),\pi(\sigma))$ is contained in $Link(p_1,\ldots,p_k)$ and, by Theorem 4.12, there exists a unique maximal path σ' of A such that $cap(\sigma') = bcap_I(\sigma)$ and $\pi(\sigma') = \pi(\sigma)$. Since $\pi(\sigma') = \pi(\sigma)$, σ and σ' have the same length and, if $\sigma' = (s'_0,\ldots,s'_h)$, there exists a one-to-one morphism φ from the subtree C_0 reduced to the unique path σ into A such that, for all $i \in \{0,\ldots,h\}$, $\varphi(s_i) = s'_i$ and, for all $i \in \{0,\ldots,h-1\}$, $cap(s'_i,s'_{i+1}) = bcap_I(s_i,s_{i+1})$. Moreover the image $s'_0 = \varphi(s_0)$ of the root of B_I is the root of A and the image $s'_h = \varphi(s_h)$ of the leaf s_h of B_I is a leaf of A. Since $bcap_I$ is discriminating we can extend φ to a one-to-one morphism φ_I from B_I into A by using the same method as in the proof (ii) of Theorem 4.12. Moreover, for every edge (s,t) of B_I, $bcap_I(s,t) = cap(\varphi_I(s,t))$ and, for every leaf f of B_I, $\varphi_I(f)$ is a leaf of A. Furthermore, the longest path τ of B_I is the product of the trunks of $A_{K\setminus I}$ and A_I. Hence, $|\tau| = \sum_{i\in K} p_i$ and $\varphi_I(\tau)$ is the trunk of A (see Fig. 4). □

Definition 5.7. Let B be the union of a family $(B_\lambda)_{\lambda\in\Lambda}$ of disjoint directed trees and, for every $\lambda \in \Lambda$, cap_λ a capacity function of B_λ. Two maximal paths σ of B_λ and σ' of B_μ are said to be equivalent relatively to $(cap_\lambda)_{\lambda\in\Lambda}$ if $\pi(\sigma) = \pi(\sigma')$ and $cap_\lambda(\sigma) = cap_\mu(\sigma')$.

Definition 5.8. Let $\sigma = (s_0,\ldots,s_h)$ and $\sigma' = (s'_0,\ldots,s'_h)$ be two paths of equal length in a graph G.

Merging σ and σ' consists in merging, for all i of $\{0,\ldots,h\}$, the vertices s_i and s'_i into a unique vertex, and, for all i of $\{0,\ldots,h-1\}$, the edges (s_i,s_{i+1}) and (s'_i,s'_{i+1}) into a unique edge.

Theorem 5.9. Let m and k be integers such that $1 \leq m < k$, $(p_1,\ldots,p_k)$ a sequence of positive integers, and $P_m(K)$ the set of subsets I of K of cardinality m.

The directed graph $A^{(m)}$ obtained by merging all equivalent maximal paths of $B^{(m)} = \bigcup_{I\in P_m(K)} B_I$ relatively to $(bcap_I)_{I\in P_m(K)}$ is isomorphic to the araucaria A of type $(p_1,\ldots,p_k)$.

Proof:
(i) Since the trees B_I are pairwise disjoint, every maximal path $\sigma = (s_0,s_1,\ldots,s_h)$ of $B^{(m)}$ belongs to a unique B_I. By the proof of Lemma 5.6, there exists a one-to-one morphism φ_I from the directed

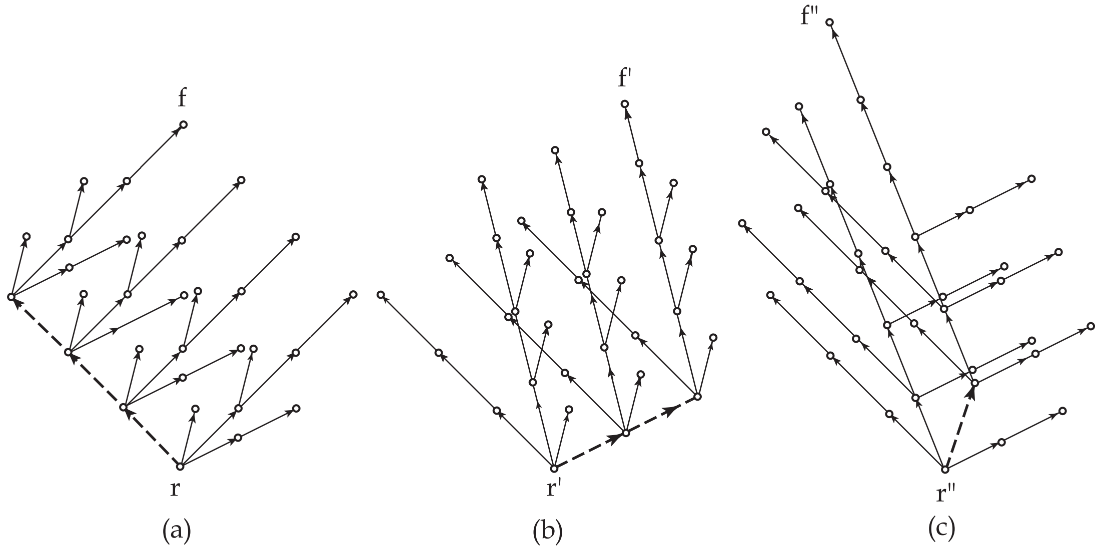


Figure 4. The araucaria $A(3,2,1)$ (see Fig. 2(b)) of type $(3,2,1)$ admits subtrees isomorphic to $B_{\{2,3\}} = \mathit{graft}(A_{\{2,3\}}/A_{\{1\}})$ (Fig. (a)), $B_{\{1,3\}} = \mathit{graft}(A_{\{1,3\}}/A_{\{2\}})$ (Fig. (b)), and $B_{\{1,2\}} = \mathit{graft}(A_{\{1,2\}}/A_{\{3\}})$ (Fig. (c)). Moreover $A(3,2,1)$ is the union of these subtrees.

tree $B_I = \mathit{graft}(A_I/A_{K\setminus I})$ into the araucaria A such that $\varphi_I(\sigma) = (\varphi_I(s_0), \ldots, \varphi_I(s_h))$ is a maximal path of A, $\mathit{cap}(\varphi_I(\sigma)) = \mathit{bcap}_I(\sigma)$, and $\pi(\varphi_I(\sigma)) = \pi(\sigma)$. The extension φ of all the mappings φ_I is a morphism from the graph $B^{(m)} = \bigcup_{I\in P_m(K)} B_I$ into A. Moreover, the extension bcap of all the capacity functions bcap_I is such that, for every edge (s,t) of $B^{(m)}$, $\mathit{bcap}(s,t) = \mathit{cap}(\varphi(s,t))$.

Since, by Theorem 4.12, every maximal path of A is characterized by its valued capacity chain, two maximal paths $\sigma = (s_0, s_1, \ldots, s_h)$ and $\sigma' = (s'_0, s'_1, \ldots, s'_h)$ of $B^{(m)}$ are equivalent if and only if $\varphi(\sigma) = \varphi(\sigma')$, that is, if and only if, for all $i \in \{0, \ldots, h\}$, $\varphi(s_i) = \varphi(s'_i)$.

(ii.1) For every vertex s of $B^{(m)}$, let $\bar{s}$ be the set of vertices of $B^{(m)}$ which are merged with s when all equivalent maximal paths of $B^{(m)}$ are merged and let $\varphi^{-1}(\varphi(s))$ be the set of vertices s' of $B^{(m)}$ such that $\varphi(s') = \varphi(s)$. It follows from (i) that $\bar{s} \subseteq \varphi^{-1}(\varphi(s))$.

We prove now that $\bar{s} = \varphi^{-1}(\varphi(s))$.

(ii.2) Let (s,t) be an edge of a tree B_I of $B^{(m)}$. From Proposition 5.5, the valued capacity chain of every maximal path of B_I containing (s,t) is of the form $((I_1, \ldots, I_g), (q_1, \ldots, q_g))$ with, for all $h \in \{1, \ldots, g\}$, either $I \subseteq I_h$ or $I_h \subseteq I$. From Definition 4.8 and from Proposition 5.5, one of these paths σ is such that $I_g = \mathit{bcap}(s,t)$.

(ii.3) Let (s',t') be an edge of $B^{(m)}$ distinct from (s,t) such that $\varphi(s',t') = \varphi(s,t)$ and let B_J be the tree of $B^{(m)}$ that contains (s',t'). Let σ_1 [resp. σ'_1] be the path from the root of B_I [resp. B_J] to t [resp. t']. By (i), $\varphi(\sigma_1) = \varphi(\sigma'_1)$, $\mathit{bcap}(\sigma_1) = \mathit{bcap}(\sigma'_1)$, and $\pi(\sigma_1) = \pi(\sigma'_1)$. As in (ii.2), B_J admits a maximal path σ' containing (s',t') whose valued capacity chain is $((I_1, \ldots, I_g), (q_1, \ldots, q_g))$. Hence, σ and σ' are equivalent, $s' \in \bar{s}$, and $t' \in \bar{t}$.

Hence, by (ii.1), $\bar{t} = \varphi^{-1}(\varphi(t))$ for every t of $B^{(m)}$.

(iii) By the isomorphism theorem for graphs, it follows that the mapping η from $B^{(m)}$ into $A^{(m)}$

such that $\eta(s) = \bar{s} = \varphi^{-1}(\varphi(s))$ for every s in $B^{(m)}$ is a surjective morphism and that there exists a one-to-one morphism $\bar{\varphi}$ from $A^{(m)}$ into the araucaria A such that, for every s of $B^{(m)}$, $\varphi(s) = \bar{\varphi}(\bar{s})$, that is, $\varphi = \bar{\varphi} \circ \eta$.

(iv) We prove now that φ is surjective.

Every vertex s of A belongs to a maximal path σ of A. From Theorem 4.12, the valued capacity chain $((I_1, \ldots, I_g), (q_1, \ldots, q_g))$ of σ belongs to $Link(p_1, \ldots, p_k)$. If $|I_g| \geq m$, there exists $I \in P_m(K)$ such that $I \subseteq I_g$ and then, for every $h \in \{1, \ldots, g\}$, $I \subseteq I_h$. If $|I_1| \leq m$, there exists $I \in P_m(K)$ such that $I_1 \subseteq I$ and then, for every $h \in \{1, \ldots, g\}$, $I_h \subseteq I$. Finally, if there exists $h \in \{1, \ldots, g-1\}$, such that $|I_{h+1}| \leq m \leq |I_h|$, there exists $I \in P_m(K)$ such that $I_{h+1} \subseteq I \subseteq I_h$. In this case, for all $j \in \{1, \ldots, h\}$, $I \subseteq I_j$ and, for all $j \in \{h+1, \ldots, g\}$, $I_j \subseteq I$. By Proposition 5.5, there exists in all three cases a maximal path σ_I in B_I with valued capacity chain $((I_1, \ldots, I_g), (q_1, \ldots, q_g))$. Hence, from (i), σ_I admits a vertex s_I such that $\varphi(s_I) = s$.

(v) Since φ is surjective, $\bar{\varphi}$ is also surjective. Hence $\bar{\varphi}$ is an isomorphism from the directed graph $A^{(m)}$ onto the araucaria A of type $(p_1, \ldots, p_k)$. □

Remark 5.10. The special case $m = 1$ of this Theorem is a result of [10], which was used for proving that every araucaria is included in the minimal automaton of the shuffle of some words. Moreover this is a new proof that our definition of araucarias is equivalent to the original one.

Remark 5.11. Theorem 5.9 can also be proved by showing that $A^{(m)}$ is a directed tree which admits a discriminating capacity function whose associated mapping is a bijection onto $Link(p_1, \ldots, p_k)$ and by then using Theorem 4.12.

Remark 5.12. In Theorem 5.9, it is not possible to replace $P_m(K)$ by one of its proper subsets $P'_m(K)$. Indeed, if $J \in P_m(K) \setminus P'_m(K)$, the union of the directed trees $(B_I)_{I \in P'_m(K)}$ contains no maximal path σ such that $cap(\sigma) = (J)$, by Proposition 5.5. However, by Theorem 4.12, the araucaria A contains such a maximal path.

Remark 5.13. An algorithm to build araucarias can be derived from Theorem 5.9. In Remark 8.11, we will give the size of the directed graph $B^{(m)}$ and the complexity of this algorithm.

6. The general grafting theorem

In this section, we extend the first grafting theorem to more than one grafting operation. As a special case of this general theorem, we obtain another method to construct araucarias by using a new family of trees generated by an iterative application of grafting to elementary trees reduced to paths.

Definition 6.1. Let $p_1, \ldots, p_k$ be $k \geq 1$ positive integers and $\mathcal{M} = (k_1, \ldots, k_m)$ a sequence of m $(1 \leq m \leq k)$ positive integers such that $k_1 + \ldots + k_m = k$.

Every sequence $\mathcal{P} = (K_1, \ldots, K_m)$ of non-empty disjoint subsets of K such that $K_1 \cup \ldots \cup K_m = K$ and, for all $i \in \{1, \ldots, m\}$, $|K_i| = k_i$ is called a partition of K with model $\mathcal{M}$. Let $\mathcal{P}_{\mathcal{M}}(K)$ be the set of these partitions.

For every partition $\mathcal{P} = (K_1, \ldots, K_m)$ of $\mathcal{P}_{\mathcal{M}}(K)$, let

$$C_{\mathcal{P}} = \mathit{graft}(A_{K_m} / \ldots / \mathit{graft}(A_{K_2} / A_{K_1}) \ldots)$$

and

$$cap_{\mathcal{P}} = graft(cap_{K_m}/\ldots/graft(cap_{K_2}/cap_{K_1})\ldots).$$

Theorem 6.2. The directed graph obtained by merging all maximal paths of $C^{(\mathcal{M})} = \bigcup_{\mathcal{P}\in\mathcal{P}_{\mathcal{M}}(K)} C_{\mathcal{P}}$ that are equivalent relatively to $(cap_{\mathcal{P}})_{\mathcal{P}\in\mathcal{P}_{\mathcal{M}}(K)}$ is the araucaria of type $(p_1,\ldots,p_k)$.

Proof:
(i) We prove this result by induction on the number m of grafting operations.

If $\mathcal{M} = (k)$, $\mathcal{P} = (K)$ is the unique partition of $\mathcal{P}_{\mathcal{M}}(K)$ and $C^{(\mathcal{M})} = C_{\mathcal{P}}$ is the araucaria $A_K = A$ with capacity function $cap_{\mathcal{P}} = cap_K = cap$.

If $m = 2$, the result is nothing else than Theorem 5.9.

Let now $\mathcal{M} = (k_1,\ldots,k_m)$ with $m > 2$ and let $\mathcal{M}' = (k_1,\ldots,k_{m-1})$. For any subset I of K of cardinality k_m, $\mathcal{P}_{\mathcal{M}'}(K\setminus I)$ is the set of partitions of $K\setminus I$ with model $\mathcal{M}'$. By induction, we can assume that the graph obtained by merging all equivalent maximal paths of $C^{(\mathcal{M}')} = \bigcup_{\mathcal{P}'\in\mathcal{P}_{\mathcal{M}'}(K\setminus I)} C_{\mathcal{P}'}$ relatively to $(cap_{\mathcal{P}'})_{\mathcal{P}'\in\mathcal{P}_{\mathcal{M}'}(K\setminus I)}$ is the araucaria $A_{K\setminus I}$ with capacity function $cap_{K\setminus I}$. Two vertices of $C^{(\mathcal{M}')}$ which are merged in a same vertex during this merging operation are said to be I-equivalent.

(ii) Let $\mathcal{P}_I$ be the set of partitions of $\mathcal{P}_{\mathcal{M}}(K)$ whose last term is I. For every $\mathcal{P} = (K_1,\ldots,K_{m-1},I)$ of $\mathcal{P}_I$, let $\mathcal{P}' = (K_1,\ldots,K_{m-1})$. In a first step, we realize the merging operation in the subgraph

$$C_I = \bigcup_{\mathcal{P}\in\mathcal{P}_I} C_{\mathcal{P}} = \bigcup_{\mathcal{P}'\in\mathcal{P}_{\mathcal{M}'}(K\setminus I)} graft(A_I/C_{\mathcal{P}'})$$

of $C^{(\mathcal{M})}$, for every subset I of K of cardinality k_m.

Let s_1 and s_2 be two I-equivalent vertices of $C^{(\mathcal{M}')}$, let $C_{\mathcal{P}'_1}$ [resp. $C_{\mathcal{P}'_2}$] be the subtree of $C^{(\mathcal{M}')}$ which contains the vertex s_1 [resp. s_2], and let σ'_1 [resp. σ'_2] be the path (possibly of length zero) from the root of $C_{\mathcal{P}'_1}$ [resp. $C_{\mathcal{P}'_2}$] to s_1 [resp. s_2]. Clearly, $cap_{\mathcal{P}'_1}(\sigma'_1) = cap_{\mathcal{P}'_2}(\sigma'_2)$ and $\pi(\sigma'_1) = \pi(\sigma'_2)$.

Then, for each maximal path σ''_1 of $A_I(s_1)$ and each maximal path σ''_2 of $A_I(s_2)$ such that $cap_I(\sigma''_1) = cap_I(\sigma''_2)$ and $\pi(\sigma''_1) = \pi(\sigma''_2)$, it follows from the proof of Lemma 5.2 that the maximal paths $\sigma_1 = \sigma'_1\sigma''_1$ and $\sigma_2 = \sigma'_2\sigma''_2$ of C_I are equivalent relatively to $(cap_{\mathcal{P}})_{\mathcal{P}\in\mathcal{P}_I}$. If we merge the paths σ_1 and σ_2 for all such pairs (σ''_1,σ''_2), σ'_1 and σ'_2 are merged relatively to $(cap_{\mathcal{P}'})_{\mathcal{P}'\in\mathcal{P}_{\mathcal{M}'}(K\setminus I)}$, s_1 and s_2 are merged in a unique vertex s, and the araucarias $A_I(s_1)$ and $A_I(s_2)$ are merged in an isomorphic araucaria $A_I(s)$.

Now, if we repeat the same operation for every pair (s_1,s_2) of I-equivalent vertices of $C^{(\mathcal{M}')}$, we merge simultaneously the equivalent maximal paths of C_I relatively to $(cap_{\mathcal{P}})_{\mathcal{P}\in\mathcal{P}_I}$ and the equivalent maximal paths of $C^{(\mathcal{M}')}$ relatively to $(cap_{\mathcal{P}'})_{\mathcal{P}'\in\mathcal{P}_{\mathcal{M}'}(K\setminus I)}$.

Hence, by the induction hypothesis, the directed graph obtained by merging all equivalent maximal paths of C_I relatively to $(cap_{\mathcal{P}})_{\mathcal{P}\in\mathcal{P}_I}$ is isomorphic to the directed tree $B_I = graft(A_I/A_{K\setminus I})$ with capacity function $bcap_I$.

(iii) To complete the merging operation for all equivalent maximal paths of $C^{(\mathcal{M})} = \bigcup_{I\in P_{k_m}(K)} C_I$ relatively to $(cap_{\mathcal{P}})_{\mathcal{P}\in\mathcal{P}_{\mathcal{M}}(K)}$, it remains to merge the maximal paths of $B^{(k_m)} = \bigcup_{I\in P_{k_m}(K)} B_I$ which are equivalent relatively to $(bcap_I)_{I\in P_{k_m}(K)}$. From Theorem 5.9, we obtain that way the araucaria of type $(p_1,\ldots,p_k)$. □

Remark 6.3. In general, $\mathcal{P}_{\mathcal{M}}(K)$ can not be replaced by one of its proper subsets. Indeed, given a model $\mathcal{M}$ and a partition $\mathcal{P} = (K_1,\ldots,K_m)$ of $\mathcal{P}_{\mathcal{M}}(K)$, it results from the definition of the capacity functions obtained by grafting, that $C_{\mathcal{P}}$ is the unique tree of $C^{(\mathcal{M})}$ that can contain a maximal path of capacity chain $(K_2 \cup ... \cup K_m, K_3 \cup \ldots \cup K_m, \ldots, K_m)$ (just use recursively the argument that a tree $C_{\mathcal{P}}$ of $C^{(\mathcal{M})}$ contains a maximal path whose last truncation has capacity K_m only if K_m is the last term of the sequence $\mathcal{P}$).

Now, from Definition 2.3, the root of the araucaria A is the root of a semi-araucaria $H_{K_2 \cup \ldots \cup K_m, 0}$ whose trunk τ has capacity $K_2 \cup \ldots \cup K_m$. From Definition 2.2, if $\sum_{i \in K_2} p_i > 1$, the vertex of τ of height 1 is the root of a semi-araucaria $H_{K_3 \cup \ldots \cup K_m, 1}$ whose trunk has capacity $K_3 \cup \ldots \cup K_m$. More generally if, for every j of $\{2, \ldots, m-1\}$, $\sum_{i \in K_j} p_i > 1$, A admits a path with capacity chain $(K_2 \cup \ldots \cup K_m, K_3 \cup \ldots \cup K_m, \ldots, K_m)$.

Since $\sum_{i \in K_j} p_i > 1$ when K_j is not reduced to a unique element $\{h\}$ with $p_h = 1$, it follows that $\mathcal{P}_{\mathcal{M}}(K)$ can not be replaced by one of its proper subsets if A is an araucaria of type $(p_1, \ldots, p_k)$ with, for all $h \in \{1, \ldots, k\}$, $p_h \neq 1$. In the converse case, not all elements of $\mathcal{P}_{\mathcal{M}}(K)$ are always necessary (see Remark 6.6).

In the special case $\mathcal{M} = (1, \ldots, 1)$, $\mathcal{P}_{\mathcal{M}}(K)$ is the set of permutations of K and Theorem 6.2 leads to a construction of araucarias by using a family of trees obtained by iteratively grafting elementary trees reduced to paths (see Fig. 5).

This result can be stated in the following way:

Theorem 6.4. Let $p_1, \ldots, p_k$ be positive integers, $A(p_1), \ldots, A(p_k)$ elementary araucarias of respective types $(p_1), \ldots, (p_k)$, $\Sigma(K)$ the set of permutations of $K = \{1, \ldots, k\}$, and, for all $\alpha \in \Sigma(K)$,

$$D_\alpha = \mathit{graft}(A(p_{\alpha(k)})/\ldots/\mathit{graft}(A(p_{\alpha(2)})/A(p_{\alpha(1)}))\ldots)$$

and

$$\mathit{cap}_\alpha = \mathit{graft}(\mathit{cap}_{\{\alpha(k)\}}/\ldots/\mathit{graft}(\mathit{cap}_{\{\alpha(2)\}}/\mathit{cap}_{\{\alpha(1)\}})\ldots).$$

The directed graph obtained by merging all equivalent maximal paths of $D = \bigcup_{\alpha \in \Sigma(K)} D_\alpha$ relatively to $(\mathit{cap}_\alpha)_{\alpha \in \Sigma(K)}$ is the araucaria of type $(p_1, \ldots, p_k)$.

By induction on the number of graftings, we can also prove that:

Proposition 6.5. There exists a maximal path σ of D_α such that $\mathit{cap}_\alpha(\sigma) = (I_1, \ldots, I_f)$ and $\pi(\sigma) = (q_1, \ldots, q_f)$ if and only if $((I_1, \ldots, I_f), (q_1, \ldots, q_f))$ belongs to $\mathit{Link}(p_1, \ldots, p_k)$ and $(I_1, \ldots, I_f)$ is a subsequence of $(\alpha(\{1, \ldots, k\}), \alpha(\{2, \ldots, k\}), \ldots, \{\alpha(k)\})$.

This result can be used for an alternative proof of Theorem 6.4.

Remark 6.6. It is not difficult to see that the araucaria of type $(p_1, p_2, p_3) = (3, 2, 1)$ can also be obtained by merging the equivalent maximal paths of $D = \bigcup_{\alpha \in \Sigma'(K)} D_\alpha$, where $\Sigma'(K)$ is the set $\Sigma(K) \setminus \{(1, 3, 2), (2, 3, 1)\}$ (see Fig. 5).

More generally, let $p_1, \ldots, p_k$ be a set of $k > 2$ positive integers, at least one of them being equal to 1. Let α be a permutation of K such that there exists $i \in \{2, \ldots, k-1\}$ for which $p_{\alpha(i)} = 1$. By definition of cap_α, for every maximal path σ of D_α, $\mathit{cap}_\alpha(\sigma)$ cannot contain $(\alpha(\{i, \ldots, k\}), \alpha(\{i+1, \ldots, k\}))$ as a subsequence. From Proposition 6.5, σ is then equivalent to a maximal path of a tree $D_{\alpha'}$ where α' is obtained from α by swapping $\alpha(i)$ either with $\alpha(i-1)$ or with $\alpha(i+1)$. It follows that the permutation α is not necessary to construct the araucaria of type $(p_1, \ldots, p_k)$ with the method of Theorem 6.4.

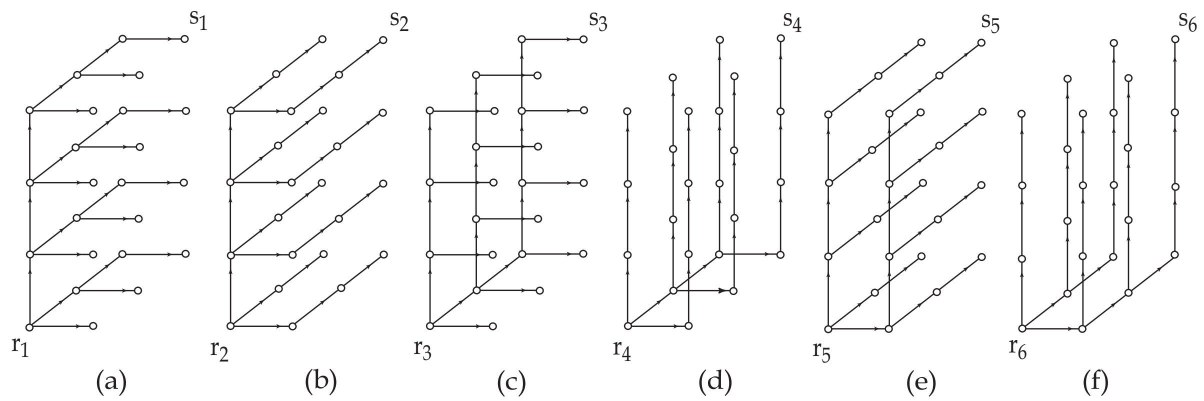


Figure 5. Given the three integers $p_1 = 3$, $p_2 = 2$, and $p_3 = 1$, figures (a), (b), (c), (d), (e), and (f) show the directed tree $D_\alpha = \mathit{graft}(A(p_{\alpha(3)})/\mathit{graft}(A(p_{\alpha(2)})/A(p_{\alpha(1)})))$ for the following respective values of α: $(1, 2, 3)$, $(1, 3, 2)$, $(2, 1, 3)$, $(2, 3, 1)$, $(3, 1, 2)$, $(3, 2, 1)$.

Remark 6.7. As in Remark 5.13, an algorithm which builds araucarias based on the construction of Theorem 6.4 can be given. The directed trees which are used here are much simpler than the trees B_I. In Remark 8.11, we will give the size of the directed graph D and the complexity of this algorithm.

7. The embedding theorem

We prove here that every directed tree can be embedded in an araucaria. Hence, despite the complexity of their definition, the family of subtrees of araucarias contains all the directed trees. This result is a simple corollary of the grafting theorems but it can also be proved by directly using Definition 2.3 (see Remark 7.3).

Theorem 7.1. Every directed tree can be embedded in an araucaria.

Proof:
If B is a subtree of an araucaria A_I, if C is a subtree of the araucaria $A_{K \setminus I}$ and, if P is a subset of the set of vertices of B, $\mathit{graft}_P(B/C)$ is a subtree of $B_I = \mathit{graft}(A_I/A_{K\setminus I})$, which is a subtree of an araucaria of type $(p_1, \ldots, p_k)$ by Lemma 5.6.

Now, every directed tree B is obtained by grafting successively its edges, which are araucarias of type (1). Hence, by induction, it follows that B is isomorphic to a subtree of some araucaria. □

Remark 7.2. Similarly, we can prove that every directed tree of size n can be embedded in the tree

$$\mathit{graft}(A(p_{n-1})/\ldots/\mathit{graft}(A(p_2)/A(p_1))\ldots)$$

where $p_1 = p_2 = \ldots = p_{n-1} = 1$.

Remark 7.3. Theorem 7.1 can also be proved directly by induction on the size of the directed tree B. As sketch of proof, it is sufficient to see that, if the root of B is removed then each resulting tree can be embedded in an araucaria by induction. Furthermore, if we add an edge (r, s) to an araucaria of root s, the resulting tree is embeddable in a semi-araucaria.

8. Some enumeration formulas for araucarias

To every vertex of an araucaria A of type $(p_1, \ldots, p_k)$, we associate a capacity which is a subset I of K. We prove that the number of vertices of capacity I is equal to $(k - |I|)! \prod_{i \in K \setminus I} p_i$. As a corollary we obtain the size of the set $Link(p_1, \ldots, p_k)$ introduced in Section 4. We prove also that the size of a semi-araucaria of type $(p_1, \ldots, p_k)$ is equal to $k! \prod_{i \in K} p_i + 1$. The size of an araucaria follows immediately from each of this two results, providing simpler proofs of a result of [10].

Definition 8.1. (i) For every leaf s of an araucaria A with capacity function cap, if p is the predecessor of s, the subset $I = cap(p, s)$ of K is called the capacity of s and is denoted $\chi(s)$.

Moreover, for every internal vertex s of A, $\chi(s) = \emptyset$ is called the capacity of s.

(ii) For every subset I of K, let $E_I(p_1, \ldots, p_k)$ be the set of vertices of capacity I of an araucaria A of type $(p_1, \ldots, p_k)$.

Lemma 8.2. For every non-empty proper subset I of K, if $m = |I|$ and $K \setminus I = \{j_1, \ldots, j_{k-m}\}$ then

$$|E_I(p_1, \ldots, p_k)| = |E_\emptyset(p_{j_1}, \ldots, p_{j_{k-m}})|.$$

Proof:
(i) Let δ be the mapping from the set $E_\emptyset(p_{j_1}, \ldots, p_{j_{k-m}})$ of internal vertices of the araucaria $A_{K \setminus I}$ into the set of leaves of the directed tree $B_I = graft(A_I / A_{K \setminus I})$ such that, for every vertex s of $E_\emptyset(p_{j_1}, \ldots, p_{j_{k-m}})$, $\delta(s)$ is the last vertex of the trunk $\tau_{(s)}$ of the araucaria $A_I(s)$. By proof of Lemma 5.2, $\tau_{(s)}$ is the terminal truncation of the maximal path σ of B_I whose last vertex is $\delta(s)$. By Lemma 5.6, $\varphi_I(\tau_{(s)})$ is the terminal truncation of the maximal path $\varphi_I(\sigma)$ of A and $\chi(\delta(s)) = cap(\varphi_I(\tau_{(s)})) = bcap_I(\tau_{(s)}) = I$. Hence $\varphi_I \circ \delta(s) \in E_I(p_1, \ldots, p_k)$ and $\varphi_I \circ \delta(E_\emptyset(p_{j_1}, \ldots, p_{j_{k-m}})) \subseteq E_I(p_1, \ldots, p_k)$.

(ii) Conversely, let t be a vertex of $E_I(p_1, \ldots, p_k)$ and $\sigma(t)$ the maximal path of A whose last vertex is t. Since, by proof of Theorem 5.9, the extension φ of all morphisms φ_I where $I \in P_m(K)$ is surjective, there exist $J \in P_m(K)$ and a maximal path σ of B_J such that $\varphi_J(\sigma) = \varphi(\sigma) = \sigma(t)$. The capacity of the terminal truncation $\tau_{(f)}$ of σ is then equal to $bcap_J(\tau_{(f)}) = cap(\varphi_J(\tau_{(f)})) = \chi(t) = I$. By Proposition 5.5, either $I \subseteq J$ or $J \subseteq I$ and, since $|I| = m = |J|$, $I = J$. Hence, by proof of Lemma 5.2, $\tau_{(f)}$ is the trunk of $A_I(s)$ for some internal vertex s of $A_{K \setminus I}$, $s \in E_\emptyset(p_{j_1}, \ldots, p_{j_{k-m}})$, and $\varphi_I \circ \delta(s) = t$. This proves that $E_I(p_1, \ldots, p_k) \subseteq \varphi_I \circ \delta(E_\emptyset(p_{j_1}, \ldots, p_{j_{k-m}}))$.

Thus, by (i), $\varphi_I \circ \delta(E_\emptyset(p_{j_1}, \ldots, p_{j_{k-m}})) = E_I(p_1, \ldots, p_k)$ and, since δ and φ_I are one-to-one,

$$|E_I(p_1, \ldots, p_k)| = |E_\emptyset(p_{j_1}, \ldots, p_{j_{k-m}})|.$$

□

Lemma 8.3. If $P(K)$ is the set of non-empty proper subsets of K,

$$|E_\emptyset(p_1, \ldots, p_k)| = 1 + \sum_{I \in P(K) \cup \{K\}} |E_I(p_1, \ldots, p_k)| \big(\sum_{i \in I} p_i - 1 \big).$$

Proof:
By Lemma 4.2, every edge of A is contained in exactly one terminal truncation. For every vertex t of the set $Leaves(A)$ of leaves of A, let $\tau(t) = (s_0, \ldots, s_{f-1}, t)$ be the terminal truncation which contains the leaf t and let $\tau^*(t)$ be the set of vertices of $\tau(t) \setminus \{s_0, t\}$. Then, if r is the root of A, $\{\tau^*(t); t \in Leaves(A)\} \cup \{r\}$ forms a partition of the set of internal vertices of A. Moreover $I = \chi(t)$ is a non-empty subset of K and $\tau^*(t)$ contains $\sum_{i\in I} p_i - 1$ internal vertices of A. It follows that

$$|E_\emptyset(p_1, \ldots, p_k)| = 1 + \sum_{I \in P(K)\cup\{K\}} |E_I(p_1, \ldots, p_k)|\big(\sum_{i\in I} p_i - 1\big).$$

□

Definition 8.4. (i) Let $\Psi_0(X_1, \ldots, X_k) = 1$ and, for each $m \in \{1, \ldots, k\}$, let

$$\Psi_m(X_1, \ldots, X_k) = \sum_{I\in P_m(K)} \prod_{i\in I} X_i$$

be the elementary symmetric polynomial of degree m on variables $X_1, \ldots, X_k$, where $P_m(K)$ is the set of subsets I of K of cardinality m.

For example, $\Psi_1(X_1, X_2, X_3) = X_1 + X_2 + X_3$, $\Psi_2(X_1, X_2, X_3) = X_1X_2 + X_2X_3 + X_3X_1$, and $\Psi_3(X_1, X_2, X_3) = X_1X_2X_3$.

(ii) The polynomial

$$\Upsilon_k(X_1, \ldots, X_k) = \sum_{m=0}^{k} m!\Psi_m(X_1, \ldots, X_k)$$

is called the araucaria polynomial in k variables.

The first araucaria polynomials are:

$\Upsilon_1(X_1) = X_1 + 1,$
$\Upsilon_2(X_1, X_2) = 2X_1X_2 + X_1 + X_2 + 1,$
$\Upsilon_3(X_1, X_2, X_3) = 6X_1X_2X_3 + 2(X_1X_2 + X_2X_3 + X_3X_1) + X_1 + X_2 + X_3 + 1.$

Lemma 8.5. For each $m \in \{1, \ldots, k-1\}$,

$$\sum_{I\in P_m(K)} \big(\prod_{i\in I} X_i\big)\big(\sum_{j\in K\setminus I} X_j\big) = (m+1)\Psi_{m+1}(X_1, \ldots, X_k).$$

Proof:
The sum

$$S = \sum_{I\in P_m(K)} \big(\prod_{i\in I} X_i\big)\big(\sum_{j\in K\setminus I} X_j\big)$$

is a symmetric function of $X_1, \ldots, X_k$. The product $\big(\prod_{i\in I} X_i\big)X_j$ is equal to $X_1 \ldots X_{m+1}$ if and only if $j \in \{1, \ldots, m+1\}$ and $I = \{1, \ldots, m+1\} \setminus \{j\}$. Therefore $X_1 \ldots X_{m+1}$ appears $m+1$ times in the sum S. By symmetry, the same thing happens for the other products and this proves the relation. □

Theorem 8.6. For every proper subset I of K, if $m = |I|$,

$$|E_I(p_1, \ldots, p_k)| = (k-m)! \prod_{i \in K \setminus I} p_i$$

and $|E_K(p_1, \ldots, p_k)| = 1$.

Proof:
For $k = 1$, every araucaria of type p_1 admits one leaf and p_1 internal vertices. Thus $|E_{\{1\}}(p_1)| = 1$ and $|E_\emptyset(p_1)| = p_1$.

Assume that the property is satisfied for every arity smaller than k and let A be an araucaria of type $(p_1, \ldots, p_k)$.

(i) If $I \neq K$, $I \neq \emptyset$, and $K \setminus I = \{j_1, \ldots, j_{k-m}\}$, by Lemma 8.2 and by the induction hypothesis,

$$|E_I(p_1, \ldots, p_k)| = |E_\emptyset(p_{j_1}, \ldots, p_{j_{k-m}})| = (k-m)! \prod_{j \in K \setminus I} p_j.$$

(ii) $|E_K(p_1, \ldots, p_k)| = 1$ since the trunk of A is the only terminal truncation of capacity K.

(iii) By (i), (ii), and Lemma 8.3,

$$|E_\emptyset(p_1, \ldots, p_k)| = \sum_{i \in K} p_i + \sum_{m=1}^{k-1} \sum_{I \in P_m(K)} (k-m)! \prod_{j \in K \setminus I} p_j \big(\sum_{i \in I} p_i - 1 \big).$$

By Lemma 8.5,

$$\sum_{I \in P_m(K)} \big(\prod_{j \in K \setminus I} p_j \big) \big(\sum_{i \in I} p_i \big) = (k-m+1) \Psi_{k-m+1}(p_1, \ldots, p_k).$$

Hence $|E_\emptyset(p_1, \ldots, p_k)|$ is equal to

$$\sum_{i \in K} p_i + \sum_{m=1}^{k-1} \big((k-m+1)! \Psi_{k-m+1}(p_1, \ldots, p_k) - (k-m)! \Psi_{k-m}(p_1, \ldots, p_k) \big)$$

and

$$|E_\emptyset(p_1, \ldots, p_k)| = k! \Psi_k(p_1, \ldots, p_k).$$

□

Remark 8.7. By setting $\prod_{j \in \emptyset} p_j = 1$, the case $I = K$ in Theorem 8.6 can be included in the general case.

Corollary 8.8. The size of the set $Link(p_1, \ldots, p_k)$ is equal to

$$\Upsilon_k(p_1, \ldots, p_k) - k! \prod_{i \in K} p_i.$$

Proof:
For every leaf t of the araucaria A of type $(p_1, \dots, p_k)$, there exists a unique maximal path of A from the root of A to t. Hence there exists a bijection from the set $Leaves(A)$ of leaves of A onto the set $MP(A)$ of maximal paths of A. By Theorem 4.12, there exists also a bijection from $MP(A)$ onto the set $Link(p_1, \dots, p_k)$. It follows that

$$|Link(p_1, \dots, p_k)| = |MP(A)| = |Leaves(A)| = \sum_{I \in P(K) \cup \{K\}} |E_I(p_1, \dots, p_k)|$$

and, by Theorem 8.6, that

$$|Link(p_1, \dots, p_k)| = \sum_{m=1}^{k-1} (k-m)! \Psi_{k-m}(p_1, \dots, p_k) = \Upsilon_k(p_1, \dots, p_k) - k! \prod_{i \in K} p_i.$$

□

Theorem 8.9. (i) The size of a semi-araucaria of type $(p_1, \dots, p_k)$ is equal to $k! \prod_{i \in K} p_i + 1$.
(ii) The size of an araucaria of type $(p_1, \dots, p_k)$ is equal to $\Upsilon_k(p_1, \dots, p_k)$.

Proof:
(i) By Definition 2.2, the size of a semi-araucaria H of type $(p_1, \dots, p_k)$ is equal to

$$|H| = 1 + \sum_{i \in K} p_i + \sum_{I \in P(K)} \big(|H_{I,h}| - 1\big)\big(\sum_{j \in K \setminus I} p_j - 1\big).$$

If we assume, by induction, that, for all $I \in P(K)$, $|H_{I,h}| = m! \prod_{i \in I} p_i + 1$, it follows that

$$|H| = 1 + \sum_{i \in K} p_i + \sum_{m=1}^{k-1} m! \Big(\sum_{I \in P_m(K)} \big(\prod_{i \in I} p_i\big)\big(\sum_{j \in K \setminus J} p_j - 1\big)\Big).$$

Hence, by Lemma 8.5,

$$|H| = 1 + \sum_{i \in K} p_i + \sum_{m=1}^{k-1} \Big((m+1)! \Psi_{m+1}(p_1, \dots, p_k) - m! \Psi_m(p_1, \dots, p_k)\Big) = k! \prod_{i \in K} p_i + 1.$$

(ii) We give here two simple proofs of the result of [10].
Proof 1.- By Definition 2.3, the size of the araucaria A of type $(p_1, \dots, p_k)$ is equal to

$$|A| = 1 + \sum_{I \in P(K) \cup \{K\}} \big(|H_{I,0}| - 1\big) = \sum_{m=0}^{k} m! \Psi_m(p_1, \dots, p_k) = \Upsilon_k(p_1, \dots, p_k).$$

Proof 2.- By Theorem 8.6, the size of A is equal to

$$|A| = \sum_{I \in P(K) \cup \{K, \emptyset\}} |E_I(p_1, \dots, p_k)| = \sum_{m=0}^{k} m! \Psi_m(p_1, \dots, p_k).$$

□

Remark 8.10. By Theorems 8.6 and 8.9, there exists a bijection from the set $E_\emptyset(p_1,\ldots,p_k)$ of internal vertices of an araucaria A of type $(p_1,\ldots,p_k)$ onto the set $H_{K,0}\setminus\{r\}$ where $H_{K,0}$ is the subsemi-araucaria of A of same type and r their common root. (We use here the same notation for a tree and for the set of its vertices.) This can also be proved directly. Indeed, if we assume the property for every arity smaller than k, for every non-empty proper subset $I=\{i_1,\ldots,i_m\}$ of K, there exist a bijection α_I from $E_\emptyset(p_{i_1},\ldots,p_{i_m})$ onto $H_{I,0}\setminus\{r\}$ and, by Lemma 8.2, a bijection ζ_I from $E_{K\setminus I}(p_1,\ldots,p_k)$ onto $E_\emptyset(p_{i_1},\ldots,p_{i_m})$. If s is the last vertex of the trunk of A, $E_K(p_1,\ldots,p_k)=\{s\}$ by Theorem 8.6. Hence there exists a bijection from

$$A\setminus E_\emptyset(p_1,\ldots,p_k)=\Big(\bigcup_{I\in P(K)} E_{K\setminus I}(p_1,\ldots,p_k)\Big)\cup\{s\}$$

onto

$$A\setminus(H_{K,0}\setminus\{r\})=\Big(\bigcup_{I\in P(K)}\big(H_{I,0}\setminus\{r\}\big)\Big)\cup\{r\}$$

which extends all the $\alpha_I\circ\zeta_I$. Thus there exists also a bijection from $E_\emptyset(p_1,\ldots,p_k)$ onto $H_{K,0}\setminus\{r\}$.

Remark 8.11. (i) For every non empty proper subset $I=\{i_1,\ldots,i_m\}$ of K, the size of the tree $B_I=\mathit{graft}(A_I/A_{K\setminus I})$ is equal to $|A_I|\cdot|A_{K\setminus I}|$. Hence, by Theorem 8.9, setting $\{j_1,\ldots,j_{k-m}\}=K\setminus I$, the size of $B^{(m)}=\bigcup_{I\in P_m(K)}B_I$ is equal to

$$\Gamma_k(p_1,\ldots,p_k)=\sum_{I\in P_m(K)}\Upsilon_m(p_{i_1},\ldots,p_{i_m})\Upsilon_{k-m}(p_{j_1},\ldots,p_{j_{k-m}}).$$

Γ_k is then a polynomial of same degree as Υ_k and, since $|P_m(K)|=\binom{k}{m}$, it has same term of highest degree. Hence, the araucaria A and the graph $B^{(m)}$ have asymptotically the same size.

(ii) Theorem 5.9 can be used to construct the araucaria A by adding and merging the trees B_I of $B^{(m)}$ one by one. This can be achieved in time $O(|B^{(m)}|)=O(|A|)$ as long as k is considered as a constant (both the size of the capacity of an edge and the number of neighbors of a vertex in A can be bounded by a value only depending on k). However, the algorithm given in Section 3 is simpler, runs faster, and remains linear even if k is not a constant.

(iii) The same remarks hold if we want to construct A using Theorem 6.4. Indeed, the size of a tree D_α is equal to $\prod_{i\in K}(p_i+1)$ and the size of $D=\bigcup_{\alpha\in\Sigma(K)}D_\alpha$ is equal to

$$\Delta_k(p_1,\ldots,p_k)=k!\prod_{i\in K}(p_i+1).$$

Hence, the polynomials Δ_k and Υ_k have same degree and same term of highest degree.

9. Conclusion

In this paper, we have given a new definition of the araucarias introduced in [10, 11] and, based on this definition, an optimal algorithm for their construction as well as a new method for calculating their size. We have also introduced a notion of capacity which leads to a characterization of the maximal paths of an araucaria. The araucarias are then characterized by these properties. The grafting operation introduced here has interesting properties in relation with capacities. In particular, we get various methods for generating araucarias of arity k by using either araucarias of arity less than k or a family of simpler trees. We also establish bijections from some sets of vertices of an araucaria of arity k onto sets of vertices of araucarias of arity smaller than k. This leads to two new proofs that the size of an araucaria of type $(p_1, \ldots, p_k)$ is a symmetric polynomial in $p_1, \ldots, p_k$. We prove also that every directed tree can be embedded in an araucaria.

These results will be used in the study of the minimal automaton of the shuffle product of a finite set of words. In particular, we hope to prove that, if the alphabets of the words $u_1, \ldots, u_k$ are pairwise disjoint up to a common letter, then the minimal automaton of $u_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup\, u_k$ can be described by using only araucarias or homomorphic images of araucarias, and that the maximum size of the minimal automaton is a polynomial function of $|u_1|, \ldots, |u_k|$ whose coefficients are exponential in k. Some properties of the minimal automaton are also expected in the general case. Moreover, we hope to be able to give an optimal algorithm to construct this automaton by using a method given in [10].

References

[1] Berstel, J., Boasson, L.: Shuffle factorization is unique, *Theoretical Computer Science*, **273**(1-2), 2002, 47–67.

[2] Biegler, F., Daley, M., McQuillan, I.: On the shuffle automaton size for words, *Journal of Automata, Languages and Combinatorics*, **15**(1/2), 2010, 53–70.

[3] Gómez, A. C., Pin, J.-E.: On a conjecture of Schnoebelen, in: *Proceedings of the 7th international conference on Developments in Language Theory*, vol. 2710 of *LNCS*, Springer, 2003, 35–54.

[4] Lothaire, M.: *Combinatorics on words*, Encyclopedia of Mathematics, Vol.17, Addison-Wesley, 1983.

[5] Lothaire, M.: *Algebraic combinatorics on words*, Cambridge University Press, 2002.

[6] Nivat, M., Ramkumar, G. D. S., Rangan, C. P., Saoudi, A., Sundaram, R.: Efficient parallel shuffle recognition, *Parallel Processing Letters*, **4**, 1994, 455–463.

[7] Schmitt, D., Spehner, J.-C.: Construction of araucaria trees, in: *Proceedings of the Kyoto international Conference of Computational Geometry and Graph Theory*, 2007, 128–129.

[8] Schnoebelen, P.: Decomposable regular languages and the shuffle operator, *EATCS Bulletin*, **67**, 1999, 283–289.

[9] Schott, R., Spehner, J.-C.: On the minimal automaton of the shuffle of words and araucarias, in: *Proceedings of the 4th international conference on Machines, Computations, and Universality 2004*, vol. 3354 of *LNCS*, Springer, 2005, 316–327.

[10] Schott, R., Spehner, J.-C.: Shuffle of words and araucaria trees, *Fundamenta Informaticae*, **74**(4), 2006, 579–601.

[11] Schott, R., Spehner, J.-C.: Erratum for "Shuffle of words and araucaria trees", *Fundamenta Informaticae*, **81**(4), 2007, 485–490.

[12] Spehner, J.-C.: Le calcul rapide des mélanges de deux mots, *Theoretical Computer Science*, **47**(2), 1986, 181–203.