On different "middle parts" of a tree

Heather Smith[*]

School of Mathematics
Georgia Institute of Technology
Atlanta, Georgia, U.S.A.

heather.smith@math.gatech.edu

László Székely[†]

Department of Mathematics
University of South Carolina
Columbia, South Carolina, U.S.A.

szekely@math.sc.edu

Hua Wang[‡]

Department of Mathematical Sciences
Georgia Southern University
Statesboro, Georgia, U.S.A.

hwang@georgiasouthern.edu

Shuai Yuan

Department of Mathematics
University of South Carolina
Columbia, South Carolina, U.S.A.

syuan@math.sc.edu

Abstract

We determine the maximum distance between any two of the center, centroid, and subtree core among trees with a given order. Corresponding results are obtained for trees with given maximum degree and also for trees with given diameter. The problem of the maximum distance between the centroid and the subtree core among trees with given order and diameter becomes difficult. It can be solved in terms of the problem of minimizing the number of root-containing subtrees in a rooted tree of given order and height. While the latter problem remains unsolved, we provide a partial characterization of the extremal structure.

Mathematics Subject Classifications: 05C05, 05C12, 05C35

1 Preliminaries

Many real-valued functions defined on the vertex set of a tree have been studied in the literature. Such functions are closely related to graph invariants motivated from practical

---

applications (such as the Wiener index in biochemistry [11]) or pure mathematical study (such as the subtree density [4]). In this paper we are interested in the *eccentricity* and *distance of a vertex*, as well as the *number of subtrees containing a vertex*. Both eccentricity and distance of a vertex are minimized at one vertex or two adjacent vertices. These vertices define the *center* and *centroid* respectively of the tree. The number of subtrees containing a vertex is maximized by one vertex or two adjacent vertices, called the *subtree core* of the tree.

The study of the center and centroid (Definitions 1.1 and 1.2) can be traced back to [5]. This paper focuses on the three different middle parts of the tree: the center, the centroid, and the subtree core. We investigate the geometry of their possible locations, in particular the extremal problem of how far they can be from each other. The resulting three problems are solved in Section 2 for arbitrary trees of fixed order, and the corresponding extremal structures are also characterized. In Section 3, the same three problems are solved for degree bounded trees and conjectures are proposed for the case of binary trees of fixed order. In Section 4, we consider the same questions for trees with fixed order and bounded diameter. In this setting, the maximum distance between the center and the centroid and between the center and the subtree core are determined, but the problem of the maximum distance between the centroid and the subtree core becomes challenging. It can be solved in terms of the problem of minimizing the number of root-containing subtrees in a rooted tree of given order and height. While this latter problem remains unsolved, in Section 5 we provide a partial characterization of the extremal structure.

Let us begin with formal definitions. The distance in the tree from u to v, denoted $d(u, v)$, is the number of edges on their unique connecting path $P(u, v)$.

Definition 1.1. The *eccentricity* of a vertex v in a tree T is

$$ecc_T(v) = \max_{u \in V(T)} d(v, u).$$

The *center* of T, denoted $C(T)$, is the set of vertices which have the minimum eccentricity among all vertices in the tree.

Definition 1.2. The *distance* of a vertex v in a tree T, denoted $d_T(v)$, is

$$d_T(v) = \sum_{u \in V(T)} d(v, u).$$

The *centroid* of T, denoted $CT(T)$, is the set of vertices which have the minimum distance among all vertices in the tree.

A *subtree* of tree T is a connected subgraph which is induced on a nonempty set of vertices. We consider T to be a subtree of itself and a single vertex is also a subtree of T.

Definition 1.3. As the name suggests, the *number of subtrees* of a vertex v in a tree T, denoted $F_T(v)$, is the number of subtrees of T which contain v. The *subtree core* of a tree T, denoted $Core(T)$, is the set of vertices that maximize the function $F_T(.)$ [8].

If H is a forest and v is a vertex in H, then $F_H(v)$ will be defined, as above, to be the number of subtrees of H which contain vertex v. In particular, all subtrees which are counted must be subtrees of the component of H which contains vertex v.

Jordan [5] found that $C(T)$ consists of either one vertex or two adjacent vertices (see also Ex. 6.21a in [6]). Given the vertices along any path of a tree, the sequence of $F_T(.)$ function values is strictly concave down ([8]), the sequence of $d(.)$ function values are strictly concave up (Ex. 6.22 in [6]; [3]), and the sequence of $ecc_T(.)$ function values are concave up (Ex. 6.21 in [6]). Strict concavity immediately implies that the sets $CT(T)$ and $Core(T)$ either consist of one vertex or two adjacent vertices.

We are specifically interested in how the middle sets are located, relative to one another. It is well-known that $C(T)$ and $CT(T)$ can be far apart (Ex. 6.22c in [6]), and that $Core(T)$ can differ from them [8].

There are some natural questions that we will explore. How close to each other can they be? How far apart can they be spread? Must they lie on a common path? Can they appear in any ordering?

It is easy to find trees where $C(T)$, $CT(T)$, and $Core(T)$ coincide, like the star and paths of even length to name a few. It is more interesting to see how far apart these middle sets can be in a single tree.

Considering one vertex from each of $C(T)$, $CT(T)$, and $Core(T)$, any two of these must lie on a common path. However, it is possible that the vertices from $C(T)$, $CT(T)$, and $Core(T)$ in the same tree T do not all lie on a common path. Figure 1 provides an example of this very situation.
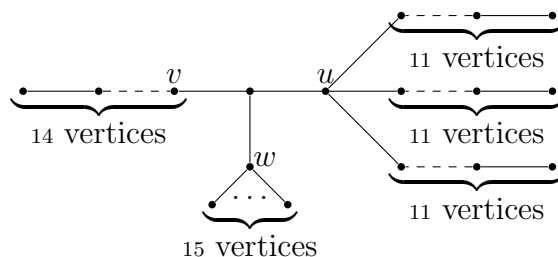

Figure 1: A tree with $v \in C(T)$, $u \in CT(T)$, $w \in Core(T)$ which do not lie on a common path. Further, all of these middle parts are singletons.

On the other hand, when the vertices of $C(T)$, $CT(T)$, $Core(T)$ happen to lie on the same path, they can appear in any order. Figure 2 provides some illustrations.

Among the examples with different ordering of middle vertices, it is interesting to observe that vertices in $Core(T)$ often have large degree; vertices in $C(T)$ often have small degree; while vertices in $CT(T)$ behave somewhat between the previous two.

Before proceeding to examining the largest distances between different middle parts, we formalize some necessary and sufficient conditions for a vertex to be in a middle part. Although not all of them are formally stated in the literature, we leave their relatively straightforward proofs to the reader.
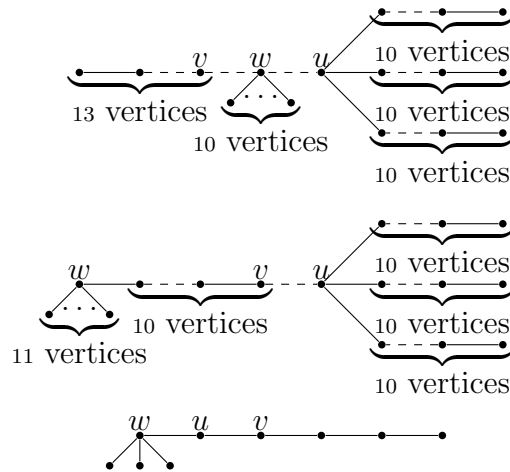
Figure 2: Trees with vertices $v \in C(T)$, $u \in CT(T)$, $w \in Core(T)$ on a common path, but in different orders. Here each middle part is a singleton.

Proposition 1.4. *Let T be a tree with at least two vertices. A vertex v is in the center $C(T)$ if and only if there are two leaves, u and w, such that $P(v,u) \cap P(v,w) = \{v\}$, $d(v,u) = ecc_T(v)$, and $d(v,w) \geqslant ecc_T(v) - 1$.*

Corollary 1.5. *If there are two leaves u, w such that $d(v,w) = d(v,u) = ecc_T(v)$, then $C(T) = \{v\}$. If no such w exists, then $|C(T)| = 2$ where the neighbor of v on $P(u,v)$ is also in the center.*

Next we give a characterization of the vertices in $CT(T)$. Note that if uv is an edge in tree T, then $T - uv$ will denote the forest that results after the deletion of edge uv from T.

Proposition 1.6. *Let T be a tree with at least two vertices. A vertex u is in the centroid $CT(T)$ if and only if for each neighbor v of u, we have*

$$n_{uv}(v) \leqslant n_{uv}(u)$$

where $n_{uv}(u)$ ($n_{uv}(v)$) denotes the number of vertices in the component containing u (v) in $T - uv$. Furthermore, if $u \in CT(T)$ and equality holds above, then $v \in CT(T)$ as well.

Lastly, Proposition 1.7 gives a characterization of $Core(T)$.

Proposition 1.7. *A vertex u is in $Core(T)$ if and only if for each neighbor v of u, we have*

$$F_{T-uv}(u) \geqslant F_{T-uv}(v).$$

Furthermore, if $u \in Core(T)$ and equality holds above, then $v \in Core(T)$ as well.

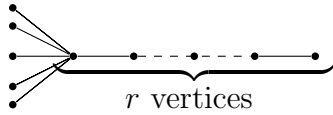For completeness, we provide a proof for the following simple fact.

Figure 3: An r-comet of order n.

Claim 1.8. *Among rooted trees of order n, the number of subtrees containing the root is at most 2^{n-1}, achieved only by the star rooted at the center; and at least n, achieved only by the path rooted at one end vertex.*

Proof. We proceed by induction on n. Let T be a tree of order n with root ρ and let ρ be of degree k with neighbors $v_1, \ldots, v_k$. Denote by T_i the connected component containing v_i in $T - \{\rho\}$ and let $n_i = |V(T_i)|$.

Then
$$F_T(r) = \prod_{i=1}^{k} \left(1 + F_{T_i}(v_i)\right) \leqslant \prod_{i=1}^{k} \left(1 + 2^{n_i-1}\right) \leqslant \prod_{i=1}^{k} 2^{n_i} = 2^{n-1},$$

where the first inequality follows from induction hypothesis and equality holds in the second inequality if and only if $n_i = 1$ for all i (and consequenly $k = n - 1$); and

$$F_T(r) = \prod_{i=1}^{k} \left(1 + F_{T_i}(v_i)\right) \geqslant \prod_{i=1}^{k} \left(1 + n_i\right) \geqslant 1 + \sum_{i=1}^{k} n_i = n,$$

where the first inequality follows from induction hypothesis and equality holds in the second inequality if and only if $k = 1$ and T_1 is a path with v_1 as a leaf. $\qquad \square$

2 Maximum distances between middle parts in general trees

In a tree T, the distance $d(S, S')$ between vertex sets S and S' is defined as the Hausdorff distance $\min\{d(u, v) : u \in S, v \in S'\}$. Fix an arbitrary $n \in \mathbb{Z}^+$. Among all trees with n vertices, we determine the maximum distance that can be realized between the center, centroid, and subtree core. We will also see that these maximum distances are achieved precisely when T has a "comet" structure.

Definition 2.1 (Barefoot, Entringer, Székely [2]). *An r-comet of order n is formed by attaching $n - r$ pendant vertices to one end vertex of a path on r vertices (Figure 3).*

2.1 Between center and centroid

Theorem 2.2. *Fix an arbitrary $n \in \mathbb{Z}^+$. For any tree T with $n \geqslant 3$ vertices,*

$$d(C(T), CT(T)) \leqslant \left\lfloor \frac{n-3}{4} \right\rfloor. \tag{1}$$

Proof. Fix a tree T on n vertices. Let $v \in C(T)$ and $u \in CT(T)$ such that the graph distance between u and v is precisely $d(C(T), CT(T))$. We assume $d(u, v) \geqslant 1$, otherwise we have nothing to prove. By the choice of u and v, no vertex on the path $P(u, v)$ other than u and v is in the center or the centroid of T.

Let $P(u, v)$ denote the path connecting u and v and let T_u denote the component containing u in $T - E(P(u, v))$. By Proposition 1.6,

$$|V(T_u)| > n - |V(T_u)|.$$

This implies

$$|V(T_u)| > \frac{n}{2}.$$

Let w be a leaf such that $P(v, w)$ and $P(u, v)$ are disjoint, except for v, and the length of $P(v, w)$ is maximum. Because $v \in C(T)$ and the neighbor of v on $P(u, v)$ is not in $C(T)$, Proposition 1.4 tells

$$d(v, w) = ecc_T(v).$$

As $n \geqslant 3$, u is not a leaf. Hence it is easy to see that

$$d(u, v) \leqslant ecc_T(v) - 1.$$

Therefore, we have

$$\frac{n}{2} > n - |V(T_u)| \geqslant d(u, v) + d(v, w) \geqslant 2d(u, v) + 1.$$

This implies

$$d(u, v) < \frac{n - 2}{4}.$$

In particular, if $n = 4k + r$ with $r \in \{0, 1, 2\}$, then

$$k - \frac{1}{2} \leqslant \frac{4k + r - 2}{4} \leqslant k.$$

Since $d(u, v) < \frac{n-2}{4}$, when $n \equiv r \mod 4$ for $r \in \{0, 1, 2\}$, $d(u, v) \leqslant k - 1$ where $k = \lfloor \frac{n}{4} \rfloor$.

When $n = 4k + 3$,

$$d(u, v) < \frac{n - 2}{4} = \frac{4k + 1}{4} = k + \frac{1}{4}.$$

As a result, $d(u, v) \leqslant k$. $\qquad\square$

Proposition 2.3. *Let $k := \lfloor \frac{n}{4} \rfloor$. Equality holds in (1) exactly when n and T fall into one of the following categories:*

- $n = 4k$ *and T is the $2k$-comet.*

- $n = 4k + 1$ *or $n = 4k + 2$ and T is one of the following trees:*

  - *$2k$-comet*

- 2k-comet on $n-1$ vertices together with one vertex pendant to one of the degree 2 vertices on the path of the comet

- a tree consisting of a path on $2k+1$ vertices with an end vertex identified with the root of a height 2 tree

- $n = 4k + 3$ and T is a $(2k + 2)$-comet.

Proof. As in the proof of Theorem 2.2, let $v \in C(T)$ and $u \in CT(T)$ such that the graph distance between u and v is precisely $d(C(T), CT(T))$. Above, we obtained $d(u, v) \leqslant \left\lfloor \frac{n-3}{4} \right\rfloor$ from

$$\left\lfloor \frac{n-1}{2} \right\rfloor \geqslant n - |V(T_u)| \tag{2}$$

$$\geqslant d(u, v) + d(v, w) \tag{3}$$

$$\geqslant 2d(u, v) + 1. \tag{4}$$

If $\left\lfloor \frac{n-1}{2} \right\rfloor$ is odd, then $d(u, v) = \left\lfloor \frac{n-3}{4} \right\rfloor$ when all of the inequalities above are tight. Equality in (3) implies that all vertices not in T_u form a path $P(u, w)$. Equality in (4) implies that the height of T_u is 1 because the neighbor of v on $P(u, v)$ is not in $C(T)$. This, together with equality in (2) characterizes the $\left\lfloor \frac{n+1}{2} \right\rfloor$-comet. In particular, when $n = 4k$, this is the 2k-comet, and when $n = 4k + 3$, this is the $2k + 2$-comet.

On the other hand, notice that $2d(u, v) + 1$ is odd. So when $\left\lfloor \frac{n-1}{2} \right\rfloor$ is even, all inequalities cannot be equalities. In particular, exactly one will be strict. If (2) is the one which is strict, then we have the $\left\lfloor \frac{n-1}{2} \right\rfloor$-comet. If (3) is the one which is strict, then we have a $\left\lfloor \frac{n+1}{2} \right\rfloor$-comet on $n - 1$ vertices with one extra vertex pendant to one of the degree 2 vertices. If (4) is the one which is not strict, then T_u has height 2 but all $\left\lfloor \frac{n-1}{2} \right\rfloor$ vertices not in T_u still lie on the path $P(u, w)$. $\square$

2.2 Between centroid and subtree core

Next we turn our attention to the centroid and the subtree core.

Theorem 2.4. *Let T be a tree with $n > 8$ vertices. If $n \geqslant 2^{\lceil \log_2 n \rceil - 1} + \lceil \log_2 n \rceil$, then*

$$d(CT(T), Core(T)) \leqslant \left\lfloor \frac{n-1}{2} \right\rfloor - \lfloor \log_2 n \rfloor - 1$$

with equality holding if and only if T is the $(n - \lfloor \log_2 n \rfloor - 1)$-comet. Otherwise

$$d(CT(T), Core(T)) \leqslant \left\lfloor \frac{n-1}{2} \right\rfloor - \lfloor \log_2 n \rfloor.$$

with equality precisely when T is the $(n - \lfloor \log_2 n \rfloor)$-comet.

Proof. Fix a tree T with $n > 8$ vertices. Let $u \in CT(T)$ and $v \in Core(T)$ where the graph distance between u and v is precisely $d(CT(T), Core(T))$. We assume $d(u,v) \geqslant 1$, otherwise we have nothing to prove. Let $P(u,v)$ denote the path connecting u and v and let T_u, T_v denote the components containing u, v respectively in $T - E(P(u,v))$. Let $T - T_v$ be the component containing v when the edges of T_v are deleted from T. Set $x := |V(T_u)|$ and $y := |V(T_v)|$. First observe

$$d(u,v) \leqslant n - x - y + 1.$$

Since $u \in CT(T)$ and the neighbor of u on $P(u,v)$ is not in $CT(T)$, Proposition 1.6 implies $x > n - x$, thus $x > \frac{n}{2}$. Since x is an integer, $x \geqslant \frac{n+1}{2}$ and consequently

$$x \geqslant \left\lceil \frac{n+1}{2} \right\rceil.$$

Similarly, since $v \in Core(T)$ and the neighbor of v on $P(u,v)$ is not in $Core(T)$, Proposition 1.7 gives

$$F_{T_v}(v) > F_{T-T_v}(w)$$

where w is the unique neighbor of v on $P(u,v)$. See Figure 4 for an illustration of how these pieces interact.
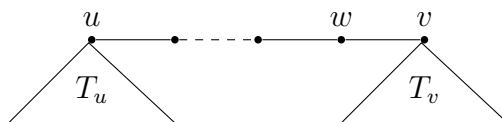

Figure 4: A representation of tree T for the proof of Theorem 2.4 with path $P(u,v)$, T_u, T_v, and w labeled.

Further note that every subtree in T_v which contains v can be uniquely identified by the set of its vertices, excluding v. Thus, by Claim 1.8,

$$F_{T_v}(v) \leqslant 2^{y-1}.$$

Note that equality holds if and only if every subset of vertices induces a tree which is the case exactly when T_v is a star centered at v. On the other hand,

$$F_{T-T_v}(w) \geqslant n - y$$

with equality if and only if $T - T_v$ is a path with w as an end vertex.

Putting these inequalities together for our specific choice of v, Proposition 1.7 yields

$$2^{y-1} > n - y.$$

As $y + 2^{y-1}$ is an increasing function, there is a unique, real-valued, $y_0 > 0$ such that $2^{y_0-1} = n - y_0$. Further, for all $0 < y < y_0$, $2^{y-1} < n - y$, and for all $y > y_0$, $2^{y-1} > n - y$.

Now let $y_0 > 0$, be the unique real value such that $2^{y_0-1} = n - y_0$. Then

$$
\begin{aligned}
y_0 &= \log_2(n - y_0) + 1 \\
&< \log_2(n) + 1.
\end{aligned}
$$

Substituting the equation $y_0 = \log_2(n - y_0) + 1$ into itself, for $n > 8$ we find

$$
\begin{aligned}
y_0 &= \log_2(n - \log_2(n - y_0) - 1) + 1 \\
&\geqslant \log_2(n - \log_2(n) - 1) + 1 \\
&= \log_2 2(n - \log_2(n) - 1) \\
&= \log_2(n + (n - 2 - 2\log_2 n)) \\
&> \log_2(n).
\end{aligned}
$$

In the last inequality, $n - 2 - 2\log_2 n > 0$ holds for $n > 8$.

As a result, we have the bounds

$$
\log_2(n) < y_0 < \log_2(n) + 1.
$$

Further, if $y_0 < \lfloor \log_2 n \rfloor + 1$, then for an integer $y > 0$, $2^y - 1 > n - y$ precisely when $y \geqslant \lfloor \log_2 n \rfloor + 1$. However, if $y_0 \geqslant \lfloor \log_2 n \rfloor + 1$ then for an integer $y > 0$, $2^y - 1 > n - y$ precisely when $y \geqslant \lfloor \log_2 n \rfloor + 2$.

When $n > 8$, our bounds for integers x and y give

$$
\begin{aligned}
d(u, v) &\leqslant n - x - y + 1 \\
&\leqslant n - \left\lceil \frac{n+1}{2} \right\rceil - \lfloor \log_2 n \rfloor \\
&= \left\lfloor \frac{n-1}{2} \right\rfloor - \lfloor \log_2 n \rfloor.
\end{aligned}
$$

As mentioned earlier, this can be strengthen to $d(u, v) \leqslant \lfloor \frac{n-1}{2} \rfloor - \lfloor \log_2 n \rfloor - 1$ if $y_0 \geqslant \lfloor \log_2 n \rfloor + 1$. However, this will only happen if $2^{\lfloor \log_2 n \rfloor} \leqslant n - \lfloor \log_2 n \rfloor - 1$ as stated in the theorem.

As for extremal trees, equality will hold in the upper bound for $d(u, v)$ exactly when T_u has $\left\lceil \frac{n+1}{2} \right\rceil$ vertices, T_v is a star, and $T - T_v$ is a path. This describes the C-comet where $C = n - \lfloor \log_2 n \rfloor$ or in the case where $n \geqslant 2^{\lfloor \log_2 n \rfloor} + \lceil \log_2 n \rceil + 1$, $C = n - \lfloor \log_2 n \rfloor - 1$. $\quad\square$

2.3 Between subtree core and center

The study of this case is similar to that in the previous section. We omit some details.

Theorem 2.5. *For any tree T on $n > 8$ vertices, if $n \geqslant 2^{\lceil \log_2 n \rceil - 1} + \lceil \log_2 n \rceil$ then*

$$
d(C(T), Core(T)) \leqslant \left\lfloor \frac{1}{2}(n - \lfloor \log_2 n \rfloor - 2) \right\rfloor
$$

which is tight for the K-comet with $K = n - \lfloor \log_2 n \rfloor + 1$. Otherwise

$$d(C(T), Core(T)) \leqslant \left\lfloor \frac{1}{2}(n - \lfloor \log_2 n \rfloor - 1) \right\rfloor$$

which is tight for the K-comet with $K = n - \lfloor \log_2 n \rfloor$.

Proof. Let $u \in Core(T)$ and $v \in C(T)$ in a tree T with $|V(T)| = n$ and the graph distance between u and v is precisely $d(C(T), Core(T))$. Use T_u (respectively T_v) to denote the component containing u (v) in $T - E(P(u,v))$ and let $y = |V(T_u)|$.

Because $v \in C(T)$ and the neighbor of v on $P(u,v)$ is not in $C(T)$, there is a leaf w in T_v with $d(v,w) = ecc_T(v)$. As argued in the proof of Theorem 2.2,

$$d(u,v) \leqslant ecc_T(v) - 1 < d(v,w),$$

$$2d(u,v) + 1 \leqslant d(u,v) + d(v,w) \leqslant n - y.$$

Note that these inequalities are tight for the $(n - y + 1)$-comet.

Because $u \in Core(T)$, we can conclude, as in the proof of Theorem 2.4,

$$2^{y-1} > n - y.$$

Consequently,

$$y \geqslant \lfloor \log_2 n \rfloor.$$

Combining inequalities, we obtain the bound in the theorem statement:

$$d(u,v) \leqslant \left\lfloor \frac{1}{2}(n - y - 1) \right\rfloor \leqslant \left\lfloor \frac{1}{2}(n - \lfloor \log_2 n \rfloor - 1) \right\rfloor.$$

Recall from Theorem 2.4 that if $n \geqslant 2^{\lceil \log_2 n \rceil - 1} + \lceil \log_2 n \rceil$, then

$$y \geqslant \lfloor \log_2 n \rfloor + 1$$

and consequently we obtain the slightly better bound

$$d(u,v) \leqslant \left\lfloor \frac{1}{2}(n - y - 1) \right\rfloor \leqslant \left\lfloor \frac{1}{2}(n - \lfloor \log_2 n \rfloor - 2) \right\rfloor. \qquad \square$$

3   Trees with degree restrictions

In Section 2, we saw that, for each pair of middle parts, the maximum distance was achieved precisely by an appropriate comet. However, the comet has a vertex of large degree. In this section, we restrict the maximum degree of the tree and ask how the extremal structures change. We will begin with a discussion of binary trees and then broaden our scope to trees with maximum degree $k > 3$.

First we state some results about the maximum or minimum number of root-containing subtrees in a tree with a specified degree sequence which will be needed later. Note that among trees (with no maximum degree condition) with n vertices, the number of root-containing subtrees is minimized by the path, rooted at one end, and maximized by the star, rooted at the center vertex (Claim 1.8).

3.1 Trees with a given degree sequence

For a rooted tree, the *height* of a vertex is the distance to the root. The *height of the tree*, $h(T)$, is the maximum of all vertex heights.

Definition 3.1. In a rooted tree T, the list of multisets $(L_0, L_1, \ldots, L_{h(T)})$, where L_i consists of the degrees of the vertices at height i (and L_0 consists of the degree of the root vertex), is called the *level-degree sequence* of the rooted tree.

Let $|L_i|$ be the number of entries in L_i counted with multiplicity. It is easy to see that a list of multisets is the level degree sequence of a rooted tree if and only if (i) the multiset $\bigcup_i L_i$ is a degree sequence for a tree, (ii) $|L_0| = 1$, and (iii) $\sum_{d \in L_0} d = |L_1|$, while $\sum_{d \in L_i}(d-1) = |L_{i+1}|$ for all $i \geqslant 1$.

In a rooted tree, the *down-degree* of the root is equal to its degree. The down degree of any other vertex is one less than its degree.

Definition 3.2. [7] Given the level-degree sequence of a rooted tree, the *level-greedy rooted tree* for this level-degree sequence is built as follows: (i) For each $i \in [n]$, place $|L_i|$ vertices in level i and to each vertex, from left to right, assign a degree from L_i in non-increasing order. (ii) For $i \in [n-1]$, from left to right, join the next vertex in L_i whose down-degree is d to the first d so far unconnected vertices on level L_{i+1}. Repeat for $i + 1$.

Definition 3.3. [10] Given a tree degree sequence $(d_1, d_2, \ldots, d_n)$ in non-increasing order, the *greedy tree* for this degree sequence is the level-greedy tree for the level-degree sequence that has $L_0 = \{d_1\}$, $L_1 = \{d_2, \ldots, d_{d_1+1}\}$ and for each $i > 1$,

$$|L_i| = \sum_{d \in L_{i-1}} (d-1)$$

with every entry in L_i at most as large as every entry in L_{i-1}.

The greedy tree frequently occurs in the study of extremal structures. A similar structure with modified root degree is crucial to our study here. Fix a degree sequence for a tree and distinguish a single value in this sequence which will be the degree of the root. Similar to the greedy tree, we define the *rooted greedy tree*.

Definition 3.4. Let $\overline{d} = (d_1, d_2, \ldots, d_n)$ be a tree degree sequence in non-increasing order with degree d_i identified as the root degree. Let

$$\hat{d} = (d'_1, d'_2, \ldots, d'_{n-1}) = (d_1, d_2, \ldots, \hat{d}_i, \ldots, d_n)$$

be the sequence $\overline{d}$ with d_i removed. The *rooted greedy tree* for the degree sequence $\overline{d}$ is the level-greedy tree for the level-degree sequence that has $L_0 = \{d_i\}$, $L_1 = \{d'_1, \ldots, d'_{d_i}\}$ and, for each $i > 1$, $|L_j| = \sum_{d \in L_{j-1}}(d-1)$, where entries of L_j are the next available elements from $\hat{d}$.
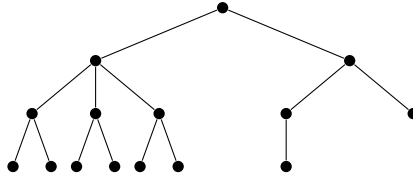
Figure 5: A rooted greedy tree with root degree 2 and degree sequence $(4, 3, 3, 3, 3, 2, 2, 1, \ldots, 1)$.

Among trees with given degree sequence, greedy trees are extremal with respect to many graph invariants. For example, the following result is for root-containing subtrees.

Theorem 3.5 (Andriantiana, Wagner, Wang [1])**.** *Fix a degree sequence $\overline{d}$ and a positive integer k. Among rooted trees with degree sequence $\overline{d}$, the greedy tree maximizes the number of subtrees with exactly k vertices which contain the root. Consequently the greedy tree maximizes the total number of root-containing subtrees.*

Fix a degree sequence, distinguish one value in the sequence as the root degree, and fix a positive integer k'. Among rooted trees with this degree sequence and the specified root degree, the corresponding rooted greedy tree maximizes the number of subtrees containing the root and which have k' vertices. Consequently, the rooted greedy tree maximizes the total number of root-containing subtrees.

3.2 Binary trees

The study of binary trees is well motivated from its applications in phylogeny. A *binary tree* is a tree in which every vertex has degree 1 or 3. A *rooted binary tree* is a rooted tree in which the root has degree 2 and all other vertices have degree 1 or 3. Székely and Wang [8] studied the number of subtrees of a binary tree with labeled vertices. They found that the extremal structures are *good trees*, *rgood trees*, and *caterpillars*. In our terms, a good binary tree is a greedy tree with root degree 3 and degree sequence

$$(3, \ldots, 3, 1, \ldots, 1)$$

and an rgood binary tree is a rooted greedy tree with root degree 2 and degree sequence

$$(3, \ldots, 3, 2, 1, \ldots, 1).$$

A binary caterpillar consists of a path P with pendant vertices that make the degree of each internal vertex 3.

Their results for the number of subtrees are as follows:

Theorem 3.6 (Székely, Wang [9])**.** *Among all binary trees with n leaves, the good binary tree minimizes the number of subtrees.*

Theorem 3.7 (Székely, Wang [8])**.** *Among all binary trees with n leaves, the binary caterpillar on n leaves minimizes the number of subtrees.*

As an immediate consequence of Theorem 3.5, we obtain the following results for root-containing subtrees.

Corollary 3.8. *Fix $n \in \mathbb{Z}^+$. Among all binary trees with n vertices and any choice of the root, the good binary tree (with the default root of the corresponding greedy structure) has the maximum number of root-containing subtrees.*

Corollary 3.9. *Fix $n \in \mathbb{Z}^+$. Among all rooted binary trees with n vertices, the rgood binary tree is the unique tree that maximizes the number of root-containing subtrees.*

For binary trees, we can examine the distance between vertices of different middle parts in much the same way that we did in Section 2. While the exact calculations are quite messy, we believe the following is true.

Conjecture 3.10. *The tree T that maximizes $d(CT(T), C(T))$ among all binary trees of order n, is formed by identifying the root of an rgood binary tree with a vertex of maximum eccentricity in a binary caterpillar (see Figure 6). The same tree structure maximizes $d(Core(T), CT(T))$ as well as $d(Core(T), C(T))$*
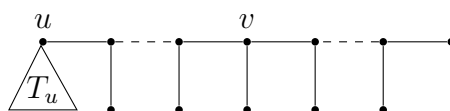

Figure 6: An extremal binary tree structure which is conjectured to maximize the distances $d(CT(T), C(T))$, $d(Core(T), CT(T))$, and $d(Core(T), C(T))$ for u and v as in Conjecture 3.10. The tree T_u is an rgood binary tree.

3.3 Trees with bounded maximum degree

We now turn our focus to trees on n vertices, all of which have degree at most k. We previously defined good binary trees and rgood binary trees. In general, for each positive integer k, a *good tree* is a greedy tree with degree sequence

$$(k, k, \ldots, k, 1, 1, \ldots 1)$$

while the *rgood trees* are rooted greedy trees with root degree $k - 1$ and degree sequence

$$(k, k, \ldots, k, k - 1, 1, 1, \ldots 1).$$

For any fixed k, these trees only exist for certain values of n. Therefore, we extend their definitions as follows so that we can create similar trees for any $n > k$.

For positive integers n, k ($n > k$), a tree with order n and maximum degree k is called an *extended good tree* if it is a greedy tree with degree sequence

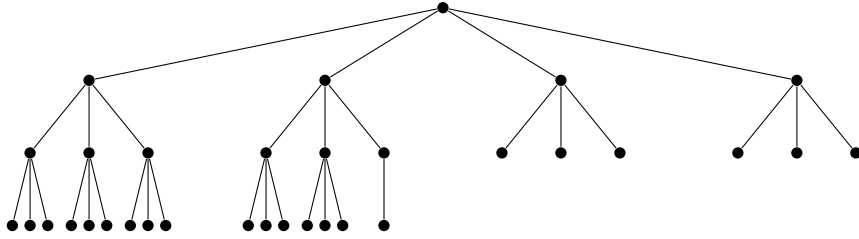$$(k, k, \ldots, k, s, 1, \ldots, 1)$$

Figure 7: An extended good tree with 33 vertices and maximum degree 4.

for some $1 \leqslant s < k$ (Figure 7). Notice that the degree sequence is determined by n and k. By the division algorithm, we can uniquely write $n - 2 = q'(k-1) + s'$ with $s' < k-1$. Thus q' will be the number of vertices of degree k, one vertex will have degree $s = s' + 1$, and the rest will be leaves.

Similarly, for positive integers n, k, the *extended rgood tree* with order n and maximum degree k, is a rooted greedy tree with root degree $k - 1$ and degree sequence

$$(k, k, \ldots, k, k-1, s, 1, \ldots, 1)$$

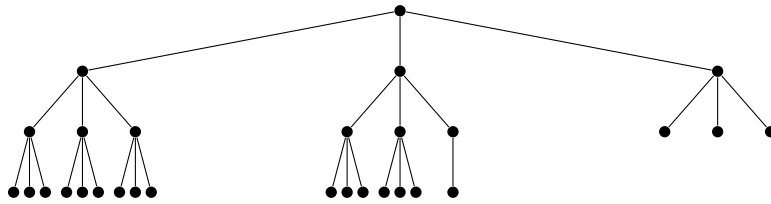for some $1 \leqslant s < k$ (Figure 8).

Figure 8: An extended rgood tree with 29 vertices and maximum degree 4.

Among all rooted trees with n vertices, maximum degree k, and root degree at most $k - 1$, we seek the one with the maximum number of root-containing subtrees.

Theorem 3.11. *Among all rooted trees with n vertices, maximum degree k, and root degree at most $k - 1$, the extended rgood tree maximizes the number of root-containing subtrees.*

Theorem 3.11 follows from Lemmas 3.12 and 3.14 below.

Lemma 3.12. *For any pair (n, k) of positive integers, $n > k$, a tree with n vertices, maximum degree k, and root degree at most $k - 1$ which maximizes the number of root-containing subtrees must have root degree $k - 1$.*

Proof. For contradiction, suppose T is such a tree with root ρ having degree at most $k - 2$. Since $n \geqslant k$, there exists a child u of ρ that is not a leaf. Let v be a child of u.

Define $T' := T - \{uv\} + \{\rho v\}$. Every root-containing subtree in T can be uniquely identified by its list of vertices. It is easy to see that each list forms a root-containing

subtree in T'. However, T' also has root-containing subtrees which contain v and not u. These do not appear in T. Therefore T' has more root-containing subtrees than T. This contradicts our choice of T. $\qquad\square$

Definition 3.13. Given two positive integer sequences $\pi = (d_0, \cdots, d_{n-1})$ and $\pi' = (d'_0, \cdots, d'_{n-1})$, we say π' *majorizes* π, denoted $\pi \lhd \pi'$, if for each $k \in \{0, \cdots, n-2\}$,

$$\sum_{i=0}^{k} d_i \leqslant \sum_{i=0}^{k} d'_i \qquad \text{and} \qquad \sum_{i=0}^{n-1} d_i = \sum_{i=0}^{n-1} d'_i.$$

The following is a simpler analogue of Theorem 11 of Andriantiana, Wagner, and Wang [1]. We skip the details.

Lemma 3.14. *Let T and T' be rooted greedy trees on n vertices with root degree $k-1$. If T has degree sequence π and T' has degree sequence π' where $\pi \lhd \pi'$, then T' has more root-containing subtrees than T.*

In the search for a tree which maximizes the number of root-containing subtrees, Lemma 3.12 implies that it is sufficient to restrict our attention to trees with root degree $k-1$. Because we are considering only degree sequences on n vertices with maximum degree k, it is easy to see that the degree sequence of the extended rgood tree majorizes all other such degree sequences. Thus, Lemma 3.14 then implies that the extended rgood tree for order n and maximum degree k as stated in Theorem 3.11.

For the purpose of our study of maximum distances between different middle parts, we also note the following fact.

Remark. Among all rooted trees of given order, root degree at most $k-1$, and maximum degree k:

- the extended rgood tree minimizes the height;

- the path (rooted at one end) minimizes the number of root-containing subtrees and maximizes the height.

3.4 Middle parts in trees with a given maximum degree

Fix $n, k \in \mathbb{Z}^+$. Similar to the binary tree case, we restrict our attention to classes of trees which have order n and maximum degree k. In this section, we detail our findings for the trees in this class which maximize the distance between different middle parts.

Theorem 3.15. *For fixed $n, k \in \mathbb{Z}^+$, each tree T with order n and maximum degree k has*

$$d(CT(T), C(T)) \leqslant \frac{n - \left\lceil \frac{n+1}{2} \right\rceil - h_u}{2}$$

where

$$h_u = \left\lceil \frac{\ln\left(\left\lceil \frac{n+1}{2} \right\rceil (k-2) + 1\right)}{\ln(k-1)} \right\rceil - 1.$$

This inequality is tight for the tree formed by identifying the root of an extended rgood tree with one end of a path of appropriate length.

Proof. For a fixed tree T, select vertices $u \in CT(T)$ and $v \in C(T)$ such that $d(u, v) = d(CT(T), C(T))$. Assume $d(u, v) \geqslant 1$, otherwise there is nothing to prove. Let T_u and T_v name the components containing u and v respectively in $T - E(P(u, v))$.

Counting the vertices in T, we obtain the inequality

$$d(u, v) \leqslant n - |V(T_u)| - |V(T_v)| + 1 \tag{5}$$

which is tight when the interior vertices of $P(u, v)$ have degree 2.

Because $u \in CT(T)$, Proposition 1.6 implies $|V(T_u)| > n - |V(T_u)|$ and hence

$$|V(T_u)| \geqslant \left\lceil \frac{n + 1}{2} \right\rceil.$$

Set h_u and h_v equal to the heights of T_u and T_v respectively. Because $v \in C(T)$, Proposition 1.4 implies $d(u, v) + h_u \leqslant h_v$ (which is tight when $C(T) = \{v\}$) and hence

$$d(u, v) \leqslant h_v - h_u \leqslant |V(T_v)| - 1 - h_u \tag{6}$$

where the last inequality is tight when T_v is a path with v being one end of that path, in which case $h_v = |V(T_v)| - 1$.

The upper bound in (6) for $d(u, v)$ is maximum when h_u is minimum. By Remark 3.3, the minimum h_u is achieved when T_u is the extended rgood tree. Since $|V(T_u)| \geqslant \left\lceil \frac{n+1}{2} \right\rceil$ and the maximum degree is k, we can determine the height of an extended rgood tree with these conditions. The extended rgood tree with maximum degree k and height h has at most $\sum_{i=0}^{h} (k-1)^i$ vertices. The height of T_u will be the smallest value h which satisfies

$$|V(T_u)| \leqslant \sum_{i=0}^{h} (k-1)^i$$
$$= \frac{(k-1)^{h+1} - 1}{k - 2}.$$

As a result,

$$h \geqslant \frac{\ln\left(|V(T_u)|(k - 2) + 1\right)}{\ln(k - 1)} - 1.$$

Since h is the smallest value that satisfies the above inequality, when $|V(T_u)| = \left\lceil \frac{n+1}{2} \right\rceil$, we can conclude

$$h = \left\lceil \frac{\ln\left(\left\lceil \frac{n+1}{2} \right\rceil(k - 2) + 1\right)}{\ln(k - 1)} \right\rceil - 1.$$

Without knowing $|V(T_v)|$ exactly, we can add (5) and (6) and solve for $d(u,v)$ to obtain the desired upper bound for $d(u,v)$:

$$2d(u,v) \leqslant n - |V(T_u)| - h_u$$
$$d(u,v) \leqslant \frac{1}{2}\left(n - |V(T_u)| - h_u\right)$$
$$\leqslant \left\lfloor \frac{1}{2}\left(n - \left\lceil \frac{n+1}{2}\right\rceil - h\right)\right\rfloor. \qquad \square$$

Theorem 3.16. *For fixed $n, k \in \mathbb{Z}^+$, each tree T with order n and maximum degree k has*

$$d(Core(T), CT(T)) \leqslant n - n' - \left\lceil \frac{n+1}{2}\right\rceil + 1$$

where n' is the minimum order of an extended rgood tree T_u with maximum degree k such that $F_{T_u}(u) \geqslant n - |V(T_u)|$. This inequality is tight for the tree formed by identifying the root of the extended rgood tree with one end of a path of appropriate length.

Proof. Let $u \in Core(T)$ and $v \in CT(T)$ such that $d(u,v) = d(Core(T), CT(T))$. Assume $d(u,v) \geqslant 1$, otherwise there is nothing to prove. Define T_u and T_v to be the components of $T - E(P(u,v))$ containing u and v respectively. Let h_u and h_v be the heights of T_u and T_v respectively.

Similar to before, Proposition 1.6 implies

$$|V(T_v)| \geqslant \left\lceil \frac{n+1}{2}\right\rceil.$$

By Proposition 1.7, $u \in Core(T)$ and its neighbor w on $P(u,v)$ is not in the subtree core precisely when

$$F_{T_u}(u) \geqslant 1 + F_{T-T_u}(w) \geqslant d(u,v) + F_{T_v}(v) \geqslant d(u,v) + |V(T_v)|,$$
$$d(u,v) \leqslant F_{T_u}(u) - F_{T_v}(v) \leqslant F_{T_u}(u) - |V(T_v)|. \qquad (7)$$

The last inequality is tight if T_v is a path.

Counting the vertices in T, we see

$$n \geqslant d(u,v) + |V(T_u)| + |V(T_v)| - 1,$$
$$d(u,v) \leqslant n - |V(T_u)| - |V(T_v)| + 1 \leqslant n - \left\lceil \frac{n+1}{2}\right\rceil - n' + 1.$$

where n' is the minimum number of vertices in a tree T_u with maximum degree k such that $F_{T_u}(u) \geqslant d(u,v) + |V(T_v)| = n - |V(T_u)|$ as in (7). Note that $F_{T_u}(u)$ is maximized by the extended rgood tree, giving the extremal tree in the theorem statement. $\qquad \square$

Theorem 3.17. *For fixed $n, k \in \mathbb{Z}^+$, each tree T with order n and maximum degree k has*

$$d(Core(T), C(T)) \leqslant n - n' - \left\lfloor \frac{1}{2}(n - n' + h')\right\rfloor$$

where $h' = \left\lceil \frac{\ln(n'(k-2)+1)}{\ln(k-1)} \right\rceil - 1$ and n' is the minimum number of vertices in the extended rgood tree T_u with maximum degree k such that $F_{T_u}(u) \geqslant n - |V(T_u)|$. This inequality is tight for the tree formed by identifying the root of the extended rgood tree with one end of a path of appropriate length.

Proof. Let $u \in Core(T)$ and $v \in C(T)$ such that $d(u,v) = d(Core(T), C(T))$. Define T_u and T_v to be the components of $T - E(P(u,v))$ containing u and v respectively. Let h_u and h_v be the heights of T_u and T_v respectively.

Because $u \in Core(T)$ and its neighbor on $P(u,v)$ is not in the subtree core, as in (7), Proposition 1.7 gives

$$d(u,v) \leqslant F_{T_u}(u) - |V(T_v)| \tag{8}$$

which is tight when T_v is a path.

Because $v \in C(T)$ and its neighbor on $P(u,v)$ is not in the center, as in the proof of Theorem 3.15, Proposition 1.4 gives

$$d(u,v) \leqslant h_v - h_u \leqslant |V(T_v)| - h_u - 1.$$

As in (6), this is also tight when T_v is a path and $C(T) = \{v\}$.

Adding these two inequalities together we obtain the following bound.

$$d(u,v) \leqslant \frac{1}{2}\left(F_{T_u}(u) - h_u - 1\right).$$

The upper bound is maximum when $F_{T_u}(u)$ is large and h_u is small which is optimized when T_u is the extended rgood tree.

If n' is the number of vertices in T_u, then because $v \in C(T)$ and T_v is a path, then $ecc_T(v)$ is at least half of the diameter of T which translates to

$$|V(T_v)| \geqslant \frac{1}{2}(n - n' + h_u).$$

Any tree on n' vertices with maximum degree at most k will have height at least the height of the corresponding extended rgood tree. As determined in the proof of Theorem 3.15,

$$h_u \geqslant \left\lceil \frac{\ln\left(n'(k-2)+1\right)}{\ln(k-1)} \right\rceil - 1 := h'.$$

In conclusion,

$$d(u,v) \leqslant n - |V(T_u)| - |V(T_v)| \leqslant n - n' - \left\lfloor \frac{1}{2}(n - n' + h') \right\rfloor.$$

Further, this upper bound is maximized when n' is minimized. However, n' must still satisfying the condition $F_{T_u}(u) \geqslant d(u,v) + |V(T_v)| = n - |V(T_u)|$ from (8). $\qquad \square$

4 Different middle parts in trees with a given diameter D

Recall that all extremal trees in Section 2 were comets. In the previous section, we explored the effect of limiting the maximum vertex degree in a tree. Here, we ask how the distances between middle parts and the corresponding extremal structures change when the diameter is limited. The next two propositions follow from exactly the same arguments as those for Theorem 2.2 and Theorem 2.5 in Section 2, we skip the proofs.

Proposition 4.1. *For fixed integers $D \geqslant 2$, there exists n_0 such that for all $n > n_0$, every tree T of order n and diameter at most D satisfies*

$$d(C(T), CT(T)) \leqslant \left\lfloor \frac{D-2}{2} \right\rfloor,$$

which is achieved by a D-comet.

Proposition 4.2. *For fixed $D \geqslant 2$, there exists n_0 such that for all $n > n_0$, every tree T of order n and diameter at most D satisfies*

$$d(C(T), Core(T)) \leqslant \left\lfloor \frac{D-2}{2} \right\rfloor,$$

which is achieved by a D-comet.

Finding the maximum of $d(CT(T), Core(T))$ is unexpectedly more difficult. Notwithstanding the progress in the remaining part of this paper, it remains unsolved. Fix integers $D \geqslant 2$ and $n \geqslant 1$. Among all trees with diameter at most D and order n, fix a tree T which realizes the maximum value for $d(CT(T), Core(T))$.

Select vertices $u \in Core(T)$ and $v \in CT(T)$ such that the graph distance between u and v is precisely $d(CT(T), Core(T))$. We assume $d(u, v) \geqslant 1$, otherwise there is nothing to prove. In $T - E(P(u, v))$, let T_u name the component containing u while T_v is the component containing v. Consider u to be the root of T_u and v to be the root of T_v.

Let w be the neighbor of u on $P(u, v)$. Because $u \in Core(T)$ and $w \notin Core(T)$, Proposition 1.7 implies

$$F_{T_u}(u) > F_{T-T_u}(w).$$

Because $v \in CT(T)$ and its neighbor on $P(u, v)$ is not in $CT(T)$, Proposition 1.6 implies

$$|V(T_v)| > n - |V(T_v)|.$$

Suppose T_u is not a star. Create a new tree T' from T by replacing T_u with a star T'_u which is rooted at u and has the same order as T_u. Using the convention that T'_u and T'_v are the components containing u and v respectively in $T' - E(P(u, v))$, we see that T'_v is isomorphic to T_v. First observe that

$$F_{T'_u}(u) \geqslant F_{T_u}(u) > F_{T-T_u}(w) = F_{T'-T'_u}(w)$$

which implies $w \notin Core(T')$ by Proposition 1.7. Further,

$$|V(T'_v)| = |V(T_v)| > n - |V(T_v)| = n - |V(T'_v)|$$

which implies the neighbor of v on $P(u, v)$ is not in the centroid of T' by Proposition 1.6. Therefore

$$d(Core(T'), CT(T')) \geqslant d_{T'}(u, v) = d_T(u, v) = d(Core(T), CT(T)).$$

By the choice of T, $d(Core(T'), CT(T')) = d(Core(T), CT(T))$. So T' is also a tree with diameter at most D and order n which maximizes $d(Core(T), CT(T))$.

Now consider the structure of T'_v in T'. Say T'_v has x vertices and height h. Suppose T'_v does not minimize the number of subtrees containing v for its height and order. Let T''_v be a tree rooted at v with height at most h and order x which minimizes $F_{T''_v}(v)$. Define T'' to be the tree created from T' by replacing T'_v with T''_v. Observe that

$$F_{T''_u}(u) = F_{T'_u}(u) > F_{T'-T'_u}(w) > F_{T''-T''_u}(w)$$

which implies $w \notin Core(T'')$ by Proposition 1.7. Further, for T''_v being the component of $T'' - E(P(u, v))$ which contains v,

$$|V(T''_v)| = |V(T'_v)| > n - |V(T'_v)| = n - |V(T''_v)|.$$

This implies, by Proposition 1.6, that the neighbor of v on $P(u, v)$ in T'' is not in $CT(T'')$ and

$$d(Core(T''), CT(T'')) \geqslant d_{T''}(u, v) = d_T(u, v) = d(Core(T), CT(T)).$$

By the choice of T, $d(Core(T''), CT(T'')) = d(Core(T), CT(T))$ which implies T'' is also a tree with diameter at most D and order n that maximizes the distance between the subtree core and the centroid.

Remark. Fix $n, D \in \mathbb{Z}^+$. Among all trees with diameter at most D and order n, one such tree T which maximizes $d(Core(T), CT(T))$ has T_u being a star rooted at u and T_v a tree which minimizes the number of subtrees containing v for its height and order. This structure of T is drawn in Figure 9.

In Section 5, we take a closer look at the structure of T_v, a tree which minimizes the number of subtrees containing v for its height and order. While we determine many necessary properties of T_v, characterizing the exact structure is still an open problem.

5   Rooted trees of given order and height

For any $n, h \in \mathbb{Z}^+$, this section is devoted to characterizing the rooted trees with n vertices and height at most h which have the minimum number of root-containing subtrees. For the remainder of this section, we will call these trees *optimal*.

To standardize some notation, we restrict our attention to trees T which are rooted at root ρ, have order n and height at most h unless mentioned otherwise. Note that $h(T) = ecc_T(\rho)$. The degree of a vertex v will be denoted $deg(v)$.
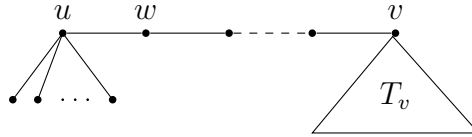
Figure 9: The structure of a tree T with diameter D and order n which maximize $d(Core(T), CT(T))$. Here, $u \in Core(T)$, $v \in CT(T)$, and T_v minimizes the number of subtrees containing v for its order and height.

For any $v \in V(T)$, let $T(v)$ denote the subtree induced by v and all of its descendants. We will view $T(v)$ as a tree rooted at v. For each neighbor v_i of ρ, set $T_i := T(v_i)$. For $f \in \{0, 1, \ldots, h-1\}$, define the f-*split* to be the tree rooted at w_1 with $h + f$ vertices, constructed from paths $P_1 = (w_1, w_2, \ldots, w_h)$, and $P_2 = (u_1, u_2, \ldots, u_f)$ by adding the edge $u_1 w_{h-f}$. In other words, the midpoint of path on $2f$ edges is joined to the root w_1 by an $(h - f - 1)$-edge path. Our main results from this section are summarized in the following theorem.

Theorem 5.1. *For positive integers n and h, there is an optimal tree T with n vertices and height at most h such that each T_i is a k_i-split with $k_1 \geqslant k_2 \geqslant \ldots \geqslant k_r$ (see Figure 10). Further, the tuple $(k_1, k_2, \ldots, k_r)$ is described by one of the following three types:*

(i) (Paths) $k_r = 0$, $k_{r-1} \in \{0, 1\}$, $k_{r-2} = \ldots = k_1 = 1$.

(ii) (One large) $k_1 > \left\lceil \sqrt{h + \frac{5}{4}} - \frac{1}{2} \right\rceil$ and $k_i = \left\lfloor \frac{h+1}{k_1+1} \right\rfloor$ for each $i \in \{2, \ldots, r\}$ provided $\left\lfloor \frac{h+1}{k_1+1} \right\rfloor \geqslant \frac{h+1}{k_1} - 1$. Further, if $n > 5h^2$, then $k_i \leqslant \ln(6h)$ for each $i \in \{2, \ldots, r\}$.

(iii) (Even distribution) $k_1 \leqslant \left\lceil \sqrt{h + \frac{5}{4}} - \frac{1}{2} \right\rceil$ and for all $i, j \in [r]$, $|k_i - k_j| \leqslant 1$. Further, if $n > 5h^2$, then $k_i \leqslant \ln(6h)$ for each $i \in \{2, \ldots, r\}$.

Here we present several lemmas regarding the characteristics of an optimal tree. These all work toward the proof of Theorem 5.1.

Lemma 5.2. *In any optimal tree T, for any $v \in V(T)$, $T(v)$ minimizes the number of root-containing subtrees among all rooted trees of the same order and height at most $h - h_T(v)$ where $h_T(v)$ is the height of v in T.*

Proof. Let T be an optimal tree. Suppose, for contradiction, that there is a vertex v for which $T(v)$ does not satisfy the lemma. In other words, there is a tree $T'(v)$, which is rooted at v, has the same order as $T(v)$, and has

$$h(T'(v)) \leqslant h - h_T(v) \quad \text{and} \quad F_{T'(v)}(v) < F_{T(v)}(v).$$

Let T' be the tree obtained from T by replacing $T(v)$ with $T'(v)$. Then T and T' have the same number of subtrees containing ρ but not v. Define $T^* := T - (T(v) - \{v\})$ and
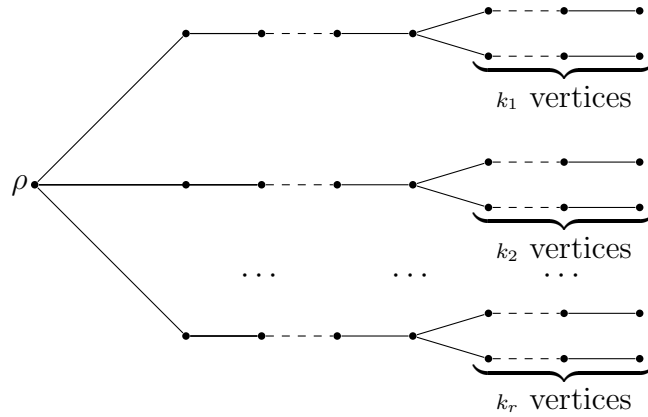
Figure 10: The structure of a tree T with height h and order n which minimizes the number of root-containing subtrees.

let $F_{T^*}(\rho, v)$ be the number of subtrees of T^* that contain both ρ and v. Because T and T' only differ in the descendants of v, we have

$$F_T(\rho) - F_{T'}(\rho) = F_{T(v)}(v)F_{T^*}(\rho, v) - F_{T'(v)}(v)F_{T^*}(\rho, v) > 0,$$

a contradiction to the optimality of T. $\qquad\square$

Lemma 5.3. *The height of any leaf in an optimal tree is h.*

Proof. If $n = h + 1$, it is straightforward to see that the path rooted at one end is the optimal tree. In the case when $n > h + 1$, some vertex must have at least 2 children. Suppose, for contradiction, that there is a leaf $v \in V(T)$ whose height is less than h. Let x be the closest ancestor (possibly the root) of v that has at least two children. Let y be a child of x that is not on $P(x, v)$ and z be the child of x on $P(x, v)$.
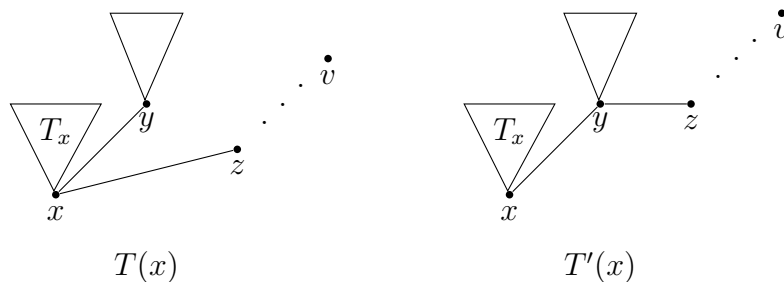

Figure 11: Trees $T(x)$ and $T'(x)$ from Lemma 5.3

Let T_x be the component containing x in $T - xy - xz$ and consider the tree

$$T'(x) := T(x) - xz + yz$$

depicted in Figure 11. Note that $T'(x)$ has the same order as $T(x)$ and has height no more than $h - h_T(x)$ because the height of v in T is less than h.

Counting the number of subtrees containing x in each tree, we obtain the following equalities:

$$
\begin{aligned}
F_{T(x)}(x) &= F_{T_x}(x)(1 + d_{T(x)}(x, v))(1 + F_{T(y)}(y)), \\
F_{T'(x)}(x) &= F_{T_x}(x)\left[1 + (1 + d_{T(x)}(x, v))F_{T(y)}(y)\right].
\end{aligned}
$$

Together, these imply

$$
F_{T(x)}(x) - F_{T'(x)}(x) = d_{T(x)}(x, v)F_{T_x}(x) > 0.
$$

Since T was optimal, this contradicts Lemma 5.2. $\qquad\square$

Lemma 5.4. *Every optimal tree has one of the following two properties:*

- *All non-root vertices have degree at most 3.*

- *All non-root vertices of height less than $h-1$ have degree at most 3. For any vertex v of height $h-1$, $deg(v) \leqslant 4$. Further, if $deg(v) = 4$, then the parent of v must have degree 2 or be the root.*

Proof. As before, this proof proceeds by contradiction. Let x be a non-root vertex in an optimal tree T with degree at least 4. Say y, z, and w are three children of x and let u be the parent of x. Denote by T_u and T_x the components containing u and x respectively in $T - ux - xy - xz - xw$. Without loss of generality, assume

$$
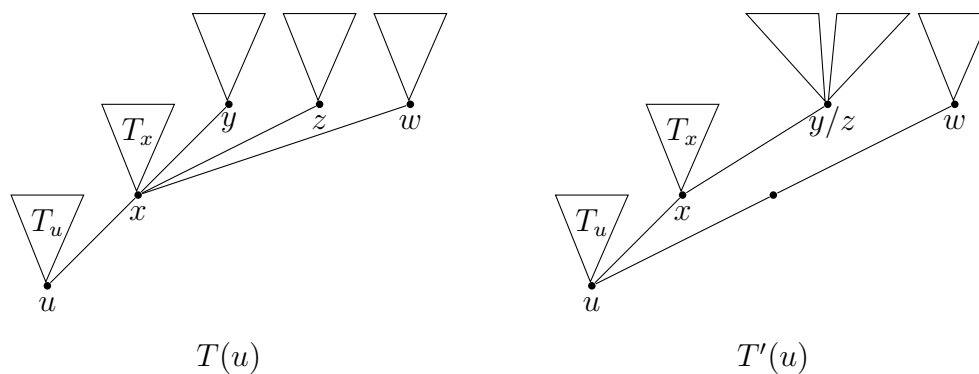F_{T(w)}(w) = \max\{F_{T(y)}(y), F_{T(z)}(z), F_{T(w)}(w)\}.
$$

Figure 12: Trees $T(u)$ and $T'(u)$ in the proof of Lemma 5.4.

Now consider the tree $T'(u)$ obtained from $T(u)$ by removing the edges xz and xw, inserting a path of length 2 between u and w, while identifying the vertices y and z (Figure 12). Note that $T'(u)$ has the same height and order as $T(u)$. Counting the number of subtrees containing u in each, we find

$$
F_{T(u)}(u) = F_{T_u}(u)\left[1 + F_{T_x}(x)\left(1 + F_{T(y)}(y)\right)\left(1 + F_{T(z)}(z)\right)\left(1 + F_{T(w)}(w)\right)\right],
$$

$$F_{T'(u)}(u) \;=\; F_{T_u}(u)\left(2 + F_{T(w)}(w)\right)\left[1 + F_{T_x}(x)\left(1 + F_{T(y)}(y)F_{T(z)}(z)\right)\right].$$

Together, these imply the following

$$
\begin{aligned}
F_{T(u)}(u) - F_{T'(u)}(u) = F_{T_u}(u)\big[& F_{T_x}(x)F_{T(y)}(y)\left(F_{T(w)}(w) - F_{T(z)}(z)\right) \\
& + F_{T_x}(x)\left(F_{T(y)}(y) - 1\right) \\
& + \left(F_{T(w)}(w) + 1\right)\left(F_{T_x}(x)F_{T(z)}(z) - 1\right)\big] \\
\geqslant 0. & \hspace{5cm} (9)
\end{aligned}
$$

Because T is an optimal tree, $T(u)$ is an optimal tree by Lemma 5.2. Therefore (9) must be equality. Note that for any tree H and vertex $a \in V(H)$, $F_H(v) \geqslant 1$ because the subtree containing only the vertex v will be counted. Therefore, equality holds in (9) exactly when $F_{T_x}(x) = F_{T(y)}(y) = F_{T(z)}(z) = F_{T(w)}(w) = 1$, or equivalently, $deg(x) = 4$ and y, z, w are all leaves so x has height $h - 1$ in T. Create T' from T by replacing $T(u)$ with $T'(u)$. Because (9) is equality, $F_{T(u)}(u) = F_{T'(u)}(u)$. Therefore T' is also an optimal tree.

In T', $deg_{T'}(x) = 3$ but $deg_{T'}(u) = deg_T(u) + 1$. Observe u has height $h - 2$ in T'. If u is not the root of T' and $deg_{T'}(u) \geqslant 4$, then we can repeat the argument for optimal tree T' and vertex u having degree at least 4. Because the height of u is $h - 2$, we will find a contradiction in the step which parallels (9). Therefore $deg_{T'}(u) \leqslant 3$ which implies $deg_T(u) \leqslant 2$. Since u is not the root of T, we can conclude $deg_T(u) = 2$ as stated in the lemma. $\qquad\square$

In the proof of Lemma 5.4, in the case where $deg_T(x) = 4$, we created another optimal tree T' where $deg_{T'}(x) = 3$ and no other degree 4 vertices where created. Hence, if an optimal tree has multiple degree 4 vertices of height $h-1$, we can repeat this procedure to obtain an optimal T' with all vertices of degree at most 3. This establishes the following observation.

Observation 5.5. *There is an optimal tree in which all non-root vertices have degree at most 3.*

We now shift our attention to the structures of T_i for $1 \leqslant i \leqslant k$.

Lemma 5.6. *In an optimal tree T, each subtree $T_i \cup \{\rho\}$ falls into one of the following three categories:*

- *There is at most one non-root vertex with degree 3.*

- *All non-root vertices of height at most $h-3$ have degree 2, the vertex of height $h-2$ has degree 3, and exactly one of its children has degree 3.*

- *All non-root vertices of height at most $h - 2$ have degree 2 and the vertex of height $h - 1$ has degree 4.*

Proof. We prove this in two pieces, considering the alternatives from Lemma 5.4 separately. We start with the optimal trees in which all vertices have degree at most 3.

For contradiction, suppose there exists a $T_i \cup \{\rho\}$ with at least two non-root vertices of degree 3. Let v be a degree 3 vertex of greatest height in T_i and let u, w be the two children of v. Let z be the closest ancestor of v such that $deg_{T_i}(z) = 3$, z has parent x, and z has child $y \notin V(P(z, v))$. Let ℓ_1 denote the distance from v to a leaf in T_i and ℓ_2 the length of $P(v, z)$. Let T_x denote the component containing x in $T(x) - xz$ (Figure 13).
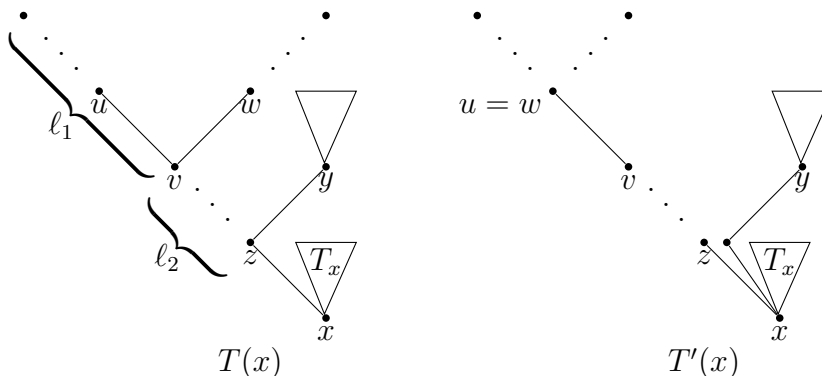

Figure 13: Transforming $T(x)$ into $T'(x)$ when $deg_T(v) = 3$ in the proof of Lemma 5.6.

Create a new tree $T'(x)$ from $T(x)$ by removing the edges vw and zy, inserting a length 2 path between x and y, and identifying u and w (Figure 13). Note that $T'(x)$ has the same height and order as $T(x)$. The number of subtrees containing x in each is

$$F_{T(x)}(x) = F_{T_x}(x) \left[1 + (1 + F_{T(y)}(y))[\ell_2 + (\ell_1 + 1)^2] \right],$$
$$F_{T'(x)}(x) = F_{T_x}(x)(2 + F_{T(y)}(y)) \left(\ell_2 + 2 + \ell_1^2 \right).$$

By Lemma 5.3, the height of each leaf in T is h, hence $V(T(y)) \geqslant \ell_1 + \ell_2$. Now we have

$$
\begin{aligned}
F_{T(x)}(x) - F_{T'(x)}(x) &= F_{T_x}(x) \left[(1 + F_{T(y)}(y))(2\ell_1 - 1) - (\ell_1^2 + \ell_2 + 1)] \right] \\
&\geqslant F_{T_x}(x) \left[(1 + \ell_1 + \ell_2)(2\ell_1 - 1) - (\ell_1^2 + \ell_2 + 1)] \right] \quad (10) \\
&= F_{T_x}(x)(\ell_1^2 + 2\ell_1\ell_2 + \ell_1 - 2\ell_2 - 2) \\
&\geqslant 0. \quad (11)
\end{aligned}
$$

When either (10) or (11) is strict inequality, we have a contradiction to the optimality of T. Equality holds exactly when $\ell_1 = \ell_2 = 1$ and $|V(T(y))| = \ell_1 + \ell_2$. In other words, $T(y)$ is a single path on two vertices with y having height $h - 1$. Since T' (constructed from T by replacing $T(x)$ with $T'(x)$) is an optimal tree, if x is not the root, $deg_{T'}(x) \leqslant 3$ since x has height $h - 3$. Therefore $deg_T(x) \leqslant 2$ as described in the second property of the lemma.

If T falls into the second category listed in Lemma 5.4, then consider a subtree T_i with a vertex v of degree 4 at height $h - 1$. We will show that all other non-root vertices in $T_i \cup \{\rho\}$ must have degree 2. Suppose to the contrary that v has an ancestor z of degree

3. (In this way, we are able to simultaneously handle the case when there are two vertices of degree 4 in $T_i \cup \{\rho\}$ because they would have to share a common ancestor of degree 3.) Label the vertices as before with s being the third child of v (Figure 14).

Create $T'(x)$ by altering $T(x)$ in a manner similar to that described above. Define

$$T'(x) = T(x) - wv - yz + xw + wy$$
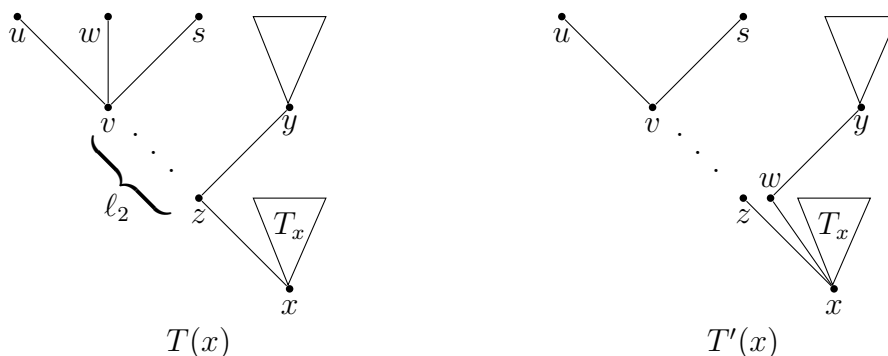
as shown in Figure 14.

Figure 14: Transforming $T(x)$ into $T'(x)$ when v has degree 4 in the proof of Lemma 5.6.

Let ℓ_2 be the distance from z to v in $T(x)$. Because all leaves have height h, $F_{T(y)}(y) \geqslant \ell_2 + 1$ which is tight when $T(y)$ is a path. Now if we calculate $F_{T(x)}(x)$ and $F_{T'(x)}(x)$ exactly and take their difference, we find

$$
\begin{aligned}
F_{T(x)}(x) &= F_{T_x}(x)\left(1 + (1 + F_{T(y)}(y))(\ell_2 + 8)\right) \\
F_{T'(x)}(x) &= F_{T_x}(x)(2 + F_{T(y)}(y))(\ell_2 + 5) \\
F_{T(x)}(x) - F_{T'(x)}(x) &= F_{T_x}(x)\left(3F_{T(y)}(y) - \ell_2 - 1\right) \\
&\geqslant F_{T_x}(x)\left(3(\ell_2 + 1) - \ell_2 - 1\right) \\
&= F_{T_x}(x)\left(2\ell_2 + 2\right) \\
&> 0.
\end{aligned}
$$

This contradicts our choice of T. Thus $T_i \cup \{\rho\}$ can have at most one vertex of degree 4 and all other non-root vertices must have degree 2 as described in the third property of the lemma. $\square$

Once again, it is useful to note that for each of the optimal trees described in the second two properties of Lemma 5.6, the proof supplies T' analogues which have the same number of root-containing subtrees and yet fall under the first property description in Lemma 5.6. This gives the following observation.

Observation 5.7. *There is an optimal tree with each $T_i \cup \{\rho\}$ having at most one non-root vertex of degree 3.*

Observation 5.7 establishes that each T_i is a k_i-split for some integer k_i with $0 \leqslant k_i \leqslant h-1$. To minimize some notation, we state the following structural observation.

Observation 5.8. *In an optimal tree T, the number of root-containing subtrees in a k_i-split together with root ρ is*

$$s_h(k_i) := h + k_i^2 + k_i + 1.$$

This definition also makes sense for the 0-split (together with ρ), which has $h + 1$ root-containing subtrees.

Lemma 5.9. *Among the T_i subtrees in an optimal tree, at most two of them can be 0-splits.*

Proof. Suppose, for contradiction, that T_i, T_j and T_k are each 0-splits in an optimal tree. Consider $S := T_i \cup T_j \cup T_k \cup \{\rho\}$. Create S' from S by replacing T_i with a 1-split, T_j with an $(h-1)$-split and deleting T_k (Figure 15).
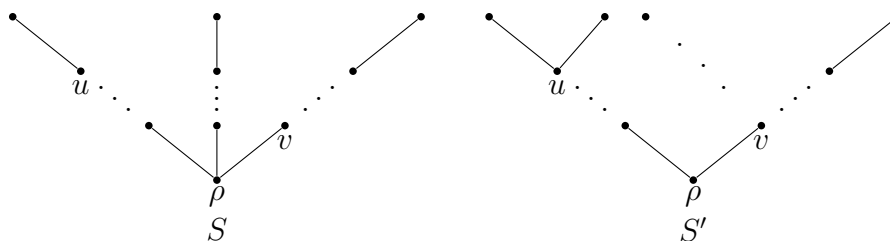

Figure 15: Trees S and S' from the proof of Lemma 5.9

The difference in the number of subtrees is

$$
\begin{aligned}
F_S(\rho) - F_{S'}(\rho) &= (s_h(0))^3 - s_h(1)s_h(h-1) \\
&= (h+1)^3 - (h+3)(2h + (h-1)^2) \\
&= 2(h-1) \\
&> 0.
\end{aligned}
$$

This contradicts the optimality of T because the tree obtained from T by replacing S with S' has fewer root-containing subtrees than T. $\square$

Lemma 5.10. *If some T_i is a 0-split, then for each $j \neq i$, T_j is either a 0-split or a 1-split.*

Proof. Suppose instead that T_i is a 0-split and T_j is a k_j-split where $1 < k_j \leqslant h-1$. Let S be the tree induced by T_i, T_j, and ρ. Construct S' from S by replacing T_i with a 1-split and replacing T_j with a $(k_j - 1)$-split (Figure 16).

Note that S' has the same height and order as S, and

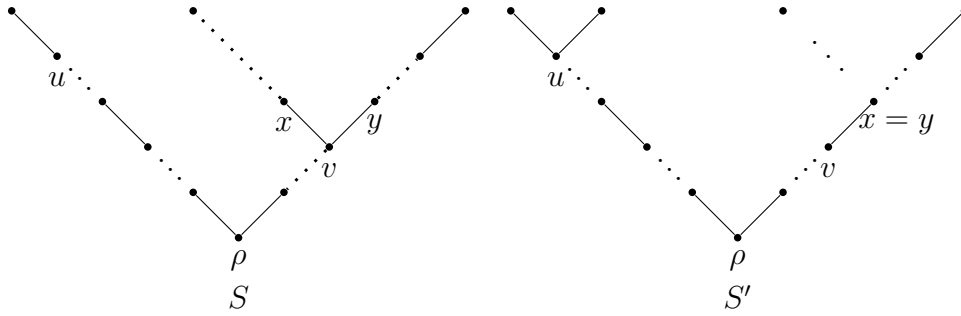$$F_S(\rho) - F_{S'}(\rho) = s_h(0)s_h(k_j) - s_h(1)s_h(k_j - 1)$$

Figure 16: Tree S and tree S' from the proof of Lemma 5.10.

$$= (h+1)\left[h + k_j^2 + k_j + 1\right] - (h+3)\left[h + (k_j-1)^2 + k_j\right]$$
$$= 2\left[(k_j-1)(h-k_j) + (k_j-1)\right]$$
$$> 0. \qquad\qquad\qquad\qquad\qquad\qquad \text{(for } k_j > 1)$$

This contradicts the optimality of T because the tree obtaining from T by replacing S with S' has fewer root-containing subtrees than T. $\qquad\square$

Lemma 5.11. *A rooted tree T is not optimal if for any T_i (k_i-split) and T_j (k_j-split), we have $k_i(1 + k_j) > h + 1$ for $1 \leqslant k_i \leqslant k_j \leqslant h - 1$.*

Proof. Define T_2 be the subtree of T which consists of the root ρ together with T_i and T_j.

Construct T_2' from T_2 by replacing T_i with a $(k_i - 1)$-split and replacing T_j with a $(k_j + 1)$-split. This construction is well-defined because $1 \leqslant k_i$ and $k_j \leqslant h - 1$.

It is easy to see that T_2' has the same height and order as T_2. We have

$$F_{T_2}(\rho) = s_h(k_i)s_h(k_j) \qquad \text{and} \qquad F_{T_2'}(\rho) = s_h(k_i - 1)s_h(k_j + 1).$$

Since $k_i \leqslant k_j$ and $k_i(1 + k_j) > h + 1$, we have

$$F_{T_2}(\rho) - F_{T_2'}(\rho) = -2(k_i - k_j - 1)(k_i + k_i k_j - h - 1) > 0,$$

which contradicts the optimality of T. $\qquad\square$

By reversing the roles of k_i and k_j in the previous lemma, we obtain the following corollary.

Corollary 5.12. *A rooted tree T is not optimal if for any T_i, which is a k_i-split, and T_j, which is a k_j-split, we have $k_j(1 + k_i) < h + 1$ and $k_i < k_j - 1$.*

Corollary 5.13. *Fix an optimal tree T in which each T_i is a k_i-split with $k_i \geqslant k_{i+1}$. If $k_1 > \sqrt{h + \frac{5}{4}} - \frac{1}{2}$, then $k_i < \sqrt{h + \frac{5}{4}} - \frac{1}{2}$ for each $i \geqslant 2$.*

Proof. Let T be an optimal tree, as described in the corollary, with $k_1 > \sqrt{h + \frac{5}{4}} - \frac{1}{2}$. For contradiction, suppose $k_2 \geqslant \sqrt{h + \frac{5}{4}} - \frac{1}{2}$. Observe

$$k_2(k_1 + 1) > \left(\sqrt{h + \frac{5}{4}} - \frac{1}{2} \right) \left(\sqrt{h + \frac{5}{4}} + \frac{1}{2} \right)$$
$$= h + 1.$$

However, this contradicts the statement of Lemma 5.11. Therefore the corollary holds. $\square$

Corollary 5.14. *Fix an optimal tree T in which each T_i is a k_i-split. For any pair $\{k_i, k_j\}$ with $k_i, k_j \leqslant \sqrt{h + \frac{5}{4}} - \frac{1}{2}$, we can conclude $|k_i - k_j| \leqslant 1$.*

Proof. Suppose $k_i, k_j \leqslant \sqrt{h + \frac{5}{4}} - \frac{1}{2}$ with $k_j \geqslant k_i + 2$. Observe

$$k_j(1 + k_i) \leqslant k_j(k_j - 1)$$
$$\leqslant \left(\sqrt{h + \frac{5}{4}} - \frac{1}{2} \right) \left(\sqrt{h + \frac{5}{4}} - \frac{3}{2} \right)$$
$$= h + 2 - 2\sqrt{h + \frac{5}{4}}$$
$$< h + 1.$$

This contradicts Corollary 5.12, finishing the proof. $\square$

Going further in this direction, we have the following even more specific statements.

Lemma 5.15. *Suppose $k_1 \geqslant k_2 \geqslant \ldots \geqslant k_k$. If $k_1 > \left\lceil \sqrt{h + \frac{5}{4}} - \frac{1}{2} \right\rceil$, then for each $i > 1$,*

$$\frac{h + 1}{k_1} - 1 \leqslant k_i \leqslant \frac{h + 1}{k_1 + 1}. \tag{12}$$

In particular, $k_2 = k_3 = \ldots = k_k = \left\lfloor \frac{h+1}{k_1 + 1} \right\rfloor$ provided $\left\lfloor \frac{h+1}{k_1 + 1} \right\rfloor \geqslant \frac{h+1}{k_1} - 1$.

Proof. Since $k_1 > \sqrt{h + \frac{5}{4}} - \frac{1}{2}$, Corollary 5.13 implies $k_2 < \sqrt{h + \frac{5}{4}} - \frac{1}{2}$. Since T is optimal, Lemma 5.11 yields $k_i(1 + k_1) \leqslant k_2(1 + k_1) \leqslant h + 1$. Thus

$$k_i \leqslant \frac{h + 1}{k_1 + 1}.$$

Because $k_1 > \left\lceil \sqrt{h + \frac{5}{4}} - \frac{1}{2} \right\rceil$, then necessarily $k_i \leqslant k_2 < k_1 - 1$. Corollary 5.12 gives

$$k_1(1 + k_i) \geqslant h + 1.$$

This is equivalent to

$$k_i \geqslant \frac{h + 1}{k_1} - 1. \qquad \square$$

We have established the three cases detailed in Theorem 5.1, aside from the further information when $n \geqslant 5h^2$. In the rest of this section, we examine the number of T_i subtrees which can be a k_i-splits in an optimal tree. This will shed light on the values of k_i in an optimal tree with a large number of vertices, compared to its height, and also lends insight into the degree of the root vertex. First we prove a technical lemma.

Lemma 5.16. *For fixed $h, n \in \mathbb{Z}$, let T be an optimal tree with n vertices, height h, and where T_i is a k_i-split. Fix $t \in \mathbb{R}$, $t \geqslant 2$ which satisfies the inequality $h^{1/(t+1)} > \ln(6h)$. For $x \in \left[h^{1/(t+1)}, h^{1/t} \right]$ with $n \geqslant (h+x)(h+x-1)+1$, then*

$$|\{i : k_i = x\}| < h + x - 1.$$

Proof. Let T be a tree such that each T_i is a k_i-split. Suppose for contradiction that $k_1 = k_2 = \ldots = k_{h+x-1} = x$ (where the k_i values are not necessarily in non-increasing order).

Let H be the subtree induced by $T_1, \ldots, T_{h+x-1}$ and the root ρ. Here, each T_i is an x-split. Thus the number of root-containing subtrees in H is

$$F_H(\rho) = (h + x^2 + x + 1)^{h+x-1}.$$

Let T_i' be an $(x-1)$-split. Define a new tree T' by replacing T_i with T_i' for each $i \in [h+x-1]$ and increasing the degree of the root by one so that the new branch T_0' is also an $(x-1)$-split. Let H' be the subtree induced by $T_0', T_1', \ldots, T_{h+x-1}'$ and the root of T'. The number of root-containing subtrees in H' is

$$F_{H'}(\rho) = (h + x^2 - x + 1)^{h+x}.$$

In order to compare the number of root-containing subtrees of T and T', it suffices to compare the number of root-containing subtrees of H and H'.

In order to compare these, consider the ratio:

$$\frac{F_T(\rho)}{F_{T'}(\rho)} = \frac{F_H(\rho)}{F_{H'}(\rho)} = \frac{(h + x^2 + x + 1)^{h+x-1}}{(h + x^2 - x + 1)^{h+x}}$$

$$= \frac{1}{h + x^2 - x + 1} \left(1 + \frac{2x}{h + x^2 - x + 1} \right)^{h+x}$$

$$\geqslant \frac{1}{h + h^{2/t} + h^{1/t} + 1} \left(1 + \frac{2h^{1/(t+1)}}{h + h^{2/t} - h^{1/(t+1)} + 1} \right)^{h + h^{1/(t+1)}}$$

since $x \in [h^{1/(t+1)}, h^{1/t}]$. The last expression can be further rewritten as

$$\frac{1}{h + h^{2/t} + h^{1/t} + 1} \left(1 + \frac{2}{h^{t/(t+1)} + h^{(t+2)/(t^2+t)} - 1 + h^{-1/(t+1)}} \right)^{h + h^{1/(t+1)}}$$

$$\geqslant \frac{1}{3h}\left(1 + \frac{2}{h^{t/(t+1)} + h^{(t+2)/(t^2+t)}}\right)^h \qquad\qquad \text{since } h^{1/(t+1)} > 1$$

$$\geqslant \frac{1}{3h}\left(1 + \frac{2}{h^{t/(t+1)} + h^{t/(t+1)}}\right)^{h^{t/(t+1)}h^{1/(t+1)}} \qquad\qquad \text{since } \tfrac{t+2}{t^2+t} < \tfrac{t}{t+1}$$

$$= \frac{1}{3h}\left(1 + \frac{1}{h^{t/(t+1)}}\right)^{h^{t/(t+1)}h^{1/(t+1)}}$$

$$\geqslant \frac{1}{3h}\cdot\frac{1}{2}e^{h^{1/(t+1)}} \qquad\qquad \text{since } \left(1 + \tfrac{1}{y}\right)^y > \tfrac{1}{2}e \text{ for } y \geqslant 1$$

$$> 1 \qquad\qquad \text{since } h^{1/(t+1)} > \ln(6h).$$

Thus, T is not an optimal tree because T' also has n vertices and height h but has fewer subtrees which contain its root. $\qquad\square$

As a consequence we obtain the following corollary:

Corollary 5.17. *Fix $h \geqslant 550$ and $n \in \mathbb{Z}$ with $n \geqslant 5h^2$. Let T be an optimal tree with subtrees T_i which are k_i-splits. Using the terminology from Theorem 5.1, if T has an "even distribution" of k_i values, then $k_i \leqslant \ln(6h)$ for all $i \in \{1, 2, \ldots, r\}$. If T has "one large" k_i values, then $k_i \leqslant \ln(6h)$ for all $i \in \{2, \ldots, r\}$.*

Proof. Fix n, h with $n \geqslant 5h^2$. Let T be an optimal tree with height h and n vertices. Toward a contradiction, suppose there is an integer $x \in (\ln(6h), \sqrt{h + \frac{5}{4}} - \frac{1}{2}]$ such that some T_i is an x-split. Consider two cases.

If $x \in \left(\ln(6h), t^{1/3}\right]$, then setting $t := \frac{\ln h}{\ln x} - 1$, we have $h^{1/(t+1)} = x > \ln(6h)$ and $t \geqslant 2$. Since $n \geqslant 5h^2$ and $x < h^{1/3}$, then clearly $n \geqslant (h+x)(h+x-1)+1$. Thus by Lemma 5.16, there are at most $h + x - 1$ subtrees T_i of T which can be x-splits.

If instead $x \in \left[h^{1/3}, \sqrt{h + \frac{5}{4}} - \frac{1}{2}\right] \subseteq \left[h^{1/3}, h^{1/2}\right]$ (where $h \geqslant 550$), then set $t := 2$. In this case, again $h^{1/(t+1)} > \ln(6h)$, $t \geqslant 2$, and $x \in [h^{1/(t+1)}, h^{1/t}]$ with $n \geqslant (h + x)(h + x - 1) + 1$. Therefore, by Lemma 5.16, there are at most $h + x - 1$ subtrees T_i of T which can be x-splits.

Now if T has an "even distribution" of k_i values, then $k_i \in \{x-1, x\}$ for some $x \leqslant \sqrt{h}$ for all $i \in \{1, 2, \ldots, r\}$. Based on the conclusions here, the number of vertices in T can be bounded as follows, which yields a contradiction:

$$\begin{aligned}
n &\leqslant (h+x)(h+x-1) + (h+x-1)(h+x-2) + 1 \\
&\leqslant (h+\sqrt{h})(h+\sqrt{h}-1)(h+\sqrt{h}-1)(h+\sqrt{h}-2) + 1 \\
&< 5h^2.
\end{aligned}$$

On the other hand, if T has "one large" k_i value, then $k_1 \leqslant h - 1$ and there is an integer $x < \sqrt{h}$ such that $k_i = x$ for $i \in \{2, 3, \ldots, r\}$. Therefore, the number of vertices in T is bounded as follows, which produces a contradiction:

$$n \leqslant (h + k_1) + (h + x)(h + x - 1) + 1 \leqslant 2h + (h + \sqrt{h})(h + \sqrt{h} - 1) + 1 < 5h^2. \quad\square$$

Thus we have established Theorem 5.1.

6 Concluding remarks

In this study, we considered distances between three fundamental concepts of middle parts of the tree: the center, centroid, and subtree core. The maximum distances between each pair of these is first examined for general trees. The extremal structures that achieve these maximum distances contained a vertex of large degree and a long path, motivating us to study the same question for trees with degree restrictions and for trees with bounded diameter. The latter leads us to an interesting and difficult problem of minimizing number of root-containing subtrees among trees with given order and height. While we do not yet have a complete characterization of such optimal trees, we have established many of their structural properties to guide our continued study of this topic.

References

[1] E. Andriantiana, S. Wagner, and H. Wang. Greedy trees, subtrees and antichains. *Electron. J. Combin.*, 20(3) #P28, 2013.

[2] C.A. Barefoot, R.C. Entringer, and L.A. Székely. Extremal values for ratios of distances in trees. *Discrete Appl. Math.*, 80:37–56, 1997.

[3] R. Entringer, D. Jackson, and D. Snyder. Distance in graphs. *Czechoslovak Math. J.*, 26:283–296, 1976.

[4] R. Jamison. On the average number of nodes in a subtree of a tree. *J. Combinatorial Theory Ser. B*, 35:207–223, 1983.

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

[6] L. Lovász. *Combinatorial Problems and Exercises.* AMS Chelsea Publishing, Providence, Rhode Island, 2 edition, 2007.

[7] N. Schmuck, S. Wagner, and H. Wang. Greedy trees, caterpillars, and wiener-type graph invariants. *MATCH Commun. Math. Comput. Chem.*, 68:273–292, 2012.

[8] L.A. Székely and H. Wang. On subtrees of trees. *Adv. in Appl. Math.*, 34:138 – 155, 2005.

[9] L.A. Székely and H. Wang. Binary trees with the largest number of subtrees. *Discrete Appl. Math.*, 155:374–385, 2007.

[10] H. Wang. The extremal values of the wiener index of a tree with given degree sequence. *Discrete Appl. Math.*, 156:2647–2654, 2008.

[11] H. Wiener. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.*, 69:17–20, 1947.