Extremal values of ratios: distance problems vs. subtree problems in trees

László A. Székely[*]

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

szekely@math.sc.edu

Hua Wang[†]

Department of Mathematical Sciences,
Georgia Southern University
Statesboro, GA 30460, USA

hwang@georgiasouthern.edu

Abstract

The authors discovered a dual behaviour of two tree indices, the Wiener index and the number of subtrees, for a number of extremal problems [*Discrete Appl. Math.* **155** (3) 2006, 374–385; *Adv. Appl. Math.* **34** (2005), 138–155]. Barefoot, Entringer and Székely [*Discrete Appl. Math.* **80**(1997), 37–56] determined extremal values of $\sigma_T(w)/\sigma_T(u)$, $\sigma_T(w)/\sigma_T(v)$, $\sigma(T)/\sigma_T(v)$, and $\sigma(T)/\sigma_T(w)$, where T is a tree on n vertices, v is in the centroid of the tree T, and u, w are leaves in T. In this paper we test how far the negative correlation between distances and subtrees go if we look for the extremal values of $F_T(w)/F_T(u)$, $F_T(w)/F_T(v)$, $F(T)/F_T(v)$, and $F(T)/F_T(w)$, where T is a tree on n vertices, v is in the subtree core of the tree T, and u, w are leaves in T—the complete analogue, changing distances to the number of subtrees. We include a number of open problems, shifting the interest towards the number of subtrees in graphs.

Keywords: Wiener index; tree; binary tree; caterpillar; star tree; good binary tree; distances in trees; subtrees of trees; extremal problems; center; centroid; subtree core.

1 Motivation

1.1 Results for the Wiener index

In chemical graph theory, the Wiener index is a topological index of a molecule, defined as the sum of the numbers of edges in the shortest paths in a chemical graph between all pairs of non-

hydrogen atoms in a molecule. Since 1947, when H. Wiener introduced this index [37] noting its correlation with boiling temperatures of paraffins, this index has turned into a frequently used graph parameter.

This paper is concerned with trees, i.e. connected and cycle-free graphs. In a tree T, for vertices $u, v \in V(T)$ let $d_T(u, v)$ denote the *distance* of the vertices, i.e. the number of edges in the unique uv path in T. The *distance of vertex w*, $\sigma_T(w)$, is defined as $\sum_{v \in V(T)} d_T(w, v)$, and the *Wiener index of the tree T*, $\sigma(T)$, is defined as $\frac{1}{2} \sum_w \sigma_T(w)$. For a survey on the Wiener index of trees, see [7]. We term degree 1 vertices in a tree as *leaves*. A tree is *binary*, if the degree of every vertex is 1 or 3. A *caterpillar tree* has a path from which the distance of every vertex is at most 1. A *complete binary tree* has an edge uv, from which all leaves have the same distance. A *good tree* is complete binary tree, or comes from a complete binary tree by deletion of some leaves in the following way. Make a planar (crossing-free) drawing of the complete binary tree such that the special uv edge is on top, and delete some pairs of leaves such that the deleted pairs of leaves make an initial segment in the left-to-right order among the leaves.
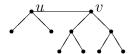

Figure 1: A good binary tree resulting after the deletion of 2 pairs of second neighbors of u.

It has been known that among trees with given number of vertices, the path maximizes and the star minimizes the Wiener index, see [8] or [22] Ex. 6.23. Regarding binary trees with given number of vertices, the *caterpillar tree* maximizes the Wiener index [9], while the *good tree* minimizes the Wiener index [9, 15]. [35, 40] generalized the concept of *good trees* to *greedy trees* and proved that greedy trees minimize the Wiener index among trees with a given degree sequence. Among trees with a given degree sequence, finding the specific caterpillar that maximizes the Wiener index turned out to be difficult [3, 41].

In [19], Lepović and Gutman conjectured that every positive integer, with only 49 explicit exceptions, is the Wiener index of some tree. This conjecture was verified independently in [32] and [36], and further extended to the class of trees with maximum degree 3 in [33]. This problem has some relevance to designing molecules with prescribed properties.

1.2 Results for subtrees of trees

In a bioinformatics paper, Knudsen [18] provided a multiple parsimony alignment algorithm, given an affine gap cost and a phylogenetic tree. In bounding the time complexity of his algorithm, a factor was the number of so-called "acceptable residue configurations", which is the number of subtrees of the phylogenetic tree containing at least one original leaf vertex. Knudsen estimated the maximum number of acceptable residue configurations among all binary trees. Let $F(T)$ denote the number of subtrees of the tree T (a subtree must have at least one vertex), and let $F_T(v)$ denote the number of subtrees of the tree T that contain vertex v.

Knudsen's work motivated our earlier papers [26, 28] to show the following results: among all trees with a given number of vertices, the *path* minimizes the number of subtrees, and the *star* maximizes the number of subtrees. Furthermore, among binary trees with a given number of vertices, the *caterpillar* minimizes the number of subtrees, and the *good tree* maximizes the number of subtrees. (We also solved exactly Knudsen's original problem, namely which binary tree maximizes the number of subtrees with at least one original leaf [27].) For trees with a given degree sequence, the number of subtrees analogue of the Wiener index result for greedy trees was shown in [42].

[5] showed that the "number of subtrees" parameter also realizes all positive integers, except 34 explicitly given numbers. The proof is somewhat number theoretic as uses representation of integers as sum of pentagonal numbers.

1.3 How far the analogy goes?

The dual behaviour of the Wiener index and the number of subtrees, shown above, is just statistical, not deterministic. Wagner [34] made an analysis of the correlation between a number of pairs of tree indices, and he found the highest (negative) correlation between the Wiener index and the number of subtrees among the indices that he considered.

Recently Taoyang Wu and us [30] found a substantial simplification and unification of the results in [26], [27], [28], regarding minimization of the Wiener index or maximization of the number of subtrees over certain families of trees, through the "semi-regular property". Making use of this general property, the extremality of the greedy tree with respect to general distance-based graph invariants are shown recently [25]. Surprisingly, the motivation for a fruitful different view came from phylogenetics, where a quantity similar to the Wiener index appeared, namely the sum of interleaf distances.

It is worthwhile to investigate how far the dual behaviour goes as the papers [26] and [28] generated considerable interest in different disciplines [4, 6, 10, 11, 12, 13, 14, 17, 20, 21, 23, 24, 31, 34, 38]. The analogy certainly goes one step further, to describe the "middle part" of the tree.

1.4 The centroid, center, and subtree core of a tree

Much research has been devoted to define the "middle part" of a tree. The first such result is due to Jordan [16]. In a tree T, the *branch weight* of a vertex v, $bw(v)$, is the maximum number of edges over all subtrees of T which contain v as a leaf. By definition, the *centroid* $C(T)$ of T is the set of vertices minimizing the branch weight. Jordan [16] showed that either $C(T) = \{c\}$, and $bw(c) \leqslant \frac{n-1}{2}$ (we always denote the order of the tree by n), or $C(T) = \{c_1, c_2\}$, where c_1 and c_2 are adjacent vertices with $bw(c_1) = bw(c_2) = \frac{n-1}{2}$, and in both cases all other vertices have branch weight strictly exceeding $\frac{n}{2}$. Zelinka [39] gave an alternative characterization of the centroid: $C(T)$ contains exactly those vertices u of $V(T)$, which *minimize* the distance function of vertices, i.e. $\sigma_T(u) = \sum_{v \in V(T)} d_T(u, v)$.

Jordan [16] also defined the *center* of a tree T, as the set of vertices minimizing the function *eccentricity* $ecc(u) = \max_{v \in V(T)} d_T(u, v)$, and showed that the center contains one vertex or two

adjacent vertices. (For a contemporary reference, see [22] 6.21 and 6.22.) Ádám [1] studied further concepts of centrality in trees.

We defined the "middle part" of a tree in a new way in [26]: the *subtree core* of T, $Core(T)$, is the set of vertices maximizing $F_T(v)$, number of subtrees of T that contain v. We showed that the subtree core of any tree T contains one or two vertices, and if the subtree core contains two vertices, then they must be adjacent. The ultimate reason for this phenomenon is that $F_T(.) : V(T) \to \mathbb{R}$ is strictly concave along any path of T, and hence F_T is maximized at a single vertex or two adjacent vertices on any path of T. This new concept differs from the concept of center and centroid.

2   The results

Barefoot, Entringer and Székely [2] determined extremal values of $\sigma_T(w)/\sigma_T(u)$, $\sigma_T(w)/\sigma_T(v)$, $\sigma(T)/\sigma_T(v)$, and $\sigma(T)/\sigma_T(w)$, where T is a tree on n vertices, v is in the centroid $C(T)$ of the tree T, and $u, w \in L(T)$ are leaves in T. (We use the notation $L(T)$ for the set of leaves of the tree T.)

In this paper we test how far the negative correlation between distances and subtrees go if we look for the extremal values of $F_T(w)/F_T(u)$, $F_T(w)/F_T(v)$, $F(T)/F_T(v)$, and $F(T)/F_T(w)$, where T is a tree on n vertices, v is in the subtree core of the tree T, and u, w are leaves in T—the complete analogue of [2], changing distances to the number of subtrees. Note that extremal behaviour of *fractions* is always more delicate than that of the numerator and denominator, therefore it is a natural step to see how far duality between Wiener index and the number of subtrees extend when we study extreme values of the ratios above. Comparison of extremal trees is particularly interesting. The Table summarizes our results.

From the two columns of "extremal tree", it is easy to see the correspondence between trees maximizing (minimizing) the distance function and the ones minimizing (maximizing) the subtree function. In rows 1, 3 and 6 of the table, the extremal trees are comets. The different path length is explained by the distance of a vertex always being polynomial in n, while the number of subtrees containing a vertex being typically exponentially large. The star is extremal in row 5. In row 2, the extremal tree is a path with an added vertex—but added at a different place. Rows 4 and 7 show different extremal trees for the two problems—still, the dumbbell contains a long path, while the other extremal tree is a path, and the spider has a central vertex like a star.

Theorems 1 and 2 seem to have the same r-comet as extremal tree, as we have to round almost the same number to obtain r. However, in neither of the two results we know when to round up or down, so the extremal r's may differ. In Theorem 7, the quantity to be rounded to obtain r for the comet, is bigger by one than in the cited two theorems.

Distance problem	Bound	Extremal tree	Subtree problem	Bound	Extremal tree
$\dfrac{\sigma_T(w)}{\sigma_T(u)} : u, w \in L(T)$ maximize	$\lesssim \dfrac{\sqrt{n}}{2}$	r-comet, $r \approx 2\sqrt{n}$ [2] Thm. 2	$\dfrac{F_T(w)}{F_T(u)} : u, w \in L(T)$ maximize	$\gtrsim \dfrac{n}{2}$	r-comet, $r \approx n - 2\log_2 n$ Thm. 1
$\dfrac{\sigma_T(w)}{\sigma_T(v)} : w \in L(T), v \in C(T)$ minimize	$\gtrsim 1$	path with a leaf added at middlepoint [2] Thm. 3	$\dfrac{F_T(w)}{F_T(v)} : w \in L(T), v \in Core(T)$ maximize	$\lesssim 2$	path with a leaf added at position $\approx 2n/3$ Thm. 3
$\dfrac{\sigma_T(w)}{\sigma_T(v)} : w \in L(T), v \in C(T)$ maximize	$\lesssim \sqrt{\dfrac{n}{2}}$	r-comet, $r \approx \sqrt{2n}$ [2] Thm. 3	$\dfrac{F_T(w)}{F_T(v)} : w \in L(T), v \in Core(T)$ minimize	$\gtrsim n$	r-comet, $r \approx n - 2\log_2 n$ Thm. 2
$\dfrac{\sigma_T}{\sigma_T(v)} : v \in C(T)$ minimize	$\gtrsim \dfrac{n}{2}$	r-dumbbell, $r \approx 2\sqrt{n}$ [2] Thm. 4	$\dfrac{F_T}{F_T(v)} : v \in Core(T)$ maximize	$\lesssim 1 + \dfrac{\lfloor \frac{n}{2}\rfloor}{\lfloor \frac{n}{2}\rfloor+1}$	path Thm. 5, proof in [29]
$\dfrac{\sigma_T}{\sigma_T(v)} : v \in C(T)$ maximize	$\lesssim n - 1$	star [2] Thm. 4	$\dfrac{F_T}{F_T(v)} : v \in Core(T)$ minimize	$\gtrsim 1 + \dfrac{n-1}{2^{n-1}}$	star Thm. 4, proof in [29]
$\dfrac{\sigma_T}{\sigma_T(w)} : w \in L(T)$ minimize	$\gtrsim \sqrt{2n}$	r-comet, $r \approx \sqrt{n}$ [2] Thm. 5	$\dfrac{F_T}{F_T(w)} : w \in L(T)$ maximize	$\lesssim n$	r-comet, $r \approx n - 2\log_2 n$ Thm. 7, proof in [29]
$\dfrac{\sigma_T}{\sigma_T(w)} : w \in L(T)$ maximize	$\lesssim n$	spider [2] Thm. 5	$\dfrac{F_T}{F_T(w)} : w \in L(T)$ minimize	$\gtrsim \dfrac{2^{n-1}+n-1}{1+2^{n-2}}$	star Thm. 6, proof in [29]

Table 1: Comparison of results.

An r-*comet* is an n-vertex tree made of an r-path with $n - r$ leaves joined to its left endpoint.

An r-*dumbbell* is an r-path with equal number of leaves joined to its endpoints to make an n-vertex tree.

A *spider* is an n-vertex tree, which has a special vertex v, has a single leaf joined to v, and has the remaining $n - 2$ vertices distributed on $\lceil \sqrt{3n}\rceil$ paths, whose length can differ by at most 1, and all start from v.

The notation $r \approx g(n)$ means that the extremal tree occurs for an r value, such that $r - g(n)$ remains bounded as $n \to \infty$. ([2] and our results are more specific about the value of r, but it is inconvenient to show in a succinct table.)

In the *Bound* columns $\underset{\sim}{\leqslant} f(n)$ means that an inequality is valid for a quantity that is asymptotically equal to $f(n)$ as $n \to \infty$.

Some of the results for distances require that n is sufficiently large.

The problem in the first row is not given a minimization version, as upper and lower bounds are reciprocals of each other.

Theorem 1 *Among trees T of order n, the maximum value of $\max_{w,u \in L(T)} \left\{ \frac{F_T(w)}{F_T(u)} \right\}$ is achieved by a tree T formed from identifying the center of the star $K_{1,x}$ with one end vertex of the path P_{n-x}, and only by such trees.*

For trees of this description the value $\frac{F_T(w)}{F_T(u)}$ can be explicitly formulated as a function of x:

$$f(x) := \frac{1 + (n-x)2^{x-1}}{n - x - 1 + 2^x}.$$

Trying to maximize $f(x)$ for real x's, from $f'(x) = 0$ we observe

$$(n-x)(n-x-1)2^{x-1}\ln 2 + 1 + 2^{x-1} = 2^{2x-1} + 2^x \ln 2. \tag{2.1}$$

Although the asymptotic calculations using iteration are routine, we show the details in this instance and skip the details later. Using the substitution $x \leftarrow n - y$, (2.1) can be rewritten as

$$2^y \left(y(y-1)\ln 2 + 1 + 2^{1+y-n} - 2\ln 2 \right) = 2^n.$$

As the left-hand side is an increasing function of y, it has a unique $y_0(n)$ solution for every positive integer n. Taking logarithm of both sides,

$$y_0(n) + \log_2 \left(y_0(n)(y_0(n) - 1)\ln 2 + 1 + 2^{1+y_0(n)-n} - 2\ln 2 \right) = n.$$

Substituting $y_0(n) \leftarrow n - \log_2 \left(y_0(n)(y_0(n)-1)\ln 2 + 1 + 2^{1+y_0(n)-n} - 2\ln 2 \right)$ in the previous equation for the occurrences of y in the logarithmic term, we obtain

$$y_0(n) + 2\log_2 n + \frac{\ln\ln 2}{\ln 2} + o(1) = n$$

and from here the $x_0(n)$ solution to (2.1) is

$$x_0(n) = 2\log_2 n + \frac{\ln\ln 2}{\ln 2} + o(1).$$

Hence, in the Table, $r = \lfloor y_0(n) \rfloor$ or $r = \lceil y_0(n) \rceil$. It is easy to see that $f(\lfloor x_0(n) \rfloor) \sim f(\lceil x_0(n) \rceil) \sim \frac{n}{2}$.

The following two theorems find extreme values of the *reciprocals* of the quantities in the second and third lines of the table:

Theorem 2 *Among trees T of order n, the maximum value of $\max_{v \in Core(T), u \in L(T)} \left\{ \frac{F_T(v)}{F_T(u)} \right\}$ is achieved by a tree T formed from identifying the center v of the star $K_{1,x-1}$ with one end vertex of the path P_{n-x+1} such that $v \in Core(T)$, and only by such trees.*

For trees of this description the value $\frac{F_T(v)}{F_T(u)}$ can be explicitly formulated as a function of x:

$$f(x) := \frac{(n-x+1)2^{x-1}}{n-x+2^{x-1}}.$$

Trying to maximize $f(x)$ for real x's, observe

$$f'(x) = \frac{(n-x+2^{x-1})(-2^{x-1}+(n-x+1)2^{x-1}\ln 2)-(n-x+1)2^{x-1}(-1+2^{x-1}\ln 2)}{(n-x+2^{x-1})^2},$$

and that the solutions of $f'(x) = 0$ are exactly the solutions of

$$(\ln 2)(n-x)(n-x+1) = 2^{x-1} - 1. \tag{2.2}$$

It is easy to see that there is only one solution $x_0(n)$ to (2.2), and standard asymptotic calculations show that

$$n - x_0(n) + 1 = n - 2\log_2 n - \frac{\ln\ln 2}{\ln 2} + o(1).$$

Hence, in the Table, $r = n+1-\lfloor x_0(n)\rfloor$ or $r = n+1-\lceil x_0(n)\rceil$. It is easy to see that $f(\lfloor x_0(n)\rfloor) \sim f(\lceil x_0(n)\rceil) \sim n$.

Theorem 3 *Among trees of order $n \geqslant 5$, $\min_{v\in Core(T),u\in L(T)}\left\{\frac{F_T(v)}{F_T(u)}\right\}$ is achieved if and only if T is formed from attaching one pendant vertex u to $v = v_x$ of a path $v_0v_1v_2\ldots v_{n-2}$. Here*

$$x = \left\lfloor \frac{2}{3}(n-1) \right\rfloor.$$

The minimum value of $\frac{F_T(v)}{F_T(u)}$ is

$$\frac{2(n-x-1)(x+1)}{1+(n-x-1)(x+1)},$$

which is asymptotically equal to 2.

The following 4 theorems are proved in a sibling paper [29], to keep the length of this paper under control:

Theorem 4 *Among trees of order n with $v \in Core(T)$,*

$$\frac{F(T)}{F_T(v)} \geqslant 1 + \frac{n-1}{2^{n-1}}$$

with equality if and only if T is a star centered at v.

Theorem 5 *Among trees T of order n, where $v \in Core(T)$, the maximum value of $\frac{F(T)}{F_T(v)}$ is obtained by a path, with*

$$\frac{F(T)}{F_T(v)} = 1 + \frac{\lfloor\frac{n}{2}\rfloor}{\lfloor\frac{n}{2}\rfloor+1}.$$

Theorem 6 *Among trees T of order n, where $u \in L(T)$,*

$$\frac{F(T)}{F_T(u)} \geqslant \frac{2^{n-1} + n - 1}{1 + 2^{n-2}}$$

with equality if and only if T is a star.

Theorem 7 *Among trees T of order n, $\max_{u \in L(T)} \left\{ \frac{F(T)}{F_T(u)} \right\}$ is achieved when T is an $(x+1)$-comet and u is a leaf that has a degree 2 neighbor, and only by such trees.*

For trees of this description the value $\frac{F(T)}{F_T(u)}$ can be explicitly formulated as a function of x:

$$f(x) := \frac{2^{n-x-1}(x+1) + \frac{1}{2}x(x-1) + (n-1)}{2^{n-x-1} + x}.$$

The function $f(x)$ is maximized when $x = n - 2\log_2 n - \frac{\ln \ln 2}{\ln 2} + o(1)$. This complicated analysis is detailed in [29].

3 Open problems

After having understood many extremal properties of the number of subtrees, next one would like to understand how they relate to global properties.

Problem 8 *Does the multiset $\{F_T(v) : v \in V(T)\}$ determine the tree T?*

It makes sense to look for a more refined information than $F_T(v)$: let $F_T(v, x)$ denote the the number of subtrees of T *of order x* that contain vertex v. Also, instead of an underlying tree T, we may consider a *graph G*. There are results on computing the number of subtrees of trees with dynamic programming [38].

Problem 9 *Are there determinant formulae to compute F_T, $F_T(v)$, $F_T(v, x)$? Or even more general, F_G, $F_G(v)$, $F_G(v, x)$? Is there a way to* approximate *these numbers in a very large graph efficiently?*

Problem 10 *Are the distributions $F_G(v, x) : x \in \mathbb{N}$ ($v \in V(G)$) connected to connectivity, expansion, or spectral properties of the underlying graph G?*

The following problem is somewhat vague:

Problem 11 *Given two graphs, G_1 and G_2, not necessarily of the same order or size, can we conclude from the "similarity" of the distributions $F_{G_i}(v, x) : x \in \mathbb{N}$ ($v \in V(G_i)$) the "similarity" of G_1 and G_2?*

4 Proof to Theorem 2

Given a tree T of order n, we want to find the maximum of

$$R(T) := \max_{v \in Core(T), u \in L(T)} \left\{ \frac{F_T(v)}{F_T(u)} \right\}.$$

First note $u = v$ is only possible in $\leqslant$ 2-vertex trees—by the strict concavity of $F_T(.)$ along paths, mentioned in Section 1.4.

Lemma 12 *If $v \in Core(T)$ and $u \in L(T)$, and $R(T)$ is maximized by $\frac{F_T(v)}{F_T(u)}$ among trees of order n, then the internal vertices of the path connecting v and u have degree 2. (See Fig. 2, we denote the uv path by $P_{u,v}$).*
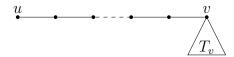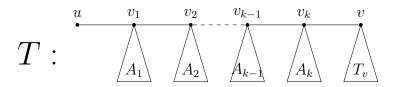

Figure 2: T_v and $P_{u,v}$.

Proof. List the vertices of the $P_{u,v}$ path as $P_{u,v} := uv_1v_2 \ldots v_kv$, and assume $k \geqslant 1$, otherwise we have nothing to prove. Let $A_0 = \{u\}, A_1, A_2, \ldots, A_k, A_{k+1} = T_v$ denote the connected components (trees) in $T - E(P_{u,v})$, with $v_i \in A_i$ for $1 \leqslant i \leqslant k$ and $v \in T_v$.

Suppose (for contradiction) that A_i is not a single vertex for some $1 \leqslant i \leqslant k$, and consider a tree T' obtained by moving all A_i's from v_i to v (see Fig. 3).
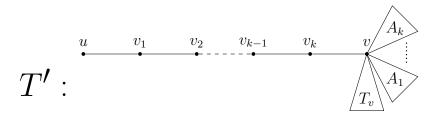

Figure 3: Making T' from T.

Simple enumerations give the following Horner's scheme like formulae:

$$F_T(u) = \left(\ldots \left(\left((F_{T_v}(v) + 1)F_{A_k}(v_k) + 1 \right) F_{A_{k-1}}(v_{k-1}) + 1 \right) \ldots + 1 \right) F_{A_1}(v_1) + 1;$$

$$F_{T'}(u) = k + 1 + F_{T_v}(v) \prod_{i=1}^{k} F_{A_i}(v_i) < F_T(u); \tag{4.3}$$

and

$$F_T(v) = F_{T_v}(v)\left(1 + F_{A_k}(v_k)\left(1 + F_{A_{k-1}}(v_{k-1})\left(1 + \dots (1 + 2F_{A_1}(v_1))\dots\right)\right)\right);$$

$$F_{T'}(v) = (k+2)F_{T_v}(v) \prod_{i=1}^{k} F_{A_i}(v_i) > F_T(v), \tag{4.4}$$

unless $F_{A_i}(v_i) = 1$ for all $1 \leqslant i \leqslant k$, which is equivalent to v_i having degree 2 in T.

Intuitively, (4.3) and (4.4) can be understood as moving any "branch" away from u and closer to v will decrease the number of subtrees containing u and increase that of v. Hence by (4.3) and (4.4), we have

$$R(T') \geqslant \frac{F_{T'}(v)}{F_{T'}(u)} > \frac{F_T(v)}{F_T(u)},$$

a contradiction. $\qquad\square$

Assuming $|V(T_v)| = x$, knowing that the tree has order n, we obtain $|V(P_{u,v})| = n - x + 1$. We can get explicitly

$$F_T(u) = n - x + F_{T_v}(v)$$

and

$$F_T(v) = (n - x + 1) \cdot F_{T_v}(v).$$

Therefore

$$\frac{F_T(u)}{F_T(v)} = \frac{1}{n - x + 1} + \frac{n - x}{n - x + 1} \cdot \frac{1}{F_{T_v}(v)}. \tag{4.5}$$

Hence, for given n and x, $\frac{F_T(v)}{F_T(u)}$ is maximized if and only if (4.5) is minimized, which is the case if and only if $F_{T_v}(v)$ is maximized. The following simple observation will be used.

Lemma 13 *Given* $x = |V(T_v)|$,
$$F_{T_v}(v) \leqslant 2^{x-1}$$
with equality if and only if T_v is a star centered at v.

Proof. This is trivial by noting that any subtree of T_v containing v induces a unique subset of $V(T_v) \setminus \{v\}$. Hence the total number of such subtrees is at most the number of such subsets. $\qquad\square$

It is easy to see that, by changing T_v to be a star centered at v to maximize $F_{T_v}(v)$, v stays in the subtree core of T. Thus we proved Theorem 2.

5 Proof to Theorem 1

Essentially the approach of the previous Section 4 will handle this question, so we skip some of the details. Given a tree T of order n, we want to find the maximum of

$$R(T) = \max_{w,u \in L(T)} \left\{ \frac{F_T(w)}{F_T(u)} \right\}.$$

First note $uw \in E(T)$ is only possible in 2-vertex trees. Choosing $w = u$ provides in any tree $R(T) = 1$, as is the case of $d_T(w, u) = 2$. Below we construct trees and leaves with higher ratios. For an optimal tree T and leaves u, w, let the path connecting w and u be $P_{w,u} = uv_1 v_2 \ldots v_k w$, and by the arguments above we can assume $k > 1$. Let A_i be defined as in Section 4 for $1 \leqslant i \leqslant k$. Like in the proof of Lemma 12, by moving all A_i's for $1 \leqslant i \leqslant k-1$ to v_k we obtain T' with

$$\frac{F_{T'}(w)}{F_{T'}(u)} > \frac{F_T(w)}{F_T(u)},$$

unless $F_{A_i}(v_i) = 1$ for all $1 \leqslant i \leqslant k$, which is equivalent to v_i having degree 2 in T. See Fig. 4 for illustration. Indeed,

$$F_T(u) = \left(\ldots \left(\left(2F_{A_k}(v_k) + 1 \right) F_{A_{k-1}}(v_{k-1}) + 1 \right) \ldots + 1 \right) F_{A_1}(v_1) + 1;$$

$$F_{T'}(u) = k + 2 \prod_{i=1}^{k} F_{A_i}(v_k) < F_T(u); \tag{5.6}$$

and

$$F_T(w) = \left(\ldots \left(\left(2F_{A_1}(v_1) + 1 \right) F_{A_2}(v_2) + 1 \right) \ldots + 1 \right) F_{A_k}(v_k) + 1;$$

$$F_{T'}(w) = 1 + (k+1) \prod_{i=1}^{k} F_{A_i}(v_k) > F_T(w), \tag{5.7}$$

unless $F_{A_i}(v_i) = 1$ for all $1 \leqslant i \leqslant k$, which is equivalent to v_i having degree 2 in T.

Thus we have the following analogue of Lemma 12.
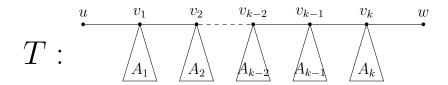
Lemma 14 $R(T)$ *is maximized by a tree T and $w, u \in L(T)$ only if T is formed by a tree T_v rooted at v with a path from u to v and an edge wv (Fig. 5).*

Assume $|V(T_v)| = x$, then $|V(P_{u,v})| = n - x$ as the tree T has order n. It is easy to see that

$$F_T(u) = n - x - 1 + 2F_{T_v}(v) = n - x - 1 + 2\left(F_{T_v}(v) + \frac{1}{n-x} \right) - \frac{2}{n-x}$$

and

$$F_T(w) = 1 + (n - x) \cdot F_{T_v}(v) = (n - x)\left(\frac{1}{n-x} + F_{T_v}(v) \right).$$
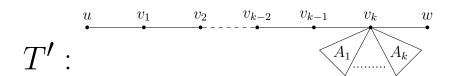
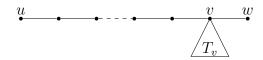Figure 4: T' and T with respect to w and u.

Figure 5: T_v and $P_{u,v}$ with respect to w and u.

Therefore

$$\frac{F_T(u)}{F_T(w)} = \frac{2}{n-x} + \frac{n-x-1-\frac{2}{n-x}}{n-x} \cdot \frac{1}{\frac{1}{n-x} + F_{T_v}(v)}. \tag{5.8}$$

Hence, with given n and x, $\frac{F_T(w)}{F_T(u)}$ is maximized if and only if (5.8) is minimized, which is the case when $F_{T_v}(v)$ is maximized. Using Lemma 13 again, we obtain Theorem 1.

6 Proof to Theorem 3

Let

$$R(T) = \min_{v \in Core(T), u \in L(T)} \left\{ \frac{F_T(v)}{F_T(u)} \right\}.$$

We need the following important Lemma, which will immediately follow from Lemmas 16 and 17:

Lemma 15 *For T with given $|V(T)| = n \geqslant 2$, $u \in L(T)$ and $v \in Core(T)$ that achieves the minimum value of $R(T)$, u and v must be adjacent (Fig. 6).*

Assume that T is the tree that minimizes $R(T)$. As $n \geqslant 2$, there is a (unique) neighbor x of u in T. For contradiction, assume $x \notin Core(T)$. Consider the tree T' obtained from T by deleting the $\{xu\}$ edge and adding the $\{vu\}$ edge. We will establish that $v \in Core(T')$ and $\frac{F_{T'}(v)}{F_{T'}(u)} < \frac{F_T(v)}{F_T(u)}$ in the following two Lemmas, providing the contradiction.

Figure 6: $uv \in E(T)$ to minimize $R(T)$.

Lemma 16 *With $u \in L(T)$ and $v \in Core(T)$ and T' defined as above, we must have*

$$v \in Core(T').$$

Lemma 17 *If the unique neighbor x of u is not a subtree core vertex of T, then*

$$\frac{F_{T'}(v)}{F_{T'}(u)} < \frac{F_T(v)}{F_T(u)}.$$

Proof. [Proof of Lemma 16] Assume $n \geqslant 3$, otherwise we have nothing to prove. To show that $v \in Core(T')$, we only need to show (by the concavity of $F_T(.)$ along any path) $F_{T'}(v) \geqslant F_{T'}(y)$ for any neighbor y of v. Evidently $y \neq v$ and $F_{T'}(v) \geqslant F_{T'}(u)$, so we can assume $y \notin \{u, v\}$. Using a generalized notation $F_H(a, b)$ for the number of subtrees of a tree H containing *both* vertices a, b, we observe

$$F_{T'}(v) = F_T(v) - F_T(u, v) + F_{T \setminus \{u\}}(v) \tag{6.9}$$

and

$$F_{T'}(y) = F_T(y) - F_T(u, y) + F_{T \setminus \{u\}}(v, y). \tag{6.10}$$

If $y \in V(P_T(u, v))$, then we have

$$F_{T \setminus \{u\}}(v) > F_{T \setminus \{u\}}(v, y) \text{ and } F_T(u, y) > F_T(u, v),$$

immediately implying that $(6.9) > (6.10)$ since $F_T(v) \geqslant F_T(y)$ (by the fact that $v \in Core(T)$). Otherwise, $v \in V(P_T(u, y))$ and $y \neq x$. Let S be the component containing v in $(T \setminus \{u\}) \setminus \{vy\}$, then

$$F_T(u, v) - F_T(u, y) = F_{T \setminus \{u\}}(x, v) - F_{T \setminus \{u\}}(x, y) = F_S(x, v) < F_S(v) = F_{T \setminus \{u\}}(v) - F_{T \setminus \{u\}}(v, y),$$

again implying $(6.9) > (6.10)$. Hence the proof of Lemma 16 is completed. $\qquad \square$

Proof. [Proof of Lemma 17] Observe

$$F_{T'}(v) = 2F_{T \setminus \{u\}}(v) < 2F_T(v). \tag{6.11}$$

Also observe

$$F_T(u) = 1 + \frac{F_T(x)}{2} \leqslant 1 + \frac{F_T(v) - 1}{2}. \tag{6.12}$$

Hence by (6.11) and (6.12),

$$F_{T'}(v)\left(F_T(u) - \frac{F_T(v)}{2}\right) \leqslant \frac{F_{T'}(v)}{2} < F_T(v). \tag{6.13}$$

Note that (6.13) implies

$$F_{T'}(v)F_T(u) < F_T(v)\left(1 + \frac{F_{T'}(v)}{2}\right) = F_T(v)F_{T'}(u), \tag{6.14}$$

finishing the proof of Lemma 17. □

Based on Lemma 15, we represent T that minimizes $R(T)$ as $T_v \cup \{uv\}$ (Fig. 6). We have

$$\frac{F_T(v)}{F_T(u)} = \frac{2F_{T_v}(v)}{1 + F_{T_v}(v)} = 2 - \frac{2}{1 + F_{T_v}(v)},$$

and hence $R(T)$ is minimized when $F_{T_v}(v)$ is minimized.

Now we focus on T_v and let the neighbors of v in T_v be $v_1, \ldots, v_k$. We claim $k \geqslant 2$. Indeed, if there is only one neighbor $v_1 = w$, which separates a subtree T_w from v, like on Fig. 6 v separates T_v from u. It is easy to see that $F_T(v) = 2 + 2F_{T_w}(w)$ and $F_T(w) = 3F_{T_w}(w)$, showing $v \notin Core(T)$ when $F_{T_w}(w) > 2$, a contradiction. This happens for all $n \geqslant 5$.

Hence $k \geqslant 2$, and let the connected components in $T_v \setminus \{v\}$ be $T_1, \ldots, T_k$ (Fig. 7), such that v_i is a vertex of T_i. Assume without loss of generality that

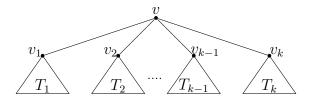$$F_{T_1}(v_1) \geqslant F_{T_i}(v_i)$$

for any i.

Figure 7: T_v and its subtrees.

First we formulate the condition for v to be in the subtree core of T:

Lemma 18 *With T and v represented above, $v \in Core(T)$ if and only if*

$$2 \prod_{1 \leqslant j \leqslant k, j \neq i} (1 + F_{T_j}(v_j)) \geqslant F_{T_i}(v_i) \tag{6.15}$$

for any $1 \leqslant i \leqslant k$.

Proof. By the strict concavity of $F_T(.)$ along any path, we have $v \in Core(T)$ if and only if $F_T(v) \geqslant F_T(y)$ for any neighbor of v. This inequality evidently holds when $y = u$.

For $y = v_i$ for some $1 \leqslant i \leqslant k$, $F_T(v) \geqslant F_T(v_i)$ if and only if the number of subtrees containing v but not v_i is at least as large as the number of subtrees containing v_i but not v. Notice that the former is

$$2 \prod_{1 \leqslant j \leqslant k, j \neq i} (1 + F_{T_j}(v_j)),$$

and the latter is

$$F_{T_i}(v_i),$$

completing the proof. $\square$

Next we claim that T_1, the heaviest branch in T_v regarding $F_{T_i}(v_i)$, must be a path.

Lemma 19 *To minimize $F_{T_v}(v)$, where $v \in Core(T)$ in Fig. 6, T_1 must be a path in Fig. 7.*

Proof. Suppose (for contradiction) that T_1 is not a path, and let

$$P_{v_1,w} = w_0(v_1)w_1 w_2 \ldots w_{l-1} w_l(w)$$

be a longest path in T_1 from $v_1 = w_0$ to some leaf $w = w_l$. (In other words, considering T_1 as a rooted tree at v_1, it is of height l and w is one of the vertices of the greatest height.) Let j ($0 \leqslant j \leqslant l - 1$) be the smallest integer such that w_j has a neighbor z *not* on $P_{v_1,w}$. Let $A_j, A_{j+1}, \ldots, A_{l-1}$ denote the connected components of $T_1 \backslash E(P_{v_1,w})$ containing $w_j, w_{j+1}, \ldots, w_{l-1}$ respectively. Define the tree $A_{\geqslant j+1} = A_{j+1} \cup A_{j+2} \cup \ldots \cup A_{l-1} \cup \{w\}$ with the edge set $w_{j+1}w_{j+2}, \ldots, w_{l-1}w_l$ added.

We will consider T_1' obtained from T_1 by moving A_j from w_j to w_{j+1} (Fig. 8), and the corresponding T_v', T' in the original problem.
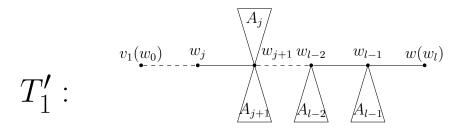

Figure 8: Generating T_1' from T_1.

Then we have

$$F_{T_1}(v_1) = j + F_{A_j}(w_j)\Big(1 + F_{A_{\geqslant j+1}}(w_{j+1})\Big) \tag{6.16}$$

and
$$F_{T_1'}(v_1) = j + 1 + F_{A_j}(w_j)F_{A_{\geqslant j+1}}(w_{j+1}). \tag{6.17}$$

From (6.16) and (6.17), it is easy to see that

$$F_{T_1'}(v_1) \geqslant \frac{1}{2}F_{T_1}(v_1).$$

Hence, for any $i \neq 1$, we have

$$F_{T_i}(v_i) \leqslant F_{T_1}(v_1) \leqslant 2F_{T_1'}(v_1),$$

maintaining the condition (6.15) in T'. Also from (6.16) and (6.17), we have

$$F_{T_1'}(v_1) < F_{T_1}(v_1) \tag{6.18}$$

since $F_{A_j}(w_j) > 1$. Hence the condition (6.15) is also maintained for $i = 1$ in T'.

Thus $v \in Core(T')$. Furthermore, (6.18) implies that $F_{T_v'}(v) < F_{T_v}(v)$, a contradiction. $\qquad\square$

We are now ready to minimize $F_{T_v}(v)$ and to characterize the structure with minimal $F_{T_v}(v)$. Assume $x = |V(T_1)|$ (and consequently $F_{T_1}(v_1) = x$) and $|V(T)| = n \geqslant 5$. Consider now $S = T_v \backslash T_1$.

Lemma 20 *The inequality $\prod_{i=2}^{k}(1 + F_{T_i}(v_i)) \geqslant n - 1 - x$ holds. Equality holds if and only if S is a path and v is an endpoint of the path S.*

Proof. It is easy to see that

$$\prod_{i=2}^{k}(1 + F_{T_i}(v_i)) \geqslant 1 + \sum_{i=2}^{k} F_{T_i}(v_i) \geqslant |V(S)| = n - 1 - x. \tag{6.19}$$

All the terms of $1 + \sum_{i=2}^{k} F_{T_i}(v_i)$ are present among the terms of expansion of the left-hand side. If equality holds, then the 2^{k-1} terms on the left side are these k terms. This happens if and only if $k = 2$. In case of $k = 2$ and equality in (6.19), the same argument applies to $S \setminus \{v\}$, etc., completing the characterization by induction. $\qquad\square$

We will need the following fact that is easy to verify by considering cases according to n modulo 3:

Lemma 21 *For a positive integer n, $n - 1 = \lceil \frac{x}{2} \rceil + x$ if and only if $x = \lfloor \frac{2}{3}(n - 1) \rfloor$.*

Now we consider cases.
Case (i): $x \geqslant \lfloor \frac{2}{3}(n - 1) \rfloor$.
As (6.15) implies

$$2\prod_{i=2}^{k}(1 + F_{T_i}(v_i)) \geqslant x,$$

we have

$$F_{T_v}(v) = \prod_{i=1}^{k}(1 + F_{T_i}(v_i)) \geqslant \left\lceil \frac{x}{2} \right\rceil (x+1) \geqslant \left\lceil \frac{\lfloor \frac{2}{3}(n-1) \rfloor}{2} \right\rceil (1 + \lfloor \frac{2}{3}(n-1) \rfloor). \qquad (6.20)$$

It is easy to see that if equation holds in (6.20), then $\left\lceil \frac{x}{2} \right\rceil = \prod_{i=2}^{k}(1 + F_{T_i}(v_i))$ and $x = \lfloor \frac{2}{3}(n-1) \rfloor$. By Lemma 20, $n - 1 - x \leqslant \prod_{i=2}^{k}(1 + F_{T_i}(v_i))$, and hence

$$n - 1 \leqslant x + \left\lceil \frac{x}{2} \right\rceil. \qquad (6.21)$$

By Lemma 21, formula (6.21) is solved by equality if and only if $x = \lfloor \frac{2}{3}(n-1) \rfloor$, and with strict inequality if and only if $x > \lfloor \frac{2}{3}(n-1) \rfloor$. Therefore, if equality holds in (6.20), then $x = \lfloor \frac{2}{3}(n-1) \rfloor$, and by Lemma 20, $k = 2$ and T_2 is a path. The description above is a construction that always realizes the lower bound in the right-hand side of (6.20).

Our goal is to conclude that the minimum of $F_{T_v}(v)$ is what we realized as the minumum for Case (i) in (6.20).

Case (ii): $\frac{1}{3}(n-1) \leqslant x \leqslant \lfloor \frac{2}{3}(n-1) \rfloor - 1$.

By Lemma 20, $F_S(v) \geqslant |V(S)| = n - 1 - x$, and hence

$$F_{T_v}(v) \geqslant (n - 1 - x)(x+1) = g(x). \qquad (6.22)$$

The lower bound for $F_{T_v}(v)$ in this case is a quadratic function of x, which is minimized in one of the endpoints of the interval. Easy calculations show

$$g\left(\frac{1}{3}(n-1)\right) > \left\lceil \frac{\lfloor \frac{2}{3}(n-1) \rfloor}{2} \right\rceil (1 + \lfloor \frac{2}{3}(n-1) \rfloor) \text{ and } g\left(\lfloor \frac{2}{3}(n-1) \rfloor - 1\right) > \left\lceil \frac{\lfloor \frac{2}{3}(n-1) \rfloor}{2} \right\rceil (1 + \lfloor \frac{2}{3}(n-1) \rfloor),$$

showing that the minimum in Case (i) cannot even be matched in Case (ii).

Case (iii): $x = 1$.

This means $T_1 = v_1$. By the choice of T_1, $T_i = v_i$ for $i = 1, 2, \ldots, k$, and hence T_v is a star. By Lemma 13, $F_{T_v}(v) = 2^{n-2}$, exceeding the right-hand side of (6.20) for $n \geqslant 5$.

Case (iv): $2 \leqslant x < \frac{1}{3}(n-1)$.

As $2 \leqslant x$, there is a leaf w in T_1 that is not v_1. We consider subcases.

Subcase (iv)(a):

For $i = 2, \ldots, k$, $T_i = v_i$. Then easy calculation gives $F_{T_v}(v) = (x+1)2^{n-2-x}$. This number exceeds the right-hand side of (6.20) for $n \geqslant 5$.

Subcase (iv)(b):

Without loss of generality, assume that there is a leaf $z \neq v_2$ in T_2 and y its unique neighbor.

Obtain a tree T_v' by removing the $\{yz\}$ edge from T_v and adding the $\{wz\}$ edge. Then (6.15) is maintained for T_v', so it is a legitimate candidate to minimize $F_{T_v}(v)$, as in Case (iv)

$$x + 1 < n - 1 - (x+1) \leqslant \prod_{i=2}^{k}(1 + F_{T_i}(v_i)).$$

Furthermore,

$$F_{T'_v}(v) \leqslant (x+2)F_{T_2}(v_2)\prod_{i=3}^{k}(1+F_{T_i}(v_i))$$

$$< (x+1)(1+F_{T_2}(v_2))\prod_{i=3}^{k}(1+F_{T_i}(v_i))$$

$$= F_{T_v}(v)$$

since $x \geqslant F_{T_2}(v_2)$. Hence this case will not obtain the minimum of $F_{T_v}(v)$.

References

[1] A. Ádám. The centrality of vertices in trees. *Studia Sci. Math. Hung.* 9:285–303, 1974.

[2] C.A. Barefoot, R.C. Entringer, and L.A. Székely. Extremal values for ratios of distances in trees. *Discrete Appl. Math.* 80:37–56, 1997.

[3] E. Cela, N.S. Schmuck, S. Wimer, and G.J. Woeginger. The Wiener maximum quadratic assignment problem. *Discrete Optim.* 8:411–416, 2011.

[4] M.H. Chehreghani, M.H. Chehreghani, C. Lucas, M. Rahgozar, and E. Ghadimi. Efficient rule based structural algorithms for classification of tree structured data. *Intelligent Data Analysis,* 13(1):165–188, 2009.

[5] É. Czabarka, L.A. Székely, and S. Wagner. The inverse problem for certain tree parameters. *Discrete Applied Math.* 157(15):3314–3319, 2009.

[6] M.J. Dinneen and M. Khosravani. Hardness of Approximation and Integer Programming Frameworks for Searching for Caterpillar Trees. CDMTCS Research Reports CDMTCS-394 (2010), https://researchspace.auckland.ac.nz/handle/2292/10547

[7] A.A. Dobrynin, R.C. Entringer, and I. Gutman. Wiener index of trees: Theory and applications. *Acta Appl. Math.* 66(3):211–249, 2001.

[8] R.C. Entringer, D.E. Jackson, and D.A. Snyder. Distance in graphs. *Czechoslovak Math. J.* 26(101):283–296, 1976.

[9] M. Fischermann, A. Hoffmann, D. Rautenbach, L. Székely, and L. Volkmann. Wiener index versus maximum degree in trees. *Discrete Appl. Math.* 122(1-3):127–137, 2002.

[10] M. Hamina and M. Peltola. Least central subtrees, center, and centroid of a tree. *Networks,* 57(4):328–332, 2011.

[11] M. Hamina and M. Peltola. Some structural properties of a least central subtree of a tree. *Algorithmic Operations Research,* 5(2), 2010.

[12] C. Heuberger and S.G. Wagner. The number of maximum matchings in a tree. *Discrete Math.* 311(21):2512–2542, 2011.

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

[14] C. Heuberger and H. Prodinger. On α-greedy expansions of numbers. *Adv. in Appl. Math.* 38(4):505–525, 2007.

[15] F. Jelen and E. Triesch. Superdominance order and distance of trees with bounded maximum degree. *Discrete Appl. Math.* 125(2–3):225–233, 2003.

[16] C. Jordan. Sur les assemblages de lignes. *J. Reine Angew. Math.* 70:185–190, 1869.

[17] M. Khosravani. Searching for Optimal Caterpillars in General and Bounded Treewidth Graphs. Ph.D. Thesis, University of Auckland, 2011.

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

[19] M. Lepović and I. Gutman. A collective property of trees and chemical trees. *J. Chem. Inf. Comput. Sci.*, 38:823–826, 1998.

[20] Shuchao Li and Shujing Wang. Further analysis on the total number of subtrees of trees. arXiv:1204.6152, 2012

[21] Shuchao Li and Shujing Wang. Enumerating the total number of subtrees of trees. arXiv:1206.2975, 2012

[22] L. Lovász. *Combinatorial Problems and Exercises*, 2nd ed., North–Holland Publishing Co., Amsterdam, 1993.

[23] A. Maunz, C. Helma, and S. Kramer. Efficient mining for structurally diverse subgraph patterns in large molecular databases. *Machine Learning,* 83(2):193–218, 2011.

[24] M. Schmidt and A. Schöbel. Location of speed-up subnetworks. Institut für Numerische und Angewandte Mathematik, Nr. 2009-13, Georg-August Universität Göttingen.

[25] 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.

[26] L.A. Székely and Hua Wang. On subtrees of trees. *Adv. Appl. Math.* 34:138–155, 2005.

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

[28] L.A. Székely and Hua Wang. Binary trees with the largest number of subtrees. *Discrete Appl. Math.* 155(3):374–385, 2006.

[29] L.A. Székely and Hua Wang. Extremal values of ratios: distance problems vs. subtree problems in trees II. To appear.

[30] L. A. Székely, Hua Wang, and Taoyang Wu. The sum of distances between the leaves of a tree and the 'semi-regular' property. *Discrete Math.* 311:1197–1203, 2011.

[31] E. Teufl and S.G. Wagner. Enumeration problems for classes of self-similar graphs *J. Combinatorial Theory, Series A*, 114(7):1254–1277, 2007.

[32] S. Wagner. A class of trees and its Wiener index. *Acta Appl. Math.*, 91(2):119–132, 2006.

[33] S. Wagner, Hua Wang, and Gang Yu. Molecular graphs and the inverse Wiener index problem. *Discrete Applied Mathematics*, 157(7):1544–1554, 2009.

[34] S. Wagner. Correlation of graph-theoretical indices. *SIAM J. Discrete Mathematics*, 21(1):33–46, 2007.

[35] Hua Wang. The extremal values of the Wiener index of a tree with given degree sequence. *Discrete Appl. Math.* 156(14):2647–2654, 2008. Corrigendum,*Discrete Appl. Math.* 157(18):3754, 2009.

[36] Hua Wang and Gang Yu. All but 49 numbers are Wiener indices of trees. *Acta Appl. Math.* 92(1):15–20, 2006.

[37] H. Wiener. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* 1(69):17–20, 1947.

[38] Weigen Yan and Yeong-Nan Yeh. Enumeration of subtrees of trees. *Theoretical Computer Science*, 369(1-3):256–268, 2006.

[39] B. Zelinka. Medians and peripherans of trees. *Arch. Math.* (Brno) 4:87–95, 1968.

[40] Xiao-Dong 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.*, 60:623–644, 2008.

[41] Xiao-Dong Zhang, Yong Liu, and Min-Xian Han. Maximum Wiener Index of Trees with Given Degree Sequence. *MATCH Commun.Math.Comput.Chem.*, 64:661–682, 2010.

[42] Xiu-Mei Zhang, Xiao-Dong Zhang, D. Gray, and Hua Wang. The number of subtrees of trees with given degree sequence. *J. Graph Theory*, (2012) DOI: 10.1002/jgt.21674