ON THE LOCAL AND GLOBAL MEANS OF SUBTREE ORDERS

STEPHAN WAGNER AND HUA WANG

Abstract. The global mean of subtrees of a tree is the average order (i.e., average number of vertices) of its subtrees. Analogously, the local mean of a vertex in a tree is the average order of subtrees containing this vertex. In the comprehensive study of these concepts by Jamison (JCT B, 1983 and 1984), several open questions were proposed. One of them asks if the largest local mean always occurs at a leaf vertex. Another asks if it is true that the local mean of any vertex of any tree is at most twice the global mean. In this note, we answer the first question by showing that the largest local mean always occurs at a leaf or a vertex of degree two and that both cases are possible. With this result, a positive answer to the second question is provided. We also show some related results on local mean and global mean of trees.

1. Introduction and statement of results

For a tree T, let $N(T)$ denote the number of subtrees of T and $R(T)$ denote the sum of the number of vertices of these subtrees. The *global mean* of T, denoted by

$$M_T = \frac{R(T)}{N(T)},$$

is the average number of vertices of subtrees of T. As trees of large order also have large global means in general, it is natural to also consider the (subtree) *density* of T, which is the normalized average order:

$$D_T = \frac{M_T}{|V(T)|}.$$

Similarly, let $N_T(v)$ denote the number of subtrees containing v in T and $R_T(v)$ denote the total order of these subtrees. The *local mean* of v in T is the average order of subtrees of T that contain v, and it is denoted by

$$\mu_T(v) = \frac{R_T(v)}{N_T(v)}.$$

The *local density* is defined in analogy to the density and denoted by

$$D_T(v) = \frac{\mu_T(v)}{|V(T)|}.$$

In 1983 and 1984, Jamison [2, 3] provided a rather comprehensive study of these concepts and related questions. For instance, it was shown that

- the global mean of a tree of order n is at least $(n+2)/3$, and equality is achieved only by a path,
- the local mean at a vertex always exceeds the global mean (with equality for the tree of order 1), and it is at least $(n+1)/2$ for a tree of order n.

2010 *Mathematics Subject Classification.* 05C05.

Key words and phrases. Subtrees, mean order, local mean, global mean, homeomorphically irreducible trees.

This material is based upon work supported financially by the National Research Foundation of South Africa under grant number 70560.

Interestingly, trees with high global mean must also have a large number of vertices of degree 2: Jamison proved in [2] that if the density $D(T_n)$ tends to 1 for a sequence T_n of trees, the proportion of vertices of degree 2 has to tend to 1 as well.

Since vertices of degree 2 are so special in this context, Jamison proposed a conjecture regarding trees without such vertices (known as homeomorphically irreducible trees, henceforth denoted by $\mathcal{T}_3$), namely that their density is at least 1/2. This conjecture was only verified recently [7]. On a related note, the order of subtrees and in particular the average total order $R(T)$ were studied for random rooted trees [5]; the set of subtrees of a tree was also considered as a poset or lattice [4, 6]. Several other interesting questions were also asked in the aforementioned papers of Jamison, most of which are still open. In this note, we are mainly interested in the following two questions:

Question 1.1. [2] *For any tree T, is the largest local mean of T always taken on at a leaf vertex?*

Question 1.2. [2] *For any tree T and vertex v of T, does $\mu_T(v) \leq 2M_T$ necessarily hold?*

We will start with some simple observations before we turn to the local mean in Section 3, where we show that every vertex of degree at least 3 has a neighbor with larger local mean. This immediately implies that the largest local mean can only occur at a leaf or a vertex of degree 2. Hence the answer to Question 1.1 is yes for the aforementioned class of homeomorphically irreducible trees. In general, however, we provide examples where the largest local mean is only attained at a vertex of degree 2. As one may expect, these examples contain many vertices of degree 2.

We deal with Question 1.2 in Section 4. It was shown in [3] that

$$M_T \leq \mu_T(v) < M_T + \frac{n-1}{2}$$

for any $v \in V(T)$. This implies a positive answer to the question for trees with density greater than 1/2, which includes for instance all trees in $\mathcal{T}_3$ [7]. Making use of our results regarding Question 1.1, we show that the answer is indeed yes for general trees. Once again, the extremal cases are trees with many vertices of degree 2. We also provide a much stronger theorem for trees in $\mathcal{T}_3$, namely that $\mu_T(v) = M_T + O(1)$ for trees in $\mathcal{T}_3$.

2. Preliminaries

Let us start with a few preliminary observations. In the following, we will repeatedly deal with *rooted* trees: let v be the root of a tree T, and denote its branches by $T_1, T_2, \ldots, T_k$, rooted at the neighbors $v_1, v_2, \ldots, v_k$ of v. Moreover, let $N_i = N_{T_i}(v_i)$ denote the number of subtrees of T_i that contain v_i, and $R_i = R_{T_i}(v_i)$ the sum of their orders. Then one has the simple recursions

$$N_T(v) = \prod_{j=1}^{k} (1 + N_j) \tag{1}$$

and

$$R_T(v) = N_T(v) + N_T(v) \sum_{i=1}^{k} \frac{R_i}{1 + N_i}, \tag{2}$$

cf. [2, 7], which follow easily from the fact that a subtree containing v induces either the empty set or a subtree containing the root in each of the branches T_i.

Similarly, we also have

$$N_T(v_i) = \frac{N_i}{1+N_i} N_T(v) + N_i, \tag{3}$$

since a subtree of T that contains v_i is either a subtree of T_i that contains v_i (N_i choices), or it is among the subtrees that contain v and do not induce the empty subtree on T_i (their proportion is $N_i/(1+N_i)$). Finally,

$$R_T(v_i) = R_i \left(1 + \frac{N_T(v)}{1+N_i}\right) + \frac{N_i}{1+N_i} N_T(v) + \frac{N_i}{1+N_i} N_T(v) \sum_{j \neq i} \frac{R_j}{1+N_j} \tag{4}$$

for $1 \leq i \leq k$. The first term is the contribution from T_i, the second term is the contribution from v (which appears in $N_i N_T(v)/(1+N_i)$ of the subtrees containing v_i), and the j-th term in the final sum is the contribution from T_j. The four identities above will be used frequently, sometimes without further mention, in the rest of this paper.

Next we state a little lemma that is essentially trivial, but will turn out surprisingly useful. It is also stated in [2, Lemma 3.2c], but we give a slightly more direct proof here for completeness.

Lemma 2.1. *For any tree T and any vertex v, the inequality*

$$R_T(v) \leq \frac{N_T(v)^2 + N_T(v)}{2}$$

holds, with equality if and only if T is a path and v is a leaf of T.

Proof. Order the $N_T(v)$ subtrees that contain v by their size. If one of these subtrees has order $k > 1$, then there must also be such a subtree of order $k-1$ (obtained by removing a leaf other than v). Hence the j-th largest subtree that contains v has order at most j, which means that the total order $R_T(v)$ of all subtrees is at most

$$1 + 2 + \cdots + N_T(v) = \frac{N_T(v)^2 + N_T(v)}{2}.$$

For equality to hold, every subtree containing v must have a unique leaf other than v, i.e., it must be a path and v one of its leaves. In particular, this must be the case for T itself, and indeed equality holds if T is a path and v one of its leaves. ∎

3. On the largest local mean

In this section, we provide the answer to Jamison's Question 1.1, which is almost affirmative: the maximum local mean occurs either at a leaf or at a vertex of degree 2. As a first step, we consider the local mean at a leaf:

Lemma 3.1. *The local mean at a leaf of a tree T is at least as large as that of its neighbor, with equality only if T is a path.*

Proof. Consider a leaf v, and let $uv \in E(T)$ be the edge incident to this leaf. Moreover, let S be the tree obtained from T by removing v, i.e., $S = T - \{v\}$. A subtree of T that contains u is either a subtree of S that contains u or obtained from such a subtree by adding v. Hence $N_T(u) = 2N_S(u)$ and $R_T(u) = 2R_S(u) + N_S(u)$ and thus

$$\mu_T(u) = \frac{R_T(u)}{N_T(u)} = \frac{2R_S(u) + N_S(u)}{2N_S(u)}.$$

Likewise, a subtree of T that contains v is either obtained from a subtree of S that contains u by adding v, or it consists only of v. Hence $N_T(v) = N_S(u) + 1$ and $R_T(v) = R_S(u) + N_S(u) + 1$, and we get

$$\mu_T(v) = \frac{R_T(v)}{N_T(v)} = \frac{R_S(u) + N_S(u) + 1}{N_S(u) + 1}.$$

The difference is
$$\mu_T(v) - \mu_T(u) = \frac{1}{2} - \frac{R_S(u)}{N_S(u)^2 + N_S(u)},$$
so the theorem follows immediately from Lemma 2.1. ■

Remark 1. If it was not for the subtree consisting of v only, the difference between $\mu_T(v)$ and $\mu_T(u)$ would be exactly $\frac{1}{2}$. The proof above confirms the intuition that this single subtree does not make a big difference.

Now we are ready to prove the main result of this section: a vertex v of degree 3 or more cannot have maximum local mean. In its essence, the idea is similar to the previous proof. We take the neighbor w for which the number of subtrees containing w but not v is smallest. If we compare the local means of v and w, then we find that the gain to the average subtree order from having to include w outweighs the loss from the small number of subtrees that contain w, but not v, in very much the same way as the single subtree $\{v\}$ did not matter in the previous lemma.

Theorem 3.2. *Let T be a tree, and let v be a vertex of degree at least* 3 *in T. There exists a neighbor w of v such that the local mean $\mu_T(w)$ is greater than the local mean $\mu_T(v)$. Hence the maximum local mean in a tree occurs either at a leaf or at a vertex of degree* 2*.*

Proof. Let v be a vertex of a tree T whose degree is $k \geq 3$, and let $v_1, v_2, \dots, v_k$ be the neighbors of v. Moreover, $T_1, \dots, T_k$ denote the trees in $T - v$ with $v_i \in V(T_i)$. Suppose (for contradiction) that v has at least as large a local mean as that of any v_i. Then by Lemma 3.1, we can assume that T_i is of order at least 2 for all $i \in \{1, 2, \dots, k\}$.

Making use of the relations (1) to (4), we find after some algebraic manipulations that our assumption,
$$\frac{R_T(v)}{N_T(v)} \geq \frac{R_T(v_i)}{N_T(v_i)} \text{ for } 1 \leq i \leq k,$$
holds if and only if the following is true for all $i \in \{1, 2, \dots, k\}$:
$$\frac{R_i}{N_i} \leq \frac{1 + \sum_{j=1}^k \frac{R_j}{1+N_j}}{1 + (1+N_i)^{-2} \prod_{j=1}^k (1+N_j)}. \tag{5}$$
Set $N = \min\{N_i : 1 \leq i \leq k\}$, and choose an index ℓ such that $N_\ell = N$. Of all subtrees of T_ℓ that contain v_ℓ, clearly only one is of order 1, all others have at least two vertices. Since $|V(T_\ell)| \geq 2$, we also have $N_\ell \geq 2$ and thus
$$\frac{R_\ell}{N_\ell} \geq \frac{1 + 2(N_\ell - 1)}{N_\ell} = 2 - \frac{1}{N_\ell} \geq \frac{3}{2}.$$
Since (5) has to hold for all i, it has to hold in particular when $i = \ell$, so we obtain
$$\frac{3}{2} \leq \frac{R_\ell}{N_\ell} \leq \frac{1 + \sum_{j=1}^k \frac{R_j}{1+N_j}}{1 + (1+N_\ell)^{-2} \prod_{j=1}^k (1+N_j)} = \frac{1 + \sum_{j=1}^k \frac{R_j}{1+N_j}}{1 + (1+N)^{-2} \prod_{j=1}^k (1+N_j)}. \tag{6}$$
Note also that (by Lemma 2.1)
$$\frac{R_j}{1+N_j} \leq \frac{N_j}{2}$$
for all j, which means that
$$\frac{3}{2} \leq \frac{1 + \frac{1}{2}\sum_{j=1}^k N_j}{1 + (1+N)^{-2} \prod_{j=1}^k (1+N_j)}.$$

Now set

$$f(x_1, x_2, \dots, x_k) = \frac{1 + \frac{1}{2}\sum_{j=1}^{k} x_j}{1 + (1+N)^{-2}\prod_{j=1}^{k}(1+x_j)}.$$

We show that the maximum of this function, subject to the condition that $x_j \geq N$ for all j, is less than 3/2, which will yield a contradiction, since $N_j \geq N$ for all j by our choice of N. The partial derivative of f with respect to any x_j is

$$\frac{\partial f}{\partial x_j} = \frac{1 - (1+N)^{-2}\prod_{i\neq j}(1+x_i)\left(1 + \sum_{i\neq j} x_i\right)}{2\left(1 + (1+N)^{-2}\prod_{j=1}^{k}(1+x_j)\right)^2},$$

which is negative when $k \geq 3$, since $(1+N)^{-2}\prod_{i\neq j}(1+x_i) \geq 1$ by our assumptions and $\sum_{i\neq j} x_i > 0$. Hence the maximum of f is attained when $x_j = N$ for all j, and this maximum is

$$\frac{1 + \frac{kN}{2}}{1 + (1+N)^{k-2}}.$$

However, Bernoulli's inequality [1, Theorem 58] yields

$$\frac{3}{2}\left(1 + (1+N)^{k-2}\right) \geq 3 + \frac{3}{2}(k-2)N > 1 + \frac{kN}{2}$$

for $k \geq 3$, hence

$$\frac{3}{2} \leq f(N_1, N_2, \dots, N_k) \leq \frac{1 + \frac{kN}{2}}{1 + (1+N)^{k-2}} < \frac{3}{2},$$

which finally gives us a contradiction. ∎

Theorem 3.2 shows that the maximum local mean of a tree must occur at either a leaf or a vertex of degree 2. Hence the answer to Jamison's Question 1.1 is affirmative for trees in $\mathcal{T}_3$ (which do not have vertices of degree 2). For general trees, however, both cases are possible. As already pointed out in [2], many trees have the maximum local mean attained at a leaf. For example, the centre and a leaf of an n-vertex star have local mean

$$\frac{n+1}{2} \qquad \text{and} \qquad \frac{n+2-n/(2^{n-2}+1)}{2}$$

respectively, and the latter is greater for all $n > 3$. On the other hand, there are infinitely many trees for which the maximum local mean occurs at an internal vertex (of degree 2), as the following construction shows:

Consider a tree T_n formed by attaching two pendant edges to each end of the path P_{2n+1} (Figure 1).

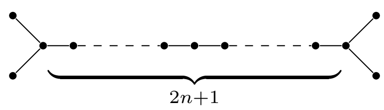

FIGURE 1. General construction of trees with maximum local mean at an internal vertex

The local mean at the i-th internal vertex is

$$(n+4)\left(1 - \frac{4}{(3+i)(2n+5-i)}\right),$$

which reaches its maximum in the middle, at $i = n+1$, where it equals $(n+2)(n+6)/(n+4)$. The local mean at a leaf, on the other hand, is $(4n^2+24n+29)/(4n+9)$, which is smaller for $n \geq 7$. Thus the maximum local mean of T_n occurs at an internal vertex for all $n \geq 7$.

In the proof of Theorem 3.2, it was crucial that subtrees contained in one of the branches T_i are outweighed by subtrees in the union of the other branches. This is not the case if we are considering a leaf (so that there is only one branch) or a vertex of degree 2 with two rather balanced branches, as in the example above, which explains in a way why the maximum local mean can occur at vertices of degree 1 or 2, but not vertices of higher degree.

4. Bounding the local with the global mean

With the help of Theorem 3.2, we can provide a positive answer to Question 1.2, since it will be sufficient to consider the local mean at vertices of degree 1 or 2. Let $\overline{N}_T(v)$ denote the number of subtrees of T not containing v and $\overline{R}_T(v)$ denote the total number of vertices of these trees. Then

$$\overline{\mu}_T(v) = \frac{\overline{R}_T(v)}{\overline{N}_T(v)}$$

is the average order of subtrees not containing v, and the global mean is

$$M_T = \frac{R_T(v) + \overline{R}_T(v)}{N_T(v) + \overline{N}_T(v)}.$$

In analogy to equations (1) and (2), we also have the following recursive formulas for a tree T rooted at v with branches $T_1, T_2, \ldots, T_k$ rooted at $v_1, v_2, \ldots, v_k$ respectively:

$$\overline{N}_T(v) = \sum_{j=1}^{k} (N_j + \overline{N}_j) \tag{7}$$

and

$$\overline{R}_T(v) = \sum_{j=1}^{k} (R_j + \overline{R}_j), \tag{8}$$

where $\overline{N}_i = \overline{N}_{T_i}(v_i)$ and $\overline{R}_i = \overline{R}_{T_i}(v_i)$ are defined in analogy to N_i and R_i. Next we need a simple lemma:

Lemma 4.1. *For any tree T and $v \in V(T)$,*

$$R_T(v) > \overline{N}_T(v).$$

Proof. We provide an injection from the set of subtrees of T not containing v into the set of pairs of a subtree containing v and one of its vertices as follows (Figure 2):

For a subtree S of T that does not contain v, let s be the vertex of S that is closest to v. Then $S \cup P_{vs}$ is a subtree of T containing v (here P_{vs} is the unique path connecting v and s in T). The mapping that maps S to $(S \cup P_{vs}, s)$ is clearly injective, which shows that $R_T(v) \geq \overline{N}_T(v)$.

Furthermore, strict inequality holds as the single-vertex subtree v does not have any preimage under this injection. ∎

Now we are finally able to answer Question 1.2 affirmatively: the key to comparing the global mean of a tree T and the local mean at a vertex v is to estimate the influence of subtrees that do not contain v.

Theorem 4.2. *For any tree T and $v \in V(T)$,*

$$\mu_T(v) < 2M_T. \tag{9}$$

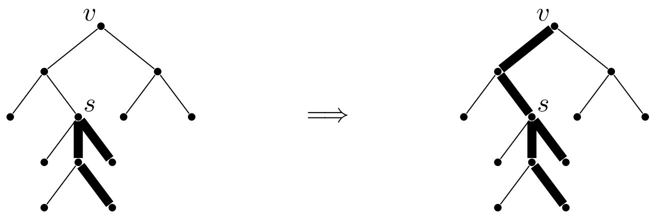

FIGURE 2. The injection

Proof. We proceed by induction on the size of T, the case $|T| = 1$ being trivial. By Theorem 3.2, we only need to consider the cases that v is of degree 1 or 2.

Case i: $\deg(v) = 1$.

Let the neighbor of v be u and $S = T - \{v\}$. For simplicity, we use N, $\overline{N}$, R and $\overline{R}$ to denote $N_S(u)$, $\overline{N}_S(u)$, $R_S(u)$ and $\overline{R}_S(u)$.

Then one easily obtains

$$\mu_T(v) = \frac{1+N+R}{1+N} = 1 + \frac{R}{1+N}$$

and

$$M_T = \frac{2R+\overline{R}+N+1}{2N+\overline{N}+1}.$$

The inequality (9) that we want to prove is equivalent to

$$(2N+\overline{N}+1)(2M_T - \mu_T(v)) > 0,$$

and the left hand side of this inequality simplifies to

$$\begin{aligned} 2(2R+\overline{R}+N+1) &- (2N+\overline{N}+1)\left(1+\frac{R}{1+N}\right) \\ &= 4R+2\overline{R}-\overline{N}+1-R\cdot\frac{2N+\overline{N}+1}{1+N} \\ &= 4R+2\overline{R}-\overline{N}+1-R\left(2+\frac{\overline{N}-1}{N+1}\right) \\ &= 2R+2\overline{R}-\overline{N}+1-\frac{\overline{N}-1}{N+1}R. \end{aligned}$$

Hence we need to prove

$$2(R+\overline{R}) > \overline{N}-1+\frac{\overline{N}-1}{N+1}R.$$

Applying the induction hypothesis (9) to S and u yields

$$2\frac{R+\overline{R}}{N+\overline{N}} > \frac{R}{N},$$

and together with Lemma 4.1, we obtain

$$2(R+\overline{R}) > \frac{R}{N}(N+\overline{N}) = R + \frac{R\overline{N}}{N} > \overline{N} + \frac{R\overline{N}}{N} > \overline{N}-1+\frac{\overline{N}-1}{N+1}R,$$

which completes the proof of (9) in this case.

Case ii: $\deg(v) = 2$.

Similarly, let v_1 and v_2 be the neighbors of v and let T_1, T_2 be the components of $T - \{v\}$ that contain v_1 and v_2 respectively. We also define R_1, R_2, $\overline{R}_1$, $\overline{R}_2$, N_1,

N_2, $\overline{N}_1$, $\overline{N}_2$ accordingly for the parameters of T_1 and T_2 with designated vertices v_1 and v_2.

Then, from (1) and (2), we obtain

$$\mu_T(v) = 1 + \frac{R_1}{1+N_1} + \frac{R_2}{1+N_2}, \tag{10}$$

and in addition we can use (7) and (8) to obtain

$$M_T = \frac{\mu_T(v)(1+N_1)(1+N_2) + R_1 + R_2 + \overline{R}_1 + \overline{R}_2}{(1+N_1)(1+N_2) + N_1 + N_2 + \overline{N}_1 + \overline{N}_2}.$$

Now after some manipulations as in the previous case, we find that (9) is equivalent to

$$\mu_T(v)(1 + N_1N_2 - \overline{N}_1 - \overline{N}_2) + 2(R_1 + \overline{R}_1) + 2(R_2 + \overline{R}_2) > 0. \tag{11}$$

Applying the induction hypothesis (9) to T_1, v_1 and T_2, v_2, we obtain, as in the first case,

$$2(R_1 + \overline{R}_1) > \frac{R_1}{N_1}(N_1 + \overline{N}_1) = R_1 + \frac{R_1\overline{N}_1}{N_1} > \overline{N}_1 + \frac{R_1\overline{N}_1}{N_1+1}$$

and

$$2(R_2 + \overline{R}_2) > \overline{N}_2 + \frac{R_2\overline{N}_2}{N_2+1}.$$

In order to obtain (11), it thus suffices to show

$$\mu_T(v)(1+N_1N_2) > \left(\mu_T(v) - 1 - \frac{R_1}{1+N_1}\right)\overline{N}_1 + \left(\mu_T(v) - 1 - \frac{R_2}{1+N_2}\right)\overline{N}_2.$$

In view of (10), this is equivalent to

$$\left(1 + \frac{R_1}{1+N_1} + \frac{R_2}{1+N_2}\right)(1+N_1N_2) > \frac{R_2\overline{N}_1}{1+N_2} + \frac{R_1\overline{N}_2}{1+N_1}. \tag{12}$$

Now simply note that by Lemma 2.1, Lemma 4.1 and the fact that $N_1 \geq 1$, we have

$$\frac{R_1N_1N_2}{1+N_1} \geq \frac{R_1N_2}{2} > \frac{\overline{N}_1R_2}{1+N_2}$$

and analogously

$$\frac{R_2N_1N_2}{1+N_2} > \frac{\overline{N}_2R_1}{1+N_1},$$

from which (12) follows trivially. ∎

Remark 2. Theorem 4.2 is asymptotically sharp (in the sense that 2 cannot be replaced by a smaller constant). To see this, consider a tree $T = T(n, m)$ consisting of a path P_n of order n and m leaves attached to one of the ends of the path (Figure 3). Then the local mean at the other end (which we denote by v) is

$$\mu_T(v) = \frac{2^{m-1}(2n+m) + n(n-1)/2}{2^m + n - 1},$$

while the global mean is

$$M_T = \frac{2^{m-1}n(m+n+1) + m + (n^3-n)/6}{2^m n + m + n(n-1)/2}.$$

If now $m \geq (1+\epsilon)\log_2 n$ for some fixed $\epsilon > 0$ and at the same time $m = o(n)$, then

$$\mu_T(v) \sim n \qquad \text{and} \qquad M_T \sim \frac{n}{2},$$

so that

$$\frac{\mu_T(v)}{M_T} \to 2$$

as the number of vertices approaches infinity.

This example explains intuitively where the constant 2 comes from: Almost all subtrees contain some of the leaves at the one end. Including the leaf at the other end forces the entire path between them to be included as well. Globally, however, only about half of this path is included on average.

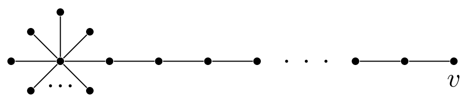

FIGURE 3. An example for the sharpness of (9)

The example that shows why the constant 2 is best-possible involves a long branchless path. Thus it is natural to ask for an improvement of our inequality (9) for trees that are more branched in some sense. Specifically, we will consider $\mathcal{T}_3$, the set of all homeomorphically irreducible trees. The statement

$$M_T \leq \mu_T(v) \leq M_T + \frac{n-1}{2}$$

that was proven in [3] for arbitrary trees can be significantly improved when we restrict our attention to $T \in \mathcal{T}_3$.

Theorem 4.3. *For any tree $T \in \mathcal{T}_3$ and any vertex $v \in V(T)$, we have*

$$M_T \leq \mu_T(v) \leq M_T + \frac{10}{3}. \tag{13}$$

To prove this theorem, we first provide an observation on rooted trees with the property that every internal vertex has at least two children. In this case, the difference between average subtree orders of subtrees containing and not containing the root is bounded by the ratio of the number of corresponding subtrees.

Lemma 4.4. *If T is a tree with root v such that all internal vertices have at least two children, then*

$$\mu_T(v) - \overline{\mu}_T(v) \leq 4\left(1 + \frac{N_T(v)}{\overline{N}_T(v)}\right). \tag{14}$$

Proof. By induction on the order of T: if T only has a single vertex, then the denominator on the right hand side of our inequality is 0, but we can interpret it as

$$\frac{R_T(v)}{N_T(v)}\overline{N}_T(v) - \overline{R}_T(v) \leq 4(\overline{N}_T(v) + N_T(v)),$$

which is true since the left hand side is 0.

For the induction step, let the neighbors of v be $v_1, \ldots, v_k$ and the components of $T-v$ be $T_1, \ldots, T_k$. We also use the abbreviations $N_i, R_i, \ldots$ for $N_{T_i}(v_i), R_{T_i}(v_i), \ldots$ as before. The induction hypothesis gives us $\mu_i\overline{N}_i - \overline{\mu}_i\overline{N}_i \leq 4(\overline{N}_i + N_i)$, which we use now to obtain

$$
\begin{aligned}
\mu_T(v) - \overline{\mu}_T(v) &= 1 + \sum_{i=1}^{k} \mu_i \frac{N_i}{1+N_i} - \frac{\sum_{i=1}^{k}(\mu_i N_i + \overline{\mu}_i \overline{N}_i)}{\sum_{i=1}^{k}(N_i + \overline{N}_i)} \\
&< 1 + \sum_{i=1}^{k} \mu_i - \frac{\sum_{i=1}^{k}(\mu_i N_i + \overline{\mu}_i \overline{N}_i)}{\sum_{i=1}^{k}(N_i + \overline{N}_i)} \\
&= \frac{\overline{N}_T(v) + \left(\sum_{i=1}^{k} \mu_i\right)\left(\sum_{i=1}^{k}(N_i + \overline{N}_i)\right) - \sum_{i=1}^{k}(\mu_i N_i + \overline{\mu}_i \overline{N}_i)}{\overline{N}_T(v)} \\
&= \frac{\overline{N}_T(v) + \sum_{i=1}^{k}\left(\mu_i \sum_{j\neq i}(N_j + \overline{N}_j)\right) + \sum_{i=1}^{k}(\mu_i \overline{N}_i - \overline{\mu}_i \overline{N}_i)}{\overline{N}_T(v)} \\
&\leq \frac{\overline{N}_T(v) + \sum_{i=1}^{k}\left(\mu_i \sum_{j\neq i}(N_j + \overline{N}_j)\right) + 4\sum_{i=1}^{k}(\overline{N}_i + N_i)}{\overline{N}_T(v)} \\
&= \frac{\overline{N}_T(v) + \sum_{i=1}^{k}\left(\mu_i \sum_{j\neq i}(N_j + \overline{N}_j)\right) + 4\overline{N}_T(v)}{\overline{N}_T(v)}.
\end{aligned}
$$

Hence to prove (14), it is sufficient to show

$$\overline{N}_T(v) + \sum_{i=1}^{k}\left(\mu_i \sum_{j\neq i}(N_j + \overline{N}_j)\right) \leq 4N_T(v). \tag{15}$$

By means of an easy induction (which is given explicitly in [7]), one can show that

$$\overline{N}_T(v) \leq N_T(v) \tag{16}$$

for all rooted trees satisfying the given condition. Finally, we need the following result:

Claim 4.5.

$$\prod_{i=1}^{k-1} a_i - \frac{k}{3}\sum_{i=1}^{k-1} a_i \geq 0$$

for any $k \geq 2$ and any sequence of real numbers $a_i \geq 2$.

Proof of Claim 4.5. The coefficient of a_j in

$$\prod_{i=1}^{k-1} a_i - \frac{k}{3}\sum_{i=1}^{k-1} a_i \tag{17}$$

is

$$\prod_{i\neq j} a_i - \frac{k}{3} \geq 2^{k-2} - \frac{k}{3} > 0.$$

Hence (17) is an increasing function with respect to each a_j, and so it is at least

$$2^{k-1} - \frac{k(2k-2)}{3},$$

which is nonnegative for $k \geq 2$. ■

Claim 4.5, in conjunction with Lemma 2.1, which implies $\mu_i \leq (N_i+1)/2$, and (16), which gives us $\overline{N}_j \leq N_j$ for all j, now yields

$$\begin{aligned}\mu_i \sum_{j\neq i}(N_j+\overline{N}_j) &\leq \frac{1+N_i}{2}\sum_{j\neq i}(N_j+\overline{N}_j)\\ &< (1+N_i)\sum_{j\neq i}(1+N_j)\\ &\leq (1+N_i)\cdot\frac{3}{k}\prod_{j\neq i}(1+N_j) = \frac{3}{k}N_T(v),\end{aligned}$$

so that we finally arrive at

$$\sum_{i=1}^{k}\left(\mu_i\sum_{j\neq i}(N_j+\overline{N}_j)\right) \leq 3N_T(v).$$

Combined with (16), this gives us exactly (15). ∎

Proof of Theorem 4.3. The inequality $M_T \leq \mu_T(v)$ has already been proven by Jamison in [2, 3], so we focus on the second inequality. By Theorem 3.2, the maximum $\mu_T(v)$ of $T \in \mathcal{T}_3$ is achieved at a leaf. Hence we only need to show $\mu_T(v) \leq M_T + 10/3$ for a leaf v. Let its only neighbor be u, set $S = T - v$, and use the same abbreviations as in the first case of the proof of Theorem 4.2. Note that S (with root u) satisfies the conditions of Lemma 4.4. Now (13) can be written as

$$\frac{1+N+R}{1+N} \leq \frac{2R+\overline{R}+N+1}{2N+\overline{N}+1}+\frac{10}{3},$$

which in turn is equivalent to

$$\frac{N^2+N\overline{N}+(R\overline{N}-N\overline{R})+N+\overline{N}-R-\overline{R}}{(N+1)(2N+\overline{N}+1)} \leq \frac{10}{3}.$$

Now Lemma 4.4 gives us $R\overline{N}-N\overline{R} \leq 4N(N+\overline{N})$, so we are done if

$$\frac{N^2+N\overline{N}+4N(N+\overline{N})+N+\overline{N}-R-\overline{R}}{(N+1)(2N+\overline{N}+1)} \leq \frac{10}{3}$$

holds. Simplifying this inequality shows that it is equivalent to

$$\frac{5N(N-\overline{N})+27N+7\overline{N}+3R+3\overline{R}+10}{3(N+1)(2N+\overline{N}+1)} \geq 0,$$

which certainly holds true in view of (16). This completes our proof. ∎

Remark 3. The constant 10/3 is most certainly not best possible, and neither is 4 in Lemma 4.4. By more careful (but probably also more complicated) estimates, it should be possible to reduce it further. The best possible constant might be 5/3: if one considers the caterpillar constructed by attaching an additional leaf to each of the interior vertices of a path, then the difference between global mean and local mean at one of the ends tends to 5/3 as the order goes to infinity. The main point of our theorem, however, is not so much the exact value of the constant, but rather the fact that the local and global mean differ at most by a fixed constant for trees in $\mathcal{T}_3$.

References

[1] G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1988. xii+324 pp.

[2] R. Jamison, On the Average Number of Nodes in a Subtree of a Tree, *J. Combin. Theory Ser. B* 35 (1983), 207–223.

[3] R. Jamison, Monotonicity of the Mean Order of Subtrees, *J. Combin. Theory Ser. B* 37 (1984), 70–78.

[4] S. Klavzar, M. Petkovsek, J. H. Steelman, Problems and solutions, solutions of elementary problems, E3281. *Amer. Math. Monthly* 97 (1990), 924–925.

[5] A. Meir, J. W. Moon, On subtrees of certain families of rooted trees, *Ars Combin.* 16 (1983), 305–318.

[6] W. T. Trotter, J. L. Moore, Some theorems on graphs and posets. *Discrete Math.* 15 (1976), 79–84.

[7] A. Vince, H. Wang, The average order of a subtree of a tree, *J. Combin. Theory Ser. B* 100 (2010), 161–170.

Stephan Wagner, Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

E-mail address: `swagner@sun.ac.za`

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

E-mail address: `hwang@georgiasouthern.edu`