Universitat Politècnica de Catalunya
Facultat de Matemàtiques i Estadística

Master in Advanced Mathematics and Mathematical Engineering

Master's thesis

Study of balance and symmetry of rooted trees

Arnau Mir Fuentes

Supervised by (Oriol Serra Albo)

September, 2020

Abstract

This work is designed to be understood within the framework of the quantitative graph theory. First, we introduce what we will later know as rooted trees, which will be our object of study for almost all the work. We are going to find a necessary and sufficient condition to know from the distances between the leaves of a rooted tree to the root if a rooted tree can be constructed as graph. Then, we will define and study the balance of a rooted tree, not only giving a formal definition that fits with our intuition of what is a balanced tree, but also we will study a very well-known index that measures the balance of tree, the Sackin Index. Finally, we will also define and study a measure of symmetry of binary rooted trees. Again we not only give a formal definition that tries to fit with our intuition of symmetry, we also define a set of indexes that try to measure the symmetry of a given binary tree.

Keywords

Balance index, Philogenetic tree, Sackin index, Symmetry index

1. Introduction

Quantitative graph theory [DES15, DESS17] is a relatively recent branch of graph theory with applications for solving problems in various disciplines such as biology, computer science, and chemistry among others. Its purpose is to quantify the structural properties of graphs by numerical invariants, rather than the classical approach of characterizing these properties by descriptions. Some active research lines in quantitative graph theory are:

- Find invariants of graphs from measurable features of the graphs to characterize their topology.
- Define distances and measures of similarity between graphs.
- Define statistical parameters associated with the characteristics of a graph from which distributions can be studied under different probabilistic models.

The impetus of quantitative graph theory has come mainly from its applications in the description, first, of the small graphs associated with molecules (with the nodes atoms and the edges the chemical bonds) and, more recently, of the huge graphs representing complex networks. In this work we focus on the study of rooted trees, which have a particular interest in computer science (see Chapter 2.3 of [Knu97]) and in phylogenetics, which is our motivation to study them. In the latter field, rooted trees are used to model, in the form of *phylogenetic trees*, the evolutionary processes of species. An ingrained phylogenetic tree represents a hypothetical evolutionary history of a set of species, located on the leaves of a tree, from a common ancestor represented by the root. In a phylogenetic tree, the edges represent the direct offspring by mutations, the interior nodes represent the different intermediate ancestors of the species we are interested in, and the evolutionary time runs from the root to the leaves.

The interest in the applications of quantitative graph theory to rooted trees in phylogenetics comes from the belief that the shape of a phylogenetic tree is a reflection of the characteristics of the underlying evolutionary process. This motivates the interest in quantifying the properties of this form by means of indices. One of the properties of a phylogenetic tree that is quantified is its balance, using the so-called *balance indices*, which measure the tendency in a tree rooted to its nodes are balanced, in the sense that, for each node, all their "children" (their direct descendants) have the same number of descendant leaves. If a tree is out of balance, it means that evolutionary processes have a tendency to occur more in some evolutionary lines than in others, and it is interesting to see if this has happened afterwards to analyze why.

Some defined indexes for rooted trees are:

- For binary trees
  - Colless index [Col82]
  - the number of "cherries" (rooted subtrees formed by a root and two leaves)
- For rooted trees in general:
  - Sackin index [Sac72, SS90]
  - Cophenetic index [MRR13]
  - generalizated Colless index [MRR18]
  - The index of the rooted quartets [CMR19]

For more balance indexes, see the section "Measures of overall asymmetry" by [Fel04] (pp. 562–563).

One of the first balance indices defined on rooted trees was what was later called the Sackin Index, proposed by Sackin in 1972 in [Sac72]. If we define the leaf depth of a rooted tree as the number of edges in the path from root to leaf, the *Sackin index* of a phylogenetic tree is the sum of the depths of all its leaves. Sackin also proposed [Sac72] to measure the balance of a rooted tree by means of the variance of leaf depths, but this idea did not prosper and instead used the average depth (which is not nothing but what we now call the Sackin Index divided by the number of leaves). Sackin showed in [Sac72] that in some binary trees the leaf depth vector characterizes the tree and showed by examples that the more unbalanced the tree, the greater this sum of depths. In particular, he stated without proof that the maximum value for a fixed number of leaves n is given to trees called *caterpillars*, or also *combs*: binary trees where each interior node has a son that is a leaf.

This statement has been maintained in many other subsequent articles, without anyone showing it until recently, we assume that because everyone believed that someone else had already shown it. Regarding trees with a minimum Sackin index, it has also been repeatedly stated that if the tree does not have to be binary, then this minimum is given to trees with a depth of 1 (consisting of a root and n leaves and nothing more) and if the tree must be binary and n is a power of 2, then the minimum is given to completely symmetrical trees, where the two trees hanging from each interior node are always isomorphic. Again, no one bothered to show this. In addition, the case of binary trees was unknown, with n not being a power of 2. This is a problem, because it is necessary to have formulas that calculate the maximum and minimum values of an balance index for each number of leaves n to normalize this index between 0 and 1 (subtracting the minimum value and dividing by the least minus the minimum), which allows to significantly compare the balance of trees with different numbers of leaves. [SS90]

Two years ago, M. Fischer [Fis18] solved all of these problems, demonstrating the results that until now had been accepted as true by characterizing the binary trees with the least Sackin index for each number n leaves. On the other hand, also last year M. Khatibi and A. Behtoi [KB18] solved the problem of finding the trees with minimum and maximum Sackin index among the rooted trees with a total of n nodes, between leaves and interior nodes, and output degree of all nodes bounded above by some number k, but allowing elementary nodes, that is, nodes with output degree 1 (which are often forbidden in phylogenetic trees), which greatly simplifies the problem.

In this work, we use a similar technique used by Fischer to attack a generalization of the problems solved by finding these extremal trees with all interior nodes that can have any output degree (prohibiting elementary interior nodes), which we will call rooted trees. In addition, so far only attempts have been made to find these minimum and maximum values in the Sackin Index by setting the number of leaves or the order. In this work, we have also raised the issue by setting the number of leaves and also the maximum depth a leaf can have in the tree.

As we have mentioned before, the symmetry have a very strong relation with the balance of a tree. In this work, we define the concept of symmetry in a very similar way that we try to understand the balance, and we try to design indexes to determine how symmetric is a tree. With symmetry, we try to define it as local property of an interior node, this way, the number and the positions of this symmetric interior nodes can help us to understand their symmetry.

The contributions of this work are divided into three fields:

Results on existence of trees. In this section we give a necessary and sufficient condition that allows us to decide whether, given a table of absolute frequencies of depths, there exists a rooted tree where leaf depths occur at these frequencies.

Extreme Sackin Index values for rooted trees. In this section we find the extreme values (maximum and minimum) of the Sackin index for rooted with a fixed number of leaves n and the trees where these values are reached maximum and minimum. In addition, we characterize the numbers that are Sackin indices of rooted trees of n leaves. Also, we find the trees that have minimum and maximum Sackin index in the set of rooted trees that have a fixed leaf number n and a maximum leaf depth set δ. In addition, we give methods for calculating the corresponding minimum and maximum values.

Brief Study of the symmetry of trees In this section, we define what is a symmetric node and we define some indices to calculate the symmetry of a tree. Also, we calculate the extremal values of one of them.

2. Rooted trees and extremal values of the Sackin Index

2.1 Some basic definitions and notation

First of all, Let us define some concepts and give some notation that we will use throughout the Chapter:

Definition 2.1. A rooted tree is a connected graph with no cycles with a distinguished node r, called the root. We denote $T_{n,r}$ a rooted tree at r with n leaves.

Definition 2.2. Let $T_{n,r}$ be a rooted tree, let $L(T_{n,r}) = \{l_1, \dots, l_n\}$ be the set of leaves and $V(T_{n,r})$ be the set of nodes. For each vertex $v \in V(T)$ we denote by

$$\delta(v) = d(r, v),$$

the distance in T from the root r to v. We also will call $\delta(v)$ the depth of v.

We denote by $\delta(T_{n,r})$, or simply δ if $T_{n,r}$ is clear from the context, the maximum depth of the leaves of the tree $T_{n,r}$.

We denote by $\mathcal{T}_{n,r}$ the set of all rooted trees with root r that have n leaves and no interior vertex of degree two. Here an interior vertex is a vertex different from the root and not a leave.

As we mentioned in the introduction, the above is a natural condition in the context of the phylogenetics trees, acondition the complexity of the problems we will adress, some of them becomng trivial without this condition

Definition 2.3. Let $T_{n,r}$ be a rooted tree of depth $\delta = \delta(T_{n,r})$. We denote by

$$S_k = S_k(T_{n,r}) = \{l \in L(T_{n,r}),\ \delta(l) = k\}$$

the set of leaves l such that $\delta(l) = k$.

The depth frequency vector of $T_{n,r}$ is $v = (v_1, \dots, v_\delta) \in \mathbb{N}^\delta$ where $v_i = |S_i|$, $1 \leq i \leq \delta$.

Definition 2.4. Let $T_{n,r}$ be a rooted tree with n leaves and let $v \in V(T_{n,r})$. We denote by T_v be the subtree rooted at v formed by all descendants of v.

We give the following easy Lemma for further use.

Lemma 2.5. *Let $T \in \mathcal{T}_n$. For each $i \in \{0, 1, \dots, \delta(T) - 1\}$ there is a vertex $v \in V(T) \setminus L(T)$ such that $\delta(v) = i$.*

Proof. Let $w \in L(T)$ such that $\delta(w) = \delta$, and consider the path

$$w \leftarrow w_{\delta-1} \leftarrow \dots \leftarrow w_i \leftarrow \dots \leftarrow w_0 = r,$$

where r is the root of T. Notice that by definition of tree there exists only one path between two nodes, therefore every node w_i is an interior node of depth i where $i \in \{0, 1, \dots, \delta - 1\}$ □

2.2 Constructible depth sequences

In this section we give a criterion to decide if a given depth frequency vector can be realized by some tree. The condition that there are no interior vertices of degree two prevents an arbitrary vector to be the depth frequency vector of a rooted tree.

Definition 2.6. Let $v = (v_1, \dots, v_\delta) \in \mathbb{N}^\delta$, we will say that v is constructible if and only if there exists a tree $T \in \mathcal{T}_{n,r}$ with $n = \sum_{i=1}^{\delta} v_i$, such that $\forall i = 1, \dots, \delta$, $|S_i| = v_i$ and if $i > \delta$ $|S_i| = 0$. In this case, we call v the depth frequency vector of T.

We start by a Lemma.

Lemma 2.7. *A vector $v = (v_1, \dots, v_\delta) \in \mathbb{N}^\delta$ is constructible if and only if $v_\delta \geq 2$ and $v' = (v_1, \dots, v_{\delta-1} + \lfloor \frac{v_\delta}{2} \rfloor) \in \mathbb{N}^{\delta-1}$ is constructible.*

Proof. Suppose that v is a constructible vector. Since there are no elementary interior nodes of degree 2 we must have $v_\delta \geq 2$.

Let T be a tree whose leaves have v as the distribution of depths of its leaves. Consider the tree T' obtained from T by deleting all the leaves of depth δ. The distribution of the depths of the leaves of T' is the vector $v' = (v_1, \dots, v_{\delta-1} + k)$, where k is the number of interior nodes of depth $\delta - 1$ in T. Notice that $k \leq \lfloor \frac{v_\delta}{2} \rfloor$ because if $k > \lfloor \frac{v_\delta}{2} \rfloor$, we would have an elementary node of degree two. By adding $\lfloor \frac{v_\delta}{2} \rfloor - k$ leaves of depth $\delta - 1$ at a node of depth $\delta - 2$ we obtained a tree such that its vector of depths is $v' = (v_1, \dots, v_{\delta-1} + \lfloor \frac{v_\delta}{2} \rfloor)$, see figure 1.

Reciprocally, let T' be a tree with vector of depths $v' = (v_1, \dots, v_{\delta-1} + \lfloor \frac{v_\delta}{2} \rfloor)$. In order to construct a tree T with distribution of depths of its leaves $v = (v_1, \dots, v_\delta)$, consider a set L' of $\lfloor \frac{v_\delta}{2} \rfloor$ leaves of depth $\delta - 1$ of T'. If v_δ is even we attach two leaves at depth δ to every leaf in L', and if v_δ is odd we add to the above one additional leaf to one of the nodes in L'.

In this way, we have constructed a tree whose vector of depths of their leaves is $v = (v_1, \dots, v_\delta)$. $\square$

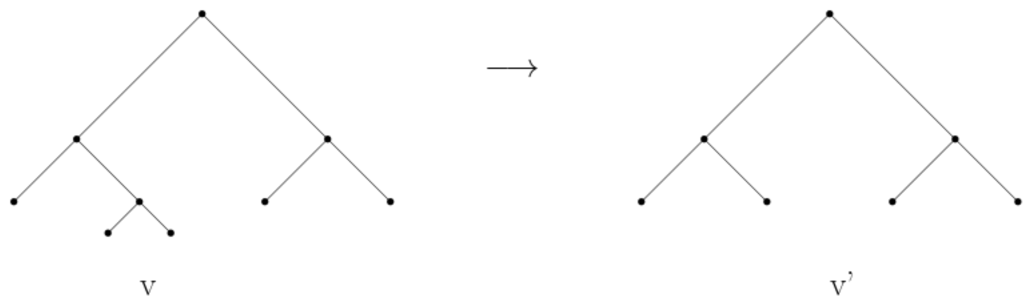


Figure 1: Construction of the tree T' with vector of depths v' from the tree T with vector of depths v in lemma 2.7

Now we can state the theorem that characterizes the constructible vectors.

Theorem 2.8. *Let δ be a positive integer. A vector $v = (v_1, \dots, v_\delta) \in \mathbb{N}^\delta$ is constructible if and only if*

$$v_\delta \geq 2,$$

and, if $\delta \geq 2$

$$v_i + p_{i+1} \geq 2, \text{ for each} i = 1, \ldots, \delta - 1,$$

where $p_\delta = v_\delta$ *and* $p_i = v_i + \lfloor \frac{p_{i+1}}{2} \rfloor$ *for* $1 \leq i < \delta$.

Proof. The proof is by induction on δ. For $\delta = 1$ every vector with a coordinate at least 2 is constructible, the star with v_δ leaves rooted at the center being a realization of the sequence. For $\delta > 1$ the induction step follows from Lemma 2.7. □

We note that every vector $v = (v_1, \ldots, v_\delta) \in \mathbb{N}^\delta$ with $v_\delta \geq 2$ an positive entries is the depth frequency vector of a rooted tree. A realization of the vector can be build from a path $P = \{x_0, x_1, \ldots, x_{\delta-1}\}$ by attaching v_i leaves to the vertex x_{i-1}. The condition $v_\delta \geq 2$ arises from our condition that no interior vertices have degree two. On the other hand the non constructible vectors of length 2 are $(0, 1)$, $(0, 2)$, $(0, 3)$ and $(1, 1)$.

2.3 Sackin Index

This section is devoted to the Sackin index, one of the first indices to measure the balance of a rooted tree. The Sackin is defined as the sum of the depths of all the leaves of the tree:

Definition 2.9. Let $T_{n,r} \in \mathcal{T}_{n,r}$ a rooted-tree with n leaves. We define $S(T_{n,r})$ the Sackin Index of $T_{n,r}$ as:

$$S(T_{n,r}) = \sum_{i=1}^{n} \delta(l_i)$$

One of our goals is to find the tree with the minimum Sackin index, which we will consider the most balanced tree and the tree with the maximum Sackin index, which we will consider the most unbalanced tree in $\mathcal{T}_{n,r}$.

Next, we will see which Sackin indices are reachable (between the minimum and the maximum), that is, for every value x between the minimum and maximum of the Sackin index in $\mathcal{T}_{n,r}$, we want to find out if there exists a tree $T_{n,r} \in \mathcal{T}_{n,r}$ such that the Sackin index of $S(T_n) = x$, and finally, we will find the maximum and the minimum Sackin index in the set of trees with n leaves but setting the maximum depth of them.

2.4 Extremal values for the Sackin index

Let us start by finding the tree with the minimum Sackin index in $\mathcal{T}_{n,r}$.

The tree with minimum Sackin index in $\mathcal{T}_{n,r}$ is the star:

Definition 2.10. Let $T_{n,r} \in \mathcal{T}_{n,r}$ be a rooted-tree with n leaves, we call $T_{n,r}$ a star-tree or n-cherry, if and only if, $\forall i = 1, \ldots, n$, $\delta(l_i) = 1$. The star in $T_{n,r}$ will be denoted by $S_{n,r}$, see figure 2.

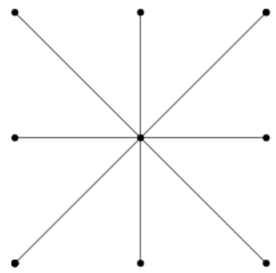


Figure 2: A star tree or an 8-cherry.

Thus, in our setting, a star will be always assumed to be rooted at its center (not at a leave).

Theorem 2.11. *For all $T \in \mathcal{T}_{n,r}$ we have*

$$S(S_{n,r}) \leq S(T).$$

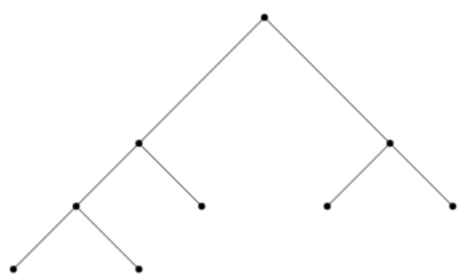
Proof. Let $T \in \mathcal{T}_{n,r}$ be a rooted tree with n leaves which is not the star, so that $\delta = \delta(T) > 1$. We construct a tree $\hat{T} \in \mathcal{T}_{n,r}$ with n leaves such that $S(\hat{T}) < S(T)$. Consider the set L_δ of the leaves of T with maximum depth δ and let p be the number of interior vertices of T with depth $\delta - 1$. We observe that, since there are no interior vertices of degree two, we have $|S_\delta| - p \geq 1$. In order to construct the new tree $\hat{T} \in \mathcal{T}_{n,r}$, remove all these leaves and add $|S_\delta| - p$ leaves to the root r. This transformation gives a new tree $\hat{T} \in \mathcal{T}_{n,r}$ such that:

$$S(\hat{T}) = S(T) - |S_\delta|\delta + (\delta - 1)p + |S_\delta| - p = S(T) + (\delta - 1)(p - |S_\delta|) - p < S(T),$$

So, if T_n is not a star then the value of $S(T_n)$ is not minimum in $\mathcal{T}_{n,r}$. Since $\mathcal{T}_{n,r}$ is finite, the function S takes its minimum at the star. □

Next, we will define a new concept in order to find the tree in $\mathcal{T}_{n,r}$ with maximum Sackin index:

Definition 2.12. Let $T_{n,r} \in \mathcal{T}_{n,r}$ be a rooted-tree with n leaves. We define $\delta'(T_{n,r})$, or simply δ', as the maximum integer such that $\exists v \neq w$ interior nodes of $T_{n,r}$ with $\delta(v) = \delta(w) = \delta'$. If these nodes don't exist, the value of δ' will be $\delta' = 0$, see figure 3.



$\delta'(T_{n,r}) = 1, \delta(T_{n,r}) = 3$

Figure 3: A tree with a $\delta' = 1$ and $\delta = 3$.

With this definition, let us prove a lemma that helps us know if the trees we define are the same:

Lemma 2.13. *Let $T, T' \in \mathcal{T}_{n,r}$ and v, v' their depth frequency vectors, respectively. If $\delta'(T) = \delta'(T') = 0$ and $v = v'$, Then, T and T' are isomorphic as graphs.*

Proof. To prove that T and T' are isomorphic, we have to define a bijective function $f : V(T) \to V(T')$ such that if $(u, v) \in E(T)$ then, $(f(u), f(v)) \in E(T')$.

To define f that fulfills the two previous condition let us make three observations:

- first of all, using that $v = v'$ and consequently, $\delta(T) = \delta(T')$ and lemma 2.5, there exists two interiors nodes $g \in V(T) \setminus L(T)$, $g' \in V(T') \setminus L(T')$ such that $\delta(g) = \delta(g') = i$. Using that $\delta'(T) = \delta'(T') = 0$, we have that these interior nodes are unique.

- secondly, all the interior nodes of the trees T and T' have at least one child that is a leaf because if not, $\delta'(T)$ or $\delta'(T')$ would not be 0,

- lastly, using the first observation, if $v = (v_1, \dots, v_\delta)$, for every v_i, there are only one interior node of T so it has v_i leaf children, because if the depth of these v_i leaves are i, the depth of their father will be $i - 1$ and we have seen that there is only one interior node of each depth. The same happens for the tree T'. In conclusion, if v and v' are interior nodes of the trees T and T', respectively of the same depth, they have the same number of children that are leaves.

Now the definition of f is clear: let $v \in V(T)$,

- if $v \in V(T) \setminus L(T)$, we define $f(v)$ as the only interior node of V' such that $\delta(v') = \delta(v)$,

- if $v \in L(T)$ such that $\delta(v) = i$, we have that $v \in S_i(T)$. Next, we consider the sets $S_i(T)$ and $S_i(T')$. We have seen that the two previous sets have the same cardinality, so we define f restricted to the set S_i as a bijection between the two sets $S_i(T)$ and $S_i(T')$. The value of $f(v)$ will be the image of that bijection.

It is straightforward to see that the previous function f preserves adjacency. let us see that if $(u, v) \in E(T)$ then $(f(u), f(v)) \in E(T')$. Consider two cases:

- $u, v \in V(T) \setminus L(T)$ We know by hypothesis that $\delta'(T) = 0$ implying that u, v are the only interior nodes of their respectively depths. Because $(u, v) \in E(T)$ we deduce $\delta(u) = \delta(v) \pm 1$, but also notice that from the first observation we can deduce also that, if $\delta(u) = \delta(v) \pm 1$ then $(u, v) \in E(T)$. By the definition of f we know that $\delta(f(u)) = \delta(u) = \delta(v) \pm 1 = \delta(f(v)) \pm 1$, and because by hypothesis $\delta'(T') = 0$ as we observe we can deduce that $(f(u), f(v)) \in E(T')$

- u or $v \in L(T)$ We can suppose without lost of generality that $u \in L(T)$. Now notice that for being $f(u)$ a leave of depth $\delta(f(u)) = \delta(u) = \delta(v) + 1 = \delta(f(v))$ and $f(v)$ the only interior node of depth $\delta(f(u)) - 1$ we deduce that $(f(u), f(v)) \in E(T')$

$\square$

Now we will state 2 lemmas that will help us to find the maximum Sackin index in $\mathcal{T}_{n,r}$:

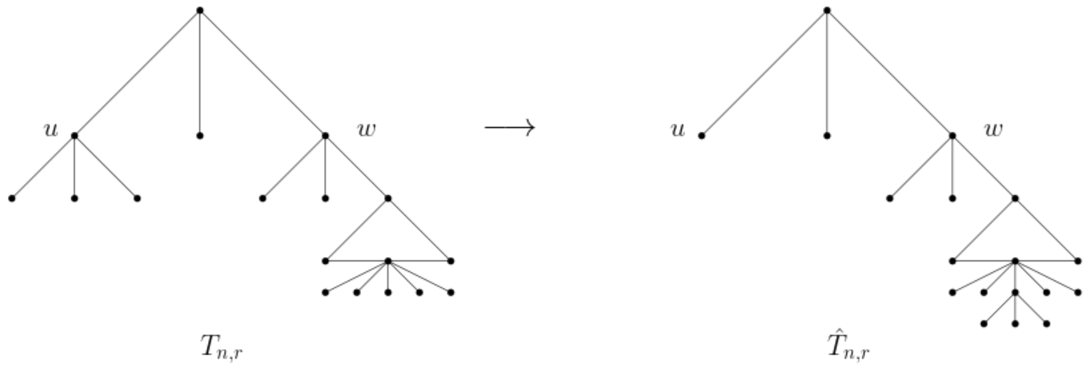


Figure 4: The result of removing all leaves that are children of u and adding those leaves to a leave of maximum depth δ. We understand a triangle as generic rooted tree that's hanging from the tree

Lemma 2.14. *Let $T_{n,r} \in \mathcal{T}_{n,r}$ be a rooted-tree with n leaves with maximum Sackin index then $\delta'(T_{n,r}) = 0$*

Proof. We will prove it by counterpositive. We assume that $\delta'(T_{n,r}) > 0$ and we will construct a tree $\hat{T}_{n,r} \in \mathcal{T}_{n,r}$ such that $S(\hat{T}_{n,r}) > S(T_{n,r})$. Because $\delta'(T_{n,r}) > 0$, there exist two interior nodes $u \neq w$ such that $\delta(u) = \delta(w) = \delta'(T_{n,r})$. Notice that the children of one of this nodes must be all leaves, because if not, let u' and w' be one child of each node u and w, respectively such that u' and w' are interior nodes. In this case, we have $\delta(u') = \delta(w') = \delta'(T_{n,r}) + 1$ and it contradicts the definition of $\delta'(T)$. Without loss of generality, we can assume that this node is u and the number of children of u is k. So, to construct a new tree $\hat{T}_{n,r}$, we remove all the children of u that recall, there are all leaves, and we add these leaves to a leaf with maximum depth δ, see figure 4 . So, we have

$$S(\hat{T}_{n,r}) - S(T_{n,r}) = k(\delta+1) + \delta' - k(\delta'+1) - \delta = (k-1)(\delta - \delta') > 0 \Rightarrow S(\hat{T}_{n,r}) > S(T_{n,r}).$$

□

Next, we will define our candidate to be the tree that has the maximum Sackin index in $\mathcal{T}_{n,r}$.

Definition 2.15. Let $T_{n,r}$ be a rooted-tree, we say that $T_{n,r}$ is a k-caterpillar, if and only if, $\delta'(T_{n,r}) = 0$ and $\forall v \notin L(T_{n,r})$, v has exactly k children, see figure 5.

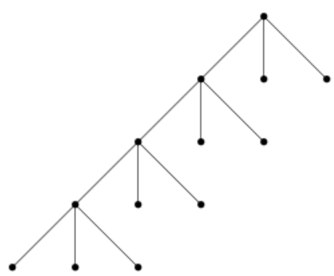


Figure 5: An example of a 3-caterpillar

The next theorem gives us the tree with the maximum Sackin index in $\mathcal{T}_{n,r}$:

Theorem 2.16. *Let $T_{n,r} \in \mathcal{T}_{n,r}$. Then, $\forall T'_{n,r}$, $S(T_{n,r}) \geq S(T'_{n,r})$, if and only if, $T_{n,r}$ is a 2-caterpillar.*

Proof. By Lemma 2.14 we know that a tree with maximum Sackin index consists of a rooted path $\{r = x_0, x_1, \dots, x_{\delta-1}\}$ of length $\delta - 1$ with some vector of depth frequency vector $v = (v_1, \dots, v_\delta)$, where v_i is the number of leaves attached to the vertex x_{i-1}. If there is a coordinate v_i with $1 \leq i \leq \delta - 1$ larger than one then the tree obtained from $T_{n,r}$ by removing a leave of S_i and adding a leave to $x_{\delta-1}$ has larger Sacking index. So we have $v_i = 1$ for all $1 \leq i \leq \delta - 1$. If $v_\delta > 2$ then deleting a leave from $x_{\delta-1}$ and adding it to a leave in S_δ produces a tree $T'_{n,r}$ with $\delta(T'_{n,r}) = \delta(T_{n,r}) + 1$, $\delta'(T'_{n,r}) = 0$ and a larger Sacking index, see figure 6. Hence we may further assume that $v_\delta = 2$, which is the depth frequency vector of 2-caterpillar. □

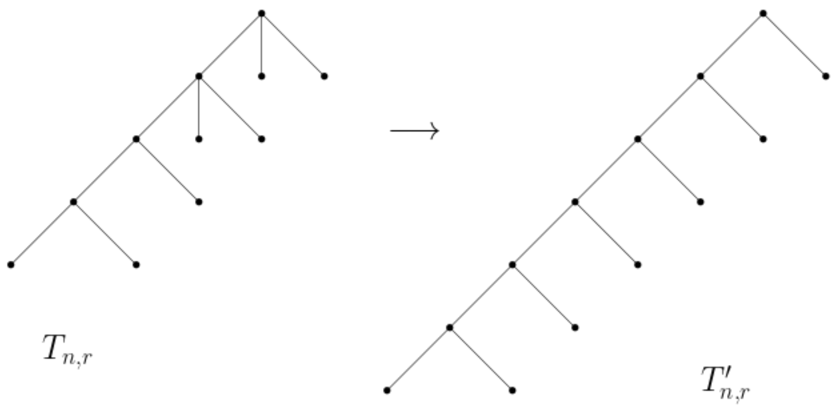


Figure 6: Construction of $T'_{n,r}$ if $T_{n,r}$ has no interior binary nodes

Notice that the value of the minimum Sackin index is $m_{n,r} = n$, and the value of the maximum Sackin index is $M_{n,r} = \frac{1}{2}(n-1)(n+2)$. The value of the minimum is straightforward to compute because all the leaves of the star-tree have depth 1. The maximum $M_{n,r}$ is calculated in the following way:

$$\sum_{i=1}^{n-1} i + n - 1 = \frac{n(n-1)}{2} + n - 1 = \frac{1}{2}(n-1)(n+2).$$

2.5 Range of Sackin indices

Another problem that we can try to solve is if the Sackin index between the minimum and the maximum is accessible for some trees. Our claim is that every value except $n+1$ can be reached:

Lemma 2.17. *Let $T_{n,r} \in \mathcal{T}_{n,r}$. Then, $S(T_{n,r}) \neq n+1$.*

Proof. let us suppose that $S(T_{n,r}) = n+1$. Because $T_{n,r}$ is not a star-tree (if it was, $S(T_{n,r}) = n$), we consider the set of all leaves of maximum depth $\delta > 1$, S_δ.

Let k the number of interior nodes of $T_{n,r}$ of depth $\delta - 1$. Next, we construct a new tree $T'_{n,r} \in \mathcal{T}_{n,r}$ in the following way:

1. first, we remove the leaves of depth δ in $T_{n,r}$,

2. due to the previous step, we have generated k new leaves of depth $\delta - 1$,

3. we add $|S_\delta| - k$ leaves to the root of $T_{n,r}$.

The value of $S(T'_{n,r})$ will be:

$$S(T'_{n,r}) = S(T_{n,r}) - |S_\delta|\delta(\text{step } 1) + (|S_\delta| - k)(\text{step } 3) + k(\delta - 1)(\text{step } 2).$$

Notice that $|S_\delta| > k$ because for each interior node of depth $\delta - 1$ there will be at least two leaves of depth δ, which implies that

$$|S_\delta|\delta - (|S_\delta| - k) - k(\delta - 1) = (\delta - 1)(|S_\delta| - k) + k \geq 2.$$

Therefore, $S(T'_{n,r}) \leq S(T_{n,r}) - 2 = n - 1$ contradicting the fact that n is the minimum Sackin Index. □

Theorem 2.18. *Let $s \in \mathbb{N}$ such that $m_{n,r} \leq s \leq M_{n,r}$ and $s \neq n + 1$. Then, exists a tree $T_{n,r} \in \mathcal{T}_{n,r}$ such that $S(T_{n,r}) = s$.*

Proof. To prove this theorem we will use the following strategy. We will consider a tree $T_{n,r} \in \mathcal{T}_{n,r}$ different from a 2-caterpillar and a star-tree and we will construct a new tree $\hat{T}_{n,r} \in \mathcal{T}_{n,r}$ such that $S(\hat{T}_{n,r}) = S(T_{n,r}) + 1$. Note that if we do that, the prove is done.

So, let $T_{n,r} \in \mathcal{T}_{n,r}$ different from a 2-caterpillar and a star-tree and we consider two cases:

$\delta'(T_{n,r}) > 0$: to construct the new tree $\hat{T}_{n,r}$, let us consider v, w interior nodes of depth δ' and we consider two subcases:

1. v, w have exactly two children. In this case, a 2-cherry is hanging on one of the two nodes because if not it would be a depth greater than δ' such that there exists two interior nodes of this depth. Without loss of generality, we assume that the 2-cherry is hanging from v. To construct $\hat{T}_{n,r}$ we remove this 2-cherry from v and we add it on a leaf of depth $\delta' + 1$ hanging from w that we know for sure exists otherwise we would find two interior nodes of depth $\delta' + 1$. Note that the new tree $\hat{T}_{n,r}$ has Sackin index

$$S(\hat{T}_{n,r}) = S(T_{n,r}) - 2(\delta' + 1) + \delta' - (\delta' + 1) + 2(\delta' + 2) = S(T_{n,r}) + 1.$$

2. one of the two nodes have more than two children. Using the same argument as in the previous case, we have that there exists a k such that a k-cherry is hanging from one of the two nodes v or w, let us suppose v. If $k = 2$, we can do the same as in the previous case. If $k > 2$, we construct our new tree $\hat{T}_{n,r}$ using two steps:

   (a) We remove all the leaves of v minus two and we add all those leaves to the node w. Notice that doing this does not change the Sackin index of the tree.

   (b) Now since v have exactly two children that are leaves, we can use the argument of the previous case.

   In summary, we have constructed a tree such that

$$S(\hat{T}_{n,r}) = S(T_{n,r}) - 2(\delta' + 1) + \delta' - (\delta' + 1) + 2(\delta' + 2) = S(T_{n,r}) + 1.$$

See figure 7.

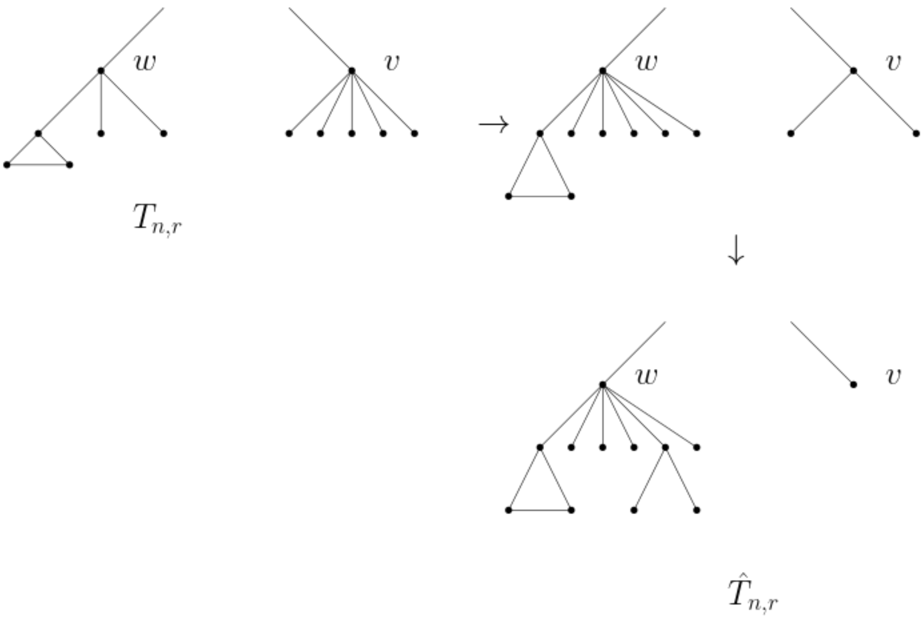


Figure 7: Construction in the case $\delta'(T_{n,r}) > 0$

$\delta'(T) = 0$: in this case, there exists v interior node with more than 2 children because if not T would be a 2-caterpillar. To construct our new tree $\hat{T}_{n,r}$, we consider two cases:

1. The depth of v is $q < \delta - 1$. In this case, notice that v must have two children that are leaves because otherwise $\delta'(T_{n,r}) > 0$. Also, it must exist an interior node w of depth $q+1$ because $q+1 < \delta$ and we know that there exists an interior node of each depth less than δ. The new tree $\hat{T}_{n,r}$ is constructed removing one of the previous two children from v and adding it to w.

2. The depth of v is $\delta - 1$. In this case, we remove a child leaf of v of depth δ and we add a 2-cherry to a brother of v which is necessarily a leaf because otherwise, $\delta'(T_{n,r}) \neq 0$. Note that $\delta \neq 1$ because we have assumed that $T_{n,r}$ is not a star-tree.

See figure 8.

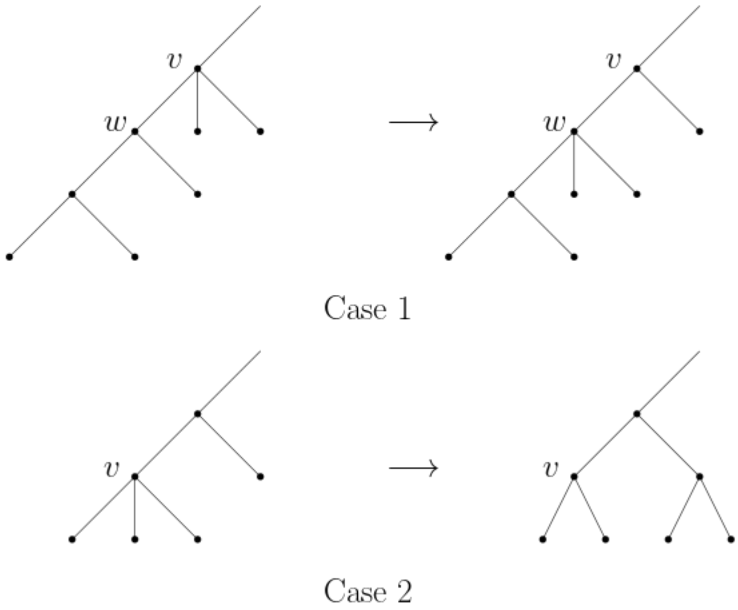


Figure 8: Construction in the case $\delta'(T_{n,r}) = 0$

In summary, in each case, we have constructed a new tree $\hat{T}_{n,r} \in \mathcal{T}_{n,r}$ such that $S(\hat{T}_{n,r}) = S(T_{n,r})+1$. $\square$

2.6 Extremal values for given depth

Now we are going to solve the problem of finding the tree with maximum and minimum Sackin Index but not in the set $\mathcal{T}_{n,r}$, instead we will solve it in the set of rooted trees of n leaves and maximum depth δ. We will call this set $\mathcal{T}_{n,\delta}$:

$$\mathcal{T}_{n,\delta} = \{T_{n,r} \in \mathcal{T}_{n,r} \text{ such that } \max_{u \in L(T_{n,r})} \delta(u) = \delta\}.$$

First, let us prove some lemmas that will help us to know in which cases $\mathcal{T}_{n,\delta} = \emptyset$.

Lemma 2.19. *Let $T_{n,r} \in \mathcal{T}_{n,r}$ such that $\delta(T_{n,r}) = \max_{T'_{n,r} \in \mathcal{T}_{n,r}} \delta(T'_{n,r})$. Then, $T_{n,r}$ is a binary tree.*

Proof. let us prove by counterpositive. So, suppose that $T_{n,r}$ is not a binary tree. Then, there exists an interior node v with more than 2 children. Next, we will construct a new tree $T'_{n,r}$ such that $\delta(T'_{n,r}) > \delta(T_{n,r})$. To do it, you have to follow the following steps:

- first, remove the subtree hanging over one of the children of v, w,
- next, remove the new leaf w,
- at the end add the leaf w and the previous subtree to a leaf of $T_{n,r}$ of depth $\delta(T_{n,r})$.

In summary, we have constructed a new tree $T'_{n,r}$ with n leaves with $\delta(T'_{n,r}) > \delta(T_{n,r})$. $\square$

Theorem 2.20. *$\mathcal{T}_{n,\delta} \neq \emptyset$, if and only if, $n \geq \delta + 1$*

Proof. We will prove the right implication by counterpositive. So, let us suppose that $n < \delta + 1$ and let us consider the tree $T_{n,r}$ such that $\delta(T_{n,r}) = \max_{T'_{n,r} \in \mathcal{T}_{n,r}} \delta(T'_{n,r})$, and we will prove that $\delta(T_{n,r}) < \delta$. First, We know that if T is a binary tree with n leaves and p interior nodes, then $n = p + 1$. Using the previous lemma, the tree $T_{n,r}$ is binary. Because, for each depth, there exists one interior node, we have that the number of interior nodes of $T_{n,r}$ must be greater that the maximum depth $\delta(T_{n,r})$:

$$p = n - 1 \geq \delta(T_{n,r}) \ \Rightarrow n \geq \delta(T_{n,r}) + 1, \ \Rightarrow \delta + 1 > n \geq \delta(T_{n,r}) + 1, \ \Rightarrow \delta > \delta(T_{n,r}).$$

Now let us prove the left implication: suppose that $n \geq \delta + 1$ and let us construct a tree with n leaves of maximum depth δ. Notice that we only have to prove that we can construct a tree of $n = \delta + 1$ and maximum depth δ because we can add the remaining leaves to the root. So, let us consider the 2-caterpillar of $\delta + 1$ leaves. It's easy to see that this tree has maximum depth δ. In conclusion, we have found a tree in $\mathcal{T}_{n,\delta}$ implying that $\mathcal{T}_{n,\delta} \neq \emptyset$. $\square$

From now on, when we consider the set $\mathcal{T}_{n,\delta}$ we will assume that δ and n fulfill the conditions of Theorem 2.20 in order to $\mathcal{T}_{n,\delta} \neq \emptyset$. The next step is to find the trees that have the maximum and the minimum Sackin index in this set $\mathcal{T}_{n,\delta}$:

Theorem 2.21 (Minimum Sackin index in $\mathcal{T}_{n,\delta}$)**.** *Let $T_{n,\delta} \in \mathcal{T}_{n,\delta}$, Then $\forall T'_{n,\delta} \in \mathcal{T}_{n,\delta}$, $S(T_{n,\delta}) \leq S(T'_{n,\delta})$ if, and only if, $\forall i = 2, \dots, \delta - 1, |S_i| = 1$, $|S_\delta| = 2$ and $|S_1| = n - \delta$.*

Proof. We prove the right implication by counterpositive. Let $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ and we consider two cases:

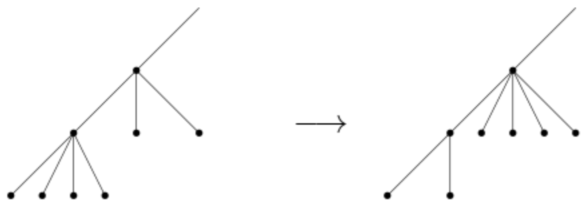
- $|S_\delta| > 2$: In this case we can construct another tree $T'_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that $S(T'_{n,\delta}) < S(T_{n,\delta})$. To do it, consider 2 more cases:

  1. All leaves of depth δ belong to the same cherry. In this case, because there are more than 2 leaves of depth δ, to construct $T'_{n,\delta}$, we remove a leaf of depth δ and hang it to an interior node of depth $\delta - 2$. So, $S(T'_{n,\delta}) = S(T_{n,\delta}) - \delta + \delta - 1 = S(T_{n,r}) - 1 < S(T_{n,\delta})$.

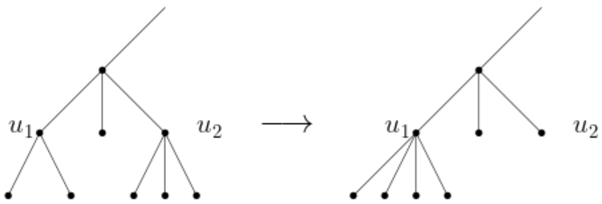
  2. Not all the leaves of depth δ belong to the same cherry. Let u_1 and u_2 the fathers of two of these cherries whose leaves have depth δ. Let $q \geq 2$ the number of leaves of the cherry of the interior node u_1. To construct $T'_{n,\delta}$, we remove q leaves of the cherry of the interior node u_1 and add $q - 1$ of them to the cherry of node u_2. Note that $T'_{n,\delta} \in \mathcal{T}_{n,\delta}$ and

  $$S(T'_{n,\delta}) = S(T_{n,\delta}) - q\delta + (q-1)\delta + \delta - 1 = S(T_{n,\delta}) - 1 < S(T_{n,\delta}).$$

  See figure 9



Case 1



Case 2

Figure 9: Construction in the case $|S_\delta| > 2$

- $|S_i| \neq 1$ for some $i = \{2, \dots, \delta - 1\}$: In this case we also will construct a tree $T'_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that $S(T'_{n,\delta}) < S(T_{n,\delta})$. To do it, consider 2 more cases:

  1. $\delta'(T_{n,\delta}) = 0$: in this case, $|S_i| > 0$ because if $|S_i| = 0$, there would be two interior nodes of depth i and $\delta'(T_{n,\delta}) > 0$, i.e., consider a node v such that depth $\delta(v) = i - 1$, this node has at least two children of depth i and because $|S_i| = 0$, they are interior nodes, so $\delta'(T_{n,\delta}) > 0$. Therefore, $|S_i| \geq 2$. Next, using that $\delta'(T_{n,\delta}) = 0$, we consider the only interior node w of depth $i - 1$, $\delta(w) = i - 1$. Because $|S_i| \geq 2$, w has at least two children that are leaves and one interior node w' of depth i. To construct the new tree $T'_{n,\delta}$, we remove one leaf of depth i of the interior node w and we add it to the only interior node of depth $i - 2$ (in case when $i = 2$, this interior node would be the root). The value of $S(T'_{n,\delta})$ will be:

  $$S(T'_{n,\delta}) = S(T_{n,\delta}) - i + (i - 1) = S(T_{n,\delta}) - 1 < S(T_{n,\delta}).$$

  2. $\delta'(T_{n,\delta}) > 0$: in this case, we consider two interior nodes v_1 and v_2 of depth $\delta'(T_{n,\delta})$. If neither of these interior nodes has a cherry hanging from it, it means that there would be two interior nodes (one child for each v_1 and v_2) of depth $\delta'(T_{n,\delta}) + 1$, which contradicts the definition of $\delta'(T_{n,\delta})$. Therefore, we suppose that v_1 is an interior node with a q-cherry hanging from it, where q is the number of leaves of that cherry. To construct the new tree $T'_{n,\delta}$ we remove all the q leaves of that cherry of depth $\delta'(T_{n,\delta}) + 1$ and we add $q - 1$ of them to the interior node v_2 of depth $\delta'(T_{n,\delta})$. Therefore, the value of $S(T'_{n,\delta})$ will be:

  $$S(T'_{n,\delta}) = S(T_{n,\delta}) - q(\delta' + 1) + (q - 1)(\delta' + 1) + \delta' = S(T_{n,\delta}) - 1 < S(T_{n,\delta}).$$

  See figure 10

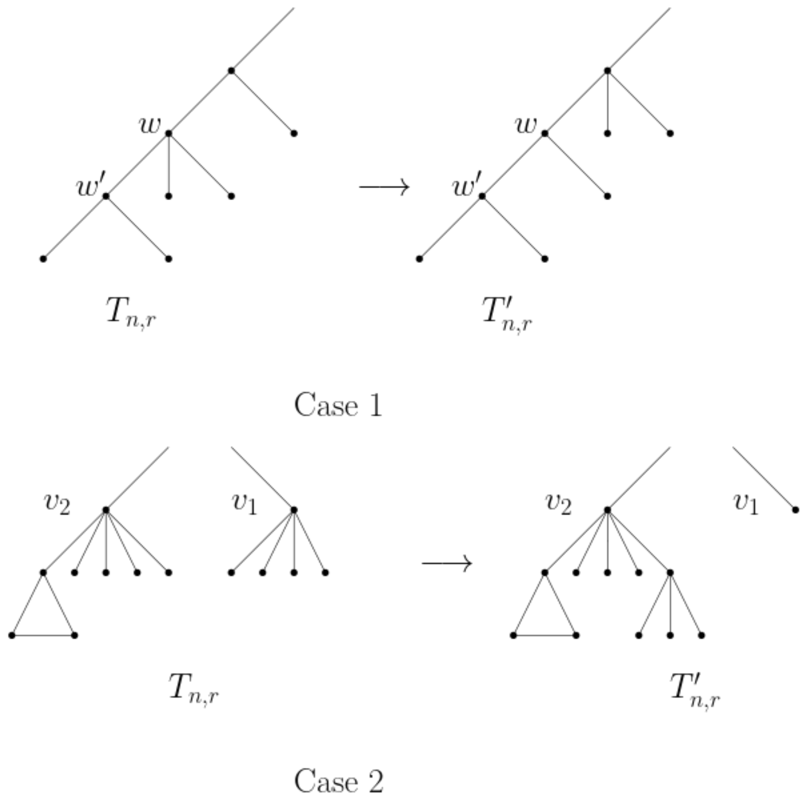


Figure 10: Construction in the case $|S_i| \neq 1$

Now it is straightforward to see that $|S_1| = n - \delta$ because:

$$|S_1| = n - \sum_{i=2}^{\delta-1} |S_i| - |S_\delta| = n - \sum_{i=2}^{\delta-1} 1 - 2 = n - (\delta - 2) + 2 = n - \delta.$$

For the proof from right to left, we use same argument that we have used before. Because we know exactly the number of leaves of each depth the Sackin index of $T_{n,\delta}$ is the following:

$$S(T_{n,\delta}) = |S_1| \cdot 1 + \sum_{i=2}^{\delta-1} |S_i| \cdot i + |S_\delta| \cdot \delta = (n - \delta) + \frac{1}{2}\delta \cdot (\delta - 1) - 1 + 2\delta = \frac{1}{2}\delta(\delta + 1) + n - 1.$$

We have proved that all trees that fulfill these conditions have the same Sackin index and that the tree with the minimum Sackin index verifies these conditions. Therefore, the tree $T_{n,\delta}$ has the minimum Sackin index in $\mathcal{T}_{n,\delta}$. $\square$

Proposition 1. *The tree $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ of theorem 2.21 is unique is the sense that if there are two trees that verify the conditions of theorem 2.21, they are isomorphic.*

Proof. Notice that $\delta'(T_{n,\delta}) = 0$ and all trees have the same depth frequency vector so using lemma 2.13 we can deduce the statement $\square$

The next step is to find the maximum Sackin index in $\mathcal{T}_{n,\delta}$: We will divide this problem in two parts, one supposing that $n \geq 2^\delta$ and for the other we will suppose $n < 2^\delta$.

Theorem 2.22. *Suppose that $n \geq 2^\delta$ and let $T_{n,\delta} \in \mathcal{T}_{n,\delta}$, $\forall T'_{n,\delta} \in \mathcal{T}_{n,\delta}$, $S(T_{n,\delta}) \geq S(T'_{n,\delta})$, if and only if, $|S_\delta| = n$*

Proof. From left to right, we will prove it by counterpositive, let $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that $|S_\delta| < n$ and let us construct a $T'_{n,\delta} \in \mathcal{T}_{n,\delta}$ with $S(T'_{n,\delta}) > S(T_{n,\delta})$. First of all, notice that $S(T_{n,\delta}) < n\delta$ because there exist an $0 < \alpha < \delta$ such that $|S_\alpha| > 0$, so $S(T_{n,\delta}) = \sum_{i=1}^{n} \delta(l_i)$ with $\delta(l_i) \leq \delta \forall 1 \leq i \leq n$ and for some j $\delta(l_j) = \alpha < \delta$ so we can conclude that $S(T_{n,\delta}) < n\delta$. Now to construct our new $T'_{n,\delta}$, we consider the full balanced tree with maximum depth δ and 2^δ leaves and we add as a child the remaining $n - 2^\delta$ leaves to any node of depth $\delta - 1$, as all the leaves are of depth δ we can conclude that $S(T'_{n,\delta}) = n\delta > S(T_{n,\delta})$.

From right to left, observe that all trees that fulfills the condition of $|S_\delta| = n$ have the same Sackin index because in this case all leaves are of depth δ implying that their Sackin index is $n\delta$, as we proved before one of this trees has to correspond to the maximum so as they all have the same Sackin index they are all maximum. ☐

Theorem 2.23. *Suppose that $n < 2^\delta$. Let $\tilde{\delta}$ be the maximum of the set $\{l, \mid n \geq 2^l + \delta - l\}$, that is $\tilde{\delta} = \max\{l, \mid n \geq 2^l + \delta - l\}$. Let $T_{n,\delta} \in \mathcal{T}_{n,\delta}$, then $\forall T'_{n,\delta} \in \mathcal{T}_{n,\delta}$, $S(T_{n,\delta}) \geq S(T'_{n,\delta})$, if and only if the following conditions are verified:*

- *$\forall k > \tilde{\delta}$, $|S_{\delta-k}| = 1$ and all interior node $v \in V(T_{n,\delta}) \setminus L(T_{n,\delta})$ with $\delta(v) = \delta - k$ is binary and there are only two nodes v_1, v_2 of depth $\delta - \tilde{\delta}$.*
- *Let $T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}}$ the tree with maximum Sackin index in the set $\mathcal{T}_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}}$:*
  - *if $(n - 2^{\tilde{\delta}} - \delta + \tilde{\delta})\tilde{\delta} \geq S(T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}})$, then $T_{v_2} = \{v_2\}$ and T_{v_1} is a tree with all their $n - \delta + \tilde{\delta}$ leaves of depth δ,*
  - *if $(n - 2^{\tilde{\delta}} - \delta + \tilde{\delta})\tilde{\delta} < S(T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}})$, then $T_{v_2} \neq \{v_2\}$, there exists a $\hat{T}_{n,\delta}$ fulfilling the previous conditions with $S(T_{n,\delta}) = S(\hat{T}_{n,\delta})$ such that the new $\hat{T}_{v_1}$ is a binary tree with only leaves of depth δ and the new $\hat{T}_{v_2} = T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}}$*

See figure 11

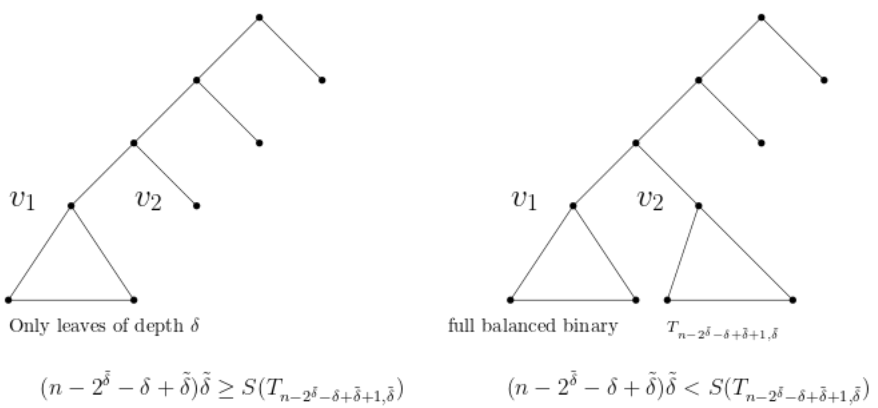


Figure 11: This is the shape of the maximum Sackin index over $\mathcal{T}_{n,\delta}$

We need the following lemmas to prove the theorem:

Lemma 2.24. *Let $T_{n,\delta} \in \mathcal{T}_{n,\delta}$, such that $\forall T'_{n,\delta} \in \mathcal{T}_{n,\delta}$, $S(T_{n,\delta}) \geq S(T'_{n,\delta})$, then, $\forall \alpha, 0 < \alpha < \delta - 1$, $|S_\alpha| \leq 1$, $|S_{\delta-1}| \leq 2$ and if $|S_{\delta-1}| = 2$, then $|S_{\delta-2}| = 0$.*

Proof. We will do this proof by counterpositive, that means, we consider a tree $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that the property of the right side of the implication is not fulfilled and we will construct a new tree $T'_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that $S(T'_{n,\delta}) \geq S(T_{n,\delta})$. Notice that not fulfilling the right side of the statement implies not fulfilling one of the three conditions of the statement, so let divide the proof in three different cases:

- Let $T_{n,\delta}$ such that $\exists \alpha$ such that $|S_\alpha| > 1$. To construct this new tree $T'_{n,\delta}$ let us consider two cases:

  1. The father of one of the leaves of depth α is binary. To construct this new tree $T'_{n,\delta}$, we perform two steps:
     - If the brother b of the leaf of depth α is also a leaf, we do nothing. If b is not a leaf, it is an interior node. In this case, we remove the subtree T_b and we hang it to another leaf of depth α. In the first step, the Sackin index of the new tree doesn't change.
     - Once we have performed the first step, we can suppose that we have a 2-cherry of leaves of depth α. So, we remove the leaves of this 2-cherry and add one leave of that 2-cherry to any interior node of depth $\delta - 1$.

     See figure 12.

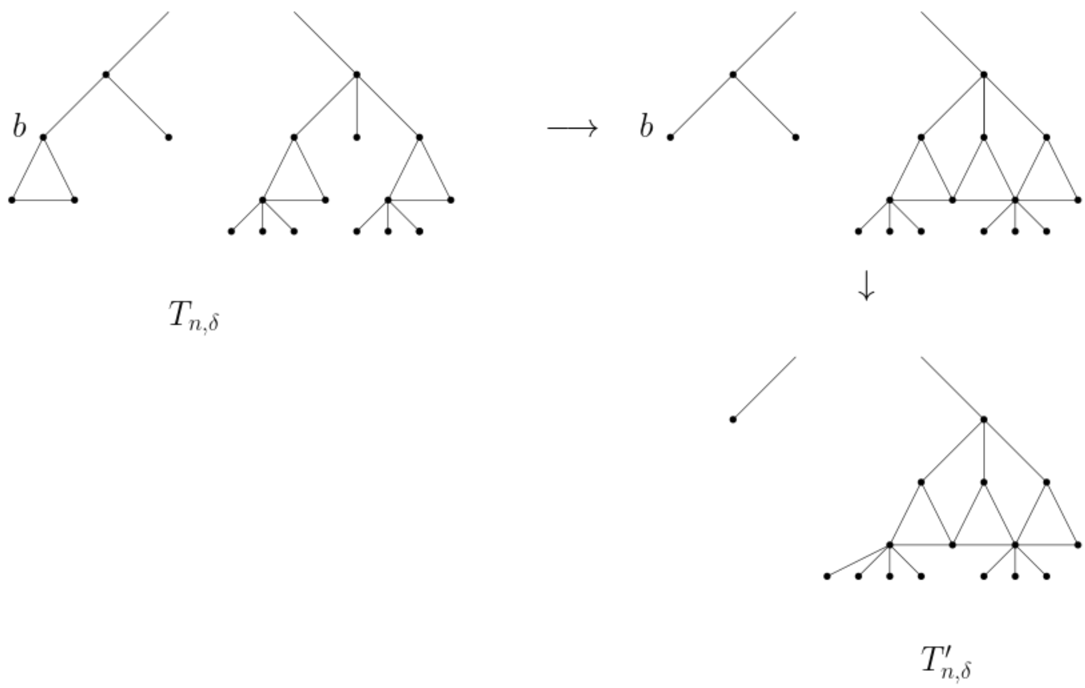


Figure 12: Construction in the case that exist $|S_\alpha| > 1$

The difference between the Sackin index of the two trees is:

$$S(T'_{n,\delta}) - S(T_{n,\delta}) = -2\alpha + \alpha - 1 + \delta = \delta - \alpha - 1 > 0$$

2. The father of any leaf of depth α is not binary. To construct this new tree $T'_{n,\delta}$, first we select one leaf of depth α, remove it and add it as a child to an interior node of depth $\delta - 1$. Notice that this construction don't generate trivial nodes because the nodes containing leaves of depth α as a child are not binary and notice that this construction remains in $\mathcal{T}_{n,\delta}$. So the difference of the Sackin index between the two trees is :

$$S(T'_{n,\delta}) - S(T_{n,\delta}) = -\alpha + \delta > 1.$$

See figure 13.

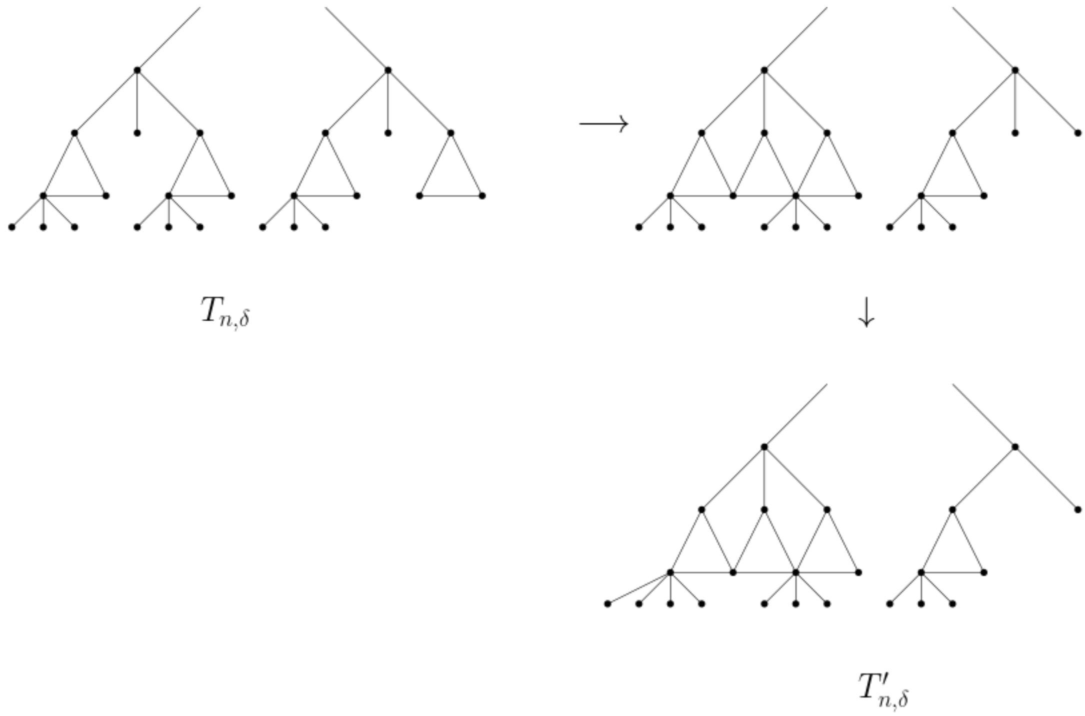


Figure 13: Construction in the case that exist $|S_\alpha| > 1$

- Let $T_{n,\delta}$ such that $|S_{\delta-1}| > 2$. To construct this new tree $T'_{n,\delta}$ let us consider two cases:

  1. The father of one of the leaves of depth $\delta - 1$ is binary. To construct this new tree $T'_{n,\delta}$, we perform two steps:

     – If the brother b of the leaf of depth $\delta - 1$ is also a leaf, we do nothing. If b is not a leaf, it is an interior node. In this case, we remove the subtree T_b (that will be a 2-cherry) and we hang it to another leaf of depth $\delta - 1$. In the first step, the Sackin index of the new tree doesn't change.

     – Once we have performed the first step, we can suppose that we have a 2-cherry of leaves of depth $\delta - 1$. So, we remove the leaves of this 2-cherry and add the leaves of that 2-cherry to any interior node of depth $\delta - 1$. (notice that we can do that because $|S_{\delta-1}| \geq 3$).

  See figure 14.

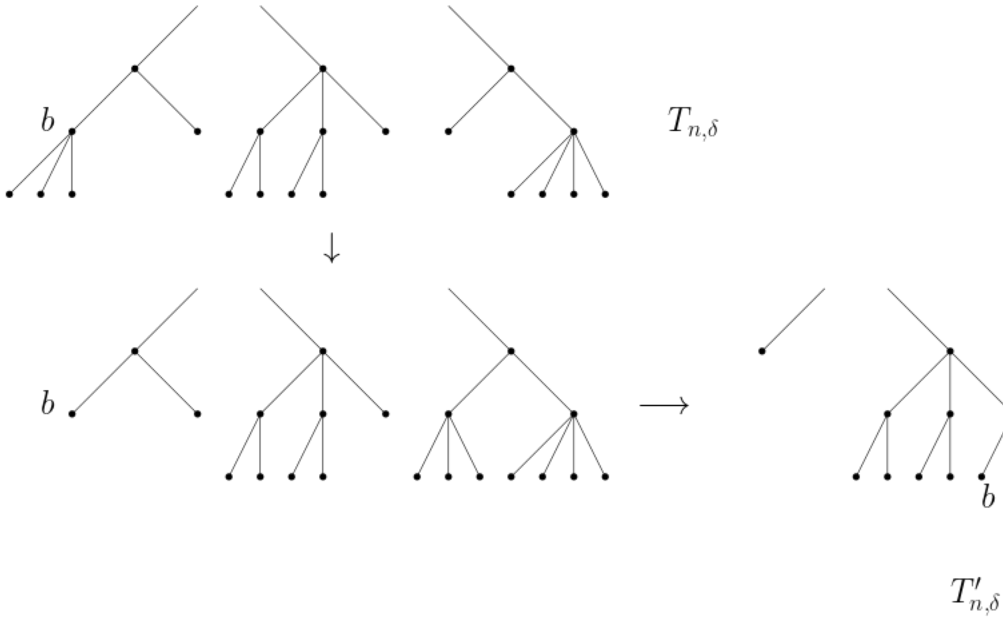


Figure 14: Construction in the case that exist $|S_{\delta-1}| > 2$

The difference between the Sackin index is:

$$S(T'_{n,\delta}) - S(T_{n,\delta}) = -2\delta + 2 + \delta - 2 - \delta + 1 + 2\delta = 1 > 0$$

2. The father of any leaf of depth α is not binary. To construct this new tree $T'_{n,\delta}$, first we select one leaf of depth $\delta - 1$, remove it and add it as a child to an interior node of depth $\delta - 1$. Notice that this construction don't generate trivial nodes because the nodes containing leaves of depth $\delta - 1$ as a child are not binary. The difference between the Sackin index of the two trees is:

$$S(T'_{n,\delta}) - S(T_{n,\delta}) = -\delta + 1 + \delta = 1 > 0.$$

See figure 15.

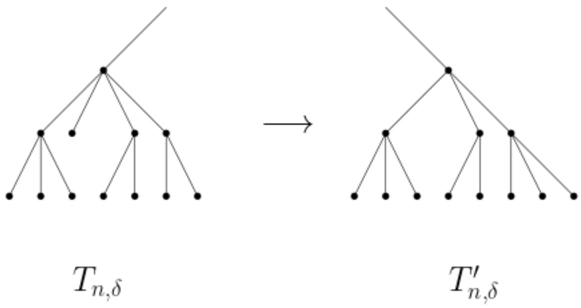


Figure 15: Construction in the case that exist $|S_{\delta-1}| > 2$

- Let $T_{n,\delta}$ such that $|S_{\delta-1}| = 2$ and $|S_{\delta-2}| \geq 1$. To construct this new tree $T'_{n,\delta}$ let us consider two cases:

1. The father of one of the leaves of depth $\delta - 1$ is binary. To construct this new tree $T'_{n,\delta}$, we perform three steps:
   - If the brother b of the leaf of dept $\delta - 1$ is also a leaf, we do nothing. If b is not a leaf, it is an interior node. In this case, we remove the subtree T_b and we hang it to another leaf of depth $\delta - 1$. In the first step, the Sackin index of the new tree doesn't change.
   - Once we have performed the first step, we can suppose that we have a 2-cherry of leaves of depth $\delta - 1$. So, we remove the leaves of this 2-cherry and add one leaf of that 2-cherry to any interior node of depth $\delta - 1$. Notice that this construction doesn't change also the Sackin index because:

   $$S(T'_{n,\delta}) - S(T_{n,\delta}) = \delta - 2 + \delta - 2(\delta - 1) = 0$$

   - Notice now that $T'_{n,\delta}$ doesn't fulfill the first condition (that we have already proved) of this lemma because $S_{\delta-2} \geq 2$ since we have generated an additional leaf of depth $\delta - 2$. In conclusion, $T'_{n,\delta}$ and $T_{n,\delta}$ cannot reach the maximum Sackin index, that is, we can construct a tree $T''_{n,\delta}$ using the same construction that we used in the first case such that $S(T''_{n,\delta}) > S(T'_{n,\delta}) = S(T_{n,\delta})$.
2. The father of any leaf of depth $\delta - 1$ is not binary. To construct this new tree $T'_{n,\delta}$, first we select one leaf of depth $\delta - 1$, remove it and add it as a child to an interior node of depth $\delta - 1$. Notice that this construction don't generate trivial nodes because the nodes containing leaves of depth $\delta - 1$ as a child are not binary. The difference between the Sackin index of the two trees is:

   $$S(T'_{n,\delta}) - S(T_{n,\delta}) = -\delta + 1 + \delta = 1 > 0$$

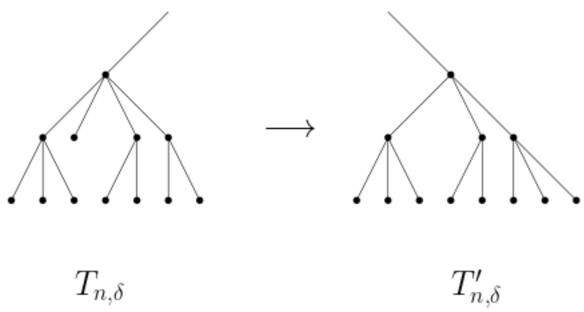


Figure 16: Construction in the case that exist $|S_{\delta-1}| = 2$

$\square$

Lemma 2.25. *Let $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that $\forall T'_{n,\delta} \in \mathcal{T}_{n,\delta}$, $S(T_{n,\delta}) \geq S(T'_{n,\delta})$. Then, $\forall v \in V(T_{n,\delta}) \setminus L(T_{n,\delta})$ the subtree T_v contains leaves of depth δ. (depth over $T_{n,\delta}$)*

Proof. All the depths in the proof are over $T_{n,\delta}$.

Using counterpositive, let us suppose that $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ contains $v \in V(T_{n,\delta}) \setminus L(T_{n,\delta})$ such that T_v doesn't have leaves of depth δ. Consider two cases:

- The maximum depth of T_v is α with $0 < \alpha < \delta - 1$. This assumption contradicts Lemma 2.24 in the sense that $|S_\alpha| \leq 1$ because the leaves of maximum depth form a k-cherry, with $k \geq 2$, so $|S_\alpha| \geq 2$.

- The maximum depth of T_v is $\delta - 1$. Notice that because the maximum depth of T_v is $\delta - 1$, if we consider an interior node w of T_v of depth $\delta - 2$, it will contain as a child at least 2 leaves of depth $\delta - 1$. Now, let w' be one brother of w. Consider two more subcases:

  1. $T_{w'}$ is a k-cherry. This assumption contradicts Lemma 2.24 because T_w is also a k-cherry and it would imply that $|S_{\delta-1}| > 2$ but $|S_{\delta-1}| \leq 2$.
  2. w' is a leaf. We have $|S_{\delta-1}| \geq 2$ because w has at least two children as leaves. If $|S_{\delta-1}| = 2$, by lemma 2.24, $|S_{\delta-2}| = 0$ but $|S_{\delta-2}| \geq 1$ because w' has depth $\delta - 2$.

  Notice that other cases do not make sense because that would imply that T_v would have leaves of depth δ.

□

Lemma 2.26. *Let $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that $\forall T'_{n,\delta} \in \mathcal{T}_{n,\delta}$, $S(T_{n,\delta}) \geq S(T'_{n,\delta})$. Consider $v \in V(T_{n,\delta}) \setminus L(T_{n,\delta})$, if v is not binary, then all the leaves of T_v have depth δ.*

Proof. We will do this proof by counterpositive, that is, we will find a new tree $T'_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that $S(T'_{n,\delta}) > S(T_{n,\delta})$.

Let $v \in V(T_{n,\delta}) \setminus L(T_{n,\delta})$ be a non-binary node. Let $v_1, \ldots, v_q$ be the children of v, for $q > 2$. Let $n_1 = |L(T_{v_1})|, \ldots, n_q = |L(T_{v_q})|$ be the number of leaves of subtrees T_{v_j}, $j = 1, \ldots, q$. Let w be a leaf of depth $\alpha < \delta$ of $T_{n,\delta}$. Without lost of generality we can suppose that this leaf w belongs to T_{v_1}. To construct our new tree $T'_{n,\delta}$, we remove the tree T_{v_1} and the remaining new leaf and we add n_1 leaves to an interior node of depth $\delta - 1$, that we can ensure that exists because by lemma 2.25 all subtrees generated by a node contain leaves of depth δ. So, the other $T_{v_2} \ldots T_{v_q}$ contain some interior node of depth $\delta - 1$ see figure 17. Next, if we compute $S(T_{n,\delta}) - S(T'_{n,\delta})$, we get:

$$
\begin{aligned}
S(T_{n,\delta}) - S(T'_{n,\delta}) &= S(T_{v_1}) + n_1(\delta(v) + 1) - n_1\delta \\
&< (\delta - \delta(v) - 1)n_1 + n_1(\delta(v) + 1) - n_1\delta \\
&= n_1(\delta - \delta(v) - 1 + \delta(v) + 1 - \delta) = 0.
\end{aligned}
$$

Notice now that at most all the leaves of T_{v_1} have the maximum depth and we can ensure that we have a leaf that does have the maximum depth. So, $S(T_1) < (\delta - \delta(v) - 1)n_1$.

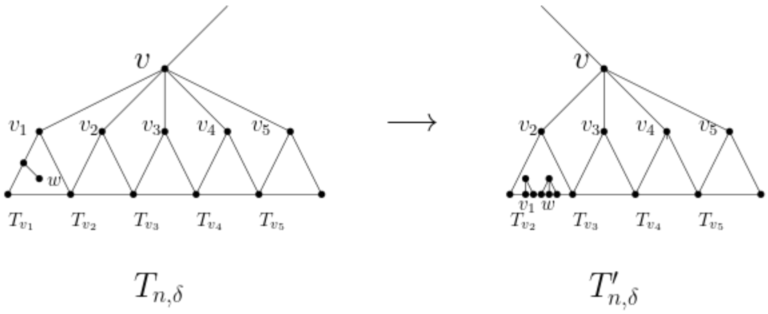


Figure 17: Construction when exist a subtree with a leave of depth different from δ

□

Lemma 2.27. *Let $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that $\forall T'_{n,\delta} \in \mathcal{T}_{n,\delta}$, $S(T_{n,\delta}) \geq S(T'_{n,\delta})$. Let $v \in V(T_{n,\delta}) \setminus L(T_{n,\delta})$ be a binary interior node. Let v_1 and v_2 be the children of v, v_1 and v_2. If $T_{v_1} \neq \{v_1\}$ and $T_{v_2} \neq \{v_2\}$, it is always possible to construct a new tree $\hat{T}_{n,\delta} \in \mathcal{T}_{n,\delta}$ from $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that:*

- *the new subtree $\hat{T}_{v_1}$ of $\hat{T}_{n,\delta}$ is a binary tree with all leaves of depth δ and*
- *the new subtree $\hat{T}_{v_2} \neq \{v_2\}$ of $\hat{T}_{n,\delta}$ is a tree such that $S(T_{n,\delta}) = S(\hat{T}_{n,\delta})$.*

Proof. To construct $\hat{T}_{n,\delta}$, we consider 3 cases:

- T_{v_1} contains a leaf of depth $\alpha < \delta - 1$. By lemma 2.24, this is the only leaf of depth α. Using Lemma 2.25, we can ensure that T_{v_2} has an interior node w of depth α. To construct our new tree $\hat{T}_{n,\delta}$, we remove T_w and we hang it to the leaf of depth α belonging to T_{v_1} see figure 18. This process doesn't generate any leaf of depth $\beta < \alpha$ and removes all leaves of depth $\alpha < \delta - 1$ of T_{v_1}. Finally, as the frequency depth vector of the two trees are the same, we can deduce that $S(T_{n,\delta}) = S(\hat{T}_{n,\delta})$.

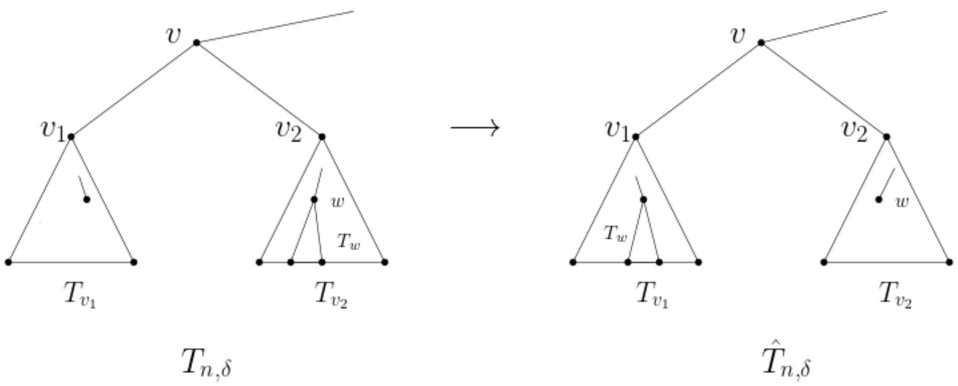


Figure 18: Construction in the case that T_{v_1} contains a leaf of depth $\alpha < \delta - 1$

- T_{v_1} contains a leaf of depth $\delta - 1$. Using lemma 2.24, there are 1 or 2 leaves of depth $\delta - 1$. let us divide this case into two subcases:
  - T_{v_1} has only one leaf of depth $\delta - 1$. Using lemma 2.25, we can deduce that there exists a k-cherry in T_{v_2} with leaves of depth δ. If it was a 2-cherry, we remove it and we hang it to the leaf of T_{v_1} of depth $\delta - 1$, if it was a k-cherry with $k \geq 3$ we remove one leaf of this k-cherry of depth δ and we add a 2-cherry to the leaf of depth $\delta - 1$ of T_{v_1}see figure 19. In the first case, because the depth frequency vector of the two trees are equal, we can deduce that $S(T_{n,\delta}) = S(\hat{T}_{n,\delta})$, and in the second case we have $S(T_{n,\delta}) - S(\hat{T}_{n,\delta}) = \delta + \delta - 1 - 2\delta = -1$. So, $S(T_{n,\delta}) \leq S(\hat{T}_{n,\delta})$ and this cannot happen because by hypothesis $\forall T'_{n,\delta} \in \mathcal{T}_{n,\delta}$, $S(T_{n,\delta}) \geq S(T'_{n,\delta})$.
  - T_{v_1} has exactly two leaves of depth $\delta - 1$. Using lemma 2.24 and lemma 2.25, T_{v_2} has at least two interior nodes of depth $\delta - 1$ because lemma 2.25 implies that T_{v_2} has at least one interior node of depth $\delta - 1$ and lemma 2.24 implies that the brother of the interior node of depth $\delta - 1$ of T_{v_2} can't be a leaf because if it happens, $|S_{\delta-1}| \geq 3$. So, T_{v_2} has 2 k-cherries with leaves of depth δ and we can apply the same process from the previous case twice.

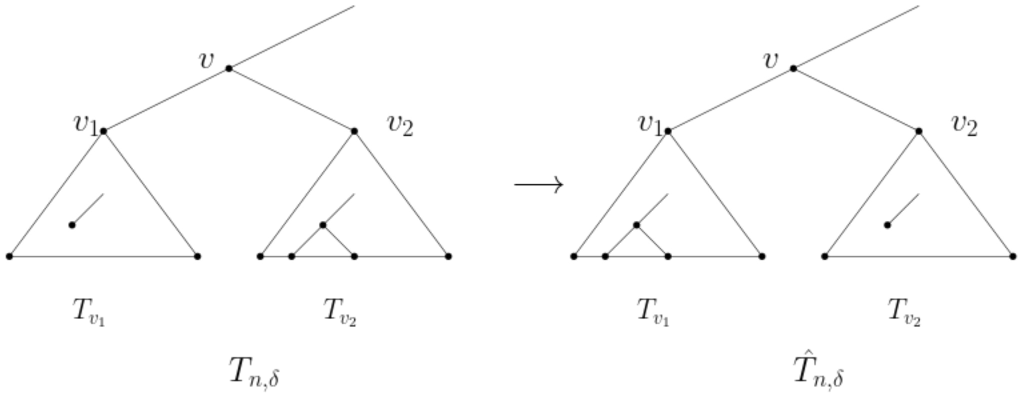


Figure 19: Construction in the case that T_{v_1} contains a leaves of depth $\delta - 1$

- T_{v_1} has a non binary interior node w. Using lemma 2.26, T_w has only leaves of depth δ. So, we can remove the tree generated by a child of v and the remaining leaf, and add the same amount of removed leaves to an interior node of depth $\delta - 1$ of T_{v_2} as children. That will construct a tree $\hat{T}_{n,\delta}$ such that $S(T_{n,\delta}) = S(\hat{T}_{n,\delta})$ because the depth frequency vector of the two trees are the same.

If we apply the previous steps as many times as necessary, the new subtree T_{v_1} will have only binary interior nodes and leaves of depth δ, that is, T_{v_1} is a binary tree where all these leaves have depth δ see figure 20. Moreover, in each step, we generate leaves in the subtree T_{v_2}, so the new $T_{v_2} \neq \{v_2\}$.

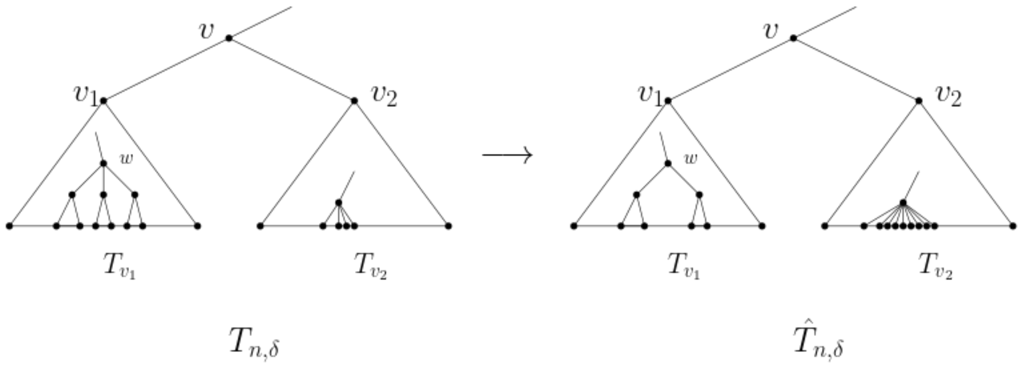


Figure 20: Construction in the case that T_{v_1} contains non binary interior nodes

□

Now we are ready to prove Theorem 2.23:

Proof. From left to right, consider $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ such that $\forall T'_{n,\delta} \in \mathcal{T}_{n,\delta}$, $S(T_{n,\delta}) \geq S(T'_{n,\delta})$. let us prove that $\forall k > \tilde{\delta}$, $|S_{\delta-k}| = 1$. Let v an interior node of depth $\delta - k$, v is binary because if not, using lemma 2.26, T_v has only leaves of depth δ. This implies that $n \geq 2^k + \delta - k$ but that contradicts that $\tilde{\delta}$ is the maximum number that fulfills that condition. There is only one interior node of depth $\delta - k$ because if there were two or more interior nodes of depth $\delta - k$, you can consider the least common ancestor of the nodes and apply lemma 2.27 again to deduce that $n \geq 2^k + \delta - k$ contradicting that $\tilde{\delta}$ is the maximum that the inequality meets. Using lemma 2.24, we have that $|S_{\delta-k}| \leq 1$ but we have shown that there is only one interior node of depth $\delta - k$. Therefore the brother of this interior node has to be a leaf. So, $|S_{\delta-k}| = 1$. The remaining leaves of $T_{n,\delta}$, that is, the leaves with depth other than $\delta - k$, for $k > \tilde{\delta}$ come from the T_v, where v is the only node of depth $\delta - \tilde{\delta} + 1$. We have already proved that v is binary because $\tilde{\delta} + 1 > \tilde{\delta}$. at this point we know for sure that the maximum tree has this structure, see figure 21.

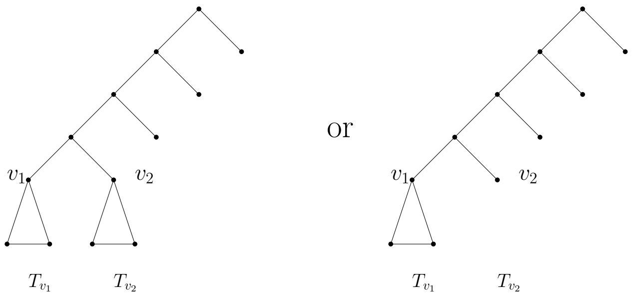


Figure 21: Construction in the case that T_{v_1} contains a leaves of depth $\delta - 1$

Let v_1, v_2 be the children of v, and consider two cases:

1. $T_{v_2} = \{v_2\}$. If this happens, as $T_{n,\delta}$ corresponds to the maximum Sackin index, using theorem 2.22, T_1 has only leaves of depth δ.

2. $T_{v_2} \neq \{v_2\}$. If this happens, we can construct by lemma 2.27 a $\hat{T}_{n,\delta}$ such that the new subtree $\hat{T}_1$ is a binary tree with only leaves of depth δ, and, finally, as $T_{n,\delta}$ corresponds to the maximum Sackin index, the new subtree $\hat{T}_{v_2} = T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}} \in \mathcal{T}_{n-2^{\tilde{\delta}}+\tilde{\delta}-\delta+1,\tilde{\delta}}$ has the maximum Sackin index in the set $\mathcal{T}_{n-2^{\tilde{\delta}}+\tilde{\delta}-\delta+1,\tilde{\delta}}$.

Let $T_{n,\delta}$ be the tree in case 1 and $\hat{T}_{n,\delta}$ be the tree in case 2 where in this case, we assume that we have applied the construction of lemma 2.27. Next, let us compute now $S(T_{n,\delta}) - S(\hat{T}_{n,\delta})$, the difference between the two Sackin index:

$$\begin{aligned} S(T_{n,\delta}) - S(\hat{T}_{n,\delta}) = \; & (n - 2^{\tilde{\delta}} - \delta + \tilde{\delta})\delta + (\delta - \tilde{\delta}) - S(T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}}) \\ & -(n - 2^{\tilde{\delta}} - \delta + \tilde{\delta} + 1)(\delta - \tilde{\delta}) \\ = \; & (n - 2^{\tilde{\delta}} - \delta + \tilde{\delta})\tilde{\delta} - S(T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}}) \end{aligned}$$

So if $(n - 2^{\tilde{\delta}} - \delta + \tilde{\delta})\tilde{\delta} - S(T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}}) \geq 0$, that is, $(n - 2^{\tilde{\delta}} - \delta + \tilde{\delta})\tilde{\delta} \geq S(T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}})$ $T_{n,\delta}$ has the maximum Sackin index, and if $(n - 2^{\tilde{\delta}} - \delta + \tilde{\delta})\tilde{\delta} < S(T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}})$ $\hat{T}_{n,\delta}$ has the maximum Sackin index, as we claim in the Theorem.

From right to left, let $T_{n,\delta} \in \mathcal{T}_{n,\delta}$ with $(n - 2^{\tilde{\delta}} - \delta + \tilde{\delta})\tilde{\delta} \geq S(T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}})$ and suppose that $T_{n,\delta}$ fulfills the first condition of theorem 2.23, since all the trees that satisfy this condition have the same Sackin index, if one of them corresponds to the maximum, all the trees have the maximum Sackin index. In the same way, if $(n - 2^{\tilde{\delta}} - \delta + \tilde{\delta})\tilde{\delta} < S(T_{n-2^{\tilde{\delta}}-\delta+\tilde{\delta}+1,\tilde{\delta}})$ we can apply the same argument as before for the second case.

□

3. Symmetry of rooted trees

In this chapter, we will study a property of the rooted trees that we call the symmetry of a tree, and try to find an answer to the questions what does it mean symmetry? and How can we quantify the symmetry of a tree? First of all we will start by the definition of symmetry. We will understand symmetry as local property of the nodes, so we have to define what does it mean for a node to be symmetric. We will explain this concept into the set of rooted binary trees, and then we will consider some index that computes the symmetry of a given tree.

In order to begin the study of the symmetry in the binary trees, we will make use of the following notation. A rooted binary tree is a rooted tree in which every node different from a leaf has precisely two children. So, the root has degree two and all other internal vertices have degree three. We denote by $\mathcal{T}^2$ the set of binary rooted trees and by $\mathcal{T}_n^2$ the set of binary rooted trees with n leaves. A binary tree is balanced if all its leaves have depth δ. A balanced binary tree has 2^δ leaves. We recall that a 2–caterpillar is a binary tree with precisely one internal node of each depth $i = 1, 2, \dots, \delta - 1$. We observe that the number m of internal vertices of a binary tree $T' \in \mathcal{T}_n^2$ satisfies $m = n - 1$, where n is the number of leaves. This follows by the handshaking Lemma, as

$$2|E(T)| = 2(n + m - 1) = \sum_{x \in V(T')} d(x) = n + 3(m - 1) + 2.$$

Therefore a binary tree with n leaves has $2n - 1$ nodes, $n - 1$ of them are internal.

Now, let us define what we will understand by a symmetric node of a binary rooted tree.

Definition 3.1. Let $T \in \mathcal{T}^2$, $v \notin L(T)$ is a symmetric node if and only if, the two children of v generate isomorphic trees. That is, if v_1,v_2 are the children of v, then $T_{v_1} \simeq T_{v_2}$

Now we have defined what is the symmetry of a node, our intuition make us think that the problem of calculating how symmetric is a tree depends in some way on the symmetric nodes of the tree. As happens with balance it's difficult to define a concept that fits exactly with our intuitive notion of symmetry of a tree so as we did with balance we will say that the 2-caterpillar it's the less symmetric trees among the binary trees because it doesn't have any symmetric node and when the number of leaves is a power of 2 the full balanced tree is the most symmetric because all nodes are symmetric. Now, we can start thinking about how to calculate how symmetric a tree is. So we can define what we will call $\mathcal{I}$, that is, a set of indices with the goal of computing the symmetry of a tree, those are of this form:

$$I_f(T) = \sum_{v \in V(T) \setminus L(T)} f(\delta(v)) S(v),$$

where $T = (V(T), E(T)) \in \mathcal{T}^2$, $f : \mathbb{N} \longrightarrow \mathbb{R}^+$ a decreasing function and we define $S(v)$ as:

$$S(v) = \begin{cases} 1 & \text{if } v \text{ is symmetric} \\ \\ 0 & \text{otherwise} \end{cases}$$

Before proving that this index can fit in some way of our notion of symmetry let us make easier the important proofs by proving first some lemmas. This lemmas will help us to understand the 2-caterpillar and the full balanced tree in terms of their interior nodes.

Lemma 3.2. *Let T be a rooted binary tree. The following are equivalent:*

(i) T is a 2–caterpillar.

(ii) $\delta'(T) = 0$.

(iii) there is an only internal node v such that T_v is a 2–cherry.

(iv) there is an only symmetric node in T.

Proof. The equivalence between (i) and (ii) and the implications (i) $\Rightarrow$ (iii) and (i) $\Rightarrow$ (iv) follow from the definitions.

Suppose that (iii) holds. Then $\delta(v) = \delta - 1$ and there is only an internal node of depth $\delta - 1$. Therefore there is only an internal node of depth $\delta - 2$ and, iteratively, an only internal node of each depth and (i) holds.

Suppose that (iv) holds. Since every 2–cherry in T gives a symmetric node, then T has only one 2–cherry and (iii) holds. $\square$

Lemma 3.3. *A tree $T \in \mathcal{T}_n^2$ with $n = 2^p$ is the full balanced tree if and only if every internal node is symmetric.*

Proof. By induction over p. If $p = 1$ then $n = 2$ and the only tree in $\mathcal{T}_n^2$ is the 2-cherry that contains only one interior node, and this node is symmetric. Let $p > 1$. Then the two children u, u' of the root generate two isomorphic balanced rooted binary trees with 2^{p-1} leaves each. By induction, all nodes in each of the two trees are symmetric. In conclusion, every node is symmetric. $\square$

3.1 Extremal values of the symmetry index

Now, let us prove that for every index I_f the 2-caterpillar is the one with smaller f–symmetry index and, when the number of leaves of the tree is a power of 2, then the full balanced tree corresponds to the maximum.

Lemma 3.4. *Let $T_{n,r} \in \mathcal{T}_{n,r}^2$, then $T_{n,r}$ is 2-caterpillar if, and only if $\forall T'_{n,r} \in \mathcal{T}_{n,r}^2$ and $\forall f : \mathbb{N} \longrightarrow \mathbb{R}^+$ a decreasing function, $I_f(T_{n,r}) \leq I_f(T'_{n,r})$.*

Before doing this proof we will need a simple observation:

Remark 3.5. Let $T_{n,r} \in \mathcal{T}_{n,r}^2$ a 2-caterpillar then $\forall T'_{n,r} \in \mathcal{T}_{n,r}^2$ that is not a 2-caterpillar $\delta(T_{n,r}) > \delta(T'_{n,r})$

The proof is very simple because $T'_{n,r}$ contains 2 nodes that generates a 2-cherry one of them with leaves of depth $\delta(T'_{n,r})$, so you can remove the other one and add a 2-cherry to a leaf of $\delta(T'_{n,r})$ constructing a tree with a higher maximum depth.

Proof. Let $T_{n,r} \in \mathcal{T}_{n,r}^2$ such that is a 2-caterpillar and $I_f \in \mathcal{I}$, let us compute $I_f(T_{n,r})$:

$$I_f(T_{n,r}) = \sum_{v \in V(T_{n,r}) \setminus L(T_{n,r})} f(\delta(v))S(v) = f(\delta(T_{n,r})) \leq \sum_{v \in V(T'_{n,r}) \setminus L(T'_{n,r})} f(\delta(v))S(v) = I_f(T'_{n,r})$$

the inequality is true because the 2-caterpillar has only, as we already proved, one symmetric node, the only one who generates a 2-cherry of the maximum depth that can be reached by an interior node v of depth $\delta(T_{n,r})-1$. So, $S(v)=1$ when v is the unique symmetric node with $\delta(v)=\delta(T_{n,r})-1$. Then consider a tree $T'_{n,r}\in\mathcal{T}^2_{n,r}$ such that $T'_{n,r}$ is not a 2-caterpillar. Because $T'_{n,r}$ is not a 2-caterpillar, there exists $v,w\in V(T'_{n,r})\setminus L(T'_{n,r})$ such that T_v,T_w are 2-caterpillars and therefore this nodes are symmetric. Then, we have f a decreasing function and $\delta(v),\delta(w)\leq\delta(T'_{n,r})<\delta(T_{n,r})$ (The last inequality is true for the remark) $f(\delta(T_{n,r}))<f(\delta(v))+f(\delta(w))$:

$$I_f(T_{n,r})=\sum_{v\in V(T_{n,r})\setminus L(T_{n,r})}f(\delta(v))S(v)=f(\delta(T_{n,r}))<f(\delta(v))+f(\delta(w))\leq I_f(T'_{n,r})$$

Proving that the 2-caterpillar is the unique that corresponds to the minimum in every index of $\mathcal{I}$, so the other implication is also true. □

Lemma 3.6. *Let $T_{n,r}\in\mathcal{T}^2_{n,r}$ with $n=2^p$, if $T_{n,r}$ is the full balanced tree then $\forall T'_{n,r}\in\mathcal{T}^2_{n,r}$ and $\forall f:\mathbb{N}\longrightarrow\mathbb{R}^+$, $I_f(T_{n,r})\geq I_f(T'_{n,r})$*

Before doing this proof we will need a simple observation:

Remark 3.7. Let $T_{n,r}\in\mathcal{T}^2_{n,r}$ with $|V(T_{n,r})\setminus L(T_{n,r})|=p$, then $n=p+1$.

The proof of this fact comes from using that in tree $N-1=M$ where N is the number of nodes and M is the number of edges, and using that for a every graph G $\sum_{v\in V(G)}deg(v)=2M$ we have that:

$$\sum_{v\in V(G)}deg(v)=2M\leftrightarrow 2+3(p-1)+n=2(N-1)\leftrightarrow 2+3(p-1)+n=2(p+n-1)\leftrightarrow n=p+1$$

Proof. Let $T_{n,r}\in\mathcal{T}^2_{n,r}$ with $n=2^p$ the full balanced tree, $I_f\in\mathcal{I}$ and $T'_{n,r}$ binary tree with n leaves, let us compute $I_f(T'_{n,r})$:

$$I_f(T'_{n,r})=\sum_{v\in V(T'_{n,r})\setminus L(T'_{n,r})}f(\delta(v))S(v)\leq\sum_{v\in V(T_{n,r})\setminus L(T_{n,r})}f(\delta(v))=\sum_{v\in V(T_{n,r})\setminus L(T_{n,r})}f(\delta(v))=I_f(T_{n,r})$$

The first inequality is correct because $S(v)$ can only have values 0 or 1 and due to $f>0$ we can deduce that at most we sum $f(\delta(v))$ for each node of the tree, that is when $\forall v\in V(T_{n,r})\setminus L(T_{n,r})S(v)=1$. The second equality is also true because we know that if the number of leaves is fixed then the number of interior nodes is fixed too so considering the most balanced tree we maximize the sum because all the nodes are symmetric. And the final equality is true because we have proved that all nodes of the most balanced binary tree for $n=2^p$ are symmetric so $S(v)=1$ $\forall v\in V(T)\setminus L(T)$. □

As you can see this collection of index allows for every tree compute a number such that we expect that will explain or graduate the symmetry of a binary tree. The most easy example that can come to our minds is only to count the number of symmetric nodes of the tree. This index is one of the already defined taking $f=1$ so this index is $I_1=\sum_{v\in V(T)\setminus L(T)}S(v)$. So now would be interesting to find which are the trees whose symmetry index I_1 is minimum or maximum, let us start first for the easy question the maximum.

Theorem 3.8. *Let $T_{n,r}\in\mathcal{T}^2_{n,r}$, $T_{n,r}$ is a 2-caterpillar if, and only if, $\forall T'_{n,r}\in\mathcal{T}^2_{n,r}$ $I_1(T_{n,r})\leq I_1(T'_{n,r})$.*

Proof. It was already proof for a general decreasing f □

Solving this problem was easy because we have a lemma to find the minimum for an arbitrary n, but to find the maximum we only have a lemma to find it for $n = 2^k$ for some k, so it seems to be more complicated. First let us prove a strong lemma that will help us a lot to find the maximum.

Lemma 3.9. *Let $T_{n,r} \in \mathcal{T}^2_{n,r}$ such that $\forall T'_{n,r} \in \mathcal{T}^2_{n,r}$ $I_1(T_{n,r}) \geq I_1(T'_{n,r})$, then all the nodes that are not symmetric forms a sub-tree in $T_{n,r}$ that only has non-symmetric nodes.*

Proof. We will do the proof by counterpositive, consider a tree $T_{n,r} \in \mathcal{T}^2_{n,r}$ such that the sub-tree generated by all the non-symmetric nodes does not form a sub-tree in $T_{n,r}$ and let us find a tree $\hat{T}_{n,r} \in \mathcal{T}^2_{n,r}$ such that $I_1(\hat{T}_{n,r}) \geq I_1(T_{n,r})$. Now, we know that there exists v, w non symmetric nodes such that the path between them contains a symmetric node, consider s the deepest one. Now consider T_s, observe that v or w belongs to the sub-tree T_s, without lost of generality we can assume that this node is v, now applying that s is symmetric we deduce that exists a node v' non symmetric such that $\delta(v') = \delta(v)$ and the two subtrees generated by the children of v' are isomorphic with a one to one correspondence with the trees generated by the children of v. So, to construct our new tree $\hat{T}_{n,r}$ we do the following:

1. Remove all the subtrees generated by the children of v and v'.

2. We add one pair of isomorphic subtrees to the children of v as a subtrees and the other to the children of v'.

See figure 22

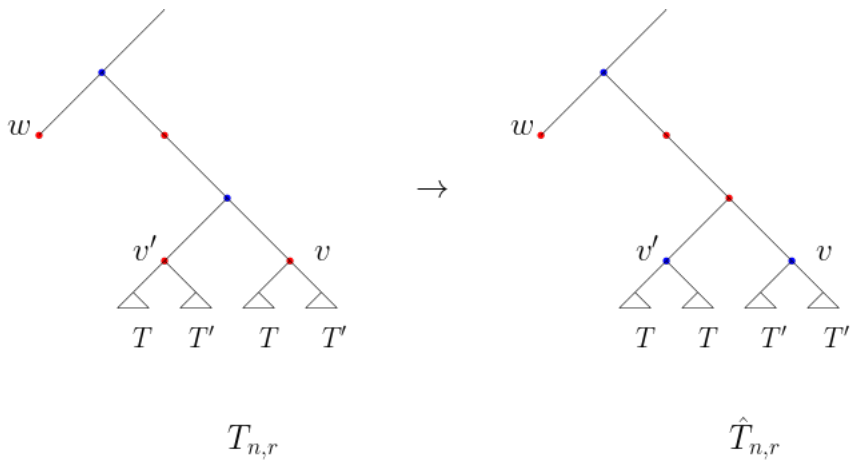


Figure 22: Construction that generates more symmetric nodes, the blues nodes are symmetric and the red nodes are non symmetric n

Notice that using this process we transform s to a non symmetric node and v and v' to symmetric nodes. Therefore, if we don't change the symmetry of any other node we can conclude that $I_1(\hat{T}_{n,r}) \geq I_1(T_{n,r})$. So, now let us prove that the symmetry of any other node have not been changed and if it would have been changed we can generate even more symmetric nodes. let us divide the nodes that can be changed in 3 subsets:

- Nodes t such that $\delta(t) \geq \delta(v)$.

The symmetry of this nodes cannot change by construction because we don't change in any moment the subtrees that generates their children.

- Nodes t such that $\delta(v) \geq \delta(t) \geq \delta(s)$.

  In this case the only nodes that we could change their condition of being symmetric is the ones between the path of v and s or the ones between the path v' and s. Notice that by construction there are no symmetric nodes between v and s, and for s being symmetric there are no symmetric nodes between v' and s. So if some of this nodes changed, they changed from non symmetric to symmetric.

- Nodes t such that $\delta(t) \leq \delta(s)$.

  Notice that if we change the symmetry of a node t, s has to be a descendant of t. In this case, if this construction transforms a symmetric node t into a non symmetric node, as t was symmetric and s is a descendant of t, we know that there exists another s' with the same conditions of s in $T_{n,r}$, so we can apply the same construction again and generate 2 more symmetric nodes, keeping t symmetric.

$\square$

With this lemma we can identify which trees correspond to the maximum. let us explain it in the following theorem:

Theorem 3.10. *Let $T_{n,r} \in \mathcal{T}^2_{n,r}$ such that $\forall T'_{n,r} \in \mathcal{T}^2_{n,r}$ $I_1(T_{n,r}) \geq I_1(T'_{n,r})$, if and only if, all nodes that are not symmetric form a sub-tree in $T_{n,r}$ of $s-1$ nodes where s is the minimum such that $n = \sum_{i=1}^{s} 2^{x_i}$, $x_i \in \mathbb{N}$.*

Proof. From left to right, we use the previous lemma 3.9 to deduce that all nodes that are not symmetric forms a sub-tree in $T_{n,r}$, now we ask ourselves which of all the trees that fulfills this property contains the minimum number of non-symmetric nodes. let us suppose that there exists a tree that fulfills this property with $s' < s$ number of interior non-symmetric nodes, consider the binary tree that generates these nodes. Notice that as the rest of the nodes are symmetric, applying lemma 3.3 we deduce that a full balanced tree are hanging from every leaf of the subtree of non-symmetric nodes considered. So if this subtree, have $s'+1$ interior nodes it means that I can express n as $s'+1 < s+1$ different sums of powers of 2, and this is a contradiction, so the minimum has to be s.

From right to left, we use as similar argument that we use from the Sackin Index, we know that the maximum corresponds with one with s non-symmetric nodes, as the number of interior nodes is determined by n and n is fixed, all the trees with s non-symmetric nodes will be maximum by the I_1 index. See figure 23

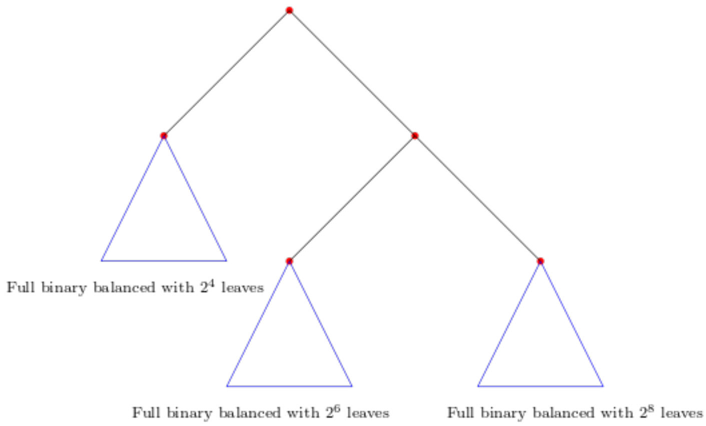


Figure 23: an example of the maximum of the index I_1 with $n = 2^4 + 2^6 + 2^8$, $s = 3$

$\square$

This we have characterized the trees that have minimum I_1 index among the binary trees with n leaves. Notice that this Index doesn't fits exactly with our intuition because f is not a strictly decreasing function, some may think about some examples of f like $f(\delta(v)) = \delta - \delta(v)$ or $f(\delta(v)) = 2^{-\delta(v)}$. The second case is particularly interesting because the root have the same value as a symmetric node as all the nodes of a given depth.

4. Conclusions

In this work, we study many things about the quantitative graph theory related with the rooted trees. It has to be understood as an contribution of the study of the extremal values of balance and symmetry index of the philogenetics trees.

Specifically, we have solved the following problems:

- Find what conditions a specific depth distribution must verify for a rooted tree to represent it.

  The key to solving this problem has been to define an operator on the depth distributions. This operator, applied to a depth distribution, constructs another depth distribution that would correspond to the depths of the subtree obtained by removing the maximum depth leaves from the tree representing the first distribution, In this way an inductive process can be performed and see if the depth distribution of a single tree of depth 1 is obtained. This process translates into formulas that characterize the distributions of depths representable by rooted trees.

- Find the extremal values of the Sackin index for rooted trees by fixing the number of leaves.

  To find the trees that reach the minimum Sackin index within the set of rooted trees of n leaves, we have combined a generalization of the techniques used in the work

  As the caterpillar is the tree that maximizes other balance indices [MRR13], it has become generalized to rooted trees, defining the k-caterpillar tree. By defining the maximum depth δ' for which there are at least 2 interior nodes of this depth, it has been seen that the maximum Sackin index of rooted trees is reached to the 2-caterpillars. The definition of δ' has been a key step as it has allowed us to solve complicated problems in a simpler way, such as characterizing natural numbers that are Sackin indices of rooted trees of n leaves.

- Find the Sackin extremal values of the rooted trees by fixing the number of leaves and the depth of the tree.

  The minimum Sackin index index of the rooted trees fixing the number of leaves and the maximum depth, was very easy because we had a intuition of which tree should be that minimum, but to find the maximum was extremely complicated, and therefore we had to do a previous work of proof and definitions, but finally we could identify which trees corresponds to the maximum Sackin index in this case.

There are some open problems directly related to the contents of this work that we were unable to solve, but we hope to resolve it soon:

- Give a formula for the maximum value of the Sackin index of the rooted trees of n leaves and depth δ. At the end of the previous chapter we have given an algorithm to calculate this maximum value recursively for all values of n, we need to find a closed formula like we did for the minimum or at least a non-recursive algorithm.

- Characterize the natural numbers that are Sackin indices of rooted trees of n leaves and depth δ. The problem is more complicated than in the case without fixing the depth, as we have not been able to find a systematic way to increase the number of leaves from 1 by 1 like we did without fixing the maximum depth.

As future work:

- Study the extremal values of the Sackin index for fixed number of leaf trees, and degree of interior nodes of at least 2 and at most k (instead of no having any condition over this degree).
- Studying how fixing the maximum depth of trees affects the behavior of the extremal values of other balance indices, for example the cophenetic [MRR13].
- Studying more indexes of the set I_f, for example, the extremal values of I_f where $f(\delta(v)) = 2^{-\delta(v)}$

Personally, this work has been a great challenge for two reasons: I have had to solve mathematical problems that have required extensive demonstrations and I have had to learn to ask myself for new problems to solve, which has allowed me to mature mathematically.

References

[CMR19] T. M. Coronado, A. Mir, and F. Rosselló, *A balance index for phylogenetic trees based on rooted quartets*, Journal of Mathematical Biology **79** (2019), no. 3, 1105–1148.

[Col82] D. H. Colless, *Review of "Phylogenetics: The theory and practice of phylogenetic systematics"*, Systematic Zoology **31** (1982), 100–104.

[DES15] M. Dehmer and F. Emmert-Streib, *Quantitative graph theory. mathematical foundations and applications*, Taylor and Francis Group, 2015.

[DESS17] Matthias Dehmer, Frank Emmert-Streib, and Yongtang Shi, *Quantitative graph theory: A new branch of graph theory and network science*, Information Sciences **418** (2017), 575–580.

[Fel04] Joseph Felsenstein, *Inferring phylogenies*, Oxford University Press, 2004.

[Fis18] M. Fischer, *Extremal values of the Sackin balance index for rooted binary trees*.

[KB18] Mahsa Khatibi and Ali Behtoei, *Extremal trees with respect to the Sackin index*, Proceedings of the 3rd International Conference on Combinatorics, Cryptography and Computation, 2018, pp. 276–280.

[Knu97] Donald Ervin Knuth, *The Art of Computer Programming, vol. 3: Sorting and searching*, Pearson Education, 1997.

[MRR13] A. Mir, F. Roselló, and L. Rotger, *A new balance index for phylogenetic trees*, Mathematical Biosciences **241** (2013), no. 1, 125–136.

[MRR18] A. Mir, L. Rotger, and F. Rosselló, *Sound Colless-like balance indices for multifurcating trees*, Plos One (2018).

[Sac72] M. J. Sackin, *"Good" and "bad" phenograms*, Systematic Zoology **21** (1972), no. 2, 225–226.

[SS90] K.T. Shao and R. Sokal, *Tree balance*, Systematic Zoology **39** (1990), 266–276.