On homometric sets in graphs

Maria Axenovich
Department of Mathematics,
Iowa State University,
Ames, IA 50011, USA
email: axenovic@iastate.edu *

Lale Özkahya
Department of Mathematics,
Iowa State University,
Ames, IA 50011, USA
email: ozkahya@illinoisalumni.org

December 5, 2012

Abstract

For a vertex set $S \subseteq V(G)$ in a graph G, the *distance multiset*, $D(S)$, is the multiset of pairwise distances between vertices of S in G. Two vertex sets are called *homometric* if their distance multisets are identical. For a graph G, the largest integer h, such that there are two disjoint homometric sets of order h in G, is denoted by $h(G)$. We slightly improve the general bound on this parameter introduced by Albertson, Pach and Young [1] and investigate it in more detail for trees and graphs of bounded diameter. In particular, we show that for any tree T on n vertices $h(T) \geq \sqrt[3]{n}$ and for any graph G of fixed diameter d, $h(G) \geq cn^{1/(2d-2)}$.

1 Introduction

For a vertex set $S \subseteq V(G)$, the *distance multiset*, $D(S)$, is the multiset of pairwise distances between vertices of S in G . We say that two vertex sets are *homometric* if their distance sets are identical. "How large could two disjoint homometric sets be in a graph?" was a question of Albertson, Pach and Young [1].

*This author's research partially supported by NSF grant DMS-0901008
Keywords: homometric, distance-set, graph.

Formally, for a graph G, the largest integer h, such that there are two disjoint homometric sets S_1, S_2 in G with $|S_1| = |S_2| = h$, is denoted by $h(G)$. For a family of graphs $\mathcal{G}$, $h(\mathcal{G})$ denotes the largest value of h such that for each graph G in $\mathcal{G}$, $h(G) \geq h$. Let $h(n)$ be $h(\mathcal{G}_n)$, where $\mathcal{G}_n$ is the set of all graphs on n vertices. In other words, $h(n) = \min\{h(G) : |V(G)| = n\}$. Albertson, Pach and Young [1] provided the most general bounds.

Theorem 1 ([1]). *$\frac{c \log n}{\log \log n} < h(n) \leq \frac{n}{4}$ for $n > 3$, and a constant c.*

It is an easy observation that $h(G) = \lfloor |V(G)|/2 \rfloor$ when G is a path or G is a cycle. However, more is known. Note that the multisets of distances for a vertex subset of a path corresponds to a multiset of pairwise differences between elements of a subset of positive integers. We shall say that two subsets of integers are homometric if their multisets of pairwise differences coincide. Among others, Rosenblatt and Seymour [8] proved that two multisets A and B of integers are homometric if and only if there are two multisets U, V of integers such that $A = U + V$ and $B = U - V$, where $U + V$ and $U - V$ are multisets, $U + V = \{u + v : \ u \in U, \ v \in V\}$, $U - V = \{u - v : \ u \in U, \ v \in V\}$. Lemke, Skiena and Smith [5] showed that if G is a cycle of length $2n$, then every subset of $V(G)$ with n vertices and its complement are homometric sets. Suprisingly, when the class $\mathcal{G}$ of graphs under consideration is not a path or a cycle, the problem of finding $h(\mathcal{G})$ becomes nontrivial. Even when $\mathcal{G}$ is a class of n-vertex graphs that are the unions of pairs of paths sharing a single point, $h(\mathcal{G})$ is not known. Here, we use standard graph-theoretic terminology, see for example [2] or [9].

The homometric set problem we consider here has its origins in Euclidean geometry, with applications in X-ray crystallography introduced in the 1930's with later applications in restriction site mapping of DNA. In particular, the fundamental problem that was considered is whether one could identify a given set of points from its multiset of distances. There are several related directions of research in the area, for example the question of recognizing the multisets corresponding to a multiset of distances realized by a set of points in the Euclidean space of given dimension, see [7].

In Section 2 we state our main results. In this paper, we provide the new bounds on $h(\mathcal{G})$ in terms of densities and diameter in Section 3, we also investigate this function for various classes of graphs, in particular for trees, in Section 4. Finally, in Section 5, we give additional bounds for $h(T)$, when T is a tree in terms of its parameters.

2 Main Results

Here, we slightly improve the known general bounds of $h(G)$, provide new bounds on $h(G)$ in terms of density, diameter, degree sequence, and give several results for $h(T)$ in case when T is a tree.

Theorem 2. *For infinitely many values of n, and a positive constant c, $h(n) \leq n/4 - c \log\log n$.*

Theorem 3. *Let G be a graph on n vertices with diameter d, maximum degree Δ, $n \geq 5$, $d \geq 2$. If for an integer k, $\binom{\binom{k}{2}+d-1}{d-1} < n - 2k + 2$, then*

$$h(G) \geq \max\{k,\ d/2,\ \sqrt{\Delta}\}.$$

In particular,

$$h(G) \geq \max\{0.5n^{1/(2d-2)},\ d/2,\ \sqrt{\Delta}\}$$

for $n \geq c(d) = d^{2d-2}$.

Since it is well known, see [2], that almost all graphs $G(n,p)$ have diameter 2, the above theorem implies that for almost every graph $G = G(n,p)$, $h(G_{n,p}) = \Omega(\sqrt{n})$.

Theorem 4 and Theorem 5 are corollaries of previously known results Theorems 10 and 9, by Caro and Yuster [3] and Karolyi [4], respectively.

Theorem 4. *For every fixed $\alpha > 0$ and for every $\epsilon > 0$ there exists $N = N(\alpha, \epsilon)$ so that for all $n > N$, if G is a graph on n vertices, diameter 2 and at most $n^{2-\alpha}$ edges or at least $\binom{n}{2} - n^{2-\alpha}$ edges, then $h(G) \geq n/2 - \epsilon n$.*

Theorem 5. *Let G be a graph with diameter 2 and even number of vertices $n \geq 90$. Let $V_i = \{v \in V : \deg(v) = i\}$. If G satisfies one of the following conditions:*
1) $|\{i : |V_i| \text{ is odd}\}| > n/2$,
2) $|V_i|$ is even for each i,
then $h(G) = n/2$.

Next, we provide the results for trees. In the following theorems, we omit ceilings and floors for simplicity. Let $\mathcal{T}_n$ be the set of all trees on n vertices. A *spider* is a tree that is a union of vertex-disjoint paths, called *legs* and a vertex that is adjacent to one of the endpoints of each leg, called the *head*. Let $\mathcal{S}_{n,k}$ be the set of n-vertex spiders with k legs and $\mathcal{S}_n$ be the set of all n-vertex spiders. A *caterpillar* is a tree, that is a union of a path, called *spine*, and leaves adjacent to the spine. Let $\mathcal{R}_n$ be the set of all caterpillars on n vertices. Finally, a *haircomb* is a tree, that consists of a path called the *spine* and a collection of vertex-disjoint paths, called legs, that have an endpoint on the spine. Let $\mathcal{H}_n$ be the set of haircombs on n vertices.

Theorem 6. *For a positive integer n, $h(\mathcal{T}_n) \geq n^{1/3} - 1$.*

Theorem 7. *For a positive integer n, $h(\mathcal{R}_n) \geq n/6$, $h(\mathcal{H}_n) \geq \sqrt{n}/2$.*

Theorem 8. *For positive integers n, k, such that $k < n$,*

$$h(\mathcal{S}_{n,k}) \geq \begin{cases} \frac{5}{12}n, & k = 3, \\ \frac{1}{3}n, & k = 4, \\ \left(\frac{1}{4} + \frac{3}{8k-12}\right)n, & k \geq 5. \end{cases}$$

Moreover, $h(\mathcal{S}_{n,n/2}) = (n+2)/4$, and $h(T) = n/2$ for any $T \in \mathcal{S}_{n,3}$ with legs on l_1, l_2, l_3 vertices if $l_1 + 1 = t(l_2 - l_3)$ for an odd integer $t \geq 1$ or if $l_1 = l_2 = l_3$.

3 Preliminary facts and proofs of general Theorems 2–5

We shall say that two integers are *almost equal* if they differ by $1, 0$, or -1.

Theorem 9 (Karolyi [4]). *Let X be a set of m integers, each between 1 and $2m-2$. If $m \geq 89$, then one can partition X into two sets, X_1 and X_2 of almost equal sizes such that the sum of elements in X_1 is almost equal to the sum of elements in X_2.*

Theorem 10 (Caro, Yuster [3]). *For every fixed $\alpha > 0$ and for every $\epsilon > 0$ there exists $N = N(\alpha, \epsilon)$ so that for all $n > N$, if G is a graph on n vertices and at most $n^{2-\alpha}$ edges then there are two vertex disjoint subgraphs of the same order and size with at least $n/2 - \epsilon n$ vertices in each of them.*

Definition 11. *For a graph H, we say that G is an (H, m, v)-**flower** with a path P if G is a vertex-disjoint union of H and an m-vertex path P with endpoint v, together with all edges between $V(H)$ and v.*

Lemma 12. *Let G be an (H, m, v)-flower with a path P. If S_1 and S_2 are homometric sets of G, where $|S_1| = |S_2| \geq 4$, then either $S_1 \cup S_2 \subseteq V(P) \cup \{u\}$, $u \in V(H)$ or $S_1 \cup S_2 \subseteq V(H) \cup \{v, v_1\}$, where v_1 is the neighbor of v in P.*

Proof. Let S_1, S_2 be homometric sets in G of size at least 4 each. Note that $D(V(H))$ consists of 1s and 2s.

Consider the vertex $x \in (S_1 \cup S_2) \cap V(P)$, farthest from v, at a distance at least 2 from v. Without loss of generality, $x \in S_1$. If there is an $x' \in S_1 \cap V(H)$ then there is no pair of vertices in S_2 with the distance equal to the distance between x and x'. Thus, $S_1 \cap V(H) = \emptyset$. Let a be the largest distance in $D(S_1)$, then $a \geq 3$ since $|S_1| \geq 4$. Let $y \in S_2 \cap V(H)$, let y' be at a distance a from y, $y' \in V(P)$. Note that a appears exactly once in $D(S_1)$ and thus it appears exactly once in $D(S_2)$. Since all vertices from $V(H)$ are at the same distance from y, there is exactly one vertex $y' \in S_2 \cap V(H)$. □

Proof of Theorem 2. For a fixed integer k, let $a_0, a_1, \ldots, a_k$ be a sequence of integers such that $a_0 = 1$, $a_1 \geq 5$ and each a_i, $i \geq 2$, is the smallest odd number satisfying

$$a_i > 4\left(1 + \sum_{j=1}^{i-1} \binom{a_j + 1}{2}\right).$$

Let $k \geq 2$ and $n = 2(a_1 + \cdots + a_k) - k/4$. Let H be a vertex-disjoint union of cliques on $a_1, a_2, \ldots, a_k$ vertices, respectively. Let the vertex sets of these cliques be $Q_1, Q_2, \ldots, Q_k$, respectively. Note that $a_j > 4\sum_{i=0}^{j-1} a_i$ for $j \geq 1$, which implies that

$$a_j \geq \frac{4}{5}\left(\sum_{i=0}^{j} a_i\right). \tag{1}$$

Let G be an $(H, n/2 - k/8, v)$-flower with a path P. Note that $k = c \log\log n$, for a constant c. This construction of G is inspired by an example given by Caro and Yuster in [3].

Let S_1, S_2 be largest homometric sets in G. If $|S_1| = |S_2| < 4$, then we are done. Otherwise, by Lemma 12, $S_1 \cup S_2 \subseteq V(P) \cup \{u\}$, $u \in V(H)$ or $S_1 \cup S_2 \subseteq V(H) \cup \{v, v_1\}$, where v_1 is the neighbor of v in P.

Case 1 $S_1 \cup S_2 \subseteq V(P) \cup \{u\}$, $u \in V(H)$.
Then $h(G) \leq (|V(P)| + 1)/2 = n/4 - k/16 + 1/2 = n/4 - c \log\log n$, for a positive constant c.

Case 2 $S_1 \cup S_2 \subseteq V(H) \cup \{v, v_1\}$, where v_1 is the neighbor of v in P.
Let $\{v_1\} = Q_0$, then $S_1 \cup S_2$ is a vertex subset of a join of vertex-disjoint cliques on sets $Q_0, \ldots, Q_k$ and

$\{v\}$. Since $G[V(H) \cup \{v, v_1\}]$ is a graph of diameter 2, S_1 and S_2 induce the same number of edges in G. For $i = 0, \ldots, k$, let $Q'_i = Q_i \cup \{v\}$ if $v \in S_1 \cup S_2$, and let $Q'_i = Q_i$ if $v \notin S_1 \cup S_2$.

If $v \notin S_1 \cup S_2$ and for all j, $|S_1 \cap Q'_j| = |S_2 \cap Q'_j|$, then, since Q_j has odd size, $(S_1 \cup S_2) \cap Q_j \neq Q_j$, and $h(G) \leq (|V(H)| - k)/2 = n/4 + k/16 - k/2 = n/2 - c' \log\log n$, for a positive constant c'. If $v \in S_1 \cup S_2$ and for all j, $|S_1 \cap Q'_j| = |S_2 \cap Q'_j|$, then $|S_1| \neq |S_2|$, since $v \in Q'_j$ for all j. So,we can assume that there is j, for which $|S_1 \cap Q'_j| \neq |S_2 \cap Q'_j|$. Let j be the largest such index and let $Q'_j = Q$.

Assume first that $j \leq k/2 - 2$ and $v \notin S_1 \cup S_2$. For each $i > j$, $(S_1 \cup S_2) \cap Q_i \neq Q_i$. Thus $h(G) \leq (n/2 + k/8 - k/2 + 2)/2 = n/4 - c \log\log n$, for a positive constant c.

Now assume that $j \leq k/2 - 2$ and $v \in S_1 \cup S_2$. Without loss of generality, let $v \in S_1$. Let $\mathbf{Q} = Q_0 \cup \cdots \cup Q_j$, $\mathbf{Q'} = Q_{j+1} \cup \cdots \cup Q_k$. Then $|S_2 \cap \mathbf{Q'}| - |S_1 \cap \mathbf{Q'}| \geq k - j - 1 \geq k/2 + 1$. Since $|S_1 \cap (\mathbf{Q} \cup \mathbf{Q'} \cup \{v\})| = |S_2 \cap (\mathbf{Q} \cup \mathbf{Q'} \cup \{v\})|$, $|\mathbf{Q}| \geq -|S_2 \cap \mathbf{Q}| + |S_1 \cap \mathbf{Q}| \geq k/2$. Therefore, (1) implies that $a_j/2 \geq \frac{4}{10}|\mathbf{Q}| \geq \frac{2}{5}(k/2) \geq k/5$.
So, we have that either $a_j/2 \geq k/5$ or that $j \geq k/2 + 2$. In any case, we have that $a_j/2 \geq k/5$.

If $|Q - (S_1 \cup S_2)| \geq a_j/2$, then $h(G) \leq (n/2 + k/8 - a_j/2)/2 \leq (n/2 + k/8 - k/5)/2 \leq n/4 - c \log\log n$, for a positive constant c. Otherwise, $|Q \cap (S_1 \cup S_2)| > a_j/2$. Then, the number of edges induced by $S_1 \cap Q$ differs from the number of edges induced by $S_2 \cap Q$ by at least $a_j/4$. The number of edges induced by $S_1 \cap (\mathbf{Q'} \cup \{v\})$ is the same as number of edges induced by $S_2 \cap (\mathbf{Q'} \cup \{v\})$. The total number of edges induced by $\mathbf{Q} \cup \{v\}$ is at most $1 + \sum_{i=1}^{j-1} \binom{a_i+1}{2} < a_j/4$. Thus S_1 and S_2 can not induce the same number of edges, a contradiction. □

Construction 13. *[1] Let G be a graph and $G' = (v_1, \ldots, v_{2t})$ be a shortest v_1,v_{2t}-path in G, for some $t \geq 1$. Let $S_1 = \{v_1, \ldots, v_t\}$, $S_2 = \{v_{t+1}, \ldots, v_{2t}\}$. Since both S_1 and S_2 induce shortest paths of the same length, they form homometric sets.*

Proof of Theorem 3. The fact that $2h(G) \geq d$ follows from Construction 13. For positive integers k, n ($k < n$), the *Kneser graph* $KG(n, k)$ is a graph on the vertex set $\binom{[n]}{k}$ whose edge set consists of pairs of disjoint k-sets. Lovász [6] proved that the chromatic number of the Kneser graph $KG(n, k)$ is $n - 2k + 2$. We fix $k \leq n/2$ and consider the Kneser graph $\mathcal{K} = KG(n, k)$. Considering a graph G with vertex set $[n]$, we define a coloring of $\mathcal{K}$ by letting the distance multiset of each k-subset of $V(G)$ be the color of the corresponding vertex in $\mathcal{K}$. Since any vertex pair in G has distance in $\{1, \ldots, d\}$, the number of possible colors that are used on $\mathcal{K}$ is at most $\binom{\binom{k}{2}+d-1}{d-1}$. If

$$\binom{\binom{k}{2}+d-1}{d-1} < n-2k+2, \tag{2}$$

then there are two adjacent vertices of the same color in $\mathcal{K}$ that correspond to a pair of disjoint k-subsets of $V(G)$ that are homometric.

Note that when $d=2$, we have $\binom{\binom{k}{2}+d-1}{d-1} = \binom{k}{2}+1 < n-2k+2$ for $k=\sqrt{n}$.

Let $v \in V(G)$ be a vertex of a maximum degree $\Delta(G)$. The closed neighborhood $N[v]$ induces a graph of diameter 2. Moreover, for any two vertices in $N[v]$, the distance between them in G is the same as in $G[N[v]]$. As before, we see that (2) with $d=2$ and $n=\Delta$ holds for $k \geq \sqrt{\Delta(G)}$. Therefore, $h(G[N[v]]) \geq \sqrt{\Delta(G)}$. This proves the first part of the theorem.

To prove the second statement of the theorem, we assume that $d \geq 3$ and $n \geq d^{2d-2}$ and show that for any k, such that $d/2 < k \leq 0.5n^{1/(2d-2)}$, the inequality (2) holds. Since $d-1 \leq k^2/2$, $2^{d-2} \leq (d-1)!$, and $k \leq 0.5n^{1/(2d-2)} \leq n/4$, we have that

$$\binom{\binom{k}{2}+d-1}{d-1} < \frac{\left(\binom{k}{2}+d-1\right)^{d-1}}{(d-1)!} \leq \frac{(k^2)^{d-1}}{(d-1)!} \leq \frac{(k^2)^{d-1}}{2^{d-2}} \leq \frac{n}{2^{d-2}} \leq \frac{n}{2} \leq n-2k+2.$$

Therefore, there are homometric sets of size k for any k, $d/2 \leq k \leq 0.5n^{1/(2d-2)}$. □

Proof of Theorem 4. In a graph with diameter at most 2, any two distinct vertices are at distance 1 or 2. Therefore, Theorem 10 implies this result. □

Proof of Theorem 5. First, we give a fact observed in [3] stating that if $A \subseteq V(G)$ and $\sum_{v\in A}\deg(v) = \sum_{v\notin A}\deg(v)$ then $|E(G[A])| = |E(G[V-A])|$. To see this, note that $|E(G[A])| = (1/2)((\sum_{v\in A}\deg(v)) - |E(A,V-A)|)$ and $|E(G[V-A])| = (1/2)((\sum_{v\in V-A}\deg(v)) - |E(A,V-A)|)$. So, the difference of these two numbers is $\frac{1}{2}[\sum_{v\in A}\deg(v) - \sum_{v\notin A}\deg(v)] = 0$.
Let $\mathcal{D} = \{0,1,\ldots,\Delta(G)\}$ and $i \in \mathcal{D}$. Let

$$V_i = \{v \in V : \deg(v) = i\}, \quad V_i = A_i \cup B_i \cup S_i,$$

where A_i, B_i, S_i are disjoint, $|S_i| \leq 1$ and $|A_i| = |B_i|$. In other words, split V_i into two equal parts if possible, and otherwise, put a remaining vertex into a set S_i.

Let

$$A' = \bigcup_{i \in \mathcal{D}} A_i, \quad B' = \bigcup_{i \in \mathcal{D}} B_i, \quad S = \bigcup_{i \in \mathcal{D}} S_i.$$

So, $V(G) = A' \cup B' \cup S$. Let $s(G) = |S|$. We shall show that there are disjoint subsets A, B of $n/2$ vertices each, such that $\sum_{v \in A} \deg(v) = \sum_{v \notin A} \deg(v)$, i.e., A and B induce the same number of edges in G.

If $|V_i|$ is even for each i, let $A = A'$ and $B = B'$.

Note that all vertices in S have distinct degrees $a_1, a_2, \cdots, a_m$, say $1 \leq a_1 < a_2 < \cdots < a_m < n$, and since $m > n/2$, $n < 2m$. Note also that $m = |S|$ is even, since n is even and $|A'| + |B'|$ is even. So, we could apply Theorem 9 and split $\{a_1, \ldots, a_m\}$ in two parts, U and U' of equal sizes and with almost equal sums. Since $\sum_{v \in V} \deg(v) = \sum_{v \in A'} \deg(v) + \sum_{v \in B'} \deg(v) + \sum_{v \in S} \deg(v) = 2 \sum_{v \in A'} \deg(v) + \sum_{v \in S} \deg(v)$, and this degree-sum is even, it follows that $\sum_{v \in S} \deg(v)$ is even. Thus $\sum_{i=1}^{m} a_i = \sum_{v \in S} \deg(v)$ is even, and the sum of elements in U is exactly equal to the sum of elements in U'. Let

$$A'' = \{v \in S : \deg(v) \in U\}, \quad B'' = \{v \in S : \deg(v) \in U'\}.$$

Finally, let $A = A' \cup A''$ and $B = B' \cup B''$. □

4 Construction and proofs of the Theorems 6–8 for trees

In the remaining part of the paper, we may omit the floor and ceiling of fractions for simplicity. For a tree T, and its vertex r, let $P = P(T, r)$ be a partial order on the vertex set of a tree T, such that $x < y$ in P if the x,r-path in T contains y. We call this an *r-order* of T and say that T is r-ordered. A vertex y is a *parent* of x in $P(T, r)$, and a vertex x is a *child* of y if $x < y$ and there is no other element z such that $x < z < y$. If neither $x < y$ nor $y < x$ for two elements x and y in $P(T, r)$, then we say that x and y are *noncomparable*. An *antichain* is defined as a set of pairwise noncomparable elements. A pair of vertices x and y are called *siblings* if they have the same parent.

Construction 14. *Let T be a tree and S be an antichain in $P(T, r)$ such that each vertex in S, has a sibling in S. Let $S'_1, \ldots, S'_k$ be the maximal families of siblings in S. Let $S'_i = A_i \cup B_i \cup C_i$, where A_i, B_i, C_i*

are disjoint, $|A_i| = |B_i|$, $|C_i| \leq 1$, $i = 1, \ldots, k$. *Let* $S_1 = A_1 \cup A_2 \cup \cdots \cup A_k$, $S_2 = B_1 \cup B_2 \cup \cdots \cup B_k$. *The sets* S_1 *and* S_2 *are homometric. See Figure 1.*

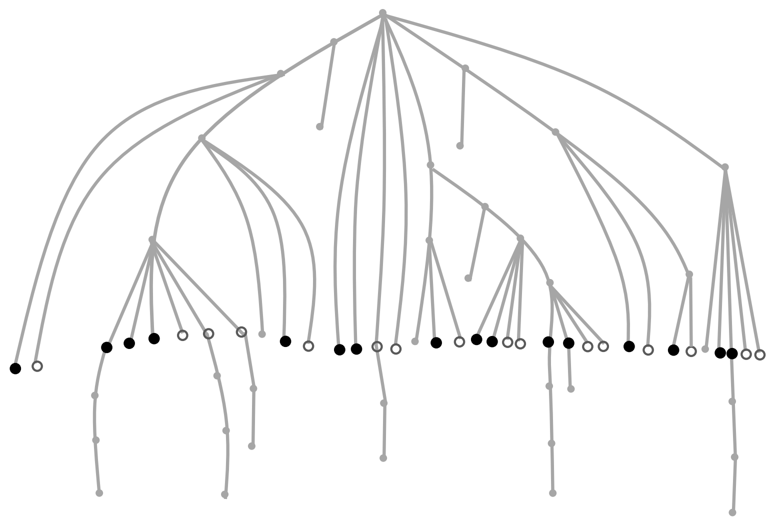

Figure 1: Example of Construction 14 with S_1 and S_2 consisting of black and hollow vertices, respectively.

Construction 15. *Let* $T \in \mathcal{S}_{n,3}$ *with head* v *and legs* L_1, L_2, L_3 *on* l_1, l_2, l_3 *vertices, respectively,* $l_1 \geq l_2$. *Let* L_3 *be a path* $v_1, \ldots, v_{l_3}$, *where* v_{l_3} *is a leaf of* T. *Let* S_1 *be the set consisting of* v, *the vertices* v_{2i}, $1 \leq i \leq \lfloor l_3/2 \rfloor$ *and* $l_2 - 1$ *vertices in* L_1 *closest to* v. *Let* S_2 *be the set consisting of the vertices* v_{2i-1}, $1 \leq i \leq \lfloor l_3/2 \rfloor$ *and* $V(L_2)$. *The sets* S_1 *and* S_2 *are homometric (see Figure 2(b)).*

Construction 16. *Let* $T \in \mathcal{S}_{n,3}$ *with head* $v = v_0$ *and legs* L_1, L_2, L_3 *on* ℓ_1, ℓ_2, ℓ_3 *vertices, respectively, and let one leg be longer than the other, say* $l_1 \geq l_2 > l_3$. *Let* L_1 *be a path* $v_1, \ldots, v_{l_1}$, *where* v_{l_1} *is a leaf of* T. *Let* $l_2 - l_3 = x > 0$, $l_1 + 1 = bx + r$, *where* b *and* r *are integers,* $0 \leq r < x$. *If* b *is odd, let* $a = b - 1$ *and if* b *is even, let* $a = b$. *Let* $P_i = \{v_{ix}, \ldots, v_{ix+x-1}\}$ *for* $0 \leq i \leq a$. *Let* S_1 *be the union of* P_{2i}, $0 \leq i \leq a/2$ *and* $V(L_3)$. *Let* S_2 *be the union of* P_{2i-1}, $1 \leq i \leq a/2$ *and* $V(L_2)$. *The sets* S_1 *and* S_2 *are homometric (see Figure 2(c)).*

Construction 16 is not used in any of the proofs here. However, it hints that for a tree $T \in \mathcal{S}_{n,3}$, $h(T)$ can be very close to $n/2$, depending on the optimized value of x, whereas other construction never provide a value close to $n/2$ when the leg lengths are different.

Construction 17. *Let T be a spider with head v and k legs $L_1, L_2, \ldots, L_k$, $k \geq 3$, on $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_k$ vertices, respectively. Let $A_i = V(L_{2i})$, let B_i be the set of $|A_i|$ vertices closest to v in L_{2i-1} , $i = 1, \ldots, \lfloor k/2 \rfloor = m$. Then let $S_1 = A_1 \cup A_2 \cup \cdots \cup A_m$, $S_2 = B_1 \cup B_2 \cup \cdots \cup B_m$. Clearly S_1 and S_2 are homometric sets (see Figure 2(d)).*

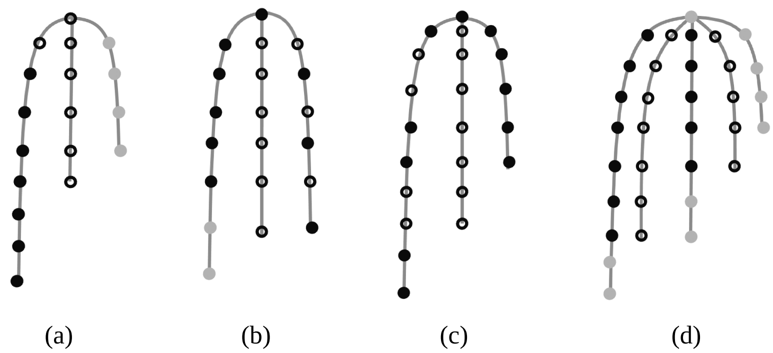

Figure 2: Examples of homometric sets in spiders using Constructions 13, 15, 16 (with $x = 2$, $a = 4$) and 17, respectively. The sets S_1 and S_2 consist of black and hollow vertices, respectively.

The following lemma will be used later in the proofs of this section.

Lemma 18. *Let T be a tree, r be a vertex, and S be an antichain in $P(T, r)$ such that each vertex in S, has a sibling in S. Let $k' = k'(S)$ be the number of maximal odd sets of siblings in S. Then*

$$2h(T) \geq \max \left\{ \operatorname{diam}(T) + 1,\ |S| - k',\ \frac{2}{3}|S| \right\}.$$

Proof. The fact that $2h(T) \geq \operatorname{diam}(T) + 1$ and $2h(T) \geq |S| - k'$ follows immediately from Constructions 13 and 14. Since each maximal family of siblings in S has at least two elements, $k' \leq |S|/3$, thus $2h(T) \geq 2|S|/3$.

□

Proof of Theorem 6. Assume that v is a vertex contained in the center of T and let $t = \lceil \operatorname{diam}(T)/2 \rceil$. Let $P = P(T, v)$. We define a partition $V(T) = N_0 \cup N_1 \cup N_2 \cup \cdots \cup N_t$, where $N_i = N_i(v)$. So, N_i is an antichain in P, $i = 1, \ldots, t$. Since $h(T) \geq t$, we can assume that $t \leq n^{1/3} - 1$, and thus

$\mathrm{diam}(T) \leq 2n^{\frac{1}{3}} - 2$. Let x be the smallest integer such that $|N_i| \leq |N_{i-1}| + x - 1$ for $1 \leq i \leq t$. Then $n = 1 + \sum_{i=1}^{t} |N_i| \leq 1 + \sum_{i=1}^{t} ix \leq x(t+1)^2/2$.

Since $t \leq n^{1/3} - 1$, by the above inequality we have $x \geq 2n^{1/3}$. For some j, $1 \leq j \leq t$, $|N_j| - |N_{j-1}| = x - 1$. Let S be a largest subset of N_j such that each vertex in S has an odd number of siblings in S. So, $|S| \geq |N_j| - |N_{j-1}| = x - 1 \geq 2n^{1/3} - 1$. By Lemma 18, $h(T) \geq n^{1/3} - 1$.

□

Proof of Theorem 7. Let T be a caterpillar on n vertices and assume that its spine P has at least $n/3$ vertices. Then, by Theorem 3, $h(T) \geq n/6$. If P has less than $n/3$ vertices, then there are at least $2n/3$ leaves. Assume that there are exactly k vertices on the spine with degree at least 4, label them as $v_1, \ldots, v_k$. Apply Lemma 18 to a set S consisting of leaves incident to a vertex in $v_1, \ldots, v_k$. Then $k'(S) \leq k$ and $2h(T) \geq |S| - k'(S) \geq (n - |V(P)|) - k \geq n - 2|V(P)| \geq \frac{n}{3}$.

Let T be a haircomb with m vertices on its spine and k spinal vertices, where $k \leq m$. We denote the length of the ith leg of T with l_i. Since $(k+1)\max(l_1, \ldots, l_k, m) \geq m + \sum_1^k l_i = n$, either $k + 1 \geq \sqrt{n}$ implying $m \geq k + 1 \geq \sqrt{n}$ or $l_i \geq \sqrt{n}$ for some i, which implies that $\mathrm{diam}(T) \geq \sqrt{n}$. By Theorem 3, $2h(T) \geq \sqrt{n}$. □

Proof of Theorem 8.
Let m be an integer such that $k = 2m$ or $k = 2m + 1$. Using Construction 17, we see that

$$2h(G) \geq n - [(l_1 - l_2) + (l_3 - l_4) + \ldots + (l_{2m-1} - l_{2m}) + x + 1] \geq n - l_1 + l_{2m} - x - 1, \tag{3}$$

where $x = 0$ if k is even and $x = l_k$ if k is odd. This observation is used in the following cases.

- Let $k = 3$. Let $\min(l_1 - l_2, l_2 - l_3) = cn$ for some $c \geq 0$, where $l_1 \geq l_2 \geq l_3$. By Construction 15, we have $2h(T) \geq n - cn$. Without loss of generality, assume that $l_1 - l_2 = cn$. Then $l_1 = l_2 + cn$ and $l_2 \geq l_3 + cn$ by our assumption and therefore, $n = l_1 + l_2 + l_3 + 1 \geq 3l_3 + 3cn$. This implies that $l_3 \leq (n - 3cn)/3$ and by Lemma 18, $2h(T) \geq l_1 + l_2 + 1 = n - l_3 \geq 2n/3 + cn$. Thus
$$2h(\mathcal{S}_{n,3}) \geq \min_{0 \leq c \leq 1} \max\left\{n - cn, \frac{2n}{3} + cn\right\} \geq 5n/6.$$
- Let $k = 4$. We have $2h(T) \geq \mathrm{diam}(T) + 1 \geq l_1 + l_2 + 1$ and (3) provides that $2h(T) \geq n - (l_1 - l_4) - 1$. Adding these inequalities gives that $4h(T) \geq n + l_2 + l_4$. By letting $l_2 + l_4 = c'n$, for some c', $0 \leq c' \leq 1$,

we rewrite this bound as $4h(T) \geq (1+c')n$. On the other hand, $2h(T) \geq \text{diam}(T)+1 \geq l_1+l_3+1 = (1-c')n$. Thus,

$$2h(\mathcal{S}_{n,4}) \geq \min_{0\leq c'\leq 1} \max\left\{\frac{(1+c')n}{2}, (1-c')n\right\} \geq 2n/3.$$

- Let $k \geq 5$. Let T be a spider on n vertices with k legs having $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_k$ vertices, respectively, and the head v. Observe that $l_2 \geq (n-l_1-1)/(k-1)$. Assume that $2h(T) = cn$ for some $c > 0$. By Lemma 18, $2h(T) = cn \geq \text{diam}(T) + 1 \geq l_1 + l_2 + 1 \geq l_1 + (n-l_1-1)/(k-1) + 1$. Thus, $-l_1 \geq n(1-ck+c)/(k-2)+1$. Using (3), $cn = 2h(T) \geq n-l_1-1 \geq n+\frac{n(1-ck+c)}{k-2}$. Since k is fixed, this implies that $c \geq 1/2+3/(4k-6)$, i.e., $h(T) \geq n/4+3n/(8k-12)$.

  The bound $h(\mathcal{S}_{n,k}) \geq (n+2)/4$ for $k \geq 5$ is attained by a spider T with $n/2$ single-vertex legs, and one leg, P, with $n/2-1$ vertices. If H is an empty graph on $n/2$ vertices, then T is an $(H, n/2-1, v)$-flower, where v is the head of T. Let S_1, S_2 be homometric sets of T. Then by Lemma 12, $S_1 \cup S_2 \subseteq V(H) \cup \{v, v_1\}$ or $S_1 \cup S_2 \subseteq V(P) \cup \{v, u\}$, where $u \in V(H)$, $v_1 \in V(P)$ is adjacent to v. In the first case, we can easily see that $v \notin S_1 \cup S_2$, so $2h(T) \leq n/2+1$.

□

5 More results on trees

For a vertex x of degree at least 3 in a tree T, the connected components of $T-x$ that are paths are called *pendent* paths of x. The endpoints of these paths adjacent to x in T are called *attachment* vertices. A vertex x of degree at least 3 is called *bad* if it has an odd number of pendent paths. Moreover, let a shortest pendent path corresponding to a bad vertex x be called a *bad path.*

We call a tree T' a *cleaned* T if T' is obtained from T by removing the vertices of all bad paths. A tree T'' is called *trimmed* T if it is obtained from T by removing the vertices of all bad paths and removing all the vertices except for the attachment vertices of all remaining pendent paths of T.

Let $bad(T)$, $bad_3(T)$ be the number of bad vertices, and the number of bad vertices of degree 3 in T, respectively. Let $bad_\Sigma(T)$ be the total number of vertices in bad paths of T. Let $N_i(x)$ be the set of vertices at distance i from a vertex x. For two disjoint sets $A', A'' \subseteq V(G)$, let $D(A', A'')$ be the multiset of distances in G between pairs u' and u'', $u' \in A'$, $u'' \in A''$.

In the following, we extend Lemma 18 to a more general case.

Lemma 19. *Let T be a tree, r be a vertex, and S be an antichain in $P(T,r)$ such that each vertex in S, has a sibling in S. Let $k' = k'(S)$ be the number of maximal odd sets of siblings in S. Then*

$$2h(T) \geq \max\left\{\mathrm{diam}(T)+1,\ |S|-k',\ \frac{2}{3}|S|,\ d_1(T)-bad(T),\ \frac{n}{\mathrm{diam}(T)}-bad(T),\ \frac{n-bad_\Sigma(T)}{\mathrm{diam}(T)}\right\}.$$

Proof. The first three bounds are proved in Lemma 18. For any tree $\mathcal{T}$, Dilworth's theorem applied to $P(\mathcal{T})$ implies that the size of a maximum antichain in $P(\mathcal{T})$ is at least $n/\mathrm{diam}(\mathcal{T})$. The set of leaves is a maximum antichain of $P(\mathcal{T})$, thus $d_1(\mathcal{T}) \geq n/\mathrm{diam}(\mathcal{T})$.

Let T' be trimmed T. Since $T' \subseteq T$ and all distances in T are preserved in T', $h(T) \geq h(T')$. Let S be the set of leaves of T'. Note that $|S| = d_1(T') = d_1(T) - bad(T)$. Moreover, since T' has no bad vertices, $k'(S) = 0$ in T'. So, $2h(T) \geq |S| - k'(S) = d_1(T) - bad(T) \geq n/\mathrm{diam}(T) - bad(T)$.

Let T'' be cleaned T. We have that $|V(T) \setminus V(T')| = bad_\Sigma(T)$, moreover, T'' has no bad vertices. Thus $2h(T) \geq 2h(T'') \geq |d_1(T'')| \geq (n - bad_\Sigma(T))/\mathrm{diam}(T'') \geq (n - bad_\Sigma(T))/\mathrm{diam}(T)$. □

For a tree T, let $d_i = d_i(T)$ be the number of vertices of degree i.

Lemma 20.
a) $d_1 - bad(T) \geq 2 + \sum_{i\geq 4} d_i$,
b) $d_1 \geq 2/3(n - d_2 - d_3)$, *and*
c) $\sum_{i\geq 4} d_i \leq (1/3)(n - d_2 - d_3)$.

Proof. Observe first that $d_1 = 2 + \sum_{i\geq 3} d_i(i-2)$. Thus $d_1 \geq 2 + \sum_{i\geq 4} 2d_i$.

Since bad vertices have degree at least 3, we have that $bad(T) \leq d_3 + \sum_{i\geq 4} d_i$, so $d_1 - bad(T) \geq 2 + \sum_{i\geq 4} d_i$. We also have that $d_1 \geq \sum_{i\geq 4} 2d_i = 2(n - d_1 - d_2 - d_3)$, thus $d_1 \geq 2/3(n - d_2 - d_3)$. To show the last inequality, observe that $\sum_{i\geq 4} d_i = n - d_1 - d_2 - d_3$, so using (b), $\sum_{i\geq 4} d_i = n - d_1 - d_2 - d_3 \geq n - 2/3(n - d_2 - d_3) - (d_2 + d_3) = 1/3(n - d_2 - d_3)$. □

Theorem 21. *Let T be a tree on n vertices. Then $h(T) = \Omega(\sqrt{n})$ if one of the following holds:*
a) $bad(T) = o(\sqrt{n})$,
b) $bad_3(T) = o(\sqrt{n})$ and $bad(T) = \Omega(\sqrt{n})$,
c) $bad_\Sigma(T) = o(n)$.
Moreover $h(T) \geq \max\{(n - d_2 - 4d_3)/6, (n - d_1 - d_2)/(\mathrm{diam}(T) + 1)\}$.

Comment. *There are trees on n vertices, such as double haircombs, for which our proof techniques do not provide bounds of the order of magnitude $\sqrt{n}$. A double haircomb is constructed from a haircomb by attaching vertex-disjoint paths to the vertices of the legs, see Figure 3. By appropriately choosing the distances between the vertices of degree 3, one can construct an n-vertex double haircomb T, such that $bad_3(T) \geq c\sqrt{n}$, $bad_\Sigma(T) \geq c'n$ for some constants $c, c' > 0$ and $\mathrm{diam}(T) = o(\sqrt{n})$.*

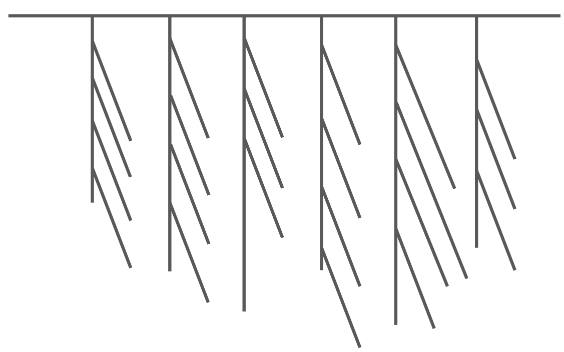

Figure 3: A double haircomb.

Proof of Theorem 21. To prove the first part of the theorem, we assume throughout the proof that $\mathrm{diam}(T) = o(\sqrt{n})$, otherwise by Lemma 19, $h(T) = \Omega(\sqrt{n})$.

- $bad(T) = o(\sqrt{n})$. By Lemma 19, $2h(T) \geq n/\mathrm{diam}(T) - bad(T) = \Omega(\sqrt{n})$.
- $bad_3(T) = o(\sqrt{n})$ and $bad(T) = \Omega(\sqrt{n})$.
  By Lemma 20, $d_1 - bad(T) \geq \sum_{i \geq 4} d_i \geq bad(T) - bad_3(T) = \Omega(\sqrt{n})$. Lemma 19 implies that $2h(T) = \Omega(\sqrt{n})$.
- $bad_\Sigma(T) = o(n)$. By Lemma 19, $h(T) \geq (n - bad_\Sigma(T))/\mathrm{diam}(T) = \Omega(\sqrt{n})$.

To prove the second part of the theorem, we use Lemma 19 and Lemma 20-(b),(c). We have that $2h(T) \geq d_1 - bad(T) \geq (d_1 - \sum_{i\geq 4} d_i) - d_3 \geq 1/3(n - d_2 - d_3) - d_3 = 1/3(n - d_2 - 4d_3)$.

Let S be the set of vertices with degree at least 3 in T, $|S| = n - (d_1 + d_2)$. Let r be a leaf in T, $V(T) = N_0 \cup N_1 \cup N_2 \cup \cdots \cup N_t$, where $N_0 = \{r\}$, $N_i = N_i(r)$ and $t = \lceil \text{diam}(T)/2 \rceil$. There is an $i \in [t]$ such that $|N_i(r) \cap S| \geq |S|/t$. Let $L = N_i(r) \cap S$. We have $|L| \geq (n - d_1 - d_2)/t \geq 2(n - d_1 - d_2)/(\text{diam}(T) + 1)$. Let the set of children of L with respect to $P(T, r)$ be L'. Since each vertex in S and thus in L has degree at least three, $|L'| \geq 2|L|$. Using Lemma 19 and the fact that $k'(L') \leq |L|$, we have $h(T) \geq (|L'| - k'(L'))/2 \geq (2|L| - |L|)/2 = |L|/2 \geq (n - d_1 - d_2)/(\text{diam}(T) + 1)$. □

6 Concluding Remarks

The definition of homometric sets allows for little control over what pairs of vertices realize what distance. This makes proving the upper bounds on $h(n)$ difficult. One may consider another definition for two sets being homometric. For a graph G, let $K(G) = (V(G), c)$ be a complete graph on vertex set $V(G)$ with an edge-coloring c, where $c(u, v)$ is equal to the distance between u and v in G. Let two disjoint sets $S_1, S_2 \subset V(G)$ be *similar* if there is an isomorphism between $K(G)[S_1]$ and $K(G)[S_2]$. Note that there may be homometric sets that are not similar. Let T be a spider with four long legs of equal length, let S_1 consist of four vertices on distinct legs with distances 2,2,2,6 to the head of the spider, respectively, let S_2 be the set of four vertices - one is the head, three other are on three distinct legs with distance 4 to the head. Although $D(S_1) = D(S_2) = \{4, 4, 4, 8, 8, 8\}$, S_1 and S_2 are not similar, since there is a triangle with all edges colored 4 in $K(G)[S_1]$ and there is a triangle with all edges colored 8 in $K(G)[S_2]$.

7 Acknowledgements

The authors thank Michael Young for pointing out one of the special cases in Theorem 8 and Radoslav Fulek and Slobodan Mitrović for finding a minor error in the proof of Theorem 5. The authors thank the referees for reading the paper carefully, spotting a minor error in Lemma 12, and providing suggestions that improved the presentation of a paper.

References

[1] M. Albertson, J. Pach, and M. Young, *Disjoint homometric sets in graphs*, Ars Math. Contemp. **4** (2011), no. 1, 1–4.

[2] B. Bollobás, *Modern graph theory*, Graduate Texts in Mathematics, vol. 184, Springer-Verlag, New York, 1998.

[3] Y. Caro and R. Yuster, *Large disjoint subgraphs with the same order and size*, European J. Combin. **30** (2009), no. 4, 813–821.

[4] G. Károlyi, *Balanced subset sums in dense sets of integers*, Integers **9** (2009), A45, 591–603.

[5] P. Lemke, S. S. Skiena, and W. D. Smith, *Reconstructing sets from interpoint distances*, Discrete and Comput. Geom., Algorithms Combin., vol. 25, Springer, Berlin, 2003, pp. 507–631.

[6] L. Lovász, *Kneser's conjecture, chromatic number, and homotopy*, J. Combin. Theory Ser. A **25** (1978), no. 3, 319–324.

[7] J. Matoušek, *Thirty-three miniatures*, Student Mathematical Library, vol. 53, American Mathematical Society, Providence, RI, 2010, Mathematical and algorithmic applications of linear algebra.

[8] J. Rosenblatt and P. D. Seymour, *The structure of homometric sets*, SIAM J. Algebraic Discrete Methods **3** (1982), no. 3, 343–350.

[9] D. B. West, *Introduction to graph theory*, Prentice Hall Inc., Upper Saddle River, NJ, 1996.