

ON LEAF-INDUCED SUBTREES IN TOPOLOGICAL TREES

AUDACE A. V. DOSSOU-OLORY AND IGNATIUS BOADI*

ABSTRACT. A subtree can be induced in a natural way by a subset of leaves of a rooted tree. We study the number of nonisomorphic subtrees induced by leaves (leaf-induced subtrees) of a rooted tree with no vertex of outdegree 1 (topological tree). We show that only stars and binary caterpillars have the minimum number of nonisomorphic leaf-induced subtrees among all topological trees with a given number of leaves. We obtain a closed formula and a recursive formula for the families of d -ary caterpillars and complete d -ary trees, respectively. An asymptotic formula is found for complete d -ary trees using polynomial recurrences. We also show that the complete binary tree of height $h > 1$ contains precisely $\lfloor 2(1.24602\dots)^{2h} \rfloor$ nonisomorphic leaf-induced subtrees.

2000 MATHEMATICS SUBJECT CLASSIFICATION 05C05, 05C10, 05C30, 05C35.

KEYWORDS AND PHRASES. topological trees, non-isomorphic leaf-induced subtrees, complete d -ary trees, d -ary caterpillars, stars.

Submission Date: 30 January 2024

*Corresponding Author.

1. INTRODUCTION

Analogous to sampling in statistics, subtrees of trees are very useful when networks are being analysed. Since networks can be large and data structures are usually complex and large, subtrees are needed because they are usually more manageable to analyse than the whole structure. Also, to determine the reliability of a network when there is a vertex or edge failure, one can look at the number of subtrees of the graph that represent the network. The larger the number of subtrees of this graph, the more reliable this network is; see, as examples, [4–6]. Furthermore, leaf-induced subtrees are applicable in web analytics, especially in analysing access patterns of visitors of a particular website [14]. EvoMiner, an efficient algorithm for frequent subtree mining (FST) of subtrees of phylogenetic trees, is introduced in [15]. FST is particularly useful when considering the ancestry of the current generation, which are naturally the leaves of a rooted tree [16].

When a vertex of a tree is designated as the root, we have a rooted tree. A topological tree is a rooted tree with no vertex of outdegree one (or degree two, except possibly the root). They are also referred to as series-reduced or homeomorphically irreducible trees [20, 21]. By definition, the tree that has only one vertex is topological. A topological tree with each vertex having outdegree at most d is a d -ary tree. In a strict or full d -ary tree, each vertex has outdegree 0 or d . They are applicable mainly in computer science, where they are used in designing and building data structures. This leads to efficient data organisation, easy access and modification. In [7], d -ary trees are used to develop an efficient solution to the Empirical Cumulative Distribution Function (ECDF) searching

problem. The ECDF searching problem arises in multivariate statistics where given N points in d -dimensional space, a point $X = \{x_1, x_2, \dots, x_d\}$ dominates another point $Y = \{y_1, y_2, \dots, y_d\}$ if for all $i \in \{1, 2, \dots, d\}$, $x_i \geq y_i$. The ECDF searching problem primarily deals with finding the ratio of the number of Y 's dominated by a given X , to N , the number of data points [11]. Data structures using d -ary trees are designed in [7], which enables efficient ways of conducting ECDF queries.

A caterpillar is a tree whose internal vertices lie on a single path (called the backbone), and all leaves are attached directly to the internal vertices. The caterpillar is found to be an extremal graph when determining the maximum or minimum of various distance-based graph invariants [9, 10]. For example, Dadedzi et al. found in [10] that the caterpillar has the greatest distance spectral radius among trees with a given degree sequence. Caterpillars are applicable in layouts of communication and electrical networks where the internal vertices represent major outlets and leaves represent consumers. They are also widely used in chemical graph theory; see, as an example, [2]. They are applied in modelling interactions and in the analysis of benzenoid hydrocarbons. Caterpillars are sometimes called Gutman trees, since Ivan Gutman introduced them to chemical graph theory and made extensive use of them in his works; see, as examples, [1, 3]. They are also sometimes called benzenoid trees [8]. When one of the internal vertices of a caterpillar is designated as the root, we have a rooted caterpillar.

We mainly focus on two families of topological trees in this work:

- (1) *d*-ary caterpillars: A *d*-ary caterpillar is defined as a strict *d*-ary tree which is also a rooted caterpillar. A *d*-ary caterpillar is denoted by F_n^d , where n is the number of leaves. In Figure 1, we show the ternary caterpillar with 7 leaves, F_7^3 . At each level of a *d*-ary caterpillar, there are $d - 1$ leaves and one internal vertex except at the highest level, where there are d leaves and one internal vertex. Thus, the number of leaves of a *d*-ary caterpillar of height h is $1 + h(d - 1)$.

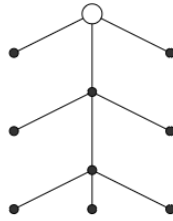


FIGURE 1. The ternary caterpillar with 7 leaves, F_7^3 .

- (2) The complete *d*-ary tree of height h , denoted by C_h^d , is a strict *d*-ary tree which has all leaves at distance h from the root, thus all leaves lie at the same level in the tree. In Figure 2, we show the complete ternary tree whose height is two, C_2^3 . In a complete *d*-ary tree, there are d vertices at level 1. Each of these vertices is then connected to d vertices at level 2, yielding d^2 vertices at level 2. Hence, at level h , there are d^h vertices. Since all leaves of a complete *d*-ary tree are at level h , the number of leaves of C_h^d is d^h . When $h = 0$, we have

the one-vertex tree, and when $h = 1$, we have the star with d leaves.

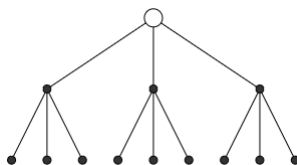


FIGURE 2. The complete ternary tree with height 2, C_2^3 .

Two graphs are isomorphic if you can label both graphs with the same labels such that any selected vertex has the same neighbours in both graphs.

By a leaf-induced subtree of a rooted tree, we mean a subtree induced by selected leaves of a tree. A subtree can be induced in a natural way by a subset of leaves of a rooted tree. The number of nonisomorphic leaf-induced subtrees of a given rooted tree is the main subject of this work. The subject of leaf-induced subtrees of rooted trees has been well studied [12,13,17–19,22–24]. Matula in [25] studied the number of nonisomorphic root-contained subtrees of complete d -ary trees of height h . The subtrees considered in [25] were, however, not induced by the leaves of the trees considered, but by any set of their vertices. Hence, the subtrees contained the root of the original tree but were not necessarily topological. For example, Matula had 66 nonisomorphic root-contained subtrees of the C_3^2 ; we obtain 17 nonisomorphic leaf-induced subtrees of this tree in this work.

To form a leaf-induced subtree of a given rooted tree T , we consider the power set of the leaves of T excluding

the null set. If S is a member of this set, then the subtree induced by the leaves in S is obtained by

- (1) extracting the smallest subtree of T containing no leaf other than those in S . To do this, we first locate the vertex which is the most recent common ancestor of the leaves in S . We then select all vertices and edges connecting the leaves in S to the vertex identified as the most recent common ancestor [13, 19],
- (2) contracting all vertices of outdegree 1 in the smallest subtree from step 1 [13, 19]. Suppose g is a vertex with outdegree 1 in the subtree obtained from step 1, and it has neighbours f and h , then we have edges $\{f, g\}$ and $\{g, h\}$; we contract g and create the edge $\{f, h\}$. This is done for all instances of vertices with outdegree 1 to create a leaf-induced subtree.

A single leaf can be selected and used to create a leaf-induced subtree, which is the one-vertex tree. By definition, any leaf-induced subtree of a rooted tree is a topological tree.

Definition 1.1. We define $N(T)$ as the number of nonisomorphic leaf-induced subtrees of a topological tree T .

Clearly, there is only one (up to isomorphism) subtree induced by a given subset of leaves of a topological tree. Hence, the total number of leaf-induced subtrees of a topological tree with n leaves is just $2^n - 1$ (excluding the empty set). We consider subtrees induced by leaves of a topological tree and select only one representative from every isomorphism class. In Example 1.1, we construct two isomorphic leaf-induced subtrees of the tree in Figure 3.

Example 1.1. Consider the tree labelled 1 in Figure 3. Let $L = \{l_1, l_3, l_4, l_6\}$. We first extract the smallest subtree containing

all the vertices and edges lying on the paths connecting the leaves in L to their most recent common ancestor. Hence, we get the leaf-induced subtree labelled 2 in Figure 3.

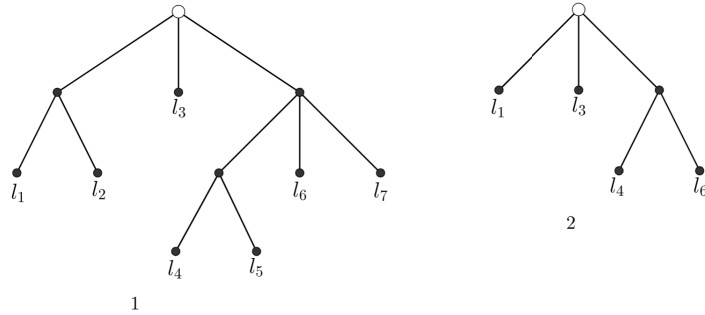


FIGURE 3. A topological tree and a subtree induced by some of its leaves.

2. MAIN RESULTS

We present formulas for the number of nonisomorphic leaf-induced subtrees of a d -ary caterpillar and a complete d -ary tree. We also give the minimum number of nonisomorphic leaf-induced subtrees of a topological tree with n leaves, together with the characterisation of all extremal trees.

All topological trees with up to four leaves are either stars or binary caterpillars except the trees shown in Figure 4.

There are 4 leaf-induced subtrees of C_2^2 and 5 of each of Q_4 and E_4^3 . Theorem 2.1 states the minimum number of nonisomorphic leaf-induced subtrees among all topological trees with $n > 4$ leaves and also states the corresponding extremal trees.

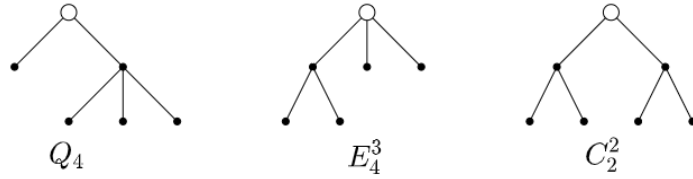


FIGURE 4. Some topological trees with 4 leaves.

Theorem 2.1. *Let $n \geq 5$ be an integer. Every n -leaf topological tree T has at least n nonisomorphic leaf-induced subtrees, with equality if and only if T is a star or a binary caterpillar.*

Proof. Consider the star S_n , with n leaves. Since all leaf-induced subtrees of S_n are themselves stars (see Figure 5), we obtain n nonisomorphic leaf-induced subtrees for S_n . Consider the binary caterpillar with n leaves f_n^2 . All leaf-induced subtrees of F_n^2 are also binary caterpillars. Hence we obtain n nonisomorphic leaf-induced subtrees for F_n^2 (see Figure 6).



FIGURE 5. All leaf-induced subtrees of S_n .

Finally, we show that a topological tree T with $n \geq 5$ leaves which is neither a star nor a binary caterpillar has at least $n+1$ nonisomorphic leaf-induced subtrees. Since T is neither a star nor a binary caterpillar, its root should have outdegree at most $n - 1$ and at most $n - 2$ leaves attached to it. The key to the proof of this part of the theorem is

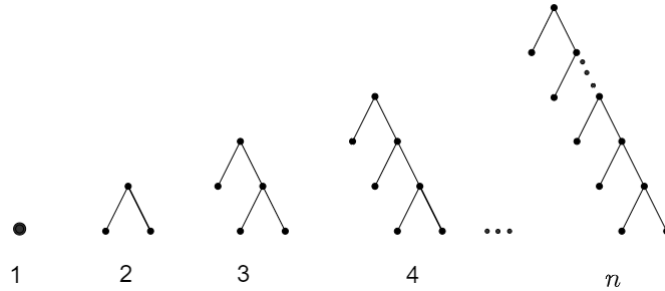


FIGURE 6. All leaf-induced subtrees of F_n^2 .

to show that there exists at least two nonisomorphic leaf-induced subtrees with k leaves for some $k \in \{3, \dots, n\}$. We consider two possible cases:

- (1) If T contains a vertex of outdegree at least three, then the leaf-induced subtrees with three leaves, S_3 and F_3^2 , are both present (see the trees labelled 3 in Figures 5 and 6 respectively). Hence, we have at least $n + 1$ nonisomorphic leaf-induced subtrees for T .
- (2) If T is a binary tree, then it must be of height at least three since it has at least five leaves. The trees C_2^2 (see Figure 4) and F_2^2 (see the tree labeled 4 in Figure 6) each with four leaves, are leaf-induced subtrees of T . Hence, we have at least $n + 1$ nonisomorphic leaf-induced subtrees for T .

□

2.1 d -ARY CATERPILLARS

Theorem 2.2. *Let $d \geq 3$ be a fixed positive integer. For the d -ary caterpillar F_n^d , we have*

$$N(F_n^d) = \frac{(d-1)^{\frac{n+d-2}{d-1}} - 1}{d-2}.$$

Proof. Since F_n^d is a strict d -ary tree, it has precisely $(n-1)/(d-1)$ internal vertices, which is also its height. Consider the internal vertex at the highest level, which has d leaves.

- (1) We can attach from 2 to d leaves to this vertex to get $d-1$ leaf-induced subtrees. Hence, in addition to the one-leaf rooted subtree, we have $1 + (d-1)$ leaf-induced subtrees.
- (2) For each of the $d-1$ leaf induced subtrees from step 1, we move to the next internal vertex above the vertex considered in step 1 and attach from 1 to $d-1$ leaves of this vertex to give $(d-1)(d-1) = (d-1)^2$ leaf-induced subtrees resulting in $1 + (d-1) + (d-1)^2$ leaf-induced subtrees.
- (3) For each of the $(d-1)^2$ leaf-induced subtrees from step 2, we move to the next vertex and attach from 1 to $d-1$ leaves to give $(d-1)^2(d-1) = (d-1)^3$ leaf-induced subtrees yielding $1 + (d-1) + (d-1)^2 + (d-1)^3$ leaf-induced subtrees in total.

This process is repeated for the new subtrees formed up the backbone of the tree until we reach the root. The number of times this operation is done is equal to the height of

the tree. We therefore have

$$\begin{aligned} N(F_n^d) &= 1 + (d - 1) + (d - 1)^2 + \dots + (d - 1)^{\frac{n-1}{d-1}} \\ &= \frac{(d - 1)^{\frac{n+d-2}{d-1}} - 1}{d - 2}, \end{aligned}$$

as required. □

2.2 COMPLETE d -ARY TREES

In this section, we develop a recursive formula and an asymptotic formula for $N(C_h^d)$. Lemma 2.3 below will be used in deriving the asymptotic formula for $N(C_h^d)$.

Lemma 2.3. *Let $d \geq 2$ be a fixed integer and $a_0, a_1, a_2, \dots, a_d$ be real numbers such that $a_i \geq 0$, and $a_d \neq 0$. If $(A_n)_{n \geq 0}$ is a polynomial sequence defined recursively by*

$$A_0 > 0, A_n = \sum_{k=0}^d a_k \cdot A_{n-1}^k,$$

then

(1)

$$\begin{aligned} \log(A_n) &= \frac{d^n - 1}{d - 1} \log(a_d) \\ &\quad + d^n \left(\log(A_0) + \sum_{j=0}^{n-1} d^{-1-j} \cdot \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_j^{k-d} \right) \right) \end{aligned}$$

for every $n \geq 1$. Moreover, if $(A_n)_{n \geq 0}$ is an increasing sequence and $\lim_{n \rightarrow \infty} A_n = \infty$, we also obtain

$$(2) \quad A_n = (1 + o(1)) a_d^{-\frac{1}{d-1}} \exp \left(d^n \left(\frac{\log(a_d)}{d-1} + \log(A_0) + \sum_{j=0}^{\infty} d^{-1-j} \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_{n-1}^{k-d} \right) \right) \right)$$

as $n \rightarrow \infty$.

Proof. To prove (2), consider the recursion

$$\begin{aligned} A_n &= a_0 + a_1 A_{n-1} + a_2 A_{n-1}^2 + \cdots + a_d A_{n-1}^d, \\ \frac{A_n}{A_{n-1}^d} &= a_d + \frac{a_{d-1}}{A_{n-1}} + \frac{a_{d-2}}{A_{n-1}^2} + \cdots + \frac{a_0}{A_{n-1}^d}, \\ \log \left(\frac{A_n}{A_{n-1}^d} \right) &= \log \left(a_d \left(1 + \frac{a_{d-1}}{a_d A_{n-1}} + \frac{a_{d-2}}{a_d A_{n-1}^2} + \cdots + \frac{a_0}{a_d A_{n-1}^d} \right) \right), \\ \log A_n - d \log A_{n-1} &= \log a_d + \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_{n-1}^{k-d} \right), \\ (3) \quad \log A_n &= d \log A_{n-1} + \log a_d \\ &\quad + \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_{n-1}^{k-d} \right). \end{aligned}$$

From (3), we get

$$(4) \quad \begin{aligned} \log A_{n-1} &= d \log A_{n-2} + \log a_d \\ &\quad + \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_{n-2}^{k-d} \right). \end{aligned}$$

Substituting (4) into (3), we get

$$\begin{aligned} \log A_n &= d^2 \log A_{n-2} + (d+1) \log a_d + d \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_{n-2}^{k-d} \right) \\ &\quad + \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_{n-1}^{k-d} \right). \end{aligned}$$

Clearly, we have

$$(5) \quad \log A_{n-j} = d \log A_{n-j-1} + \log a_d + \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_{n-j-1}^{k-d} \right)$$

for every j such that $n-j > 0$. Using (5) to find expressions for A_{n-j} , $j = 1, 2, \dots, n-1$ and substituting into (3), we get

$$\begin{aligned} \log A_n &= (d^{n-1} + d^{n-2} + \dots + d + 1) \log a_d + d^n \log A_0 \\ &\quad + d^{n-1} \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_0^{k-d} \right) + \dots \\ &\quad + d \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_{n-2}^{k-d} \right) + \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_{n-1}^{k-d} \right) \\ (6) \quad &= \frac{d^n - 1}{d - 1} \log a_d + d^n \log A_0 \\ &\quad + d^n \left(\sum_{j=0}^{n-1} d^{-1-j} \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_j^{k-d} \right) \right). \end{aligned}$$

This proves (2). Consider the series

$$d^n \sum_{j=n}^{\infty} d^{-1-j} \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_j^{k-d} \right).$$

Note that we have $k < d$. Recall that $a_k/a_d \geq 0$. Since $\sup_{j \geq n} A_j^{k-d} = A_n^{k-d}$, we have

$$\begin{aligned} 0 &\leq d^n \sum_{j=n}^{\infty} d^{-1-j} \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_j^{k-d} \right) \\ &\leq d^n \sum_{j=n}^{\infty} d^{-1-j} \cdot \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_n^{k-d} \right) \\ &= \frac{1}{d-1} \cdot \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_n^{k-d} \right) \end{aligned}$$

and

$$\lim_{n \rightarrow \infty} \left(\frac{1}{d-1} \cdot \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_n^{k-d} \right) \right) = 0.$$

Hence

$$d^n \sum_{j=n}^{\infty} d^{-1-j} \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_j^{k-d} \right) = o(1).$$

Now, let

$$(7) \quad R_n(d) = \sum_{j=n}^{\infty} d^{-1-j} \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_j^{k-d} \right).$$

We rewrite (6) as

$$\begin{aligned} \log A_n &= \frac{d^n - 1}{d - 1} \log a_d + d^n \log A_0 \\ &\quad + d^n \left(\sum_{j=0}^{n-1} d^{-1-j} \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_j^{k-d} \right) + R_n(d) - R_n(d) \right) \\ &= \frac{d^n - 1}{d - 1} \log a_d + d^n \log A_0 \\ &\quad + d^n \left(\sum_{j=0}^{\infty} d^{-1-j} \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_j^{k-d} \right) - R_n(d) \right), \end{aligned}$$

which holds for all $n \geq 1$. Let

$$K(d) = \sum_{j=0}^{\infty} d^{-1-j} \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_j^{k-d} \right).$$

Then we have

$$\begin{aligned} \log A_n &= d^n \left(\log(A_0) + \frac{\log(a_d)}{d - 1} + K(d) \right) - \frac{\log(a_d)}{d - 1} + o(1), \\ A_n &= \exp \left(d^n \left(\log(A_0) + \frac{\log(a_d)}{d - 1} + K(d) \right) - \frac{\log(a_d)}{d - 1} + o(1) \right) \\ &= \exp(o(1)) \cdot \exp \left(\log \left(a_d^{\frac{-1}{d-1}} \right) \right) \cdot \exp \left(d^n \left(\log(A_0) \right. \right. \\ &\quad \left. \left. + \frac{\log(a_d)}{d - 1} + K(d) \right) \right). \end{aligned}$$

Using Taylor series expansion for $\exp(o(1))$, we have

$$= (1 + o(1)) a_d^{-\frac{1}{d-1}} \exp \left(d^n \left(\log(A_0) + \frac{\log(a_d)}{d - 1} + K(d) \right) \right)$$

as $n \rightarrow \infty$. This completes the proof of the lemma. \square

Theorem 2.4. *The number of nonisomorphic leaf-induced subtrees of the complete d -ary tree, C_h^d , is given by*

$$(8) \quad N(C_h^d) = -N(C_{h-1}^d) + \binom{d + N(C_{h-1}^d)}{d}$$

with $N(C_0^d) = 1$. Furthermore,

$$(9) \quad N(C_h^d) = (1 + o(1))(d!)^{\frac{1}{d-1}} \kappa(d)^{d^h}$$

for some effectively computable constant $\kappa(d) > 1$ as $h \rightarrow \infty$.

Proof. Fix $h \geq 1$. Call v the root of C_h^d and A_h the set of all nonisomorphic leaf-induced subtrees of C_h^d whose root is v and has root degree r is uniquely determined by a choice of a r -element multiset A_{h-1} (since all r branches lie in A_{h-1}). So for every $r \in \{2, 3, \dots, d\}$, there are precisely

$$\binom{N(C_{h-1}^d) + r - 1}{r}$$

such subtrees (with root degree r). Therefore, taking into consideration the subtree consisting only of the single leaf, and using the fact that the elements of A_{h-1} are already counted this way, we get

$$N(C_h^d) = 1 + \sum_{r=2}^d \binom{N(C_{h-1}^d) + r - 1}{r}$$

nonisomorphic leaf-induced subtrees. Let $j = N(C_{h-1}^d) + r - 1$. We get

$$\begin{aligned} N(C_h^d) &= 1 + \sum_{j=N(C_{h-1}^d)+1}^{N(C_{h-1}^d)+d-1} \binom{j}{j+1-N(C_{h-1}^d)} \\ &= 1 + \sum_{j=N(C_{h-1}^d)+1}^{N(C_{h-1}^d)+d-1} \binom{j}{N(C_{h-1}^d)-1} \quad \text{since } \binom{n}{r} = \binom{n}{n-r} \\ &= -N(C_{h-1}^d) + N(C_{h-1}^d) + 1 + \sum_{j=N(C_{h-1}^d)+1}^{N(C_{h-1}^d)+d-1} \binom{j}{N(C_{h-1}^d)-1} \\ &= -N(C_{h-1}^d) + \sum_{j=-1+N(C_{h-1}^d)}^{N(C_{h-1}^d)+d-1} \binom{j}{N(C_{h-1}^d)-1}. \end{aligned}$$

It can be shown by induction on d that

$$\sum_{j=-1+N(C_{h-1}^d)}^{N(C_{h-1}^d)+d-1} \binom{j}{N(C_{h-1}^d)-1} = \binom{d+N(C_{h-1}^d)}{N(C_{h-1}^d)}.$$

Hence we have

$$N(C_h^d) = -N(C_{h-1}^d) + \binom{d+N(C_{h-1}^d)}{N(C_{h-1}^d)} = -N(C_{h-1}^d) + \binom{d+N(C_{h-1}^d)}{d}.$$

Observe that $N(C_h^d)$ increases with h since the nonisomorphic leaf-induced subtrees of C_h^d include all nonisomorphic leaf-induced subtrees of C_{h-1}^d . To prove (9), first set

$A_h = N(C_h^d)$. We have

$$\begin{aligned}
 A_h &= -A_{h-1} + \binom{d + A_{h-1}}{d} \\
 &= -A_{h-1} + \frac{(A_{h-1} + d)!}{(A_{h-1}!)(d!)} \\
 &= -A_{h-1} \\
 &\quad + \frac{(A_{h-1} + d)(A_{h-1} + d - 1)(A_{h-1} + d - 2) \cdots (A_{h-1} + 1)}{d!} \\
 &= -A_{h-1} \\
 &\quad + \frac{A_{h-1}^d + \left(\sum_{i=1}^d i\right)A_{h-1}^{d-1} + \cdots + \left(\sum_{i=1}^d \prod_{j=1, j \neq i}^d j\right)A_{h-1} + d!}{d!} \\
 &= \frac{-A_{h-1}d! + A_{h-1}^d + \left(\sum_{i=1}^d i\right)A_{h-1}^{d-1} + \cdots + \left(\sum_{i=2}^d \prod_{j=1, j \neq i}^d j\right)A_{h-1}}{d!} \\
 &\quad + \frac{d!A_{h-1} + d!}{d!} \\
 &= \frac{A_{h-1}^d + \left(\sum_{i=1}^d i\right)A_{h-1}^{d-1} + \cdots + \left(\sum_{i=2}^d \prod_{j=1, j \neq i}^d j\right)A_{h-1} + d!}{d!}.
 \end{aligned}$$

By setting a_i as the coefficients of A_{h-1}^i , $i \in \{0, 1, 2, 3, \dots, d\}$, noting that $A_0 = 1$ and $a_d = 1/d!$ and using the identity derived in (3), we have

$$(10) \quad \log A_h = d \log A_{h-1} - \log d! + \log \left(1 + d! \sum_{k=0}^{d-1} a_k A_{h-1}^{k-d} \right).$$

From Lemma 2.3, we get

$$N(C_h^d) = (1 + o(1))(d!)^{\frac{1}{d-1}} \kappa(d)^{d^h},$$

where

$$(11) \quad \kappa(d) = \exp\left(-\frac{\log d!}{d-1} + K(d)\right)$$

and

$$(12) \quad K(d) = \sum_{j=0}^{\infty} d^{-1-j} \cdot \log\left(1 + d! \sum_{k=0}^{d-1} a_k (N(C_j^d))^{k-d}\right).$$

From (10), we get

$$(13) \quad \log A_{h+1} = d \log A_h - \log d! + \log\left(1 + d! \sum_{k=0}^{d-1} a_k A_h^{k-d}\right),$$

$$\log\left(\frac{d! A_{h+1}}{A_h^d}\right) = \log\left(1 + d! \sum_{k=0}^{d-1} a_k A_h^{k-d}\right).$$

Recall that $A_h = N(C_h^d)$. Substituting (13) into (12), we get

$$(14) \quad K(d) = \sum_{j=0}^{\infty} d^{-1-j} \cdot \log\left(d! \cdot \frac{N(C_{j+1}^d)}{N(C_j^d)^d}\right).$$

We have

$$-\frac{\log d!}{d-1} + K(d) = -\frac{\log d!}{d-1} + \sum_{j=0}^{\infty} d^{-1-j} \cdot \log\left(d! \cdot \frac{N(C_{j+1}^d)}{N(C_j^d)^d}\right).$$

Since $N(C_{j+1}^d) > N(C_j^d)^d$, we obtain

$$\begin{aligned} -\frac{\log d!}{d-1} + K(d) &> -\frac{\log d!}{d-1} + \sum_{j=0}^{\infty} d^{-1-j} \cdot \log d! \\ &= -\frac{\log d!}{d-1} + \frac{\log d!}{d-1} \\ &= 0. \end{aligned}$$

Hence $\kappa(d) > 1$. □

The value of $\kappa(d)$ derived in (11) for every d can be numerically evaluated. Table 1 gives values of $\kappa(d)$ for $d \in \{2, 3, \dots, 10\}$.

TABLE 1. Approximated values of $\kappa(d)$.

d	$\kappa(d)$
2	1.246020832983661
3	1.254860390384554
4	1.2189114976086313
5	1.1888457507131132
6	1.165394603276801
7	1.1469724134908297
8	1.1322182196849957
9	1.1201639471936817
10	1.1101387293827483

From Table 1, when $d = 2$, we get

$$N(C_h^2) \sim 2(1.246020832983661)^{2^h}$$

as $h \rightarrow \infty$.

Propositions 2.1 and 2.2 below give further results about $\kappa(d)$.

Proposition 2.1. *The sequence $\kappa(d)_{d \geq 2}$ converges to 1 as $d \rightarrow \infty$.*

Proof. From equations (11) and (12), we have

(15)

$$\begin{aligned} \log(\kappa(d)) &= -\frac{\log d!}{d-1} + \frac{\log d! + \log(d)}{d} \\ &\quad + \sum_{j \geq 1} \frac{\log d! + \log(N(C_{j+1}^d)) - d \cdot \log(N(C_j^d))}{d^{j+1}}. \end{aligned}$$

From (8), we can write

$$\begin{aligned} N(C_{j+1}^d) &= -N(C_j^d) + \binom{d + N(C_j^d)}{d} = \sum_{k=0}^d a_k \cdot N(C_j^d)^k \\ &\leq \sum_{k=0}^d a_k \cdot N(C_j^d)^d \end{aligned}$$

$$N(C_{j+1}^d) \leq N(C_j^d)^d \sum_{k=0}^d a_k$$

$$(16) \quad \frac{N(C_{j+1}^d)}{N(C_j^d)^d} \leq \sum_{k=0}^d a_k = d.$$

We make a substitution of (16) into (15). We have

$$\begin{aligned} 0 \leq \log(\kappa(d)) &\leq -\frac{\log d!}{d-1} + \frac{\log d! + \log(d)}{d} + \sum_{j \geq 1} \frac{\log d! + \log(d)}{d^{j+1}} \\ &= -\frac{\log d!}{d-1} + \frac{\log d! + \log(d)}{d} + (\log d! + \log(d)) \sum_{j \geq 1} \frac{1}{d^{j+1}} \\ &= -\frac{\log d!}{d-1} + \frac{\log d! + \log(d)}{d} + \frac{\log d! + \log(d)}{d(d-1)} \\ &= \frac{\log d}{d-1} \end{aligned}$$

It is clear that

$$\lim_{d \rightarrow \infty} \frac{\log d}{d-1} = 0.$$

Hence $\kappa(d) \rightarrow 1$ as $d \rightarrow \infty$. □

Proposition 2.2. *Given $d \geq 2$, there exists a positive integer $H \geq 2$ such that*

$$N(C_h^d) = \lfloor d!^{\frac{1}{d-1}} \kappa(d)^{d^h} \rfloor$$

for every $h \geq H$.

Proof. From (7), we can show that the relation

$$0 < R_n(d) \leq \frac{d^{-n}}{d-1} \cdot \log \left(1 + \sum_{k=0}^{d-1} \frac{a_k}{a_d} \cdot A_n^{k-d} \right)$$

holds for every $d \geq 2$ and every $n \geq 1$. We deduce the double inequality

$$\begin{aligned} (17) \quad & d^n \left(\sum_{j=0}^{\infty} d^{-1-j} \cdot \log \left(1 + d! \sum_{k=0}^{d-1} a_k \cdot A_j^{k-d} \right) \right. \\ & \left. - \frac{d^{-n}}{d-1} \cdot d! \sum_{k=0}^{d-1} a_k \cdot A_n^{k-d} \right) - \frac{d^n - 1}{d-1} \log d! \leq \log(A_n) \\ & \leq d^n \left(\sum_{j=0}^{\infty} d^{-1-j} \cdot \log \left(1 + d! \sum_{k=0}^{d-1} a_k \cdot A_j^{k-d} \right) \right) - \frac{d^n - 1}{d-1} \log d! \end{aligned}$$

From (17), we obtain

$$\begin{aligned}
 (18) \quad & d^n \left(\sum_{j=0}^{\infty} d^{-1-j} \cdot \log \left(1 + d! \sum_{k=0}^{d-1} a_k \cdot A_j^{k-d} \right) \right) - \frac{d!}{d-1} A_n^{-1} \sum_{k=0}^{d-1} a_k \\
 & - \frac{d^n}{d-1} \log d! + \frac{1}{d-1} \log d! \leq \log(A_n) \leq d^n \left(\sum_{j=0}^{\infty} d^{-1-j} \cdot \log \left(1 \right. \right. \\
 & \left. \left. + d! \sum_{k=0}^{d-1} a_k \cdot A_j^{k-d} \right) \right) - \frac{d^n}{d-1} \log d! + \frac{1}{d-1} \log d!
 \end{aligned}$$

Given that $\sum_{k=0}^{d-1} a_k = \frac{dd!-1}{d!}$ and substituting (14) into (18), we obtain

$$\begin{aligned}
 & d^n \left(-\frac{\log d!}{d-1} + K(d) \right) - \frac{d!}{d-1} \cdot A_n^{-1} \cdot \frac{dd!-1}{d!} + \log d!^{\frac{1}{d-1}} \\
 & \leq \log(A_n) \leq d^n \left(-\frac{\log d!}{d-1} + K(d) \right) + \log d!^{\frac{1}{d-1}} \\
 & d!^{\frac{1}{d-1}} \kappa(d)^{d^n} \cdot \exp \left(-\frac{d!}{d-1} \cdot A_n^{-1} \cdot \frac{dd!-1}{d!} \right) \leq A_n \leq d!^{\frac{1}{d-1}} \kappa(d)^{d^n}
 \end{aligned}$$

Replacing A_n with $N(C_h^d)$ and using the basic inequality $e^t \geq 1+t$, we obtain

$$(19) \quad d!^{\frac{1}{d-1}} \kappa(d)^{d^h} - \frac{d!^{\frac{d}{d-1}} \cdot \kappa(d)^{d^h}}{d-1} \cdot \frac{1}{N(C_h^d)} \left(d - \frac{1}{d!} \right) \leq N(C_h^d) \leq d!^{\frac{1}{d-1}} \kappa(d)^{d^h}$$

We substitute $N(C_h^d)$ in the left-hand side of (19) with

$$d!^{\frac{1}{d-1}} \kappa(d)^{d^h} - \frac{d!^{\frac{d}{d-1}} \cdot \kappa(d)^{d^h}}{d-1} \cdot \frac{1}{N(C_h^d)} \left(d - \frac{1}{d!} \right).$$

Iterating this substitution process a (finite) number of times leads to

$$d!^{\frac{1}{d-1}} \kappa(d)^{d^h} - \delta(d) \leq N(C_h^d) \leq d!^{\frac{1}{d-1}} \kappa(d)^{d^h}$$

for some constant $0 < \delta(d) < 1$. An immediate consequence of this double inequality is that

$$N(C_h^d) = \lfloor d!^{\frac{1}{d-1}} \kappa(d)^{d^h} \rfloor$$

holds for all $h \geq H$ for certain H . □

For $d = 2$, the previous condition $A_h > (dd! - 1)(d - 1)$ becomes $A_h > 3$, which means that $N(C_h^2) \geq 4$, hence $H = 2$. We deduce that

$$N(C_h^2) = \lfloor 2(1.2460208329836625089431529441999359284665241772983812581\dots)^{2^h} \rfloor$$

for every $h \geq 2$.

CONCLUSION

In this work, we determined formulas for the number of nonisomorphic leaf-induced subtrees of d -ary caterpillars and complete d -ary trees. We began by showing how to derive the nonisomorphic leaf-induced subtrees of a topological tree. A closed formula was derived for d -ary caterpillars only. For the complete d -ary trees, however, we have an asymptotic formula. Closed formulas have not been found yet, and they are nearly impossible to obtain. This can be a basis for future work.

REFERENCES

- [1] Ivan Gutman and Edward C Kirby. Benzenoids with branching graphs that are trees. *Croatica Chemica Acta*, 73(4): 943–955, 2000.
- [2] Stephan Wagner and Hua Wang. Introduction to chemical graph theory. *CRC Press*, 2018.
- [3] Ivan Gutman and Sven J. Cyvin. Introduction to the theory of benzenoid hydrocarbons. *Springer Science & Business Media*, 2012.

- [4] Yuzhi Xiao, Haixing Zhao, Zhen Liu and Yaping Mao. Trees with large numbers of subtrees. *International Journal of Computer Mathematics*, 94(2): 372–385, 2017.
- [5] Laszlo A. Székely and Hua Wang. Binary trees with the largest number of subtrees. *Discrete applied mathematics*, 155(3): 374–385, 2007.
- [6] Laszlo A. Székely and Hua Wang. On subtrees of trees. *Advances in Applied Mathematics*, 34(1): 138–155, 2005.
- [7] Maurizio Talamo and Giorgio Gambosi. An application of m-ary trees to the design of data structures for geometric searching problems. *RAIRO-Theoretical Informatics and Applications*, 23(2): 165–176, 1989.
- [8] Sherif El-Basil. Applications of caterpillar trees in chemistry and physics. *Journal of Mathematical Chemistry*, 1(2): 153–174, 1987.
- [9] Nina S. Haug, Stephan Wagner and Hua Wang. Greedy trees, caterpillars, and Wiener-type graph invariants. *MATCH Communications in Mathematical and Computer Chemistry*, 68(1): 273–292, 2012.
- [10] Kenneth Dadedzi, Valisoa R. Misanantenaina and Stephan Wagner. On the distance spectral radius of trees with given degree sequence. *Discussiones Mathematicae Graph Theory*, 40(2): 295–524, 2020.
- [11] Jon L. Bentley and Michael I. Shamos. A problem in multivariate statistics: Algorithm, data structure and applications. *Carnegie-Mellon Univ Pittsburg PA Dept of Computer Science*, 1978.
- [12] Audace A. V. Dossou-Olory. Leaf-induced subtrees of leaf-Fibonacci trees. *Discrete Mathematics Letter*, 1: 1–7, 2019.
- [13] Éva Czabarka, Audace A. V. Dossou-Olory and Laszlo A. Székely. Inducibility of d -ary trees. *Discrete Mathematics*, 343(2): 111671, 2020.
- [14] Alexandre Blondin Massé et al. Fully leafed induced subtrees. *In International Workshop on Combinatorial Algorithms*, 90–101, 2018.
- [15] Akshay Deepak et al. EvoMiner: frequent subtree mining in phylogenetic databases. *Knowledge and Information Systems*, 41(3): 559–590, 2014.
- [16] Xin-Hui Zou et al. Analysis of 142 genes resolves the rapid diversification of the rice genus. *Genome biology*, 9(3): 1–13, 2008.

- [17] Audace A. V. Dossou-Olory and Stephan Wagner. On the inducibility of small trees. *Discrete Mathematical and Theoretical Computer Science*, 21(4): 5804, 2019.
- [18] Audace A. V. Dossou-Olory. The minimum asymptotic density of binary caterpillars. *Graphs and Combinatorics*, 35(1): 303–320, 2019.
- [19] Éva Czabarka, Laszlo A. Székely and Stephan Wagner. Inducibility in binary trees and crossings in random tanglegrams. *Society for Industrial and Applied Mathematics Journal on Discrete Mathematics*, 31(3): 1732–1750, 2017.
- [20] Francois Bergeron, Gilbert Labelle and Pierre Leroux. Combinatorial species and tree-like structures. *Cambridge University Press*, No. 67, 1998.
- [21] Elizabeth S. Allman and John A. Rhodes. Mathematical models in biology: an introduction. *Cambridge University Press*, 2004.
- [22] Audace A. V. Dossou-Olory and Stephan Wagner. Inducibility of topological trees. *Quaestiones Mathematicae* 42(6): 749–764, 2019.
- [23] Audace A. V. Dossou-Olory and Stephan Wagner. Further results on the inducibility of d -ary trees. *Australas J. Comb.* 81: 1–4, 2021.
- [24] Audace A. V. Dossou-Olory. On the inducibility of rooted trees. PhD dissertation, *Stellenbosch: Stellenbosch University*, 2018.
- [25] David Matula On the number of subtrees of a symmetric n -ary tree. *SIAM Journal on Applied Mathematics* 18(3): 688–703, 1970.

AUDACE A. V. DOSSOU-OLORY, INSTITUT DE MATHÉMATIQUES ET DE SCIENCES PHYSIQUES (IMSP), DANGBO,, AND INSTITUT NATIONAL DE L'EAU (INE) / CENTRE D'EXCELLENCE D'AFRIQUE POUR L'EAU ET L'ASSAINISSEMENT, UNIVERSITY OF ABOMEY-CALAVI, BENIN.

Email address: audace@aims.ac.za

IGNATIUS BOADI, AFRICAN INSTITUTE FOR MATHEMATICAL SCIENCES, GHANA, SUMMERHILL ESTATES, EAST LEGON HILLS, SANTOE ACCRA, GHANA.

Email address: ignatius@aims.edu.gh