Leaf-induced subtrees of leaf-Fibonacci trees

Audace A. V. Dossou-Olory*

Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa
Institut de Mathématiques et de Sciences Physiques (01 BP 603 Porto-Novo), Université d'Abomey-Calavi, Bénin

(Received: 15 November 2018. Received in revised form: 27 December 2018. Accepted: 28 December 2018. Published online: 3 January 2019.)

© 2019 the author. This is an open access article under the CC BY (International 4.0) license (https://creativecommons.org/licenses/by/4.0/).

Abstract

In analogy to a concept of Fibonacci trees, we define the leaf-Fibonacci tree of size $n$ and investigate its number of nonisomorphic leaf-induced subtrees. Denote by $f_0$ the one vertex tree and by $f_1$ the tree that consists of a root with two leaves attached to it; the leaf-Fibonacci tree $f_n$ of size $n \geq 2$ is the binary tree whose branches are $f_{n-1}$ and $f_{n-2}$. We derive a nonlinear difference equation for the number $N(f_n)$ of nonisomorphic leaf-induced subtrees (subtrees induced by leaves) of $f_n$, and also prove that $N(f_n)$ is asymptotic to $K_1 \cdot K_2 \phi^n$ as $n$ tends to infinity, where $\phi$ is the golden ratio and $K_1, K_2$ are explicitly calculated constants.

Keywords: leaf-induced subtrees; nonisomorphic subtrees; leaf-Fibonacci trees.

2010 Mathematics Subject Classification: 05C05, 05C30.

1. Introduction

Fibonacci trees provide an alternative approach to a binary search in computer science and information processing [12, p. 417]. The Fibonacci tree of order $n$ is defined as the binary tree whose left branch is the Fibonacci tree of order $n-1$ and right branch is the Fibonacci tree of order $n-2$, while the Fibonacci tree of order 0 or 1 is the tree with only one vertex [12]. We show in Figure 1 the Fibonacci tree of order 5.

[Diagram of the Fibonacci tree of order 5]

Thus, the Fibonacci tree of order $n$ has precisely $F_{n+1}$ leaves (so $2F_{n+1} - 1$ vertices), where $F_n$ denotes the $n$-th Fibonacci number:

$$F_0 = 0, \quad F_1 = 1, \quad F_n = F_{n-1} + F_{n-2} \text{ for } n > 1.$$ 

Fibonacci trees form a subclass of the so-called AVL (“Adel’son-Vel’skii and Landis”—named after the inventors) trees [1]; these trees have the defining property that for every internal vertex $v$, the heights (i.e., the greatest distance to a leaf from the root) of the left and right branches of the subtree rooted at $v$ (consisting of $v$ and all its descendants) differ by at most one. Figure 2 shows an AVL tree of height 3. For more information on Fibonacci trees and their uses, we refer to [9–11,15].

In analogy to the concept of Fibonacci trees, we define the leaf-Fibonacci tree of size $n$ as follows:

- Denote by $f_0$ the tree with only one vertex and by $f_1$ the tree that consists of a root with two leaves attached to it;
- For $n \geq 2$, connect the roots of the trees $f_{n-1}$ and $f_{n-2}$ to a new common vertex to obtain the tree $f_n$.

*E-mail address: audace@aims.ac.za
In other words, the leaf-Fibonacci tree $f_n$ of size $n \geq 2$ is the binary tree whose branches are the leaf-Fibonacci trees $f_{n-1}$ and $f_{n-2}$. Hence, $f_n$ has precisely $F_{n+2}$ leaves, where $F_n$ is the $n$-th Fibonacci number ($F_1 = 1$, $F_2 = 1$, $F_3 = 2$, $F_4 = 3$, $F_5 = 5$, $F_6 = 8$, $F_7 = 13$, ...). Figure 3 shows the leaf-Fibonacci tree of size 5.

Let $T$ be a rooted tree without vertices of outdegree 1 (also known as topological or series-reduced or homeomorphically irreducible trees [2,3,6,14]). Every choice of $k$ leaves of $T$ induces another topological tree, which is obtained by extracting the minimal subtree of $T$ that contains all the $k$ leaves and contracting (if any) all vertices of outdegree 1; see Figure 4 for an illustration.

Every subtree obtained through this operation is sometimes referred to as a leaf-induced subtree [4–6]. We note that the study of leaf-induced subtrees of binary trees finds a noteworthy relevance in phylogenetics—see Semple and Steel’s book [13] which describes the mathematical foundations of phylogenetics.

In this note, we shall be interested in the number of nonisomorphic leaf-induced subtrees of a leaf-Fibonacci tree of size $n$. Two rooted trees are said to be isomorphic if there is a graph isomorphism (preserving adjacency) between them that maps the root of one to the root of the other. It is important to note that the problem of enumerating leaf-induced subtrees becomes trivial if isomorphisms are not taken into account: in fact, it is clear that every topological tree with $n$ leaves has exactly $2^n - 1$ leaf-induced subtrees.

We mention that nonisomorphic leaf-induced subtrees of a topological tree have been studied only very recently: Wagner and Dossou-Olory [7] obtained exact and asymptotic enumeration results on the number of nonisomorphic
leaf-induced subtrees of two classes of \(d\)-ary trees, namely the so-called \(d\)-ary caterpillars and even \(d\)-ary trees. In [7], the authors also derived extremal results for the number of root containing leaf-induced subtrees of a topological tree.

We shall denote the number of nonisomorphic leaf-induced subtrees of the leaf-Fibonacci tree \(f_n\) by \(N(f_n)\). Our main results are a recurrence relation and an asymptotic formula for \(N(f_n)\). As it turns out, the plan to compute \(N(f_n)\) will be to consider the number of root containing leaf-induced subtrees of \(f_n\).

In [16], Wagner studied the number of independent vertex subsets (sets of vertices containing no pair of adjacent vertices) of a Fibonacci tree of order \(n\) with the notable difference that in his context, the Fibonacci tree of order 0 has no vertices. Wagner derived a system of recurrence relations for the number of independent vertex subsets of a Fibonacci tree of an arbitrary order \(n\), and also proved that there are positive constants \(A, B > 0\) such that the number of independent vertex subsets of a Fibonacci tree of order \(n\) is asymptotic to \(A \cdot B^n\) as \(n\) tends to infinity. In the present study, we obtain a similar asymptotic formula for the number \(N(f_n)\) of nonisomorphic leaf-induced subtrees of the leaf-Fibonacci tree of size \(n\): we demonstrate that for some effectively computable constants \(A_1, A_2 > 0\),

\[
N(f_n) \sim A_1 \cdot A_2^n \quad \text{as} \quad n \to \infty.
\]

2. Main results

We note from the recursive definition of the tree \(f_n\) that \(f_m\) is a leaf-induced subtree of \(f_n\) for every \(m \leq n\). However, not every leaf-induced subtree of \(f_n\) is again a leaf-Fibonacci tree: in fact, by repeatedly removing leaves from \(f_n\), one easily sees that \(f_n\) has leaf-induced subtrees of every number of leaves \(k\) between 1 and \(n\).

As was mentioned in the introduction, the plan to compute \(N(f_n)\) will be to consider the number of root containing leaf-induced subtrees of \(f_n\). In Lemma 2.1, we show that among all isomorphic leaf-induced subtrees with two or more leaves of \(f_n\), there is always at least one of them that contains the root of \(f_n\).

**Lemma 2.1.** All nonisomorphic leaf-induced subtrees with two or more leaves of the leaf-Fibonacci tree \(f_n\) can be identified as containing the root of \(f_n\).

**Proof.** The tree \(f_0\) has only one vertex which is also its leaf and root, so the statement holds vacuously for \(n = 0\). The statement is trivial for \(n = 1\) (\(f_1\) is the only leaf-induced subtree in this case). Let \(n > 1\) and consider a subset of \(k > 1\) leaves of \(f_n\). We argue by double induction on \(n\) and \(k\):

- If all \(k\) leaves belong to \(f_{n-1}\) then by the induction hypothesis on \(n\), the induced subtree with \(k\) leaves contains the root of \(f_{n-1}\). Moreover, by the induction hypothesis on \(k\), the tree \(f_{n-1}\) can be identified as containing the root of \(f_n\) (as \(f_{n-1}\) is clearly a leaf-induced subtree of \(f_n\)). Hence, the induced subtree with \(k\) leaves can be identified as containing the root of \(f_n\).

- If all \(k\) leaves belong to \(f_{n-2}\), then we also deduce by the induction hypothesis that the induced subtree with \(k\) leaves is a root containing leaf-induced subtree of \(f_n\).

- If \(k_1\) leaves belong to \(f_{n-1}\) and \(k - k_1 \) leaves belong to \(f_{n-2}\), then by the induction hypothesis, the induced subtrees with \(k_1\) and \(k - k_1\) leaves are root containing leaf-induced subtrees of \(f_{n-1}\) and \(f_{n-2}\), respectively. Consequently, the root of the induced subtree with \(k\) leaves coincides with the root of \(f_n\).

This completes the induction step as well as the proof of the lemma. \(\square\)

We then obtain the following proposition.

**Proposition 2.1.** The number \(N(f_n)\) of nonisomorphic leaf-induced subtrees of the leaf-Fibonacci tree \(f_n\) satisfies the following nonlinear recurrence relation:

\[
N(f_n) = 1 + \frac{1}{2}N(f_{n-2}) - \frac{1}{2}N(f_{n-2})^2 + N(f_{n-2}) \cdot N(f_{n-1})
\]  \(\text{(1)}\)

with initial values \(N(f_0) = 1, N(f_1) = 2\).

**Proof.** It is obvious that \(N(f_0) = 1\) and \(N(f_1) = 2\). Let \(n > 1\). By Lemma 2.1, \(N(f_n)\) is precisely one more the number of nonisomorphic root containing leaf-induced subtrees of \(f_n\) (the subtree with only one vertex has been included). Since all leaf-induced subtrees of the leaf-Fibonacci tree \(f_{n-2}\) are again leaf-induced subtrees of \(f_{n-1}\), the nonisomorphic root containing leaf-induced subtrees of \(f_n\) can be categorised by two types of enumeration:
• Both branches of the induced subtree are leaf-induced subtrees of \( f_{n-2} \). The total number of these possibilities is \( \frac{1 + N(f_{n-2})}{2} \) as the induced subtrees have to be nonisomorphic.

• One of the branches of the induced subtree is a leaf-induced subtree of \( f_{n-2} \) while the other branch is a leaf-induced subtree of \( f_{n-1} \) but does not belong to the set of leaf-induced subtrees of \( f_{n-2} \). The total number of these possibilities is \( N(f_{n-2}) \cdot N(f_{n-1} - N(f_{n-2})) \).

Therefore, we obtain

\[
N(f_n) = 1 + \left( \frac{1 + N(f_{n-2})}{2} \right) + N(f_{n-2})(N(f_{n-1}) - N(f_{n-2}))
\]

\[
= 1 + \frac{1}{2}N(f_{n-2}) - \frac{1}{2}N(f_{n-2})^2 + N(f_{n-2}) \cdot N(f_{n-1}),
\]

which completes the proof of the proposition.

The sequence \( (N(f_n))_{n \geq 0} \) starts as

\[
N(f_0) = 1, N(f_1) = 2, N(f_2) = 3, N(f_3) = 6, N(f_4) = 16, N(f_5) = 82, N(f_6) = 1193,
N(f_7) = 94506, N(f_8) = 112034631, \ldots
\]

It seems that recursion (1) cannot be solved explicitly: in fact, we fail to find a transformation strategy that can reduce (1) to a linear recursion. Therefore, finding an asymptotic formula should be in order. In the following theorem, we show that \( N(f_n) \) grows doubly exponentially in \( n \).

**Theorem 2.1.** There are two positive constants \( K_1, K_2 > 0 \) (both solely depending on the first terms of \( (N(f_n))_{n \geq 0} \)) such that

\[
N(f_n) = (1 + o(1))K_1 \cdot K_2 \left( \frac{\lambda}{e} \right)^n
\]

as \( n \to \infty \).

**Proof.** For ease of notation, set \( A_n := N(f_n) \). Then we have

\[
A_n = 1 + \frac{1}{2}A_{n-2} - \frac{1}{2}A_{n-2}^2 + A_{n-2} \cdot A_{n-1}
\]

with initial values \( A_0 = 1, A_1 = 2 \). Since the sequence \( (A_n)_{n \geq 0} \) increases with \( n \), it is not difficult to note that

\[
A_n \geq \frac{1}{2}A_{n-1} \cdot A_{n-2}
\]

for all \( n \geq 2 \). Also, since \( A_n \geq A_2 = 3 \) for all \( n \geq 2 \) and \( 1 + A_1/2 - A_1^2/2 = 0 \), it is not difficult to see that

\[
A_n \leq A_{n-1} \cdot A_{n-2}
\]

for all \( n \geq 3 \). Thus, we have

\[
\lim_{n \to \infty} \frac{A_{n-1}}{A_n} = 0,
\]

which also implies that the sequence \( (A_{n-1}/A_n)_{n \geq 1} \) is bounded for every \( n \geq 1 \). We may then start by proceeding as in [8, Section 2.2.3]. Let us use \( Q_n \) as a shorthand for \( \log(A_n) \) and \( E_n \) as a shorthand for

\[
\log \left( 1 + \frac{1}{2A_{n-1}} - \frac{A_{n-2}}{2A_{n-1}} + \frac{1}{A_{n-1} \cdot A_{n-2}} \right).
\]

With these notations, we have

\[
Q_n = Q_{n-1} + Q_{n-2} + E_n.
\]

By setting \( R_{n-1} := Q_{n-2} \), we obtain the following system (written in matrix form) of two linear difference equations:

\[
\begin{pmatrix}
Q_n \\
R_n
\end{pmatrix}
= \begin{pmatrix}
1 & 1 \\
1 & 0
\end{pmatrix}
\begin{pmatrix}
Q_{n-1} \\
R_{n-1}
\end{pmatrix}
+ \begin{pmatrix}
E_n \\
0
\end{pmatrix}.
\]
By iteration on $n$, one gets

$$
\begin{pmatrix}
Q_n \\
R_n
\end{pmatrix} = \left( \begin{array}{cc}
1 & 1 \\
1 & 0
\end{array} \right)^{n-1} \begin{pmatrix}
Q_1 \\
Q_0
\end{pmatrix} + \sum_{i=2}^{n} \left( \begin{array}{cc}
1 & 1 \\
1 & 0
\end{array} \right)^{n-1} \begin{pmatrix}
E_i \\
0
\end{pmatrix}
\right)
$$

for all $n \geq 2$, as $R_1 = Q_0$. The eigenvalue decomposition gives us

$$
\begin{pmatrix}
1 & 1 \\
1 & 0
\end{pmatrix} = \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix}
\lambda_1 & \lambda_2 \\
1 & 1
\end{pmatrix} \left( \begin{array}{cc}
\lambda_1 & 0 \\
0 & \lambda_2
\end{array} \right) \begin{pmatrix}
1 & -\lambda_2 \\
-1 & \lambda_1
\end{pmatrix}
$$

with

$$
\lambda_1 = \frac{1 - \sqrt{5}}{2} \quad \text{and} \quad \lambda_2 = \frac{1 + \sqrt{5}}{2}.
$$

It follows that

$$
\begin{pmatrix}
1 & 1 \\
1 & 0
\end{pmatrix}^m = \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix}
\lambda_1 & \lambda_2 \\
1 & 1
\end{pmatrix} \left( \begin{array}{cc}
\lambda_1^m & 0 \\
0 & \lambda_2^m
\end{array} \right) \begin{pmatrix}
1 & -\lambda_2 \\
-1 & \lambda_1
\end{pmatrix}
$$

for all integer values of $m$. Consequently, we have

$$
\begin{pmatrix}
Q_n \\
R_n
\end{pmatrix} = \frac{\log(2)}{\lambda_2 - \lambda_1} \left( \begin{array}{cc}
\lambda_2^n & \lambda_1^n \\
\lambda_2^n & \lambda_2^n
\end{array} \right) + \sum_{i=2}^{n} E_i \left( \begin{array}{cc}
\lambda_1^{n-i+1} - \lambda_2^{n-i+1} \\
\lambda_1^{n-i} - \lambda_2^{n-i}
\end{array} \right)
$$

for all $n \geq 2$ as $Q_0 = 0$ and $Q_1 = \log(2)$. Therefore, we obtain

$$
Q_n = \frac{\log(2)}{\lambda_2 - \lambda_1} (\lambda_2^n - \lambda_1^n) + \frac{1}{\lambda_2 - \lambda_1} \sum_{i=2}^{n} E_i (\lambda_2^{n-i+1} - \lambda_1^{n-i+1})
$$

for all $n \geq 2$. Since the sequence $(E_n)_{n \geq 2}$ is bounded for every $n \geq 2$ (as $\lim_{n \to \infty} E_n = 0$ by virtue of (2) and (3)), we derive that

$$
\sum_{i=2}^{n} |E_i| \cdot |\lambda_1|^{n-i+1} \leq \frac{|\lambda_1^n - |\lambda_1|}{|\lambda_1| - 1} \cdot \sup_{2 \leq m \leq n} |E_m|
$$

for all $n \geq 2$. This implies that the quantity

$$
\frac{1}{\lambda_2 - \lambda_1} \sum_{i=2}^{n} E_i \cdot \lambda_1^{n-i+1}
$$

converges to a definite limit as $n \to \infty$ (note that $|\lambda_1| < 1$ and $\sup_{2 \leq m \leq n} |E_m|$ is finite for every $n \geq 2$). On the other hand, we have

$$
0 \leq \sum_{i=n+1}^{\infty} E_i \cdot \lambda_2^{n-i+1} \leq \frac{\lambda_2}{\lambda_2 - 1} \cdot \sup_{m \geq n+1} |E_m|
$$

for all $n \geq 2$ (note that $\lambda_2 > 1$), which implies that

$$
\frac{1}{\lambda_2 - \lambda_1} \sum_{i=n+1}^{\infty} E_i \cdot \lambda_2^{n-i+1} = O\left( \sup_{m \geq n+1} |E_m| \right) = o(1)
$$

as $n \to \infty$. Putting everything together, we arrive at

$$
Q_n = \frac{\lambda_2^n}{\lambda_2 - \lambda_1} \left( \log(2) + \sum_{i=2}^{\infty} E_i \cdot \lambda_2^{-i+1} \right) - \frac{1}{\lambda_2 - \lambda_1} \sum_{i=2}^{n} E_i \cdot \lambda_1^{n-i+1}
$$

as $n \to \infty$. We deduce that

$$
A_n = \left( 1 + O\left( \lambda_1^n + \sup_{m \geq n+1} |E_m| \right) \right)
$$

as $n \to \infty$. We deduce that
\[
\exp \left( \frac{\lambda_1^2}{\lambda_2 - \lambda_1} \left( \log(2) + \sum_{i=2}^{\infty} E_i \cdot \lambda_2^{-i+1} \right) - \frac{1}{\lambda_2 - \lambda_1} \sum_{i=2}^{\infty} E_i \cdot \lambda_1^{n-i+1} \right) \\
= (1 + o(1)) \cdot \exp \left( \frac{\lambda_1^2}{\lambda_2 - \lambda_1} \left( \log(2) + \sum_{i=2}^{\infty} E_i \cdot \lambda_2^{-i+1} \right) - \frac{1}{\lambda_2 - \lambda_1} \sum_{i=2}^{\infty} E_i \cdot \lambda_1^{n-i+1} \right)
\]
as \( n \to \infty \). Denote by \( K_2 \) the quantity
\[
\exp \left( - \frac{1}{\lambda_2 - \lambda_1} \cdot \lim_{n \to \infty} \left( \sum_{i=2}^{n} E_i \cdot \lambda_2^{-i+1} \right) \right)
\]
and by \( K_1 \) the quantity
\[
\exp \left( - \frac{1}{\lambda_2 - \lambda_1} \cdot \lim_{n \to \infty} \left( \sum_{i=2}^{n} E_i \cdot \lambda_1^{n-i+1} \right) \right).
\]
Thus,
\[
A_n = N(f_n) = (1 + o(1)) K_1 \cdot K_2 \lambda_2^n = (1 + o(1)) K_1 \cdot K_2 \left( \frac{1 + \sqrt{5}}{2} \right)^n
\]
as \( n \to \infty \), where \( K_1 \) and \( K_2 \) can now be written as
\[
K_2 = \exp \left( \frac{1}{\sqrt{5}} \left( \log(2) + \sum_{i=2}^{\infty} \left( \frac{1 + \sqrt{5}}{2} \right)^{-i+1} \right) \cdot \log \left( 1 + \frac{1}{2N(f_{i-2})} - \frac{N(f_{i-2})}{2N(f_{i-3})} + \frac{1}{N(f_{i-2})} \right) \right)
\]
and
\[
K_1 = \exp \left( - \frac{1}{\sqrt{5}} \cdot \lim_{n \to \infty} \left( \sum_{i=2}^{n} \left( \frac{1 - \sqrt{5}}{2} \right)^{n-i+1} \right) \cdot \log \left( 1 + \frac{1}{2N(f_{i-2})} - \frac{N(f_{i-2})}{2N(f_{i-3})} + \frac{1}{N(f_{i-2})} \right) \right).
\]
By (numerically) evaluating \( K_1 \) and \( K_2 \), we obtain that the number of nonisomorphic leaf-induced subtrees of the leaf-Fibonacci tree \( f_n \) is asymptotically
\[
1.00001887227319 \cdots (1.48369689570172 \cdots) \left( \frac{1 + \sqrt{5}}{2} \right)^n
\]
as \( n \to \infty \). This completes the proof of the theorem. \( \square \)

The asymptotic formula from Theorem 2.1 can also be written in terms of the Fibonacci number \( F_n \): indeed, the number of leaves of \( f_n \) is given by
\[
|f_n| = F_{n+2} = \frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} \right)^{2+n} - \left( \frac{1 - \sqrt{5}}{2} \right)^{2+n}
\]
for every \( n \); so we deduce that
\[
\frac{10}{5 + 3\sqrt{5}} \cdot |f_n| \sim \left( \frac{1 + \sqrt{5}}{2} \right)^n
\]
as \( n \to \infty \). This implies that
\[
N(f_n) \sim K_1 \cdot K_2 \left( \frac{1 + \sqrt{5}}{2} \right)^n |f_n|
\]
\[
= 1.00001887227319 \cdots (1.48369689570172 \cdots) \left( \frac{5 + 3\sqrt{5}}{2} \right) |f_n|
\]
as \( n \to \infty \).

**Acknowledgments**

The author was supported by Stellenbosch University in association with African Institute for Mathematical Sciences (AIMS) South Africa. The author thanks Stephan Wagner for a suggestion in the proof of the asymptotic formula.
References


