# EXTREMAL PROBLEMS FOR TREES WITH GIVEN SEGMENT SEQUENCE

ERIC OULD DADAH ANDRIANTIANA, STEPHAN WAGNER, AND HUA WANG

ABSTRACT. A segment of a tree $T$ is a path whose end vertices have degree 1 or at least 3, while all internal vertices have degree 2. The lengths of all the segments of $T$ form its segment sequence, in analogy to the degree sequence. We address a number of extremal problems for the class of all trees with a given segment sequence. In particular, we determine the extremal trees for the number of subtrees, the number of matchings and independent sets, the graph energy, and spectral moments.

## 1. INTRODUCTION

So-called topological indices are graph invariants that map a graph to a real number, usually serving as descriptors of the graph structure. Throughout the years numerous topological indices have been introduced, motivated by various applications. To give one example, the *Wiener index* of a graph $G$ is defined as the sum of all distances between pairs of vertices in $G$:

$$W(G) = \sum_{\{u,v\}\subseteq V(G)} d_G(u, v),$$

where $d_G(u, v)$ (or simply $d(u, v)$ when there is no ambiguity) is the distance between $u$ and $v$. Several further examples will be given later.

Particular attention has been given to extremal problems. Here, the general question is: given a family of graphs, what can be said about the maximum and minimum values of a given graph invariant and the graphs for which these are attained? To give a simple example, it is known that the maximum and minimum of the Wiener index are attained by the path and the star respectively (see [8, Equation (3)]) if the family of all trees is considered.

A recent paper by Lin and Song [19] considered the family of all trees with a given *segment sequence*, and this family will also be the main object of study in our paper. Let us first define the concept of a segment sequence. We write $P(v, w)$ for the unique path between two vertices $v$ and $w$ of a tree $T$, and $E(P(v, w))$ for its edge set. A *segment* of a tree $T$ is a path $P(v, w)$ in $T$ with the property that each of the ends ($v$ and $w$) is either a leaf or a *branching vertex* (vertex whose degree is at least 3) and that all internal vertices of the path have degree 2. The segment sequence of $T$ is the non-increasing sequence of the lengths of all segments of $T$, see Figure 1 for an example. If $(l_1, l_2, \ldots, l_m)$ is the segment sequence of a tree, then the number of edges is $l_1 + l_2 + \ldots + l_m$, and consequently the number of vertices $l_1 + l_2 + \ldots + l_m + 1$.

Segments can be thought of as building blocks of a tree, and there are, for example, several formulas that allow for the efficient calculation of the Wiener index

2010 *Mathematics Subject Classification.* 05C05; 05C35, 05C69, 05C70.

*Key words and phrases.* trees, segment sequence, subtrees, independent sets, matchings, walks, Merrifield-Simmons index, Hosoya index, graph energy, Estrada index.

This work was supported by grants from the Simons Foundation (#245307) and the National Research Foundation of South Africa (grants 70560 and 96310).

based on segment lengths, see [8, Section 5]. This was also one of the motivations to study extremal problems for trees with given segment sequence.

*Starlike trees* play a special role in this context. For a given segment sequence $(l_1, l_2, \ldots, l_m)$, the starlike tree $S(l_1, l_2, \ldots, l_m)$ is the tree with exactly one vertex of degree $\geq 3$ formed by identifying one end of each of the $m$ segments. It was shown in [19] that $S(l_1, l_2, \ldots, l_m)$ minimizes the Wiener index among all trees with segment sequence $(l_1, l_2, \ldots, l_m)$.

To find the tree with segment sequence $(l_1, l_2, \ldots, l_m)$ that *maximizes* the Wiener index turns out to be somewhat more complicated, and it leads naturally to the notion of *quasi-caterpillars*. A quasi-caterpillar is a tree with the property that all its branching vertices (vertices of degree greater than 2) lie on a path, see Figure 1. It was shown in [4] that the tree maximizing the Wiener index among all trees with a given segment sequence is necessarily always a quasi-caterpillar, answering a question that was posed in [19]. Some further properties of the maximizing quasi-caterpillar were determined in [4] as well.

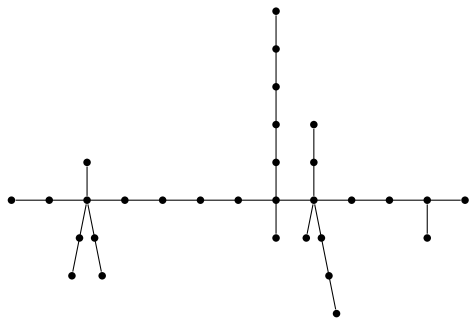


FIGURE 1. A quasi-caterpillar with segment sequence (5,5,3,3,2,2,2,2,2,1,1,1,1,1,1).

Similar questions on trees with given number of segments were also discussed in [19]. It was shown that, among trees with $m$ segments, the Wiener index is minimized by the balanced starlike tree $ST(n, m)$, defined as the unique starlike tree $S(l_1, \ldots, l_m)$ of order $n$ that satisfies $|l_i - l_j| \leq 1$ for all $i, j \in \{1, 2, \ldots, m\}$.

Again, the maximization problem is slightly more complicated and was settled in [4]. For given $n$ and $m$, we define trees $O(n, m)$ (for odd $m$) and $E(n, m)$ (for even $m$) respectively. The graph $O(n, m)$ is obtained from a path $v_0 v_1 \ldots v_\ell$ of length $\ell = n - \frac{m+1}{2}$ by attaching a total of $\frac{m-1}{2}$ leaves to vertices $v_1, v_2, \ldots, v_{\lfloor (m-1)/4 \rfloor}$ and $v_{\ell-1}, v_{\ell-2}, \ldots, v_{\ell - \lceil (m-1)/4 \rceil}$, see Figure 2 (left) for the case $n = 11$, $m = 7$. Note that $O(n, m)$ has exactly $m$ segments.

Likewise, $E(n, m)$ is a tree with $n$ vertices and $m$ segments ($m$ even) obtained from a path $v_0 v_1 \ldots v_\ell$ of length $\ell = n - \frac{m}{2} - 1$ by attaching a total of $\frac{m}{2}$ leaves to vertices $v_1, v_2, \ldots, v_{\lfloor (m-2)/4 \rfloor}$ and $v_{\ell-1}, v_{\ell-2}, \ldots, v_{\ell - \lceil (m-2)/4 \rceil}$, where two leaves are attached to vertex $v_1$ (so that it becomes the only vertex of degree 4), see Figure 2 (right) for the case $n = 11$, $m = 8$.

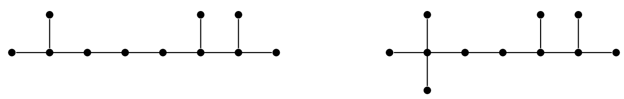


FIGURE 2. The trees $O(11, 7)$ and $E(11, 8)$.

It was proven in [4] that $O(n, m)$ (if $m$ is odd) and $E(n, m)$ (if $m$ is even) always have the greatest Wiener index among trees with $n$ vertices and $m$ segments.

The main aim of this paper is to prove several similar results for other graph invariants. The first of these invariants, and the one we will study most thoroughly, is the *number of subtrees* of a tree $T$, denoted by $F(T)$. The study of extremal problems involving $F(T)$ started in [23,24]. The star was found to have the greatest number of subtrees, while the path has the least number of subtrees. Special families of trees have been considered as well, notably trees with given degree sequence [4,22,30,31], but also others (trees with given number of leaves, bipartition, domination number, etc.) [16].

It is known that a not yet fully understood relation between $F(T)$ and the Wiener index $W(T)$ exists. Indeed, for the family of all trees and many special families of trees (e.g. trees with given degree sequence), it is known that the extremal structure that minimizes $F(T)$ also maximizes $W(T)$ and/or vice versa. The correlation between various graph invariants was studied in [25], and $F(T)$ and $W(T)$ were found to be strongly negatively correlated.

Just like the number of subtrees, most of the other invariants we study in this paper are also based on counting certain substructures. For a rooted tree $B$, let $\mathrm{m}(B,k)$ be the number of matchings of $B$ with cardinality $k$ and $\mathrm{M}(B,x) = \sum_{k\geq 0} \mathrm{m}(B,k)x^k$ for any real $x$ (the matching generating polynomial). The *Hosoya index* $\mathrm{Z}(T)$, which is simply the total number of matchings, is given by

$$\mathrm{Z}(T) = \mathrm{M}(T,1). \tag{1}$$

The *graph energy* $\mathrm{En}(T)$ is defined as the sum of the absolute values of the eigenvalues of a graph. It is strongly related to the Hosoya index by virtue of the remarkable identity [17]

$$\mathrm{En}(T) = \frac{2}{\pi}\int_0^\infty \frac{dx}{x^2}\log \mathrm{M}(T,x^2). \tag{2}$$

The *Merrifield-Simmons* index $\sigma(G)$ is defined as the total number of independent sets of $G$, in analogy to the Hosoya index. Indeed, the Merrifield-Simmons index and the Hosoya index do not only have similar definitions, they are also known to be correlated in a similar way as the Wiener index and the number of subtrees [25].

Extremal problems for the Merrifield-Simmons index and the Hosoya index have been studied vigorously; see [27] for a survey on such questions. Specificlally, several families of trees have been studied in this regard. It is fairly easy to see that the star and the path are extremal again, and extremal structures are also known for instance for trees with given diameter [5, 15, 18, 20, 21], number of leaves [20, 28], maximum degree [12, 26] and degree sequence [2]. The same can be said of the graph energy, see the recent book [17] and the references therein.

Our last section will be concerned with the *number of walks*, which is connected to the spectrum. For a graph $G$ on $n$ vertices with eigenvalues $\lambda_1, \dots, \lambda_n$, the $k$-th spectral moment is defined as

$$C(k,G) = \sum_{i=1}^{n} \lambda_i^k,$$

which is also equal to the trace of the $k$-th power of the adjacency matrix. It is well known [6] that it coincides with the number of closed walks of length $k$. If $G$ is a tree (or any other bipartite graph), then $C(2k+1,G) = 0$ for every integer $k$. Finally, the *Estrada index* of $G$ is defined as

$$EE(G) = \sum_{i=1}^{n} e^{\lambda_i},$$

which can be rewritten as

$$EE(G) = \sum_{k\geq 0} \frac{C(k,G)}{k!} \tag{3}$$

by expanding the exponential function into a power series. See [10] for a survey on the Estrada index. Recent extremal results on the Estrada index of trees under various side conditions can be found in [9, 14, 29]. The approach to the Estrada index via the number of closed walks was recently used in [3] to study trees with given degree sequence, and it will be important in this paper as well.

Before we start with the presentation of our results, let us give a brief outline of this paper. First, we consider subtrees of trees. In Section 2, we first show that the starlike trees maximize the number of subtrees of any order, a rather strong statement that implies the same result for the total number of subtrees. By comparing extremal starlike trees associated with different segment sequences (analogous to what has been done for degree sequences), we show that, among trees with a given number of vertices and segments, the number of subtrees of any given order is maximized when the Wiener index is minimized. In Section 3, we consider the corresponding minimization problem. We will show that the extremal trees, given the segment sequence, are always quasi-caterpillars. Moreover, we prove that $E(n,m)$ and $O(m,n)$ minimize the number of subtrees among trees with a given number of vertices and segments. These results further confirm the interesting negative correlation between the number of subtrees and the Wiener index that was also observed in several other papers.

Thereafter, we focus on the Merrifield-Simmons index (number of independent sets) and the Hosoya index (number of matchings). Once again, starlike trees are found to be extremal. The graph energy, although defined in terms of eigenvalues, is closely related to the Hosoya index in view of the relation (2), hence it is discussed together with those two invariants.

In our final section, we study walks in trees with a given segment sequence or number of segments, and obtain further analogous results. Since walks are closely related to the graph spectrum, we will be able to deduce some results on invariants such as the Estrada index, which are defined in terms of the spectrum, as corollaries.

## 2. Maximum number of subtrees

In this section, we consider the general question of maximizing the number of subtrees of any given order among trees on $n$ vertices with given segment sequence or number of segments. For a positive integer $k$, let $n_k(T)$ denote the number of subtrees of order $k$ in $T$, and let $n_k(T,v)$ denote the number of subtrees of order $k$ in $T$ containing vertex $v$. We first establish the following lemma.

**Lemma 2.1.** *Let $v$ and $w$ be vertices of a tree $T$, such that each internal vertex of the unique path $P_{v,w}$ joining $v$ to $w$ has degree 2 as a vertex of $T$. Let $w_1, w_2, \dots, w_l$ be the neighbors of $w$ that are not on the path $P_{v,w}$. Define*

$$T' = T - ww_1 - ww_2 - \cdots - ww_l + vw_1 + vw_2 + \cdots + vw_l,$$

*as in Figure 3. Then*

$$n_k(T') \geq n_k(T)$$

*for every positive integer $k \leq |T|$, with equality if and only if $v$ or $w$ is a leaf of $T$ (in which case $T$ and $T'$ are isomorphic) or $k \in \{1, 2, |T|\}$.*

*Proof.* Let $T_v$ and $T_w$ be the connected components of $T - E(P_{v,w})$ that contain $v$ and $w$ respectively. All subtrees that contain no edge in $T_v$ or no edge in $T_w$ are preserved by the operation that yields $T'$. A subtree $S$ of $T$ that contains edges of

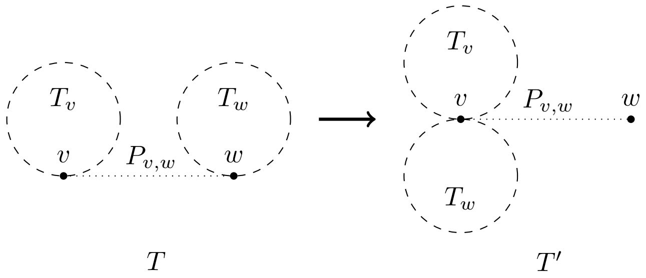



FIGURE 3. Generating $T'$ from $T$.

both $T_v$ and $T_w$ has to contain the entire path $P_{v,w}$, including $v$ and $w$. It is clear that the vertices of $S$ still induce a tree in $T'$. Thus the desired inequality follows immediately.

Conversely, if $k$ is greater than 2 but less than $|T|$, then there are subtrees of $T'$ with $k$ vertices that contain $v$ and edges of both $T_v$ and $T_w$, but not $w$. These subtrees have no corresponding subtrees in $T$, so the inequality is strict for these values of $k$. ∎

Theorem 2.2 below is an immediate consequence of Lemma 2.1.

**Theorem 2.2.** *Among all trees with segment sequence $(l_1, \dots, l_m)$, the number of subtrees of order $k$ is maximized by the starlike tree $S(l_1, \dots, l_m)$. The starlike tree is the unique tree with this property for $k \notin \{1, 2, l_1 + l_2 + \cdots + l_m + 1\}$.*

*Proof.* If $k \in \{1, 2, l_1 + l_2 + \cdots + l_m + 1\}$, then there is nothing to show, since all possible trees have the same number of subtrees with $k$ vertices. Hence assume that $2 < k < l_1 + l_2 + \cdots + l_m + 1$. Let $T$ be a tree with segment sequence $(l_1, \dots, l_m)$ that maximizes $n_k(T)$. It suffices to show that $T$ has exactly one branching vertex (if $m \geq 3$).

Otherwise, let $v$ and $w$ be any two branching vertices with no branching vertex on the path $P_{v,w}$. Applying Lemma 2.1 immediately yields a contradiction. ∎

As a next step, we compare the extremal trees for different segment sequences.

**Definition 1.** *Given two non-increasing sequences $\tau = (l_1, \cdots, l_m)$ and $\tau' = (l'_1, \cdots, l'_m)$, $\tau'$ is said to majorize $\tau$, denoted $\tau \lhd \tau'$, if*

$$\sum_{i=0}^{k} l_i \leq \sum_{i=0}^{k} l'_i \text{ for } k \in \{1, 2, \dots, m-1\}, \text{ and } \qquad \sum_{i=0}^{m} l_i = \sum_{i=0}^{m} l'_i.$$

*We extend the definition to sequences that are not necessarily of the same length by assuming that the shorter sequence is padded with zeros at the end.*

Let us first formulate and prove a technical lemma on subtrees of paths, which in turn implies a second lemma on trees obtained from merging a given tree with a path.

**Lemma 2.3.** *Let $v_1, v_2, \dots, v_n$ be the vertices of a path $P_n$ (in this order). For every positive integer $k \leq n$, we have*

$$n_k(P_n, v_1) \leq n_k(P_n, v_2) \leq \cdots \leq n_k(P_n, v_{\lceil n/2 \rceil}).$$

*Proof.* Note that every subtree of a path is again a path; if it has $k$ vertices, then it consists of vertices $v_i, v_{i+1}, \ldots, v_{i+k-1}$ for some $i$. It follows easily that

$$n_k(P_n, v_j) = \begin{cases} j & j < k, \\ k & k \le j \le n-k+1, \\ n-j+1 & j > n-k+1, \end{cases}$$

if $k \le \frac{n}{2}$, and

$$n_k(P_n, v_j) = \begin{cases} j & j < n-k+1, \\ n-k+1 & n-k+1 \le j \le k, \\ n-j+1 & j > k, \end{cases}$$

if $k > \frac{n}{2}$. The stated inequalities follow in both cases. ∎

**Lemma 2.4.** *Let $T$ be a tree and $v$ one of its vertices. Let $P(n, \ell, T, v)$ denote the graph obtained by identifying $v$ with the $\ell$-th vertex of an $n$-vertex path ($\ell \in \{1, 2, \ldots, n\}$), see Figure 4. The following inequalities hold for all positive integers $k$:*

$$n_k(P(n, 1, T, v)) \le n_k(P(n, 2, T, v)) \le \cdots \le n_k(P(n, \lceil \tfrac{n}{2} \rceil, T, v)). \tag{4}$$

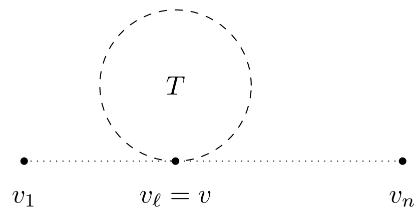



FIGURE 4. Illustration of Lemma 2.4.

*Proof.* Note that the number of $k$-vertex subtrees using only edges of the path or only edges of $T$ is the same for all trees of the form $P(n, \ell, T, v)$. Thus it remains to consider those subtrees that use edges of both $T$ and the path, whose vertices we denote by $v_1, v_2, \ldots, v_n$ as in the previous lemma. The number of $k$-vertex subtrees of $P(n, \ell, T, v)$ that contain edges of the path and $T$ is given by

$$\sum_{r=2}^{k-1} n_r(P_n, v_\ell) n_{k+1-r}(T, v),$$

so the claim follows immediately from Lemma 2.3. ∎

**Theorem 2.5.** *Given two segment sequences $\tau$ and $\tau'$ such that $\tau' \triangleleft \tau$, we have*

$$n_k(S(\tau)) \le n_k(S(\tau'))$$

*for every positive integer $k$.*

*Proof.* The inequality is trivial for $k \in \{1, 2\}$. Hence we assume that $k \ge 3$ for the rest of the proof. We also assume that the two sequences have the same length, padding the shorter sequence with zeros if necessary.

For any two segment sequences $\tau$ and $\tau'$ with $\tau' \triangleleft \tau$, there exists a sequence of segment sequences $\tau^{(i)} = (l_1^{(i)}, \ldots, l_n^{(i)})$ for $1 \le i \le r$, with

$$\tau' = \tau^{(1)} \triangleleft \tau^{(2)} \triangleleft \cdots \triangleleft \tau^{(r-1)} \triangleleft \tau^{(r)} = \tau,$$

where two consecutive sequences $\tau^{(i)}$ and $\tau^{(i+1)}$ differ at exactly two entries, say the $j$-th and $k$-th entry, with $j < k$, in such a way that $l_j^{(i+1)} = l_j^{(i)} + 1$ and

$l_k^{(i+1)} = l_k^{(i)} - 1$. One possibility to construct such a sequence of segment sequences is as follows: to get from $\tau' = \tau^{(1)}$ to $\tau^{(2)}$, we can take $j$ and $k$ to be the first and last index where $\tau$ and $\tau'$ differ; then we repeat the procedure with $\tau^{(2)}$ and $\tau$, etc.

Hence we may assume that $\tau'$ and $\tau$ only differ in two entries, and that the difference is 1 for both of them. Suppose that the segment lengths in $\tau$ that are replaced to obtain $\tau'$ are $p$ and $q$ respectively, where $p \geq q+2$. The new segment lengths in $\tau'$ are $p-1$ and $q+1$. Together, two segments of length $p$ and $q$ in $S(\tau)$ form a path of $p+q+1$ vertices. The starlike graph obtained from $S(\tau)$ by removing the two segments is denoted by $S$, and the center vertex by $v$. Note that $S(\tau)$ and $S(\tau')$ are isomorphic to $P(p+q+1, q+1, S, v)$ and $P(p+q+1, q+2, S, v)$ in the notation of Lemma 2.4, so the desired inequality is in fact a special case of (4). ■

It is easy to see that among all segment sequences of at most $m$ segment lengths with fixed sum of lengths, the unique segment sequence $(l_1, \ldots, l_m)$ with $|l_i - l_j| \leq 1$ for all $1 \leq i, j \leq m$ is majorized by every other segment sequence. From Theorems 2.2 and 2.5, we immediately obtain the following:

**Corollary 2.6.** *Among all trees of order $n$ with at most $m$ segments, the number of subtrees of any fixed order is maximized by the balanced starlike tree $ST(n, m)$.*

Likewise, among all segment sequences with maximum segment length $L$, the sequence $(L, 1, 1, \ldots, 1)$ is clearly majorized by any other segment sequence, which gives us another corollary. Let us write $B(n, L) = S(L, 1, 1 \ldots, 1)$ for the "broom", the starlike tree of order $n$ with one segment of length $L$ and $n - L - 1$ segments of length 1.

**Corollary 2.7.** *Among all trees of order $n$ whose longest segment consists of $L$ edges, the broom $B(n, L)$ has the greatest number of subtrees of any fixed order.*

## 3. Minimum number of subtrees

In this section, we consider the minimization problem for the number of subtrees. The results we obtain parallel those for the Wiener index in [4] and are slightly less precise than those in the previous section.

We will only consider the total number of subtrees, denoted by $F(T)$, rather than the number for each fixed order $k$. We start with the following technical lemma, which is a slight variation of Lemma 3.3 of [24], but follows from exactly the same arguments. It provides some information on how the number of subtrees behaves as branches are moved. In order to formulate it, we have to define one additional quantity: $F_T(v)$ is the number of subtrees of $T$ that contain the vertex $v$.

**Lemma 3.1.** *Given a tree $R$ with vertices $x$ and $y$, and two rooted trees $X$ and $Y$ (with roots $x'$ and $y'$, respectively), let $T$ be obtained by identifying the root of $X$ with $x$ and the root of $Y$ with $y$, and let $T'$ be obtained by identifying the root of $X$ with $y$ and the root of $Y$ with $x$. If*

$$F_R(x) > F_R(y) \text{ and } F_X(x') > F_Y(y'),$$

*then*

$$F(T') < F(T).$$

*Proof.* Note that the number of single-vertex subtrees and the number of subtrees that only contain edges of $R$, only edges of $X$, or only edges of $Y$ is the same in both trees; likewise, the number of subtrees that contain edges of both $X$ and $Y$ (thus necessarily also edges of $R$) is the same in both trees. We have to compare the numbers of other subtrees.

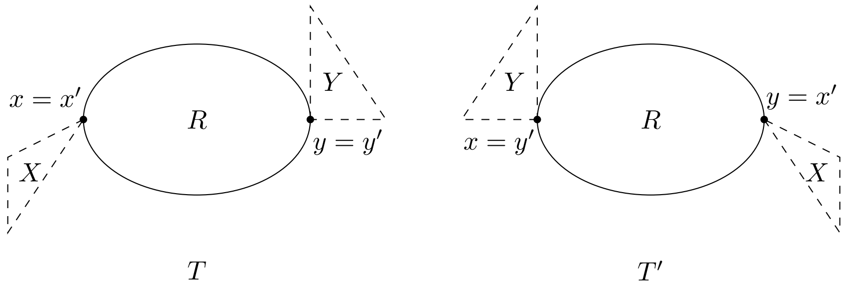



FIGURE 5. Illustration of Lemma 3.1.

- Subtrees that contain edges of $X$ and $R$, but no edges of $Y$: such subtrees necessarily contain $x'$, so their number is $(F_x(R)-1)(F_{x'}(X)-1)$ in $T$ and $(F_y(R)-1)(F_{x'}(X)-1)$ in $T'$.
- Subtrees that contain edges of $Y$ and $R$, but no edges of $X$: such subtrees necessarily contain $y'$, so their number is $(F_y(R)-1)(F_{y'}(Y)-1)$ in $T$ and $(F_x(R)-1)(F_{y'}(Y)-1)$ in $T'$.

So we find that the difference between $F(T)$ and $F(T')$ is

$$\begin{aligned} F(T)-F(T') &= (F_x(R)-F_y(R))(F_{x'}(X)-1)+(F_y(R)-F_x(R))(F_{y'}(Y)-1) \\ &= (F_x(R)-F_y(R))(F_{x'}(X)-F_{y'}(Y)), \end{aligned}$$

which is positive by our assumptions. This completes the proof. ■

3.1. **Trees with given segment sequence.** Let us first show that the minimum number of subtrees among all trees with a given segment sequence is always attained by a quasi-caterpillar. Some further properties of such extremal quasi-caterpillars will be provided before the analogous problem is considered for trees with a given number of segments.

**Theorem 3.2.** *If a tree $T$ minimizes the number of subtrees among all trees with the same segment sequence, then it must be a quasi-caterpillar.*

*Proof.* Let $T$ be such an optimal tree (i.e., $T$ minimizes the number of subtrees), and let $P$ be a path with the greatest possible number of segments on it. Clearly, the two ends of $P$ have to be leaves; we will denote the ends of $P$ by $v_0$ and $v_k$ and the branching vertices on $P$ by $v_1, v_2, \ldots, v_{k-1}$ (in the order of their distances from $v_0$). For each $i$ $(1 \le i \le k-1)$, let the neighbors of $v_i$ that do not lie on $P$ be $v_{i1}, \ldots, v_{il_i}$, and let $T_{ij}$ $(1 \le j \le l_i)$ denote the component containing $v_{ij}$ after removing the edge between $v_i$ and $v_{ij}$.

In each of the subtrees $T_{ij}$, consider the branching vertex (or leaf if there is no branching vertex) closest to $v_i$ and call it $u_{ij}$. Finally, we write $S_{ij}$ for the component containing $u_{ij}$ in $T - E(P(v_i, u_{ij}))$ (Figure 6).

If $S_{ij}$ is a single vertex for every $i$ and $j$, then $T$ is a quasi-caterpillar, and we are done.

Otherwise, let $S = S_{i_0 j_0}$ be such that $F_{S_{i_0 j_0}}(u_{i_0 j_0})$ is maximal, i.e.

$$F_{S_{i_0 j_0}}(u_{i_0 j_0}) \ge F_{S_{ij}}(u_{ij})$$

for every pair $(i,j)$ with $1 \le i \le k-1$, $1 \le j \le l_i$.

Let $T_{\le i_0}$ and $T_{>i_0}$ denote the components of $T - E(P(v_{i_0}, v_{i_0+1}))$ containing $v_{i_0}$ and $v_{i_0+1}$ respectively. The subtrees $T_{<i_0}$ and $T_{\ge i_0}$ are defined analogously; note that $T_{\le i_0} = T_{<i_0+1}$ and $T_{>i_0} = T_{\ge i_0+1}$. Suppose, without loss of generality (which is possible by symmetry), that

$$F_{T_{<i_0}}(v_{i_0-1}) \ge F_{T_{>i_0}}(v_{i_0+1}). \tag{5}$$

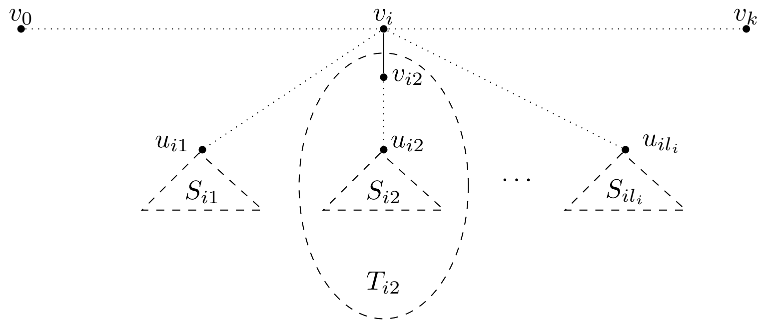



Figure 6. The labeling of $T$.

Moreover, we can assume that $F_S(u_{i_0j_0}) > F_{S_{ij}}(u_{ij})$ for all $i > i_0$ and all $j$. Otherwise we can simply consider, instead of $S$, a subtree $S_{ij}$ with $F_{S_{ij}}(u_{ij}) = F_S(u_{i_0j_0})$ for which the index $i$ is maximal; note that (5) still holds in this case.

By our choice of the path $P = P(v_0, v_k)$ as a path with the greatest possible number of segments on it, $i_0 \neq k-1$, i.e., $v_{i_0}$ cannot be the last branching vertex (since then there would be a path through $u_{i_0j_0}$ rather than $v_k$ that contains more segments). Thus $v_{i_0+1}$ is still a branching vertex, not a leaf.

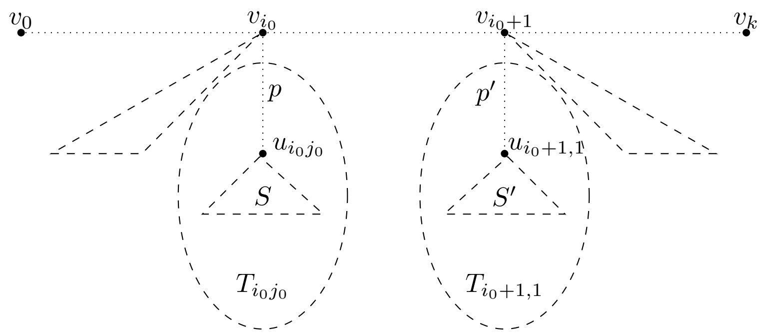



Figure 7. The branches that are switched.

Hence we can consider the subtree $T_{i_0+1,1}$ consisting of the path from $v_{i_0+1,1}$ to $u_{i_0+1,1}$ and the subtree $S' = S_{i_0+1,1}$ (Figure 7). We distinguish two cases, depending on the lengths of the paths $P(v_{i_0}, u_{i_0j_0})$ and $P(v_{i_0+1}, u_{i_0+1,1})$, which we denote by $p$ and $p'$ respectively:

(1) If $p \geq p'$, let $T'$ be obtained from $T$ by switching $T_{i_0j_0}$ and $T_{i_0+1,1}$.
(2) If $p < p'$, let $T'$ be obtained from $T$ by switching $S$ and $S'$.

It is easy to see that in either case $T'$ has the same segment sequence as $T$.

- In the first case, let $R$ be the tree obtained from $T$ by removing $T_{i_0j_0}$ and $T_{i_0+1,1}$. The number of subtrees of $R$ that contain $v_{i_0}$, but not $v_{i_0+1}$, is $F_{R_{\leq i_0}}(v_{i_0}) \cdot d(v_{i_0}, v_{i_0+1})$, since any subtree of $R_{\leq i_0}$ containing $v_{i_0}$ can be augmented by up to $d(v_{i_0}, v_{i_0+1}) - 1$ vertices on the path between $v_{i_0}$ and $v_{i_0+1}$. Similarly, the number of subtrees of $R$ that contain $v_{i_0+1}$, but not $v_{i_0}$, is $F_{R_{\geq i_0+1}}(v_{i_0+1}) \cdot d(v_{i_0}, v_{i_0+1})$. Since we have

$$F_{R_{\leq i_0}}(v_{i_0}) > F_{R_{<i_0}}(v_{i_0-1}) = F_{T_{<i_0}}(v_{i_0-1}) \geq F_{T_{>i_0}}(v_{i_0+1}) > F_{R_{\geq i_0+1}}(v_{i_0+1})$$

by our assumption (5), it follows that

$$F_R(v_{i_0}) > F_R(v_{i_0+1}).$$

Likewise, let $X$ be the tree obtained from $T_{i_0j_0}$ by adding the vertex $v_{i_0}$ and the edge connecting it to $v_{i_0j_0}$, and let $Y$ be the tree obtained from $T_{i_0+1,1}$ by adding the vertex $v_{i_0+1}$ and the edge connecting it to $v_{i_0+1,1}$. We have

$$F_X(v_{i_0}) = p + F_S(u_{i_0j_0}) > p' + F_{S'}(u_{i_0+1,1}) = F_Y(v_{i_0+1}).$$

Now Lemma 3.1 implies that $F(T') < F(T)$, a contradiction.

- In the second case, we apply a similar argument, now to the tree $R'$ that is obtained from $T$ by removing $S_{i_0j_0}$ and $S_{i_0+1,1}$, except for the vertices $u_{i_0j_0}$ and $u_{i_0+1,1}$. A similar argument as before shows that

$$F_{R'}(u_{i_0j_0}) > F_{R'}(u_{i_0+1,1}).$$

Together with the inequality $F_S(u_{i_0j_0}) > F_{S'}(u_{i_0+1,1})$, Lemma 3.1 implies again that $F(T') < F(T)$, and we reach the same contradiction as before.

In both cases, we see that $T$ cannot be optimal, which completes our proof. ■

3.2. **Further characterization of the optimal quasi-caterpillar.** Let the longest path of a quasi-caterpillar containing all the branching vertices be called the *backbone*; all segments that do not lie on the backbone (and thus connect a leaf with a branching vertex) are called "pendant segments". Let a segment sequence $(l_1, l_2, \ldots, l_m)$ be given; we know from Theorem 3.2 that the minimum of the number of subtrees can only be attained for a quasi-caterpillar. In the following we present some further characteristics of extremal quasi-caterpillars.

**Theorem 3.3.** *A quasi-caterpillar $T$ that minimizes the number of subtrees among trees with segment sequence $(l_1, l_2, \ldots, l_m)$ must satisfy the following:*

(1) *If the number of segments is odd, all branching vertices have degree exactly 3. If the number of segments is even, all but one of the branching vertices have degree 3. The only exception must be a branching vertex of degree 4, which must be the first (or last) branching vertex on the backbone. This also means that the number of segments on the backbone is $k = \lfloor (m+1)/2 \rfloor$, the number of pendant segments is $k' = \lceil (m-1)/2 \rceil$.*

(2) *The lengths of the segments on the backbone, listed from one end to the other, form a unimodal sequence $r_1, r_2, \ldots, r_k$, i.e.,*

$$r_1 \le r_2 \le \cdots \le r_j \ge \cdots \ge r_k$$

*for some $j \in \{1, 2, \ldots, k\}$.*

(3) *The lengths of the pendant segments, starting from one end of the backbone towards the other (ordered in the same direction as the $r_i$'s), form a sequence of values $s_1, s_2, \ldots, s_{k-1}$ such that*

$$r_1 \ge s_1 \ge s_2 \ge \cdots \ge s_h \le \cdots \le s_{k'} \le r_k$$

*for some $h \in \{1, 2, \ldots, k'\}$.*

*Proof.* **(1)** Let the backbone be the path $P(v_0, v_k)$ between leaves $v_0$ and $v_k$ with branching vertices $v_1, v_2, \ldots, v_{k-1}$ (in the order of their distances from $v_0$). We also define $T_{<j}$, $T_{>j}$, $T_{\le j}$, $T_{\ge j}$ in the same way as before.

First we claim that there is no branching vertex of degree greater than 4. Otherwise, let $v_i$ be a vertex of degree at least 5 with neighbors $v_{i1}, v_{i2}, v_{i3}, \ldots$ not on $P(v_0, v_k)$. Finally, let $T_{i1}, T_{i2}, T_{i3}$ be the pendant segments at $v_i$ containing $v_{i1}, v_{i2}, v_{i3}$ respectively.

Suppose, without loss of generality, that

$$F_{T_{<i}}(v_{i-1}) \ge F_{T_{>i}}(v_{i+1}).$$

It follows that

$$F_{T_{\leq i}-T_{i1}-T_{i2}}(v_i) > F_{T_{>i}}(v_{i+1}), \tag{6}$$

since for every subtree of $T_{<i}$ that contains $v_{i-1}$, there is a corresponding subtree of $T_{\leq i} - T_{i1} - T_{i2}$ that contains $v_i$, obtained by adding the segment between $v_{i-1}$ and $v_i$.

Let $T'$ be obtained from $T$ by detaching $T_{i1}$ and $T_{i2}$ from $v_i$ and reattaching them to $v_{i+1}$. Note that $T'$ has the same segment sequence as $T$, even if $i = k-1$. Let $R$ be obtained from $T$ by removing $T_{i1}$ and $T_{i2}$. It follows from (6) that

$$F_R(v_i) > F_R(v_{i+1}),$$

by a similar argument as in the proof of the previous theorem. Thus we can apply Lemma 3.1, where we take $X$ to be the tree consisting of $T_{i1}$, $T_{i2}$ and $v_i$, and $Y$ to be a single vertex. We find that the tree $T'$ has a smaller number of subtrees than $T$, which contradicts our assumption on $T$.

Now we know that all branching vertices are of degree 3 or 4. We can repeat the same argument as before with a vertex $v_i$ of degree 4 (moving only one segment instead of two) to obtain a contradiction, unless $v_i = v_1$ or $v_i = v_{k-1}$ (in which case we would have to move a single segment to the end of the backbone, which changes the segment sequence). Thus the only branching vertices that could possibly have degree 4 are $v_1$ and $v_{k-1}$.

Now assume that both $v_1$ and $v_{k-1}$ are vertices of degree 4. Let $S$ and $S'$ be two segments attached to $v_1$ and $v_{k-1}$ respectively, other than those connecting them to other backbone vertices. Moreover, let $R$ be obtained from $T$ by removing $S$ and $S'$ (Figure 8). Suppose, without loss of generality, that

$$F_R(v_1) \leq F_R(v_{k-1}).$$

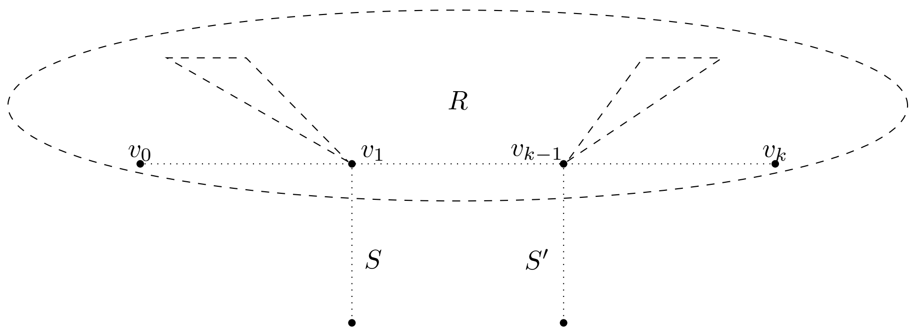



FIGURE 8. The segments $S$ and $S'$ and the rest of the tree (denoted $R$).

Let the lengths of $S$ and $S'$ be $s$ and $s'$ respectively, let $x$ be the length of the segment between $v_0$ and $v_1$ (note that $x \geq s$ by definition, since the path from $v_0$ to $v_k$ was assumed to be longest possible), and let finally $R'$ be obtained from $R$ by removing the segment from $v_0$ to $v_1$. Once again, we construct a tree $T'$ with the same segment sequence as $T$, but fewer subtrees: this tree $T'$ is obtained from $T$ by removing both $S$ and $S'$ and attaching them to $v_0$. Clearly, $T'$ and $T$ have the same segment sequence.

Now we determine the change in the number of subtrees. Note that the subtrees that lie entirely in $R$ or one of the segments $S$ and $S'$ (excluding their endpoints in $R$) remain the same. In $T$, the number of subtrees containing vertices of $R$ as well as either $S$ or $S'$ (but not both) is

$$s \cdot F_R(v_1) + s' \cdot F_R(v_{k-1}).$$

In $T'$, this number is

$$(s+s')\cdot F_R(v_0).$$

The number of subtrees of $T$ that contain vertices of $R$, $S$ and $S'$ is

$$s\cdot s'\cdot F_R(v_1,v_{k-1}),$$

where $F_R(v_1,v_{k-1})$ is the number of subtrees of $R$ containing both $v_1$ and $v_{k-1}$, thus the entire path $P(v_1,v_{k-1})$. On the other hand, the number of such subtrees in $T'$ is

$$s\cdot s'\cdot F_R(v_0).$$

Finally, observe that $F_R(v_0)=x+F_{R'}(v_1)$ and $F_R(v_1)=(x+1)F_{R'}(v_1)$. Putting everything together, we see that

$$\begin{aligned}F(T)-F(T')&=s\cdot F_R(v_1)+s'\cdot F_R(v_{k-1})-(s+s')\cdot F_R(v_0)\\&\quad+s\cdot s'\cdot F_R(v_1,v_{k-1})-s\cdot s'\cdot F_R(v_0)\\&\geq(s+s')F_R(v_1)-(s+s')F_R(v_0)-ss'F_R(v_0)\\&=(s+s')(x+1)F_{R'}(v_1)-(ss'+s+s')(x+F_{R'}(v_1))\\&=(sx+s'x-ss')F_{R'}(v_1)-x(ss'+s+s').\end{aligned}$$

Recall that $x\geq s$. Moreover, one can easily obtain a lower bound for $F_{R'}(v_1)$: every subpath of the path from $v_1$ to $v_k$ that starts at $v_1$ is counted by $F_{R'}(v_1)$. The length of this path is at least $1+d(v_{k-1},v_k)$, which is greater or equal to $1+s'$ by the same argument that gave us the inequality $x\geq s$. Thus there are at least $2+s'$ such subpaths. Since $v_1$ was assumed to have degree 4, it also has at least one more neighbor that neither lies on the backbone nor on $S$. It can be combined with any of the aforementioned subpaths of $P(v_1,v_k)$ to yield a subtree of $R'$ that contains $v_1$. Consequently, we have $F_{R'}(v_1)\geq 2(2+s')$ and thus

$$\begin{aligned}F(T)-F(T')&\geq(sx+s'x-ss')F_{R'}(v_1)-x(ss'+s+s')\\&\geq sxF_{R'}(v_1)-x(ss'+s+s')\\&\geq 2sx(2+s')-x(ss'+s+s')\\&=x(ss'+3s-s')\geq 3sx>0.\end{aligned}$$

We see that $F(T)>F(T')$, so we have a contradiction once again. Thus there is at most one vertex of degree 4, and it has to be either $v_1$ or $v_{k-1}$, if there is such a vertex at all. This happens if and only if the number of segments is even, since the total number of segments is $2k-1$ if all branching vertices have degree 3, and $2k$ if there is a single vertex of degree 4.

**(2)** Consider the segments $P(v_0,v_1),P(v_1,v_2),\ldots,P(v_{k-1},v_k)$ on the backbone, let $r_1,r_2,\ldots,r_k$ be the lengths of these segments, and let $M$ be the maximum length of a backbone segment. Let $j$ be the smallest index such that $r_j=d(v_{j-1},v_j)=M>r_{j+1}=d(v_j,v_{j+1})$. Such an index always exists (if necessary, after reversing the backbone) unless all segments on the backbone have the same length. In this case, however, there is nothing to prove.

Moreover, let $T_{\leq j-1}$, $T_j$ and $T_{\geq j+1}$ denote the components containing $v_{j-1}$, $v_j$ and $v_{j+1}$ respectively in $T-E(P(v_{j-1},v_{j+1}))$ (Figure 9). Consider the tree $R$ consisting of the path from $v_{j-1}$ to $v_{j+1}$ and the tree $T_j$. We have

$$F_R(v_{j-1})=r_j+(1+r_{j+1})F_{T_j}(v_j)<r_{j+1}+(1+r_j)F_{T_j}(v_j)=F_R(v_{j+1}),$$

and it follows that

$$F_{T_{\leq j-1}}(v_{j-1})\geq F_{T_{\geq j+1}}(v_{j+1}),$$

for otherwise interchanging $T_{\leq j-1}$ and $T_{\geq j+1}$ will decrease the number of subtrees (by Lemma 3.1).

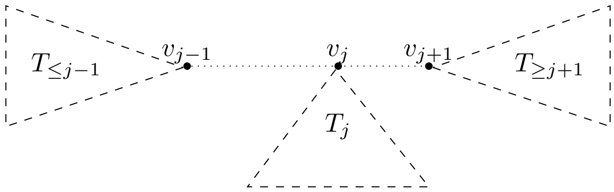



FIGURE 9. The subtrees $T_{\le j-1}$, $T_j$ and $T_{\ge j+1}$.

Consequently, we must also have

$$F_{T_{\le i-1}}(v_{i-1}) > F_{T_{\le j-1}}(v_{j-1}) \ge F_{T_{\ge j+1}}(v_{j+1}) > F_{T_{\ge i+1}}(v_{i+1})$$

for any $i > j$, implying that $r_i \ge r_{i+1}$ by the same argument. It follows that $r_j > r_{j+1} \ge \cdots \ge r_k$. Similarly, one can show that $r_1 \le \cdots \le r_j$.

**(3)** The two inequalities $r_1 \ge s_1$ and $s_{k-1} \le r_k$ follow from the definition of the backbone as longest path.

For the rest of the proof, we only consider the case that all branching vertices have degree 3, the other case being similar. Let $S_i$ denote the pendant segment at $v_i$, and $s_i$ its length. $T_{\le i}$ and $T_{\ge i}$ are defined in the same way as before, and $U_{\le i}$ and $U_{\ge i}$ are obtained from them by removing segment $S_i$ (excluding $v_i$).

Let $\mu$ be the minimum length of all pendant segments, and let $h$ be the smallest index such that $s_h = \mu < s_{h+1}$ (again, such an index exists, if necessary after reversing the backbone, unless all pendant segments have the same length).

In the tree $R$ obtained by removing $S_h$ and $S_{h+1}$ from $T$, we have

$$F_R(v_h) - F_R(v_{h+1}) = r_{h+1} F_{U_{\le h}}(v_h) - r_{h+1} F_{U_{\ge h+1}}(v_{h+1}).$$

Thus $F_{U_{\le h}}(v_h) \ge F_{U_{\ge h+1}}(v_{h+1})$, for otherwise $F_R(v_h) < F_R(v_{h+1})$, and interchanging $S_h$ and $S_{h+1}$ would decrease the number of subtrees by Lemma 3.1. Thus, for any $i > h$,

$$F_{U_{\le i}}(v_i) \ge F_{T_{\le h}}(v_h) > F_{U_{\le h}}(v_h) \ge F_{U_{\ge h+1}}(v_{h+1}) \ge F_{T_{\ge i+1}}(v_{i+1}) > F_{U_{\ge i+1}}(v_{i+1}),$$

which implies that $s_{i+1} \ge s_i$ by the same argument. It follows that $s_h < s_{h+1} \le \cdots \le s_{k-1}$. Similarly, one can show that $s_1 \ge \cdots \ge s_h$. ■

### 3.3. **Trees with a given number of segments.**

**Theorem 3.4.** *Among all trees of order $n$ with $m$ segments, $O(n,m)$ ($E(n,m)$) minimizes the number of subtrees if $m$ is odd (even).*

*Proof.* We only consider the case of odd $m$, the other case is similar. Let $T$ be an optimal tree, given the number of vertices and segments.

From Theorems 3.2 and 3.3, it is clear that $T$ has to be a quasi-caterpillar, and that every branching vertex has degree 3. Let the backbone be the path $P(v_0, v_k)$, and let $v_1, v_2, \ldots, v_{k-1}$ be the branching vertices on the backbone. Note that the total number of segments is $m = 2k - 1$. Moreover, let $a$ and $b$ be the lengths of $P(v_0, v_1)$ and the other pendant segment ending at $v_1$, and let $R$ be the path formed by these two segments. Finally, $X$ is obtained from $T$ by removing the two segments (except for the vertex $v_1$).

Suppose that $\min\{a, b\} > 1$, and let $w$ be a vertex in $R$ that is adjacent to a leaf. We have

$$F_R(v_1) = (a+1)(b+1) > 2(a+b) = F_R(w),$$

so by Lemma 3.1 the number of subtrees decreases if we move $X$ from $v_1$ to $w$ (which amounts to replacing the two segments by segments of length 1 and $a+b-1$ respectively). This contradicts our choice of $T$, so $\min\{a, b\} = 1$. By the results of the previous section, we know more specifically that the pendant segment at $v_1$ (and by the same argument, the pendant segment at $v_{k-1}$) has to have length 1, and by statement (3) of Theorem 3.3, all pendant segments have length 1. In other words, $T$ is a caterpillar.

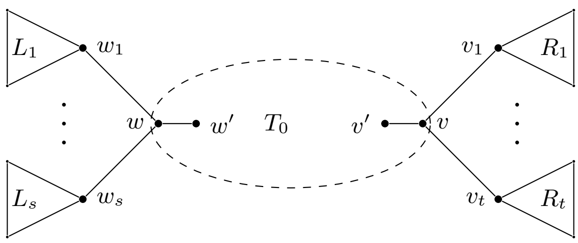
From the (partial) characterization of trees with given degree sequence that minimize the number of subtrees (see [30]), we also know that the degrees of the internal vertices along the backbone have to be decreasing at first, then increasing, i.e., the sequence of degrees has to be of the form $3, 3, \dots, 3, 2, 2, \dots, 2, 2, 3, 3, \dots, 3$. It only remains to show that the number of vertices of degree 3 on the two sides is as equal as possible (difference at most 1). Let us rename the vertices on the backbone as follows: $u_0 = v_0, u_1, u_2, \dots, u_{n-k} = v_k$; this includes all vertices, not just the branching vertices. Assume that there is a leaf attached to $u_1, u_2, \dots, u_x$ and $u_{n-k-1}, u_{n-k-2}, \dots, u_{n-k-y}$, where $x + y = k - 1$. If $k - 1 = n - k - 1$ (equivalently, $n = 2k = m + 1$), there is nothing to prove, as there is only one possibility left for $T$: all vertices on the backbone have to have degree 3. Otherwise, assume that $|x - y| > 1$; without loss of generality, $x > y + 1$. If we move the one leaf from $u_x$ to $u_{n-k-y-1}$, the number of subtrees decreases by a simple application of Lemma 3.1 (taking $R$ to be the path formed by $u_x, u_{x+1}, \dots, u_{n-k-y-1}$ and the leaf adjacent to $u_x$), so we reach yet another contradiction. Thus $|x-y| \leq 1$, which means that $T$ is isomorphic to $O(n, m)$. ■

## 4. Energy, Hosoya index and related graph invariants

In this section, we focus on two invariants based on counting independent sets (Merrifield-Simmons index) and matchings (Hosoya index). In view of the connection (2) between energy and matchings of trees, we can treat the energy along with these two. As it turns out, the starlike trees encountered in Section 2 are extremal again: they maximize the number of independent sets and minimize the number of matchings as well as the energy.

We first require the following lemma, a modified weaker version of the exchange lemma that can be found in [12, 13].

**Lemma 4.1** (cf. [12, 13])**.** *Suppose that a tree $T$ can be decomposed as follows:*




*where $s, t \geq 1$. Define the trees $T_v$ and $T_w$ as follows:*

$$T_v = T - ww_1 - ww_2 - \cdots - ww_s + vw_1 + vw_2 + \cdots + vw_s$$

*and*

$$T_w = T - vv_1 - vv_2 - \cdots - vv_t + wv_1 + wv_2 + \cdots + wv_t.$$

*Then we have*

$$\sigma(T) < \max\{\sigma(T_v), \sigma(T_w)\}$$

*and*

$$\mathrm{M}(T, x) > \min\{\mathrm{M}(T_v, x), \mathrm{M}(T_w, x)\} \tag{7}$$

*for every positive real number $x$.*

We observe that analogous statements hold for the *matching number* $\nu$ and the *independence number* $\alpha$. Recall that the matching number is the greatest cardinality of a matching, and the independence number is the greatest cardinality of an independent set.

**Lemma 4.2.** *For $T, T_v, T_w$ as in Lemma 4.1, we have*

$$\nu(T) \geq \min\{\nu(T_v), \nu(T_w)\} \tag{8}$$

*and*

$$\alpha(T) \leq \max\{\alpha(T_v), \alpha(T_w)\}. \tag{9}$$

*Proof.* First, (8) is actually a consequence of the previous lemma. If $\nu(T) < \nu(T_v)$ and $\nu(T) < \nu(T_w)$, then

$$\lim_{x\to\infty} \frac{\mathrm{M}(T,x)}{\mathrm{M}(T_v,x)} = \lim_{x\to\infty} \frac{\mathrm{m}(T,\nu(T))x^{\nu(T)}}{\mathrm{m}(T_v,\nu(T_v))x^{\nu(T_v)}} = 0$$

and analogously

$$\lim_{x\to\infty} \frac{\mathrm{M}(T,x)}{\mathrm{M}(T_w,x)} = 0,$$

contradicting (7).

To prove (9), we observe that every independent set of $T$ is either an independent set of $T_v$ or an independent set of $T_w$: if $v$ is not contained in it, it is an independent set of $T_v$, and if $w$ is not contained in it, it is an independent set of $T_w$. If both $v$ and $w$ are contained in it, it is an independent set of both $T_v$ and $T_w$. ■

**Remark 1.** *By the König-Egerváry theorem, the sum $\alpha(T) + \nu(T)$ only depends on the number of vertices of $T$, so the two statements of the lemma are actually equivalent.*

If a tree $T$ has two vertices of degree at least 3, then we can apply Lemma 4.1 to obtain a new tree $T'$ (either $T_v$ or $T_w$) with the same segment sequence and more independent sets. The same is true for the number of matchings, and by Lemma 4.2 also for the independence number and matching number (in the latter cases not necessarily with strict inequality). Iterating the argument, we necessarily arrive at a tree that has only one vertex of degree 3 or greater, since the number of such vertices decreases with each step. For any given segment sequence, there is only one such tree, namely the starlike tree $S(l_1, \ldots, l_m)$. Thus we have:

**Theorem 4.3.** *If $T$ is a tree with segment sequence $(l_1, \ldots, l_m)$, then*

$$\sigma(S(l_1,\ldots,l_m)) \geq \sigma(T), \quad \alpha(S(l_1,\ldots,l_m)) \geq \alpha(T), \text{ and } \nu(S(l_1,\ldots,l_m)) \leq \nu(T),$$

*and* $\mathrm{M}(S(l_1,\ldots,l_m),x) \leq \mathrm{M}(T,x)$ *for every positive real number $x$.*

Since the Hosoya index $Z(T)$ equals $\mathrm{M}(T,1)$ and the energy $\mathrm{En}(T)$ is equal to the integral in (2), we obtain the following corollary immediately:

**Corollary 4.4.** *If $T$ is a tree with segment sequence $(l_1, \ldots, l_m)$, then*

$$\mathrm{Z}(S(l_1,\ldots,l_m)) \leq \mathrm{Z}(T)$$

*and*

$$\mathrm{En}(S(l_1,\ldots,l_m)) \leq \mathrm{En}(T).$$

Unfortunately, there is no majorization result analogous to Theorem 2.5 for independent sets and matchings. Given segment sequences $L = (l_1, \ldots, l_m)$ and $H = (h_1, \ldots, h_m)$ such that $L$ majorizes $H$, it is not always true that $\mathrm{M}(S(L),x) \leq \mathrm{M}(S(H),x)$, nor is it always true that $\mathrm{M}(S(L),x) \geq \mathrm{M}(S(H),x)$. The same applies to the Merrifield-Simmons index $\sigma$, see for instance the following examples:

$$\sigma(S(2,2,4)) < \sigma(S(2,3,3)) \text{ and } \sigma(S(1,2,3)) > \sigma(S(2,2,2)),$$

$$\mathrm{M}(S(2,2,4),x) > \mathrm{M}(S(2,3,3),x) \text{ and } \mathrm{M}(S(1,2,3),x) < \mathrm{M}(S(2,2,2),x).$$

Nevertheless, the extremal tree with given number of segments can be determined using established tools: it is no longer the balanced starlike tree, but rather the broom, which consists of a star and a path attached to the center. The following classical lemma, which parallels Lemma 2.4, is needed for this purpose:

**Lemma 4.5** ([11, 26, 32])**.** *Let $G$ be a connected graph with at least two vertices, and $v \in V(G)$. Let $P(n,k,G,v)$ denote the graph obtained by identifying $v$ with the $k$-th vertex of an $n$-vertex path ($k \in \{1,2,\ldots,n\}$), as in Lemma 2.4. The following inequalities hold:*

$$\begin{aligned} &M(P(n,1,G,v),x) > M(P(n,3,G,v),x) > \cdots > M(P(n,2\lfloor \tfrac{n-2}{4} \rfloor + 1,G,v),x) \\ &\quad > M(P(n,2\lfloor \tfrac{n}{4} \rfloor,G,v),x) > \cdots > M(P(n,4,G,v),x) > M(P(n,2,G,v),x) \end{aligned}$$

*for all positive real numbers $x$, and*

$$\begin{aligned} &\sigma(P(n,1,G,v),x) < \sigma(P(n,3,G,v),x) < \cdots < \sigma(P(n,2\lfloor \tfrac{n-2}{4} \rfloor + 1,G,v),x) \\ &\quad < \sigma(P(n,2\lfloor \tfrac{n}{4} \rfloor,G,v),x) < \cdots < \sigma(P(n,4,G,v),x) < \sigma(P(n,2,G,v),x). \end{aligned}$$

**Remark 2.** *The first inequality even holds for all coefficients of $M(.,x)$ (albeit not necessarily with strict inequality), as shown in [11].*

Suppose that a starlike tree $S(l_1,l_2,\ldots,l_m)$ with $m$ segments has at least two segments whose lengths are greater than 1, and let us denote these lengths by $a$ and $b$. Replacing them by segments of length 1 and $a+b-1$ amounts to moving the graph formed by the remaining segments along a path of length $a+b$, as in Lemma 4.5. As the lemma shows, $M(.,x)$ is decreased by this move for every positive $x$, while $\sigma$ is increased. The following result is therefore immediate:

**Theorem 4.6.** *If $T$ is a tree with exactly $m$ segments and $n$ vertices, then we have*

$$\mathrm{M}(T,x) \geq \mathrm{M}(S(n-m,1,1,\ldots,1),x)$$

*for all positive real $x$, and*

$$\sigma(T) \leq \sigma(S(n-m,1,1,\ldots,1)).$$

**Remark 3.** *Lemma 4.5 also shows that $\mathrm{M}(S(n-m,1,1,\ldots,1),x)$ is decreasing in $m$, while $\sigma(S(n-m,1,1,\ldots,1))$ is increasing in $m$. One can therefore also replace "exactly $m$ segments" by "at most $m$ segments" in the formulation of the theorem.*

Finally, we also have the following corollary, in analogy to Corollary 4.4:

**Corollary 4.7.** *If $T$ is a tree with at most $m$ segments and $n$ vertices, then*

$$\mathrm{Z}(T) \geq \mathrm{Z}(S(n-m,1,1,\ldots,1))$$

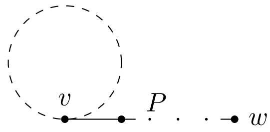
*and*

$$\mathrm{En}(T) \geq \mathrm{En}(S(n-m,1,1,\ldots,1)).$$

## 5. Maximum spectral moment and Estrada index

We follow similar ideas as in [7] to show that starlike trees also maximize the spectral moments. The approach is based on walk enumeration. Let $\mathcal{C}_w(k;T)$ denote the set of closed walks of length $k$ in $T$ which start from $w$. Since there is no closed walk of odd length in a tree, we only need to consider even lengths.

**Lemma 5.1.** *Suppose that a graph $H$ contains a path $P$ whose endpoints are a vertex $w$ of degree 1 and an arbitrary vertex $v$, and whose internal vertices all have degree 2. In other words, $H$ is of the following form:*




*Then there is an injection*

$$\xi : \mathcal{C}_w(2k; H) \longrightarrow \mathcal{C}_v(2k; H)$$

*for any integer* $k \geq 0$.

*Proof.* Let the vertices on the path $P$ be $u_0, u_1, \ldots, u_l$, where $u_0 = v$ and $u_l = w$. To any walk $W = u_{j_0}, u_{j_1}, \ldots, u_{j_r}$ of length $r$ that only uses vertices of $P$, we associate the "mirrored" walk

$$\xi_P(W) = u_{l-j_0}, u_{l-j_1}, \ldots, u_{l-j_r}.$$

The map $\xi_P$ is clearly an involution on the set of walks in $P$ that bijectively maps walks starting at $w$ to walks starting at $v$. Moreover, it maps closed walks to closed walks.

Now let us define a map $\xi : \mathcal{C}_w(2k; H) \longrightarrow \mathcal{C}_v(2k; H)$:

- For a walk $W \in \mathcal{C}_w(2k; P - v)$ (i.e., a closed walk of length $2k$ starting at $w$ that never reaches $v$), we set $\xi(W) = \xi_P(W)$.
- Now let $W = z_0, z_1, \ldots, z_{2k} \in \mathcal{C}_w(2k; H)$ be a closed walk starting at $w$ that reaches $v$ at some point. Let $i$ be the smallest index such that $z_i = v$. Then we set

$$\xi(W) = z_i, z_{i+1}, \ldots, z_{2k-1}, \xi_P(z_i, z_{i-1}, \ldots, z_0).$$

  In words, we remove the initial part up to the first time that $v$ is reached, reverse its direction (turning it into a walk from $v$ to $w$ that does not visit $v$ again), then take its mirror image (which yields a walk from $w$ to $v$ that never visits $w$ again) and append it at the end, which is possible since $z_{2k} = w$.

It is easy to see that $\xi$ is indeed injective: if $W' = \xi(W)$ lies entirely in $P$ and never visits $w$, then we have $W = \xi^{-1}(W') = \xi_P^{-1}(W') = \xi_P(W')$, i.e., the preimage is obtained by reflection. Otherwise, $W' = \xi(W) = z'_0, z'_1, \ldots, z'_{2k}$ has to visit $w$ at least once, and we can consider the largest index $j$ such that $z'_j = w$. The walk $z'_j, z'_{j+1}, \ldots, z'_{2k}$ must be the final part that is appended at the end of the construction of $\xi$. The original walk $W$ must therefore be

$$\xi_P(z'_{2k}, z'_{2k-1}, \ldots, z'_j) z'_1, z'_2, \ldots, z'_j.$$

■

Next we provide a generalization of Lemma 2 in [7], which parallels our Lemma 2.1.

**Lemma 5.2.** *Let* $v$ *and* $w$ *be vertices of a graph* $G$ *that consists of two connected graphs* $G_v$ *and* $G_w$ *and a path* $P_{v,w}$ *joining two vertices* $v$ *and* $w$ *in* $G_v$ *and* $G_w$ *respectively. Let* $w_1, w_2, \ldots, w_l$ *be the neighbors of* $w$ *that are not on the path* $P_{v,w}$*, and define*

$$G' = G - ww_1 - ww_2 - \cdots - ww_l + vw_1 + vw_2 + \cdots + vw_l,$$

*as in Lemma 2.1, cf. Figure 3. Then*

$$C(2k, G) \leq C(2k, G')$$

*for every integer* $k \geq 0$. *The inequality is strict if* $k \geq 2$ *and both* $G_v$ *and* $G_w$ *consist of more than one vertex.*

*Proof.* Let $H$ be obtained from $G$ by removing $G_w$, except for the vertex $w$. Note that $H$ has the required shape for Lemma 5.1. Define $\xi$ as in that lemma; with some minor abuse of notation, we suppress the dependence of $\xi$ on $k$.

We only have to compare the number of walks using edges of both $G_v$ and $G_w$. Let the set of closed walks of length $2k$ with this property in $G$ be denoted by $\mathcal{C}'(2k, G)$; $\mathcal{C}'(2k, G')$ is defined analogously. We define an injection $\eta : \mathcal{C}'(2k, G) \longrightarrow \mathcal{C}'(2k, G')$ as follows.

Given a walk

$$W = w_0, w_1, \ldots, w_{2k},$$

split it into walks that alternatingly lie entirely in $G_w$ and $H$. Those pieces that lie in $G_w$ are kept the same, while the pieces in $H$ are closed walks starting and ending at $w$; they are mapped by $\xi$ to walks starting and ending at $v$ that take their place in $\eta(W)$. The only possible exception are the first and last piece; if they lie in $G_w$, they are kept as well. If however they lie in $H$, we have to combine them to a closed walk first: let these pieces be $w_0, w_1, \ldots, w_{j_0}$ and $w_{j_1}, w_{j_1+1}, \ldots, w_{2k}$ respectively, where necessarily $w_{j_0} = w_{j_1} = w$. The walk

$$w_{j_1}, w_{j_1+1}, \ldots, w_{2k} = w_0, w_1, \ldots, w_{j_0}$$

must then be an element of $\mathcal{C}_w(j_0 + 2k - j_1, H)$, so $\xi$ maps it to a closed walk

$$z_{j_1}, z_{j_1+1}, \ldots, z_{2k} = z_0, z_1, \ldots, z_{j_0-1}, z_{j_0}$$

with $z_{j_0} = z_{j_1} = v$. Now we split it into the two walks $z_{j_1}, z_{j_1+1}, \ldots, z_{2k}$ and $z_0, z_1, \ldots, z_{j_0-1}, z_{j_0}$ and replace $w_1, w_2, \ldots, w_{j_0}$ and $w_{j_1}, w_{j_1+1}, \ldots, w_{2k}$ by them to complete the construction of $\eta(W)$.

Since $\xi$ as constructed in Lemma 5.1 is injective, so is $\eta$, which proves the desired inequality. The inequality is strict if $k \geq 2$ and both $G_v$ and $G_w$ contain at least one edge, because then $\mathcal{C}'(2k, G')$ contains walks that do not use edges of $P_{v,w}$ while $\eta(W)$ uses at least one edge of the path $P_{v,w}$ for any given $W$. ∎

Now we obtain the main theorem of this section:

**Theorem 5.3.** *If $T$ is a tree with segment sequence $(l_1, \ldots, l_m)$, then*

$$C(2k, S(l_1, \ldots, l_m)) \geq C(2k, T)$$

*for any integer $k \geq 0$.*

*Furthermore, the inequality is strict if $T \neq S(l_1, \ldots, l_m)$ and $k \geq 2$.*

*Proof.* This follows from Lemma 5.2 in exactly the same way as Theorem 2.2 followed from Lemma 2.1. ∎

In view of the formula (3) that relates the Estrada index to spectral moments, it follows immediately that the same conclusion holds for the Estrada index:

**Corollary 5.4.** *If $T$ is a tree with segment sequence $(l_1, \ldots, l_m)$, then*

$$EE(S(l_1, \ldots, l_m)) \geq EE(T).$$

*Furthermore, the inequality is strict if $T \neq S(l_1, \ldots, l_m)$.*

In order to compare trees with different segment sequences, we make use of the following lemma taken from [1], which is analogous to Lemma 2.4 and Lemma 4.5.

**Lemma 5.5.** *Let $T$ be a tree and $v$ one of its vertices. Let $P(n, \ell, T, v)$ denote the graph obtained by identifying $v$ with the $\ell$-th vertex of an $n$-vertex path ($\ell \in \{1, 2, \ldots, n\}$), as in Lemma 2.4. The following inequalities hold for all $k \geq 1$:*

$$C(2k, P(n, 1, T, v)) \leq C(2k, P(n, 2, T, v)) \leq \cdots \leq C(2k, P(n, \lceil \tfrac{n}{2} \rceil, T, v)). \quad (10)$$

The following majorization theorem now follows in exactly the same way from Lemma 5.5 as Theorem 2.5 followed from Lemma 2.4.

**Theorem 5.6.** *Given two segment sequences $\tau$ and $\tau'$ such that $\tau' \lhd \tau$, we have*

$$C(2k, S(\tau)) \leq C(2k, S(\tau'))$$

*for every integer $k \geq 0$.*

Once again, we obtain a number of corollaries. It was already shown in [1] that the balanced starlike trees $ST(n,m)$ maximize all even spectral moments and thus the Estrada index among all starlike trees with a given center degree (equivalently, number of segments). Combining Theorem 5.3 and Theorem 5.6, we get the following (in analogy to Corollary 2.6):

**Corollary 5.7.** *If $T$ is a tree with at most $m$ segments and $n$ vertices, then*

$$C(2k, T) \leq C(2k, ST(n,m))$$

*for every integer $k \geq 0$, and hence*

$$EE(T) \leq EE(ST(n,m)).$$

Similarly, the analog of Corollary 2.7 reads as follows:

**Corollary 5.8.** *If $T$ is a tree of order $n$ whose longest segment consists of $L$ edges, then*

$$C(2k, T) \leq C(2k, B(n,L))$$

*for every integer $k \geq 0$, and hence*

$$EE(T) \leq EE(B(n,L)).$$

## References


[1] I. Aleksandar and S. Dragan. The Estrada index of chemical trees. *J. Math. Chem.*, 47:305–314, 2010.

[2] E. O. D. Andriantiana. Energy, Hosoya index and Merrifield-Simmons index of trees with prescribed degree sequence. *Discrete Appl. Math.*, 161(6):724–741, 2013.

[3] E. O. D. Andriantiana and S. Wagner. Spectral moments of trees with given degree sequence. *Linear Algebra Appl.*, 439(12):3980–4002, 2013.

[4] E. O. D. Andriantiana, S. Wagner, and H. Wang. Maximum Wiener index of trees with given segment sequence. *MATCH Commun. Math. Comput. Chem.*, 75(1):91–104, 2016.

[5] Y. Chen and S. Wen. On the total number of matchings of trees with prescribed diameter. *Int. J. Nonlinear Sci.*, 4(1):37–43, 2007.

[6] D. Cvetković, M. Doob, and H. Sachs. *Spectra of graphs – Theory and application*. Johann Ambrosius Barth Verlag, Heidelberg, Leipzig, 3rd edition, 1995.

[7] H. Deng. A proof of a conjecture on the Estrada index. *MATCH Commun. Math. Comput. Chem.*, 62:599–606, 2009.

[8] A. Dobrynin, R. Entringer, and I. Gutman. Wiener index of trees: Theory and applications. *Acta Appl. Math.*, 66:211–249, 2001.

[9] Z. Du and B. Zhou. The Estrada index of trees. *Linear Algebra Appl.*, 435(10):2462–2467, 2011.

[10] I. Gutman, H. Deng, and S. Radenković. The Estrada index: an updated survey. *Zb. Rad. (Beogr.)*, 14(22)(Selected topics on applications of graph spectra):155–174, 2011.

[11] I. Gutman and F. J. Zhang. On the ordering of graphs with respect to their matching numbers. *Discrete Appl. Math.*, 15(1):25–33, 1986.

[12] C. Heuberger and S. Wagner. Maximizing the number of independent subsets over trees with bounded degree. *J. Graph Theory*, 58(1):49–68, 2008.

[13] C. Heuberger and S. Wagner. Chemical trees minimizing energy and Hosoya index. *J. Math. Chem.*, 46:214–230, 2009.

[14] A. Ilić and D. Stevanović. The Estrada index of chemical trees. *J. Math. Chem.*, 47(1):305–314, 2010.

[15] A. Knopfmacher, R. F. Tichy, S. Wagner, and V. Ziegler. Graphs, partitions and Fibonacci numbers. *Discrete Appl. Math.*, 155(10):1175–1187, 2007.

[16] S. Li and S. Wang. Further analysis on the total number of subtrees of trees. *Electron. J. Combin.*, 19(4):Research Paper 48, 14 pp., 2012.
[17] X. Li, Y. Shi, and I. Gutman. *Graph energy.* Springer, New York, 2012.
[18] X. Li, H. Zhao, and I. Gutman. On the Merrifield-Simmons index of trees. *MATCH Commun. Math. Comput. Chem.*, 54(2):389–402, 2005.
[19] H. Lin and M. Song. On segment sequences and the Wiener index of trees. *MATCH Commun. Math. Comput. Chem.*, 75(1):81–89, 2016.
[20] X.-F. Pan, J.-M. Xu, C. Yang, and M.-J. Zhou. Some graphs with minimum Hosoya index and maximum Merrifield-Simmons index. *MATCH Commun. Math. Comput. Chem.*, 57(1):235–242, 2007.
[21] A. S. Pedersen and P. D. Vestergaard. An upper bound on the number of independent sets in a tree. *Ars Combin.*, 84:85–96, 2007.
[22] A. V. Sills and H. Wang. The minimal number of subtrees of a tree. *Graphs Combin.*, 31(1):255–264, 2015.
[23] L. A. Székely and H. Wang. On subtrees of trees. *Adv. in Appl. Math.*, 34(1):138–155, 2005.
[24] L. A. Székely and H. Wang. Binary trees with the largest number of subtrees. *Discrete Appl. Math.*, 155:374–385, 2007.
[25] S. Wagner. Correlation of graph-theoretical indices. *SIAM Journal on Discrete Mathematics*, 21:33–46, 2007.
[26] S. Wagner. Extremal trees with respect to Hosoya index and Merrifield-Simmons index. *MATCH Commun. Math. Comput. Chem.*, 57(1):221–233, 2007.
[27] S. Wagner and I. Gutman. Maxima and minima of the Hosoya index and the Merrifield-Simmons index: a survey of results and techniques. *Acta Appl. Math.*, 112(3):323–346, 2010.
[28] A. Yu and X. Lv. The Merrifield-Simmons indices and Hosoya indices of trees with $k$ pendant vertices. *J. Math. Chem.*, 41(1):33–43, 2007.
[29] J. Zhang, B. Zhou, and J. Li. On Estrada index of trees. *Linear Algebra Appl.*, 434(1):215–223, 2011.
[30] X.-M. Zhang and X.-D. Zhang. The minimal number of subtrees with a given degree sequence. *Graphs Combin.*, 31:309–318, 2015.
[31] X.-M. Zhang, X.-D. Zhang, D. Gray, and H. Wang. The number of subtrees of trees with given degree sequence. *J. Graph Theory*, 73(3):280–295, 2013.
[32] H. Zhao and X. Li. On the Fibonacci numbers of trees. *Fibonacci Quart.*, 44(1):32–38, 2006.

Eric Ould Dadah Andriantiana, Department of Mathematics (Pure and Applied), Rhodes University, PO Box 94, 6140 Grahamstown, South Africa
*E-mail address*: E.Andriantiana@ru.ac.za

Stephan Wagner, Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa
*E-mail address*: swagner@sun.ac.za

Hua Wang, Department of Mathematical Sciences, Georgia Southern University, Statesboro, GA 30460, USA
*E-mail address*: hwang@georgiasouthern.edu