The Number of Subtrees of Trees with Given Degree Sequence *

Xiu-Mei Zhang[1,2], Xiao-Dong Zhang[1,†], Daniel Gray[3] and Hua Wang[4]

[1]Department of Mathematics, Shanghai Jiao Tong University

800 Dongchuan road, Shanghai, 200240, P. R. China

[2]Department of Mathematics,Shanghai sandau University

2727 jinhai road, Shanghai, 201209, P. R. China

[3]Department of Mathematics, University of Florida

Gainesville, FL 32611, United States

[4]Department of Mathematical Sciences, Georgia Southern University

Stateboro, GA 30460, United States

Abstract

This paper investigates some properties of the number of subtrees of a tree with given degree sequence. These results are used to characterize trees with the given degree sequence that have the largest number of subtrees, which generalizes the recent results of Kirk and Wang. These trees coincide with those which were proven by Wang and independently Zhang et al. to minimize the Wiener index. We also provide a partial ordering of the extremal trees with different degree sequences, some extremal results follow as corrollaries.

Key words: Tree; subtree; degree sequence; majorization;

AMS Classifications: 05C05, 05C30

*This work is supported by the National Natural Science Foundation of China (No:10971137), the National Basic Research Program (973) of China (No.2006CB805900), and a grant of Science and Technology Commission of Shanghai Municipality (STCSM, No: 09XD1402500) .

†Corresponding author: Xiao-Dong Zhang (Email: xiaodong@sjtu.edu.cn)

1 Introduction

All graphs in this paper will be finite, simple and undirected. A *tree* $T = (V, E)$ is a connected, acyclic graph where $V(T)$ and $E(T)$ denote the vertex set and edge set respectively. We refer to vertices of degree 1 of T as *leaves*. The unique path connecting two vertices u, v in T will be denoted by $P_T(u, v)$. The number of edges on $P(u, v)$ is called distance $dist_T(u, v)$, or for short $dist(u, v)$ between them. We call a tree (T, r) rooted at the vertex r (or just by T if it is clear what the root is) by specifying a vertex $r \in V(T)$. The *height* of a vertex v of a rooted tree T with root r is $h_T(v) = dist_T(r, v)$. For any two different vertices u, v in a rooted tree (T, r), we say that v is a *successor* of u and u is an *ancestor* of v if $P_T(r, u) \subset P_T(r, v)$. Furthermore, if u and v are adjacent to each other and $dist_T(r, u) = dist_T(r, v) - 1$, we say that u is the *parent* of v and v is a *child* of u. Two vertices u, v are siblings of each other if they share the same parent. A subtree of a tree will often be described by its vertex set.

The number of subtrees of a tree has received much attention. It is well known that the path P_n and the star $K_{1,n-1}$ have the most and least subtrees among all trees of order n, respectively. The binary trees that maximize or minimize the number of subtrees are characterized in [5, 7].

Formulas are given to calculate the number of subtrees of these extremal binary trees. These formulas use a new representation of integers as a sum of powers of 2. Number theorists have already started investigating this new binary representation [1]. Also, the sequence of the number of subtrees of these extremal binary trees (with $2l$ leaves, $l = 1, 2, \cdots$) appears to be new [4]. Later, a linear-time algorithm to count the subtrees of a tree is provided in [11].

In a related paper [6], the number of leaf-containing subtrees are studied for binary trees. The results turn out to be useful in bounding the number of acceptable residue configurations. See [3] for details.

An interesting fact is that among binary trees of the same size, the extremal one that minimizes the number of subtrees is exactly the one that maximizes some chemical indices such as the well known Wiener index, and vice versa. In [2], subtrees of trees with given order and maximum vertex degree are studied. The extremal trees coincide with the ones for the Wiener index as well. Such correlations between different topological indices of trees are studied in [8].

Recently, in [13] and [9] respectively, extremal trees are characterized regarding the Wiener index with a given degree sequence. Then it is natural to consider the following question.

Problem 1.1 *Given the degree sequence and the number of vertices of a tree, find the upper bound for the number of subtrees, and characterize all extremal trees that attain this bound.*

It will not be a surprise to see that such extremal trees coincide with the ones that attain the minimum Wiener index. Along this line, we also provide an ordering of the degree sequences according to the largest number of subtrees. With our main results, Theorems 2.3 and 2.4, one can deduce extremal graphs with the largest number of subtrees in some classes of graphs. This generalizes the results of [5], [2], etc.

The rest of this paper is organized as follows: In Section 2, some notations and the main theorems are stated. In Section 3, we present some observations regarding the structure of the extremal trees. In Section 4, we present the proofs of the main theorems. In Section 5, we show, as corollaries, characterizations of the extremal trees in different categories of trees including previously known results.

2 Preliminaries

For a nonincreasing sequence of positive integers $\pi = (d_0, \cdots, d_{n-1})$ with $n \geq 3$, let $\mathcal{T}_\pi$ denote the set of all trees with π as its degree sequence. We can construct a special tree $T^*_\pi \in \mathcal{T}_\pi$ by using breadth-first search method as follows. Firstly, label the vertex with the largest degree d_0 as v_{01} (the root). Secondly, label the neighbors of v_0 as $v_{11}, v_{12}, \ldots, v_{1d_0}$ from left to right and let $d(v_{1i}) = d_i$ for $i = 1, \cdots, d_0$. Then repeat the second step for all newly labeled vertices until all degrees are assigned. For example, if $\pi = (4, 4, 3, 3, 3, 3, 3, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$, T^*_π is shown in Fig. 1. There is a vertex v_{01} (the root) in layer 0 with the largest degree 4; its four neighbors are labeled as $v_{11}, v_{12}, v_{13}, v_{14}$ in layer 1, with degrees 4, 3, 3, 3 from left to right; nine vertices $v_{21}, v_{22}, \cdots, v_{29}$ in layer 2; five vertices $v_{31}, v_{32}, v_{33}, v_{34}, v_{35}$ in layer 3. The number of vertices in each layer i, denoted by s_i can be easily calculated as $s_0 = 1$, $s_1 = d_0 = 4$, $s_2 = d_1 + d_2 + d_3 + d_4 - s_1 = 4 + 3 + 3 + 3 - 4 = 9$, and $s_3 = d_5 + \cdots + d_{13} - s_2 = 5$.

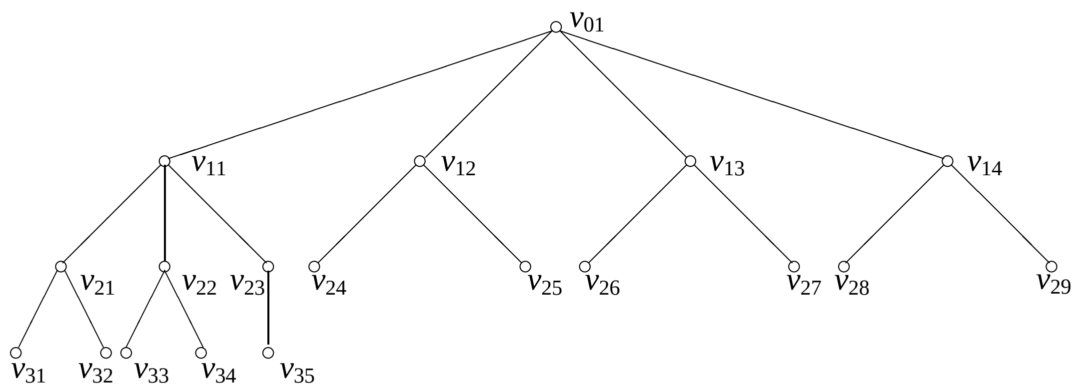

Figure 1

To explain the structure and properties of T^*_π, we need the following notation from [12].

Definition 2.1 *([12]) Let $T = (V, E)$ be a tree with root v_0. A well-ordering $\prec$ of the vertices is called a BFS-ordering if $\prec$ satisfies the following properties.*

(1) If $u, v \in V$, and $u \prec v$, then $h(u) \leq h(v)$ and $d(u) \geq d(v)$;

(2) If there are two edges $uu_1 \in E(T)$ and $vv_1 \in E(T)$ such that $u \prec v$, $h(u) = h(u_1) - 1$ and $h(v) = h(v_1) - 1$, then $u_1 \prec v_1$.

We call trees that have a BFS-ordering of its vertices a BFS-tree.

It is easy to see that T^*_π has a BFS-ordering and any two BFS-trees with degree sequence π are isomorphic (for example, see [12]). And the BFS-trees are extremal with respect to the Laplacian spectral radius.

Let $\pi = (d_0, \cdots, d_{n-1})$ and $\pi' = (d'_0, \cdots, d'_{n-1})$ be two nonincreasing sequences. If $\sum_{i=0}^{k} d_i \leq \sum_{i=0}^{k} d'_i$ for $k = 0, \cdots, n-2$ and $\sum_{i=0}^{n-1} d_i = \sum_{i=0}^{n-1} d'_i$, then the sequence π' is said to *major* the sequence π and denoted by $\pi \triangleleft \pi'$. It is known that the following holds (for example, see [10] or [12]).

Proposition 2.2 *(Wei [10]) Let $\pi = (d_0, \cdots d_{n-1})$ and $\pi' = (d'_0, \cdots, d'_{n-1})$ be two nonincreasing graphic degree sequences. If $\pi \triangleleft \pi'$, then there exists a series of graphic degree sequences $\pi_1, \cdots, \pi_k$ such that $\pi \triangleleft \pi_1 \triangleleft \cdots \triangleleft \pi_k \triangleleft \pi'$, where π_i and π_{i+1} differ at exactly two entries, say d_j (d'_j) and d_k (d'_k) of π_i (π_{i+1}), with $d'_j = d_j + 1$, $d'_k = d_k - 1$ and $j < k$.*

The main results of this paper can be stated as follows.

Theorem 2.3 *With a given degree sequence π, T_π^* is the unique tree with the largest number of subtrees in $\mathcal{T}_\pi$.*

Theorem 2.4 *Given two different degree sequences π and π_1. If $\pi \triangleleft \pi_1$, then the number of subtrees of T_π^* is less than the number of subtrees of T_{π_1}.*

3 Some Observations

In order to prove Theorems 2.3 and 2.4, we need to introduce some more terminologies. For a vertex v of a rooted tree (T, r), let $T(v)$, the subtree induced by v, denote the subtree of T (rooted at v) that is induced by v and all its successors. For a tree T and vertices $v_1, v_2, \ldots, v_{m-1}, v_m$ of T, let $f_T(v_1, v_2, \ldots, v_{m-1}, v_m)$ denote the number of subtrees of T that contain the vertices $v_1, v_2, \ldots, v_{m-1}, v_m$. In particular, $f_T(v)$ denotes the number of subtrees of T that contain v. Let $\varphi(T)$ denote the number of non-empty subtrees of T.

Let W be a tree and x, y be two vertices of W. The path $P_W(x, y)$ from x to y can be denoted by $x_m x_{m-1} \ldots x_2 x_1 y_1 y_2 \ldots y_{m-1} y_m$ for odd $dist(x, y)$ or $x_m x_{m-1} \ldots x_2 x_1 z y_1 y_2 \ldots y_{m-1} y_m$ for even $dist(x, y)$, where $x_m \equiv x, y_m \equiv y$. Let G_1 be the graph resulted from W by deleting all edges in $P_W(x, y)$. The connected components (in G_1) containing x_i, y_i and z are denoted by X_i, Y_i and Z, respectively, for $i = 1, 2, \ldots, m$. We also let $X_{\geq k}$ be the connected component of W containing x_k after deleting the edge $x_{k-1}x_k$ and $Y_{\geq k}$ be the connected component of W containing y_k after deleting the edge $y_{k-1}y_k$, for $k = 1, \cdots, m$. Figure 2 shows such a labelling according to a path of odd length (without z).

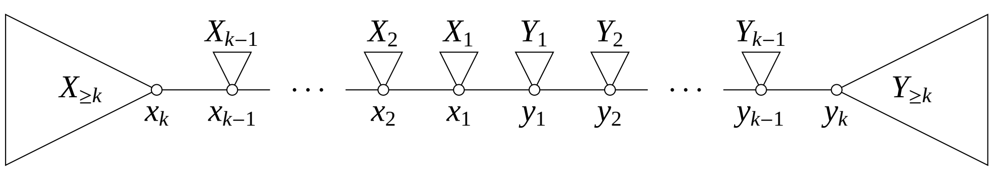

Figure 2: Labelling of a path and the components

We need the next two lemmas from [2] to proceed.

Lemma 3.1 *([2]) Let W be a tree with a path $P_W(x_m, y_m) = x_m x_{m-1} \ldots x_2 x_1 (z) y_1 y_2 \ldots y_{m-1} y_m$ from x_m to y_m. If $f_{X_i}(x_i) \geq f_{Y_i}(y_i)$ for $i = 1, 2, \ldots, m$, then $f_W(x_m) \geq f_W(y_m)$. Furthermore, if this inequality holds, then $f_W(x_m) = f_W(y_m)$ if and only if $f_{X_i}(x_i) = f_{Y_i}(y_i)$ for $i = 1, 2, \ldots, m$.*

Now let X and Y be two rooted trees with roots x' and y'. Let T be a tree containing vertices x and y. Then we can build T' by identifying the root x' of X with x of T and the root y' of Y with y of T, and T'' by identifying the root x' of X with y of T and the root y' of Y with x of T.

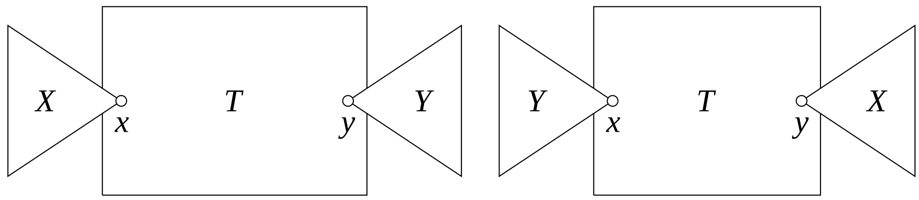

Figure 3: Constructing T' (left) and T'' (right)

Lemma 3.2 *([2]) Let T ,T',T'' be as in Figure 2. If $f(x) \geq f_W(y)$ and $f_X(x) \leq f_Y(y)$, then $\varphi(T'') \geq \varphi(T')$ with equality if and only if $f_T(x) = f_T(y)$ or $f_X(x') = f_Y(y')$.*

From Lemmas 3.1 and 3.2, we immediately achieve the following observation. We leave the proof to the reader.

Lemma 3.3 *Let T be a tree in $\mathcal{T}_\pi$ and $P(x_m, y_m) = x_m x_{m-1} \ldots x_2 x_1 (z) y_1 y_2 \ldots y_{m-1} y_m$ be a path of T. Let T' be the tree from T by deleting the two edges $x_k x_{k+1}$ and $y_k y_{k+1}$ and adding two edges $x_{k+1} y_k$ and $y_{k+1} x_k$. If $f_{X_i}(x_i) \geq f_{Y_i}(y_i)$ for $i = 1, \cdots, k$ and $1 \leq k \leq m-1$, and $f_{X_{\geq k+1}}(x_{k+1}) \leq f_{Y_{\geq k+1}}(y_{k+1})$, then*

$$\varphi(T) \leq \varphi(T')$$

with equality if and only if $f_{X_{\geq k+1}}(x_{k+1}) = f_{Y_{\geq k+1}}(y_{k+1})$ or $f_{X_i}(x_i) = f_{Y_i}(y_i)$ for $i = 1, \cdots, k$.

For convenience, we refer to trees that maximize the number of subtrees as *optimal*. In terms of the structure of the optimal tree, we have the following version of Lemma 3.3.

Corollary 3.4 *Let T be an optimal tree in $\mathcal{T}_\pi$ and $P(x_m, y_m) = x_m x_{m-1} \ldots x_2 x_1 (z) y_1 y_2 \ldots y_{m-1} y_m$ be a path of T. If $f_{X_i}(x_i) \geq f_{Y_i}(y_i)$ for $i = 1, \cdots, k$ with at least one strict inequality and $1 \leq k \leq m-1$, then $f_{X_{\geq k+1}}(x_{k+1}) \geq f_{Y_{\geq k+1}}(y_{k+1})$.*

Lemma 3.5 *Let T be an optimal tree in $\mathcal{T}_\pi$ and $P(x_m, y_m) = x_m x_{m-1} \ldots x_2 x_1 (z) y_1 y_2 \ldots y_{m-1} y_m$ be a path of T. If $f_{X_i}(x_i) \geq f_{Y_i}(y_i)$ for $i = 1, \cdots, k$ with at least one strict inequality and $1 \leq k \leq m-1$, then $f_{X_{k+1}}(x_{k+1}) \geq f_{Y_{k+1}}(y_{k+1})$.*

Proof. If $k = m-1$, then by Corollary 3.4, the assertion holds since $f_{X\geq m}(x_m) = f_{X_m}(x_m)$ and $f_{Y\geq m}(y_m) = f_{Y_m}(y_m)$. Hence we assume that $1 \leq k \leq m-2$. Suppose that $f_{X_{k+1}}(x_{k+1}) < f_{Y_{k+1}}(y_{k+1})$. Denote by M the number of subtrees of T not containing vertices x_k and y_k. Let W be the connected component of T by deleting the two edges x_kx_{k+1} and y_ky_{k+1} containing vertices x_k and y_k. Then

$$\begin{aligned}\varphi(T) &= \{1 + f_{X_{k+1}}(x_{k+1})[1 + f_{X_{\geq k+2}}(x_{k+2})]\}\,[f_W(x_k) - f_W(x_k, y_k)] + \\ &\quad \{1 + f_{Y_{k+1}}(y_{k+1})[1 + f_{Y_{\geq k+2}}(y_{k+2})]\}\,[f_W(y_k) - f_W(x_k, y_k)] + \\ &\quad \{1 + f_{X_{k+1}}(x_{k+1})[1 + f_{X_{\geq k+2}}(x_{k+2})]\}\,\{1 + f_{Y_{k+1}}(y_{k+1})[1 + f_{Y_{\geq k+2}}(y_{k+2})]\}\, f_W(x_k, y_k) + M.\end{aligned}$$

Oh the other hand, let T' be the tree from T by deleting four edges x_kx_{k+1}, $x_{k+1}x_{k+2}$, y_ky_{k+1} and $y_{k+1}y_{k+2}$ and adding four edges x_ky_{k+1}, $y_{k+1}x_{k+2}$, y_kx_{k+1} and $x_{k+1}y_{k+2}$. Clearly, $T' \in \mathcal{T}_\pi$ and

$$\begin{aligned}\varphi(T') &= \{1 + f_{Y_{k+1}}(y_{k+1})[1 + f_{X_{\geq k+2}}(x_{k+2})]\}\,[f_W(x_k) - f_W(x_k, y_k)] + \\ &\quad \{1 + f_{X_{k+1}}(x_{k+1})[1 + f_{Y_{\geq k+2}}(y_{k+2})]\}\,[f_W(y_k) - f_W(x_k, y_k)] + \\ &\quad \{(1 + f_{Y_{k+1}}(y_{k+1})[1 + f_{X_{\geq k+2}}(x_{k+2})]\}\,\{1 + f_{X_{k+1}}(x_{k+1})[1 + f_{Y_{\geq k+2}}(y_{k+2})]\}\, f_W(x_k, y_k) + M.\end{aligned}$$

Hence

$$\begin{aligned}\varphi(T') - \varphi(T) &= (f_{Y_{k+1}}(y_{k+1}) - f_{X_{k+1}}(x_{k+1}))\{[1 + f_{X_{\geq k+2}}(x_{k+2})][f_W(x_k) - f_W(x_k, y_k)] - \\ &\quad [1 + f_{Y_{\geq k+2}}(y_{k+2})][f_W(y_k) - f_W(x_k, y_k)] + f_w(x_k, y_k)(f_{X_{\geq k+2}}(x_{k+2}) - f_{Y_{\geq k+2}}(y_{k+2}))\}.\end{aligned}$$

Obviously, we have $f_W(y_k) > f_W(x_k, y_k)$ and $f_W(x_k) > f_W(x_k, y_k)$. By Lemma 3.1, we have $f_W(x_k) > f_W(y_k)$. Further by Corollary 3.4, we have $f_{X_{\geq k+1}}(x_{k+1}) \geq f_{Y_{\geq k+1}}(y_{k+1})$. Since $f_{X_{\geq k+1}}(x_{k+1}) = f_{X_{k+1}}(x_{k+1})(1 + f_{X_{\geq k+2}}(x_{k+2}))$ and $f_{Y_{\geq k+1}}(y_{k+1}) = f_{Y_{k+1}}(y_{k+1})(1 + f_{Y_{\geq k+2}}(y_{k+2}))$, we have $f_{X_{\geq k+2}}(x_{k+2}) \geq f_{Y_{\geq k+2}}(y_{k+2})$ since we assumed $f_{X_{k+1}}(x_{k+1}) < f_{Y_{k+1}}(y_{k+1})$. Therefore, $\varphi(T') > \varphi(T) > 0$, contradicting to the optimality of T. So the assertion holds. ■

Lemma 3.6 *Let P be a path of an optimal T in $\mathcal{T}_\pi$ whose end vertices are leaves.*
(i) *If the length of P is odd $(2m-1)$, then the vertices of P can be labeled as $x_m x_{m-1} \cdots x_1$ $y_1 y_2 \cdots y_m$ such that*

$$f_{X_1}(x_1) \geq f_{Y_1}(y_1) \geq f_{X_2}(x_2) \geq f_{Y_2}(y_2) \geq \cdots \geq f_{X_m}(x_m) = f_{Y_m}(y_m) = 1.$$

(ii) *If the length of P is even $(2m)$, then the vertices of P can be labeled as $x_{m+1} x_m x_{m-1} \cdots x_1$ $y_1 y_2 \cdots y_m$ such that*

$$f_{X_1}(x_1) \geq f_{Y_1}(y_1) \geq f_{X_2}(x_2) \geq f_{Y_2}(y_2) \geq \cdots \geq f_{X_m}(x_m) \geq f_{Y_m}(y_m) = f_{X_{m+1}}(x_{m+1}) = 1.$$

Proof. We provide the proof of part (i), part (ii) can be shown in a similar manner.

Obviously, the vertices of P may be labeled as $x_r x_{r-1} \cdots x_1 y_1 y_2 \cdots y_s$ such that f_{X_1} the maximum among f_{X_i} and f_{X_j} for $i = 1, 2, \cdots, r$ and $j = 1, 2, \cdots, s$, where $r + s = 2m$. Therefore, there is only one of the following three cases:

Case 1: If the number of the maximum components is one, then there exists a $1 \leq k \leq m$ such that

$$f_{X_1}(x_1) > f_{Y_1}(y_1), \quad f_{Y_1}(y_1) = f_{X_2}(x_2), \cdots, f_{Y_{k-1}}(y_{k-1}) = f_{X_k}(x_k), f_{Y_k}(y_k) > f_{X_{k+1}}(x_{k+1}) \qquad (1)$$

Next we will prove (1). It is divided into three subcases.

Case 1.1: If $f_{Y_1}(y_1) > f_{X_2}(x_2)$, then we have $k = 1$ and (1)holds.

Case 1.2: If $f_{Y_1}(y_1) < f_{X_2}(x_2)$, then the vertices of P may be relabeled such thaty_i is instead by x_{i+1} for $i = 1, \cdots, s$ and X_i is instead by y_{i-1} for $i = 2, \cdots, r$. Hence it is the same as the subcase 1.1.

Case 1.3: If $f_{Y_1}(y_1) = f_{X_2}(x_2)$. Then we must have $f_{Y_2}(y_2) > f_{X_3}(x_3)$ or $f_{Y_2}(y_2) < f_{X_3}(x_3)$ or $f_{Y_2}(y_2) = f_{X_3}(x_3)$.

Case 1.3.1: If $f_{Y_2}(y_2) > f_{X_3}(x_3)$, then we have $k = 2$ and (1)holds.

Case 1.3.2: If $f_{Y_2}(y_2) < f_{X_3}(x_3)$, then the vertices of P may be relabeled such thaty_i is instead by x_{i+1} for $i = 1, \cdots, s$ and X_i is instead by y_{i-1} for $i = 2, \cdots, r$. Hance, the case is the same as the subcase 1.3.1.

Case 1.3.3: If $f_{Y_2}(y_2) = f_{X_3}(x_3)$, we can continue to analyze like $f_{Y_1}(y_1) = f_{X_2}(x_2)$. Then we have $k \geq 3$ and (1)holds. Next we will prove that if (1) holds, then we must have

$$r = s = m.$$

Otherwise, if $r < s$, then by Lemma 3.5, $f_{X_i}(x_i) \geq f_{Y_i}(y_i)$ for $i = 1, \cdots, r$. Hence by Corollary 3.4 we have $f_{X_{\geq r}}(x_r) \geq f_{Y_{\geq r}}(y_r)$. On the other hand, it is clear that $f_{X_{\geq r}}(x_r) = 1$ and $f_{Y_{\geq r}}(y_r) \geq 2$, contradiction.

If $r > s$, then $r \geq s+2$ since $r+s = 2m$. Now we consider the path from vertex x_{s+1} to y_s. By Lemma 3.5, we have $f_{Y_i}(y_i) \geq f_{X_{i+1}}(x_{i+1})$ for $i = 1, \cdots, s$. Further, by Corollary 3.4, we have $f_{Y_{\geq s}}(y_s) \geq f_{X_{\geq s+1}}(x_{s+1})$. Similarly, since $f_{Y_{\geq s}}(y_s) = 1$ and $f_{X_{\geq s+1}}(x_{s+1}) \geq 2$, contradiction. Therefore $r = s = m$.

Now by Lemma 3.5 applied to the path from x_m to y_m, we have $f_{X_i}(x_i) \geq f_{Y_i}(y_i)$ for $i = 1, \cdots, m$. On the other hand, by Lemma 3.5 applied to the path from y_{m-1} to x_m, we have $f_{Y_i}(y_i) \geq f_{X_{i+1}}(x_{i+1})$ for $i = 1, 2, \ldots, m-1$. Hence the assertion holds.

Case 2: If the number of the maximum components is $2k \geq 2$. Then the path P can be labeled as $x_m x_{m-1} \cdots x_1 y_1 y_2 \cdots y_m$ such that

$$f_{X_1}(x_1) = f_{Y_1}(y_1) = \cdots = f_{X_k}(x_k) = f_{Y_k}(y_k) > f_{X_{k+1}}(x_{k+1}) \geq f_{Y_{k+1}}(y_{k+1}), \tag{2}$$

and the vertices $x_1, x_2, \cdots, x_k, y_1, y_2, \cdots, y_k$ are in the maximum components respectively. That is to say all the maximum components are adjoining. Otherwise, there must be two pair vertices satisfying the first inequality in (1). Hence either of them, the vertices of P may be labeled as $x_{r_1} x_{r_1-1} \cdots x_1 y_1 y_2 \cdots y_{s_1}$ or $x_{r_2} x_{r_1-1} \cdots x_1 y_1 y_2 \cdots y_{s_2}$. By the case 1, we can have $r_1 = r_2 = s_1 = s_2 = m$. But it is impossible.Therefore, if there are more than one component with the most subtrees containing the vertex on the path P, then all of them must adjoin.

Case 3: If the number of the maximum components is $2k + 1 > 2$. Then the path P can be labeled as $x_m x_{m-1} \cdots x_1 y_1 y_2 \cdots y_m$ such that

$$f_{X_1}(x_1) = f_{Y_1}(y_1) = \cdots = f_{X_k}(x_k) = f_{Y_k}(y_k) = f_{X_{k+1}}(x_{k+1}) > f_{Y_{k+1}}(y_{k+1}), \tag{3}$$

We omits the details.

Then Cases (2) or (3) can be handled in the same manner, we omit the details here. ■

Following the conditions in Lemma 3.6, we have the following.

Lemma 3.7 **(i)** *If case (i) of Lemma 3.6 holds, then*

$$f_T(x_1) \geq f_T(y_1) > f_T(x_2) \geq f_T(y_2) > \cdots > f_T(x_m) \geq f_T(y_m).$$

Moreover, if $f_T(x_k) = f_T(y_k)$ *for some* $1 \leq k \leq m$, *then* $f_T(x_i) = f_T(y_i)$ *for* $i = k, \cdots, m$.

(ii) *If case (ii) of Lemma 3.6 holds, then*

$$f_T(x_1) > f_T(y_1) \geq f_T(x_2) > f_T(y_2) \geq \cdots \geq f_T(x_m) > f_T(y_m) \geq f_T(x_{m+1}).$$

Moreover, if $f_T(y_k) = f_T(x_{k+1})$ *for some* $1 \leq k \leq m$, *then* $f_T(y_i) = f_T(x_{i+1})$ *for* $i = k, \cdots m$.

Proof. We only prove part (i), part (ii) is similar.

For any $2 \leq k \leq m$, let W_{k-1} be the connected component of T containing vertices x_{k-1} and y_{k-1} after removing the edges $x_{k-1}x_k$ and $y_{k-1}y_k$. For $k = 1$ and $k = m$, it is easy to see

$$f_T(x_1) - f_T(y_1) = f_{X_1}(x_1)(1 + f_{X \geq 2}(x_2)) - f_{Y_1}(y_1)(1 + f_{Y \geq 2}(y_2))$$

and

$$f_T(x_m) - f_T(y_m) = f_{X_m}(x_m)(1 + f_{W_{m-1}}(x_{m-1})) - f_{Y_m}(y_m)(1 + f_{W_{m-1}}(y_{m-1})).$$

Moreover,

$$f_T(x_k) = f_{X_k}(x_k)(1+f_{X_{\geq k+1}}(x_{k+1}))(1+f_{W_{k-1}}(x_{k-1})+f_{W_{k-1}}(x_{k-1},\cdots,y_{k-1})f_{Y_k}(y_k)(1+f_{Y_{\geq k+1}}(y_{k+1}))) \tag{4}$$

and

$$f_T(y_k) = f_{Y_k}(y_k)(1+f_{Y_{\geq k+1}}(y_{k+1}))(1+f_{W_{k-1}}(y_{k-1})+f_{W_{k-1}}(y_{k-1},\cdots,x_{k-1})f_{X_k}(x_k)(1+f_{X_{\geq k+1}}(x_{k+1}))). \tag{5}$$

By equations (4) and (5), we have

$$\begin{aligned} f_T(x_k) - f_T(y_k) &= f_{X_k}(x_k)(1 + f_{W_{k-1}}(x_{k-1}))(1 + f_{X_{\geq k+1}}(x_{k+1})) \\ &\quad - f_{Y_k}(y_k)(1 + f_{W_{k-1}}(y_{k-1}))(1 + f_{Y_{\geq k+1}}(y_{k+1})). \end{aligned} \tag{6}$$

Now we claim that for $1 \leq k \leq m-1$,

$$f_{X_{\geq k+1}}(x_{k+1}) \geq f_{Y_{\geq k+1}}(y_{k+1}), \tag{7}$$

If there is at least one strict inequality in $f_{X_i}(x_i) \geq f_{Y_i}(y_i)$ for $i = 1, \cdots, k$, then by Lemma 3.5, (7) holds.

If $f_{X_i}(x_i) = f_{Y_i}(y_i)$ for $i = 1, \cdots, k$ and there exists a $k < l < m$ such that $f_{X_i}(x_i) = f_{Y_i}(y_i)$ for $i = 1, \cdots, l-1$ and $f_{X_l}(x_l) > f_{Y_l}(y_l)$. Then by Lemma 3.5, we have $f_{X_{\geq l+1}}(x_{l+1}) \geq f_{Y_{\geq l+1}}(y_{l+1})$. Moreover,

$$f_{X_{\geq k+1}}(x_{k+1}) = \sum_{j=k+1}^{l} \prod_{i=k+1}^{j} f_{X_i}(x_i) + f_{X_{\geq l+1}}(x_{l+1}) \prod_{i=k+1}^{l} f_{X_i}(x_i) \tag{8}$$

and

$$f_{Y_{\geq k+1}}(y_{k+1}) = \sum_{j=k+1}^{l} \prod_{i=k+1}^{j} f_{Y_i}(y_i) + f_{Y_{\geq l+1}}(y_{l+1}) \prod_{i=k+1}^{l} f_{Y_i}(y_i). \tag{9}$$

By equations (8) and (9), the claim holds.

If $f_{X_i}(x_i) = f_{Y_i}(y_i)$ for $i = 1, \cdots, m$, then by equations (8) and (9), we have $f_{X_{\geq k+1}}(x_{k+1}) = f_{Y_{\geq k+1}}(y_{k+1})$ and the claim holds.

Hence (7) is proved.

On the other hand, by Lemma 3.1, we have $f_{W_{k-1}}(x_{k-1}) \geq f_{W_{k-1}}(y_{k-1})$. Together with (7), we see that (6) ≥ 0. Then $f_T(x_k) \geq f_T(y_k)$.

Now we prove $f_T(y_k) \geq f_T(x_{k+1})$ for any $1 \leq k \leq m-1$. Let U_k be the connected component of T containing vertex x_k after removing the edges $y_{k-1}y_k$ (if $k = 1$, let $y_0 = x_1$)

and $x_k x_{k+1}$. Then

$$f_T(y_k) = f_{Y_k}(y_k)(1 + f_{Y_{\geq k+1}}(y_{k+1}))(1 + f_{U_k}(y_{k-1}) + f_{U_k}(y_{k-1}, \cdots, x_k) f_{X_{k+1}}(x_{k+1})(1 + f_{X_{\geq k+2}}(y_{k+2}))) \tag{10}$$

and

$$f_T(x_{k+1}) = f_{X_{k+1}}(x_{k+1})(1 + f_{X_{\geq k+2}}(x_{k+2}))(1 + f_{U_k}(x_k) + f_{U_k}(x_k, \cdots, y_{k-1}) f_{Y_k}(y_k)(1 + f_{Y_{\geq k+1}}(y_{k+1}))). \tag{11}$$

Similar to (7), we can show that $f_{Y_{\geq k+1}}(y_{k+1}) \geq f_{X_{\geq k+2}}(x_{k+2})$. By Lemma 3.1, we have $f_{U_k}(y_{k-1}) \geq f_{U_k}(x_k)$. Hence (10) and (11) imply that

$$\begin{aligned} f_T(y_k) - f_T(x_{k+1}) &= f_{Y_k}(y_k)(1 + f_{U_k}(y_{k-1}))(1 + f_{Y_{\geq k+1}}(y_{k+1})) \\ &\quad - f_{X_{k+1}}(x_{k+1})(1 + f_{U_k}(x_k))(1 + f_{X_{\geq k+2}}(x_{k+2})) \\ &= f_{Y_{\geq k}}(y_k)(1 + f_{U_k}(y_{k-1})) - f_{X_{\geq k+1}}(x_{k+1})(1 + f_{U_k}(x_k)) \geq 0. \end{aligned} \tag{12}$$

Moreover, if $f_T(x_k) = f_T(y_k)$ for some $1 \leq k \leq m$, then by (6), we have

$$f_{X_k}(x_k) = f_{Y_k}(y_k), \quad f_{X_{\geq k}}(x_k) = f_{Y_{\geq k}}(y_k), \quad f_{W_{k-1}}(x_{k-1}) = f_{W_{k-1}}(y_{k-1}). \tag{13}$$

Since $f_{X_{\geq k}} = f_{X_k}(x_k)(1+f_{X_{\geq k+1}}(x_{k+1}))$ and $f_{Y_{\geq k}} = f_{Y_k}(y_k)(1+f_{Y_{\geq k+1}}(y_{k+1}))$, we have $f_{X_{\geq k+1}}(x_{k+1}) = f_{Y_{\geq k+1}}(y_{k+1})$ by (13). On the other hand, since

$$f_{W_k}(x_k) = f_{X_k}(x_k)(1 + f_{W_{k-1}}(x_{k-1}) + f_{W_{k-1}}(x_{k-1}, \cdots, y_{k-1}) f_{Y_k}(y_k))$$

and

$$f_{W_k}(y_k) = f_{Y_k}(y_k)(1 + f_{W_{k-1}}(y_{k-1}) + f_{W_{k-1}}(y_{k-1}, \cdots, x_{k-1}) f_{X_k}(x_k)),$$

we have $f_{W_k}(x_k) = f_{W_k}(y_k)$ by (13). Hence

$$\begin{aligned} f_T(x_{k+1}) &= f_{X_{\geq k+1}}(x_{k+1})(1 + f_{W_k}(x_k) + f_{W_k}(x_k, \cdots, y_k) f_{Y_{\geq k+1}}(y_{k+1})) \\ &= f_{Y_{\geq k+1}}(y_{k+1})(1 + f_{W_k}(y_k) + f_{W_k}(x_k, \cdots, y_k) f_{X_{\geq k+1}}(x_{k+1})) = f_T(y_{k+1}). \end{aligned} \tag{14}$$

Therefore we have $f_T(x_i) = f_T(y_i)$ for $i = k, \cdots, m$.

Finally, we prove that $f_T(y_i) > f_T(x_{i+1})$ for $i = 1, \cdots, m-1$. Suppose that $f_T(y_k) = f_T(x_{k+1})$ for some $1 \leq k \leq m$. Then by equation (12), we have $f_{Y_k}(y_k) = f_{X_{k+1}}(x_{k+1})$ and $f_{Y_{\geq k}}(y_k) = f_{X_{\geq k+1}}(x_{k+1})$. Moreover,

$$f_{Y_k}(y_k)(1 + f_{Y_{\geq k+1}}(y_{k+1})) = f_{Y_{\geq k}}(y_k) = f_{X_{\geq k+1}}(x_{k+1}) = f_{X_{k+1}}(x_{k+1})(1 + f_{X_{\geq k+2}}(x_{k+2})).$$

Hence $f_{Y_{\geq k+1}}(y_{k+1}) = f_{X_{\geq k+2}}(x_{k+2})$. Continuing this way in an inductive manner, we have $f_{Y_{\geq m-1}}(y_{m-1}) = f_{X_{\geq m}}(x_m)$. But $f_{Y_{\geq m-1}}(y_{m-1}) \geq 2$ and $f_{X_{\geq m}}(x_m) = 1$, contradiction.

Combining the above results, we have proved part (i). ■

The next Lemma relates the number of subtrees to the structure of the tree.

Lemma 3.8 *For a path $P(x_m, y_m) = x_m x_{m-1} \ldots x_2 x_1 (z) y_1 y_2 \ldots y_{m-1} y_m$ in an optimal tree T, if $f_{X_i}(x_i) \geq f_{Y_i}(y_i)$ for $i = 1, \cdots, k$, $1 \leq k \leq m-1$, then $d(x_k) \geq d(y_k)$.*

Moreover, if $f_{X_i}(x_i) = f_{Y_i}(y_i)$ for $i = 1, \cdots, k$, $1 \leq k \leq m-1$, then $d(x_k) = d(y_k)$.

Proof. Suppose that $d(x_k) < d(y_k)$, let $r = d(y_k) - d(x_k) \geq 1$ and $y_k u_i \in Y_{\geq k}$ for $i = 1, \cdots, r$.

Further let W be the connected component of T containing vertices x_k and y_k after removing the r edges $y_k u_1, \cdots, y_k u_r$. Let X be the single vertex x_k and let Y be the connected component of T containing vertex y_k after removing all edges incident to y_k except for the r edges $y_k u_1, \cdots, y_k u_r$. Since $f_{X_i}(x_i) \geq f_{Y_i}(y_i)$ for $i = 1, \cdots, k$, it is easy to see that $f_W(x_k) > f_W(y_k)$ and $f_X(x_k) = 1 < 2 \leq f_Y(y_k)$. By Lemma 3.2, there exists another tree $T' \in \mathcal{T}_\pi$ such that $\varphi(T) < \varphi(T')$, contradicting to the optimality of T.

Therefore the assertion holds. The case of equality is similar. ■

From Lemmas 3.6, 3.7 and 3.8 we have the following Lemma that decides the 'center' of the optimal tree.

Lemma 3.9 *Let T be an optimal tree in $\mathcal{T}_\pi$. If $f_T(v_0) = \max\{f_T(v), v \in V(T)\}$, then $d(v_0) = \max\{d(v), v \in V(T)\}$.*

Proof. The assertion clearly holds for small trees, so we assume that $|V(T)| \geq 4$. Suppose that $d(v_0) < \max\{d(v), v \in V(T)\}$. Then there exists a vertex w such that $d(v_0) < d(w)$. By Theorem 9.1 in [5], $f_T(v)$ is maximized at one or two adjacent vertices of T. Thus we have $f_T(v_0) > f_T(v)$ for $v \in V(T) \setminus \{v_0\}$, or $f_T(v_0) = f_T(v_1) > f_T(v)$ for $v \in V(T) \setminus \{v_0, v_1\}$ and $v_0 v_1 \in E(T)$.

Case 1: $f_T(v_0) > f_T(v)$ for $v \in V(T) \setminus \{v_0\}$. Hence, $f_T(v_0) > f_T(w)$. It is easy to see that v_0 is not a leaf (otherwise, let u be a neighbor of v_0 and we have $f_T(u) > f_T(v_0)$). Let P be a path containing vertex v_0 and w whose end vertices are leaves. Let the length of P be $2m-1$ (the even length case is similar). Then by Lemma 3.6, the vertices of P can be labeled as $P = x_m \cdots x_1 y_1 \cdots y_m$ such that

$$f_{X_1}(x_1) \geq f_{Y_1}(y_1) \geq f_{X_2}(x_2) \geq f_{Y_2}(y_2) \geq \cdots \geq f_{X_m}(x_m) = f_{Y_m}(y_m) = 1.$$

Hence by Lemma 3.7, we have

$$f_T(x_1) \geq f_T(y_1) \geq f_T(x_2) \geq f_T(y_2) \geq \ldots \geq f_T(x_m) \geq f_T(y_m).$$

Therefore x_1 must be v_0 and w must be x_k for $2 \leq k \leq m$ or y_j for $1 \leq j \leq m$. By Lemma 3.8, we have $d(v_0) = d(x_1) \geq d(x_k) = d(w)$ or $d(v_0) = d(x_1) \geq d(y_j) = d(w)$, contradiction. Hence the assertion holds.

Case 2: $f_T(v_0) = f_T(v_1) > f_T(v)$ for $v \in V(T) \setminus \{v_0, v_1\}$ and $v_0v_1 \in E(T)$. If $w = v_1$, then by Lemma 3.8, we have $d(w) = d(v_1) = d(v_0) < d(w)$, contradiction.

Hence we assume that $w \neq v_1$. First note that v_0 and v_1 are not leaves. Let P be a path containing vertices v_0, v_1 and w whose end vertices are leaves. Let the length of P be $2m-1$ (the even case is similar), then by Lemma 3.6, the vertices of P can be labeled as $P = x_m \cdots x_1 y_1 \cdots y_m$ such that

$$f_{X_1}(x_1) \geq f_{Y_1}(y_1) \geq f_{X_2}(x_2) \geq f_{Y_2}(y_2) \geq \cdots \geq f_{X_m}(x_m) \geq f_{Y_m}(y_m) = 1.$$

Hence by Lemma 3.7, we have

$$f_T(x_1) \geq f_T(y_1) \geq f_T(x_2) \geq f_T(y_2) \geq \ldots \geq f_T(x_m) \geq f_T(y_m).$$

Therefore $\{x_1, y_1\} = \{v_0, v_1\}$ and w must be x_k or y_k for $1 < k \leq m$. By Lemma 3.8, $d(v_0) \geq d(w)$ and $d(v_1) \geq d(w)$, contradiction.

Combining cases (1) and (2), the assertion is proved. ■

Lemma 3.10 *Let T be an optimal tree in $\mathcal{T}_\pi$. If there is a path $P = u_l u_{l-1} \cdots u_1 v_0 v_1 \cdots v_k$ with $f_T(v_0) = \max\{f_T(v) : \ v \in V(P)\}$, $f_T(u_1) \geq f_T(v_1)$, and $l = k$ (or $l = k+1$), then*

$$f_T(u_1) \geq f_T(v_1) \geq f_T(u_2) \geq \cdots \geq f_T(u_k) \geq f_T(v_k) \text{ (or } \ \geq f_T(u_{k+1}))$$

and

$$d(u_1) \geq d(v_1) \geq d(u_2) \geq \cdots \geq d(u_k) \geq d(v_k) \text{ (or } \ \geq d(u_{k+1})).$$

Proof. Clearly, there exists a path Q that contains the path P and its end vertices are leaves. We assume $l = k$ (the $l = k+1$ case is similar).

Let the length of Q be $2m-1$ (the even length case is similar). By Lemmas 3.7 and 3.8, The vertices of Q can be labeled as $Q = x_m x_{m-1} \cdots x_1 y_1 \cdots y_m$ such that

$$f_T(x_1) \geq f_T(y_1) > f_T(x_2) \geq f_T(y_2) > \ldots > f_T(x_m) \geq f_T(y_m)$$

and

$$d(x_1) \geq d(y_1) \geq d(x_2) \geq d(y_2) \geq \cdots \geq d(x_m) = d(y_m) = 1.$$

Case 1: $v_0 = x_1$. We must have $u_1 = y_1$ and $v_1 = x_2$. Then $u_i = y_i$ and $v_i = x_{i+1}$ for $i = 1, \cdots, k$. Hence the assertion holds.

Case 2: $v_0 = x_i$ for $i > 1$. Then $f_T(v_0) \geq f_T(x_1) \geq f_T(y_1) \geq f_T(x_i) = f_T(v_0)$, which implies $f_T(x_1) = f_T(y_1) = f_T(v_0)$ and contradicts to Theorem 9.1 in [5].

Case 3: $v_0 = y_i$. Then $i = 1$ and $f_T(x_1) = f_T(y_1) = f_T(v_0)$. We must have $u_1 = x_1$ and $v_1 = y_2$. Then $u_i = x_i$ and $v_i = y_{i+1}$ for $i = 1, \cdots, k$. So the assertion holds. ■

Now for an optimal tree T in $\mathcal{T}_\pi$, let $v_0 \in V(T)$ be the root of T with $f_T(v_0) = \max\{f_T(v) : v \in V(T)\}$ and $d(v_0) = \max\{d(v) : v \in V(T)\}$.

Corollary 3.11 *If there is a path $P = u_k \cdots u_1 w v_1 v_2 \cdots v_k$ with $dist(u_k, v_0) = dist(v_k, v_0) = dist(w, v_0) + k$ and $f_T(u_1) \geq f_T(v_1)$, then*

$$f_T(u_1) \geq f_T(v_1) \geq f_T(u_2) \geq \cdots \geq f_T(u_k) \geq f_T(v_k)$$

and

$$d(u_1) \geq d(v_1) \geq d(u_2) \geq \cdots \geq d(u_k) \geq d(v_k).$$

If there is a path $P = u_{k+1} \cdots u_1 w v_1 v_2 \cdots v_k$ with $dist(u_{k+1}, v_0) = dist(v_k, v_0) + 1 = dist(w, v_0) + k + 1$ and $f_T(u_1) \geq f_T(v_1)$, then

$$f_T(u_1) \geq f_T(v_1) \geq f_T(u_2) \geq \cdots \geq f_T(u_k) \geq f_T(v_k) \geq f_T(u_{k+1})$$

and

$$d(u_1) \geq d(v_1) \geq d(u_2) \geq \cdots \geq d(u_k) \geq d(v_k) \geq d(u_{k+1}).$$

Proof. If $w = v_0$, then the assertion follows from Lemma 3.10. If $w \neq v_0$, then there exists a path Q containing vertices $u_k, \cdots, u_1, w, v_0$ whose end vertices are leaves. By Lemma 3.10, we have $f_T(w) \geq f_T(u_1) \geq \cdots \geq f_T(u_k)$. Similarly, there exists a path R containing vertices $v_k, \cdots, v_1, w, v_0$ whose end vertices are leaves and we have $f_T(w) \geq f_T(v_1) \geq \cdots \geq f_T(v_k)$. Therefore $f_T(w) = \max\{f_T(v) : v \in V(P)\}$, the assertion follows from Lemma 3.10. ■

4 Proofs of Theorems 2.3 and 2.4

Now we are ready to prove Theorems 2.3 and 2.4.

Proof. of Theorem 2.3. Let T be an optimal tree in $\mathcal{T}_\pi$. By Lemma 3.9, there exists a vertex v_0 such that $f_T(v_0) = \max\{f_T(v) : \ v \in V(T)\}$ and $d(v_0) = \max\{d(v) : \ v \in V(T)\}$. Let v_0 be the root of T and put $V_i = \{v : \ dist(v, v_0) = i\}$ for $i = 0, \cdots, p+1$ with $V(T) = \bigcup_{i=0}^{p+1} V_i$. Denote by $|V_i| = s_i$ for $i = 1, \cdots, p+1$. We now can relabel the vertices of $V(T)$ by the recursion method. For V_0, relabel v_0 by v_{01} as the root of tree T. The vertices of V_1 (consisting of all neighbors v_{01}) are relabeled as $v_{11}, \cdots, v_{1,s_1}$, satisfying:

$$f_T(v_{11}) \geq f_T(v_{12}) \geq \cdots \geq f_T(v_{1,s_1})$$

and

$$f_T(v_{1i}) = f_T(v_{1j}) \ \text{implies} \ d(v_{1i}) \geq d(v_{1j}) \text{ for } 1 \leq i < j \leq s_1.$$

Generally, we assume that all vertices of V_i are relabeled as $\{v_{i1}, \cdots, v_{i,s_i}\}$ for $i = 1, \cdots, t$. Now consider all vertices in V_{t+1}. Since T is tree, it is easy to see that $s_1 = d(v_{01})$ and

$$s_{t+1} = |V_{t+1}| = d(v_{t1}) + \cdots + d(v_{t,s_t}) - s_t.$$

Hence for $1 \leq r \leq s_t$, all neighbors in V_{t+1} of v_{tr} are relabeled as

$$v_{t+1,d(v_{t1})+\cdots+d(v_{t,r-1})-(r-1)+1}, \cdots, v_{t+1,d(v_{t1})+\cdots+d(v_{t,r})-r}$$

and satisfy the conditions:

$$f_T(v_{t+1,i}) \geq f_T(v_{t+1,j}) \tag{15}$$

and

$$f_T(v_{t+1,i}) = f_T(v_{t+1,j}) \ \text{implies} \ d(v_{t+1,i}) \geq d(v_{t+1,j}) \tag{16}$$

for $d(v_{t1}) + \cdots + d(v_{t,r-1}) - (r-1) + 1 \leq i < j \leq d(v_{t1}) + \cdots + d(v_{t,r}) - r$. In this way, we have relabeled all vertices of $V(T) = \bigcup_{i=0}^{p+1} V_i$. Therefore, we are able to define a well-ordering of vertices in $V(T)$ as follows:

$$v_{ik} \prec v_{jl}, \ \text{ if } 0 \leq i < j \leq p+1 \ \text{ or } \ i = j \text{ and } 1 \leq k < l \leq s_i. \tag{17}$$

We need to prove that this well-ordering is a BFS-ordering of T. In other words, T is isomorphic to T^*_π.

We first prove, for $t = 0, \cdots, p+1$, the following inequalities.

$$f_T(v_{t1}) \geq f_T(v_{t2}) \geq \cdots \geq f_T(v_{t,s_t}) \geq f_T(v_{t+1,1}) \tag{18}$$

and

$$d(v_{t1}) \geq d(v_{t2}) \geq \cdots \geq d(v_{t,s_t}) \geq d(v_{t+1,1}). \qquad (19)$$

For any two vertices v_{ti} and v_{tj} with $1 \leq i < j \leq s_t$, there exists a path $P = v_{ti} \cdots v_{k+1,l}$ $w_k v_{k+1,r} \cdots v_{tj}$ with $l < r$, where $dist(v_{ti}, v_{01}) = dist_(v_{tj}, v_{01}) = dist(w_k, v_{01}) + t - k$. Then we have $f_T(v_{k+1,l}) \geq f_T(v_{k+1,r})$, $f_T(v_{ti}) \geq f_T(v_{tj})$ and $d(v_{ti}) \geq d(v_{tj})$ by Corollary 3.11. On the other hand, we consider the path $Q = v_{t+1,1}v_{t1} \cdots v_{11}v_{01}v_{1s_1} \cdots v_{t,s_t}$. Then $f_T(v_{t,s_t}) \geq f_T(v_{t+1,1})$ and $d(v_{t,s_t}) \geq d(v_{t+1,1})$ by Corollary 3.11. Therefore (18) and (19) hold for $t = 0, \cdots, p+1$. That is

$$f_T(v_{01}) \geq f_T(v_{11}) \geq \cdots \geq f_T(v_{1,s_1}) \geq f_T(v_{21}) \geq \cdots \geq f_T(v_{2,s_2}) \geq \cdots \geq f_T(v_{p+1,s_{p+1}}) \qquad (20)$$

and

$$d(v_{01}) \geq d(v_{11}) \geq d(v_{1,s_1}) \geq d(v_{21}) \geq \cdots \geq d(v_{2,s_2}) \geq d(v_{p+1,1}) \geq d(v_{p+1,s_{p+1}}). \qquad (21)$$

By (17), (20) and (21), it is easy to see that this well ordering satisfies all conditions in Definition 2.1. Hence T has a BFS-ordering. Further, by Proposition 2.2 in [12], T is isomorphic to T_π^*. So T_π^* is the unique optimal tree in $\mathcal{T}_\pi$ having the largest number of subtrees. ■

Proof. of Theorem 2.4. By proposition 2.2, without loss of generality, we assume that $\pi = (d_0, d_1, \cdots, d_i, \cdots, d_j, \cdots, d_{n-1})$ and $\pi_1 = (d_0, d_1, \cdots, d_i + 1, \cdots, d_j - 1, \cdots, d_{n-1})$ with $i < j$, then we have $\pi \triangleleft \pi_1$. Let T_π^* be the optimal tree in $\mathcal{T}_\pi$. By the proof of Theorem 2.3, the vertices of T_π^* can be labeled as the $V = \{v_0, \cdots, v_{n-1}\}$ such that

$$f_{T_\pi^*}(v_0) \geq f_{T_\pi^*}(v_1) \geq \cdots \geq f_{T_\pi^*}(v_{n-1})$$

and

$$d(v_0) \geq d(v_1) \geq \cdots \geq d(v_{n-1}),$$

where $d(v_l) = d_l$ for $l = 0, \cdots, n-1$. Moreover, v_0 is the root of T_π^*. There exists a vertex v_k such that $v_j v_k \in E(T_\pi^*)$ with $k > j$. Let W be the tree achieved from T_π^* by removing the subtree induced by v_k. Moreover, let X be the single vertex v_i and Y be the subtree induced by v_k with the edge $v_j v_k$ added, respectively. Clearly, $f_T(v_i) = f_W(v_i) + f_W(v_i, v_j)(f_Y(v_j) - 1)$ and $f_T(v_j) = f_W(v_j) + f_W(v_j)(f_Y(v_j) - 1)$. Hence by $f_W(v_i, v_j) < f_W(v_j)$ and $f_T(v_i) \geq f_T(v_j)$, we have $f_W(v_i) > f_W(v_j)$. On the other hand, let T_1 be the tree from T by deleting the edge $v_j v_k$ and adding the edge $v_i v_k$. Then the degree sequence of T_1 is π_1. By Lemma 3.2, we have $\varphi(T_\pi^*) < \varphi(T_1)$. Hence $\varphi(T_\pi^*) < \varphi(T_1) \leq \varphi(T_{\pi_1}^*)$.

The assertion is then proved. ■

5 Applications of the Main Theorems

In the end we use Theorems 2.3 and 2.4 to achieve extremal graphs with the largest number of subtrees in some classes of graphs. As corollaries, we provide proofs to some results in [5], [2], etc.

Let $\mathcal{T}^{(1)}_{n,\Delta}$ be the set of all trees of order n with the largest degree Δ, $\mathcal{T}^{(2)}_{n,s}$ be the set of all trees of order n with s leaves, $\mathcal{T}^{(3)}_{n,\alpha}$ be the set of all trees of order n with the independence number α and $\mathcal{T}^{(4)}_{n,\beta}$ be the set of all trees of order n with the matching number β.

Corollary 5.1 *([5]) Let T be any tree of order n. Then*

$$\binom{n+1}{2} \leq \varphi(T) \leq 2^{n-1} + n - 1$$

with left equality if and only if T is a path of order n and the right equality if and only if T is the star $K_{1,n-1}$.

Proof. Let T be a tree of order n with degree sequence τ. Let $\pi_1 = (2, \cdots, 2, 1, 1)$ and $\pi_2 = (n-1, 1, \cdots, 1)$ with n terms. Clearly the path P of order n is the only tree with the degree sequence π_1 and the star $K_{1,n-1}$ of order n is the only tree with degree sequence π_2. Furthermore, $\pi_1 \triangleleft \tau \triangleleft \pi_2$. Hence by Theorems 2.3 and 2.4, the assertion holds. ■

Corollary 5.2 *([2]) There is only one optimal tree T^*_Δ in $\mathcal{T}^{(1)}_{n,\Delta}$ with $\Delta \geq 3$, where T^*_Δ is T^*_π with degree sequence π as follows: Denote $p = \lceil \log_{(\Delta-1)} \frac{n(\Delta-2)+2}{\Delta} \rceil - 1$ and $n - \frac{\Delta(\Delta-1)^p-2}{\Delta-2} = (\Delta-1)r + q$ for $0 \leq q < \Delta - 1$. If $q = 0$, put $\pi = (\Delta, \cdots, \Delta, 1, \cdots, 1)$ with the number $\frac{\Delta(\Delta-1)^{p-1}-2}{\Delta-2} + r$ of degree Δ. If $q \geq 1$, put $\pi = (\Delta, \cdots, \Delta, q, 1, \cdots, 1)$ with the number $\frac{\Delta(\Delta-1)^{p-1}-2}{\Delta-2} + r$ of degree Δ.*

Proof. For any tree T of order n with the largest degree Δ, let $\pi_1 = (d_0, \cdots, d_{n-1})$ be the nonincreasing degree sequence of T. Assume that T^*_Δ has $p+2$ layers. Then there is a vertex in layer 0 (i.e. the root), there are Δ vertices in layer 1, there are $\Delta(\Delta-1)$ vertices in layer 2, $\cdots$, there are $\Delta(\Delta-1)^{p-1}$ vertices in layer p, there are at most $\Delta(\Delta-1)^p$ vertices in layer $p+1$. Hence

$$1 + \Delta + \Delta(\Delta-1) + \cdots + \Delta(\Delta-1)^{p-1} < n \leq 1 + \Delta + \Delta(\Delta-1) + \cdots + \Delta(\Delta-1)^p.$$

Thus

$$\frac{\Delta(\Delta-1)^p - 2}{\Delta - 2} < n \leq \frac{\Delta(\Delta-1)^{p+1} - 2}{\Delta - 2}.$$

Hence

$$p = \lceil \log_{(\Delta-1)} \frac{n(\Delta-2)+2}{\Delta} \rceil - 1$$

and there exist integers r and $0 \le q < \Delta - 1$ such that

$$n - \frac{\Delta(\Delta-1)^p - 2}{\Delta - 2} = (\Delta-1)r + q.$$

Therefore the degrees of all vertices from layer 0 to layer $p-1$ are Δ and there are r vertices in layer p with degree Δ. Denote by $m = \frac{\Delta(\Delta-1)^{p-1}-2}{\Delta-2} + r - 1$. Then there are $m+1$ vertices with degree Δ in T^*_Δ. Hence the degree sequence of $T^*_\Delta \in \mathcal{T}_{n,\Delta}$ is $\pi = (d'_0, \cdots, d'_{n-1})$ with $d'_0 = \cdots = d'_m = \Delta$, $d'_{m+1} = \cdots = d'_{n-1} = 1$ for $q = 0$; and is $\pi = (d'_0, \cdots, d'_{n-1})$ with $d'_0 = \cdots = d'_m = \Delta$, $d'_{m+1} = q$, $d'_{m+2} = \cdots = d'_{n-1} = 1$ for $q = 1$. It follows from $d_i \le \Delta$ that $\sum_{i=0}^k d_i \le \sum_{i=0}^k d'_i$ for $k = 0, \cdots, m$. Further by $d'_i = 1 \le d_i$ for $k = m+2, \cdots, n-1$, we have

$$\sum_{i=0}^{k} d_i = 2(n-1) - \sum_{i=k+1}^{n-1} d_i \le 2(n-1) - \sum_{i=k+1}^{n-1} d'_i = \sum_{i=0}^{k} d'_i$$

for $k = m+1, \cdots n-1$. Thus $\pi_1 \triangleleft \pi$. Hence by Theorems 2.3 and 2.4, $\varphi(T) \le \varphi(T^*_\Delta)$ with equality if and only if $T = T^*_\Delta$. ■

Remark If $\Delta = 3$ in Corollary 5.2, then the result is precisely Theorem 2.1 in [7].

Corollary 5.3 *There is only one optimal tree T^*_s in $\mathcal{T}^{(2)}_{n,s}$ where T^*_s is obtained from t paths of order $q+2$ and $s-t$ paths of order $q+1$ by identifying one end of the s paths. Here $n-1 = sq+t, 0 \le t < s$. In other words, for any tree of order n with s leaves,*

$$\varphi(T) \le (q+2)^t + (q+1)^{s-t+2}$$

with equality if and only if T is T^*_s.

Proof. Let T be any tree in $\mathcal{T}^{(2)}_{n,s}$ with the nonincreasing degree sequence $\pi_1 = (d_0, \cdots, d_{n-1})$. Thus $d_{n-s-1} > 1$ and $d_{n-s} = \cdots = d_{n-1} = 1$. Let T^*_π be a BFS-tree with degree sequence $\pi = (s, 2, \cdots, 2, 1, \cdots, 1)$, where there are the number s of $1'$s in π. It is easy to see that $\pi_1 \triangleleft \pi$. By Theorem 2.4, the assertion holds. ■

Corollary 5.4 *There is only one optimal tree T^*_α in $\mathcal{T}^{(3)}_{n,\alpha}$, where T^*_α is T^*_π with degree sequence $\pi = (\alpha, 2, \cdots, 2, 1, \cdots, 1)$ with numbers $n-\alpha-1$ of $2'$s and α of $1'$s, i.e., T^*_π is obtained from the star $K_{1,\alpha}$ by adding $n-\alpha-1$ pendent edges to $n-\alpha-1$ leaves of $K_{1,\alpha}$. In other words, for any tree of order n with the independence number α,*

$$\varphi(T) \le 2^{2\alpha-n+1}3^{n-\alpha-1} + 2n - \alpha - 2$$

with equality if and only if T is T^*_α.

Proof. For any tree T of order n with the independence number α, let I be an independent set of T with size α and $\tau = (d_0, \cdots, d_{n-1})$ be the degree sequence of T. If there exists a leaf u with $u \notin I$, then there exists a vertex $v \in I$ with $(u, v) \in E(T)$. Hence $I \bigcup \{u\} \setminus \{v\}$ is an independent set of T with size α. Therefore, one can always construct an independent set of T with size α that contains all leaves of T. Hence there are at most α leaves. Then $d_{n-\alpha-1} \geq 2$ and $\tau \triangleleft \pi$. By Theorems 2.3 and 2.4, the assertion holds. ■

Corollary 5.5 *There is only one optimal tree T^*_β in $\mathcal{T}^{(4)}_{n,\beta}$, where T^*_β is T^*_π with degree sequence $\pi = (n - \beta, 2, \cdots, 2, 1, \cdots, 1)$. Here the number of $1'$s is $n - \beta$. That is, T^*_π is obtained from the star $K_{1,n-\beta}$ by adding $\beta - 1$ pendent edges to $\beta - 1$ leaves of $K_{1,n-\beta}$. In other words, for any $T \in \mathcal{T}^{(4)}_{n,\beta}$,*

$$\varphi(T) \leq 2^{n-2\beta+1}3^{\beta-1} + n - \beta - 2$$

with equality if and only if T is T^*_β.

Proof. For any tree T of order n with matching number β, let $\tau = (d_0, \cdots, d_{n-1})$ be the degree sequence of T. Let M be a matching of T with size β. Since T is connected, there are at least β vertices in T such that their degrees are at least two. Hence $d_{\beta-1} \geq 2$ and $\tau \triangleleft \pi$. By Theorems 2.3 and 2.4, the assertion holds. ■

References

[1] C. Heuberger and H. Prodinger, On α-greedy expansions of numbers, *Adv. in Appl. Math.*, 38(2007), 505–525.

[2] R. Kirk and H. Wang, Largest number of subtrees of trees with a given maximum degree, *SIAM J. Discrete Math.*, 22(2008) 985–995.

[3] B. Knudsen, Optimal multiple parsimony alignment with affine gap cost using a phylogenetic tree, in *Lecture Notes in Bioinformatics,* Vol.2812, Springer-Verlag, 2003, pp. 433–446.

[4] The On-Line Encyclopedia of Integer Sequences, A092781, http://www.research.att.com/ njas/sequences.

[5] L. A. Székely and H. Wang, On subtrees of trees, *Advances.in Applied Mathematics*, 34(2005) 138–155.

[6] L. A. Székely and H. Wang, Binary trees with the largest number of subtrees with at least one leaf, *Congr. Numer.*, 177(2005) 147–169.

[7] L. A. Székely and H. Wang, Binary trees with the largest number of subtrees, *Discrete Applied Mathematics*, 155(2007) 374–385.

[8] S. G. Wagner, Correlation of graph-theoretical indeces, *SIAM J. Discrete Math.* 21(2007) 33–46.

[9] H. Wang, The extremal values of the Wiener index of a tree with given degree sequence, *Discrete Applied Mathematics*, 156(2008)2647–2654.

[10] W.-D.Wei, The class $\mathcal{U}(R, S)$ of (0,1) matrices, *Discrete Mathematics* 39(1982) 201–205.

[11] W.-G. Yan and Y.-N. Yeh, Enumeration of subtrees of trees, *Theoretical Comuter Science*, 369(2006) 256–268.

[12] X.-D. Zhang, The Laplacian spectral radii of trees with degree sequences *Discrete Mathematics* 308(2008) 3143–3150.

[13] 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(2008) 623–644.

[14] X.-D. Zhang, Y. Liu and M.-X. Han, Maximum Wiener Index of Trees with Given Degree Sequence, *MATCH Commun.Math.Comput.Chem.*, 64(2010) 661–682.