

Greedy trees, subtrees and antichains

Eric Ould Dadah Andriantiana

Department of Mathematical Sciences
Stellenbosch University
Private Bag X1, Matieland 7602, South Africa
`ericoda@sun.ac.za`

Stephan Wagner

Department of Mathematical Sciences
Stellenbosch University
Private Bag X1, Matieland 7602, South Africa
`swagner@sun.ac.za`

Hua Wang

Department of Mathematical Sciences
Georgia Southern University
Statesboro, GA 30460, USA
`hwang@georgiasouthern.edu`

Submitted: Jan 28, 2013; Accepted: Aug 20, 2013; Published: Aug 30, 2013

Mathematics Subject Classifications: 05C05, 05C35

Abstract

Greedy trees are constructed from a given degree sequence by a simple greedy algorithm that assigns the highest degree to the root, the second-, third-, ... highest degrees to the root's neighbors, and so on.

They have been shown to maximize or minimize a number of different graph invariants among trees with a given degree sequence. In particular, the total number of subtrees of a tree is maximized by the greedy tree. In this work, we show that in fact a much stronger statement holds true: greedy trees maximize the number of subtrees of any given order. This parallels recent results on distance-based graph invariants.

We obtain a number of corollaries from this fact and also prove analogous results for related invariants, most notably the number of antichains of given cardinality in a rooted tree.

1 Introduction and statement of results

The question of finding extremal graph structures that maximize or minimize a certain graph invariant is a recurring theme throughout graph theory. For trees, the so-called greedy trees have been shown to be extremal among trees with a given degree sequence with respect to many graph invariants such as the Wiener index (sum of all distances) [17, 20] and related distance-based invariants [11], the spectral radius [2] and Laplacian spectral radius [1, 19], etc. These trees are constructed from a given degree sequence by a

simple greedy algorithm that assigns the highest degree to the root, the second-, third-, ... highest degrees to the root's neighbors, and so on – a formal definition will be given later in this section.

The number of subtrees of a tree, being an interesting topic on its own mathematical right [12, 18], also plays a role in other fields such as phylogenetic reconstruction [7]. In many questions concerning subtrees of trees, the number of subtrees of a specific order plays an important role. For instance, the average subtree order [4, 5, 16] and the subtree poset [3, 14, 15] have been considered in recent works. The concept of a subtree polynomial akin to the matching polynomial, the independence polynomial and other polynomials associated to a graph, has been brought forward as well [4]: if $n_k(T)$ is the number of subtrees of order k in a tree T of order n , then the associated polynomial is

$$\Phi_T(x) = \sum_{k=0}^n n_k(T)x^k.$$

More generally, a weighted version (also including edge weights) is studied in [18], and a bivariate version, where a second variable marks the number of leaves, is considered in [10].

The greedy trees were shown in [22] to have the maximum number of subtrees among all trees with given degree sequence. The analogous problem for the minimum was studied recently in [21]. Further recent results on maxima and minima of the number of subtrees under various restrictions can be found in [8, 9]. The main result of this paper is the fact that the greedy tree not only maximizes the total number of subtrees (in terms of the aforementioned polynomial, the value $\Phi_T(1)$), but actually *every single coefficient*, i.e., the number of subtrees of any given order. A similar result was achieved recently for distance-based graph invariants: in [11] it was shown that the number of pairs of vertices whose distance is at most k is maximized by the greedy tree (given the degree sequence) for every k .

To formalize our results, we start with a few definitions.

Definition 1. Let F be a rooted forest where the maximum height of any component is $k - 1$. The *leveled degree sequence* of F is the sequence

$$D = (V_1, \dots, V_k), \tag{1}$$

where, for any $1 \leq i \leq k$, V_i is the non-increasing sequence formed by the degrees of the vertices of F at the i^{th} level (i.e., vertices of distance $i - 1$ from the root in the respective component).

For convenience, we denote the “number of levels” in D by $L(D)$ (maximum height plus one), evidently $L(D) = k$ in (1).

The greedy trees have been defined in various equivalent ways in previous works [2, 11, 17, 19]. For our purposes, we begin with the definitions of level greedy trees and forests.

Definition 2. The *level greedy forest* with leveled degree sequence

$$D = ((i_{1,1}, \dots, i_{1,k_1}), (i_{2,1}, \dots, i_{2,k_2}), \dots, (i_{n,1}, \dots, i_{n,k_n})) \tag{2}$$

is obtained using the following “greedy algorithm”:

- (i) Label the vertices of the first level by $g_1^1, \dots, g_{k_1}^1$, and assign degrees to these vertices such that $\deg g_j^1 = i_{1,j}$ for all j .
- (ii) Assume that the vertices of the h^{th} level have been labeled $g_1^h, \dots, g_{k_h}^h$ and a degree has been assigned to each of them. Then for all $1 \leq j \leq k_h$ label the neighbors of g_j^h at the $(h+1)^{\text{th}}$ level, if any, by

$$g_{1+\sum_{m=1}^{j-1}(i_{h,m}-1)}^{h+1}, \dots, g_{\sum_{m=1}^j(i_{h,m}-1)}^{h+1},$$

and assign degrees to the newly labeled vertices such that $\deg g_j^{h+1} = i_{h+1,j}$ for all j .

The level greedy forest with leveled degree sequence D is denoted by $G(D)$ (Figure 1).

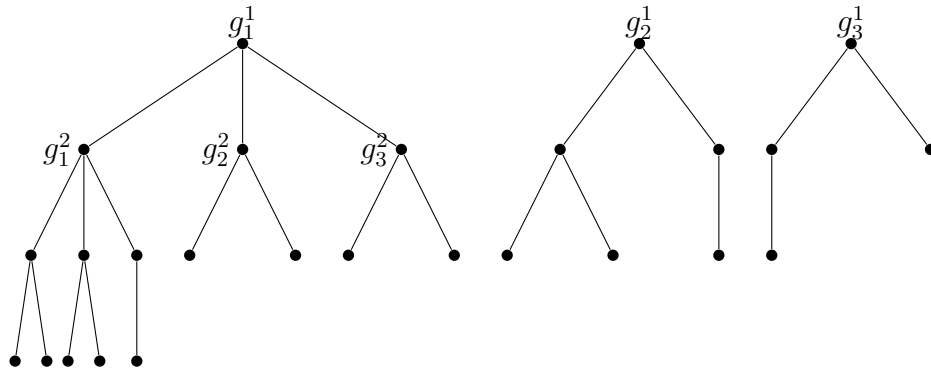


Figure 1: A level greedy forest (only the labels of the first six vertices are shown).

Definition 3. A connected level greedy forest is called a *level greedy tree*.

In analogy to (rooted) level greedy trees, we will also need an edge-rooted version:

Definition 4. The *edge-rooted* level greedy tree with leveled degree sequence

$$D = ((i_{1,1}, i_{1,2}), (i_{2,1}, \dots, i_{2,k_2}), \dots, (i_{n,1}, \dots, i_{n,k_n}))$$

is obtained from the two-component level greedy forest with leveled degree sequence

$$((i_{1,1} - 1, i_{1,2} - 1), (i_{2,1}, \dots, i_{2,k_2}), \dots, (i_{n,1}, \dots, i_{n,k_n}))$$

by joining the two roots.

Finally, we define greedy trees and greedy forests:

Definition 5. If a root in a tree can be chosen such that it becomes a level greedy tree whose leveled degree sequence, as given in (2), satisfies

$$\min(i_{j,1}, \dots, i_{j,k_j}) \geq \max(i_{j+1,1}, \dots, i_{j+1,k_{j+1}})$$

for all $1 \leq j \leq n-1$, then it is called a *greedy tree* (Figure 2). In the case that D is a degree sequence (as opposed to a leveled degree sequence), we use $G(D)$ to denote the greedy tree with degree sequence D .

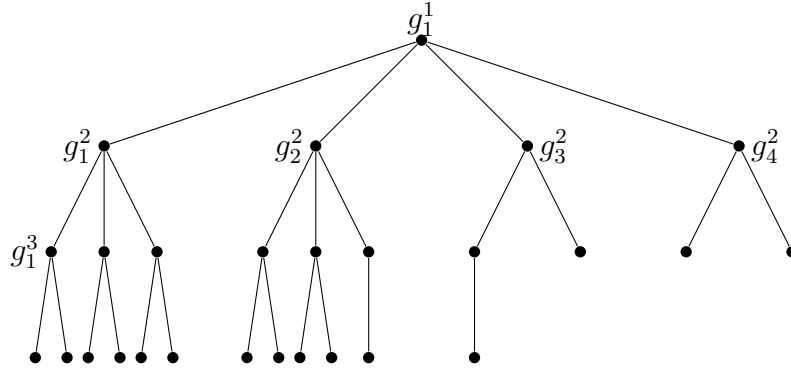


Figure 2: A greedy tree (only the labels of the first six vertices are shown).

Definition 6. A forest with components $F_1, \dots, F_t$ each of which is a greedy tree is called *greedy forest* if

$$\min\{\deg v : v \in F_i\} \geq \max\{\deg v : v \in F_{i+1}\}$$

for all $1 \leq i \leq t - 1$.

Remark 1. All the components of a greedy forest, except possibly one, have only vertices of degree 1 or 0.

As mentioned earlier, the first main theorem of this paper shows that the greedy tree maximizes the number of subtrees of any given order.

Theorem 1. *Among all trees T with degree sequence D , the number $n_k(T)$ of subtrees of order k attains its maximum when T is the greedy tree $G(D)$.*

It should be remarked that for specific k , the greedy tree is not necessarily the only tree for which $n_k(T)$ reaches its maximum (for instance, when $k = 1$, then $n_k(T)$ is simply the order of T and thus equal for all trees with degree sequence D). The important point of Theorem 1 is the fact that $n_k(T) \leq n_k(G(D))$ for *all* k simultaneously and all trees T with degree sequence D .

Theorem 1 will be proven in Section 3. In fact, a slightly more general result holds:

Theorem 2. *Among all forests F with given degree sequence, the number $n_k(F)$ of subtrees of order k attains its maximum when F is the greedy forest.*

In Section 4, we compare greedy trees with different degree sequences, which yields a number of corollaries such as:

Corollary 1. *Among trees with given order n and maximum degree Δ , the number $n_k(T)$ of subtrees of order k attains its maximum when T is the greedy tree $G(\Delta, \Delta, \dots, \Delta, d, 1, 1, \dots, 1)$, where $1 \leq d < \Delta$ is chosen in such a way that $d \equiv n - 1 \pmod{\Delta - 1}$.*

Similar results are obtained for trees with given number of leaves, independence number or matching number.

In Section 5, we study the number of subtrees containing a specific vertex (which we can assume to be the root). One of the motivations is a connection to a different counting problem: a rooted tree can be regarded as the Hasse diagram of a poset. There is a natural bijection between antichains and subtrees containing the root (see Figure 3): to each subtree that contains the root, we can associate the antichain that is formed by the leaves (excluding the root unless it is the only vertex of the subtree).

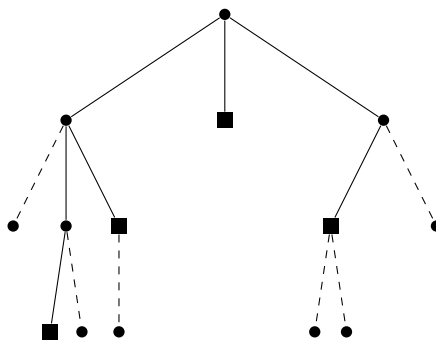


Figure 3: The correspondence between antichains (indicated by square nodes) and subtrees that contain the root (solid edges).

Therefore, the total number of subtrees that contain the root is the same as the number of (nonempty) antichains in the associated poset. It was pointed out by Klazar [6] that the number of antichains in a rooted tree of order n is at most $2^{n-1} + 1$ (with equality when the tree is a star rooted at its center) and at least n (with equality when the tree is a path rooted at one of its ends). Apart from this, not much seems to be known about extremal values of the number of antichains in a rooted tree.

The main result of Section 5 reads as follows:

Theorem 3. *Let $n_k(T, v)$ denote the number of subtrees of order k in T that contain the vertex v . For any tree T with degree sequence D , any vertex v of T and any $k \geq 1$, the inequality*

$$n_k(T, v) \leq n_k(G(D), r(G(D)))$$

holds, where $r(G(D))$ is the canonical root of the greedy tree, as chosen in Definition 5.

This implies that the greedy tree, rooted in the canonical way, also has the greatest number of antichains among all rooted trees with given degree sequence. A more general statement, where the degree of the root can be prescribed as well, also holds (see Theorem 12). Moreover, we also prove analogous statements for subtrees with a given number of leaves, which corresponds to antichains of given cardinality.

2 Preliminaries

We start with some preliminary considerations and further definitions and notation. We denote by $\mathbb{T}_D$ the set of all rooted (or edge-rooted) trees with leveled degree sequence

D. Similarly, let $\mathbb{F}_D$ be the set of all rooted forests with two components and with given leveled degree sequence *D*.

If *T* is a rooted tree, then we denote its root by $r(T)$. Whenever we consider $T - r(T)$ as rooted forest, we take as root in each connected component the unique neighbor of $r(T)$ in *T*. We denote by $C(T)$ the set of the connected components of $T - r(T)$.

Let T_1 and T_2 be two rooted trees. For $j \in \{1, 2\}$ and $l \geq 1$ let $\mathcal{V}_{l,j} = \{v_{j,1}^l, \dots, v_{j,k_{l,j}}^l\}$ be the set of vertices at the l^{th} level of T_j . We write $T_1 \triangleright T_2$ if the height of T_1 is at least that of T_2 and for any $l \geq 1$ we have

$$\min \left\{ \deg v_{1,1}^l, \dots, \deg v_{1,k_{l,1}}^l \right\} \geq \max \left\{ \deg v_{2,1}^l, \dots, \deg v_{2,k_{l,2}}^l \right\}$$

if $\mathcal{V}_{l,2}$ is not empty. The relation $\triangleright$ is easily seen to be transitive.

Remark 2. Let *F* be a rooted forest. *F* is a level greedy forest if and only if its components can be labeled as $F_1, \dots, F_t$ such that each of $F_1, \dots, F_t$ is a level greedy tree and $F_1 \triangleright \dots \triangleright F_t$. A tree *T* rooted at *v* is a level greedy tree if and only if $T - v$ is a level greedy forest.

The following simple observation turns out to be extremely useful in our study.

Lemma 4. *Let *F* be a rooted forest with $t \geq 2$ components. *F* is a level greedy forest if and only if any two components of *F* form a level greedy forest.*

Proof. As mentioned in Remark 2, a rooted forest is a level greedy forest if and only if $\triangleright$ induces a total order on its components (it is possible that two components F_i and F_j are isomorphic, but this can only happen if their degrees are constant on each level, in which case $F_i \triangleright F_j$ holds as well as $F_j \triangleright F_i$). Equivalently, any two components should be comparable by $\triangleright$, which in turn is equivalent to the statement that any two components form a level greedy forest. $\square$

Remark 3. The key to our proofs is the following observation made in [11]: if a tree is a level greedy tree with respect to every possible root vertex as well as a level greedy edge-rooted tree with respect to every possible root edge, then it must be a greedy tree. The proof is based on the “semi-regular” property defined in [13].

Moreover, if a rooted (or edge-rooted) tree is transformed to a level greedy tree with the same leveled degree sequence, then the Wiener index (sum of distances between all pairs of vertices) decreases, which means that one always reaches the greedy tree by repeatedly applying such transformations (no infinite loops are possible).

3 The number of subtrees of given order

The main objective of this section is to show that, given the degree sequence, the greedy tree maximizes the number of subtrees of any order *k*. We first consider two-component forests with a given leveled degree sequence, the case of a tree can then be considered as a special case.

We denote by $G_1(D)$ and $G_2(D)$ the two connected components of the level greedy forest $G(D)$, where we assume that

$$|V(G_1(D))| \geq |V(G_2(D))|.$$

Similarly, we write F_1 and F_2 for the components of a two-component rooted forest. In order to formulate our key lemma (Lemma 6 below), we need a few more definitions:

Definition 7. Let F be a rooted forest which has n levels of vertices. The *level sequence* of a subforest F' of F is the sequence $(s_1, \dots, s_n)$, where s_i is the number of vertices of F' at the i^{th} level in F .

We denote by $n_S(F)$ the number of subtrees in F with level sequence S . Clearly, for any integer $k \geq 1$, the number $n_k(F)$ of subtrees of order k in F can be written as the sum of $n_S(F)$ over all possible level sequences S that sum to k . Given a level sequence S , we write S^- for the level sequence obtained from S by removing the first term (i.e., if $S = (s_1, \dots, s_n)$, then $S^- = (s_2, \dots, s_n)$).

For any two pairs (a_1, a_2) and (b_1, b_2) of nonnegative real numbers we write $(b_1, b_2) \blacktriangleleft (a_1, a_2)$ if and only if

$$a_1 \geq \max\{b_1, b_2\}$$

and

$$a_1 + a_2 \geq b_1 + b_2.$$

The following simple fact will be used repeatedly in our proofs.

Lemma 5 (cf. [11]). *If $(b_1, b_2) \blacktriangleleft (a_1, a_2)$ and $(d_1, d_2) \blacktriangleleft (c_1, c_2)$, then we have $b_1d_1 + b_2d_2 \leq a_1c_1 + a_2c_2$.*

Now we are ready to formulate and prove the key lemma of this section.

Lemma 6. *Let D be a given leveled degree sequence of a two-component forest. For any level sequence $S = (s_1, s_2, \dots, s_{L(D)})$ and for any $F \in \mathbb{F}_D$ we have*

$$(n_S(F_1), n_S(F_2)) \blacktriangleleft (n_S(G_1(D)), n_S(G_2(D))). \quad (3)$$

Similarly, if D is a leveled degree sequence of a (vertex) rooted tree, then for any $T \in \mathbb{T}_D$ and for any level sequence $S = (s_1, s_2, \dots, s_{L(D)})$ we have

$$n_S(T) \leq n_S(G(D)). \quad (4)$$

Remark 4. Note that (4) is equivalent to $(n_S(T), 0) \blacktriangleleft (n_S(G(D)), 0)$. Hence we will only show (3) where F_2 is allowed to be empty.

Proof. We prove (3) by induction on $L(D)$. The case of $L(D) = 1$, as well as that of $L(D) = 2$, is trivial: in either case, the corresponding sets $\mathbb{F}_D$ and $\mathbb{T}_D$ contain only one element. Assume that the lemma is true whenever $L(D) \leq k$ for some integer $k \geq 2$. Now assume that $L(D) = k + 1$. We can also assume that $n_S(F_1) \geq n_S(F_2)$. There are two cases:

Case 1: $s_1 = 0$. In this case we have

$$n_S(G(D)) = \sum_{X \in C(G_1(D)) \cup C(G_2(D))} n_{S^-}(X)$$

and

$$n_S(F) = \sum_{X \in C(F_1) \cup C(F_2)} n_{S^-}(X).$$

Assume that there are elements H_1 and H_2 of $C(F_1) \cup C(F_2)$ such that $H_1 \cup H_2$ is not a (rooted) greedy forest, and let B be the leveled degree sequence of $H_1 \cup H_2$. We know by the induction hypothesis that

$$(n_{S^-}(H_1), n_{S^-}(H_2)) \blacktriangleleft (n_{S^-}(G_1(B)), n_{S^-}(G_2(B))).$$

By replacing H_1 and H_2 by $G_1(B)$ and $G_2(B)$, respectively, we obtain a new rooted forest F^1 with the same leveled degree sequence: if H_1 and H_2 both belong to $C(F_1)$, then $G_1(B)$ and $G_2(B)$ become part of $C(F_1^1)$, and the same applies to $C(F_2)$ and $C(F_2^1)$. If H_1 and H_2 belong to $C(F_1)$ and $C(F_2)$ respectively, then the larger component $G_1(B)$ becomes part of $C(F_1^1)$ and $G_2(B)$ becomes part of $C(F_2^1)$. It follows that

$$(n_S(F_1), n_S(F_2)) \blacktriangleleft (n_S(F_1^1), n_S(F_2^1)).$$

We iterate the process to obtain a sequence (with $F = F^0$)

$$(n_S(F_1^0), n_S(F_2^0)) \blacktriangleleft (n_S(F_1^1), n_S(F_2^1)) \blacktriangleleft \cdots \blacktriangleleft (n_S(F_1^K), n_S(F_2^K)).$$

This process always terminates, since the vector of all component sizes of $F^{t+1} - r(F^{t+1})$ (sorted in descending order) is lexicographically greater than that of $F^t - r(F^t)$. At the end, any two elements of $C(F_1^K) \cup C(F_2^K)$ form a greedy forest. By Lemma 4, such a situation is reached only when $F^K - \{r(F_1^K), r(F_2^K)\}$ is level greedy. Thus the branches of F^K are the same as those of $G(D)$, which means that

$$n_S(F_1^K) + n_S(F_2^K) = n_S(G_1(D)) + n_S(G_2(D)).$$

Since the largest branches are all part of $G_1(D)$ in the greedy tree, we also have

$$(n_S(F_1^K), n_S(F_2^K)) \blacktriangleleft (n_S(G_1(D)), n_S(G_2(D))),$$

which completes the proof in this case.

Case 2: $s_1 = 1$. Let r_i be the degree of the root of F_i for $i = 1, 2$. We reason by induction on $r = \max\{r_1, r_2\}$. If $r = 1$ then there are two subcases:

- If $s_2 = 0$, there is nothing to prove: all potential subtrees only consist of a root, so that $n_S(F)$ does not actually depend on F . Likewise, if $s_2 \geq 2$, then there are no subtrees with level sequence S in view of the assumption $r = 1$, and this is independent of the shape of F .

- Assume that $s_2 = 1$. Let G'_1, G'_2, F'_1, F'_2 be the trees obtained by removing the roots from $G_1(D), G_2(D), F_1, F_2$ respectively (G'_2 and F'_2 are empty if $G_2(D)$ and F_2 are). We have

$$n_S(G_1(D)) = n_{S^-}(G'_1), \quad n_S(G_2(D)) = n_{S^-}(G'_2),$$

and

$$n_S(F_1) = n_{S^-}(F'_1), \quad n_S(F_2) = n_{S^-}(F'_2).$$

Hence (3) follows (by our outer induction hypothesis with respect to $L(D)$) from

$$(n_{S^-}(F'_1), n_{S^-}(F'_2)) \blacktriangleleft (n_{S^-}(G'_1), n_{S^-}(G'_2)).$$

Assume (3) holds for $r \leq l$ for some $l \geq 1$. Let $r = l + 1$, and let A, B, A', B' be subtrees of F_1 and F_2 as shown in Figure 4. Note that if F_2 is an isolated vertex then B is empty and if F_2 is empty then so are B and B' .

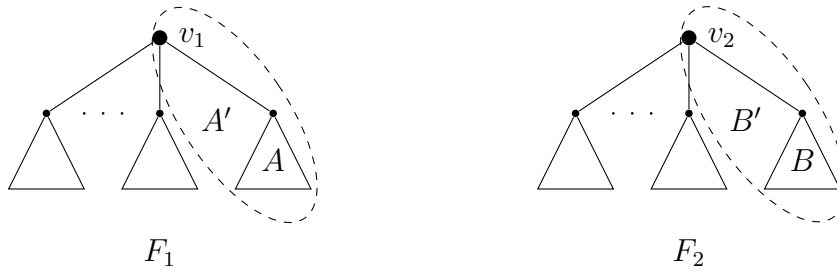


Figure 4: Decomposition of F .

Let $\mathbb{S}$ be the set of all possible level sequences whose first term is 1, so that we have

$$\begin{aligned} n_S(F) &= \sum_{\substack{S_1, S_2 \in \mathbb{S} \\ S_1^- + S_2^- = S^-}} (n_{S_1}(A')n_{S_2}(F_1 - A) + n_{S_1}(B')n_{S_2}(F_2 - B)) \\ &= \sum_{\substack{S_1, S_2 \in \mathbb{S} \\ S_1^- + S_2^- = S^-}} (n_{S_1^-}(A)n_{S_2}(F_1 - A) + n_{S_1^-}(B)n_{S_2}(F_2 - B)). \end{aligned}$$

We consider this sum term by term for any two given level sequences S_1 and S_2 . Let D_1 be the leveled degree sequence of $A \cup B$, and let D_2 be the leveled degree sequence of $(F_1 - A) \cup (F_2 - B)$. By applying the induction hypothesis with respect to $L(D)$ to A and B , we get

$$n_{S_1^-}(G_1(D_1)) \geq \max\{n_{S_1^-}(A), n_{S_1^-}(B)\} \quad (5)$$

and

$$n_{S_1^-}(G_1(D_1)) + n_{S_1^-}(G_2(D_1)) \geq n_{S_1^-}(A) + n_{S_1^-}(B). \quad (6)$$

On the other hand, applying the induction hypothesis with respect to r to $F_1 - A$ and $F_2 - B$ yields

$$n_{S_2}(G_1(D_2)) \geq \max\{n_{S_2}(F_1 - A), n_{S_2}(F_2 - B)\} \quad (7)$$

and

$$n_{S_2}(G_1(D_2)) + n_{S_2}(G_2(D_2)) \geq n_{S_2}(F_1 - A) + n_{S_2}(F_2 - B). \quad (8)$$

Equations (5) and (7) imply

$$\begin{aligned} & n_{S_1^-}(G_1(D_1))n_{S_2}(G_1(D_2)) \\ & \geq (\max\{n_{S_1^-}(A), n_{S_1^-}(B)\}) \cdot (\max\{n_{S_2}(F_1 - A), n_{S_2}(F_2 - B)\}) \\ & \geq \max\{n_{S_1^-}(A)n_{S_2}(F_1 - A), n_{S_1^-}(B)n_{S_2}(F_2 - B)\} \\ & = \max\{n_{S_1}(A')n_{S_2}(F_1 - A), n_{S_1}(B')n_{S_2}(F_2 - B)\}. \end{aligned} \quad (9)$$

By Lemma 5, we have

$$\begin{aligned} & n_{S_1^-}(G_1(D_1))n_{S_2}(G_1(D_2)) + n_{S_1^-}(G_2(D_1))n_{S_2}(G_2(D_2)) \\ & \geq n_{S_1^-}(A)n_{S_2}(F_1 - A) + n_{S_1^-}(B)n_{S_2}(F_2 - B) \end{aligned} \quad (10)$$

from (5), (6), (7) and (8). Let F^1 be the rooted forest whose first component F_1^1 is obtained by adding an edge joining the two roots of $G_1(D_2)$ and $G_1(D_1)$ and taking the root of $G_1(D_2)$ as root of F_1^1 , and the second component F_2^1 is constructed analogously by adding an edge joining the roots of $G_2(D_2)$ and $G_2(D_1)$ and taking the root of $G_2(D_2)$ as root of F_2^1 (if $G_2(D_1)$ is empty, then $F_2^1 = G_2(D_2)$). See Figure 5.

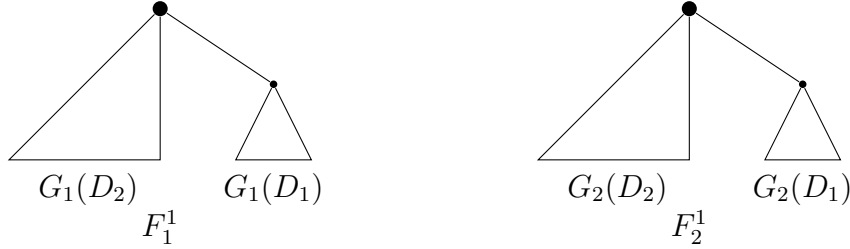


Figure 5: The rooted forest F^1 .

Since (9) and (10) are valid for arbitrary S_1 and S_2 satisfying the relation $S^- = S_1^- + S_2^-$, they imply that

$$n_S(F_1^1) \geq \max\{n_S(F_1), n_S(F_2)\}$$

and

$$n_S(F_1^1) + n_S(F_2^1) \geq n_S(F_1) + n_S(F_2).$$

We iterate this process to obtain a sequence of the form (with $F = F^0$)

$$(n_S(F_1^0), n_S(F_2^0)) \blacktriangleleft (n_S(F_1^1), n_S(F_2^1)) \blacktriangleleft \cdots \blacktriangleleft (n_S(F_1^K), n_S(F_2^K)).$$

The process terminates when it is no longer possible to find suitable branches A and B to replace. Clearly F^K must satisfy $F_1^K \supset F_2^K$. This means that F_1^K and F_2^K have the same leveled degree sequences as $G_1(D)$ and $G_2(D)$, respectively.

As our last step, we show that

$$n_S(F_1^K) \leq n_S(G_1(D)) \text{ and } n_S(F_2^K) \leq n_S(G_2(D)), \quad (11)$$

which implies

$$(n_S(F_1^K), n_S(F_2^K)) \blacktriangleleft (n_S(G_1(D)), n_S(G_2(D))).$$

Recall that r_i is the degree of the root of F_i^K and let $C(F_i^K) = \{H_1, \dots, H_{r_i}\}$ and $C(G_i(D)) = \{H'_1, \dots, H'_{r_i}\}$. From the way F_i^K is formed we know that any $r_i - 1$ elements of $C(F_i^K)$ form a level greedy forest.

In particular, if $r_i \geq 3$, then any two elements of $C(F_i^K)$ form a level greedy forest. In view of Lemma 4, we conclude that the elements of $C(F_i^K)$ form a level greedy forest. Hence F_i^K is a level greedy tree, and (11) follows trivially.

If $r_i = 2$, we have:

- If $s_2 \geq 3$, then $n_S(F_i^K) = n_S(G_i(D)) = 0$.
- If $s_2 = 2$, let T and U be the trees obtained from F_i^K and $G_i(D)$ respectively by merging the root and its neighbors. Let $S' = (s_1, s_3, s_4, \dots, s_n)$. Then we can use the outer induction hypothesis to get

$$n_S(F_i^K) = n_{S'}(T) \leq n_{S'}(U) = n_S(G_i(D)).$$

- If $s_2 = 1$, then we use again the (outer) induction hypothesis to get

$$n_S(F_i^K) = n_{S^-}(H_1) + n_{S^-}(H_2) \leq n_{S^-}(H'_1) + n_{S^-}(H'_2) = n_S(G_i(D)).$$

- The case $s_2 = 0$ is trivial.

If $r_i = 1$, the induction hypothesis of the first case gives us

$$n_S(F_i^K) = n_{S^-}(H_1) \leq n_{S^-}(H'_1) = n_S(G_i(D))$$

if $s_2 \geq 1$, and

$$n_S(F_i^K) = n_S(G_i(D)) = 1$$

if $s_2 = 0$.

Finally, if $r_i = 0$, then the isolated vertex F_i^K is clearly a greedy tree. Thus we have shown (11) in all possible cases, so that

$$(n_S(F_1), n_S(F_2)) \blacktriangleleft (n_S(F_1^K), n_S(F_2^K)) \blacktriangleleft (n_S(G_1(D)), n_S(G_2(D))),$$

which completes the induction and thus our entire proof. $\square$

A similar lemma also holds for edge-rooted trees in a completely analogous way.

Lemma 7. *Let D be the leveled degree sequence of an edge-rooted tree. For any $T \in \mathbb{T}_D$ we have $n_S(T) \leq n_S(G(D))$ for any level sequence $S = (s_1, s_2, \dots, s_{L(D)})$.*

Proof. Let $D = ((i_{1,1}, i_{1,2}), (i_{2,1}, \dots, i_{2,k_2}), \dots, (i_{n,1}, \dots, i_{n,k_n}))$. If $s_1 \leq 1$, then the lemma follows clearly from Lemma 6 (the edge between the roots does not play a role in this case). The case $s_1 = 2$ is obtained by another application of Lemma 6, since in the case we have

$$n_S(G(D)) = n_{(s_1-1, s_2, \dots, s_{L(D)})}(G(D'))$$

for $D' = ((i_{1,1} + i_{1,2} - 2), (i_{2,1}, \dots, i_{2,k_2}), \dots, (i_{n,1}, \dots, i_{n,k_n}))$, and

$$n_S(T) = n_{(s_1-1, s_2, \dots, s_{L(D)})}(T')$$

where T' is the tree obtained by merging the ends of the edge root to obtain a vertex root. Note that $G(D')$ and T' are elements of $\mathbb{T}_{D'}$. Finally, if $s_1 \geq 3$, then clearly $n_S(T) = n_S(G(D)) = 0$. $\square$

Our main result of this section follows as an immediate consequence of Lemmas 6 and 7. As explained in Remark 3, a tree that is level greedy with respect to all possible vertex or edge roots also satisfies the “semi-regular” property from [13] and is thus a greedy tree. Hence we have the following theorem:

Theorem 8. *Among trees with a given degree sequence, $n_k(T)$ is maximized when T is the greedy tree.*

Corollary 2. *Among forests with a given degree sequence, $n_k(F)$ is maximized when F is the greedy forest.*

Proof. Let F be a forest whose components are $F_1, \dots, F_t$ ordered by non-increasing diameters. Whenever there is a possible choice of roots for F so that it has a leveled degree sequence B and it is not a level greedy forest, we have $n_k(F) \leq n_k(G(B))$. Hence we can choose F to be level greedy with respect to any choice of (vertex) roots.

Let v be a leaf end of a longest path in F_1 . Assume that F_2 has a vertex w whose degree is larger than 1. Then the forest $F_1 \cup F_2$ considered to be rooted at v and w would not be level greedy: $\deg v < \deg w$ but the height of F_1 is larger than the height of F_2 . Hence, F_1 is the only component of F that can possibly have vertices with degree greater than 1.

By Theorem 8 choosing F_1 to be greedy leaves unchanged or increases the number of subtrees of order k . With Remark 1, this completes the proof. $\square$

4 Comparing different degree sequences

Comparing the greedy trees associated to different degree sequences, we will be able to determine the extremal trees with respect to the number of subtrees (of any given order) in a variety of different tree classes, cf. [22]. We use $D = (d_1, \dots, d_n)$ and $B = (b_1, \dots, b_n)$ to denote two different degree sequences (as opposed to leveled degree sequences unless otherwise mentioned) of trees.

We write

$$B \prec D \tag{12}$$

and say that D majorizes B if for any $1 \leq \ell \leq n$ we have

$$\sum_{i=1}^{\ell} b_i \leq \sum_{i=1}^{\ell} d_i.$$

The main goal of this section is to compare $n_k(G(B))$ and $n_k(G(D))$ if (12) is satisfied. It turns out that $n_k(G(B)) \leq n_k(G(D))$ in this case, from which a number of corollaries can be deduced.

For a vertex v in a rooted or edge-rooted tree T , let T_v denote the subtree induced by v and its descendants (Figure 6). Let $n_k(T, v)$ denote the number of subtrees of order k in T that contain the vertex v . The following lemma compares the values of n_k for all vertices on the same level of a level greedy tree.

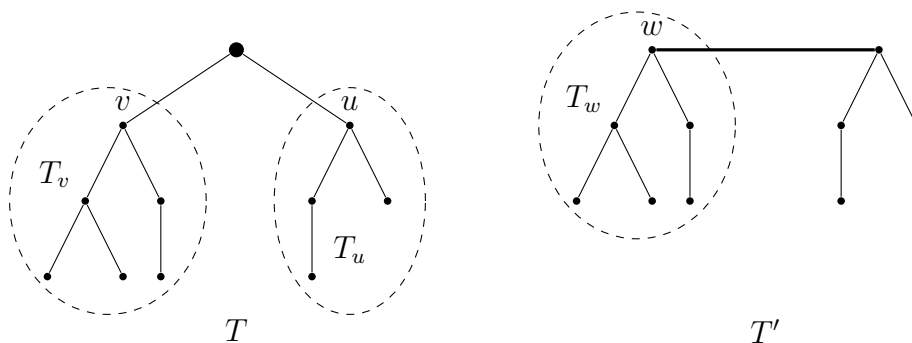


Figure 6: Definition of T_v .

Lemma 9. Let $D = ((i_{1,1}), (i_{2,1}, \dots, i_{2,k_2}), \dots, (i_{n,1}, \dots, i_{n,k_n}))$ be a leveled degree sequence of a tree. Then for all $1 \leq l \leq L(D)$ and $k \geq 1$ we have

$$n_k(G(D), g_1^l) \geq n_k(G(D), g_2^l) \geq \dots \geq n_k(G(D), g_{k_l-1}^l) \geq n_k(G(D), g_{k_l}^l).$$

Proof. Let $u = g_i^l$ and $v = g_j^l$ with $i < j$ be two vertices on the same level l of $T = G(D)$, and let w be their first (i.e., closest) common ancestor. We define a size-preserving injection from the set of all subtrees of $G(D)$ that contain v to the set of all subtrees of $G(D)$ that contain u . To this end, let u' and v' be the children of w for which $u \in T_{u'}$ and $v \in T_{v'}$. By the greedy construction, all vertices in $T_{u'}$ have greater or equal degree than all vertices in $T_{v'}$ on the same level. Hence there is an isomorphic embedding Φ of $T_{v'}$ into $T_{u'}$ that maps v to u , see Figure 7 for an example.

Let us now describe the size-preserving injection that maps a subtree R of $T = G(D)$ containing v to a subtree R' of $G(D)$ containing u . We distinguish three different cases:

1. If R contains both u and v , then we simply set $R' = R$.
2. If R does not contain u and also does not contain w , then we set $R' = \Phi(R)$.

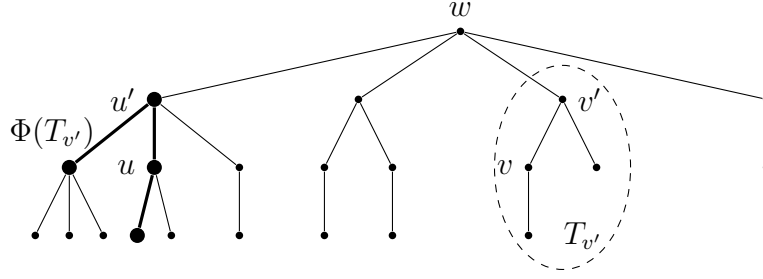


Figure 7: The tree $T_{v'}$ and its image $\Phi(T_{v'})$ (bold).

3. If R does not contain u , but it does contain w , then let x be the first vertex (i.e. closest to w) on the path from w to u that is not contained in R , and let y be the vertex on the path from w to v that lies on the same level as x . Replace $R \cap T_y$ by $\Phi(R \cap T_y)$ (note that Φ maps the path from w to v to the path from w to u , thus x to y) to obtain R' , see Figure 8.

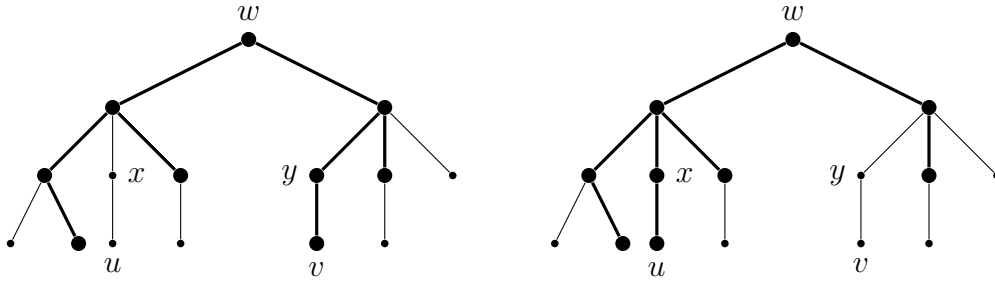


Figure 8: R (left) and R' (right) in Case (3).

It is easy to see that this is an injection that preserves the size of subtrees, so it follows immediately that $n_k(T, u) = n_k(G(D), g_i^l) \geq n_k(G(D), g_j^l) = n_k(T, v)$. $\square$

The same result holds for edge-rooted trees, and the proof is analogous:

Lemma 10. *Let $D = ((i_{1,1}, i_{1,2}), (i_{2,1}, \dots, i_{2,k_2}), \dots, (i_{n,1}, \dots, i_{n,k_n}))$ be a leveled degree sequence of an edge-rooted tree. Then we have*

$$n_k(G(D), g_1^l) \geq n_k(G(D), g_2^l) \geq \dots \geq n_k(G(D), g_{k_l-1}^l) \geq n_k(G(D), g_{k_l}^l).$$

for any positive integer k and $1 \leq l \leq L(D)$.

We are now able to prove the main theorem of this section:

Theorem 11. *Let $D = (d_1, \dots, d_n)$ and $B = (b_1, \dots, b_n)$ be degree sequences of trees of the same order such that $B \prec D$. Then for any positive integer k we have*

$$n_k(G(B)) \leq n_k(G(D)).$$

Proof. If $B = D$, the statement is trivial. Otherwise, there exists i_0 such that $d_{i_0} \neq b_{i_0}$. Since

$$\sum_{i=1}^n b_i = \sum_{i=1}^n d_i, \quad (13)$$

we know that the set $\{i : d_i \neq b_i\}$ must have at least two elements. Let $l = \min\{i : d_i \neq b_i\}$ and $m = \max\{i : d_i \neq b_i\}$. We must have $b_l < d_l$ and $b_m > d_m$. Note first that

$$B_1 = (b_1, \dots, b_{l-1}, b_l + 1, b_{l+1}, \dots, b_{m-1}, b_m - 1, b_{m+1}, \dots, b_n)$$

is a valid degree sequence, because $b_{l-1} = d_{l-1} \geq d_l \geq b_l + 1$ and $b_{m+1} = d_{m+1} \leq d_m \leq b_m - 1$. It is easy to see that $B \prec B_1$. Consider two vertices u and v in the greedy tree $G(B)$ such that $\deg u = b_l$ and $\deg v = b_m$.

Case 1: The length of the path in $G(B)$ joining u and v is even. Let w be the middle vertex of this path. Consider $G(B)$ as a level greedy tree whose root is w , then u and v are on the same level, say level h . We have $u = g_i^h$ and $v = g_j^h$ for some $i < j$. Without loss of generality, we may assume that j is the largest index such that $\deg g_j^h = b_m$ (otherwise, replace v by the vertex $g_{j'}^h$ which has this property). Let $x = g_r^{h+1}$ be a child of $v = g_j^h$, and let $H = G(B)_x$ be the branch rooted at x . Then $G(B) - H$ is still a level greedy tree (by maximality of j).

Consider the tree $T = G(B) - vx + ux$ with degree sequence B_1 . Subtrees of $G(B)$ are still subtrees in T except for those that contain both v and x , but not u . On the other hand, we gain subtrees that contain u and x , but not v . This yields

$$n_k(T) - n_k(G(B)) = \sum_{k_1+k_2=k} n_{k_1}(H, x) \left(n_{k_2}(G(B) - H, u) - n_{k_2}(G(B) - H, v) \right),$$

which is nonnegative in view of Lemma 9. It follows that

$$n_k(G(B_1)) \geq n_k(T) \geq n_k(G(B))$$

for all $k > 0$.

Case 2: The length of the path in $G(B)$ joining u and v is odd. The argument is analogous to the previous case, but we use Lemma 10 instead of Lemma 9.

In either case, we have

$$n_k(G(B_1)) \geq n_k(G(B))$$

for all $k > 0$. We repeat this process to obtain a sequence of degree sequences $B_0 = B, B_1, B_2, \dots, B_r = D$ such that

$$B = B_0 \prec B_1 \prec \dots \prec B_r = D$$

and

$$n_k(G(B)) = n_k(G(B_0)) \leq n_k(G(B_1)) \leq \dots \leq n_k(G(B_r)) = n_k(G(D))$$

for all $k > 0$, which proves the theorem. $\square$

A number of corollaries follow in the same way as Corollaries 5.1 – 5.5 of [22]. Corollary 1, which has already been mentioned in the introduction, is such an instance. Let us mention a few more; the proofs are very similar, so we only give a proof of the first corollary.

Corollary 3. *For any tree T of order n , we have*

$$n_k(T) \leq \begin{cases} \binom{n-1}{k-1} & k > 1, \\ n & k = 1. \end{cases}$$

Proof. Note that the degree sequence $(n-1, 1, 1, \dots, 1)$ of the star S_n majorizes all other degree sequences, and that

$$n_k(S_n) = \begin{cases} \binom{n-1}{k-1} & k > 1, \\ n & k = 1. \end{cases}$$

□

Corollary 4. *Among trees T of order n with s leaves, the number $n_k(T)$ is maximized by the greedy tree $G(s, 2, 2, \dots, 2, 1, 1, \dots, 1)$ (the number of $2s$ is $n-s-1$, the number of $1s$ is s) for any $k \geq 1$.*

Corollary 5. *Among trees T of order n with independence number $\alpha \geq n/2$, the number $n_k(T)$ is maximized by the greedy tree $G(\alpha, 2, 2, \dots, 2, 1, 1, \dots, 1)$ (the number of $2s$ is $n-\alpha-1$, the number of $1s$ is α) for any $k \geq 1$.*

Corollary 6. *Among trees T of order n with matching number $\beta \leq n/2$, the number $n_k(T)$ is maximized by the greedy tree $G(n-\beta, 2, 2, \dots, 2, 1, 1, \dots, 1)$ (the number of $2s$ is $\beta-1$, the number of $1s$ is $n-\beta$) for any $k \geq 1$.*

5 Subtrees containing a given vertex

This section is devoted to subtrees containing a given vertex, which, as explained in the introduction, are strongly related to antichains in rooted trees. One of the consequences of Theorem 1 is the following:

Corollary 7. *Let $\rho_k(T)$ be the average number of subtrees of size k containing a randomly chosen vertex of T . The inequality*

$$\rho_k(T) \leq \rho_k(G(D))$$

holds for all $k \geq 1$ and all trees T of degree sequence D .

Proof. If we denote the order of T by n as usual, we have

$$\rho_k(T) = \frac{\sum_{v \in V(T)} n_k(T, v)}{n} = \frac{kn_k(T)}{n},$$

since each subtree of order k is counted k times. The desired inequality follows immediately. □

It is natural to assume that the greedy tree also maximizes $n_k(T, v)$ if we choose v to be the canonical root. This fact, which has already been stated in the introduction (Theorem 3), is the main result of this section.

Proof of Theorem 3. Fix k , and let T be a rooted tree with degree sequence D . Consider a path $P = u_1 \dots u_m$ such that $m \geq 2$ and $u_1 = r(T)$ is the root of T . Let B be one of the branches attached to u_m by an edge such that $V(P) \cap V(B) = \emptyset$. Let T_i be the rooted tree obtained by removing B from u_m and attaching it to u_i for some $1 \leq i \leq m - 1$ (the root stays the same, see Figure 9).

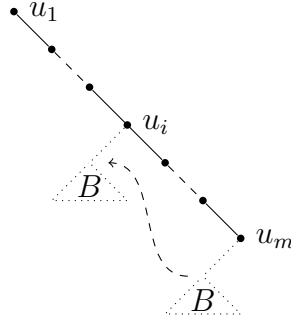


Figure 9: Moving the branch B from u_m to u_i .

Then we have $n_k(T_i, r(T_i)) \geq n_k(T, r(T))$ for any $k \geq 1$ and $1 \leq i \leq m - 1$: the number of subtrees which contain no vertex of B stays unchanged, and any subtree that contains the root as well as some vertices of B must contain the whole path P , thus it just gets transformed to a new subtree of the same order.

Now let T_M be a rooted tree with degree sequence D such that whenever a rooted tree T has degree sequence D , we always have $n_k(T_M, r(T_M)) \geq n_k(T, r(T))$. By the observation above, we can choose T_M such that $r(T_M)$ has maximum degree and the degrees of the vertices decrease as we move away from the root following a path. Hence if $D = (d_1, \dots, d_n)$, then $\deg r(T_M) = d_1$ and there exists a neighbor v of $r(T_M)$ with $\deg v = d_2$. By Lemma 6, T_M can be chosen to be a level greedy tree. Let $n_k(T_M, e)$ be the number of subtrees of order k in T_M that contain the edge $e := r(T_M)v$. If T'_M is the component of $T_M - e$ that contains the root $r(T_M)$ (see Figure 10), then we have

$$n_k(T_M, r(T_M)) = n_k(T_M, e) + n_k(T'_M, r(T'_M)). \quad (14)$$

By Lemma 7, we can reshuffle the branches in T_M to become a level greedy tree with edge root e , without decreasing $n_k(T_M, e)$ or $n_k(T'_M, r(T'_M))$: Note that the new T'_M obtained after reshuffling has the old one as a subgraph, given the fact that both of them are level greedy. Thus we also assume that T_M is level greedy with respect to the edge e .

For contradiction, assume that T_M (vertex rooted) is not isomorphic as rooted tree to $G(D)$. Then for some $i \geq 2$, there exist vertices u_i and u_{i+1} at levels i and $i+1$ respectively such that $\deg u_i < \deg u_{i+1}$. Using the fact that T_M is vertex rooted level greedy, we have the following where w_i and w_{i+1} are vertices at the level i and $i+1$ respectively:

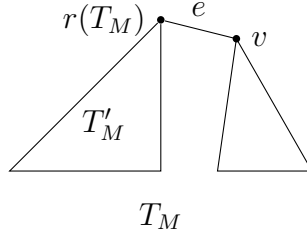


Figure 10: The tree T_M

Case 1: If $u_i, u_{i+1} \in V(T'_M)$, then there exists a vertex $w_{i+1} \in V(T_M - T'_M)$ such that $\deg u_i < \deg u_{i+1} \leq \deg w_{i+1}$.

Case 2: If $u_i, u_{i+1} \in V(T_M - T'_M)$, then there exists a vertex $w_i \in V(T'_M)$ such that $\deg w_i \leq \deg u_i < \deg u_{i+1}$. If level i of T'_M is already empty, we set $\deg w_i = 0$, and the argument that follows is still valid.

Case 3: If $u_i \in V(T_M - T'_M)$ and $u_{i+1} \in V(T'_M)$, then there exist vertices $w_i \in V(T'_M)$ and $w_{i+1} \in V(T_M - T'_M)$ such that

$$\deg w_i \leq \deg u_i < \deg u_{i+1} \leq \deg w_{i+1}.$$

The case that level i of T'_M is empty is treated in the same way as before. Hence all the three cases above can be reduced to the following fourth case:

Case 4: $u_i \in V(T'_M)$ and $u_{i+1} \in V(T_M - T'_M)$, but this contradicts the fact that T_M is level greedy as edge rooted tree with root e . $\square$

Before extending Theorem 3 a little further, we introduce the following related concepts.

Definition 8. A level greedy tree with leveled degree sequence

$$D = ((i_{1,1}), (i_{2,1}, \dots, i_{2,k_2}), \dots, (i_{n,1}, \dots, i_{n,k_n}))$$

is called a *sliced greedy tree* if

$$\min(i_{j,1}, \dots, i_{j,k_j}) \geq \max(i_{j+1,1}, \dots, i_{j+1,k_{j+1}})$$

for all $2 \leq j \leq n-1$. If furthermore we also have $i_{1,1} + 1 \geq \max(i_{2,1}, \dots, i_{2,k_2})$, then we say that the tree is a *branch greedy tree*.

In particular, a greedy tree is always a sliced greedy tree and a branch greedy tree. It is not hard to see that any sliced greedy tree can always be completed by adding further branches to turn it into a greedy tree, and that every branch of a greedy (sliced greedy, branch greedy) tree is branch greedy.

We now aim to extend Theorem 3, and show that among all rooted trees with given degree sequence and given degree of the root, the corresponding sliced greedy tree has the maximum number of subtrees of any given order containing the root.

Theorem 12. Let $D = (d_1, \dots, d_n)$ be a degree sequence of a tree. Let $\mathbb{T}_D^d$ be the set of all rooted trees with degree sequence D and root of degree d . Let $G(D, d)$ be the sliced greedy tree in $\mathbb{T}_D^d$. For any $T \in \mathbb{T}_D^d$ and for any positive integer k we have

$$n_k(T, r(T)) \leq n_k(G(D, d), r(G(D, d))).$$

Proof. The case where $d = d_1$ coincides with Theorem 3. Hence, in the rest of the proof we assume that $d \leq d_2$. Since we know that $n_0(T, r(T)) = n_0(G(D, d), r(G(D, d))) = 0$ and $n_1(T, r(T)) = n_1(G(D, d), r(G(D, d))) = 1$, we only have to check the case where $k \geq 2$. We use an induction with respect to n . For the case where $n = 1, 2, 3$, the theorem clearly holds since $|\mathbb{T}_D^d| \leq 1$. Assume it holds whenever $1 \leq n \leq h$ for some integer $h \geq 3$. Now, consider the case where $n = h + 1$. By the same reasoning as in the first paragraph in the proof of Theorem 3, we can move branches in T closer to $r(T)$ without decreasing $n_k(T, r(T))$. Therefore, we can assume that if u is a neighbor of $r(T)$ and u' is any vertex of T which is in the branch of $r(T)$ containing u then $\deg u \geq \deg u'$. This and the induction hypothesis allow us to assume that each branch of $r(T)$ is branch greedy. In particular, $r(T)$ must have a neighbor v such that $\deg v = d_1$.

Let us start another induction on d . If $d = 1$, then for all $k \geq 1$ we have

$$\begin{aligned} n_k(T, r(T)) &= n_{k-1}(T - r(T), v) \\ &\leq n_{k-1}(G(D', d_1 - 1), r(G(D', d_1 - 1))) \\ &= n_k(G(D, d), r(G(D, d))), \end{aligned}$$

where D' is the degree sequence of $T - r(T)$.

Next we consider the case $d = 2$. Let T_1 and T_2 be the components of $T - r(T)$, where the neighbors of $r(T)$ are considered as roots and v is in $V(T_1)$. By Lemma 6, we can assume that $T_1 \triangleright T_2$. If T and $G(D, d)$ have the same leveled degree sequence, then the two are isomorphic and we are done. Otherwise, there is an integer $i \geq 2$ such that there are two vertices u_i and u_{i+1} at the i^{th} and $(i + 1)^{\text{th}}$ levels of T respectively, which satisfy

$$\deg u_i < \deg u_{i+1}. \quad (15)$$

We choose u_{i+1} such that its degree is maximum among all vertices at the $(i + 1)^{\text{th}}$ level. Since both T_1 and T_2 are branch greedy, u_i and u_{i+1} must belong to different branches of $r(T)$. But it is impossible that $u_i \in V(T_1)$ and $u_{i+1} \in V(T_2)$, since if we let $w_i \in V(T_2)$ be a vertex at the i^{th} level in T_2 , then using the relation $T_1 \triangleright T_2$ and the fact that T_2 is branch greedy we would have $\deg u_i \geq \deg w_i \geq \deg u_{i+1}$, contradicting (15). Hence, we must have $u_i \in V(T_2)$ and $u_{i+1} \in V(T_1)$. Let T' be the tree obtained from T by merging v and $r(T)$ to become the new root, see Figure 11. Then we have (recall that we are assuming $k \geq 2$)

$$n_k(T, r(T)) = n_{k-1}(T', r(T')) + n_{k-1}(T_2, r(T_2)). \quad (16)$$

Note that $n_{k-1}(T', r(T'))$ counts the subtrees of order k in T which contain the edge $vr(T)$, and $n_{k-1}(T_2, r(T_2))$ counts those that do not contain $vr(T)$.

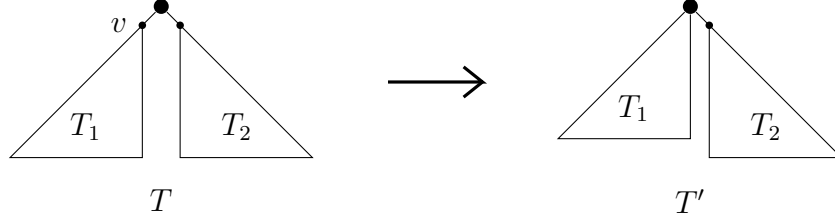


Figure 11: The tree T' in the proof of Theorem 12

Let $x_1, \dots, x_{d_1-1}$ be the neighbors of $r(T_1)$ in T_1 . We permute the vertices of T' to obtain a new, level greedy tree T'' with the same leveled degree sequence as T' but such that if $B_1, \dots, B_{d_1}$ are the branches of $r(T'')$ containing $x_1, \dots, x_{d_1-1}, r(T_2)$ respectively, then we have $B_{d_1} \triangleright \dots \triangleright B_1$. Set $T_2'' := B_{d_1}$ and $T_1'' := T'' - B_{d_1}$. Let T''' be a tree obtained from T by replacing T_1 and T_2 by T_1'' and T_2'' , respectively. Note that T_2 is isomorphic to a subgraph of T_2'' . Let T^1 be the level greedy tree with the same leveled degree sequence as T''' . From Lemma 6, we deduce that $n_{k-1}(T', r(T')) \leq n_{k-1}(T'', r(T''))$, $n_{k-1}(T_2, r(T_2)) \leq n_{k-1}(T_2'', r(T_2''))$ and consequently

$$\begin{aligned} n_k(T, r(T)) &= n_{k-1}(T', r(T')) + n_{k-1}(T_2, r(T_2)) \\ &\leq n_{k-1}(T'', r(T'')) + n_{k-1}(T_2'', r(T_2'')) \\ &= n_k(T''', r(T''')) \leq n_k(T^1, r(T^1)). \end{aligned}$$

Along this process, at least one vertex with maximum degree at the $(i+1)^{\text{th}}$ level in T (hence the same degree as u_{i+1}) is transferred to the i^{th} level in T'', T''' and T^1 . We can iterate the same process until we reach a tree that is isomorphic to $G(D, d)$.

Now we can resume our induction with respect to the degree d . Assume that the Theorem holds for $d = m$ for some integer $m \geq 2$. Now, consider the case where $d = m+1$. By the induction hypothesis with respect to d , we can assume that whenever B is a branch of $r(T)$ then $T - B$ is a sliced greedy tree. For any vertex u_i at the i^{th} level and u_{i+1} at the $(i+1)^{\text{th}}$ level in T for some $i \geq 2$, there exists a branch B of $r(T)$ such that $\{u_i, u_{i+1}\} \cap V(B) = \emptyset$, since $d \geq 3$. Since $T - B$ is a sliced greedy tree, we must have $\deg u_i \geq \deg u_{i+1}$. This means that T has the same leveled degree sequence as $G(D, d)$. Hence, by Lemma 6 we get $n_k(G(D, d), r(G(D, d))) \geq n_k(T, r(T))$. $\square$

Recall that subtrees containing the root correspond to antichains when a rooted tree is regarded as a Hasse diagram of a poset. Since the cardinality of the antichain corresponds to the number of leaves (counting the root as a leaf only if it is the only vertex of the subtree), it is natural to ask whether similar statements as Theorem 3 and Theorem 12 remain true if the number of subtrees with a fixed number l of leaves is considered instead. This turns out to be the case.

For any (vertex) rooted forest F , let $\eta_l(F)$ denote the number of subtrees in F which contain one of the roots and have l leaves (as before, the root is only counted as a leaf if it is the only vertex). It is convenient to set $\eta_0(F) = 1$ and $\eta_l(F) = 0$ for negative l .

Moreover, it is easy to see that $\eta_1(F) = |F|$ only depends on the order of F , so we will focus on the case $l \geq 2$ in the following.

Lemma 13. *Let D be a given leveled degree sequence of a two-component forest. For any positive integer l and for any $F \in \mathbb{F}_D$ we have*

$$(\eta_l(F_1), \eta_l(F_2)) \blacktriangleleft (\eta_l(G_1(D)), \eta_l(G_2(D))). \quad (17)$$

Similarly, if D is a leveled degree sequence of a (vertex) rooted tree, then for any $T \in \mathbb{T}_D$ and for any positive integer l we have

$$\eta_l(T) \leq \eta_l(G(D)). \quad (18)$$

Proof. By the same reasoning as in Remark 4, we only have to show (17), where we allow F_2 to be empty. We use an induction on $L(D)$. The cases $L(D) = 1, 2$ are trivial, since the degree sequence uniquely characterizes the tree in these cases. Assume that the lemma is true whenever $L(D) \leq k$ for some integer $k \geq 2$, and suppose that $L(D) = k + 1$. Let r_i be the degree of the root of F_i for $i = 1, 2$, and assume that $r_1 \geq r_2$. We start a new induction, this time with respect to r_1 . For $r_1 = 1$ and any $l \geq 2$, we have

$$\eta_l(G_1(D)) = \eta_l(G_1(D) - r(G_1(D))) \text{ and } \eta_l(F_1) = \eta_l(F_1 - r(F_1)),$$

for non-empty $G_2(D)$ we get

$$\eta_l(G_2(D)) = \eta_l(G_2(D) - r(G_2(D))) \text{ and } \eta_l(F_2) = \eta_l(F_2 - r(F_2)),$$

and for empty $G_2(D)$ we still have

$$\eta_l(F_2) = \eta_l(F_2 - r(F_2)) = 0 \text{ and } \eta_l(G_2(D)) = \eta_l(G_2(D) - r(G_2(D))) = 0.$$

Hence (17) follows (by the induction hypothesis) from

$$(\eta_l(F_1 - r(F_1)), \eta_l(F_2 - r(F_2))) \blacktriangleleft (\eta_l(G_1(D) - r(G_1(D))), \eta_l(G_2(D) - r(G_2(D)))).$$

Assume (17) holds whenever $r_1 \leq m$ for some $m \geq 1$, and let $r_1 = m + 1$. Let A and B be subtrees of F_1 and F_2 as shown in Figure 4, where B is empty if F_2 is an isolated vertex or if F_2 is empty. Then the following relation holds:

$$\begin{aligned} \eta_l(F) &= \eta_{l-1}(A)(\eta_1(F_1 - A) - 1) + \eta_{l-1}(B)(\eta_1(F_2 - B) - 1) \\ &\quad + \sum_{\substack{l_1, l_2 \geq 0, l_2 \neq 1 \\ l_1 + l_2 = l}} \eta_{l_1}(A)\eta_{l_2}(F_1 - A) + \eta_{l_1}(B)\eta_{l_2}(F_2 - B). \end{aligned}$$

This follows from the fact that the l leaves have to be divided into l_1 leaves in A (or B) and l_2 leaves in $F_1 - A$ (or $F_2 - B$) respectively. The only exception is the case $l_2 = 1$: the subtree of $F_1 - A$ ($F_2 - B$) that only consists of the root counts with 0 leaves. The next step of the proof follows the same lines as the proof of Lemma 6, so we skip the details:

the induction hypothesis shows that each term in the above sum is maximized when A , B , $F_1 - A$ and $F_2 - B$ are components of greedy forests. This argument gives us a forest F^K whose components F_1^K and F_2^K have the same leveled degree sequences as $G_1(D)$ and $G_2(D)$ respectively, and which satisfies

$$(\eta_l(F_1), \eta_l(F_2)) \blacktriangleleft (\eta_l(F_1^K), \eta_l(F_2^K)).$$

Moreover, if d_i denotes the root degree of F_i^K , we can assume that any $d_i - 1$ elements of $C(F_i^K)$ form a level greedy forest. If $d_i \neq 2$, then it follows immediately that F_i^K is a greedy tree (as in Lemma 6). Thus we only consider the case $d_i = 2$.

Let $C(F_i^K) = \{H_1, H_2\}$ and $C(G_i(D)) = \{H'_1, H'_2\}$, and let $F_i'^K$ and $G'_i(D)$ be, respectively, the trees obtained from F_i^K by merging $r(F_i^K)$ with its two neighbors and from $G_i(D)$ by merging $r(G_i(D))$ with its neighbors. Then the induction hypothesis with respect to $L(D)$ applied to $H_1 \cup H_2$ yields

$$\begin{aligned} \eta_2(F_i^K) &= \eta_2(F_i'^K) + \eta_1(H_1) + \eta_1(H_2) - 1 \\ &\leq \eta_2(G'_i(D)) + \eta_1(H'_1) + \eta_1(H'_2) - 1 = \eta_2(G_i(D)), \end{aligned}$$

where the term -1 is due to an over-count of the subtree having the two neighbors of $r(F_i)$ or of $r(G_i(D))$ as leaves. For all $l \geq 3$ we have

$$\begin{aligned} \eta_l(F_i^K) &= \eta_l(F_i'^K) + \eta_{l-1}(H_1) + \eta_{l-1}(H_2) \\ &\leq \eta_l(G'_i(D)) + \eta_{l-1}(H'_1) + \eta_{l-1}(H'_2) = \eta_l(G_i(D)). \end{aligned} \tag{19}$$

Note here that $\eta_l(F_i'^K)$ is the number of subtrees in F_i^K containing $r(F_i^K)$ and having l leaves none of which is a neighbor of $r(F_i^K)$, while $\eta_{l-1}(H_1)$ and $\eta_{l-1}(H_2)$ count subtrees with l leaves one of which is a neighbor of $r(F_i^K)$. $\square$

Lemma 14. *Let T be a vertex rooted tree, and let $P = u_1 \dots u_m$ ($m \geq 2$) be a path starting at the root (i.e., $u_1 = r(T)$). Let B be a branch attached by an edge to u_m such that $V(B) \cap V(P) = \emptyset$. Let v be the neighbor of u_m in B , and let T' be a tree obtained from T by removing the edge vu_m and adding a new edge vu_{m-1} (see Figure 9 with $i = m - 1$). Then we have $\eta_l(T) \leq \eta_l(T')$.*

Proof. Clearly we have $\eta_l(T - B) = \eta_l(T' - B)$. Hence we only have to compare the number of subtrees containing some vertices of B . To any subtree S in T which contains $r(T)$ and such that $V(S) \cap V(B) \neq \emptyset$, we associate $f(S)$ defined as follows:

- $f(S) = S - vu_m + vu_{m-1}$ if u_m is not a leaf in $S - vu_m + vu_{m-1}$;
- otherwise we have $f(S) = S - u_m + vu_{m-1}$.

Clearly $f(S)$ is a subtree of T' , it has the same number of leaves as S , and if $f(S) = f(S')$ then we also have $S = S'$: add u_m if it is missing, add vu_m and remove vu_{m-1} to obtain S and S' from $F(S)$ and $F(S')$, respectively. Hence f is injective, and we are done. $\square$

Theorem 15. *For all rooted trees T with degree sequence D and for all positive integers l we have $\eta_l(G(D)) \geq \eta_l(T)$, where $G(D)$ is rooted in the canonical way as in Definition 5.*

Proof. Let T_M be a rooted tree with degree sequence $D = (d_1, \dots, d_n)$, and such that for all rooted trees T with degree sequence D we have $\eta_l(T_M) \geq \eta_l(T)$. By the two Lemmas 13 and 14, we can choose T_M to be level greedy with $\deg r(T_M) = d_1$, and such that $r(T_M)$ has a neighbor v whose degree is d_2 . Let T'_M be the component of $T_M - vr(T_M)$ that contains $r(T_M)$, and let T''_M be the tree obtained from T_M by merging v and $r(T_M)$ to become the new root. Let A be the set of subtrees of T_M containing $r(T_M)$ and having l leaves such that one of them is v , and let B be the set of the subtrees of T_M containing $r(T_M)$ and l leaves none of which is v . Then we have

$$\eta_l(T_M) = |A| + |B| = \eta_{l-1}(T'_M) + \eta_l(T''_M)$$

for $l \geq 3$; for $l = 2$, one has to subtract 1 (for the subtree that only consists of $r(T_M)$ and v). If we reshuffle the branches of v and $r(T_M)$ such that T_M considered as edge rooted tree with $vr(T_M)$ as root is level greedy, the values of $\eta_{l-1}(T'_M)$ and $\eta_l(T''_M)$ will not decrease. This is because T''_M would then be level greedy as well, and the new T'_M has the old one as subgraph (the fact that T'_M is level greedy is used here). Hence, the theorem follows by the same argument that we have already seen in the proof of Theorem 3. $\square$

More generally, we also have a theorem analogous to Theorem 12:

Theorem 16. *Let T be a rooted tree with degree sequence $D = (d_1, \dots, d_n)$ and root degree $\deg r(T) = d$. Then for any positive integer l , we have $\eta_l(T) \leq \eta_l(G(D, d))$, where $G(D, d)$ denotes the sliced greedy tree whose root has degree d and whose degree sequence is D .*

Proof. See Theorem 15 for the case where $d = d_1$. As mentioned earlier, the case $l = 1$ is trivial, so we assume that $d \leq d_2$ and $l \geq 2$. By Lemma 14, we can restrict ourselves to the case where $r(T)$ has a neighbor v with $\deg v = d_1$. Let T' be the tree obtained from T by merging $r(T)$ and v and let T'' be the connected component of $T - vr(T)$ that contains $r(T)$. Then for all $l \geq 3$ we have

$$\eta_l(T) = \eta_l(T') + \eta_{l-1}(T'') \tag{20}$$

as in the previous proof (for $l = 2$, we have to subtract 1). Here, $\eta_l(T')$ counts the number of subtrees of T containing $r(T)$ and having l leaves none of which is v , and $\eta_{l-1}(T'')$ counts the subtrees of T containing $r(T)$ such that one of its l leaves is v . The rest of the proof is exactly as we have seen in the proof of Theorem 12, but we use (20) instead of (16), and we use Lemma 13 in the place of Lemma 6. $\square$

Acknowledgments

The first author was supported by the German Academic Exchange Service (DAAD), in association with the African Institute for Mathematical Sciences (AIMS). Code No.: A/11/07858.

The second author was supported by the National Research Foundation of South Africa under grant number 70560.

The third author was supported by the Simons Foundation under grant number 245307.

References

- [1] T. Biyikoğlu, M. Hellmuth, and J. Leydold. Largest eigenvalues of the discrete p -Laplacian of trees with degree sequences. *Electron. J. Linear Algebra*, 18:202–210, 2009.
- [2] T. Biyikoğlu and J. Leydold. Graphs with given degree sequence and maximal spectral radius. *Electron. J. Combin.*, 15(1):Research Paper 119, 9 pp., 2008.
- [3] M. S. Jacobson, A. E. Kézdy, and S. Seif. The poset on connected induced subgraphs of a graph need not be Sperner. *Order*, 12(3):315–318, 1995.
- [4] R. E. Jamison. On the average number of nodes in a subtree of a tree. *J. Combin. Theory Ser. B*, 35(3):207–223, 1983.
- [5] R. E. Jamison. Monotonicity of the mean order of subtrees. *J. Combin. Theory Ser. B*, 37(1):70–78, 1984.
- [6] M. Klazar. Twelve countings with rooted plane trees. *European J. Combin.*, 18(2):195–210, 1997.
- [7] B. Knudsen. Optimal multiple parsimony alignment with affine gap cost using a phylogenetic tree. *Lecture Notes in Bioinformatics*, 2812:433–446, 2003.
- [8] S. Li and S. Wang. Enumerating the total number of subtrees of trees. Preprint, <http://arxiv.org/abs/1206.2975>.
- [9] 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.
- [10] J. L. Martin, M. Morin, and J. D. Wagner. On distinguishing trees by their chromatic symmetric functions. *J. Combin. Theory Ser. A*, 115(2):237–253, 2008.
- [11] N. Schmuck, S. Wagner, and H. Wang. Greedy trees, Caterpillars, and Wiener-type graph invariants. *MATCH Commun. Math. Comput. Chem.*, 68(1):273–292, 2012.
- [12] L. A. Székely and H. Wang. On subtrees of trees. *Adv. in Appl. Math.*, 34(1):138–155, 2005.
- [13] L. A. Székely, H. Wang, and T. Wu. The sum of the distances between the leaves of a tree and the ‘semi-regular’ property. *Discrete Math.*, 311(13):1197–1203, 2011.

- [14] W. T. Trotter, Jr. and J. I. Moore, Jr. Some theorems on graphs and posets. *Discrete Math.*, 15(1):79–84, 1976.
- [15] A. Vince and H. Wang. Infinitely many trees have non-Sperner subtree poset. *Order*, 24(2):133–138, 2007.
- [16] A. Vince and H. Wang. The average order of a subtree of a tree. *J. Combin. Theory Ser. B*, 100(2):161–170, 2010.
- [17] H. Wang. The extremal values of the Wiener index of a tree with given degree sequence. *Discrete Appl. Math.*, 156(14):2647–2654, 2008.
- [18] W. Yan and Y.-N. Yeh. Enumeration of subtrees of trees. *Theoret. Comput. Sci.*, 369(1-3):256–268, 2006.
- [19] X.-D. Zhang. The Laplacian spectral radii of trees with degree sequences. *Discrete Math.*, 308(15):3143–3150, 2008.
- [20] X.-D. Zhang, Q.-Y. Xiang, L.-Q. Xu, and R.-Y. Pan. The Wiener index of trees with given degree sequences. *MATCH Commun. Math. Comput. Chem.*, 60(2):623–644, 2008.
- [21] X.-M. Zhang and X.-D. Zhang. Trees with given degree sequences that have minimal subtrees. Preprint, <http://arxiv.org/abs/1209.0273>.
- [22] 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.