Extrema of Local Mean and Local Density in a Tree

Ruoyu Wang

Submitted: Jan 27, 2025; Accepted: Mar 2, 2026; Published: Mar 27, 2026
© The author. Released under the CC BY license (International 4.0).

Abstract

Given a tree T and a subtree S of T, one can define the local mean at S, $\mu_T(S)$, to be the average order of the subtrees of T containing S. In 1983, Jamison showed that $\mu_T(S) < \mu_T(S')$ if $S \subset S'$ as subtrees of T. Therefore, it is natural to ask the following question. Among all the k-subtrees (subtrees of order k), which one achieves the maximal/minimal local mean and what properties does it have? We call such k-subtrees k-maximal/k-minimal. Wagner and H. Wang showed in 2016 that a 1-maximal subtree has degree 1 or 2. In this paper, we show that if T is not a path, a 1-minimal subtree of T has degree at least 3. For $k \geqslant 2$, we show that a k-maximal subtree has at most one leaf whose degree in T is greater than 2, and that such a leaf can only occur when all other leaves in S are also leaves in T. Parallel results hold for k-minimal subtrees. Roughly speaking, the leaves of a k-maximal subtree tend to have degree 1 or 2 in T, while the leaves of a k-minimal subtree tend to have degree at least 3 in T.

In the second part, this paper introduces the local density as a normalization of local means, for the sake of comparing subtrees of different orders. We show that the local density at subtree S is lower-bounded by $1/2$ with equality if and only if S contains all the vertices of degree at least 3 in T. On the other hand, local density can be arbitrarily close to 1.

Mathematics Subject Classifications: 05C05, 05C30

1 Introduction

In the early 1980s, Jamison [1, 2] initiated the study of the mean order of subtrees and density of a tree T. There are two types of means studied therein. The first is the *global mean*, μ_T, which is the average order of all subtrees of T. The second type is the *local mean at a subtree* S of T, $\mu_T(S)$, which equals the average order of all subtrees of T containing S. Local means can be compared among all the subtrees of the same order. Denote the order of T by $|T|$. Throughout this paper, a subtree S is called *k-maximal*

Department of Mathematics, Uppsala University, Sweden (ruoyu.wang@math.uu.se).

 https://doi.org/10.37236/13813

if $|S| = k$ and $\mu_T(S) = \max\{\mu_T(U) \mid U$ is a subtree of T and $|U| = k\}$, and *k-minimal* subtrees are defined analogously.

Jamison [1, 2] established fundamental results for subsequent research in this area, proving among other results that:

- the global mean is no greater than the local mean at a vertex,
- the local mean respects inclusion, i.e., for subtrees $S \subset S'$, the local mean at S is less than the local mean at S',
- the local mean at S is lower-bounded by $(|S| + |T|)/2$.

Several conjectures and open questions were raised in these two papers, most of which were solved in the 2010s. In 2010, Vince and H. Wang [4] showed that for series-reduced trees (i.e., trees with no vertex of degree 2), the local mean is at least half of $|T|$ and less than $\frac{3}{4}$ of $|T|$. Later on, in 2014, Haslegrave [7] provided necessary and sufficient conditions for the local mean tending to each of the two bounds in series-reduced trees. Wagner and H. Wang [3] studied the ratio of local mean (at a vertex) to global mean and showed that it is bounded above by 2.

The latest problem that has been solved [2, (5.6)] is the conjecture that the decrease in global mean under the contraction of an edge is at least $\frac{1}{3}$. Luo, Xu, Wagner and H. Wang [6] proved the special case when the edge contains a leaf, while the general case was settled by R. Wang [9] in 2024. The only question that still remains open on the list [1, 2] is the one that asks whether the tree(s) of maximal global mean of any given order n is a caterpillar. Jamison [1, (7.1)] conjectured a positive answer to this question, known as the caterpillar conjecture. Attempts and progress [8, 5] were made on this question, but a final answer is still unknown.

This paper mainly concerns the extremal local mean in a tree. The idea originated from the following question [1, (7.5)]:

- For any tree T, is a 1-maximal subtree necessarily a leaf?

This question itself is solved by Wagner and H. Wang [3]. They showed that a 1-maximal subtree is a vertex whose degree is either 1 or 2, and that both cases are possible. Herein, we provide a simpler proof of this result as part of Theorem 2. Moreover, we ask the following general question for any $k \in \{1, \ldots, n\}$:

- If S is a k-maximal/k-minimal subtree of T, what can one say about S?

The first part of this article answers the above question through the following results.

For a finite tree T, we define the *body* of T, denoted by T^*, to be the smallest subtree that contains every vertex v of T with $\deg_T(v) \geqslant 3$, or to be $\emptyset$ if no such vertex exists, i.e., if T is a path.

Theorem 1. *Let T be a tree that is not a path and T^* its body. Let S be a k-minimal subtree of T with $k \leqslant |T^*|$. Then $S \subseteq T^*$ and every vertex of S has degree at least* 2 *in T.*

Furthermore, if $k \geqslant 2$*, then* S *contains at most one leaf of degree* 2 *in* T*, and all other leaves of* S *have degree at least* 3 *in* T*. If* $k = 1$*, then the vertex of* S *has degree at least* 3 *in* T*.*

Theorem 2. *Let* S *be a* k*-maximal subtree of* T*. If* $k \geqslant 2$*, then* S *contains at most one leaf of degree at least* 3 *in* T*, and all other leaves of* S *have degree* 1 *or* 2 *in* T*. If* $k = 1$*, then the vertex of* S *has degree* 1 *or* 2 *in* T*.*

Theorem 3. *Let* S *be a* k*-maximal subtree of* T*. If there exists a leaf of* S *of degree at least* 3 *in* T *then all other leaves of* S *have degree* 1 *in* T*.*

To facilitate the proof, we introduce a new notion, the *index*, which reflects the change of local mean in the sense that when a new neighbour is added to a subtree, its local mean increases by the index of that neighbour and when a leaf is removed from a subtree, its local mean decreases by the index of that leaf. The index not only greatly simplifies some proofs but also provides insight into the behaviour of local mean and density.

In the last part of this paper, we introduce and study the *local density* at a proper subtree S of T, denoted by $D_T(S)$, for the sake of normalizing the local means, so that one can compare subtrees of T at different orders. Hence it is valid to ask which subtree of T attains the maximal/minimal local density. Jamison [1] defined the *density* of a tree T, D_T, as the quotient of the global mean μ_T to the order $|T|$. Just as the density of a tree can be interpreted as the probability of a random vertex being contained in a random subtree, the local density at subtree S represents the probability of a random vertex *outside* S being contained in a random subtree *containing* S. It will be shown that local density can be arbitrarily close to 1. On the other hand, the lower bound is characterized in the following theorem.

Theorem 4. *Let* S *be a proper subtree of* T *and* T^* *be the body of* T*. Then* $D_T(S) \geqslant \frac{1}{2}$*, with equality if and only if* $S \supseteq T^*$*.*

2 Notations and Preliminaries

Let $T = (V, E)$ be a tree, where $V \neq \emptyset$ is the set of vertices and E is the set of edges. A subtree S of T, denoted by $S \subseteq T$, is a connected subgraph $S = (V', E')$ where $\emptyset \neq V' \subseteq V$ and $E' \subseteq E$. To specify a vertex, we will use notations such as $v \in T$ and $v \in T - S$ to indicate $v \in V$ and $v \in V \backslash V'$, respectively, without referring to the sets of vertices. Denote the edge joining adjacent vertices v and w by vw. For $S \subseteq T$ and $v \in S$, let $\deg_S(v)$ be the degree of v in S and $\deg_T(v)$ the degree of v in T. Denote the set of leaves of T, $\{v \in T | \deg_T(v) = 1\}$, by $L(T)$. If vertex $w \in T$ lies outside of subtree S and is adjacent to some vertex in S, then we say that w is a neighbour of S or w is adjacent to S, and denote the subtree formed by including w and the induced edge $S + w$. Similarly, the subtree obtained by removing a leaf v of S is denoted by $S - v$. Following the conventions by Mol and Oellermann [5], for a non-path T, we call a maximal path in T comprised of one leaf and vertices of degree 2 a *limb*. Note that for such T, the remaining part of the tree is exactly T^*, the body of T.

Let $S \subseteq T$, and $v \in (T - S) \cup L(S)$, i.e., v is either a vertex that lies outside of S or is a leaf of S. We define two subtrees associated to S and v. Let S_v be the minimal subtree in $\{U \subseteq T \mid S \subseteq U, v \in U\}$ and $T_{v,S}$ be the maximal subtree in $\{U \subseteq T \mid U \cap S_v = \{v\}\}$, where minimal and maximal are defined with respect to inclusion on sets of vertices. Intuitively, S_v is the smallest subtree containing both S and v, and $T_{v,S}$ is the branch at v that "grows away" from S. In particular, in the case of a singleton subtree S, we have $L(S) = \emptyset$, therefore S_v and $T_{v,S}$ are only well-defined for $v \in T - S$. Furthermore, we will use the following notation.

$$\begin{aligned}
N_T: &\quad \text{number of subtrees of } T,\\
R_T: &\quad \text{total order of all subtrees of } T,\\
N_T(S): &\quad \text{number of subtrees containing } S,\\
R_T(S): &\quad \text{total order of all subtrees containing } S,\\
N_T(v;S): &\quad \text{number of subtrees of } T_{v,S} \text{ containing } v, \text{ which equals } N_{T_{v,S}}(v),\\
R_T(v;S): &\quad \text{total order of all subtrees counted by } N_T(v;S).
\end{aligned}$$

Note that the last two quantities are only defined for $v \in (T - S) \cup L(S)$.

With the notations above, we write $\mu_T(S) = R_T(S)/N_T(S)$, $\mu_T = R_T/N_T$ and $D_T = \mu_T/|T|$. For the sake of simplicity, when S has order 1 or 2, say $V(S) = \{v\}$ or $V(S) = \{v, w\}$, we will slightly abuse notation by writing $N_T(v)$ or $N_T(v, w)$, respectively, instead of $N_T(S)$. Similarly, the same applies to μ_T and R_T.

Next, we introduce some preliminary results which will be used throughout this article.

Let $v_1, \ldots, v_d$ be the neighbours of v in T, and $T_1, \ldots, T_d$ be the corresponding components of $T - v$. Write $N_i = N_{T_i}(v_i)$ and $R_i = R_{T_i}(v_i)$, for all $i \in \{1, \ldots, d\}$.

First, one has

$$N_T(v) = \prod_{i=1}^{d} (N_i + 1). \tag{1}$$

Indeed, to form a subtree containing v, one simply chooses, for each i, any subtree containing v_i from T_i (there are N_i ways to complete this step) or leaves it empty (contributing the $+1$).

Next,

$$\mu_T(v) = 1 + \sum_{i=1}^{d} \frac{R_i}{N_i + 1}. \tag{2}$$

This equality reflects the following fact: the local mean at v can be calculated by 1, which corresponds to the vertex v, plus the adjusted local mean (adjusted to include the empty set, with the $+1$ in the denominator) at each v_i of T_i. Indeed, any subtree containing v can be decomposed as the union of v, and one subtree containing v_i or the empty set for each branch T_i. A more detailed proof can be found in Jamison's forementioned articles [1, Lemma 3.2 (b)] [2, Lemma 1].

One has the following lower bound for the local mean at v [1, Theorem 3.6],

$$\mu_T(v) \geqslant \frac{|T| + 1}{2}, \tag{3}$$

with equality if and only if T is astral over v (T is either a path or v is the only vertex of degree greater than 2 in T).

Now, let v, w be two adjacent vertices of T. Denote by T_v and T_w the two components of $T - vw$ containing v and w respectively, as in Figure 1.

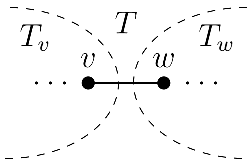

Figure 1: T_v and T_w

Then we have
$$R_T(v,w) = N_{T_w}(w)R_{T_v}(v) + R_{T_w}(w)N_{T_v}(v). \tag{4}$$
To see this, we first observe that $N_T(v,w) = N_{T_w}(w)N_{T_v}(v)$. Dividing by $N_T(v,w)$ on both sides of the equation, (4) is equivalent to
$$\mu_T(v,w) = \mu_{T_v}(v) + \mu_{T_w}(w).$$
This equation holds as any subtree of T containing v and w can be decomposed as the union of a subtree of T_v containing v and a subtree of T_w containing w.

The last observations are concerned with contraction with respect to a subtree. For a tree T and a subtree S, let T/S be the new tree obtained by collapsing S to a single vertex s and keeping all the adjacency relations in the natural way: a vertex is adjacent to s in T/S if and only if it is adjacent to S in T.

Now let S be a subtree of T and $U \subseteq S$ a subtree of S. Then one has a natural bijection induced by contraction with respect to U :
$$\{\text{subtrees of } T \text{ containing } S\} \leftrightarrow \{\text{subtrees of } T/U \text{ containing } S/U\}.$$
Furthermore, through the same bijection, one sees that each subtree of T in the left set has an order greater than the corresponding subtree of T/U in the right set by exactly $|U| - 1$. Hence, the same applies to the local means, namely
$$\mu_T(S) = \mu_{T/U}(S/U) + |U| - 1. \tag{5}$$
This equality was first proved by Jamison [1, (4.4)].

3 Auxiliary Results: Index

In this section, we define the index and prove three lemmas on the index.

Definition 5. Let $v \in T$. Define the *index* of v in T to be

$$i_T(v) := \frac{R_T(v)}{N_T(v)(1 + N_T(v))}.$$

Furthermore, let $S \subseteq T$. For any vertex $v \in (T - S) \cup L(S)$, define the *index* of v in T with respect to S to be

$$i_T(v; S) := \frac{R_T(v; S)}{N_T(v; S)(1 + N_T(v; S))}. \tag{6}$$

Note that $i_T(v; S) = i_{T_{v,S}}(v)$, the index of v in $T_{v,S}$.

Throughout this paper, we will omit the reference to T when it is safe from ambiguity and write $i(v)$ and $i(v; S)$, respectively.

Let v be a vertex in T. We have the inequality

$$i_T(v) \leqslant \frac{1}{2}, \tag{7}$$

with equality if and only if T is a path and v is an endpoint. This is equivalent to the upper bound on the difference between the local mean and the adjusted local mean established by Jamison [1, Lemma 3.2c]. Furthermore, if w is a neighbour of v in T and $S = \{w\}$, we have

$$i_T(v; w) = i_{T_{v,S}}(v) \leqslant \frac{1}{2}, \tag{8}$$

with equality if and only if $T_{v,S}$ is a path, as in Figure 2. Note that v is a limb vertex, i.e., v belongs to a limb of T, if and only if there exists a neighbour w of v such that $i(v; w) = \frac{1}{2}$.

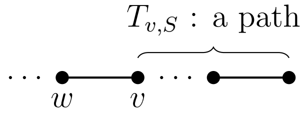

Figure 2: $i(v; w) = \frac{1}{2}$.

Now we are ready to state and prove the Index Lemma.

Lemma 6 (Index Lemma)**.** *Let S be a subtree of T. If w is a neighbour of S, then*

$$\mu_T(S + w) = \mu_T(S) + i(w; S).$$

Equivalently, if v is a leaf of subtree S, then

$$\mu_T(S - v) = \mu_T(S) - i(v; S).$$

Proof. By (5), one can always contract S and it suffices to show the case when S is a single vertex, say $S = \{v\}$. Let w be a neighbour of v, and T_v, T_w be the components of $T - vw$ that contain v, w respectively. We need to show

$$\mu_T(v,w) - \mu_T(v) = i(w;v),$$

which is equivalent to

$$\frac{R_T(v,w)}{N_T(v,w)} - \frac{R_T(v)}{N_T(v)} = i(w;v).$$

Note that any subtree of T containing v is either a subtree of T_v containing v or a subtree of T containing both v and w. Hence, $N_T(v) = N_{T_v}(v)\left(N_{T_w}(w)+1\right)$ and $R_T(v) = R_{T_v}(v) + R_T(v,w)$. Then using (4), one has

$$\begin{aligned}\frac{R_T(v,w)}{N_T(v,w)} - \frac{R_T(v)}{N_T(v)} &= \frac{N_{T_v}(v)R_{T_w}(w) + R_{T_v}(v)N_{T_w}(w)}{N_{T_v}(v)N_{T_w}(w)} \\ &\quad - \frac{R_{T_v}(v) + \left(N_{T_v}(v)R_{T_w}(w) + R_{T_v}(v)N_{T_w}(w)\right)}{N_{T_v}(v)\left(N_{T_w}(w)+1\right)} \\ &= \frac{N_{T_v}(v)R_{T_w}(w)}{N_{T_v}(v)N_{T_w}(w)\left(N_{T_w}(w)+1\right)} \\ &= \frac{R_{T_w}(w)}{N_{T_w}(w)\left(N_{T_w}(w)+1\right)} \\ &= i(w;v).\end{aligned}$$

□

Remark 7. Since the index is a positive number whose value lies in $(0, \frac{1}{2}]$, the previous lemma immediately implies

$$\mu_T(v) < \mu_T(v,w) \leqslant \mu_T(v) + \frac{1}{2},$$

for any adjacent vertices v, w in T ([1, Lemma 4.1]), and other 'continuity' results [1, Theorem 4.3 and 4.5].

Lemma 8. *Let S be a proper subtree of T and $v \in (T - S) \cup L(S)$. Let $v_1, \ldots, v_d$ be the neighbours of v in $T_{v,S}$ and $d \geqslant 2$, as in Figure 3b. Then $i(v;S) < i(v_j;S)$ for any $j \in \{1, \ldots, d\}$.*

Proof. Fix $j \in \{1, \ldots, d\}$. Let $R_i := R(v_i;S)$ and $N_i := N(v_i;S)$ for all $i \in \{1, \ldots, d\}$. The claim $i(v;S) < i(v_j;S)$ is, by (1) and (2), equivalent to

$$\frac{1 + \sum_{i=1}^{d} \frac{R_i}{N_i+1}}{\prod_{i=1}^{d}(N_i+1) + 1} < \frac{R_j}{N_j(N_j+1)}. \tag{9}$$

To start, we consider the case that the component in $T - v$ at v_j is a limb of T. In this case, the right hand side of (9) is $\frac{1}{2}$, which is the maximal value of the index. As the maximal value of the index only occurs for a path, the condition $d \geqslant 2$ excludes the possibility of it being achieved by $i(v;S)$. Therefore, (9) holds for this case.

With the case above excluded, we claim $R_j/N_j \geqslant 2$. Indeed, among all the subtrees that contain v_j but not v, we have one singleton subtree $\{v_j\}$, at least one subtree has order 3, and all other such subtrees have order at least 2. Hence,

$$R_j \geqslant 1 + 3 + 2(N_j - 2) = 2N_j.$$

With this inequality established, we now relax the left hand side of (9),

$$\text{LHS} < \frac{1 + \sum \frac{R_i}{N_i+1}}{\prod (N_i+1)} = \frac{1 + \frac{R_j}{N_j+1} + \sum_{i\neq j} \frac{R_i}{N_i+1}}{\prod (N_i+1)} \leqslant \frac{1 + \frac{R_j}{N_j+1} + \sum_{i\neq j} \frac{N_i}{2}}{\prod (N_i+1)},$$

where the summations and multiplications are taken over $i = 1, \ldots, d$, and the last inequality uses (8). Therefore, it suffices to prove

$$\begin{aligned}
& \frac{1 + \frac{R_j}{N_j+1} + \sum_{i\neq j} \frac{N_i}{2}}{\prod (N_i+1)} \leqslant \frac{R_j}{N_j(N_j+1)} \\
\Leftrightarrow \quad & 1 + \frac{R_j}{N_j+1} + \sum_{i\neq j} \frac{N_i}{2} \leqslant \frac{R_j}{N_j} \prod_{i\neq j} (N_i+1) \\
\Leftrightarrow \quad & 1 + \frac{1}{2}\sum_{i\neq j} N_i \leqslant \frac{R_j}{N_j} - \frac{R_j}{N_j+1} + \frac{R_j}{N_j}\left(\sum_{i\neq j} N_i + [\text{higher order terms}]\right) \\
\Leftrightarrow \quad & 1 \leqslant i(v_j;S) + \left(\frac{R_j}{N_j} - \frac{1}{2}\right)\sum_{i\neq j} N_i + \frac{R_j}{N_j}\cdot [\text{higher order terms}].
\end{aligned}$$

Since $R_j/N_j \geqslant 2$ and $d \geqslant 2$ ensures the existence of at least one $i \neq j$, we have

$$\left(\frac{R_j}{N_j} - \frac{1}{2}\right) N_i \geqslant \frac{3}{2}\cdot 1 > 1,$$

and the main inequality follows. □

Lemma 9. *With notations as above, if* $d = 1$*, i.e.,* v *has only one neighbour* w *in* $T_{v,S}$*, as in Figure 3a, then* $i(v;S) \geqslant i(w;S)$. *The equality holds if and only if* $i(v;S) = i(w;S) = \frac{1}{2}$.

Proof. Let $N := N(w;S)$ and $R := R(w;S)$. By reversing the inequality in (9) and

(a) Only one neighbour of v in $T_{v,S}$: $i(v;S) \geqslant i(w;S)$

(b) Several neighbours of v in $T_{v,S}$: $i(v;S) < i(v_j;S)$

Figure 3: Change of indices with respect to S.

setting $d = 1$, we see that $i\,(v;S) \geqslant i\,(w;S)$ is equivalent to

$$\begin{aligned}
& & \frac{1+\frac{R}{N+1}}{(N+1)+1} &\geqslant \frac{R}{N\,(N+1)} \\
&\Leftrightarrow & \frac{R+N+1}{N+2} &\geqslant \frac{R}{N} \\
&\Leftrightarrow & RN+N^2+N-RN-2R &\geqslant 0 \\
&\Leftrightarrow & N^2+N-2R &\geqslant 0 \\
&\Leftrightarrow & \frac{1}{2} &\geqslant \frac{R}{N\,(N+1)} \\
&\Leftrightarrow & \frac{1}{2} &\geqslant i\,(w;S)\,.
\end{aligned}$$

The claim follows by (7). □

4 Extrema of Local Means

We start the section by proving Theorem 1.

Proof of Theorem 1. Let S be k-minimal in T. First, we show that $S \subseteq T^*$. Assume that S contains leaves that lie outside of T^*, i.e., in some limb of T. It is not hard to see that there exists at least one such leaf v with $i(v;S) = \frac{1}{2}$. Note that $k \leqslant |T^*|$ implies that the rest of S cannot contain the entire body, which further implies, by the non-emptiness of T^*, that there exists a neighbour w of S such that the component of w in $T - S$ contains at least one vertex whose degree in T is at least 3. Therefore, $i(w;S) < \frac{1}{2}$. Thus one can include the neighbour of w outside of S and delete v, resulting in a subtree of the same order but lower local mean, which leads to a contradiction. So, $S \subseteq T^*$ and all vertices of S have degree at least 2.

Now consider $k \geqslant 2$. Assume, for the purpose of contradiction, that S has two leaves v and w of degree 2, and without loss of generality, $i\,(v;S) \leqslant i\,(w;S) < \frac{1}{2}$. Let v' be the only neighbour of v in $T_{v,S}$. By Lemma 9, we know that v' has a lower index than v, hence lower than the index of w. Including v' and deleting w then results in a subtree of order k but with local mean lower than S, and the desired contradiction follows. Therefore, for $k \geqslant 2$, S has at most one leaf of degree 2. On the other hand, as $k \geqslant 2$, the previous argument immediately implies that S has at least one leaf of degree at least 3, since S has at least two leaves and as $S \subseteq T^*$ each leaf of S has degree at least 2 in T.

For $k = 1$, let $S = \{a\}$. Since we already know $\deg_T(a) \geqslant 2$, assume $\deg_T(a) = 2$. Let v, w be the two neighbours of a. Without loss of generality, let $i(v; a) \geqslant i(w; a)$. Observe that both indices are strictly less than $\frac{1}{2}$ since $a \in T^*$. First, we include w and obtain a 2-subtree $\{a, w\}$ with local mean $\mu_T(a) + i(w; a)$. Then we delete a and obtain a singleton subtree $\{w\}$ with local mean $\mu_T(a) + i(w; a) - i(a; w)$. However, $i(a; w) > i(v; w) = i(v; a)$ by Lemma 9. Hence w has a lower local mean than a, which contradicts that $S = \{a\}$ is 1-minimal. □

Note that the result above does not apply to paths. However, it follows immediately from the Index Lemma that local means of subtrees of the same order are equal in a path.

Briefly, Theorem 1 says that a subtree that minimizes the local mean tends to have its leaves located at vertices of degree at least 3. However, it may depend on the order of subtrees k and the order of the body of T. When k is large enough, all k-subtrees, including the minimizing one(s), may have to contain more than one vertex of degree 1 or 2. For example, in the case of a star with n leaves and $k > 2$, any k-subtree has $k - 1$ leaves.

Next, we prove Theorem 2 and 3 which describe the subtrees of maximal local mean. In the case of a singleton subtree, i.e. $k = 1$, Theorem 2 provides an easier proof to the fact that the vertex of the greatest local mean has degree 1 or 2, which was first proved by Wagner and H. Wang [3, Theorem 3.2].

Proof of Theorem 2. When $k \geqslant 2$, S has at least two leaves. Assume that S has at least two leaves v and w of degree at least 3. Without loss of generality, let $i(v; S) \geqslant i(w; S)$. Since v has at least two neighbours outside of S, Lemma 8 applies and all neighbours of v in $T_{v,S}$ have higher indices than that of v, hence higher than that of w. Then Lemma 6 implies that including any of v's neighbours and excluding w increases the local mean, contradicting that S is k-maximal.

When $k = 1$, let $S = \{v\}$ and assume, for the purpose of contradiction, that v has at least three neighbours a, b and c. Without loss of generality, let a have the highest index with respect to v, i.e. $i(a; v) \geqslant i(b; v)$ and $i(a; v) \geqslant i(c; v)$. Observe that $i(b; v) = i(b; a)$, which implies $i(a; v) \geqslant i(b; a)$. As v has at least two neighbours b and c in the component of v in $T - a$, Lemma 8 applies and yields $i(b; a) > i(v; a)$. Then $i(a; v) > i(v; a)$ and by Lemma 6, $\mu_T(a) = \mu_T(a, v) - i(v; a) = \mu_T(v) + i(a; v) - i(v; a) > \mu_T(v)$, which contradicts the assumption that S is 1-maximal. □

Proof of Theorem 3. Let w be a leaf of S and $\deg_T(w) \geqslant 3$. Let $w_1, \ldots, w_d$ be the neighbours of w outside of S. Note that $d \geqslant 2$ since $\deg_T(w) \geqslant 3$. Denote their corresponding components in $T - S$ by $T_j, j = 1, \ldots, d$. Assume, for the purpose of contradiction, that $v \neq w$ is a leaf of S with $\deg_T(v) \geqslant 2$. By Theorem 2, we know $\deg_T(v) = 2$. Denote the only neighbour of v outside of S by v_1, and its corresponding component in $T - S$ by C_1, as shown in Figure 4. Let $\bar{\mu}_{T_j}(w_j)$ be the adjusted local mean of T_j at w_j, namely

$$\bar{\mu}_{T_j}(w_j) = \frac{R(w_j; S)}{N(w_j; S) + 1}.$$

Note that $i(w_j;S) = \bar{\mu}_{T_j}(w_j)/N(w_j;S)$. By (1) and (2),

$$i(w;S) = \frac{\mu_{T_{w,S}}(w)}{1+N(w;S)} = \frac{1+\sum_{j=1}^{d}\bar{\mu}_{T_j}(w_j)}{1+\prod_{j=1}^{d}(N(w_j;S)+1)}.$$

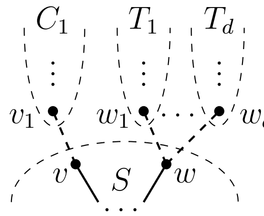

Figure 4: $\deg_T(v) = 2$ and $\deg_T(w) > 2$.

First, observe that C_1 cannot be a path. Indeed, if C_1 is a path, then $i(v_1;S) = \frac{1}{2}$, however, $d \geqslant 2$ implies $i(w;S) < \frac{1}{2} = i(v_1;S)$, which contradicts the assumption that S is k-maximal. We can further exclude the case that T_j is a path, since in that case $i(w_j;S) = \frac{1}{2} > i(v;S)$ and we have another contradiction. Now, as S is k-maximal, by the Index Lemma, we have $i(v;S) \geqslant i(w_j;S)$, $j = 1,\ldots,d$. Since C_1 is not a path, one has $|C_1| \geqslant 3$ and $N(v_1;S) \geqslant 4$, and (3) further implies $R(v_1;S)/N(v_1;S) \geqslant (|C_1|+1)/2 \geqslant 2$. Combining these inequalities together, we have

$$\begin{aligned}
\frac{i(v;S)}{i(v_1;S)} &= \frac{R(v_1;S)+N(v_1;S)+1}{R(v_1;S)} \cdot \frac{N(v_1;S)}{N(v_1;S)+2} \\
&< \frac{R(v_1;S)+N(v_1;S)+1}{R(v_1;S)} \\
&= 1 + \frac{N(v_1;S)+1}{R(v_1;S)} \\
&\leqslant 1 + \frac{N(v_1;S)+1}{2N(v_1;S)} \\
&\leqslant 1 + \frac{1}{2} + \frac{1}{8} = \frac{13}{8}.
\end{aligned}$$

Next, let k be an index such that $\bar{\mu}_{T_k}(w_k) \geqslant \bar{\mu}_{T_j}(w_j), \forall j \in \{1,\ldots,d\}$. Since T_j is not a path, we have $\bar{\mu}_{T_k}(w_k) \geqslant \bar{\mu}_{T_j}(w_j) > 1$ and $N(w_j;S) \geqslant 4$. Thus,

$$\begin{aligned}\frac{i\left(w_k;S\right)}{i\left(w;S\right)} &= \frac{\bar{\mu}_{T_k}(w_k)}{1+\sum_{j=1}^{d}\bar{\mu}_{T_j}(w_j)} \cdot \frac{1+\prod_{j=1}^{d}\left(N(w_j;S)+1\right)}{N(w_k;S)} \\ &\geqslant \frac{1}{d+1} \cdot \frac{1+\prod_{j=1}^{d}\left(N(w_j;S)+1\right)}{N(w_k;S)} \\ &> \frac{\prod_{j=1}^{d}\left(N(w_j;S)+1\right)}{(d+1)\,N(w_k;S)} \\ &> \frac{\prod_{j=1,j\neq k}^{d}\left(N(w_j;S)+1\right)}{d+1} \geqslant \frac{5}{3}.\end{aligned}$$

Combining the above two inequalities on ratios of indices and $i\left(v;S\right) \geqslant i\left(w_k;S\right)$, we get

$$\frac{i\left(v;S\right)}{i\left(v_1;S\right)} < \frac{13}{8} < \frac{5}{3} < \frac{i\left(w_k;S\right)}{i\left(w;S\right)} \Rightarrow i\left(v_1;S\right) > i\left(w;S\right),$$

which contradicts S being k-maximal. □

Some examples of extremal local means are given below. In Figure 5, T is constructed by connecting the centres of two stars $K_{1,n}$ by a path of $(k-2)$ vertices, with $k = 2, 3, \ldots$. There are two 2-minimal subtrees in T : the two edges at the two ends of the path in the middle, cd and $c'd'$. To see this, first observe that a 2-subtree with a leaf, say v_1c, has a higher local mean than dc, as $i\left(v_1;c\right) = \frac{1}{2} > i\left(d;c\right)$. Then, for any edge vv' in the path cc', direct calculation yields the local mean

$$\mu\left(v,v'\right) = \frac{k}{2} + \frac{(n+a)\,2^{n-1}}{2^n+a-1} + \frac{(n+b)\,2^{n-1}}{2^n+b-1},$$

where a and b are the numbers of vertices in the paths cv and $v'c'$, respectively. Therefore, $1 \leqslant a, b \leqslant k-1$ and $a+b=k$. It can be shown that $\mu\left(v,v'\right)$ attains its minimum either at $a=1, b=k-1$ or $a=k-1, b=1$, and attains its maximum in the middle of the path.

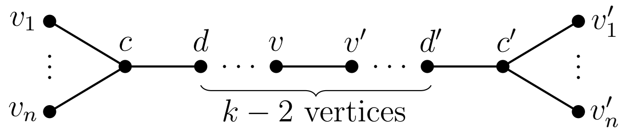

Figure 5: Two 2-minimal subtrees cd and $c'd'$.

In Figure 6, we connect the centres of two stars $K_{1,2}$ by a path of $2m$ vertices, $m = 1, 2, \ldots$. Let vw be the edge in the middle of the path and cd be any edge containing a leaf. Then cd is 2-maximal when $m = 1$, and vw is 2-maximal when $m > 1$. To see this, first we observe that, from the last example, vw maximizes the local mean among all 2-subtrees on the path. Hence, one only needs to compare $\mu(v,w)$ and $\mu(c,d)$. Then, by direct calculation and results from Wagner and H. Wang [3], we have

$$\mu(v,w) = 1 + \frac{(m+2)(m+6)}{m+4}, \ \mu(c,d) = \frac{4m^2+28m+41}{4m+10}.$$

The conclusion follows readily.

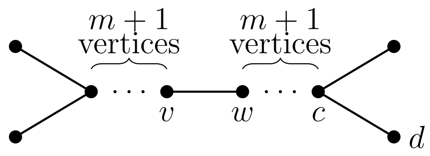

Figure 6: $m = 1$: cd is 2-maximal; $m > 1$: vw is 2-maximal.

In the last part of this section, we give two immediate refinements of the description of k-maximal subtrees.

Corollary 10. *Let T be a tree that is not a path. Let $v \in T$ be a vertex of maximal local mean. Then v is either a leaf, or it lies in a path inside T^*.*

Proof. We only need to eliminate the possibility of v being a non-leaf limb vertex. Suppose v lies in a limb and v is not a leaf. Let w be the leaf in the limb. Denote the neighbour of v in path vw by u. Observe that $T - v$ contains two components. The component containing u is the path uw, while the other is not a path, as T is not a path. So we have $i(u;v) = 1/2$ and $i(v;u) < 1/2$. Then by the Index Lemma,

$$\mu(u) = \mu(v,u) - i(v;u) > \mu(v,u) - i(u;v) = \mu(v),$$

which contradicts the assumption that v achieves maximal local mean. □

Corollary 11. *Let T be a tree that is not a path and S a k-maximal subtree of T. Then the leaves of S of degree 2 either all lie in paths inside T^*, or all lie in limbs of T.*

Proof. Suppose v and w are leaves of S both of degree 2, and v lies in a body path, w lies in a limb. Let w' be the neighbour of w that does not lie in S. We know $i(v;S) < 1/2$ and $i(w';S) = 1/2$. Then removing v and including w' yields a k-subtree with higher local mean, which contradicts the assumption. □

5 On density and its localization

For the sake of comparing local means of subtrees of different orders, we normalize the local mean and define the local density of T which generalizes the (global) density [1].

Definition 12. Let S be a proper subtree of T, where $|T| = n$ and $|S| = k$. Let $\mu_T(S)$ be the local mean at S. Define the *local density at* S as

$$D_T(S) = \frac{\mu_T(S) - k}{n - k}.$$

Just as density can be interpreted as the probability of a randomly picked vertex being in a randomly picked subtree, local density at S can be interpreted as the probability of a random vertex *outside* of S being in a random subtree *containing* S.

Remark 13. If we extend the definition of the local density to allow the case that S is the null graph, namely $k = 0$, then the local density $D_T(S)$ reduces to the global density D_T. At the other end, when $k = n - 1$, there are only two subtrees containing S, therefore $\mu_T(S) = k + \frac{1}{2}$ and $D_T(S) = \frac{1}{2}$.

Theorem 14. *Let S be a proper subtree of T, where $|T| = n$ and $|S| = k$. If $2 \leqslant k \leqslant n-1$ and w is a leaf of S, then*

$$D(S) \geqslant D(S - w) \Leftrightarrow D(S) \geqslant 1 - i(w; S).$$

Similarly, if $1 \leqslant k \leqslant n - 2$ and v is a neighbour of S, then

$$D(S + v) \geqslant D(S) \Leftrightarrow D(S) \geqslant 1 - i(v; S).$$

Proof. By the Index Lemma, we have the following equivalent inequalities:

$$\begin{aligned} & D(S) \geqslant D(S - w) \\ \Leftrightarrow \quad & \frac{\mu(S) - k}{n - k} \geqslant \frac{\mu(S) - i(w) - (k - 1)}{n - (k - 1)} \\ \Leftrightarrow \quad & \frac{n - k + 1}{n - k} \geqslant \frac{\mu(S) - k - i(w) + 1}{\mu(S) - k} \\ \Leftrightarrow \quad & \frac{1}{n - k} \geqslant \frac{1 - i(w)}{\mu(S) - k} \\ \Leftrightarrow \quad & D(S) \geqslant 1 - i(w). \end{aligned}$$

The same calculation yields the other equivalence. □

We know that the local mean trivially increases if the subtree absorbs a neighbour, but for local density, this is no longer the case. Local density can increase after deleting a leaf or decrease after including a neighbour. An advantage, however, is that local density allows one to compare subtrees of *different* orders. Hence one can say "the subtree that attains maximal local density in T" without specifying the order of subtree unlike the situation for the local mean.

Example 15. Consider the tree T of order 6 shown in Figure 7.

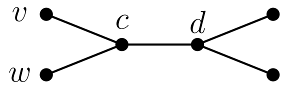

Figure 7: A tree of order 6.

Then $T^* = \{c, d\}$ and we have

$$D(v,c) = \frac{21}{40}, D(c) = \frac{13}{25}, D(v) = \frac{31}{55}.$$

Thus

$$D(c) < D(v,c) < D(v).$$

As this example shows, local density can change in both directions by removing a vertex, in other words, local density does not respect set-inclusion in general. There are four subtrees, the four singleton leaves, that attain maximal local density. Furthermore, direct calculation shows $D(v,w,c) = \frac{8}{15}$, while the local densities of the remaining subtrees (those not isomorphic to $\{v,w,c\}, \{v,c\}, \{c\}$ and $\{v\}$) are indeed all equal to $\frac{1}{2}$, which is justified by Theorem 4, which establishes a lower bound on the local density. We now prove the theorem.

Proof of Theorem 4. Let $k = |S|$. By contraction and its induced bijection and the equality (5), we have

$$D_T(S) = \frac{\mu_T(S) - k}{n-k} = \frac{\mu_{T/S}(S/S) - 1}{(n-k+1)-1} = D_{T/S}(S/S),$$

i.e., the local density is invariant under contraction. Then (3) implies

$$D_{T/S}(S/S) = \frac{\mu_{T/S}(S/S) - 1}{n-(k-1)-1} \geqslant \frac{\frac{n-(k-1)+1}{2} - 1}{n-k} = \frac{1}{2},$$

with equality if and only if T/S is astral over S/S, which is equivalent to $S \supseteq T^*$. ☐

Despite the fact that it is possible for the local density to increase or decrease in removing a vertex, with the above theorem, one can now say that the local density (at S) does not increase in removing a leaf of S of index $\frac{1}{2}$ (with respect to S).

Proposition 16. *Let S consist of two adjacent vertices v, w with $i(w;v) = \frac{1}{2}$. Then*

$$D_T(v) \leqslant D_T(v,w),$$

where equality holds if and only if T is astral over v.

Proof. This follows directly from the first half of Theorem 14 and Theorem 4. ☐

This proposition can easily be extended to a more general setting as follows.

Theorem 17. *Let* $S \subset T$ *and* $S' = S \cap T^*$. *If* $S' \neq \emptyset$, *then*

$$D(S') \leqslant D(S),$$

where equality holds if and only if $S \subseteq T^*$ *or* $T^* \subseteq S$.

Proof. Let $L = S - S'$ be the part of S consisting of limb vertices of T. If $L = \emptyset$, then $S = S'$, i.e., $S \subseteq T^*$, and the conclusion follows trivially. Now let $v \in L$ be a leaf of S in some limb. By Theorem 4, $D(S) \geqslant \frac{1}{2}$ with equality if and only if $T^* \subseteq S$. On the other hand, $i(v;S) = \frac{1}{2}$. Therefore, $D(S) \geqslant 1 - i(v;S)$. Then Theorem 14 yields $D(S-v) \leqslant D(S)$, with equality if and only if $T^* \subseteq S$. One can repeat this process of deleting limb leaves of S and eventually reach the conclusion. □

From the results so far, we know that the lower bound $D(S) = \frac{1}{2}$ is attained if and only if S contains the body of T. Any other subtree has $D(S) > \frac{1}{2}$.

On the other hand, local density can be arbitrarily close to 1. Let T be a tree of order n. Indeed, we know that local mean at a vertex is no less than global mean [1, Theorem 3.9], and that there exist trees with global mean arbitrarily close to n. Combining the two facts, there exist trees whose local densities are arbitrarily close to 1,

$$1 \geqslant D_T(v) = \frac{\mu_T(v) - 1}{n-1} \geqslant \frac{\mu_T - 1}{n-1} \to 1,\ n \to \infty.$$

However, the story of local density on the maximal side is still far from complete. We have not yet found a characterization for subtrees that attain maximal local density as we have for the minimal case. What is known is that if S attains the maximal local density in T then S necessarily attains the maximal local mean at order $|S|$. This observation follows immediately from the definition of local density. Hence the properties described in Section 4, those of k-maximal subtrees, also apply to subtrees of maximal local density. However, it is reasonable to expect that there exist properties that *only* subtrees of maximal local density have and k-maximal subtrees generally do not.

Theorem 17 reveals some information about the structure of the subtree attaining maximal local density. Let us call a body vertex v that is adjacent to some limb vertex a *joint vertex*. Now if a subtree (that does not contain the body, as this makes its local density minimal) contains some joint vertex v, then including all limbs connected to v would only increase the local density. So for a subtree of maximal local density and any joint vertex v, it would either avoid v or contain v along with all its limbs. Therefore, there are three possibilities for the structure of a subtree with maximal local density:

1. A proper subtree of the body that does not contain any joint vertex.

2. A subtree that can be decomposed into two parts: a proper subtree of the body containing some joint vertices, and their corresponding whole limbs.

3. A proper subtree (path) of some limb that contains the leaf.

Example 18. Let us take a look at some concrete examples on maximal local density. We start with paths and stars. The case of paths is straightforward since the path has no body. For a path of order n, by Theorem 4, any proper subtree S has $D(S) = \frac{1}{2}$. Now consider the star $T = K_{1,n-1}$ with centre c and a leaf v. For any subtree S with order at least 2, it must contain $\{c\} = T^*$. Hence, Theorem 4 yields $D(c) = \frac{1}{2}$ and $D(S) = \frac{1}{2}$ for $|S| \geqslant 2$. By symmetry, the only non-trivial local density is $D(v)$. Direct calculations give

$$D(v) = \frac{n2^{n-3}}{(n-1)(2^{n-2}+1)} \geqslant \frac{1}{2} = D(c), \tag{10}$$

for $n \geqslant 2$, and equality holds in the middle inequality only when $n = 2$ or 3.

For general trees, we use a computer program to exhaust all trees and subtrees up to order 15. Most maximizers (subtree(s) that attains maximal local density in its ambient tree) are singleton leaves or path segments in limbs containing leaves. In fact, all maximizers in trees of order up to 11 are of one of the two types; only at order 12 are there two trees whose maximizers contain vertices of degree greater than 2. For orders 13 and 14, the quantities of such trees increase to 8 and 36, respectively. Up to order 14, all maximizers either do not intersect T^* or intersect T^* only at joint vertices. For those maximizers containing joint vertices, they also contain the entire branch of limbs, as predicted. At order 15, there is one maximizer (the only one among trees of order 15 or less) that contains non-joint vertices in T^*, as indicated by circled vertices and thick edges in Figure 8.

To summarize, among the three cases described in the paragraph preceding this example, only the last two are observed in all trees up to order 15. Besides the absence of the first case, the theory does not exclude the possibility that maximizers contain more than one joint vertex (and thus also contain all branches of limbs incident to them), but no such instances were found.

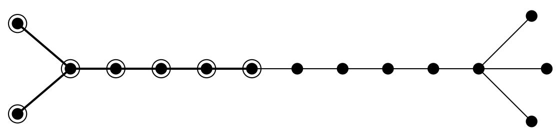

Figure 8: The only maximizer that contains non-joint vertices in the body in all trees of order up to 15.

The last part of this section answers the following question:

Does local density always exceed global density?

We know that local mean (at a vertex) exceeds global mean [1], for any tree with order greater than one. The next example eliminates the possibility for the analogue to hold for the density.

Example 19. Consider the star $T = K_{1,n-1}$ with $n \geqslant 2$. Let v be a leaf and c the centre. We will compare the three quantities, the global density D_T, the local density at a leaf $D_T(v)$ and the local density at the centre $D_T(c)$. For the global density D_T, the number of all subtrees is $2^{n-1} + n - 1$, and the total cardinality is

$$\begin{aligned} &\sum_{i=0}^{n-1} \binom{n-1}{i} (i+1) + n - 1 \\ &= 2^{n-1} + (n-1)\, 2^{n-2} + n - 1 \\ &= (n+1)\, 2^{n-2} + n - 1. \end{aligned}$$

So

$$D_T = \frac{\mu_T}{n} = \frac{(n+1)\, 2^{n-2} + n - 1}{n 2^{n-1} + n^2 - n}.$$

Observe that $D_T > \frac{1}{2}$, since the inequality is equivalent to $2^{n-1} > (n-1)(n-2)$, which holds for all $n \geqslant 1$. Then, on the local densities, we use the result (10) from the previous example:

$$D_T\,(v) = \frac{n 2^{n-3}}{(n-1)\,(2^{n-2}+1)} \geqslant \frac{1}{2} = D_T\,(c)\,,$$

where inequality holds strictly for $n \geqslant 4$, and $D_T(v) = \frac{1}{2} = D_T(c)$ for $n = 2$ or 3.What is left is to compare $D_T\,(v)$ and D_T. Simplification yields that

$$D_T\,(v) > D_T \text{ if and only if } \; 4^{n-2} + n(n-1)(n-4)2^{n-3} > (n-1)^2,$$

where the latter is true for all $n \geqslant 4$. To conclude, the three quantities satisfy

$$D_T\,(c) < D_T < D_T\,(v)$$

for $n \geqslant 4$, and

$$D_T\,(c) = D_T\,(v) < D_T$$

for $n = 2$ or 3. Hence the example shows that there is no general order relation between local densities and the global density.

6 Conclusion

We introduce the index to quantify incremental changes of local mean, which streamlines several arguments and yields structural results for extremal local means (Theorems 1-3). For local density, we prove the sharp lower bound $D_T(S) \geqslant \frac{1}{2}$ with equality if and only if $S \supseteq T^*$ (Theorem 4), give comparison principles that relate density changes to indices (Theorem 14) and to the interaction with the body T^* (Theorem 17). Computations over all trees up to order 15 show that maximizers of local density are mostly limb-paths containing leaves (including singleton leaves); when they touch T^*, they do so mostly only at joint vertices and include the entire attached limbs. Notably, among the three structural scenarios listed before Example 18, only cases (2) and (3) occur up to order 15. Therefore, it seems appropriate to conclude with the following conjecture.

Conjecture 20. Let T be a finite tree. Then the subtree(s) that attains maximal local density contains at least one leaf of T.

Acknowledgements

Supported by the Swedish research council (VR), grant 2022-04030.

I thank Professor Stephan Wagner and Professor John Haslegrave for all the discussions and advice.

References

[1] R. E. Jamison, On the average number of nodes in a subtree of a tree, J. Combin. Theory Ser. B 35(3):207–223, 1983.

[2] R. E. Jamison, Monotonicity of the mean order of subtrees, J. Combin. Theory Ser. B 37(1):70–78, 1984.

[3] S. Wagner and H. Wang, On the local and global means of subtree orders, J. Graph Theory 81(2):154–166, 2016.

[4] A. Vince and H. Wang, The average order of a subtree of a tree, J. Combin. Theory Ser. B 100(2):161–170, 2010.

[5] L. Mol and O. R. Oellermann, Maximizing the mean subtree order, J. Graph Theory 91(4):326–352, 2019.

[6] Z. Luo, K. Xu, S. Wagner, and H. Wang, On the mean subtree order of trees under edge contraction, J. Graph Theory 102(3):535–551, 2023.

[7] J. Haslegrave, Extremal results on average subtree density of series-reduced trees, J. Combin. Theory Ser. B 107:26–41, 2014.

[8] S. Cambie, S. Wagner, and H. Wang, On the maximum mean subtree order of trees, European J. Combin. 97:Paper No. 103388, 19 pp., 2021.

[9] R. Wang, On the difference of mean subtree orders under edge contraction, J. Combin. Theory Ser. B 169:45–62, 2024.