Decreasing the mean subtree order by adding k edges

Stijin Cambie[2*], Guantao Chen[1†], Yanli Hao[1‡], Nizamettin Tokar[1§]
[1]Department of Mathematics and Statistics, Georgia State University
Atlanta, GA 30303, USA
[2]Extremal Combinatorics and Probability Group (ECOPRO)
Institute for Basic Science (IBS), Daejeon, KR

Abstract

The *mean subtree order* of a given graph G, denoted $\mu(G)$, is the average number of vertices in a subtree of G. Let G be a connected graph. Chin, Gordon, MacPhee, and Vincent [J. Graph Theory, 89(4): 413-438, 2018] conjectured that if H is a proper spanning supergraph of G, then $\mu(H) > \mu(G)$. However, Cameron and Mol [J. Graph Theory, 96(3): 403-413, 2021] have disproved this conjecture by showing that there are infinitely many pairs of graphs H and G with $H \supset G$, $V(H) = V(G)$ and $|E(H)| = |E(G)| + 1$ such that $\mu(H) < \mu(G)$. They also conjectured that for every positive integer k, there exists a pair of graphs G and H with $H \supset G$, $V(H) = V(G)$ and $|E(H)| = |E(G)| + k$ such that $\mu(H) < \mu(G)$. Furthermore, they proposed that $\mu(K_m + nK_1) < \mu(K_{m,n})$ provided $n \gg m$. In this note, we confirm these two conjectures.

Keywords: Mean subtree order; Subtree

1 Introduction

Graphs in this paper are simple unless otherwise specified. Let G be a graph with vertex set $V(G)$ and edge set $E(G)$. The *order* of G, denoted by $|G|$, is the number of vertices in G, that is, $|G| = |V(G)|$. The *complement* of G, denoted by $\overline{G}$, is the graph on the same vertex set as G such that two distinct vertices of $\overline{G}$ are adjacent if and only if they are not adjacent in G. For an edge subset $F \subseteq E(\overline{G})$, denote by $G + F$ the graph obtained from G by adding the edges of F. For a vertex subset $U \subseteq V(G)$, denote by $G - U$ the graph obtained from G by deleting the vertices of U and all edges incident with them. For any two graphs G_1, G_2 with $V(G_1) \cap V(G_2) = \emptyset$, denote by $G_1 + G_2$ the graph obtained from G_1, G_2 by adding an edge between any two vertices $v_1 \in V(G_1)$ and $v_2 \in V(G_2)$.

*supported by the Institute for Basic Science (IBS-R029-C4), current affiliation: KU Leuven, stijn.cambie@hotmail.com
†Partially supported by NSF grant DMS-1855716, gchen@gsu.edu
‡yhao4@gsu.edu
§ntokar1@gsu.edu

A tree is a graph in which every pair of distinct vertices is connected by exactly one path. A subtree of a graph G is a subgraph of G that is a tree. By convention, the empty graph is not regarded as a subtree of any graph. The *mean subtree order* of G, denoted $\mu(G)$, is the average order of a subtree of G. Jamison [5, 6] initiated the study of the mean subtree order in the 1980s, considering only the case that G is a tree. In [5], he proved that $\mu(T) \geq \frac{n+2}{3}$ for any tree T of order n, with this minimum achieved if and only if T is a path; and $\mu(T)$ could be very close to its order n. Jamison's work on the mean order of subtrees of a tree has received considerable attention [4, 8, 9, 10, 11]. At the 2019 Spring Section AMS meeting in Auburn, Jamison presented a survey that provided an overview of the current state of open questions concerning the mean subtree order of a tree, some of which have been resolved [1, 7].

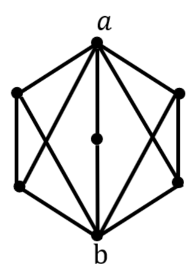

Figure 1: Adding the edge between a and b decreases the mean subtree order

Recently, Chin, Gordon, MacPhee, and Vincent [3] initiated the study of subtrees of graphs in general. They believed that the parameter μ is monotonic with respect to the inclusion relationship of subgraphs. More specifically, they [3, Conjecture 7.4] conjectured that for any simple connected graph G, adding any edge to G will increase the mean subtree order. Clearly, the truth of this conjecture implies that $\mu(K_n)$ is the maximum among all connected simple graphs of order n, but it's unknown if $\mu(K_n)$ is the maximum. Cameron and Mol [2] constructed some counterexamples to this conjecture by a computer search. Moreover, they found that the graph depicted in Figure 1 is the smallest counterexample to this conjecture and there are infinitely many graphs G with $xy \in E(\overline{G})$ such that $\mu(G+xy) < \mu(G)$. In their paper, Cameron and Mol [2] initially focused on the case of adding a single edge, but they also made the following conjecture regarding adding several edges.

Conjecture 1.1. *For every positive integer k, there are two connected graphs G and H with $G \subset H$, $V(G) = V(H)$ and $|E(H)\backslash E(G)| = k$ such that $\mu(H) < \mu(G)$.*

We will confirm Conjecture 1.1 by proving the following theorem, which will be presented in Section 2.

Theorem 1.2. *For every positive integer k, there exist infinitely many pairs of connected graphs G and H with $G \subset H$, $V(G) = V(H)$ and $|E(H)\backslash E(G)| = k$ such that $\mu(H) < \mu(G)$.*

In the same paper, Cameron and Mol [2] also proposed the following conjecture.

Conjecture 1.3. *Let m, n be two positive integers. If $n \gg m$, then we have $\mu(K_m + nK_1) < \mu(K_{m,n})$.*

We can derive Conjecture 1.1 from Conjecture 1.3, the proof of which is presented in Section 3, by observing that when $m = 2k$, the binomial coefficient $\binom{m}{2}$ is divisible by k. With $2k-1$ steps, we add k edges in each step, and eventually the mean subtree order decreases, so it must have decreased in some intermediate step.

2 Theorem 1.2

Let G be a graph of order n, and let $\mathcal{T}_G$ be the family of subtrees of G. By definition, we have $\mu(G) = (\sum_{T \in \mathcal{T}_G} |T|)/|\mathcal{T}_G|$. The *density* of G is defined by $\sigma(G) = \mu(G)/n$. More generally, for any subfamily $\mathcal{T} \subseteq \mathcal{T}_G$, we define $\mu(\mathcal{T}) = (\sum_{T \in \mathcal{T}} |T|)/|\mathcal{T}|$ and $\sigma(\mathcal{T}) = \mu(\mathcal{T})/n$. Clearly, $1 \leq \mu(G) \leq n$ and $0 < \sigma(G) \leq 1$.

2.1 The Construction

Fix a positive integer k. For some integer m, let $\{s_n\}_{n \geq m}$ be a sequence of non-negative integers satisfying: (1) $2s_n \leq n - k - 1$ for all $n \geq m$; (2) $s_n = o(n)$, i.e., $\lim_{n \to \infty} s_n/n = 0$; and (3) $2^{s_n} \geq n^2$ for all $n \geq m$. Notice that many such sequences exist. Take, for instance, the sequence $\{\lceil 2\log_2(n) \rceil\}_{n \geq m}$, as in [2], where m is the least positive integer such that $m - 2\lceil 2\log_2(m) \rceil \geq k+1$.

In the remainder of this paper, we fix P for a path $v_1 v_2 \cdots v_{n-2s_n}$ of order $n - 2s_n$. Clearly, $|P| \geq k+1$. Furthermore, let $P^* := P - \{v_1, \ldots, v_{k-1}\} = v_k \cdots v_{n-2s_n}$.

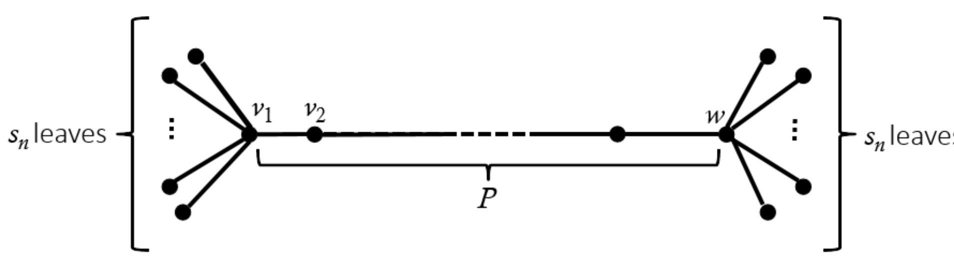

Figure 2: G_n

Let G_n be the graph obtained from the path P by joining s_n leaves to each of the two endpoints v_1 and $w := v_{n-2s_n}$ of P (see Figure 2). Let $G_{n,k} := G_n + \{v_1 w, v_2 w, \ldots, v_k w\}$, that is, $G_{n,k}$ is the graph obtained from G_n by adding k new edges $e_1 := v_1 w, e_2 := v_2 w, \ldots, e_k := v_k w$ (see Figure 3).

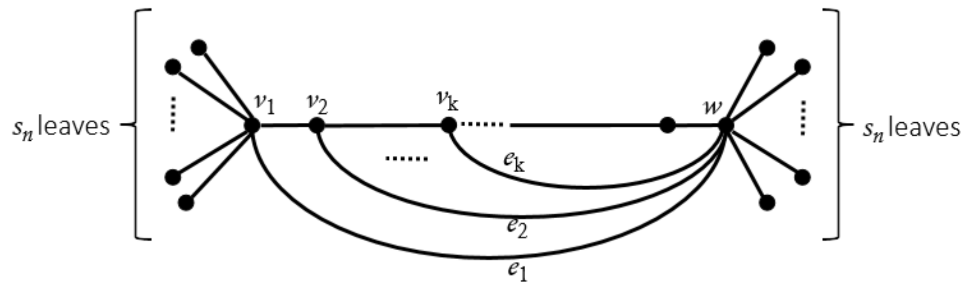

Figure 3: $G_{n,k}$

Let $\mathcal{T}_{n,k}$ be the family of subtrees of $G_{n,k}$ containing the vertex set $\{v_1, v_k, w\}$ but not containing the path $P^* = v_k \cdots w$. It is worth noting that $\mathcal{T}_{n,1}$ is the family of subtrees of $G_{n,1}$ containing edge $v_1 w$. Note that the graphs G_n and $G_{n,1}$ defined above are actually the graphs T_n and G_n constructed by Cameron and Mol in [2], respectively. From the proof of Theorem 3.1 in [2], we obtain the following two results regarding the density of $G_n, G_{n,1}, \mathcal{T}_{n,1}$.

Lemma 2.1. $\lim\limits_{n\to\infty} \sigma(G_n) = 1$.

Lemma 2.2. $\lim\limits_{n\to\infty} \sigma(G_{n,1}) = \lim\limits_{n\to\infty} \sigma(\mathcal{T}_{n,1}) = \frac{2}{3}$.

The following two technical results concerning the density of $\mathcal{T}_{n,k}$ are crucial in the proof of Theorem 1.2. The proofs of these results will be presented in Subsubsection 2.1.1 and Subsubsection 2.1.2, respectively.

Lemma 2.3. *For any fixed positive integer k,* $\lim\limits_{n\to\infty} \sigma(\mathcal{T}_{n,k}) = \lim\limits_{n\to\infty} \sigma(\mathcal{T}_{n-k+1,1})$.

Lemma 2.4. *For any fixed positive integer k,* $\lim\limits_{n\to\infty} \sigma(\mathcal{T}_{n,k}) = \lim\limits_{n\to\infty} \sigma(G_{n,k})$.

The combination of Lemma 2.2, Lemma 2.3 and Lemma 2.4 immediately yields the following result.

Corollary 2.5. *For any fixed positive integer k,* $\lim\limits_{n\to\infty} \sigma(G_{n,k}) = \frac{2}{3}$.

Combining Lemma 2.1 and Corollary 2.5, we have that $\lim\limits_{n\to\infty} \sigma(G_{n,k}) = \frac{2}{3} < 1 = \lim\limits_{n\to\infty} \sigma(G_n)$ for any fixed positive integer k. By definition, we gain that $\sigma(G_{n,k}) = \mu(G_{n,k})/|G_{n,k}|$ and $\sigma(G_n) = \mu(G_n)/|G_n|$. Since $|G_{n,k}| = |G_n|$, it follows that $\mu(G_{n,k}) < \mu(G_n)$ for n sufficiently large, which in turn gives Theorem 1.2.

The following result presented in [2, page 408, line -2] will be used in our proof.

Lemma 2.6. $|\mathcal{T}_{n,1}| = 2^{2s_n} \cdot \binom{n-2s_n}{2}$.

2.1.1 Proof of Lemma 2.3

Let H be the subgraph of $G_{n,k}$ induced by vertex set $\{v_1, \ldots, v_k, w\}$ (see Figure 4). Furthermore, set $n_1 = n - k + 1$, and let $G^+_{n_1}$ be the graph obtained from $G_{n,k}$ by contracting vertex set $\{v_1, \ldots, v_k\}$ into vertex v_1 and removing any resulting loops and multiple edges (see Figure 5). Clearly, $G^+_{n_1}$ is isomorphic to $G_{n_1,1}$.

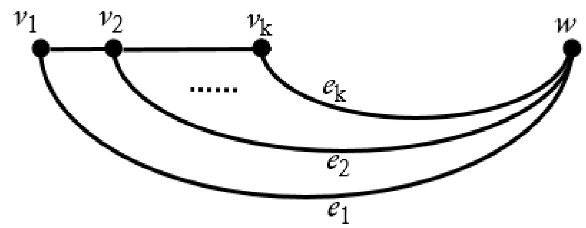

Figure 4: H

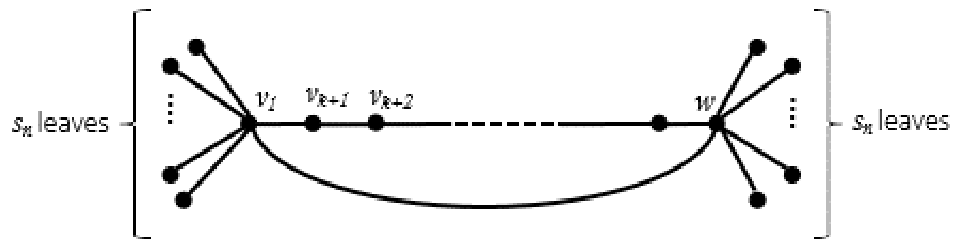

Figure 5: $G^+_{n_1}$

Let $T \in \mathcal{T}_{n,k}$, that is, T is a subtree of $G_{n,k}$ containing the vertex set $\{v_1, v_k, w\}$ but not containing the path $P^* = v_k \cdots w$. Let T_1 be the subgraph of H induced by $E(H) \cap E(T)$. Since T does not contain the path P^*, we have that T_1 is connected, and so it is a subtree of H. Let T_2 be the graph obtained from T by contracting vertex set $\{v_1, \ldots, v_k\}$ into the vertex v_1 and removing any resulting loops and multiple edges. Since T_1 is connected and contains vertex set $\{v_1, v_k, w\}$, it follows that T_2 is a subtree of $G^+_{n_1}$ containing edge v_1w. So, each $T \in \mathcal{T}_{n,k}$ corresponds to a unique pair (T_1, T_2) of trees, where T_1 is a subtree of H containing vertex set $\{v_1, v_k, w\}$, and $T_2 \in \mathcal{T}_{n_1,1}$. We also notice that $|T| = |T_1| + |T_2| - 2$, where the -2 arises due to the fact that T_1 and T_2 share exactly two vertices v_1 and w.

Let $\mathcal{T}'_H \subseteq \mathcal{T}_H$ be the family of subtrees of H containing vertex set $\{v_1, v_k, w\}$. By the corresponding relationship above, we have $|\mathcal{T}_{n,k}| = |\mathcal{T}'_H| \cdot |\mathcal{T}_{n_1,1}|$. Hence, we obtain that

$$
\begin{aligned}
\mu(\mathcal{T}_{n,k}) &= \frac{\sum\limits_{T\in\mathcal{T}_{n,k}}|T|}{|\mathcal{T}_{n,k}|} = \frac{\sum\limits_{T_1\in\mathcal{T}'_H}\sum\limits_{T_2\in\mathcal{T}_{n_1,1}}(|T_1|+|T_2|-2)}{|\mathcal{T}'_H|\cdot|\mathcal{T}_{n_1,1}|}\\
&= \frac{|\mathcal{T}'_H|\cdot\sum\limits_{T_2\in\mathcal{T}_{n_1,1}}|T_2| + |\mathcal{T}_{n_1,1}|\cdot\sum\limits_{T_1\in\mathcal{T}'_H}|T_1| - 2|\mathcal{T}_{n_1,1}|\cdot|\mathcal{T}'_H|}{|\mathcal{T}'_H|\cdot|\mathcal{T}_{n_1,1}|}\\
&= \mu(\mathcal{T}_{n_1,1}) + \mu(\mathcal{T}'_H) - 2.
\end{aligned}
$$

Dividing through by n, we further gain that

$$
\sigma(\mathcal{T}_{n,k}) = \frac{n_1}{n}\cdot\sigma(\mathcal{T}_{n_1,1}) + \frac{k+1}{n}\cdot\sigma(\mathcal{T}'_H) - \frac{2}{n}.
$$

Since $\sigma(\mathcal{T}'_H)$ is always bounded by 1, it follows that $\lim\limits_{n\to\infty}\frac{k+1}{n}\cdot\sigma(\mathcal{T}'_H)=0$. Combining this with $\lim\limits_{n\to\infty}\frac{n_1}{n}=1$ and $\lim\limits_{n\to\infty}\frac{2}{n}=0$, we get $\lim\limits_{n\to\infty}\sigma(\mathcal{T}_{n,k}) = \lim\limits_{n\to\infty}\sigma(\mathcal{T}_{n_1,1}) = \frac{2}{3}$ (by Lemma 2.2), which completes the proof of Lemma 2.3. □

2.1.2 Proof of Lemma 2.4

Let $\overline{\mathcal{T}}_{n,k} := \mathcal{T}_{G_{n,k}}\backslash\mathcal{T}_{n,k}$. If $\lim\limits_{n\to\infty}|\overline{\mathcal{T}}_{n,k}|/|\mathcal{T}_{n,k}|=0$, then $\lim\limits_{n\to\infty}\frac{|\overline{\mathcal{T}}_{n,k}|}{|\mathcal{T}_{n,k}|+|\overline{\mathcal{T}}_{n,k}|}=0$ because $\frac{|\overline{\mathcal{T}}_{n,k}|}{|\mathcal{T}_{n,k}|+|\overline{\mathcal{T}}_{n,k}|}\le |\overline{\mathcal{T}}_{n,k}|/|\mathcal{T}_{n,k}|$, and so $\lim\limits_{n\to\infty}\frac{|\mathcal{T}_{n,k}|}{|\mathcal{T}_{n,k}|+|\overline{\mathcal{T}}_{n,k}|}=1$. Hence,

$$
\begin{aligned}
\lim_{n\to\infty}\sigma(G_{n,k}) &= \lim_{n\to\infty}\frac{\mu(G_{n,k})}{n} = \lim_{n\to\infty}\frac{1}{n}\cdot\left(\frac{\sum\limits_{T\in\mathcal{T}_{n,k}}|T|}{|\mathcal{T}_{n,k}|+|\overline{\mathcal{T}}_{n,k}|}+\frac{\sum\limits_{T\in\overline{\mathcal{T}}_{n,k}}|T|}{|\mathcal{T}_{n,k}|+|\overline{\mathcal{T}}_{n,k}|}\right)\\
&= \lim_{n\to\infty}\left(\sigma(\mathcal{T}_{n,k})\cdot\frac{|\mathcal{T}_{n,k}|}{|\mathcal{T}_{n,k}|+|\overline{\mathcal{T}}_{n,k}|}+\sigma(\overline{\mathcal{T}}_{n,k})\cdot\frac{|\overline{\mathcal{T}}_{n,k}|}{|\mathcal{T}_{n,k}|+|\overline{\mathcal{T}}_{n,k}|}\right) = \lim_{n\to\infty}\sigma(\mathcal{T}_{n,k}).
\end{aligned}
$$

Thus, to complete the proof, it suffices to show that $\lim\limits_{n\to\infty}|\overline{\mathcal{T}}_{n,k}|/|\mathcal{T}_{n,k}|=0$. We now define the following two subfamilies of $\mathcal{T}_{G_{n,k}}$.

- $\mathcal{B}_1=\{T\in\mathcal{T}_{G_{n,k}}\ :\ v_1\notin V(T)\text{ or }w\notin V(T)\}$; and
- $\mathcal{B}_2=\{T\in\mathcal{T}_{G_{n,k}}\ :\ T\cap P^*\text{ is a path, and }T\text{ contains }w\}$.

Recall that $\mathcal{T}_{n,k}$ is the family of subtrees of $G_{n,k}$ containing vertex set $\{v_1,v_k,w\}$ and not containing the path $P^*=v_k\cdots w$. For any $T\in\overline{\mathcal{T}}_{n,k}$, by definition, we have the following scenarios: $v_1\notin V(T)$, and so $T\in\mathcal{B}_1$ in this case; $w\notin V(T)$, and so $T\in\mathcal{B}_1$ in this case; $v_k\notin V(T)$ and $w\in V(T)$, then $T\cap P^*$ is a path, and so $T\in\mathcal{B}_2$ in this case; $P^*\subseteq T$, and so $T\in\mathcal{B}_2$ in this case. Consequently, $\overline{\mathcal{T}}_{n,k}\subseteq\mathcal{B}_1\cup\mathcal{B}_2$, which in turn gives that

$$
|\overline{\mathcal{T}}_{n,k}|\le|\mathcal{B}_1|+|\mathcal{B}_2|. \tag{1}
$$

Let S_{v_1} denote the star centered at v_1 with the s_n leaves attached to it and S_w denote the star centered at w with the s_n leaves attached to it. Then $G_{n,k}$ is the union of four subgraphs S_{v_1}, S_w, H, and P^*.

- Considering the subtrees of S_{v_1} with at least two vertices and the subtrees of S_{v_1} with a single vertex, we get $|\mathcal{T}_{S_{v_1}}| = (2^{s_n} - 1) + (s_n + 1) = 2^{s_n} + s_n = 2^{s_n} + o(2^{s_n})$.
- Considering the subtrees of S_w with at least two vertices and the subtrees of S_w with a single vertex, we get $|\mathcal{T}_{S_w}| = (2^{s_n} - 1) + (s_n + 1) = 2^{s_n} + s_n = 2^{s_n} + o(2^{s_n})$.
- Considering the subpaths of P^* with at least two vertices and the subpaths of P^* with a single vertex, we get $|\mathcal{T}_{P^*}| = \binom{|P^*|}{2} + |P^*| = \binom{|P^*|+1}{2} = \binom{n-2s_n-k+2}{2} \leq \frac{n^2}{2}$.
- The number of subpaths of P^* containing w is bounded above by $|P^*| = n - 2s_n - k + 1 \leq n$.

Since $s_n = o(n)$, we have the following two inequalities

$$\begin{aligned}
|\mathcal{B}_1| &\leq (s_n + |\mathcal{T}_H| \cdot |\mathcal{T}_{P^*}| \cdot |\mathcal{T}_{S_w}|) + (s_n + |\mathcal{T}_H| \cdot |\mathcal{T}_{P^*}| \cdot |\mathcal{T}_{S_{v_1}}|) \\
&\leq 2\left[s_n + |\mathcal{T}_H| \cdot (2^{s_n} + o(2^{s_n})) \cdot \frac{n^2}{2}\right] = |\mathcal{T}_H| \cdot \left(2^{s_n} \cdot n^2 + o(2^{s_n} \cdot n^2)\right) \\
|\mathcal{B}_2| &\leq |\mathcal{T}_{S_{v_1}}| \cdot |\mathcal{T}_{S_w}| \cdot |P^*| \cdot |\mathcal{T}_H| = \left(2^{2s_n} \cdot n + o(2^{2s_n} \cdot n)\right) \cdot |\mathcal{T}_H|.
\end{aligned}$$

Recall that $n_1 = n - k + 1$. Applying Lemma 2.6, we have

$$|\mathcal{T}_{n,k}| = |\mathcal{T}'_H| \cdot |\mathcal{T}_{n_1,1}| = |\mathcal{T}'_H| \cdot 2^{2s_n}\binom{n_1 - 2s_n}{2} = |\mathcal{T}'_H| \cdot 2^{2s_n} \cdot \left(\frac{n^2}{2} - o(n^2)\right).$$

Recall that $2^{s_n} \geq n^2$. Since $|\mathcal{T}_H|$ is bounded by a function of k because $|H| = k+1$, we have the following two inequalities.

$$\lim_{n\to\infty} \frac{|\mathcal{B}_1|}{|\mathcal{T}_{n,k}|} = \lim_{n\to\infty} \frac{|\mathcal{T}_H| \cdot 2^{s_n} \cdot n^2}{|\mathcal{T}'_H| \cdot 2^{2s_n} \cdot \frac{n^2}{2}} = \lim_{n\to\infty} \frac{2|\mathcal{T}_H|}{|\mathcal{T}'_H| \cdot 2^{s_n}} = 0$$

and

$$\lim_{n\to\infty} \frac{|\mathcal{B}_2|}{|\mathcal{T}_{n,k}|} = \lim_{n\to\infty} \frac{2^{2s_n} \cdot n \cdot |\mathcal{T}_H|}{|\mathcal{T}'_H| \cdot 2^{2s_n} \cdot \frac{n^2}{2}} = \lim_{n\to\infty} \frac{2 \cdot |\mathcal{T}_H|}{|\mathcal{T}'_H| \cdot n} = 0.$$

Hence, we conclude that

$$\lim_{n\to\infty} \frac{|\mathcal{B}_1| + |\mathcal{B}_2|}{|\mathcal{T}_{n,k}|} = 0$$

Combining this with (1), we have that $\lim_{n\to\infty} |\overline{\mathcal{T}}_{n,k}|/|\mathcal{T}_{n,k}| = 0$, which completes the proof of Lemma 2.4. □

2.2 An Alternative Construction

The graphs we constructed in order to prove Theorem 1.2, and the sets of k edges that were added to them, are certainly not the only examples that could be used to prove Theorem 1.2. For example, the k-edge set $\{v_1w, v_2w, \ldots, v_kw\}$ can be replaced by the k-edge set $\{v_1v_{n-2s_n}, v_2v_{n-2s_n-1}, \ldots, v_kv_{n-2s_n-k+1}\}$.

Fix a positive integer k and let n be an integer much larger than k. We follow the notation given in Section 2. Recall that G_n is obtained from a path $P := v_1v_2\cdots v_{n-2s_n}$ by attaching two stars centered at v_1 and v_{n-2s_n}, and $\lim_{n\to\infty}\sigma(G_n) = 1$. Let $E_k := \{v_{i_1}v_{j_1}, v_{i_2}v_{j_2}, \ldots, v_{i_k}v_{j_k}\}$ be a set of k edges in $\overline{G_n}$ such that $1 \le i_1 < j_1 \le i_2 < j_2 \le \cdots \le i_k < j_k \le n - 2s_n$. Let $H_{n,k} = G_n + E_k$. For convenience, we assume that $j_\ell - i_\ell$ have the same value, say p, for $\ell \in \{1, \ldots, k\}$.

A simple calculation shows that for each path Q of order q, we have $\mu(Q) = (q+2)/3$ (See Jamison [5]), and so $\lim_{q\to\infty}\sigma(Q) = 1/3$. For any non-empty subset $F \subseteq E_k$, we define $\mathcal{T}_F = \{T \in \mathcal{T}_{H_{n,k}} : E(T) \cap E_k = F\}$. For any edge $v_{i_\ell}v_{j_\ell} \in F$, let $e_\ell = v_{i_\ell}v_{j_\ell}$ and $P_\ell = v_{i_\ell}v_{i_\ell+1}\cdots v_{j_\ell}$. Note that every tree $T \in \mathcal{T}_F$ is a union of a subtree of $H_{n,k} - \cup_{e_\ell\in F}(V(P_\ell)\backslash\{v_{i_\ell}, v_{j_\ell}\})$ containing F and $\cup_{e_\ell\in F}(E(P_\ell) - E(P_\ell^*))$ for some path $P_\ell^* \subseteq P_\ell$ containing at least one edge. Since $|E(P_\ell)| = p$, the line graph of P_ℓ is a path of order p. Consequently, the mean of $|E(P_\ell^*)|$ over subpaths of P_ℓ is $(p+2)/3$. Hence, the mean of $|E(P_\ell) - E(P_\ell^*)|$ over all subpaths P_ℓ^* of P_ℓ is $p - (p+2)/3 = 2(p-1)/3$ for each $e_\ell \in F$. Let $s = |F|$. Since every subtree $T \in \mathcal{T}_F$ has at most $n - s(p-1)$ vertices outside $\cup_{e_\ell\in F}(P_\ell - v_{i_\ell} - v_{j_\ell})$, we get the following inequality.

$$\mu(\mathcal{T}_F) \le n - s(p-1) + s\cdot\frac{2(p-1)}{3} \le n - \frac{s(p-1)}{3}.$$

By taking p as a linear value of n, say $p = \alpha n$ ($\alpha < \frac{1}{k}$), we get $\sigma(\mathcal{T}_F) \le 1 - s\alpha/3 + s/3n < \sigma(G_n)$ since we assume that n is much larger than k. Since $\mathcal{T}_{H_{n,k}} = \bigcup_{F\subseteq E_k}\mathcal{T}_F$, we have $\sigma(H_{n,k}) < \sigma(G_n)$, and so $\mu(H_{n,k}) < \mu(G_n)$.

3 Proof of Conjecture 1.3

To simplify notation, we let $G := K_m + nK_1$, where $V(G) = V(K_{m,n})$. Denote by A and B the two color classes of $K_{m,n}$ with $|A| = m$ and $|B| = n$, respectively. For each tree $T \subseteq G$, we have $E(T) \cap E(K_m) = \emptyset$ or $E(T) \cap E(K_m) \ne \emptyset$. This implies that the family of subtrees of G consists of the subtrees of $K_{m,n}$ and the subtrees sharing at least one edge with K_m. For each tree $T \subseteq G$, let $A(T) = V(T) \cap A$ and $B(T) = V(T) \cap B$. Then, $|T| = |A(T)| + |B(T)|$. Furthermore, let $B_2(T)$ and $B_{\ge 2}(T)$ be the sets of vertices $v \in B(T)$ such that $d_T(v) = 2$ and $d_T(v) \ge 2$, respectively. Clearly, $B_2(T) \subseteq B_{\ge 2}(T) \subseteq B(T)$. We define a subtree $T \in \mathcal{T}_G$ to be a *b-stem* if $B_{\geq 2}(T) = B(T)$, which means that $d_T(v) \geq 2$ for any $v \in B(T)$.

Let T be a b-stem and assume that T contains f edges in K_m. Counting the number of edges in T, we obtain $|E(T)| = f + \sum_{v \in B(T)} d_T(v)$. Since T is a tree, we have $|E(T)| = |T| - 1 = |A(T)| + |B(T)| - 1$. Therefore, we gain

$$|B(T)| = |A(T)| - 1 - \left(f + \sum_{v \in B(T)} (d_T(v) - 2) \right). \tag{2}$$

Since T is a b-stem, we have $\sum_{v \in B(T)} (d_T(v) - 2) \geq 0$, which implies that $|B(T)| \leq |A(T)| - 1 \leq m - 1$. Thus, $|T| = 2|A(T)| - \left(1 + f + \sum_{v \in B(T)} (d_T(v) - 2)\right) \leq 2|A(T)| - 1$. It follows that a b-stem $T \in \mathcal{T}_G$ is the *max b-stem*, i.e., the b-stem with the maximum order among all b-stems in $\mathcal{T}_G$, if and only if $A(T) = A$, $E(T) \cap E(K_m) = \emptyset$, and $B_2(T) = B_{\geq 2}(T)$. This is equivalent to saying that T is a max b-stem if and only if $|A(T)| = m$ and $|B(T)| = m - 1$.

The b-stem of a tree $T \subset G$ is the subgraph induced by $A(T) \cup B_{\geq 2}(T)$, and it is a subtree in $\mathcal{T}_G$. It is worth noting that the b-stem of every subtree $T \subset G$ exists, except for the case when T is a tree with only one vertex belonging to B. Conversely, given a b-stem T_0, a tree $T \subset G$ contains T_0 as its b-stem if and only if $T_0 \subseteq T$, $A(T) = A(T_0)$, and $B(T) \backslash B(T_0)$ is a set of vertices with degree 1 in T. Equivalently, T can be obtained from T_0 by adding vertices in $B(T) \backslash B(T_0)$ as leaves. So, there are exactly $(|A(T_0)| + 1)^{n - |B(T_0)|}$ trees containing T_0 as their b-stem.

For two non-negative integers a, b, let $\mathcal{T}_G(a, b)$ (resp. $\mathcal{T}_{K_{m,n}}(a, b)$) be the family of subtrees in $\mathcal{T}_G$ (resp. $\mathcal{T}_{K_{m,n}}$) whose b-stems T_0 satisfy $|A(T_0)| = a$ and $|B(T_0)| = b$. For any $A_0 \subseteq A$ and $B_0 \subseteq B$, let $f_G(A_0, B_0)$ (resp. $f_{K_{m,n}}(A_0, B_0)$) denote the number of b-stems T_0 spanned by $A_0 \cup B_0$; that is, $A(T_0) = A_0$ and $B_{\geq 2}(T_0) = B_0$. Clearly, $f_G(A_0, B_0)$ and $f_{K_{m,n}}(A_0, B_0)$ depend only on $|A_0|$ and $|B_0|$, so we can denote them by $f_G(|A_0|, |B_0|)$ and $f_{K_{m,n}}(|A_0|, |B_0|)$, respectively. By counting, we have $|\mathcal{T}_G(a, b)| = \binom{m}{a} \cdot \binom{n}{b} \cdot f_G(a, b) \cdot (a+1)^{n-b}$ and $|\mathcal{T}_{K_{m,n}}(a, b)| = \binom{m}{a} \cdot \binom{n}{b} \cdot f_{K_{m,n}}(a, b) \cdot (a+1)^{n-b}$, due to the fact that there are $\binom{m}{a}$ ways to pick an a-set in A and $\binom{n}{b}$ ways to pick a b-set in B. Since $a \leq m$ and $b \leq m - 1$, there exist positive numbers c_1 and c_2 that depend only on m, such that

$$c_1 n^b (a+1)^{n-b} \leq |\mathcal{T}_G(a, b)| \leq c_2 n^b (a+1)^{n-b} \tag{3}$$

Note that if $(a, b) \neq (m, m-1)$, then we have $b \leq m - 2$. Applying inequality (3), we get $|\cup_{(a,b) \neq (m,m-1)} \mathcal{T}_G(a, b)| \leq c_3 |\mathcal{T}_G(m, m-1)|/n$ for some constant $c_3 > 0$ depending only on m.

Given a b-stem T_0 with $|A(T_0)| = a$ and $|B(T_0)| = b$, let T be a tree chosen uniformly at random from $\mathcal{T}_G$ (resp. $\mathcal{T}_{K_{m,n}}$) that contains T_0 as its b-stem. Then, the probability of a vertex $v \in B \backslash B(T_0)$ in T is $\frac{a}{a+1}$. This shows that the mean order of trees containing T_0 as their b-stem is $(n - b)\frac{a}{a+1} + a + b$, denoted by $\mu(a, b)$. Note that $\sum_{T \in \mathcal{T}_G(a,b)} |T| = \mu(a, b) \cdot |\mathcal{T}_G(a, b)|$

and $\sum_{T\in\mathcal{T}_{K_{m,n}}(a,b)}|T| = \mu(a,b)\cdot|\mathcal{T}_{K_{m,n}}(a,b)|$. Assume that T_0 has f edges in K_m, and set $c = \sum_{v\in B(T_0)}(d_{T_0}(v)-2)$. Using (2), we have $b = a-(1+f+c)$. Hence, $\mu(a,b) = \frac{(n+2+a)\cdot a}{a+1} - \frac{1+f+c}{a+1}$, which reaches its maximum value when $a = m$ and $f = c = 0$, i.e., when T_0 is a max b-stem. We then have:

$$\mu(G) = \frac{\mu(m,m-1)|\mathcal{T}_G(m,m-1)| + \sum_{(a,b)\neq(m,m-1)}\mu(a,b)|\mathcal{T}_G(a,b)| + n}{|\mathcal{T}_G(m,m-1)| + \sum_{(a,b)\neq(m,m-1)}|\mathcal{T}_G(a,b)|},$$

$$\mu(K_{m,n}) = \frac{\mu(m,m-1)|\mathcal{T}_{K_{m,n}}(m,m-1)| + \sum_{(a,b)\neq(m,m-1)}\mu(a,b)|\mathcal{T}_{K_{m,n}}(a,b)| + n}{|\mathcal{T}_{K_{m,n}}(m,m-1)| + \sum_{(a,b)\neq(m,m-1)}|\mathcal{T}_{K_{m,n}}(a,b)|},$$

where n denotes the number of subtrees with a single vertex in B.

Note that $|\mathcal{T}_G(a,b)| \geq |\mathcal{T}_{K_{m,n}}(a,b)|$, with equality holding if and only if $a = b-1$, and so in particular when $(a,b) = (m,m-1)$. We have derived before that $0 < \mu(a,b) < \mu(m,m-1)$ when $(a,b) \neq (m,m-1)$. Using the inequality $|\cup_{(a,b)\neq(m,m-1)}\mathcal{T}_G(a,b)| \leq c_3|\mathcal{T}_G(m,m-1)|/n$, we conclude that $\mu(G) > \frac{n}{n+c_3}\mu(m,m-1) > \max_{(a,b)\neq(m,m-1)}\mu(a,b)$ for n sufficiently large (for fixed m).

Since $\mu(K_{m,n})$ is the average of the same terms, as well as some additional terms of the form $\mu(a,b)$, which are smaller than $\mu(G)$, we conclude that $\mu(G) < \mu(K_{m,n})$. This completes the proof. □

Acknowledgments

We would like to express our sincere gratitude to the two anonymous referees for their valuable comments and suggestions that improved this manuscript.

References

[1] Stijn Cambie, Stephan Wagner, and Hua Wang. On the maximum mean subtree order of trees. *European Journal of Combinatorics*, 97:103388, 2021.

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

[3] Alex J. Chin, Gary Gordon, Kellie J. MacPhee, and Charles Vincent. Subtrees of graphs. *J. Graph Theory*, 89(4):413–438, 2018.

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

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

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

[7] Zuwen Luo, Kexiang Xu, Stephan Wagner, and Hua Wang. On the mean subtree order of trees under edge contraction. *Journal of Graph Theory*, 102(3):535–551, 2023.

[8] Lucas Mol and Ortrud R. Oellermann. Maximizing the mean subtree order. *J. Graph Theory*, 91(4):326–352, 2019.

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

[10] Stephan Wagner and Hua Wang. Indistinguishable trees and graphs. *Graphs Combin.*, 30(6):1593–1605, 2014.

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