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

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

December 28, 2012

Abstract

We 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]. We introduced the concept of *subtree core* : the *subtree core* of a tree consists of one or two adjacent vertices of a tree that are contained in the largest number of subtrees. 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 Part I of this paper we tested how far the negative correlation between distances and subtrees go if we look for (and characterize) the extremal values of $F_T(w)/F_T(u)$, $F_T(w)/F_T(v)$. In this paper we characterize the extremal values of $F(T)/F_T(v)$, 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 w is a leaf in T—completing the analogy, changing distances to the number of subtrees.

AMS Subject Classification (2010): 05C05; 05C12; 05C35; 92E10
Keywords: Wiener index; tree; star tree; distances in trees; subtrees of trees; extremal problems; center; centroid; subtree core
*This author acknowledges financial support from the grant #FA9550-12-1-0405 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA) and from the grant 1000475 of the NSF DMS.
†This author acknowledges financial support from the grant 245307 from the Simons Foundation.

1 Introduction

For a tree T, let $F(T)$ denote the number of subtrees of T and $F_T(v)$ denote the number of subtrees of T containing the vertex $v \in V(T)$. We [5] introduced a new centrality concept for trees: the set of vertices that maximize $F_T(v)$ among all vertices of T. This is called the *subtree core* of T and will be denoted by $Core(T)$. The concept of subtree core is the natural subtree analogue of the well-known concepts of *center* [3] and *centroid* [3, 9] of a tree.

Theorem 1 [5] *The function $F_T(.)$ is strictly concave along any path of T, and hence the subtree core of any tree T contains one or two vertices. If the subtree core contains two vertices, then they must be adjacent.*

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 [2].) We term degree 1 vertices in a tree as *leaves*. We use the notation $L(T)$ for the set of leaves of the tree T.

Knudsen [4] 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's work motivated our earlier papers [5, 6]. Since the appearance of these papers, extensive work has been done in the study of the number of subtrees of a tree. In particular, the extremal values of this number are determined for many different categories of trees. An intriguing fact is that these extremal trees that maximize or minimize the number of subtrees coincide with the extremal trees that minimize or maximize the Wiener index of a tree (even though there is no functional relation between the number of subtrees and the Wiener index of a tree). Wagner [8] made an analysis of correlation the 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. For a summary of this progress see [7].

Barefoot, Entringer and Székely [1] 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. 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 the duality (negative correlation) between the Wiener index and the number of subtrees extend when we study extreme values of the ratios above.

In [7] we started investigating exactly this problem and we conclude the investigation in the present paper. [7] characterized the extremal values of $F_T(w)/F_T(u)$, $F_T(w)/F_T(v)$, 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 the first two problems of [1], changing distances to the number of subtrees. In this paper we characterize the extremal values of $F(T)/F_T(v)$, 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 w is a leaf in T—the complete analogue of the last two problems of [1], changing distances to the number of subtrees.

2 The results

First note the identity

$$\frac{F(T)}{F_T(v)} = 1 + \frac{\overline{F_T}(v)}{F_T(v)},$$

where $\overline{F_T}(v)$ denotes the *number of subtrees of T not containing v*. As it is a useful observation, we spell out the following immediate consequence of Theorem 1:

Corollary 2 *Both $\frac{F(T)}{F_T(v)}$ and $\frac{\overline{F_T}(v)}{F_T(v)}$ are strictly convex, when v runs along along any path of a tree T. Both ratios are maximized at a leaf vertex and minimized at subtree core vertices.*

Theorem 3 *For any tree T with $|V(T)| = n$ and any $v \in Core(T)$, we have*

(i) $F_T(v) \leq 2^{n-1}$ *with equality if and only if T is a star centered at v.*

(ii) $F(T) \leq 2^{n-1} + n - 1$ *with equality if and only if T is a star* [5].

(iii) $\frac{F(T)}{F_T(v)} \geq 1 + \frac{n-1}{2^{n-1}}$ *with equality if and only if T is a star centered at v.*

Proof. Simply note that

$$\overline{F_T}(v) \geq |V(T \setminus \{v\})| = n - 1$$

and

$$F_T(v) \leq \text{ the number of subsets of } V(T \setminus \{v\}) \ = 2^{n-1}$$

with both equalities if and only if T is a star. Hence (i) follows. For (iii), observe

$$\frac{F(T)}{F_T(v)} = 1 + \frac{\overline{F_T}(v)}{F_T(v)} \geq 1 + \frac{n-1}{2^{n-1}}.$$

□

Now take an arbitrary $u \in L(T)$ and let v be its unique neighbor. Consider $T_v = T \setminus \{u\}$, and observe

$$\frac{F(T)}{F_T(u)} = 1 + \frac{\overline{F_T}(u)}{F_T(u)} = 1 + \frac{F(T_v)}{1 + F_{T_v}(v)} = 1 + \frac{F_{T_v}(v) + \overline{F_{T_v}}(v)}{1 + F_{T_v}(v)} = 2 + \frac{\overline{F_{T_v}}(v) - 1}{1 + F_{T_v}(v)}.$$

As in the proof of Theorem 3, $\overline{F_{T_v}}(v) - 1$ is minimized and $1 + F_{T_v}(v)$ is maximized when T_v is a star centered at v. Hence we obtained:

Theorem 4 *For a given $n = |V(T)|$ and any $u \in L(T)$,*

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

with equality if and only if T is a star.

Theorem 5 *With given $|V(T)| = n$ and for $u \in L(T)$, the maximum value of $\frac{F(T)}{F_T(u)}$ is obtained when u is an endvertex of a path of length $x \geq x_0$, and the other endvertex of the path is identified with the center of a star on $n - x$ vertices. With these parameters, we have*

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

(In the terminology of [7], this tree is an $(x+1)$-comet.) The maximum of $\frac{F(T)}{F_T(u)}$ is attained when $x = n - 2\log_2 n - \frac{\ln\ln 2}{\ln 2} + o(1)$ is rounded up or down, and the maximum value of $\frac{F(T)}{F_T(u)}$ is $n(1+o(1))$.

The extremal tree in Theorem 5 seems to be extremely similar as the optimum tree for Theorem 1 in [7]. We attempt to provide some explanation for this coincidence in [7].

Theorem 6 *For a given $n = |V(T)|$, the maximum $\frac{F(T)}{F_T(v)}$ for $v \in Core(T)$, is obtained by a path, with*

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

It is easy to verify that for a path P_n on n vertices, we have

$$\frac{F(P_n)}{F_{P_n}(v)} = \frac{\binom{n}{2} + n}{\lceil \frac{n}{2} \rceil \left(\lfloor \frac{n}{2} \rfloor + 1 \right)} = \begin{cases} \frac{2n}{n+1}, & \text{if } n \text{ odd} \\ \frac{2n+2}{n+2}, & \text{if } n \text{ even} \end{cases} \tag{2.1}$$

as in the theorem. The proofs to Theorems 5 and 6 will be given in the next two sections.

3 Proof to Theorem 5: tight upper bound for $\frac{F(T)}{F_T(u)}$, $u \in L(T)$

Fix a positive integer n. Let us be given a tree T with $|V(T)| = n$ and an arbitrary $u \in L(T)$. Let u_1 denote the unique neighbor of u and let $T_1 = T \setminus \{u\}$. We start with an observation:

Lemma 7 *If a tree T and $u \in L(T)$ maximize $\frac{F(T)}{F_T(u)}$, then*

$$u_1 \in L(T_1),$$

with u_1 and T_1 defined above the Lemma.

Proof. As in the proof of Theorem 4, we obtain

$$\frac{F(T)}{F_T(u)} = 2 + \frac{\overline{F_{T_1}}(u_1) - 1}{1 + F_{T_1}(u_1)}. \tag{3.2}$$

For $n \geq 3$, we have $\overline{F_{T_1}}(u_1) \geq 1$. Suppose (for contradiction) that u_1 is not a leaf in T_1, and we will show that (3.2) can be increased.

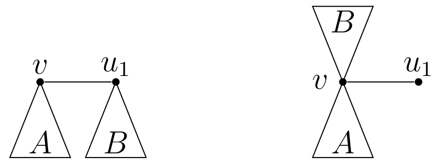

Figure 1: T_1 (left) and T_1' (right).

Focusing now on T_1, let v be a neighbor of u_1 in T_1 and A (resp. B) denote the component containing v (resp. u_1) in $T_1 \setminus u_1v$, i.e. after the deletion of the u_1v edge. The assumption that u_1 is not a leaf in T_1 will be used in the form $F_B(u_1) > 1$. Consider now T_1', which is obtained from T_1 by "moving" B from u_1 to v, see Fig. 1. Simple calculations yield

$$F_{T_1}(u_1) = (1 + F_A(v))F_B(u_1) > 1 + F_A(v)F_B(v) = F_{T_1'}(u_1)$$

and

$$\overline{F_{T_1}}(u_1) = F(A) + F(B) - F_B(u_1) < F(A) - F_A(v) + F(B) - F_B(v) + F_{A\cup B}(v) = \overline{F_{T_1'}}(u_1),$$

where in the right-hand side terms refer to T_1', in the left-hand terms refer to T_1. Note that $F_B(v)$ in T_1' has the same value as $F_B(u_1)$ in T_1, and the terms $F(A)$ and $F(B)$ are the same in the two trees. Also, $F_{A\cup B}(v) > F_A(v)$ in T_1'. Hence (3.2) is increased when T_1 is replaced with T_1'. □

Now we are going to repeat this argument. Assume that the unique neighbor of u_1 in T_1 is u_2, and consider $T_2 = T_1 \setminus \{u_1\}$. Repeating our arguments, we obtain

$$\frac{F(T)}{F_T(u)} = 2 + \frac{\overline{F_{T_1}}(u_1) - 1}{1 + F_{T_1}(u_1)} = 2 + \frac{F_{T_2}(u_2) + \overline{F_{T_2}}(u_2) - 1}{2 + F_{T_2}(u_2)} = 3 + \frac{\overline{F_{T_2}}(u_2) - 3}{2 + F_{T_2}(u_2)}.$$

Suppose that in the tree T that maximizes $\frac{F(T)}{F_T(u)}$ we have a path $uu_1u_2\dots u_x$ where u_x is the vertex

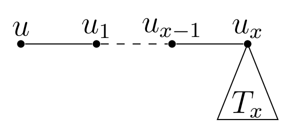

Figure 2: Structures of the optimal T and T_x.

closest to u that is of degree > 2, and consider $T_x = T \setminus \{u, u_1, ..., u_{x-1}\}$, see Fig. 2. Repeating this argument yields

$$\frac{F(T)}{F_T(u)} = x + 1 + \frac{\overline{F_{T_x}}(u_x) - \frac{1}{2}x(x+1)}{x + F_{T_x}(u_x)}. \tag{3.3}$$

Define $x_0 = x_0(n)$ to be the largest integer satisfying

$$2^{n-x-2} + (n - x - 2) \geq \frac{1}{2}x(x+1). \tag{3.4}$$

Clearly $0 < x_0 < n$ for all $n \geq 3$, and every integers in $[0, x_0]$ satisfy the inequality (3.4), while no integer in $[x_0 + 1, n]$ satisfy (3.4). The reason is that the left-hand side of (3.4) is decreasing in x, while the right-hand side of (3.4) is increasing in x. The following lemma describes the structure of T that maximizes $\frac{F(T)}{F_T(u)}$ without telling the optimal value of the parameter x.

Lemma 8 *For a given $n = |V(T)|$, the tree T and $u \in L(T)$ (see Fig. 2) that obtain the maximum value of $\frac{F(T)}{F_T(u)}$ satisfy these two properties:*
(i) $x \geq x_0 + 1$; and
(ii) T_x is a star centered at u_x.

Proof. First assume $x \leq x_0$ for contradiction. Recall (3.3)

$$\frac{F(T)}{F_T(u)} = x + 1 + \frac{\overline{F_{T_x}}(u_x) - \frac{1}{2}x(x+1)}{x + F_{T_x}(u_x)}$$

and consider two cases:
(a) If

$$\overline{F_{T_x}}(u_x) \geq \frac{1}{2}x(x+1),$$

then

$$\frac{\overline{F_{T_x}}(u_x) - \frac{1}{2}x(x+1)}{x + F_{T_x}(u_x)} \tag{3.5}$$

is nonnegative. Since u_x has degree at least 2 in T_x, let a be one of the neighbors (in T_x) of u_x and A, B be the components in $T_x \setminus \{u_x a\}$ containing a, u_x respectively (left of Fig. 3).

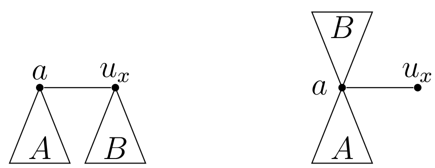

Figure 3: T_x (left) and T'_x (right).

Consider T'_x generated from T_x by moving B from u_x to a (right of Fig. 3), we have (by the same reasoning as Lemma 7)

$$F_{T_x}(u_x) = (1 + F_A(a))F_B(u_x) > 1 + F_A(a)F_B(a) = F_{T'_x}(u_x)$$

and

$$\overline{F_{T_x}}(u_x) = F(A) + F(B) - F_B(u_x) < F(A) - F_A(a) + F(B) - F_B(a) + F_{A\cup B}(a) = \overline{F_{T'_x}}(u_x).$$

Thus the numerator of (3.5) increases and the denominator of (3.5) decreases when T_x is replaced by T'_x. Consequently

$$\frac{F(T)}{F_T(u)} = x + 1 + \frac{\overline{F_{T_x}}(u_x) - \frac{1}{2}x(x+1)}{x + F_{T_x}(u_x)}$$

increases, a contradiction.

(b) If

$$\overline{F_{T_x}}(u_x) < \frac{1}{2}x(x+1),$$

then by (3.3), we have

$$\frac{F(T)}{F_T(u)} < x+1.$$

Consider T' obtained by replacing T_x by a star T'_x on $n-x$ vertices with u_x being a leaf of it. Now

$$\begin{aligned}\frac{F(T')}{F_{T'}(u)} &= x+1+\frac{\overline{F_{T'_x}}(u_x)-\frac{1}{2}x(x+1)}{x+F_{T'_x}(u_x)} \\ &= x+1+\frac{2^{n-x-2}+(n-x-2)-\frac{1}{2}x(x+1)}{x+F_{T'_x}(u_x)} \\ &\geq x+1,\end{aligned}$$

a contradiction. Hence (i) is proved.

If $x \geq x_0+2$, then

$$\frac{F(T)}{F_T(u)} = x+1+\frac{\overline{F_{T_x}}(u_x)-\frac{1}{2}x(x+1)}{x+F_{T_x}(u_x)} = x+\frac{F(T_x)-\frac{1}{2}x(x-1)}{x+F_{T_x}(u_x)}. \quad (3.6)$$

Note that

$$F(T_x) \leq 2^{n-x-1}+(n-x-1) < \frac{1}{2}x(x-1),$$

where the first inequality follows from Theorem 3 (ii), and the second inequality follows from $x-1 > x_0$. Also note that the second term in (3.6)

$$\frac{F(T_x)-\frac{1}{2}x(x-1)}{x+F_{T_x}(u_x)} < 0$$

independently of what T_x actually is, and to maximize $\frac{F(T)}{F_T(u)}$, we want to maximize both $F(T_x)$ and $F_{T_x}(u_x)$. This is achieved when T_x is a star according to Theorem 3 (i) and (ii).

Finally let us assume (for contradiction) that $x = x_0+1$, u_x is of degree at least 3 and T_x is not a star. Then T is represented by Fig. 4. Here a is a neighbor of u_{x_0+1} that is not a leaf in

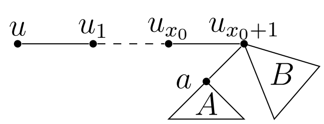

Figure 4: Structures of T and T_x when $x = x_0+1$.

T_{x_0+1} (such an a exists, since T_x is not a star) and B contains at least one vertex different from u_{x_0+1} (since $u_{x_0+1} = u_x$ is of degree at least 3). Now

$$\frac{F(T)}{F_T(u)} = x_0+1+\frac{F(T_{x_0+1})-\frac{1}{2}x_0(x_0+1)}{(x_0+1)+F_{T_{x_0+1}}(u_{x_0+1})}.$$

(a) If $F(T_{x_0+1}) < \frac{1}{2}x_0(x_0+1)$, then

$$\frac{F(T)}{F_T(u)} < x_0 + 1.$$

Consider a T' obtained from T by replacing T_{x_0+1} with a star on $n - x_0 - 1$ vertices centered at u_{x_0+1}. Simple calculation shows that

$$\frac{F(T')}{F_{T'}(u)} = x_0 + 1 + \frac{2^{n-x_0-2} + n - x_0 - 2 - \frac{1}{2}x_0(x_0+1)}{(x_0+1) + 2^{n-x_0-2}},$$

which is $\geq x_0 + 1$ by the definition of x_0, a contradiction.

(b) If $F(T_{x_0+1}) \geq \frac{1}{2}x_0(x_0+1)$, then consider T'_{x_0+1} obtained by "moving" B from u_{x_0+1} to a but keeping u_{x_0+1} in the tree (see Fig. 5), and the corresponding T' obtained from T.

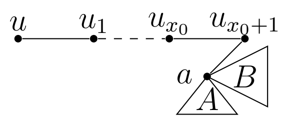

Figure 5: Structures of T' and T'_{x_0+1}.

Then, as $F_A(a) > 1$ and $F_B(a) > 1$, we have

$$\begin{aligned} F(T'_{x_0+1}) &= 1 + 2F_A(a)F_B(a) + \overline{F_A}(a) + \overline{F_B}(a) \\ &> (F_A(a)+1)F_B(u_{x_0+1}) + F_A(a) + \overline{F_A}(a) + \overline{F_B}(u_{x_0+1}) \\ &= F(T_{x_0+1}); \end{aligned}$$

and, as $F_B(a) > 1$, we have

$$F_{T'_{x_0+1}}(u_{x_0+1}) = 1 + F_A(a)F_B(a) < (F_A(a)+1)F_B(u_{x_0+1}) = F_{T_{x_0+1}}(u_{x_0+1}).$$

Hence

$$\frac{F(T')}{F_{T'}(u)} = x_0 + 1 + \frac{F(T'_{x_0+1}) - \frac{1}{2}x_0(x_0+1)}{(x_0+1) + F_{T'_{x_0+1}}(u_{x_0+1})} > x_0 + 1 + \frac{F(T_{x_0+1}) - \frac{1}{2}x_0(x_0+1)}{(x_0+1) + F_{T_{x_0+1}}(u_{x_0+1})} = \frac{F(T)}{F_T(u)},$$

a contradiction. □

Lemma 8 immediately forces the structure required in Theorem 5. At this point we are left with computing the asymptotics for the optimal x and the corresponding $\frac{F(T)}{F_T(u)}$ value. Relaxing the optimization problem for real x's, we obtain

$$f'(x) = \frac{2^{2(n-x-1)} + 2^{n-x-1}\left(-\frac{\ln 2}{2}x^2 + \frac{2-3\ln 2}{2}x + (n-1)\ln 2 - \frac{3}{2}\right) + \frac{x^2}{2} - (n-1)}{(2^{n-x-1}+x)^2}.$$

We try to solve $f'(x)=0$ for real x (in terms of n). Considering it as a quadratic equation in 2^{n-x-1}, we obtain

$$2^{n-x-1} = \frac{1}{2}\left(\frac{\ln 2}{2}x^2+\frac{3\ln 2-2}{2}x+\frac{3}{2}-(n-1)\ln 2\right) \tag{3.7}$$

$$\pm \sqrt{n-1-\frac{x^2}{2}+\frac{1}{4}\left(\frac{\ln 2}{2}x^2+\frac{3\ln 2-2}{2}x+\frac{3}{2}-(n-1)\ln 2\right)^2}. \tag{3.8}$$

As (3.4) is satisfied by $x=\lceil n/2\rceil$, hence $n/2\le x_0$, and we have to maximize $f(x)$, and hence solve $f'(x)=0$, on $\frac{n}{2}\le x\le n$, unless the optimum occurs at the endpoints. Regarding the endpoints of the intervals, it is easy to see that

$$x\sim\frac{n}{2} \text{ implies } f(x)\sim\frac{n}{2}; \text{ and } x\sim n \text{ implies } f(x)\sim\frac{n}{2}.$$

Set $A(x)=\frac{x^2}{2}-(n-1)$ and $B(x)=\frac{1}{2}\left(\frac{\ln 2}{2}x^2+\frac{3\ln 2-2}{2}x+\frac{3}{2}-(n-1)\ln 2\right)$.

Consider the positive sign in (3.8). As in $[n/2,n]$ the function 2^{n-x-1} decreases and $B(x)+\sqrt{B^2(x)-A(x)}$ increases; for $x=\frac{n}{2}-O(1)$ the function 2^{n-x-1} is exponential while $B(x)+\sqrt{B^2(x)-A(x)}$ is quadratic; and for $x=n-O(1)$ the function 2^{n-x-1} is bounded while $B(x)+\sqrt{B^2(x)-A(x)}$ is quadratic; we conclude that for n sufficiently large, there is a unique x_a in the interval $[n/2,n]$ that solves the equation

$$2^{n-x-1}=B(x)+\sqrt{B^2(x)-A(x)}. \tag{3.9}$$

Using Newton's Binomial Theorem for $\sqrt{1-x}$ and the fact that a convergent Taylor series of a function around 0 provides asymptotic expansion of the function at 0, we get

$$\begin{aligned}B(x)+\sqrt{B^2(x)-A(x)} &= B(x)\left(1+\sqrt{1-\frac{A(x)}{B^2(x)}}\right)=B(x)\left(2+O\Big(\frac{A(x)}{B^2(x)}\Big)\right)\\ &= 2B(x)(1+o(1))=\frac{\ln 2}{2}x^2(1+o(1)).\end{aligned}$$

Taking the logarithm of equation (3.9), which is spelled out in (3.7,3.8), we obtain asymptotics for x_a as

$$n-x_a-1=\log_2\left(\Big(\frac{\ln 2}{2}x_a^2\Big)(1+o(1))\right)=2\log_2 x_a+\frac{\ln\ln 2}{\ln 2}-1+o(1)$$

and

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

Substituting the right-hand side of the previous formula into the place of x_a in the same right-hand side, we obtain

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

It is easy to see that $f(\lfloor x_a \rfloor) \sim f(\lceil x_a \rceil) \sim n$.

Turning to the negative squareroot, if it gives an optimum x_b with $n/2 \le x_b \le b$, then with A and B above we have

$$2^{n-x_b-1} = B(x_b) - \sqrt{B^2(x_b) - A(x_b)}.$$

Using again Newton's Binomial Theorem for $\sqrt{1-x}$ and the fact that a convergent Taylor series of a function around 0 provides asymptotic expansion of the function at 0, we get

$$\begin{aligned} B(x) - \sqrt{B^2(x) - A(x)} &= B(x)\left(1 - \sqrt{1 - \frac{A(x)}{B^2(x)}}\right) \\ &= B(x)\left(1 - 1 + \frac{1}{2}\frac{A(x)}{B^2(x)} + O\Big(\frac{A^2(x)}{B^4(x)}\Big)\right) = \frac{A(x)}{2B(x)} + O\Big(\frac{A^2(x)}{B^3(x)}\Big), \end{aligned}$$

where $\frac{A(x)}{2B(x)} \sim \frac{1}{2}\frac{\frac{x^2}{2}-(n-1)}{\frac{\ln 2}{4}x^2} \sim \frac{1}{\ln 2}$ for all $n/2 \le x \le n$. Therefore $2^{n-x_b-1} = \frac{1}{\ln 2} + o(1)$, and hence $x_b = n - O(1)$. It is easy to check that $f(\lfloor x_b \rfloor) \sim f(\lceil x_b \rceil) \sim n/2$, therefore the maximum place among real numbers is at x_a in the interval $[n/2, n]$.

4 Proof to Theorem 6: tight upper bound on $\frac{F(T)}{F_T(v)}$, $v \in Core(T)$

For a given $n = |V(T)|$ and $v \in Core(T)$, by the identity

$$R(T) := \frac{F(T)}{F_T(v)} = 1 + \frac{\overline{F_T}(v)}{F_T(v)},$$

$R(T)$ is maximized if and only if

$$R'(T) := \frac{\overline{F_T}(v)}{F_T(v)}$$

is maximized.

Now let v be of degree k with neighbors v_i $(i = 1, \dots, k)$ and T_i $(i = 1, \dots, k)$ be the corresponding branches (see Fig. 6). Then

$$R'(T) = \frac{\sum_{i=1}^{k} F(T_i)}{\prod_{i=1}^{k}(1 + F_{T_i}(v_i))} = \frac{\sum_{i=1}^{k}(F_{T_i}(v_i) + \overline{F_{T_i}}(v_i))}{\prod_{i=1}^{k}(1 + F_{T_i}(v_i))}. \tag{4.10}$$

First we establish a key Lemma that characterizes paths.

Lemma 9

$$\overline{F_T}(v) \le \binom{F_T(v)}{2} = \frac{F_T(v)^2 - F_T(v)}{2}$$

for any T and $v \in T$. Equality is obtained if and only if v is an endvertex of the path.

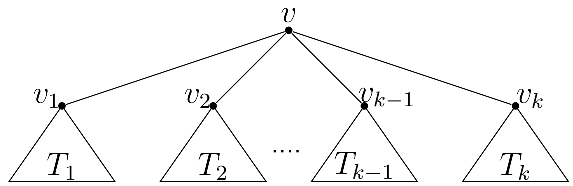

Figure 6: T and its subtrees.

Proof. For $n \geq 1$, this can be proved by establishing an injection from the set of subtrees not containing v to the set of unordered pairs of subtrees containing v. If T' is a subtree in the first set, we define its image as a pair $\{T_1, T_2\}$ in the second set, where T_2 is the smallest subtree of T containing both v and T', and T_1 is the path connecting v to the closest vertex of T' in T. This is an injection as T' can be recovered from its image in the following way: let T_1 denote the element of the image pair that is contained in the other element. T' is obtained by removing the vertices of T_1 with the exception of the endpoint different from v from T_1. (See Fig. 7.) If T is a path, then this injection is clearly a surjection as well. If T is not a path, or if T is a path but v is not an endvertex of the path, then there are a pair subtrees of T, both containing v, such that none of them is a subtree of the other. Such a pair is not an image under the injection.

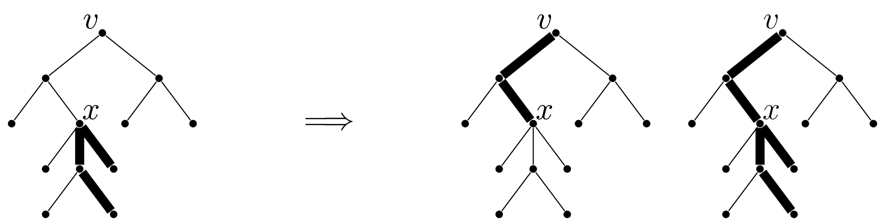

Figure 7: The injection from T' (left) to T_1 (middle) and T_2 (right).

□

We claim that the maximum of $R'(T)$ is obtained when $k = 2$.

Lemma 10 *For $n \geq 6$, the degree of $v \in Core(T)$ is 2 in any tree that maximizes $R'(T)$.*

Proof. If $k = 1$, then $n = 2$. Suppose for contradiction that $k \geq 3$, and assume without loss of generality that

$$F_{T_1}(v_1) \geq F_{T_2}(v_2) \geq F_{T_i}(v_i)$$

for any $i \neq 1, 2$. By Theorem 1 we have $v_2 \notin Core(T)$. By applying Lemma 9 to each T_i and v_i in

(4.10), we have

$$
\begin{aligned}
R'(T) &\le \frac{\sum_{i=1}^{k}\left(\binom{F_{T_i}(v_i)}{2}+F_{T_i}(v_i)\right)}{\prod_{i=1}^{k}(1+F_{T_i}(v_i))} && (4.11)\\
&= \frac{1}{2}\sum_{i=1}^{k}\frac{F_{T_i}(v_i)}{\prod_{j:j\neq i}(F_{T_j}(v_j)+1)}\\
&\le \frac{1}{2}\frac{F_{T_1}(v_1)}{\prod_{i>1}(F_{T_i}(v_i)+1)}+\frac{1}{2}\frac{\sum_{i>1}F_{T_i}(v_i)}{(F_{T_1}(v_1)+1)\prod_{i>2}(F_{T_i}(v_i)+1)}\\
&< \frac{1}{2}\frac{F_{T_1}(v_1)}{\prod_{i>1}(F_{T_i}(v_i)+1)}+\frac{1}{2}\frac{\prod_{i>1}(F_{T_i}(v_i)+1)}{(F_{T_1}(v_1)+1)\prod_{i>2}(F_{T_i}(v_i)+1)}\\
&\le \frac{1}{2}\frac{F_{T_1}(v_1)}{\prod_{i>1}(F_{T_i}(v_i)+1)}+\frac{1}{2}\frac{F_{T_2}(v_2)+1}{F_{T_1}(v_1)+1}. && (4.12)
\end{aligned}
$$

Note that since v is in $Core(T)$, we must have

$$F_T(v) \ge F_T(v_i)$$

for any i. By comparing the number of subtrees containing v but not v_i and the number of subtrees containing v_i but not v, we have

$$\prod_{j:j\neq i}(1+F_{T_j}(v_j)) \ge F_{T_i}(v_i)$$

for any i. Particularly, for $i = 1$, setting

$$x = F_{T_1}(v_1),$$

$$y = F_{T_2}(v_2) + 1,$$

and

$$z = \prod_{i>2}(F_{T_i}(v_i) + 1)$$

yields

$$yz = ax \text{ with } a \ge 1, z \ge 2, x \ge y - 1.$$

Now from (4.11—4.12),

$$R'(T) < \frac{1}{2}\left(\frac{y}{x+1} + \frac{1}{a}\right). \tag{4.13}$$

If $x = y - 1$, then $a = \frac{y}{x}z > z \ge 2$ and

$$R'(T) < \frac{1}{2}\left(1 + \frac{1}{a}\right) < \frac{1}{2}\left(1 + \frac{1}{2}\right) = \frac{3}{4}. \tag{4.14}$$

Otherwise, $x \ge y$ and $a \le z$, we have

$$R'(T) < \frac{1}{2}\left(\frac{y}{x+1} + \frac{1}{a}\right) \le \frac{1}{2}\left(\frac{y}{x} + \frac{1}{a}\right) = \frac{1}{2}\left(\frac{a}{z} + \frac{1}{a}\right). \tag{4.15}$$

Consider the function $f(a) = \frac{1}{2}(\frac{a}{z} + \frac{1}{a})$ in $1 \le a \le z$. This function takes its maximum at $a = 1$ and $a = z$ with the value $f(1) = f(z) = \frac{1}{2}(1 + \frac{1}{z})$, and its minimum at $a = \sqrt{z}$ with the value $f(\sqrt{z}) = \frac{1}{\sqrt{z}}$.

By (4.14) and (4.15), $R'(T) < 3/4$. By (2.1), $n \le 5$. □

Therefore in a tree maximizing $R'(T)$, $v \in Core(T)$ must be of degree 2. Observe that

$$F_{T_2}(v_2) \le F_{T_1}(v_1) \le 1 + F_{T_2}(v_2). \tag{4.16}$$

The reason is the following:

$$F_T(v) = (F_{T_1}(v_1) + 1)(F_{T_2}(v_2) + 1) = 1 + F_{T_1}(v_1) + F_{T_2}(v_2) + F_{T_1}(v_1)F_{T_2}(v_2)$$

and

$$F_T(v) \ge F_T(v_1) = 2F_{T_1}(v_1) + F_{T_1}(v_1)F_{T_2}(v_2).$$

Lemma 11 *For $n \ge 6$, if $|V(T)| = n$, T and $v \in Core(T)$ maximize $R'(T)$, then both neighbors of v in T, v_1 and v_2, are of degree 2.*

Proof. We no longer make the assumption that $F_{T_1}(v_1) \ge F_{T_2}(v_2)$, so the numbering of v_1 and v_2 is arbitrary. If v_1 has degree 1, then it is easy to see that $n \le 4$. Suppose for contradiction that v_2 is of degree at least three. Then T can be represented as in Fig. 8 where both A and B contain more than one vertex, in other words $a := F_A(v_2) \ge 2$ and $b := F_B(v_2) \ge 2$, and $F_{T_2}(v_2) = F_A(v_2)F_B(v_2) = ab$. By (4.16), $F_{T_1}(v_1) \ge F_{T_2}(v_2) - 1 = ab - 1$.

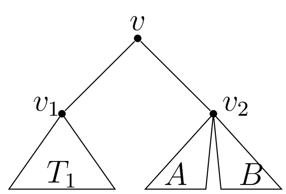

Figure 8: T, T_1, T_2 and the subtrees.

Now we have

$$\begin{aligned}
R'(T) &= \frac{F_{T_1}(v_1) + \overline{F_{T_1}}(v_1) + F_A(v_2)F_B(v_2) + \overline{F_A}(v_2) + \overline{F_B}(v_2)}{(F_{T_1}(v_1)+1)(F_{T_2}(v_2)+1)} \\
&\quad \text{(by applying Lemma 9 to } T_1, A \text{ and } B) \\
&\le \frac{\frac{1}{2}F_{T_1}(v_1)(F_{T_1}(v_1)+1) + F_A(v_2)F_B(v_2) + \frac{1}{2}(F_A(v_2)^2 - F_A(v_2)) + \frac{1}{2}(F_B(v_2)^2 - F_B(v_2))}{(F_{T_1}(v_1)+1)(F_{T_2}(v_2)+1)} \\
&\le \frac{1}{2}\frac{F_{T_1}(v_1)}{F_{T_2}(v_2)+1} + \frac{ab + \frac{1}{2}(a^2 - a) + \frac{1}{2}(b^2 - b)}{ab(ab+1)}. \qquad (4.17)
\end{aligned}$$

Note that $\frac{1}{2}\frac{F_{T_1}(v_1)}{F_{T_2}(v_2)+1} \le \frac{1}{2}$ by (4.16). We consider cases.

Case $a = 2, b = 2$:

As $F_{T_2}(v_2) = 4$, we have $F_{T_1}(v_1) \in \{3, 4, 5\}$. We list the possible trees below in Fig. 9.

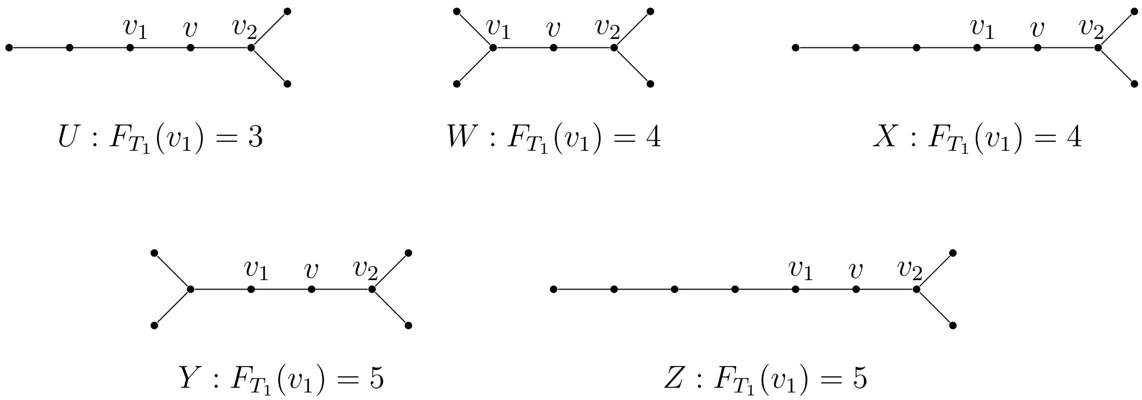

Figure 9: Cases for $a = 2, b = 2$.

For these trees, easy calculation shows $R'(U) = \frac{3}{5} < R'(P_7) = \frac{3}{4}$, $R'(W) = \frac{12}{25} < R'(P_7) = \frac{3}{4}$, $R'(X) = \frac{16}{25} < R'(P_8) = \frac{4}{5}$, $R'(Y) = \frac{17}{30} < R'(P_8) = \frac{4}{5}$, and $R'(Z) = \frac{7}{10} < R'(P_9) = \frac{4}{5}$, contradicting the maximality of the respective R' fractions.

Case $a = 2, b = 3$**:**
Based on (4.17), $R'(T) \le \frac{1}{2} + \frac{10}{6 \cdot 7} < \frac{3}{4}$, implying $n \le 5$.

Case $a \ge 3, b \ge 3$**:**
We have $\frac{ab + \frac{1}{2}(a^2 - a) + \frac{1}{2}(b^2 - b)}{ab(ab+1)} = \frac{1}{1+ab} + \frac{a-1}{2b(ab+1)} + \frac{b-1}{2a(ab+1)} \le \frac{1}{ab} + \frac{1}{2b^2} + \frac{1}{2a^2} = \frac{1}{2}\left(\frac{1}{a} + \frac{1}{b}\right)^2$, which is maximal at $a = 3, b = 3$. Hence $R'(T) \le \frac{1}{2} + \frac{1}{2} \cdot \frac{4}{9} = \frac{13}{18} < \frac{3}{4}$, implying $n \le 5$. □

Now we are finally ready to prove Theorem 6. The proof goes by induction on n. The base cases are $1 \le n \le 5$. For $n = 1, 2, 3$, the only tree is a path, therefore there is nothing to prove. For $n = 4$, there are two trees, P_4 and the star S_4. The star is out by Theorem 3, as it is the single minimizer of $R = R' + 1$. For $n = 5$, there are three trees, P_5, the star S_5, and Q_5, which comes from a P_4 by attaching a leaf to an interior vertex. The star is out again, and $R'(Q_5) = \frac{5}{12} < R'(P_5) = \frac{2}{3}$, as required.

With Lemma 11, we can represent the optimal T as in Fig. 10 and write

$$R'(T) = \frac{2F_A(a) + 2F_B(b) + 2 + \overline{F_A}(a) + \overline{F_B}(b)}{(F_A(a) + 2)(F_B(b) + 2)}. \tag{4.18}$$

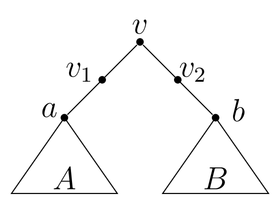

Figure 10: Another representation of T.

By (4.16), we can assume $F_B(b) = F_A(a) + \epsilon$ where $\epsilon \in \{0, 1\}$. Then (4.18) can be written as

$$R'(T) = \frac{2(F_A(a)+1)+\epsilon}{(F_A(a)+2)(F_A(a)+2+\epsilon)} + \frac{F_A(a)+F_B(b)+\overline{F_A}(a)+\overline{F_B}(b)}{(F_A(a)+2)(F_A(a)+2+\epsilon)}. \tag{4.19}$$

Note that by removing v_1, v_2 from T and joining a, b to v, we get a tree T' on $n-2$ vertices (see Fig. 11), for which

$$R'(T') = \frac{F_A(a)+F_B(b)+\overline{F_A}(a)+\overline{F_B}(b)}{(F_A(a)+1)(F_A(a)+1+\epsilon)}.$$

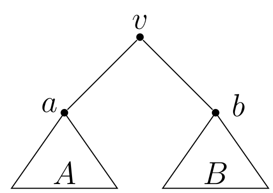

Figure 11: T'.

We are going to show

$$R'(T) \le \frac{\lfloor \frac{n}{2} \rfloor}{\lfloor \frac{n}{2} \rfloor + 1} \tag{4.20}$$

by induction on n. Rewrite (4.19) as

$$\begin{aligned} &R'(T) \\ =&\frac{2(F_A(a)+1)+\epsilon}{(F_A(a)+2)(F_A(a)+2+\epsilon)} + \frac{F_A(a)+F_B(b)+\overline{F_A}(a)+\overline{F_B}(b)}{(F_A(a)+1)(F_A(a)+1+\epsilon)} \cdot \frac{(F_A(a)+1)(F_A(a)+1+\epsilon)}{(F_A(a)+2)(F_A(a)+2+\epsilon)} \\ \le&\frac{2(F_A(a)+1)+\epsilon}{(F_A(a)+2)(F_A(a)+2+\epsilon)} + \frac{\lfloor \frac{n}{2} \rfloor - 1}{\lfloor \frac{n}{2} \rfloor} \cdot \frac{(F_A(a)+1)(F_A(a)+1+\epsilon)}{(F_A(a)+2)(F_A(a)+2+\epsilon)}. \end{aligned}$$

Hence, with

$$f(x) = \frac{2x + \epsilon + \frac{\lfloor \frac{n}{2} \rfloor - 1}{\lfloor \frac{n}{2} \rfloor} x(x+\epsilon)}{(x+1)(x+1+\epsilon)},$$

we have $R'(T) \le f(F_A(a)+1)$. It is simple algebra to check that for both $\epsilon \in \{0, 1\}$, we have

$$f\big(\lfloor \frac{n}{2} \rfloor\big) = \frac{\lfloor \frac{n}{2} \rfloor}{\lfloor \frac{n}{2} \rfloor + 1}.$$

Therefore, to show (4.20), it suffices to show that for all integer x

$$f(x) \le f\big(\lfloor \frac{n}{2} \rfloor\big)$$

holds. Using the definition of $f(x)$ as a fraction, this is equivalent to

$$\left(2x+\epsilon+\left(1-\frac{1}{\lfloor\frac{n}{2}\rfloor}\right)x(x+\epsilon)\right)\left(\lfloor\frac{n}{2}\rfloor+1\right)\left(\lfloor\frac{n}{2}\rfloor+1+\epsilon\right)$$

$$\leq\left(2\lfloor\frac{n}{2}\rfloor+\epsilon+\left(\lfloor\frac{n}{2}\rfloor-1\right)(\lfloor\frac{n}{2}\rfloor+\epsilon)\right)\left(x+1\right)\left(x+1+\epsilon\right),$$

which is, in turn, equivalent to

$$0\leq\frac{\left(\lfloor\frac{n}{2}\rfloor+1+\epsilon\right)\left(\left(x-\lfloor\frac{n}{2}\rfloor\right)^2+\epsilon\left(x-\lfloor\frac{n}{2}\rfloor\right)\right)}{\lfloor\frac{n}{2}\rfloor}.$$

The last inequality holds as $x-\lfloor\frac{n}{2}\rfloor$ is an integer. Therefore, (4.20) is proved by induction. Furthermore, if equality holds for T in (4.20), then equality must hold for T' in its corresponding (4.20), with $n\leftarrow n-2$. Eventually, by induction, T must be a path.

References

[1] C.A. Barefoot, R.C. Entringer, L.A. Székely, Extremal values for ratios of distances in trees, *Disc. Appl. Math.* **80** (1997) 37–56.

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

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

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

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

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

[7] L.A. Székely and Hua Wang, Extremal values of ratios: distance problems vs. subtree problems in trees I, to appear.

[8] S. Wagner, Correlation of graph-theoretical indices, *SIAM J. Discrete Mathematics* **21** (2007), no. 1, 33–46.

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