Bounding Mean Orders of Sub-k-Trees of k-Trees

Stijn Cambie[a] Bradley McCoy[b] Stephan Wagner[c,d] Corrine Yap[e]

Submitted: Oct 2, 2023; Accepted: Feb 22, 2024; Published: Mar 22, 2024

Abstract

For a k-tree T, we prove that the maximum local mean order is attained in a k-clique of degree 1 and that it is not more than twice the global mean order. We also bound the global mean order if T has no k-cliques of degree 2 and prove that for large order, the k-star attains the minimum global mean order. These results solve the remaining problems of Stephens and Oellermann [*J. Graph Theory* 88 (2018), 61–79] concerning the mean order of sub-k-trees of k-trees.

Mathematics Subject Classifications: 05C05, 05C35

1 Introduction

In [10] and [11] Jamison considered the mean number of nodes in subtrees of a given tree. He showed that for trees of order n, the average number of nodes in a subtree of T is at least $(n+2)/3$, with this minimum achieved if and only if T is a path. He also showed that the average number of nodes in a subtree containing a root is at least $(n+1)/2$ and always exceeds the average over all unrooted subtrees. The mean subtree order in trees was further investigated, e.g. in [3, 7, 14, 20, 22], as well as extensions to arbitrary graphs [1, 4–6] and the mean order of the connected induced subgraphs of a graph [8, 9, 17–19].

In [16], Stephens and Oellermann extended the study to k-trees and families of sub-k-trees. A *k-tree* is a generalization of a tree that has the following recursive construction.

Definition 1 (**k-tree**)**.** Let k be a fixed positive integer.

1. The complete graph K_k is a k-tree.

[a] Department of Computer Science, KU Leuven Campus Kulak, Kortrijk, Belgium (`stijn.cambie@hotmail.com`).

[b] School of Computing, Montana State University, Bozeman, MT, U.S.A. (`bradleymccoy@montana.edu`).

[c] Department of Mathematics, Uppsala University, Uppsala, Sweden (`stephan.wagner@math.uu.se`).

[d] Institute of Discrete Mathematics, TU Graz, Graz, Austria (`stephan.wagner@tugraz.at`).

[e] School of Mathematics, Georgia Institute of Technology, Atlanta, GA, U.S.A. (`math@corrineyap.com`).

2. If T is a k-tree, then so is the graph obtained from T by joining a new vertex to all vertices of some k-clique of T.

3. There are no other k-trees.

Note that for $k = 1$ we have the standard recursive construction of trees. A *sub-k-tree* of a k-tree T is a subgraph that is itself a k-tree. Let $S(T)$ denote the collection of all sub-k-trees of T and let $N(T) := |S(T)|$ be the number of sub-k-trees. We denote by $R(T) = \sum_{X \in S(T)} |X|$ the total number of vertices in all sub-k-trees (where for a graph G the notation $|G|$ is used throughout to mean the number of vertices in G). The *global mean (sub-k-tree) order* is

$$\mu(T) = \frac{R(T)}{N(T)}.$$

For an arbitrary k-clique C of T, let $S(T; C)$ denote the collection of sub-k-trees containing C and let $N(T; C) := |S(T; C)|$. The *local clique number* is $R(T; C) = \sum_{X \in S(T;C)} |X|$ and the *local mean (sub-k-tree) order* is

$$\mu(T; C) = \frac{R(T; C)}{N(T; C)}.$$

The *degree* of C is the number of $(k + 1)$-cliques that contain C.
Stephens and Oellermann concluded their study of the mean order of sub-k-trees of k-trees with several open questions. Our main contribution here is to answer three of them.
It was conjectured by Jamison in [10] and proven by Vince and Wang in [20] that for trees of order n without vertices of degree 2—called *series-reduced* trees—the global mean subtree order is between $\frac{n}{2}$ and $\frac{3n}{4}$. For k-trees we provide similar asymptotically sharp bounds, answering [16, Problem 6].

Theorem 2. *For every k-tree T without k-cliques of degree 2, the global mean sub-k-tree order satisfies*

$$\frac{n + k}{2} - o_n(1) < \mu(T) < \frac{3n + k - 3}{4}.$$

These bounds are asymptotically sharp. In particular, for large k, $\frac{3n}{4}$ is not an upper bound. For large n, the k-star is the unique extremal k-tree for the lower bound.

Wagner and Wang proved in [21] that the maximum local mean subtree order occurs at a leaf or a vertex of degree 2. We prove an analogous result for k-trees, answering [16, Problem 4]. In contrast to the result for trees, it turns out that for $k \geqslant 2$, the maximum can only occur at a k-clique of degree 1, unless T is a k-tree of order $k + 2$.

Theorem 3. *Suppose that $k \geqslant 2$. For a k-tree T of order $n \neq k + 2$, if a k-clique C maximizes $\mu(T; C)$, then C must be a k-clique of degree 1. For $n = k + 2$, every k-clique C satisfies $\mu(T; C) = k + 1$.*

Lastly, Jamison [10] conjectured that for a given tree T and any vertex v, the local mean order is at most twice the global mean order of all subtrees in T. Wagner and Wang [21] proved that this is true. Answering [16, Problem 2] affirmatively, we show

Theorem 4. *The local mean order of the sub-k-trees containing a fixed k-clique C is less than twice the global mean order of all sub-k-trees of T.*

1.1 Related Results

A total of six questions were posed in [16]. Problems 1 and 3 were solved by Luo and Xu [13]. Regarding the first problem, Jamison [10] showed that for any tree T and any vertex v of T, the local mean order of subtrees containing v is an upper bound on the global mean order of subtrees of T. Stephens and Oellermann asked about a generalization to k-trees, to which Luo and Xu showed:

Theorem 5 ([13])**.** *For any k-tree T of order n with a k-clique C, we have $\mu(T;C) \geqslant \mu(T)$ with equality if and only if $T \cong K_k$.*

For the third problem, it was shown in [10] that paths have the smallest global mean subtree order. For k-trees we have:

Theorem 6 ([13])**.** *For any k-tree T of order n, we have*

$$\mu(T) \geqslant \frac{\binom{n-k+2}{3}}{\binom{n-k+1}{2} + (n-k)k + 1} + k$$

with equality if and only if T is a path-type k-tree.

Very recently, Li, Ma, Dong, and Jin [12] gave a partial proof of Theorem 3, showing that the maximum local mean order always occurs at a k-clique of degree 1 or 2, thus also solving [16, Problem 4]. They did this by combining the fact that the 1-characteristic trees of adjacent k-cliques can be obtained from each other by a partial Kelmans operation [12, Lem. 4.6], an inequality between local orders of neighboring vertices after performing a partial Kelmans operation [12, Thm. 3.3], as well as [16, Lem. 11] and [21, Thm. 3.2]. Theorem 3 also solves Problem 5.4 from [12], asking whether the maximum local mean order can ever occur at a k-clique of degree 2, but not at a k-clique of degree 1.
Outline: In Section 2, we go over definitions and notation. Theorems 2, 3, 4 are proven in Sections 3, 4, 5 respectively. We additionally address [16, Problem 5], which is a more general question asking what one can say about the local mean order of sub-k-trees containing a fixed r-clique for $1 \leqslant r \leqslant k$. We give a possible direction and partial results in the concluding section.

2 Notation and Definitions

The global mean sub-k-tree order $\mu(T)$ and the local mean sub-k-tree order $\mu(T;C)$ are defined in the introduction. The local mean order always counts the k vertices from C and it sometimes will be more convenient to work with the average number of additional vertices in a (uniform random) sub-k-tree of T containing C, in which case we use the notation $\mu^\bullet(T;C) = \mu(T;C) - k$. Moreover, the number of sub-k-trees not containing C

will be denoted by $\overline{N}(T; C) = N(T) - N(T; C)$ and the total number of vertices in the sub-k-trees that do not contain C will be denoted by $\overline{R}(T; C) = R(T) - R(T; C)$.
A *k-leaf* or *simplicial vertex* is a vertex belonging to exactly one $(k+1)$-clique of T, i.e., a vertex of degree k. A *simplicial k-clique* is a k-clique containing a k-leaf. Note that a k-clique of degree 1 is not necessarily simplicial. A *major k-clique* is a k-clique with degree at least 3. Two k-cliques are *adjacent* if they share a $(k-1)$-clique.
The *stem* of a k-tree T is the k-tree obtained by deleting all k-leaves from T.
Subclasses of trees generalize to subclasses of k-trees. We will reference two in particular: paths generalize to *path-type k-trees* that are either isomorphic to K_k or K_{k+1} or have precisely two k-leaves. Note that for every $n \geqslant k+4$ and $k \geqslant 2$, there are multiple non-isomorphic path-type k-trees of order n.
A *k-star* is either K_k or K_{k+1}, or it is the unique k-tree with $n-k$ simplicial vertices when $n \geqslant k+2$.
Furthermore, the combination of a k-star and a k-path is called a k-broom: take a k-path of a certain length and add some simplicial vertices to a simplicial k-clique of the k-path.

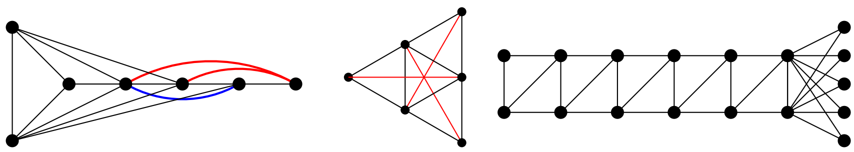


Figure 1: Examples of a 3-path, 3-star and 2-broom.

For a given k-clique C in T, it is often useful to decompose T into C and the sub-k-trees that result from deleting C. Let $B_1, B_2, \ldots, B_d$ be the $(k+1)$-cliques that contain C, and let v_i be the vertex of $B_i \setminus C$. Moreover, let $C_{i,1}, \ldots, C_{i,k}$ be the k-subcliques of B_i other than C. The k-tree T can be *decomposed* into C and k-trees $T_{1,1}, \ldots, T_{d,k}$, rooted at $C_{1,1}, \ldots, C_{d,k}$ respectively, that are pairwise disjoint except for the vertices of the cliques $C_{i,j}$.
In a 1-tree, any two vertices are connected by a unique path. This fact generalizes to k-trees through the construction of the 1-characteristic tree of a k-tree T [13, 16]. For a k-clique C in T, a *perfect elimination ordering of T to C* is an ordering $v_1, v_2, \ldots, v_{n-k}$ of its vertices other than $V(C) = \{c_1, \ldots, c_k\}$ such that each vertex v_i is simplicial in the k-tree spanned by C and $v_j, 1 \leqslant j \leqslant i$. In [16] it is shown that for any $v \notin C$, there is a unique sequence of vertices that along with C induce a path-type k-tree $P(C, v)$ and that form a perfect elimination ordering of $P(C, v)$ to C. It is also proven that T can be written as $\bigcup P(C, v)$ where the union is taken over all k-leaves $v \in V$. Each k-tree $P(C, v)$ has an associated 1-tree $P'(C, v)$ where the vertices consist of a single vertex representing the entire clique C, along with the remaining non-C vertices of $P(C, v)$. The edges are consecutive pairs from the perfect elimination ordering. Taking $\bigcup P'(C, v)$ over all k-leaves v gives us what is called the 1-*characteristic tree* of T, which we will denote T'_C. See Figure 2 for an example.

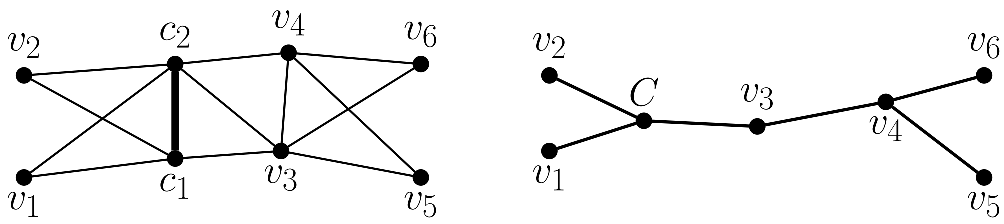


Figure 2: A 2-tree (left) with 2-clique $C = \{c_1, c_2\}$ and the 1-characteristic tree (right).

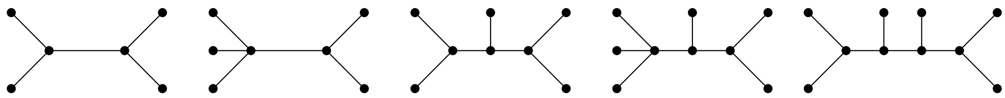


Figure 3: Series-reduced trees with minimum mean subtree order for $6 \leqslant n \leqslant 10$.

3 Excluding k-cliques of degree 2

In this section, we prove

Theorem 2. *For every k-tree T without k-cliques of degree 2, the global mean sub-k-tree order satisfies*

$$\frac{n+k}{2} - o_n(1) < \mu(T) < \frac{3n+k-3}{4}.$$

These bounds are asymptotically sharp. In particular, for large k, $\frac{3n}{4}$ is not an upper bound. For large n, the k-star is the unique extremal k-tree for the lower bound.

3.1 The lower bound

Among series-reduced trees of order n, for $4 \leqslant n \leqslant 8$ the star attains the maximum mean subtree order, but for $n \geqslant 11$ it attains the minimum mean subtree order. It can be derived from [7, Lem. 12] and an adapted version of [7, Cor. 11] (with 2 replaced by any $\varepsilon > 0$, at the cost of replacing 30 by n_ε) that this is indeed the case for n sufficiently large.[1]
In [2], it is shown that for $6 \leqslant n \leqslant 10$, the series-reduced trees attaining the minimum mean subtree order are those presented in Figure 3, and for $n \geqslant 11$, S_n is always the unique extremal graph.
In this subsection, we will prove that the above extremal statement generalizes to k-trees without a k-clique of degree 2.
First, we prove the following lemma, which states that every k-tree contains a $(k+1)$-clique C that plays the role of a centroid in a tree (a vertex or edge whose removal splits the tree into components of size at most $\frac{n}{2}$). Figure 4 is an example of a 2-tree demonstrating why we must take a $(k+1)$-clique and not a k-clique.

Lemma 7. *Any k-tree T (of order $n \geqslant k+1$) has a $(k+1)$-clique C such that the order of all components of $T \setminus C$ is at most $\left\lceil \frac{n-(k+1)}{2} \right\rceil$.*

[1] We thank John Haslegrave for this remark.

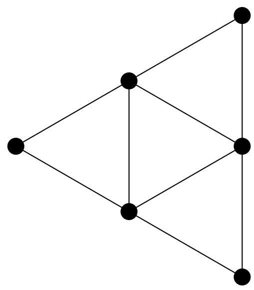


Figure 4: 2-tree with a centroid 3-clique but no centroid 2-clique.

Proof. Suppose for contradiction that such a $(k+1)$-clique does not exist. Consider the $(k+1)$-clique C for which the largest component of $T \setminus C$ has minimum order over all choices of C. By assumption, the largest component of $T \setminus C$ has order $n_0 > \left\lceil \frac{n-(k+1)}{2} \right\rceil$, and by the construction of a k-tree, there exist vertices $u \in C$ and v in the largest component T_0 of $T \setminus C$ such that u has no neighbor in T_0 and $v \cup (C \setminus u)$ forms a $(k+1)$-clique C'. Now $T \setminus C'$ has components whose sizes are bounded by $\max\{n - n_0 - k, n_0 - 1\} < n_0$, which gives a contradiction. Hence C satisfies the stated property. $\square$

The following lemma and its proof are similar to [15, Lemma 5.1].

Lemma 8. *For any k-tree T without k-cliques of degree 2, there is a k-clique C' for which $\mu(T; C') - \mu(T) \leqslant \frac{n^3}{2^{(n-k)/4+1}} = o_n(1)$.*

Proof. Take C as in Lemma 7. Let its vertices be $\{u_1, u_2, \ldots, u_{k+1}\}$.
For every component of $T \setminus C$, there is a unique vertex $u_i, 1 \leqslant i \leqslant k+1$, such that the component together with $C \setminus u_i$ forms a k-tree. Now we consider two cases that are handled analogously.
Case 1: There is some u_i such that the union of components that form a k-tree when adding $C' = C \setminus u_i$ has order at least $\frac{n-k}{2}$.
In this case, we consider the $r \leqslant n-k$ components of $T \setminus C'$ and for every $1 \leqslant j \leqslant r$, we let T_j be the union of such a component and C'.
We then apply the following claim:

Claim 9. *Let C' be a k-clique in a k-tree T of order n with no k-cliques of degree 2. Then T has at least $\frac{n-k+1}{2}$ simplicial vertices that are not in C'.*

Proof. Consider the 1-characteristic tree $T'_{C'}$, whose order is $n-k+1$. For every $u \in T \setminus C'$, either all the k-cliques containing u have degree 1, in which case u has degree 1 in $T'_{C'}$, or (at least) one of them has degree at least 3 and so does u in $T'_{C'}$. Thus $T'_{C'}$ is series-reduced and thus has at least $\frac{n-k+1}{2} + 1$ leaves. The latter implies that T has at least $\frac{n-k+1}{2}$ simplicial vertices that are not in C'. $\diamondsuit$

By the choice of C, for every $1 \leqslant j \leqslant r$, there are at least $\left\lfloor \frac{n-k+1}{2} \right\rfloor$ vertices that do not belong to T_j. By Claim 9, among them are at least $\frac{n-k}{4}$ simplicial vertices of T. Observe that given any sub-k-tree of T containing C', we can map it to its sub-k-tree intersection

with T_j. If two elements of $S(T;C')$ differ only in some subset of k-leaves of $\bigcup_{h\neq j} T_h$, then they map to the same element of $S(T_j;C')$. Thus, each element of $S(T_j;C')$ is mapped to at least $2^{\frac{n-k}{4}}$ times, and $N(T;C') \geqslant 2^{\frac{n-k}{4}} N(T_j;C')$.
Using the perfect elimination ordering, every sub-k-tree S in T_j not containing C' can be extended in a minimal way into a sub-k-tree containing C'. Furthermore, by considering the 1-characteristic tree $T'_{C'}$ of C', it is clear that there are no more than $|T_j| - k \leqslant \left\lceil \frac{n-(k+1)}{2} \right\rceil \leqslant \frac{n}{2}$ k-trees that extend to the same tree. Here we have used Lemma 7 again. Thus, if we define a map from $\overline{S}(T_j;C')$ to $S(T_j;C')$ using the minimal extension, every element of $S(T_j;C')$ is mapped to at most $\frac{n}{2}$ times.
Putting the previous two observations together, we have that the number of sub-k-trees containing C' is $N(T;C') \geqslant 2^{\frac{n-k}{4}} N(T_j;C') \geqslant \frac{1}{n} 2^{\frac{n-k}{4}+1} \overline{N}(T_j;C')$.
By summing over all j, we obtain that

$$rN(T;C') \geqslant \frac{1}{n} 2^{\frac{n-k}{4}+1} \sum_{j=1}^{r} \overline{N}(T_j;C') = \frac{1}{n} 2^{\frac{n-k}{4}+1} \overline{N}(T;C').$$

Since $r \leqslant n$, we have that $N(T;C') \geqslant \frac{1}{n^2} 2^{\frac{n-k}{4}+1} \overline{N}(T;C')$. This implies that $\frac{\mu(T)}{\mu(T;C')} \geqslant 1 - \frac{n^2}{2^{\frac{n-k}{4}+1}}$. The result follows now from $\mu(T;C') \leqslant n$.
Case 2: For every u_i the union of components in $T \setminus C$ that form a k-tree when adding $C \setminus u_i$ has order smaller than $\frac{n-k}{2}$.
Let T_i, $1 \leqslant i \leqslant k+1$, be the k-trees obtained above when adding $C \setminus u_i$. By Claim 9, for each i there are at least $\frac{n-k}{4}$ k-leaves not belonging to T_i. Thus, the same computations apply and in particular we have that $N(T;C') \geqslant \frac{1}{n^2} 2^{\frac{n-k}{4}+1} \overline{N}(T;C')$ as before. Now for $C' = C \setminus u_1$, we conclude by inclusion monotonicity [13, Thm. 33] that $\mu(T;C') - \mu(T) \leqslant \mu(T;C) - \mu(T) \leqslant \frac{n^3}{2^{(n-k)/4+1}}$. □

Proof of Theorem 2, lower bound. Take a k-clique C' which satisfies Lemma 8. Let $T'_{C'}$ be the 1-characteristic tree of T with respect to C'. By [13, Thm. 33] and [7, Lem. 12], we conclude that $\mu(T;C') = \mu(T'_{C'};C') + k - 1 \geqslant \frac{n+k}{2} + \frac{i-1}{10}$, where i is the number of internal vertices in $T'_{C'}$. By Lemma 8, we conclude.
For sharpness, observe that if T is not a k-star, we have $i \geqslant 2$ and the lower bound inequality is strict. When T is a k-star, it contains no k-cliques of degree 2 provided that $n > k+2$. As computed in [16] (page 64), $\mu(T) = \frac{R(T)}{N(T)} = \frac{(n+k)2^{n-k-1}}{2^{n-k}+(n-k)k} = \frac{n+k}{2} - o_n(1)$. □

3.2 The upper bound

In this subsection, we generalize to k-trees the statement that a series-reduced tree has average subtree order at most $\frac{3n}{4}$ by giving a lower bound for the number of k-leaves and proving that k-leaves belong to at most half of the sub-k-trees. This idea was also used in [7]. Note that the upper bound is slightly larger than $\frac{3n}{4}$ for larger k, which intuitively can be explained by the fact that the smallest sub-k-tree already has k vertices, and more precisely the vertices in the base k-clique will all be major vertices.

Proof of Theorem 2, upper bound. Our upper bound will come from the observation that $\mu(T) = \sum_{v \in T} p(v)$ where $p(v)$ is the fraction of sub-k-trees containing v. We will specifically consider when v is a k-leaf and bound the corresponding terms in the summation.

We first prove that the 1-characteristic tree T'_C of a k-tree T without k-cliques of degree 2 is a series-reduced tree, for any k-clique C of T. Indeed, given a k-clique C, there is either exactly one vertex adjacent to C or at least 3. As such, the degree of C in T'_C is not 2. For any other vertex $v \in T \setminus C$, either it is a leaf or some vertex w was added to a k-clique C', where v is the latest vertex of C'. In the latter case, note that v was added to some k-clique of the form $C' \cup \{u\} \setminus \{v\}$. This implies that u and w are adjacent to all of C', and since T has no k-cliques of degree 2, there must be at least another vertex w' with this property. Now v is adjacent to w, w', and some earlier vertex, so its degree in T'_C is at least 3.

Thus T'_C has at least $\frac{n-k+1}{2} + 1$ leaves, which implies that T contains at least this many k-leaves (here one must also observe that if C has degree 1 in T'_C, some vertex of C is simplicial in T).

Now fix a k-leaf v. Since $n \geqslant k+2$, there is a vertex u (different from v) such that $N(v) \cup \{u\}$ spans a K_{k+1}. Define a function f on sub-k-trees containing v such that $f(C') = (C' \setminus \{v\}) \cup \{u\}$ for a k-clique C' and $f(T') = T' \setminus v$ otherwise. We can check that f maps k-cliques to k-cliques and is in fact an injection from sub-k-trees containing v to sub-k-trees not containing v. Indeed, because v is a k-leaf, $C' \cup \{u\}$ is not a $(k+1)$-clique and thus not a sub-k-tree. Hence there does not exist a $(k+1)$-clique C'' for which $C'' \setminus \{v\} = (C' \setminus \{v\}) \cup \{u\}$.

This implies that every k-leaf belongs to at most half of the sub-k-trees in T. Remembering that there are at least $\frac{n-k+3}{2}$ k-leaves, the global mean order of T is then

$$\mu(T) = \sum_{v \text{ non-}k\text{-leaf}} p(v) + \sum_{v \ k\text{-leaf}} p(v) \leqslant n - \frac{1}{2} \cdot \frac{n-k+3}{2} = \frac{3n+k-3}{4}.$$

For sharpness, let $n = 2s+3-k$ for an integer s. We construct T by first constructing a caterpillar T' which consists of a path-type k-tree P_s^{k+1} on vertices $v_1, v_2, \ldots, v_s$, for which every $k+1$ consecutive vertices form a clique, and adding a k-leaf connected to every k consecutive vertices. To obtain T, we extend T' by adding two k-leaves, which are connected to $\{v_1, \ldots, v_k\}$ and $\{v_{s-k+1}, \ldots, v_s\}$ respectively. Note that T has a "stem" of s vertices and the number of k-leaves is $\ell + 2 = (s-k+1) + 2 = s-k+3$.

This k-tree T has the property that none of the k-cliques has degree 2: if a k-clique contains one of the k-leaves, its degree is 1, since every k-leaf only belongs to one $(k+1)$-clique. Otherwise, a k-clique consists entirely of vertices of the stem. If they are consecutive vertices $v_i, v_{i+1}, \ldots, v_{i+k-1}$, the degree of the clique is 3: the clique can be extended to a $(k+1)$-clique by adding v_{i-1}, v_{i+k}, or the additional k-leaf that is adjacent to it. If the k-clique consists of stem vertices that are non-consecutive, then it must be a subset of a clique of $k+1$ consecutive vertices $v_i, v_{i+1}, \ldots, v_{i+k}$ (otherwise its first and last vertex would not be adjacent). This only extends to one $(k+1)$-clique (by adding the missing vertex), so the degree is 1.

For every $1 \leqslant i \leqslant s-k$, there are i sub-k-trees of the stem each of order $s+1-i$. Each of these can be extended by adding any subset of the $\ell+1-i$ neighboring k-leaves, or even one or two more if some of the end-vertices of the stem are involved.
Now we can compute that

$$\begin{aligned} N(T) &= (k(n-k)+1-\ell) + 2^{\ell+2} + 2\sum_{i=2}^{\ell} 2^i + \sum_{i=3}^{s-k+1} (i-2)\cdot 2^{\ell+1-i} \\ &\sim 9 \cdot 2^{\ell}. \end{aligned}$$

The first expression $k(n-k)+1-\ell$ counts the number of simplicial k-cliques different from the ones consisting of k consecutive vertices in the stem. $2^{\ell+2}$ is the number of sub-k-trees containing the whole stem. The third term counts the number of sub-k-trees containing v_1 or v_s but not both and at least k vertices of the stem. The last summation counts the sub-k-trees containing at least k vertices of the stem and none of v_1 and v_s. We can also compute $R(T)$ by summing the total size of the respective sub-k-trees.

$$\begin{aligned} R(T) &= (k(n-k)+1-\ell)k + 2^{\ell+2}\left(s+\frac{\ell+2}{2}\right) + 2\sum_{i=2}^{\ell} 2^i \left(k+i-2+\frac{i}{2}\right) \\ &\quad + \sum_{i=3}^{s-k+1} (i-2)\cdot 2^{\ell+1-i}\left(s+1-i+\frac{\ell+1-i}{2}\right) \\ &\sim 2^{\ell+2}\left(s+\frac{\ell+2}{2}\right) + 2^{\ell+2}(k-2) + 3\cdot 2^{\ell+1}(\ell-1) + 2^{\ell}\left(s+1+\frac{\ell+1}{2}\right) - 15\cdot 2^{\ell-1} \\ &= \left(5s+\frac{17}{2}\ell+4k-16\right)2^{\ell} \\ &= \left(\frac{27}{4}n+\frac{9}{4}k-\frac{111}{4}\right)2^{\ell}. \end{aligned}$$

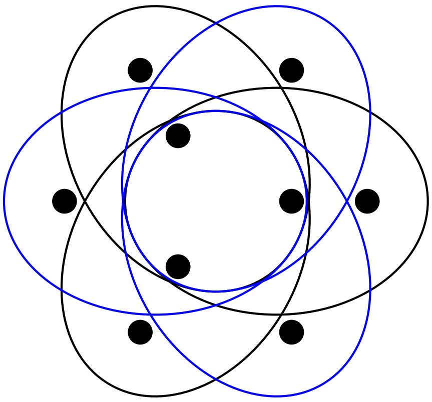
Finally, we conclude that $\mu(T) = \frac{R(T)}{N(T)} \sim \frac{3}{4}n + \frac{1}{4}k - \frac{37}{12}$, which is only $\frac{7}{3}$ away from the upper bound. These computations have also been verified in https://github.com/StijnCambie/AvSubOrder_ktree/blob/main/M_comb_ktree.mw. $\square$

4 The maximum local mean order

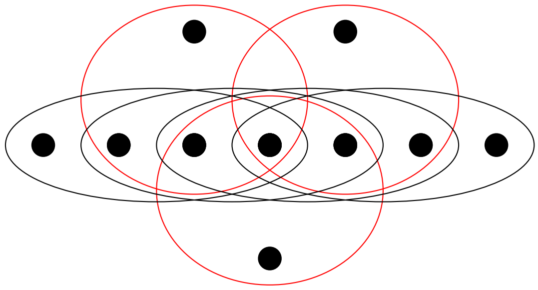
In this section, we prove

Theorem 3. *Suppose that $k \geqslant 2$. For a k-tree T of order $n \neq k+2$, if a k-clique C maximizes $\mu(T;C)$, then C must be a k-clique of degree 1. For $n = k+2$, every k-clique C satisfies $\mu(T;C) = k+1$.*

We consider the k-clique in a k-tree T for which the local mean order is greatest. Our aim is to show that its degree cannot be too large, specifically at most 2. To do so,



(a) A k-star



(b) A k-caterpillar.

Figure 5: Sketch of k-trees with mean sub-k-tree order roughly $\frac{n}{2}$ and $\frac{3n}{4}$ for $k = 3$.

we prove that if a k-clique C has degree at least 2, then there is a neighboring k-clique whose local mean order is not smaller, with strict inequality when C has degree at least 3. From this, it already follows that any k-clique attaining the maximum has degree at most 2. Moreover, in the case where the maximum is attained at a k-clique of degree 2, we will show that one can always move along maximizing k-cliques without doubling back, which means that one always ends up in at least one k-clique with degree 1 in which the maximum is attained.

The proof requires some technical inequalities, which we prove first. We start with a generalization of [21, Lemma 2.1] to the k-tree case as follows.

Lemma 10. *For a sub-k-tree T and a k-clique C, we have*

$$R(T; C) \leqslant \frac{N(T; C)^2 + (2k-1)N(T; C)}{2}.$$

Equality holds if and only if T is a path-type k-tree and C is a simplicial k-clique.

Proof. The statement is clearly true when $T = C$, so assume $|T| > k$. Observe that every sub-k-tree T' of T containing C has a vertex v (not belonging to C) which is simplicial within T' (consider a perfect elimination order where C is taken as the base k-clique). This implies that $T' \backslash v$ is a sub-k-tree containing C as well.

If there exists a sub-k-tree containing C of order $\ell > k$, then the above implies that there must also exist a sub-k-tree of order $\ell - 1$ containing C. Thus, if we list the sub-k-trees in $S(T; C)$ from smallest to largest order, we see that

$$R(T; C) \leqslant k + (k+1) + \cdots + (k + N(T; C) - 1) = \frac{1}{2}(N(T; C)^2 + (2k-1)N(T; C))$$

as desired.

The equality case is clear, since every sub-k-tree $T' \neq C$ of T must have exactly one k-leaf not belonging to C, and equality is attained when T is a path-type k-tree. $\square$

Using the previous lemma, we can now bound the local mean order in terms of the number of sub-k-trees.

Lemma 11. *For a k-tree T and one of its k-cliques C, we have*

$$k + \frac{\log_2 N(T;C)}{2} \leqslant \mu(T;C) \leqslant \frac{N(T;C) + (2k-1)}{2}.$$

The minimum occurs if and only if T is a k-star and C its base k-clique. The maximum is attained exactly when T is a path-type k-tree and C is simplicial.

Proof. The upper bound follows immediately from Lemma 10 by dividing both sides of the inequality by $N(T;C)$. The equality cases are the same.
By [16, Thm. 12] we have $\mu(T;C) \geqslant \frac{|T|+k}{2}$. Since a sub-$k$-tree is determined by its vertices, we also have

$$N(T;C) \leqslant 2^{|T|-k}. \tag{1}$$

Combining these two inequalities, we get $\mu(T;C) \geqslant \frac{\log_2 N(T;C)}{2} + k$.
To attain the lower bound, equality must hold for (1). This is the case if and only if T is a k-star and C its base k-clique. Indeed, for a k-star of order n with base clique C, we have $N(T;C) = 2^{n-k}$ and $\mu(T;C) = \frac{n+k}{2}$. In the other direction, every vertex together with C needs to form a k-tree and thus a $(k+1)$-clique, which is possible only for a k-star. □

We will also make use of the following elementary inequality, which can be considered as the opposite statement of the inequality between the arithmetic mean and geometric mean (AM-GM).

Lemma 12. *Let $x_1, x_2, \ldots, x_n \in \mathbb{R}_{\geqslant 1}$, and let $P = \prod_i x_i$. Then $\sum_i x_i \leqslant P + (n-1)$. Furthermore, equality is attained if and only if $x_i = P$ for some i and $x_j = 1$ for all $j \neq i$.*

Proof. This can be proven in multiple ways. The most elementary way is to observe that if $x_i, x_j > 1$, then $x_i x_j + 1 > x_i + x_j$ since $(x_i - 1)(x_j - 1) > 0$. Repeating this with pairs of elements which are strictly larger than 1 gives the result. Alternatively, one could consider the variables $\alpha_i = \log(x_i) \geqslant 0$. Since their sum is the fixed constant $\log P$ and the exponential function is convex, as a corollary of Karamata's inequality $\sum_i \exp(\alpha_i)$ is maximized when all except one are equal to 0. □

We are now ready to prove Theorem 3. Recall that T can be decomposed into C and k-trees $T_{1,1}, \ldots, T_{d,k}$ rooted at $C_{1,1}, \ldots, C_{d,k}$ respectively that are pairwise disjoint except for the vertices of the cliques $C_{i,j}$. Let B_i denote the $(k+1)$-clique that contains C as well as $C_{i,1}, \ldots, C_{i,k}$, and let v_i be the vertex of $B_i \setminus C$. Finally, set $N_{i,j} = N(T_{i,j}; C_{i,j})$ and $\mu^\bullet_{i,j} = \mu^\bullet(T_{i,j}; C_{i,j})$.

Proof of Theorem 3. We first observe that

$$N(T;C) = \prod_{i=1}^{d} \left(1 + \prod_{j=1}^{k} N_{i,j}\right). \tag{2}$$

This is because a sub-k-tree S of T containing C is specified as follows: given $1 \leqslant i \leqslant d$, choose the sub-$k$-tree intersection of S with $T_{i,1}, \ldots, T_{i,k}$. There are $\prod_{j=1}^{k} N_{i,j}$ ways to do this if $v_i \in S$ and one if $v_i \notin S$. We do this independently for each i, resulting in the product above.

Next, we would like to express the local mean order at C in terms of the quantities $N_{i,j}$ and $\mu^{\bullet}_{i,j}$: it is given by

$$\mu^{\bullet}(T; C) = \sum_{i=1}^{d} \frac{1 + \sum_{j=1}^{k} \mu^{\bullet}_{i,j}}{1 + \prod_{j=1}^{k} N_{i,j}^{-1}}. \tag{3}$$

To see this, note that we can interpret $\mu(T; C)$ as the expected size of a random sub-k-tree chosen from $S(T; C)$, which can then be written as the sum of the expected sizes of the intersection with each component $T_{i,j}$ in the decomposition. We have that $\frac{1}{1+\prod_{j=1}^{k} N_{i,j}^{-1}} = \frac{\prod_{j=1}^{k} N_{i,j}}{1+\prod_{j=1}^{k} N_{i,j}}$ is the probability that a randomly chosen sub-k-tree of T that contains C also contains v_i (by the same reasoning that gave us (2)). Once v_i is included, it adds 1 to the number of vertices, and an average total of $\sum_{j=1}^{k} \mu^{\bullet}_{i,j}$ is added from the extensions in $T_{i,1}, \ldots, T_{i,k}$.

Without loss of generality, we can assume that $N_{1,1} = \min_{i,j} N_{i,j}$. We want to compare $\mu^{\bullet}(T; C)$ to $\mu^{\bullet}(T; C_{1,1})$ and prove that $\mu^{\bullet}(T; C_{1,1}) \geqslant \mu^{\bullet}(T; C)$ provided that $d \geqslant 2$, with strict inequality if $d > 2$. Observe that this will be enough to prove our claim: no clique with degree greater than 2 can attain the maximum local mean order, and starting from any clique, we may repeatedly apply the inequality above to obtain a sequence of neighboring cliques whose local mean orders are weakly increasing and the last of which is a degree-1 k-clique. (More precisely, for the first step replacing C with $C' := C_{1,1}$ and considering the decomposition $\{C'_{i,j}\}$ and $\{T'_{i,j}\}$ with respect to C', the clique adjacent to C' with equal or larger $\mu^{\bullet}$ cannot be the original clique C as C will not correspond to $\min_{i,j} N'_{i,j}$. So in repeatedly applying the inequality, we will obtain a sequence of distinct cliques which must terminate but can only terminate once we have reached a clique of degree 1.)

Let us first express $\mu^{\bullet}(T; C_{1,1})$ in terms of the $N_{i,j}$ and $\mu^{\bullet}_{i,j}$ as well. First, we have

$$N(T; C_{1,1}) = N_{1,1} + \prod_{j=1}^{k} N_{1,j} \prod_{i=2}^{d} \left(1 + \prod_{j=1}^{k} N_{i,j}\right).$$

The reasoning is similar to (2): there are $N_{1,1}$ sub-k-trees that contain $C_{1,1}$, but not the full $(k+1)$-clique B_1, and the remaining product counts sub-k-trees containing B_1. We also have

$$\mu^{\bullet}(T; C_{1,1}) = \frac{N_{1,1}\mu^{\bullet}_{1,1}}{N(T; C_{1,1})} + \left(1 - \frac{N_{1,1}}{N(T; C_{1,1})}\right)\left(1 + \sum_{j=1}^{k} \mu^{\bullet}_{1,j} + \sum_{i=2}^{d} \frac{1 + \sum_{j=1}^{k} \mu^{\bullet}_{i,j}}{1 + \prod_{j=1}^{k} N_{i,j}^{-1}}\right),$$

using the fact that $\frac{N_{1,1}}{N(T; C_{1,1})}$ is the probability that a random sub-k-tree containing $C_{1,1}$ does *not* contain B_1. Let us now take the difference $\mu^{\bullet}(T; C_{1,1}) - \mu^{\bullet}(T; C)$: we have, after some manipulations,

$$\begin{aligned}\mu^\bullet(T;C_{1,1})-\mu^\bullet(T;C) &= \frac{N_{1,1}}{N(T;C_{1,1})}\left(\mu^\bullet_{1,1}-\mu^\bullet(T;C)\right)\\ &\quad+\left(1-\frac{N_{1,1}}{N(T;C_{1,1})}\right)\left(1+\sum_{j=1}^k \mu^\bullet_{1,j}-\frac{1+\sum_{j=1}^k \mu^\bullet_{1,j}}{1+\prod_{j=1}^k N_{i,j}^{-1}}\right)\\ &= \frac{N_{1,1}}{N(T;C_{1,1})}\left(\mu^\bullet_{1,1}-\mu^\bullet(T;C)\right)\\ &\quad+\left(1-\frac{N_{1,1}}{N(T;C_{1,1})}\right)\left(\frac{1+\sum_{j=1}^k \mu^\bullet_{1,j}}{1+\prod_{j=1}^k N_{1,j}}\right).\end{aligned}$$

Now let $F := \frac{1+\sum_{j=1}^k \mu^\bullet_{1,j}}{1+\prod_{j=1}^k N_{1,j}}$. We want to show that $\mu^\bullet(T;C_{1,1})-\mu^\bullet(T;C)\geqslant 0$, i.e., that

$$\frac{N_{1,1}}{N(T;C_{1,1})}\mu^\bullet_{1,1}+F\geqslant \frac{N_{1,1}}{N(T;C_{1,1})}\left(\mu^\bullet(T;C)+F\right).$$

Equivalently,

$$\mu^\bullet_{1,1}+\frac{N(T;C_{1,1})}{N_{1,1}}\cdot F\geqslant \mu^\bullet(T;C)+F. \tag{4}$$

Using the previous computations, we know that

$$\frac{N(T;C_{1,1})}{N_{1,1}}=1+\prod_{j=2}^k N_{1,j}\prod_{i=2}^d\left(1+\prod_{j=1}^k N_{i,j}\right).$$

So the left-hand side of (4) is equal to

$$\mu^\bullet_{1,1}+F+\frac{\prod_{j=2}^k N_{1,j}}{1+\prod_{j=1}^k N_{1,j}}\left(1+\sum_{j=1}^k \mu^\bullet_{1,j}\right)\prod_{i=2}^d\left(1+\prod_{j=1}^k N_{i,j}\right).$$

We can subtract F from both sides and use (3) to replace $\mu^\bullet(T;C)$. Taking into account that $\mu^\bullet_{1,1}\geqslant 0$, it is sufficient to prove

$$\frac{\prod_{j=2}^k N_{1,j}}{1+\prod_{j=1}^k N_{1,j}}\left(1+\sum_{j=1}^k \mu^\bullet_{1,j}\right)\prod_{i=2}^d\left(1+\prod_{j=1}^k N_{i,j}\right)\geqslant \sum_{i=1}^d \frac{1+\sum_{j=1}^k \mu^\bullet_{i,j}}{1+\prod_{j=1}^k N_{i,j}^{-1}}. \tag{5}$$

We first prove (5) for $d=2$, in which case it can be rewritten as

$$\begin{aligned}&\frac{\prod_{j=1}^k N_{1,j}}{N_{1,1}(1+\prod_{j=1}^k N_{1,j})}\left(1+\sum_{j=1}^k \mu^\bullet_{1,j}\right)\left(1-N_{1,1}+\prod_{j=1}^k N_{2,j}\right)\\ &\qquad\geqslant \frac{\prod_{j=1}^k N_{2,j}}{1+\prod_{j=1}^k N_{2,j}}\left(1+\sum_{j=1}^k \mu^\bullet_{2,j}\right).\end{aligned}$$

Write $\prod_{j=1}^k N_{1,j} = N_{1,1}^{k-1} y$ and $\prod_{j=1}^k N_{2,j} = N_{1,1}^{k-1} z$ where $y, z \geqslant N_{1,1}$. By Lemma 11, we have $\mu^\bullet_{1,j} \geqslant \frac{1}{2}\log_2 N_{1,j}$ and $\mu^\bullet_{2,j} \leqslant \frac{1}{2}(N_{2,j}-1)$. Applying Lemma 12 to the numbers $\frac{N_{2,j}}{N_{1,1}}$, $1 \leqslant j \leqslant k$, gives us that $\sum_{j=1}^k N_{2,j} \leqslant (k-1)N_{1,1} + z$. Hence we find that it suffices to prove that

$$\frac{y\left(1 - N_{1,1} + N_{1,1}^{k-1} z\right)}{N_{1,1}(1 + N_{1,1}^{k-1} y)} \left(1 + \frac{(k-1)\log_2 N_{1,1} + \log_2 y}{2}\right) \geqslant \frac{z}{1 + N_{1,1}^{k-1} z}\left(1 + \frac{(k-1)N_{1,1} + z - k}{2}\right).$$

We note that the left-hand side is strictly increasing in y and the right-hand side is independent of y, which implies that it is sufficient to prove the equality when $y = N_{1,1}$. That is, we want to prove that for every $z \geqslant y \geqslant 1$

$$\frac{1 - y + y^{k-1} z}{1 + y^k}\left(1 + \frac{k}{2}\log_2 y\right) \geqslant \frac{z}{1 + y^{k-1} z}\left(1 + \frac{(k-1)y + z - k}{2}\right). \tag{6}$$

If $y = N_{1,1} = 1$, (6) reduces to $\frac{z}{2} \geqslant \frac{z}{2}$.
If $y = N_{1,1} \geqslant 2$, $z \geqslant y$ and $k \geqslant 2$ imply that $2 + (k-1)y + z - k \leqslant kz$. Together with $\log_2 y \geqslant 1$, we conclude that it is sufficient to prove that

$$\frac{1 - y + y^{k-1} z}{1 + y^k}\frac{k+2}{2} \geqslant \frac{z}{1 + y^{k-1} z}\frac{kz}{2}$$

which is equivalent to

$$(1 - y + y^{k-1} z)(k+2)(1 + y^{k-1} z) \geqslant (1 + y^k) k z^2.$$

For $k = 2$, the difference between the two sides is an increasing function in y, and for $y = 2$ it reduces to $3z^2 - 2 \geqslant 0$, which holds.
For $k \geqslant 3$, the inequality is immediate, using that $y^{2(k-1)} z^2 \geqslant (1.5y^k + 1) z^2 \geqslant y^k z + y^k z^2 + z^2$ (remember that $z \geqslant y \geqslant 2$).
Once (5) has been verified for $d = 2$, we can apply induction to prove it for $d \geqslant 3$. Let $C = \frac{\prod_{j=2}^k N_{1,j}}{1 + \prod_{j=1}^k N_{1,j}}\left(1 + \sum_{j=1}^k \mu^\bullet_{1,j}\right)$, $g_i = 1 + \prod_{j=1}^k N_{i,j}$ and $f_i = \frac{1 + \sum_{j=1}^k \mu^\bullet_{i,j}}{1 + \prod_{j=1}^k N_{i,j}^{-1}}$. We can then rewrite (5) as

$$C \prod_{i=2}^d g_i \geqslant \sum_{i=1}^d f_i$$

By the induction hypothesis, we have

$$\frac{C}{g_m}\prod_{i=2}^d g_i \geqslant \left(\sum_{i=1}^d f_i\right) - f_m$$

for every $2 \leqslant m \leqslant d$. Summing over m, we obtain that

$$C \prod_{i=2}^{d} g_i \left(\sum_{m=2}^{d} \frac{1}{g_m} \right) \geqslant f_1 + (d-2) \sum_{i=1}^{d} f_i.$$

Since we have $g_m \geqslant 2$ for every m, the conclusion now follows as

$$\sum_{m=2}^{d} \frac{1}{g_m} \leqslant \frac{(d-1)}{2} \leqslant d-2,$$

and since $f_1 > 0$, inequality (5) is even strict in the case that $d > 2$.
We conclude that if C has degree $d \geqslant 2$, then $\mu^\bullet(T; C) \leqslant \mu^\bullet(T; C_{1,1})$. Equality can only be attained when $d = 2$. Looking back over the proof of (5), we also see that for equality to hold, all $N_{i,j}$ except for $N_{2,2}$ (up to renaming) must be equal to 1, and due to Lemma 11, $T_{2,2}$ has to be a path-type k-tree.
For $n \geqslant k+3$, we compute that in a path-type k-tree T the maximum among the degree-1 k-cliques is attained by a central one, which implies that no degree-2 k-cliques can be extremal. Let B_1 be the unique $(k+1)$-clique containing C, and assume that $T \setminus B_1$ contains components of size a and b. Thus $a + b = n - (k+1)$. Then $\mu(T; C) = \frac{k+(a+1)(b+1)\frac{n+k}{2}}{(a+1)(b+1)+1} = \frac{n+k}{2} - \frac{n-k}{2((a+1)(b+1)+1)}$, and this is maximized if and only if $|a-b| \leqslant 1$. The latter can also be derived from considering the 1-characteristic tree. This is illustrated in Figure 6 below. We emphasize that the maximizing k-clique of degree 1 is not simplicial. When $n \leqslant k+2$, every k-clique has the same local mean sub-k-tree order, and so the (unique) degree-2 k-clique when $n = k+2$ is the only case where equality can occur at a k-clique with degree 2. This concludes the proof of Theorem 3. $\square$

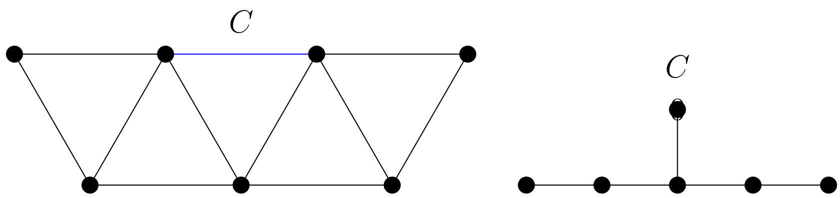


Figure 6: 2-path for which the maximum local mean is attained in a non-simplicial clique C, with its 1-characteristic tree T'_C.

5 Bounding local mean order by global mean order

In this section, we prove

Theorem 4. *The local mean order of the sub-k-trees containing a fixed k-clique C is less than twice the global mean order of all sub-k-trees of T.*

As before, given a k-tree T and a k-clique C in T, we utilize the decomposition of T into C and k-trees $T_{1,1}, \ldots, T_{d,k}$ rooted at $C_{1,1}, \ldots, C_{d,k}$. Set $N_{i,j} = N(T_{i,j}; C_{i,j})$, $\overline{N}_{i,j} = \overline{N}(T_{i,j}; C_{i,j})$, $\mu_{i,j} = \mu(T_{i,j}; C_{i,j})$, $\overline{\mu}_{i,j} = \overline{\mu}(T_{i,j}; C_{i,j})$, $\mu^\bullet_{i,j} = \mu^\bullet(T_{i,j}; C_{i,j})$ and $R_{i,j} = R(T_{i,j}; C_{i,j})$, $\overline{R}_{i,j} = \overline{R}(T_{i,j}; C_{i,j})$.
Since a sub-k-tree not containing C needs to be a sub-k-tree of some k-tree $T_{i,j}$, we have

$$\overline{N}(T; C) = \sum_{i=1}^{d} \sum_{j=1}^{k} (N_{i,j} + \overline{N}_{i,j})$$

and

$$\overline{R}(T; C) = \sum_{i=1}^{d} \sum_{j=1}^{k} (R_{i,j} + \overline{R}_{i,j}).$$

Following the proof for trees, we show

Lemma 13. *For any k-tree T and k-clique $C \in T$,*

$$R(T; C) > \overline{N}(T; C).$$

Proof. Assume to the contrary that there exists a minimum counterexample T. Since the statement is true when $T = C$, we have $|T| > k$ and we can consider the decomposition as before.
Note that if $N_{i,j} = 1$, we have that $\overline{N}_{i,j} = 0$, and otherwise we have $\overline{N}_{i,j} \leqslant R_{i,j} = \mu_{i,j} N_{i,j} = (k + \mu^\bullet_{i,j}) N_{i,j}$.
We can rewrite $R(T; C) - \overline{N}(T; C)$ as

$$N(T; C) \left(k + \mu^\bullet(T; C)\right) - \overline{N}(T; C).$$

Thus it suffices to show that

$$N(T; C)(k + \mu^\bullet(T; C)) - \sum_{i=1}^{d} \sum_{j=1}^{k} N_{i,j} - \sum_{i=1}^{d} \sum_{\substack{j=1 \\ N_{i,j} > 1}}^{k} (k + \mu^\bullet_{i,j}) N_{i,j}$$

is nonnegative. Expanding using (2) and (3) gives an expression that is increasing in $\mu^\bullet_{i,j}$ for each i, j, and so it suffices to prove that this expression is nonnegative when each of these is 0.
Let f be a function on the positive integers defined by $f(x) = \begin{cases} 1 & \text{if } x = 1, \\ 1 + kx & \text{if } x > 1. \end{cases}$
We now want to prove that

$$\left(k + \sum_{i=1}^{d} \frac{1}{1 + \prod_{j=1}^{k} N_{i,j}^{-1}} \right) N(T; C) \geqslant \sum_{i=1}^{d} \sum_{j=1}^{k} f(N_{i,j}). \tag{7}$$

When $d = 1$, this becomes $k + (k+1)\prod_{j=1}^{k} N_{1,j} \geqslant \sum_{j=1}^{k} f(N_{1,j})$. When increasing a value $N_{1,j}$ which is at least equal to 2, the left-hand side increases more than the right-hand side. As such, it is sufficient to consider the case where a of the terms $N_{1,j}$ equal 2, while the other $k-a$ terms equal 1. In this case, the desired inequality holds since $k + (k+1)2^a > k + 2ak$ for every integer $0 \leqslant a \leqslant k$.
Next, we consider the case $d \geqslant 2$. In this case, when $N_{i,j}$ increases by 1, the left-hand side of (7) increases by at least $2k$ and the right-hand side by at most $2k$. When all $N_{i,j}$ are equal to 1, the conclusion follows from $\left(k + \frac{d}{2}\right) 2^d > k2^d > dk$. □

We now bound the local mean order by the global mean order.

Proof of Theorem 4. Let T be a k-tree and C a k-clique in T. We want to prove that

$$\mu(T; C) < 2\mu(T).$$

We proceed by induction on the number of vertices in T. Note first that the inequality is trivial if $|T| \leqslant 2k$: since the mean is taken over sub-k-trees, which have at least k vertices each, we have $\mu(T) \geqslant k$. On the other hand, we clearly have $\mu(T; C) \leqslant |T|$, and both inequalities hold with equality only if $|T| = k$.
We thus proceed to the induction step, and assume that $|T| > 2k$. We have two cases with respect to C.
Case 1: C is simplicial.
Let v be a k-leaf in C, let C' denote the clique adjacent to C, and let $T' = T - \{v\}$. Moreover, let $N, \overline{N}, R$, and $\overline{R}$ denote $N(T'; C')$, $\overline{N}(T; C')$, $R(T'; C')$ and $\overline{R}(T'; C')$, respectively. We have

$$\mu(T; C) = \frac{N + R + k}{N + 1}$$

and

$$\mu(T) = \frac{2R + \overline{R} + N + k^2}{2N + \overline{N} + k}.$$

We want to prove that

$$(2N + \overline{N} + k)(2\mu(T) - \mu(T; C)) > 0,$$

which is equivalent to

$$2R + 2\overline{R} + 2k^2 - 2k - \frac{(N + R + k)(\overline{N} + k - 2)}{N + 1} > 0.$$

By the induction hypothesis, we have

$$2\frac{R + \overline{R}}{N + \overline{N}} = 2\mu(T') > \mu(T'; C') = \frac{R}{N},$$

so it is sufficient to prove that

$$\frac{R(N+\overline{N})}{N}+2k^2-2k-\frac{(N+R+k)(\overline{N}+k-2)}{N+1}>0.$$

Multiplying by $\frac{N+1}{N}$, this is seen to be equivalent to

$$R-\overline{N}+2k^2-3k+2-\frac{(k-3)R+k\overline{N}-k^2}{N}+\frac{\overline{N}R}{N^2}>0.$$

This can be broken up into three terms as follows:

$$\frac{(N^2-kN+3N+\overline{N})(R-\overline{N})}{N^2}+\Big(\frac{\overline{N}}{N}-k\Big)^2+\Big((k-1)(k-2)+\frac{k^2}{N}+\frac{3\overline{N}}{N}\Big)>0.$$

Note here that the second term is trivially nonnegative, and the last term trivially positive. Since $|T'|\geqslant 2k$, we have $N>k$ (one gets at least $k+1$ sub-k-trees containing C' by successively adding vertices); hence $N^2-kN+3N+\overline{N}>0$. Thus the first term is positive by Lemma 13, completing the induction step in this case.

Case 2: C is not simplicial.

By Theorem 3, we only need to consider the case where C has degree 1. Let v be the unique common neighbor of C, and let C_i, $1\leqslant i\leqslant k$, be the other k-cliques in the $(k+1)$-clique spanned by $C\cup\{v\}$.

Let T_i, $1\leqslant i\leqslant k$, be the sub-k-trees rooted at C_i (pairwise disjoint except for the vertices of the cliques C_i). Let $N_i=N(T_i;C_i)$, $\overline{N}_i=\overline{N}(T_i;C_i)$, $R_i=R(T_i;C_i)$, $\overline{R}_i=\overline{R}(T_i;C_i)$, $\mu_i=\mu(T_i;C_i)$ and $\mu_i^\bullet=\mu^\bullet(T_i;C_i)$. We can assume without loss of generality that $N_1\geqslant N_2\geqslant\cdots\geqslant N_j>1=N_{j+1}=\cdots=N_k$, where $j\geqslant 2$ since C is not simplicial.

We can now express the local and global mean in a similar way to Case 1. Here $N(T;C)=\prod_{i=1}^k N_i+1$, and all the sub-$k$-trees counted here, except for C, contain v. We have

$$\mu(T;C)=\frac{\prod_{i=1}^k N_i(1+\sum_{i=1}^k\mu_i^\bullet)}{\prod_{i=1}^k N_i+1}+k,$$

$$\mu(T)=\frac{\prod_{i=1}^k N_i(1+\sum_{i=1}^k\mu_i^\bullet+k)+\sum_{i=1}^k(R_i+\overline{R}_i)+k}{\prod_{i=1}^k N_i+\sum_{i=1}^k(N_i+\overline{N}_i)+1}.$$

In the remainder of this section, we will omit the bounds in products and sums if they are over the entire range from 1 to k: $\sum N_i$ and $\prod N_i$ mean $\sum_{i=1}^k N_i$ and $\prod_{i=1}^k N_i$ respectively. Then $\big(\prod N_i+\sum(N_i+\overline{N}_i)+1\big)\,(2\mu(T)-\mu(T;C))$ equals

$$(\prod N_i)(1+\sum\mu_i^\bullet+k)+k+2\sum(R_i+\overline{R}_i)-k\sum(N_i+\overline{N}_i)-\frac{(\prod N_i)(1+\sum\mu_i^\bullet)\sum(N_i+\overline{N}_i)}{\prod N_i+1}. \tag{8}$$

We want to show that this expression is positive. By induction, we know that $2(R_i+\overline{R}_i)>(k+\mu_i^\bullet)(N_i+\overline{N}_i)$, thus $2(R_i+\overline{R}_i)-k(N_i+\overline{N}_i)>\mu_i^\bullet(N_i+\overline{N}_i)$. It follows that (8) is greater than

$$(\prod N_i)(1+\sum\mu_i^\bullet+k)+k+\sum\mu_i^\bullet(N_i+\overline{N}_i)-\frac{(\prod N_i)(1+\sum\mu_i^\bullet)\sum(N_i+\overline{N}_i)}{\prod N_i+1}.$$

Multiplying by $\frac{\prod N_i+1}{\prod N_i}$ and observing that this factor is greater than 1, we find that (8) is indeed positive if we can prove that

$$(\prod N_i+1)(1+\sum\mu_i^\bullet+k)+k+\sum\mu_i^\bullet(N_i+\overline{N}_i)\geqslant(1+\sum\mu_i^\bullet)\sum(N_i+\overline{N}_i). \quad (9)$$

In particular, a potential counterexample would have to satisfy

$$\frac{(1+\sum\mu_i^\bullet)\sum(N_i+\overline{N}_i)}{\prod N_i+1}>1+\sum\mu_i^\bullet+k. \quad (10)$$

To simplify proving eq. (9), we first note that it is sufficient to consider the case where $k=j$. Once $N_i,\mu_i^\bullet$ and $\overline{N}_i$ are fixed for $1\leqslant i\leqslant j$, the terms that are dependent on k are $(\prod_{i\leqslant j}N_i+1)k+k$ on the left, and $(1+\sum_{i\leqslant j}\mu_i^\bullet)(k-j)$ on the right. The latter since if $N_i=1$, then T_i only consists of C_i, and thus $\overline{N}_i=\mu_i^\bullet=0$. Now since $\prod_{i\leqslant j}N_i+1\geqslant 1+\sum_{i\leqslant j}N_i>1+\sum_{i\leqslant j}\mu_i^\bullet$, the increase of the left side is larger than the increase on the right side. Here $\prod_{i\leqslant j}N_i\geqslant\sum_{i\leqslant j}N_i$ is true since the product of $j\geqslant 2$ numbers, each greater than or equal to 2, is at least equal the sum of the same j numbers. Moreover, $N_i>\mu_i^\bullet$ follows from Lemma 11.
So from now on, we can assume that $j=k$ and all N_i are at least equal to 2.
By lemma 13, $\overline{N}_i\leqslant R_i=\mu_iN_i=(k+\mu_i^\bullet)N_i$. Since (9) is a linear inequality in each $\overline{N}_i$ and the coefficient on the right-hand side is always greater than the coefficient on the left-hand side, we can reduce eq. (9) to a sufficient inequality that is only dependent on k, N_i and $\mu_i^\bullet$ for $1\leqslant i\leqslant k$ by taking $\overline{N}_i=(k+\mu_i^\bullet)N_i$. This will be assumed in the following.
If now all the parameters in (9) are fixed except for one $\mu_i^\bullet$, we have a linear inequality in $\mu_i^\bullet$: the quadratic terms stemming from $\mu_i^\bullet\overline{N}_i$ are equal on both sides and cancel. As such, it is sufficient to prove the inequality for the extremal values of $\mu_i^\bullet$. Here we use the trivial inequality $\mu_i^\bullet\geqslant 0$ as well as the upper bound $\mu_i^\bullet\leqslant\frac{N_i-1}{2}$, which is taken from lemma 11. So if (9) can be proven in the case where $\overline{N}_i=(k+\mu_i^\bullet)N_i$ and $\mu_i^\bullet\in\{0,\frac{N_i-1}{2}\}$ for all i with $1\leqslant i\leqslant k$, we are done.
For $2\leqslant k\leqslant 5$, this is achieved by exhaustively checking all 2^k cases that result (using symmetry, there are actually only $k+1$ cases to consider). See the detailed verifications in `https://github.com/StijnCambie/AvSubOrder_ktree`. So for the rest of the proof, we assume that $k\geqslant 6$, and we will use the slightly weaker bound $\mu_i^\bullet\leqslant\frac{N_i}{2}$ instead of $\mu_i^\bullet\leqslant\frac{N_i-1}{2}$ for $1\leqslant i\leqslant k$ in a few cases.
We distinguish two further cases, depending on the value of $\mu_1^\bullet$. In these cases, we will use the following two inequalities.

Claim 14. *Let $k\geqslant 6$ and $N_1\geqslant N_2\geqslant\ldots\geqslant N_k\geqslant 2$. Then*

$$\prod_{i=2}^k N_i\geqslant 3\sum_{2\leqslant i\leqslant k}N_i, \quad (11)$$

$$\frac{5}{3}(\prod_{i=1}^k N_i+1)\geqslant\left(\sum_{2\leqslant i\leqslant k}N_i-2\right)\sum_{1\leqslant i\leqslant k}N_i. \quad (12)$$

Proof. The first inequality, eq. (11), is true if all the N_i are equal to 2, since $2^{k-1} > 6(k-1)$ for every $k \geqslant 6$. Increasing some N_i by 1 increases the product by at least 2^{k-2}, while the sum increases by only 3. So the inequality holds by a straightforward inductive argument. Next, we prove eq. (12). When all N_i are equal to 2, it becomes $\frac{5}{3}(2^k+1) \geqslant 4(k-2)k$. This is easily checked for $k \in \{6,7\}$, and for $k \geqslant 8$, the stronger inequality $2^k \geqslant 4k^2$ can be shown by induction.

Now observe that the difference between the left- and right-hand sides is increasing with respect to N_1, since $\prod_{i\geqslant 2} N_i > \sum_{i\geqslant 2} N_i$ by the first inequality. It is also increasing in the other N_i's; for example, we can see this is true for N_2 since

$$\frac{5}{3}\prod_{i\neq 2} N_i \geqslant 5\sum_{i\neq 2} N_i > 2\sum_{1\leqslant i\leqslant k} N_i > \left(\sum_{2\leqslant i\leqslant k} N_i - 2\right) + \sum_{1\leqslant i\leqslant k} N_i.$$

In the first step, we have applied eq. (11) but replacing N_1 with N_2. Again, we may conclude using induction. ♢

Claim 15. *Given $\mu_1^\bullet = \frac{N_1}{2}$, it is sufficient to consider the case where $\mu_i^\bullet = \frac{N_i}{2}$ for all $1 \leqslant i \leqslant k$.*

Proof. Starting from any counterexample to (9), we can iteratively change $\mu_i^\bullet$ (considered as variables) for $1 \leqslant i \leqslant k$ based on the worst case of the linearization (to 0 if the coefficient on the left-hand side is greater and to $\frac{N_i}{2}$ if the coefficient on the right-hand side is greater) to obtain further counterexamples.

To show that it suffices to consider $\mu_i^\bullet = \frac{N_i}{2}$ for every $1 \leqslant i \leqslant k$, we prove that, given $\mu_1^\bullet = \frac{N_1}{2}$, the coefficient of $\mu_2^\bullet$ on the left-hand side in eq. (9) is not greater than the coefficient on the right-hand side. This then implies that $\mu_2^\bullet = \frac{N_2}{2}$ is indeed the worst case. For $3 \leqslant i \leqslant k$, we can argue in the same fashion.

Assume for sake of contradiction that the coefficient on the left-hand side is greater. Recall that $\overline{N}_2 = (k+\mu_2^\bullet)N_2$. After subtracting $\mu_2^\bullet(N_2+\overline{N}_2)$ from both sides, the coefficient of $\mu_2^\bullet$ on the left is $\prod N_i + 1$, while on the right it is $\sum_{i\neq 2}(N_i+\overline{N}_i) + (1+\sum_{i\neq 2}\mu_i^\bullet)N_2$. Thus we must have

$$\prod N_i + 1 > \sum_{i\neq 2}(N_i+\overline{N}_i) + (1+\sum_{i\neq 2}\mu_i^\bullet)N_2 \geqslant \sum_{i\neq 2}(N_i+\overline{N}_i) + (1+\mu_1^\bullet)N_2.$$

Using that

$$(1+\mu_1^\bullet)N_2 = \left(1+\frac{N_1}{2}\right)N_2 \geqslant \left(1+\frac{N_2}{2}\right)N_2 \geqslant (1+\mu_2^\bullet)N_2 = N_2 + \overline{N}_2 - kN_2$$

(recall here that we are assuming without loss of generality that $N_1 \geqslant N_2 \geqslant \cdots \geqslant N_k$), this implies that $\prod N_i + 1 \geqslant \sum(N_i+\overline{N}_i) - kN_2$. Adding kN_2 to both sides and multiplying both sides by $1+\sum\mu_i^\bullet$ results in

$$\left(\prod N_i + 1 + kN_2\right)(1+\sum\mu_i^\bullet) \geqslant (1+\sum\mu_i^\bullet)\sum(N_i+\overline{N}_i) > (1+\sum\mu_i^\bullet+k)(\prod N_i+1).$$

In the second inequality, we applied (10) which we may do since we began with the assumption that we have a counterexample to (9). After simplification, we get that $N_2(1+\sum \mu_i^\bullet) > \prod N_i$, and thus

$$\sum_{i\neq 2} N_i > 1 + \frac{1}{2}\sum N_i \geqslant 1 + \sum \mu_i^\bullet > \prod_{i\neq 2} N_i.$$

Since $k \geqslant 6$, this is a clear contradiction to eq. (11). $\diamondsuit$

Having proven Claim 15, we are left with two cases to consider: $\mu_i^\bullet = \frac{N_i}{2}$ for all $1 \leqslant i \leqslant k$, or $\mu_1^\bullet = 0$. It is easy to conclude in the former case, except when $k = 6$ and at least 5 values N_i are equal to 2, which has to be handled separately. See https://github.com/StijnCambie/AvSubOrder_ktree/blob/main/2M-mu_j_large_case1.mw for details. The final remaining case is when $\mu_1^\bullet = 0$. We obtain two new inequalities by multiplying eq. (12) with $\frac{k+1}{2}$ and eq. (11) with $(1+\sum \mu_i^\bullet)\frac{N_1}{6}$, and use that $N_1 = \max\{N_i\}$ and $\mu_i^\bullet \leqslant \frac{N_i-1}{2}$.

$$\frac{5(k+1)}{6}(\prod N_i + 1) \geqslant \frac{k+1}{2}\left(\sum_{2\leqslant i\leqslant k} N_i - 2\right)\sum_{1\leqslant i\leqslant k} N_i \geqslant (1+\sum \mu_i^\bullet)\sum_{1\leqslant i\leqslant k}(1+k)N_i, \tag{13}$$

$$\frac{\prod N_i}{6}(1+\sum \mu_i^\bullet) \geqslant (1+\sum \mu_i^\bullet)\frac{N_1}{2}\sum_{2\leqslant i\leqslant k} N_i \geqslant (1+\sum \mu_i^\bullet)\sum_{2\leqslant i\leqslant k} N_i\mu_i^\bullet. \tag{14}$$

Summing these two inequalities together, we have that eq. (9) holds as a corollary of

$$(\prod N_i + 1)(1+\sum \mu_i^\bullet + k) \geqslant (1+\sum \mu_i^\bullet)\sum\left(N_i(1+k+\mu_i^\bullet)\right). \qquad \square$$

We remark that by considering a suitable k-broom, one can show that theorem 4 is sharp, as was also the case for trees.

6 Conclusion

This paper together with [13, Thm. 18 & 20] answers all of the open questions from [16] except for one, which was stated as rather general and open-ended:

Problem 16. For a given r-clique R, $1 \leqslant r < k$, what is the (local) mean order of all sub-k-trees containing R?

One natural version of this question is to consider the local mean sub-k-tree order over sub-k-trees that contain a fixed vertex. In this direction, we prove the following result, which can be considered as another monotonicity result related to [13, Thm. 23].

Theorem 17. *Let T be a k-tree, $k \geqslant 2$, C a k-clique of T, and v a vertex in C. Then $\mu(T; v) < \mu(T; C)$.*

Proof. The statement is trivially true if $|T| \leqslant k+1$. So assume that T with $|T| \geqslant k+2$ is a minimum counterexample to the statement. Recalling the decomposition into trees $T_{i,j}$ used earlier, note that all sub-k-trees containing v and not C are part of $T_{i,j}$ for some i, j. Without loss of generality, we can assume that $\mu(T_{1,1}; v) = \max_{i,j} \mu(T_{i,j}; v)$. It suffices to prove that $\mu(T_{1,1}; v) < \mu(T; C)$. Taking into account (3), it is sufficient to consider the case where C is a simplicial k-clique of T. Let u be the simplicial vertex of B_1. Let $T' = T \setminus \{u\}$ and $C' = B_1 \setminus \{u\}$.

Claim 18. *We have $\mu(T'; C') < \mu(T; C)$.*

Proof. Let $R = R(T'; C')$ and $N = N(T'; C')$. We now need to prove that $\mu(T; C) = \frac{2R+N+k}{2N+1} > \frac{R}{N} = \mu(T'; C')$, which is equivalent to $N + k > \frac{R}{N}$. The latter is immediate since $N \geqslant |T'| - (k-1)$ and $\frac{R}{N} = \mu(T'; C') \leqslant |T'|$. ♢

Since the sub-k-trees containing v are exactly those that contain C, or sub-k-trees of T' containing C', or k-cliques within B_1 different from C, we conclude that $\mu(T; v) < \mu(T; C)$. □

Luo and Xu [13, Ques. 35] also asked if for a given order, a k-tree attaining the largest global mean sub-k-tree order is necessarily a caterpillar-type k-tree. In contrast with the questions in [16], this question is still open for trees. We prove that the local version (proven in [3, Thm. 3]), which states that for fixed order, the maximum is attained by a broom, is also true for the generalization of k-trees. This is almost immediate by observations from [16].

Proposition 19. *If a k-tree T of order n and k-clique C of T attain the maximum possible value of $\mu(T; C)$, then T has to be a k-broom with C being one of its simplicial k-cliques.*

Proof. Let T'_C be the characteristic 1-tree of T with respect to C. Then by [16, Lem. 11] $\mu(T; C) = \mu(T'_C; C) + k - 1$. Since T'_C is a tree on $n - (k-1)$ vertices, by [3, Thm. 3], the maximum local mean subtree order is attained by a broom B. Since there is a k-broom T with C being a simplicial k-clique for which $T'_C \cong B$ with C as root, this maximum can be attained. Reversely, if $\mu(T'_C; C)$ is maximized, then T'_C is a broom where C is its root (and thus simplicial). □

As such, we conclude that the k-tree variants of many results on the average subtree order for trees are now also proven. The analogue of [10, Ques. (7.5)] can be considered as the only question among them where the answer is slightly different for k-trees: in contrast with the case of trees ($k = 1$), the maximum local mean sub-k-tree order cannot occur in a k-clique with degree 2 when $k \geqslant 2$ (with one small exception).

Acknowledgments

The authors would like to express their gratitude to the American Mathematical Society for making a research visit possible related to the Mathematics Research Community workshop "Trees in Many Contexts". This was supported by the National Science Foundation under Grant Number DMS 1916439. The first author has been supported by internal Funds of KU Leuven (PDM fellowship PDMT1/22/005). The third author is supported by the Swedish research council (VR), grant 2022-04030. The fourth author is supported by the National Institutes of Health (R01GM126554).

References

[1] S. Cambie, G. Chen, Y. Hao, and N. Tokar. Decreasing the mean subtree order by adding k edges. *J. Graph Theory*, 105(3):357–366, 2024.

[2] S. Cambie and S. Wagner. Mean subtree order under contraction. *In preparation*, 2024.

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

[4] B. Cameron and L. Mol. On the mean subtree order of graphs under edge addition. *J. Graph Theory*, 96(3):403–413, 2021.

[5] X. Chen, G. Wei, and H. Lian. Solution to a conjecture on the mean subtree order of graphs under edge addition. *J. Graph Theory*, 104(3):645–658, 2023.

[6] A. J. Chin, G. Gordon, K. J. MacPhee, and C. Vincent. Subtrees of graphs. *J. Graph Theory*, 89(4):413–438, 2018.

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

[8] J. Haslegrave. The number and average size of connected sets in graphs with degree constraints. *J. Graph Theory*, 100(3):530–542, 2022.

[9] J. Haslegrave. The path minimises the average size of a connected induced subgraph. *Discrete Math.*, 345(5):Paper No. 112799, 7, 2022.

[10] 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.

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

[12] Z. Li, T. Ma, F. Dong, and X. Jin. On the maximum local mean order of sub-k-trees of a k-tree. arXiv:2309.11885, 2023.

[13] Z. Luo and K. Xu. On the Local and Global Mean Orders of Sub-k-Trees of k-Trees. *Electron. J. Combin.*, 30(1):Paper No. 1.43, 2023.

[14] 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.

[15] D. Ralaivaosaona and S. Wagner. On the distribution of subtree orders of a tree. *Ars Math. Contemp.*, 14(1):129–156, 2018.

[16] A. M. Stephens and O. R. Oellermann. The mean order of sub-k-trees of k-trees. *J. Graph Theory*, 88(1):61–79, 2018.

[17] A. Vince. The average size of a connected vertex set of a graph—explicit formulas and open problems. *J. Graph Theory*, 97(1):82–103, 2021.

[18] A. Vince. The average size of a connected vertex set of a k-connected graph. *Discrete Math.*, 344(10):Paper No. 112523, 4, 2021.

[19] A. Vince. A lower bound on the average size of a connected vertex set of a graph. *J. Combin. Theory Ser. B*, 152:153–170, 2022.

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

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

[22] R. Wang. Extrema of local mean and local density in a tree. arXiv:2306.13422, 2023.