

Inducibility of d -ary trees

Éva Czabarka^{a,b}, Audace A.V. Dossou-Olory^{c,1}, László A. Székely^{a,b,2},
Stephan Wagner^{c,*,3}



^a Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA

^b Department of Mathematics and Applied Mathematics, University of Johannesburg, P.O. Box 524, Auckland Park, Johannesburg 2006, South Africa

^c Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

ARTICLE INFO

Article history:

Received 11 February 2018

Received in revised form 8 April 2019

Accepted 12 September 2019

Available online 26 September 2019

Keywords:

d -ary trees

Leaf-induced subtrees

Inducibility

Maximum density

Stars

Caterpillars

ABSTRACT

Imitating the binary inducibility, a recently introduced invariant of binary trees (Czabarka et al., 2017), we initiate the study of the inducibility of d -ary trees (rooted trees whose vertex outdegrees are bounded from above by $d \geq 2$). We determine the exact inducibility for stars and binary caterpillars. For T in the family of strictly d -ary trees (every vertex has 0 or d children), we prove that the difference between the maximum density of a d -ary tree D in T and the inducibility of D is of order $\mathcal{O}(|T|^{-1/2})$ compared to the general case where it is shown that the difference is $\mathcal{O}(|T|^{-1})$ which, in particular, responds positively to a conjecture on the inducibility in binary trees. We also discover that the inducibility of a binary tree in d -ary trees is independent of d . Furthermore, we establish a general lower bound on the inducibility and also provide a bound for some special trees. Moreover, we find that the maximum inducibility is attained for binary caterpillars for every d .

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

The inducibility of graphs is an old topic in extremal graph theory. However, the inducibility of trees with a given number of leaves is a novel concept that has been put forward recently in [6], inspired by a problem stemming from phylogenetics. The specific problem that is addressed in [6] can be stated as maximizing the asymptotic density of appearances of ‘small’ binary trees in binary trees with ‘large’ number of leaves. The present work is aimed at extending the concept from binary trees to rooted trees with bounded degrees in which no vertex has degree 2 (except possibly for the root).

The problem of maximizing the density of graphs in larger graphs is a key concept that has been explored starting from the work [15] of Golumbic and Pippenger and has ever since been of interest to graph theorists. In [15], the authors

* Corresponding author.

E-mail addresses: czabarka@math.sc.edu (É. Czabarka), audace@aims.ac.za (A.A.V. Dossou-Olory), szekely@math.sc.edu (L.A. Székely), swagner@sun.ac.za (S. Wagner).

¹ The second author was supported by Stellenbosch University, South Africa and African Institute for Mathematical Sciences (AIMS) South Africa.

² The third author was supported in part by the NSF DMS, grant number 1600811.

³ The fourth author was supported by the National Research Foundation of South Africa, grant number 96236.

introduced the subject for simple graphs: for finite and simple graphs G and H with k and n vertices, respectively, let $\mathcal{I}(G, H)$ be the number of distinct subgraphs induced by k distinct vertices of H which are isomorphic to G . The limit

$$I(G) := \lim_{n \rightarrow \infty} \left(\max_{|V(H)=n} \mathcal{I}(G, H) / \binom{n}{k} \right),$$

where the maximum runs over all finite and simple graphs H on $n = |V(H)| \geq k$ vertices, always exists and is called the *inducibility* of the graph G . In this setting, it is the asymptotic behavior of the maximum number of appearances of G as subgraph in an arbitrary n -vertex graph as $n \rightarrow \infty$ which is captured by the inducibility $I(G)$.

There have been numerous investigations on $I(G)$ in special cases. The maximum number of induced subgraphs of H isomorphic to the complete bipartite graph $K_{k,k}$ has been studied in a 1986 paper [3] by Bollobás et al. A great deal of work (see [2,4,10]) followed [3] afterwards. As such, Brown and Sidorenko [4] computed explicitly $I(K_{k,k+l})$ for all $l \geq 1$ but under the restriction $k \geq l(l-1)/2$. For $k < l(l-1)/2$, they stated the result as a function of the maximum over $[0, 1]$ of a certain polynomial in a single variable.

In 2014, Hirst [13] determined, employing Razborov’s flag algebra method and semi-definite programming techniques, the inducibility of two 4-vertex graphs: the complete tripartite graph $K_{1,1,2}$ and the so-called paw graph (graph constructed from a triangle by appending a pendant edge). The concept of inducibility is still gaining consideration from several research groups; see [9,12] for some recent results on blow-up of graphs and graphs on four vertices, respectively. The language of flag algebras was also employed recently in [1] to derive the inducibility of the cycle on five vertices, thereby settling a particular case of a conjecture formulated in [15].

Let us mention that there has also been interest in oriented graphs. For instance, in 2011, Sperfeld [17] explored, by means of flag algebra, the inducibility of (monodirected) graphs with at most four vertices. Three years later, Huang [14] determined, by means of a different approach, the maximum induced proportion of directed star graphs.

2. Inducibility of trees

One of the most widely used classes of graphs is the class of trees. Due to their prevalence in many situations, the study of the inducibility of trees emerged in 2016 [5,6], and has been addressed in two different settings.

In [5], Bubeck and Linial investigated what they called *k-profile of trees*, which is the vector of induced proportions of trees with k vertices classified according to their isomorphism type. At the end of their paper, Bubeck and Linial briefly defined a notion of inducibility of trees in this specific situation. Analogously, originally motivated by a question from phylogenetics, the recent paper [6] introduced a further variant of the concept by studying the inducibility for rooted binary trees.

A natural related problem for further study is the inducibility of degree-bounded rooted trees. The special case of binary trees is also known as phylogenetic trees—these trees are widely used in biology to describe how species are evolutionarily linked [16]. The present work investigates the concept of inducibility for degree-bounded rooted trees. Among other results, it provides an affirmative answer to a conjecture from [6].

Let $d \geq 2$ be an arbitrary but fixed positive integer. A rooted tree will be called a *d-ary tree* in this paper if each of its non-leaf vertices has between 2 and d children. For a rooted tree D , we write $|D|$ for the number of leaves of D .

If S is a subset of the leaf set of a d -ary tree T , then the unique subtree obtained by

- first extracting the minimal subtree of T containing all leaves in S , and then
- contracting, except possibly for the root, the vertices of degree 2 in the induced tree, which means that the two incident edges of each such vertex are merged and the vertex itself is suppressed,

is called a *leaf-induced subtree* of T ; see Fig. 1 for an illustration. We define the root of the resulting tree as the closest common ancestor of the leaves in S .

By a *copy* of D in T , we mean any leaf-induced subtree of T isomorphic (in the sense of rooted trees) to D . The total number of copies of D in T will be denoted by $c(D, T)$, and the quotient

$$\frac{c(D, T)}{\binom{|T|}{|D|}}$$

by $\gamma(D, T)$. In words, $\gamma(D, T)$ is the proportion of all subsets of $|D|$ leaves of T that induce a copy of D .

We define the *inducibility of a d-ary tree D in d-ary trees T* as being the limit superior, taken over all d -ary trees, of the density of all subsets of $|D|$ leaves of T that induce a copy of D :

$$I_d(D) := \limsup_{\substack{|T| \rightarrow \infty \\ T \text{ d-ary tree}}} \gamma(D, T) = \limsup_{n \rightarrow \infty} \max_{\substack{|T|=n \\ T \text{ d-ary tree}}} \gamma(D, T). \tag{1}$$

This novel concept of inducibility is, of course, in analogy to the very first one [15] for simple graphs. However, it is not clear at this point whether the sequence

$$\left(\max_{\substack{|T|=n \\ T \text{ d-ary tree}}} \gamma(D, T) \right)_{n \geq |D|}$$

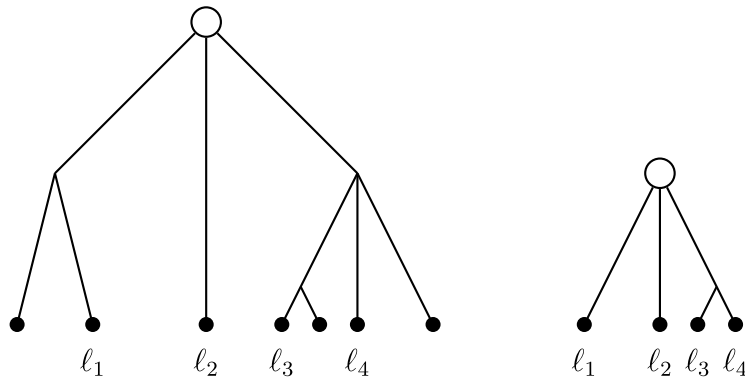


Fig. 1. A ternary tree and the subtree induced by four leaves (l_1, l_2, l_3, l_4) .

converges for every d -ary tree D , in which case we could simply write

$$I_d(D) = \lim_{n \rightarrow \infty} \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T).$$

Notice that Eq. (1) already tells us that the maximum number of copies of D in an arbitrary n -leaf d -ary tree is at most

$$\binom{n}{|D|} (I_d(D) + o(1))$$

as $n \rightarrow \infty$. In fact, we shall prove later that $I_d(D)$ is always a positive real number for every d -ary tree D —see Proposition 8. For the particular case where $d = 2$, it was conjectured in [6] that the asymptotic formula

$$\max_{\substack{|T|=n \\ T \text{ binary tree}}} \gamma(B, T) = I_2(B) + \mathcal{O}(n^{-1}) \tag{2}$$

holds for every binary tree B . In the present work, we affirm this conjecture, and even generalize the result to every $d \geq 2$ (Theorem 3).

A d -ary tree in which every vertex has exactly 0 or d children will be called a *strictly d -ary tree*. It turns out that $I_d(D)$ can also be computed by merely taking the limit superior in the family of strictly d -ary trees. Define the *inducibility of a d -ary tree D in strictly d -ary trees* as being the limit superior, taken over all strictly d -ary trees, of the density of subsets of $|D|$ leaves that induce a copy of D . That is, we have

$$i_d(D) := \limsup_{\substack{|T| \rightarrow \infty \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D, T) = \limsup_{n \rightarrow \infty} \max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D, T).$$

We note that $0 \leq i_d(D) \leq I_d(D) \leq 1$ by definition. We demonstrate in Section 4 that the identity $I_d(D) = i_d(D)$ holds for every d -ary tree D and every d —see Theorem 5. As counterpart of (2), we show in Corollary 6 that

$$\max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D, T) = i_d(D) + \mathcal{O}(n^{-1/2}).$$

In Section 5 we show that the inducibility of any fixed binary tree is the same among binary trees and d -ary trees, for any $d \geq 2$; in Section 6 we show that the inducibility of any d -ary tree is positive; in Section 7 we prove that the only d -ary trees that have inducibility 1 are the binary caterpillars; and finally, we mention in Section 8 further problems on the inducibility of d -ary trees.

From now we always assume that $d \geq 2$ and $|D| \geq 2$.

3. On stars and binary caterpillars

For technical reasons, it is consistent and convenient to treat the single vertex of a one-vertex tree as both leaf and root. By joining leaves to a single vertex, one obtains a *star*. We denote the k -leaf star by S_k as suggested by Fig. 2. Obviously, our definitions bring $I_d(S_2) = i_d(S_2) = 1$ for every d .

A *complete d -ary tree* is a strictly d -ary tree in which all the leaves (d^h in total) are at the same distance h from the root; see Fig. 3 for the complete 4-ary tree of height 2.

For our next theorem, which characterizes the inducibility of stars, we need Muirhead’s inequality [11]: let (l_1, l_2, \dots, l_d) and (m_1, m_2, \dots, m_d) be two vectors of nonnegative real numbers. Assume that $l_1 \geq l_2 \geq \dots \geq l_d$ and $m_1 \geq m_2 \geq \dots \geq m_d$.

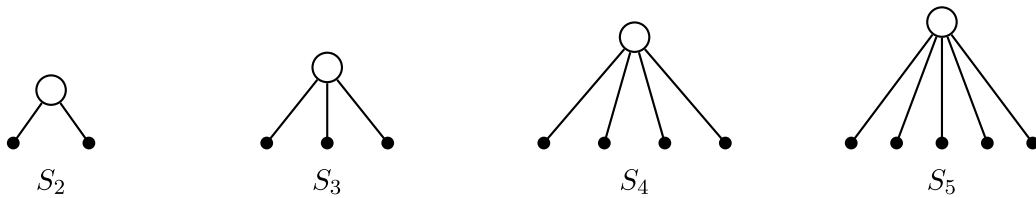


Fig. 2. Stars S_k .

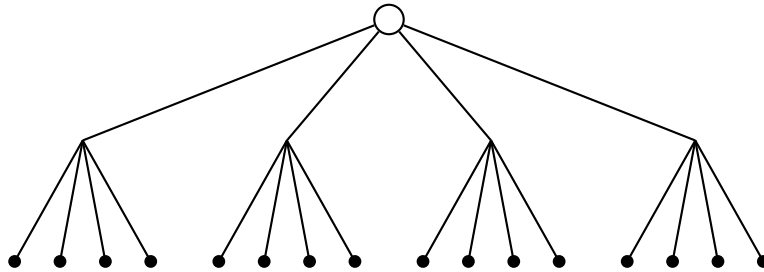


Fig. 3. The complete 4-ary tree of height 2.

m_d in this order. The vector (l_1, l_2, \dots, l_d) majorizes the vector (m_1, m_2, \dots, m_d) if $l_1 + l_2 + \dots + l_d = m_1 + m_2 + \dots + m_d$, and for every $k \in \{1, 2, \dots, d - 1\}$,

$$\sum_{i=1}^k l_i \geq \sum_{i=1}^k m_i .$$

Muirhead's inequality states that

$$\sum_{\pi \in S_d} \prod_{i=1}^d x_{\pi(i)}^{l_i} \geq \sum_{\pi \in S_d} \prod_{i=1}^d x_{\pi(i)}^{m_i}$$

for all nonnegative real numbers x_1, x_2, \dots, x_d whenever the vector (l_1, l_2, \dots, l_d) majorizes the vector (m_1, m_2, \dots, m_d) . The sums on both sides are taken over all permutations of the indices $1, 2, \dots, d$.

Theorem 1. For every fixed positive integer $d \geq 2$ and every $k \in \{2, 3, \dots, d\}$, the inducibility of the k -leaf star S_k in strictly d -ary trees is

$$i_d(S_k) = \frac{d!}{(d - k)!(d^k - d)} .$$

Moreover, $i_d(S_k)$ is an increasing sequence in d for $k \geq 3$, starting with $d = k$.

Proof. Fix $d \geq 2$, and let us prove that for a strictly d -ary tree T with n leaves, we have

$$c(S_k, T) \leq \binom{d}{k} \cdot \frac{n^k - n}{d^k - d}$$

for every $k \geq 2$, with equality if T is a complete d -ary tree.

For $k = 2$, the assertion is immediate as $c(S_2, T) = \binom{d}{2}$. For $k \geq 3$, we use induction on n . The base cases with $n < k$ are trivial as there cannot be any copies of S_k in T . For the induction step, consider the d branches T_1, T_2, \dots, T_d of T with n_1, n_2, \dots, n_d leaves, respectively. So we have $n_1 + n_2 + \dots + n_d = n$. We distinguish possible scenarios that can occur for a subset of k leaves:

- all k leaves belong to the same branch of T . The total number of these subsets of leaves that induce S_k is given by $c(S_k, T_1) + c(S_k, T_2) + \dots + c(S_k, T_d)$,
- one of the branches of T contains more than one of the k leaves, but not all of them. In this case the leaf-induced subtree is not isomorphic to S_k ,
- k of the branches of T , say (without loss of generality) T_1, T_2, \dots, T_k , contain exactly one of the leaves each. In this case, the k leaves always induce S_k , yielding $n_1 \cdot n_2 \cdot \dots \cdot n_k$ copies of S_k .

Therefore, we establish that

$$c(S_k, T) = \sum_{i=1}^d c(S_k, T_i) + \sum_{\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, d\}} \prod_{j=1}^k n_{i_j}.$$

The induction hypothesis gives

$$c(S_k, T) \leq \frac{\binom{d}{k}}{d^k - d} \cdot \sum_{i=1}^d (n_i^k - n_i) + \sum_{\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, d\}} \prod_{j=1}^k n_{i_j}. \tag{3}$$

On the other hand, the multinomial theorem yields the following decomposition:

$$\begin{aligned} n^k &= \left(\sum_{i=1}^d n_i \right)^k = \sum_{\substack{l_1, l_2, \dots, l_d \geq 0 \\ l_1 + l_2 + \dots + l_d = k}} \binom{k}{l_1, l_2, \dots, l_d} \prod_{i=1}^d n_i^{l_i} \\ &= \sum_{\substack{l_1, l_2, \dots, l_d \geq 0 \\ l_1 + l_2 + \dots + l_d = k}} \binom{k}{l_1, l_2, \dots, l_d} \frac{1}{d!} \sum_{\pi \in S_d} \prod_{i=1}^d n_{\pi(i)}^{l_i} \\ &= \sum_{i=1}^d n_i^k + \sum_{\substack{0 \leq l_1, l_2, \dots, l_d < k \\ l_1 + l_2 + \dots + l_d = k}} \binom{k}{l_1, l_2, \dots, l_d} \frac{1}{d!} \sum_{\pi \in S_d} \prod_{i=1}^d n_{\pi(i)}^{l_i}. \end{aligned}$$

Since every vector (l_1, l_2, \dots, l_d) of nonnegative integers with $l_1 + l_2 + \dots + l_d = k$ majorizes the vector $(1, 1, \dots, 1, 0, 0, \dots, 0)$ (k ones, followed by $d - k$ zeros), we can apply Muirhead's inequality to every term in the second sum of this decomposition:

$$\sum_{\pi \in S_d} \prod_{i=1}^d n_{\pi(i)}^{l_i} \geq \sum_{\pi \in S_d} \prod_{i=1}^k n_{\pi(i)} = k!(d - k)! \sum_{\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, d\}} \prod_{j=1}^k n_{i_j}.$$

This gives us

$$\begin{aligned} n^k &\geq \sum_{i=1}^d n_i^k + \sum_{\substack{0 \leq l_1, l_2, \dots, l_d < k \\ l_1 + l_2 + \dots + l_d = k}} \binom{k}{l_1, l_2, \dots, l_d} \frac{k!(d - k)!}{d!} \sum_{\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, d\}} \prod_{j=1}^k n_{i_j} \\ &= \sum_{i=1}^d n_i^k + (d^k - d) \frac{k!(d - k)!}{d!} \sum_{\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, d\}} \prod_{j=1}^k n_{i_j}, \end{aligned}$$

using the multinomial theorem in the opposite direction now. We can rewrite this as

$$\sum_{\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, d\}} \prod_{j=1}^k n_{i_j} \leq \frac{\binom{d}{k}}{d^k - d} \cdot \left(n^k - \sum_{i=1}^d n_i^k \right).$$

Plugging this into (3) yields

$$\begin{aligned} c(S_k, T) &\leq \frac{\binom{d}{k}}{d^k - d} \cdot \sum_{i=1}^d (n_i^k - n_i) + \frac{\binom{d}{k}}{d^k - d} \cdot \left(n^k - \sum_{i=1}^d n_i^k \right) \\ &= \frac{\binom{d}{k}}{d^k - d} (n^k - n), \end{aligned}$$

completing the induction. Furthermore, equality can only arise in this context if

$$n_1 = n_2 = \dots = n_d.$$

Thus the inequality holds with equality if and only if, for every internal vertex v of T , the number of leaves in the d branches of the subtree of T rooted at v are the same: in this case, T is a complete d -ary tree. The assertion on the inducibility follows by passing to the density and taking the limit:

$$i_d(S_k) = \lim_{n \rightarrow \infty} \left\{ \frac{\binom{d}{k} \cdot (n^k - n) / (d^k - d)}{\binom{n}{k}} \right\} = \frac{d!}{(d - k)! \cdot (d^k - d)}.$$

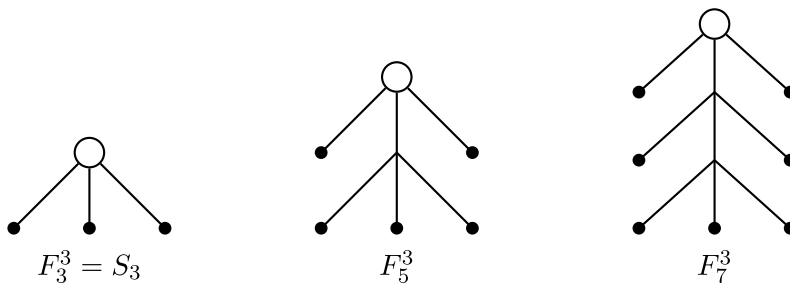


Fig. 4. Ternary caterpillars F_k^3 .

We now turn to the second assertion of the theorem.

Claim: For every given positive integer $k \geq 3$,

$$u_k(x) = \frac{(x - 1)(x - 2) \cdots (x - k + 1)}{x^{k-1} - 1}$$

is an increasing function for $x \geq k$.

Proof of the Claim: Indeed, we have

$$\log u_k(x) = -\log(x^{k-1} - 1) + \sum_{i=1}^{k-1} \log(x - i)$$

so that

$$\begin{aligned} \frac{d}{dx}(\log u_k(x)) &= -\frac{k-1}{x^{k-1}-1} \cdot x^{k-2} + \sum_{i=1}^{k-1} \frac{1}{x-i} \\ &= \sum_{i=1}^{k-1} \left(\frac{1}{x-i} - \frac{x^{k-2}}{x^{k-1}-1} \right) \\ &= \frac{1}{x^{k-1}-1} \cdot \sum_{i=1}^{k-1} \left(\frac{-1+i \cdot x^{k-2}}{x-i} \right) > 0 \end{aligned}$$

as $x \geq k \geq 3$. We conclude that $(i_d(S_k))_{d \geq k}$ is an increasing sequence as soon as $k \geq 3$. \square

By a d -ary caterpillar, we mean a strictly d -ary tree in which every non-leaf vertex has $d - 1$ children that are leaves, except for the lowest, which has d children that are leaves. Note that the non-leaf vertices must lie on a single path starting at the root. We denote the d -ary caterpillar with k leaves by F_k^d (refer to Fig. 4 for ternary caterpillars).

Theorem 2. Let k and $d \geq 2$ be fixed positive integers. For the binary caterpillar F_k^2 , we have

$$\max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \frac{c(F_k^2, T)}{\binom{n}{k}} = 1 - \mathcal{O}(n^{-1}).$$

In particular, $i_d(F_k^2) = I_d(F_k^2) = 1$.

Proof. Fix $d \geq 2$. We use a direct counting argument which gives us a lower bound on the number of copies of F_k^2 in F_n^d . A binary caterpillar can be constructed by attaching exactly one pendant edge to all vertices of a path, except the last one (that is furthest away from the root).

Thus, a copy of F_k^2 in F_n^d can be obtained by first choosing a subset of k vertices from the set of internal vertices of F_n^d , and then for each of them, choosing one of its (at least) $d - 1$ children that are leaves. Since every strictly d -ary tree T has exactly $(|T| - 1)/(d - 1)$ internal vertices, we deduce that there are at least

$$\binom{\frac{n-1}{d-1}}{k} \cdot (d - 1)^k$$

copies of F_k^2 in F_n^d . Therefore, we get

$$c(F_k^2, F_n^d) \geq \binom{\frac{n-1}{d-1}}{k} \cdot (d - 1)^k = \frac{n^k}{k!} - \mathcal{O}(n^{k-1}).$$

Hence, because we have $c(F_k^2, F_n^d) \leq \binom{n}{k} \leq n^k/k!$ by definition, the assertion on the maximum density of F_k^2 in strictly d -ary trees and thus the inducibility $i_d(F_k^2)$ follows. In particular, we obtain $i_d(F_k^2) = I_d(F_k^2) = 1$. \square

It can actually be shown that

$$c(F_k^2, F_n^d) = (d - 1)^{k-1} \binom{\frac{n-1}{d-1}}{k-1} \cdot \frac{2n - (d - 2)(k - 2)}{2k}$$

for $k, n > 1$, but this precision is not required if one is only interested in the inducibility $i_d(F_k^2)$.

4. Inducibility vs. Maximum density

Our first goal in this section is to prove a conjecture from [6] which states that the maximum of $\gamma(D, T)$ over d -ary trees T with n leaves converges fairly quickly to the inducibility $I_d(D)$, namely at a rate of n^{-1} . To be precise, the conjecture was only made in the binary case.

Conjecture 1 ([6]). *For a binary tree B , we have*

$$\max_{\substack{|T|=n \\ T \text{ binary tree}}} \gamma(B, T) = i_2(B) + \mathcal{O}(n^{-1}).$$

We prove:

Theorem 3. *For every fixed positive integer $d \geq 2$ and every d -ary tree D , we have*

$$\max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) = I_d(D) + \mathcal{O}(n^{-1})$$

for all $n \geq |D|$. In particular,

$$\lim_{n \rightarrow \infty} \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) = I_d(D).$$

Moreover, the sequence given by

$$\max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T)$$

is nonincreasing in n .

Proof of Theorem 3. Fix $d \geq 2$. Let D and T be arbitrary d -ary trees such that $|D| \leq |T|$. Since $c(D, T)$ represents the number of subsets of $|D|$ leaves of T that induce a copy of D , we immediately deduce that the $|T|$ leaves of T are contained in $|D| \cdot c(D, T)/|T|$ copies of D on average. Thus, there exist leaves l_1 and l_2 of T satisfying the relation

$$c_{l_1}(D, T) \leq \frac{|D| \cdot c(D, T)}{|T|} \leq c_{l_2}(D, T), \tag{4}$$

where $c_l(D, T)$ stands for the number of subsets of $|D|$ leaves of T that contain l and induce a copy of D .

- We can assume $|T| \geq 2$. The number of subsets of $|D|$ leaves of T not involving the leaf l_1 that induce a tree isomorphic to D is exactly $c(D, T) - c_{l_1}(D, T)$, which is therefore at least $(1 - \frac{|D|}{|T|})c(D, T)$ in view of relation (4). One can construct a new tree T' from T by deleting the leaf l_1 and contracting the vertex adjacent to l_1 if its degree in T' is 2 in the same way as in the construction of leaf-induced subtrees. We have

$$c(D, T') = c(D, T) - c_{l_1}(D, T) \geq \left(1 - \frac{|D|}{|T|}\right)c(D, T),$$

and it follows that

$$\max_{|T|=n-1} c(D, T) \geq \left(1 - \frac{|D|}{n}\right) \max_{|T|=n} c(D, T)$$

as T was taken to be an arbitrary d -ary tree. Hence, by passing to the density by dividing by $\binom{n-1}{|D|}$, we obtain

$$\max_{\substack{|T|=n-1 \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) \geq \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T)$$

for every $n \geq |D| + 1$. So the sequence

$$\left(\max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}} \gamma(D, T) \right)_{n \geq |D|}$$

is nonincreasing and bounded from below, which means that the limit

$$\lim_{n \rightarrow \infty} \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T)$$

exists and is $I_d(D)$.

- Denote by T^+ the d -ary tree obtained by replacing the leaf l_2 by an internal vertex with two leaves l_2, l'_2 attached to it. That way, the number of copies of D in T^+ not involving l_2 is just $c(D, T)$, whereas the number of copies of D in T^+ involving l_2 is no less than the number of copies of D in T involving l_2 . Therefore, by relation (4), the quantity $(1 + \frac{|D|}{|T|})c(D, T)$ provides a lower bound on $c(D, T^+)$:

$$\left(1 + \frac{|D|}{|T|}\right)c(D, T) \leq c(D, T^+).$$

It follows that

$$\left(1 + \frac{|D|}{n}\right) \max_{|T|=n} c(D, T) \leq \max_{|T|=n+1} c(D, T)$$

as T was assumed to be an arbitrary n -leaf d -ary tree. If we divide this inequality by $\binom{n+1}{|D|}$, we obtain

$$\max_{\substack{|T|=n+1 \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) \geq \left(1 - \frac{|D|(|D| - 1)}{n(n + 1)}\right) \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T)$$

for every $n \geq |D|$, and by p -fold iteration

$$\max_{\substack{|T|=n+p \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) \geq \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) \cdot \prod_{j=0}^{p-1} \left(1 - \frac{|D|(|D| - 1)}{(n + p - j)(n + p - j - 1)}\right)$$

for all n, p with $p \geq 1$ and $n \geq |D|$. Now, for $n \geq |D|$ and $p \geq 1$, we have

$$0 \leq \frac{|D|(|D| - 1)}{(n + p - j)(n + p - j - 1)} < 1$$

for every $0 \leq j \leq p - 1$. A simple induction on p yields

$$\prod_{j=0}^{p-1} \left(1 - \frac{|D|(|D| - 1)}{(n + p - j)(n + p - j - 1)}\right) \geq 1 - \sum_{j=0}^{p-1} \frac{|D|(|D| - 1)}{(n + p - j)(n + p - j - 1)}.$$

Letting $p \rightarrow \infty$ instantly gives the estimate

$$I_d(D) \geq \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) \cdot \left(1 - \sum_{i=0}^{\infty} \frac{|D|(|D| - 1)}{(n + i + 1)(n + i)}\right)$$

for every $n \geq |D|$. Since

$$\sum_{i=0}^{\infty} \frac{1}{(n + i + 1)(n + i)} = \sum_{i=0}^{\infty} \left(\frac{1}{n + i} - \frac{1}{n + i + 1}\right) = \frac{1}{n},$$

this shows that

$$I_d(D) \geq \left(1 - \frac{|D|(|D| - 1)}{n}\right) \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T).$$

Now we combine the two contributions to obtain

$$0 \leq \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) - I_d(D) \leq \frac{|D|(|D| - 1)}{n}$$

for every $n \geq |D|$. The desired asymptotic formula follows immediately. \square

We remark that the averaging argument employed in the proof of [Theorem 3](#) does not work for strictly d -ary trees T . For example, deleting one leaf (and the single edge incident to it) from a strictly d -ary tree yields a tree that is no longer strictly d -ary as soon as $d \geq 3$. Moreover, the error term $\mathcal{O}(n^{-1})$ is generally best possible, as e.g. the discussion of the complete binary tree of height 2 (with four leaves) in [\[6\]](#) shows.

Our next focus is to prove that $i_d(D)$ and $I_d(D)$ are always equal for every d -ary tree D and every d .

Lemma 4. For a fixed positive integer k , we have

$$\frac{\binom{n}{k}}{\binom{p+n}{k}} = 1 - \mathcal{O}(p/n)$$

as $p \geq 1$ and $n/p \rightarrow \infty$.

Proof. We have

$$\begin{aligned} \frac{\binom{n}{k}}{\binom{p+n}{k}} &= \frac{n^k \left(1 - \mathcal{O}\left(\frac{1}{n}\right)\right)}{(p+n)^k \left(1 - \mathcal{O}\left(\frac{1}{p+n}\right)\right)} \\ &= \left(1 + \frac{p}{n}\right)^{-k} \left(1 - \mathcal{O}\left(\frac{1}{n}\right)\right) \\ &= 1 - \mathcal{O}\left(\frac{p}{n}\right) \end{aligned}$$

as $p \geq 1$ and $n/p \rightarrow \infty$. \square

Theorem 5. Fix a positive integer $d \geq 2$, and let D be a d -ary tree. For every positive integer $n \equiv 1 \pmod{d-1}$ and every d -ary tree T with $\lfloor n^{\frac{1}{2}} \rfloor$ leaves, there exists a strictly d -ary tree T^* with n leaves such that the asymptotic formula

$$\gamma(D, T) = \gamma(D, T^*) + \mathcal{O}(n^{-1/2})$$

holds as $n \rightarrow \infty$, and the \mathcal{O} -constant depends on d and $|D|$ only. In particular, we have

$$i_d(D) = I_d(D).$$

Proof. Consider an arbitrary d -ary tree T (not necessarily a strictly d -ary tree) such that $|T| = \lfloor n^{\frac{1}{2}} \rfloor = n^{\frac{1}{2}} - \mathcal{O}(1)$. We describe an explicit construction for T^* ; the line of reasoning follows probabilistic ideas.

- For every $r \in \{2, 3, \dots, d-1\}$, add $d-r$ branches of one leaf each to every internal vertex of T whose number of children is r . Call the augmented tree T' , and denote by $\tilde{L}(T')$ the set of the additional leaves added to T to obtain the strictly d -ary tree T' . If $|T|_r$ stands for the number of internal vertices of T with r children, then we have

$$|\tilde{L}(T')| = \sum_{r=2}^{d-1} (d-r)|T|_r < (d-2)|T|, \text{ thus } |\tilde{L}(T')| = \mathcal{O}(n^{\frac{1}{2}}),$$

because it is easy to see that the total number of internal vertices of T is less than its number of leaves. Note that

$$|T'| = |T| + |\tilde{L}(T')| = \Theta(n^{\frac{1}{2}}).$$

- Consider the tree T' . Let m be the greatest positive integer satisfying both $m \equiv 1 \pmod{d-1}$ and $m \leq n^{\frac{1}{2}}$. So we have $\lfloor n^{\frac{1}{2}} \rfloor - (d-2) \leq m \leq \lfloor n^{\frac{1}{2}} \rfloor$. Since it suffices to prove the statement for sufficiently large n , we may assume that $m \geq d$. Choose an arbitrary strictly d -ary tree S with $m - (d-1)$ leaves so that

$$\lfloor n^{\frac{1}{2}} \rfloor + 3 - 2d \leq |S| \leq \lfloor n^{\frac{1}{2}} \rfloor + 1 - d.$$

Append a copy of S to every leaf l of T' that does not belong to $\tilde{L}(T')$ by identifying its root with l . Call the resulting tree T'' . We shall refer to the tree S as a ‘dangling’ tree of T'' . Note that

$$|T''| = |T'| \cdot |S| + |\tilde{L}(T')| = n - \mathcal{O}(n^{\frac{1}{2}})$$

and that $|T''| < n$ in view of the inequalities

$$|T| \leq n^{\frac{1}{2}}, |S| \leq n^{\frac{1}{2}} - (d-1), |\tilde{L}(T')| < (d-1)|T|.$$

- Additionally, pick an arbitrary strictly d -ary tree S_p with

$$1 + n - |T''| = \mathcal{O}(n^{\frac{1}{2}})$$

leaves and append the root of T'' to a leaf of S_p . Denote by T^* the strictly d -ary tree that results from this construction. Note that

$$|T^*| = |T''| + |S_p| - 1 = n.$$

Now, let $1 \leq k \leq |T|$ be an arbitrary but fixed positive integer, and pick k leaves of T^* uniformly at random. The probability that none of the k randomly chosen leaves of T^* lies in $L(T')$ or S_p and no two of them belong to the same dangling tree of T^* is exactly

$$\frac{\binom{|T|}{k}}{\binom{|T^*|}{k}} \cdot |S|^k.$$

In words: since there are exactly $|T|$ dangling trees in T^* , we choose k of them and one leaf from each of the k chosen dangling trees to obtain such a subset of k leaves of T^* .

With Lemma 4 at our disposal, we obtain

$$\begin{aligned} \frac{\binom{|T|}{k}}{\binom{|T^*|}{k}} \cdot |S|^k &= \frac{(|T| \cdot |S|)^k (1 - \mathcal{O}(|T|^{-1}))}{n^k (1 - \mathcal{O}(n^{-1}))} \\ &= 1 - \mathcal{O}(n^{-\frac{1}{2}}) \end{aligned} \tag{5}$$

as $n \rightarrow \infty$. Now let D be a d -ary tree with k leaves. To simplify matters, define the following events: E_1 is the event that the k distinct randomly chosen leaves of T^* induce a copy of D ; E_2 is the event that the k leaves are randomly chosen from distinct dangling trees of T^* .

Note that the tree induced by k leaves of T^* that belong to k distinct dangling trees is equal to the tree induced by the k leaves of T to which these k dangling trees were attached. Hence the probability $P(E_1 \cap E_2)$ that k randomly chosen leaves of T^* are from distinct dangling trees of T^* and induce a copy of D is given by

$$P(E_1 \cap E_2) = P(E_1|E_2) \cdot P(E_2) = \frac{c(D, T)}{\binom{|T|}{|D|}} \cdot \frac{\binom{|T|}{k} \cdot |S|^k}{\binom{|T^*|}{k}} = \frac{c(D, T)}{\binom{n}{k}} \cdot |S|^k$$

(recall that T^* has n leaves). From this observation and the fact that $\gamma(D, T^*)$ is precisely the probability $P(E_1)$ that k distinct randomly chosen leaves of T^* induce a copy of D , we deduce that

$$\begin{aligned} \gamma(D, T^*) &= P(E_1) = P(E_1 \cap E_2) + P(E_1|E_2^c)(1 - P(E_2)) \\ &= \frac{c(D, T)}{\binom{n}{k}} \cdot |S|^k + P(E_1|E_2^c) \cdot \left(1 - \frac{\binom{|T|}{k} \cdot |S|^k}{\binom{n}{k}}\right) \end{aligned}$$

by virtue of the law of total probability, where $P(E_1|E_2^c)$ stands for the probability that k distinct leaves of T^* induce a copy of D under the condition that the event E_2 has not occurred.

This implies that

$$\gamma(D, T^*) = \frac{c(D, T)}{\binom{|T|}{k}} + \left(1 - \frac{\binom{|T|}{k} \cdot |S|^k}{\binom{n}{k}}\right) \cdot \left(P(E_1|E_2^c) - \frac{c(D, T)}{\binom{|T|}{k}}\right),$$

and since $P(E_1|E_2^c)$ and $\gamma(D, T) = c(D, T)/\binom{|T|}{k}$ are both between 0 and 1, it follows from the asymptotic formula (5) that

$$\gamma(D, T^*) - \gamma(D, T) = \mathcal{O}(n^{-1/2}).$$

This finishes the proof of the first part of the theorem.

Finally, the immediate consequence we obtain is that

$$I_d(D) = \limsup_{\substack{|T| \rightarrow \infty \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) = \limsup_{|T^*| \rightarrow \infty} \gamma(D, T^*) \leq \limsup_{\substack{|T| \rightarrow \infty \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D, T) = i_d(D).$$

In other words, this shows that $I_d(D) \leq i_d(D)$. Thus, the proof of the second part of the theorem is completed as well because we have $I_d(D) \geq i_d(D)$ by definition. \square

With Theorem 5 and its proof at hand, we can now prove an analogue of Theorem 3 for the maximum density in strictly d -ary trees.

Corollary 6. For every fixed positive integer $d \geq 2$ and every d -ary tree D , we have

$$\max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D, T) = i_d(D) + \mathcal{O}(n^{-1/2})$$

for all $n \geq |D|$. In particular, we have

$$\lim_{n \rightarrow \infty} \max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D, T) = i_d(D).$$

Proof. Using the identity $i_d(D) = I_d(D)$ from [Theorem 5](#) together with the first claim of [Theorem 3](#), we immediately obtain

$$\max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D, T) \leq \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) = i_d(D) + \mathcal{O}(n^{-1}). \tag{6}$$

Also, the second part of [Theorem 3](#) gives

$$i_d(D) = I_d(D) \leq \max_{\substack{|T|=\lfloor n^{1/2} \rfloor \\ T \text{ } d\text{-ary tree}}} \gamma(D, T), \tag{7}$$

while the first claim of [Theorem 5](#) guarantees that we can construct a strictly d -ary tree T^* with n leaves from the tree T_{\max} that attains the maximum in (7), such that

$$\gamma(D, T_{\max}) = \gamma(D, T^*) + \mathcal{O}(n^{-\frac{1}{2}}). \tag{8}$$

Formulae (7) and (8) immediately give

$$i_d(D) \leq \gamma(D, T^*) + \mathcal{O}(n^{-\frac{1}{2}}) \leq \max_{\substack{|T|=n \\ T \text{ strictly } d\text{-ary tree}}} \gamma(D, T) + \mathcal{O}(n^{-\frac{1}{2}}),$$

which, together with (6), completes the proof of the corollary. \square

In certain cases, we suspect a stronger asymptotic result on the maximum density in strictly d -ary trees [8]. This, in particular, happens when

$$\max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} c(D, T)$$

is attained by strictly d -ary trees for every $n \equiv 1 \pmod{d-1}$. Finally, a natural question to ask at this point is the following:

Question 1. Can the \mathcal{O} -term in [Corollary 6](#) be improved further for general d -ary trees D ?

5. Inducibility of a binary tree in d -ary trees

Our aim in this section is to compare the inducibilities of a binary tree B in d -ary trees, for different values of d . It turns out that $i_d(B)$ is independent of d .

Theorem 7. Every binary tree B satisfies

$$i_d(B) = i_2(B) = I_d(B) = I_2(B)$$

for every d .

Proof. The following correspondence is our stepping stone for proving the assertion. Let $d \geq 3$ be fixed, and fix an arbitrary total order $<$ on the set of strictly d -ary trees. For a strictly d -ary tree T , we always order the branches T_1, T_2, \dots, T_d in such a way that

$$T_1 \leq T_2 \leq \dots \leq T_d.$$

From the tree T , we build a binary tree $G(T)$ by means of the recursive algorithm depicted in [Fig. 5](#), starting with the single leaf being invariant under G . More specifically, for $|T| > 1$, the tree $G(T)$ is obtained as follows:

- draw a path on $d - 1$ vertices v_2, v_3, \dots, v_d in this order (v_2 and v_d are the endvertices of the path);
- attach a leaf l_i (by dropping a pendant edge) to every vertex v_i of the path except for the lowest vertex v_2 ;
- replace the vertex v_2 of the path by an internal vertex with two leaves l_1 and l_2 attached to it;
- append the root of $G(T_i)$ to leaf l_i for every $i \in \{1, 2, \dots, d\}$. The vertex v_d is the root of $G(T)$.

For example, G maps the star S_d to the binary caterpillar F_d^2 . With this transformation at hand, let us prove that the inequality

$$c(B, G(T)) \geq c(B, T)$$

holds for every binary tree B and every strictly d -ary tree T .

The construction of $G(T)$ also yields a natural bijection between the leaves of T and those of $G(T)$. We can show by induction that if a set of leaves of T induces a copy of a binary tree B , then so do the corresponding leaves in $G(T)$. This is clearly true when $|T| = 1$. For the induction step, we have two cases:

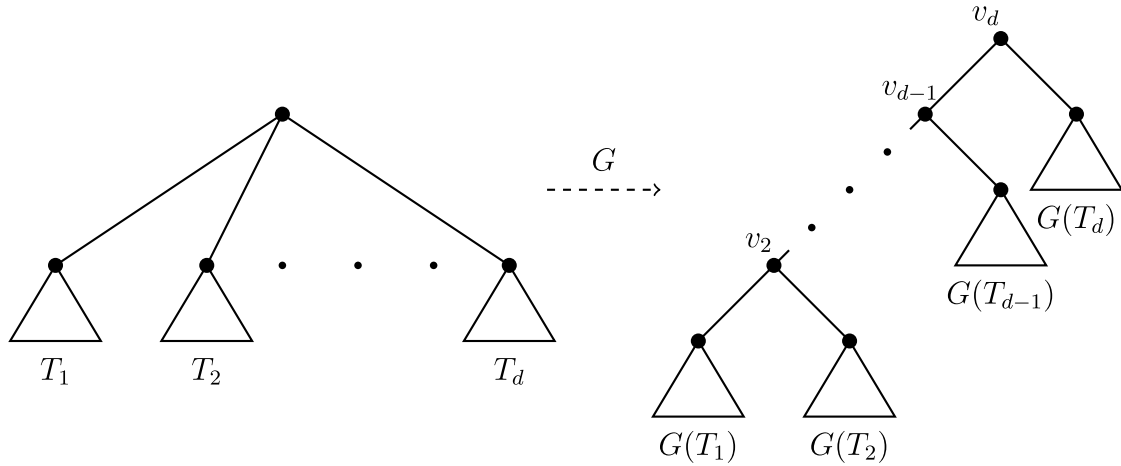


Fig. 5. T is a strictly d -ary tree, and $G(T)$ is its corresponding binary tree under the tree transformation G described in the proof of Theorem 7.

- If the leaves that induce B all lie in one branch T_i for some $i \in \{1, 2, \dots, d\}$, then the corresponding leaves lie in $G(T_i)$, and we are done by the induction hypothesis.
- Otherwise, they lie in exactly two branches T_i and T_j for some $i \neq j \in \{1, 2, \dots, d\}$, and the two leaf sets induce the two branches of B . By the induction hypothesis, this is also true for the corresponding leaves in $G(T_i)$ and $G(T_j)$, and we are done again.

This shows that every subset of leaves of T that induces a copy of B corresponds to a unique subset of leaves of $G(T)$ that induces a copy of B . Therefore, there is an injection from the copies of B in T to the copies of B in $G(T)$, and we have $c(B, T) \leq c(B, G(T))$ as claimed. Consequently, we arrive at

$$i_d(B) \leq \limsup_{\substack{|T| \rightarrow \infty \\ T \text{ strictly } d\text{-ary tree}}} \gamma(B, G(T)) \leq \limsup_{\substack{|T| \rightarrow \infty \\ T \text{ binary tree}}} \gamma(B, T) = i_2(B).$$

On the other hand, we have both $i_2(B) = I_2(B)$ and $I_d(B) \geq I_2(B)$ by definition, while Corollary 6 gives us $i_d(B) = I_d(B)$. Hence, the statement of the theorem follows. \square

We point out that the analogue of Theorem 7 for non-binary trees is not true in general: as Theorem 1 shows, the inducibility $i_d(S_k)$ of the k -leaf star is a strictly increasing function of d for $k \geq 3$.

From here onwards, we shall use only $i_d(D)$. Our next section deals with some bounds on the inducibility.

6. Some general results

Let us say something about how small $i_d(D)$ can be:

Proposition 8. *Let $d \geq 2$ be an arbitrary but fixed positive integer. Every d -ary tree D with at least two leaves satisfies*

$$i_d(D) \geq \frac{(|D| - 1)!}{|D|^{|D|-1} - 1}.$$

In particular, every d -ary tree has positive inducibility.

Proof. We employ a tree construction similar to the one used for proving that $i_d(D) = I_d(D)$. Fix $d \geq 2$. For two strictly d -ary trees S_1, S_2 , denote by $\mathcal{F}(S_1; S_2)$ the unique strictly d -ary tree that is formed by appending the root of S_2 to every leaf of S_1 . See Fig. 6 for a picture. So we have $|\mathcal{F}(S_1; S_2)| = |S_1| \cdot |S_2|$.

We define a sequence of strictly d -ary trees $T^{[n]}$ as follows: $T^{[1]}$ is just a single vertex, and $T^{[n+1]} = \mathcal{F}(D; T^{[n]})$ for $n \geq 1$. It is clear that $T^{[2]} = D$ and $|T^{[n]}| = |D|^{n-1}$. In this fashion, there are $|D| \cdot c(D, T^{[n]})$ copies of D in the $|D|$ copies of $T^{[n]}$ that are attached to the leaves of D to obtain $T^{[n+1]}$. Also, if one takes one arbitrary leaf from each of these $|D|$ copies, one obtains a total of another $|T^{[n]}|^{|D|}$ copies of D . Thus, we get

$$c(D, T^{[1+n]}) \geq |D| \cdot c(D, T^{[n]}) + |T^{[n]}|^{|D|},$$

which gives

$$c(D, T^{[1+n]}) \geq \sum_{j=0}^{n-1} |D|^j \cdot |T^{[n-j]}|^{|D|}$$

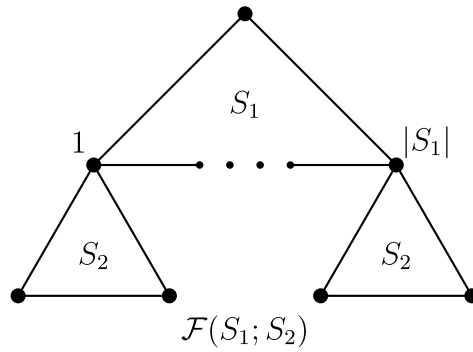


Fig. 6. A rough picture of the tree $\mathcal{F}(S_1; S_2)$ defined in the proof of Proposition 8.

by iteration as $c(D, T^{[1]}) = 0$. Since $|T^{[n-j]}| = |D|^{n-j-1}$, we establish the inequality

$$c(D, T^{[n]}) \geq |D|^{(n-2)|D|} \cdot \sum_{j=0}^{n-2} |D|^{j(1-|D|)},$$

which becomes

$$\gamma(D, T^{[n]}) \geq \frac{|D|^{(n-1)|D|} - |D|^{n-1}}{(|D|^{|D|} - |D|) \binom{|D|^{n-1}}{|D|}}.$$

Letting $n \rightarrow \infty$ produces

$$\limsup_{n \rightarrow \infty} \gamma(D, T^{[n]}) \geq \frac{(|D| - 1)!}{|D|^{|D|-1} - 1}$$

and, in particular, the assertion of the proposition. \square

The bound obtained in Proposition 8 seems to be weak in general. Possibly, S_d is even the only d -ary tree that attains this bound (see Theorem 1).

The tree $\mathcal{F}(S_1; S_2)$ (as constructed in the proof of Proposition 8) offers another special result:

Theorem 9. Let $d \geq 2$ be an arbitrary but fixed positive integer. For two d -ary trees S_1, S_2 , let $\mathcal{F}(S_1; S_2)$ be as defined in the proof of Proposition 8. Then we have

$$i_d(\mathcal{F}(S_1; S_2)) \geq \frac{(|S_1| \cdot |S_2|)!}{(|S_2|!)^{|S_1|} \cdot |S_1|^{|S_1| \cdot |S_2|}} \cdot (i_d(S_2))^{|S_1|}.$$

Proof. Fix $d \geq 2$. For a d -ary tree T , it is easy to see that a copy of $\mathcal{F}(S_1; S_2)$ in $\mathcal{F}(S_1; T)$ can be obtained by taking one copy of S_2 in T for each of the $|S_1|$ copies of T planted in $\mathcal{F}(S_1; T)$. Thus, the inequality

$$c(\mathcal{F}(S_1; S_2), \mathcal{F}(S_1; T)) \geq c(S_2, T)^{|S_1|}$$

is valid so that we obtain

$$\max_{\substack{|T'|=|S_1| \cdot |T| \\ T' \text{ } d\text{-ary tree}}} \frac{c(\mathcal{F}(S_1; S_2), T')}{\binom{|S_1| \cdot |T|}{|S_1| \cdot |S_2|}} \geq \frac{\binom{|T|}{|S_2|}^{|S_1|}}{\binom{|S_1| \cdot |T|}{|S_1| \cdot |S_2|}} \cdot \gamma(S_2, T)^{|S_1|}.$$

Now we choose T in such a way that

$$|T| = n, \quad c(S_2, T) = \max_{\substack{|T'|=n \\ T' \text{ } d\text{-ary tree}}} c(S_2, T').$$

Letting $n \rightarrow \infty$ yields

$$\limsup_{n \rightarrow \infty} \max_{\substack{|T'|=|S_1| \cdot n \\ T' \text{ } d\text{-ary tree}}} \gamma(\mathcal{F}(S_1; S_2), T') \geq \frac{(|S_1| \cdot |S_2|)!}{(|S_2|!)^{|S_1|} \cdot |S_1|^{|S_1| \cdot |S_2|}} \cdot (i_d(S_2))^{|S_1|}.$$

Consequently, we obtain the desired inequality. \square

In some special cases, there is an upper bound that asymptotically matches the lower bound on $i_d(\mathcal{F}(S_1; S_2))$ as $|\mathcal{F}(S_1; S_2)|$ gets large—see [8].

7. The trees with the maximal inducibility

The first natural question to pose regarding a graph invariant concerns its extreme values and the extremal graphs. In our case, we have already proved that every tree has positive inducibility (Proposition 8) and that $i_d(F_k^2) = 1$ for every k and every d (Theorem 2). But besides binary caterpillars, are there other d -ary trees with inducibility 1? The answer turns out to be negative. This extends the result for binary trees in [6].

Theorem 10. *Let $d \geq 2$ be an arbitrary but fixed positive integer. Among all d -ary trees, only binary caterpillars have (d -ary) inducibility 1.*

Proof. Fix $d \geq 2$ and let us prove that for any fixed positive integers $k > 1$ and $n > d^{k-2}$, every d -ary tree with n leaves contains a copy of the k -leaf binary caterpillar F_k^2 .

The key observation is that if a d -ary tree has height at least $k - 1$, then it must have F_k^2 as a leaf-induced subtree. So fix $k > 1$, $n > d^{k-2}$, and consider a d -ary tree T with n leaves. It is easy to see by induction on h that for every fixed integer $h \geq 0$, a d -ary tree with height at most h can never have more than d^h leaves. Since T has more than d^{k-2} leaves, its height must be at least $k - 1$, so it contains F_k^2 . In conclusion, for $k > 1$, every n -leaf d -ary tree must contain a copy of the binary caterpillar F_k^2 as soon as $n > d^{k-2}$.

Now, consider a d -ary tree D that has at least three leaves. If D is different from $F_{|D|}^2$, then we obtain $c(D, T) < \binom{|T|}{|D|}$ for every d -ary tree T with n leaves provided that $n > d^{k-2}$. Therefore, fixing $n > d^{k-2}$, we get

$$\max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T) < 1.$$

On the other hand, from the second part of Theorem 3, we have

$$i_d(D) \leq \max_{\substack{|T|=n \\ T \text{ } d\text{-ary tree}}} \gamma(D, T),$$

which now completes the proof of the theorem. \square

As one might expect, the limit

$$\liminf_{\substack{|T| \rightarrow \infty \\ T \text{ } d\text{-ary tree}}} \gamma(F_k^2, T)$$

is positive for every d and every k . An exact formula is available in [7].

8. Concluding comments

It seems appropriate to close with further problems on the inducibility of d -ary trees. Since we already have a general lower bound on $i_d(D)$, it would be nice to have a better understanding of the following general problem:

Problem 1. Given positive integers $d \geq 2$ and $k \geq 5$, find

$$\min_{\substack{|D|=k \\ D \text{ } d\text{-ary tree}}} i_d(D),$$

and furthermore, characterize the d -ary trees that attain this minimum.

We conjecture that except for binary trees, the inducibility in d -ary trees always depends on d (in contrast to binary trees, where it does not; see Theorem 7 and the discussion thereafter).

Conjecture 2. *Let $d \geq 2$ be an arbitrary but fixed positive integer. Among all d -ary trees D , only binary trees satisfy*

$$i_d(D) = i_{d+1}(D) = i_{d+2}(D) = \dots$$

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

References

- [1] József Balogh, Ping Hu, Bernard Lidický, Florian Pfender, Maximum density of induced 5-cycle is achieved by an iterated blow-up of 5-cycle, *European J. Combin.* 52 (2016) 47–58.
- [2] Béla Bollobás, Yoshimi Egawa, Andrew Harris, Guoping Jin, The maximal number of induced r -partite subgraphs, *Graphs Combin.* 11 (1) (1995) 1–19.
- [3] Béla Bollobás, Chiê Nara, Shun-ichi Tachibana, The maximal number of induced complete bipartite graphs, *Discrete Math.* 62 (3) (1986) 271–275.
- [4] Jason I. Brown, Alexander Sidorenko, The inducibility of complete bipartite graphs, *J. Graph Theory* 18 (6) (1994) 629–645.
- [5] Sébastien Bubeck, Nati Linial, On the local profiles of trees, *J. Graph Theory* 81 (2) (2016) 109–119.
- [6] Éva Czabarka, László A. Székely, Stephan Wagner, Inducibility in binary trees and crossings in random tanglegrams, *SIAM J. Discrete Math.* 31 (3) (2017) 1732–1750.
- [7] Audace A.V. Dossou-Olory, The minimum asymptotic density of binary caterpillars, *Graphs Combin.* 35 (1) (2019) 303–320.
- [8] Audace A.V. Dossou-Olory, Stephan Wagner, Further results on the inducibility of d -ary trees, 2018, ArXiv preprint, arXiv:1811.11235.
- [9] Chaim Even-Zohar, Nati Linial, A note on the inducibility of 4-vertex graphs, *Graphs Combin.* 31 (5) (2015) 1367–1380.
- [10] Geoffrey Exoo, Dense packings of induced subgraphs, *Ars Combin.* 22 (1986) 5–10.
- [11] Godfrey H. Hardy, John E. Littlewood, George Pólya, *Inequalities*, Cambridge University Press, Cambridge, 1952.
- [12] Hamed Hatami, James Hirst, Serguei Norine, The inducibility of blow-up graphs, *J. Combin. Theory Ser. B* 109 (2014) 196–212.
- [13] James Hirst, The inducibility of graphs on four vertices, *J. Graph Theory* 75 (3) (2014) 231–243.
- [14] Hao Huang, On the maximum induced density of directed stars and related problems, *SIAM J. Discrete Math.* 28 (1) (2014) 92–98.
- [15] Nicholas Pippenger, Martin C. Golumbic, The inducibility of graphs, *J. Combin. Theory Ser. B* 19 (3) (1975) 189–203.
- [16] Charles Semple, Mike A. Steel, *Phylogenetics*, Oxford University Press, 2003.
- [17] Konrad Sperfeld, The inducibility of small oriented graphs, 2011, ArXiv preprint, arXiv:1111.4813.