Ph.D. Working Notes

February 24, 2023

Contents

1 Crescent Labelled Trees

Let T be a tree of order n. A crescent labelling of T is a map $L : E(T) \mapsto \{1, 2, \dots, t\}$, such that the distance multiset of $L(T)$ is of the form $\{d_1^1, d_2^2, \dots, d_{n-1}^{n-1}\}$. The diameter of T, denoted $\mathrm{Diam}(T)$, is the length of the (u, v)-path in T. The max degree of T is denoted $\Delta(T)$.

Lemma 1 (Basic Diameter Lower Bound)**.** *Let t be a positive integer. If $L(T)$ is a crescent labelling of the tree T with weights $\{1, 2, \dots, t\}$, then $\mathrm{Diam}(T) \geq \frac{n-1}{t}$.*

Proof. Since there are at least $n - 1$ distinct distances, there is a distance d with value at least $n - 1$. Let $u, v \in V(T)$ such that $d(u, v) = d$, then since t is the max edge weight, this means that the number of edges on a (u, v)-path is at least $\frac{d}{t} \geq \frac{n-1}{t}$. □

For a pair of vertices $u, v \in V(T)$, we denote the (u, v)-path in T as $P(u, v)$. Lemma 2 below generalizes the observation underlying the maximum degree upper bound of $\sim \sqrt{2n}$.

Lemma 2. *Let T be a tree of order n. For every $i \in [1, n - 1]$, $M \in V(T)$, and $j \in \mathcal{N}(M)$, define*

$$D_j := \{u \in V(T) \setminus \{M\} : d(u, M) = d_i, j \in P(u, M)\}.$$

Then distance $2d_i$ occurs with multiplicity at least $\sum_{j<k}^{\deg(M)} |D_j| \cdot |D_k|$.

Proof. Let $M \in V(T)$ and $i \in [1, n - 1]$. Since T is a tree, there is always a unique (u, v)-path for all $u, v \in V(T)$. So, for each $u \in D_j$ and $v \in D_k$, the (u, v)-path must go through M, which means $d(u, v) = d(u, M) + d(M, v) = 2d_i$. There are $|D_j| \cdot |D_k|$ such u and v pairs, so indeed $2d_i$ has multiplicity at least $\sum_{j<k}^{\deg(M)} |D_j| \cdot |D_k|$. □

Now we apply the lemma to get a condition on crescent labelled trees.

Proposition 1.1 (Max Multiplicity Condition)**.** *Let $L(T)$ be a crescent labelling of a tree T. Then for every $i \in [1, n-1]$, $M \in V(T)$, and $j \in \mathcal{N}(M)$,*

$$\sum_{j<k}^{\deg(M)} |D_j| \cdot |D_k| < n.$$

Proof. Since $L(T)$ is a crescent labelling of T, no distance can have multiplicity greater than $n-1$and T is a tree. Since T is a tree, it follows by Lemma 2 that for each vertex $M \in V(T)$, $i \in [1, n-1]$, $\sum_{j<k}^{\deg(M)} |D_j| \cdot |D_k| < n$. □

Next is a general lemma on $\{1, 2, \dots, t\}$-words containing subwords with $t-1$ consecutive $1s$.

Lemma 3 (Arithmetic Condition)**.** *Let $k \geq 2t$. Let $\mathbf{w}$ be a $\{1, 2, \dots, t\}$-word with length k. If $w_{a-t+2} = w_{a-t+3} = \cdots = w_a = 1$ for some $a \in \{1, 2, \dots, k\}$, then each value*

$$1, 2, \dots, \max\left\{\sum_{i=1}^{a} w_i, \sum_{i=a-t+2}^{k} w_i\right\}$$

occurs as a partial sum in $\mathbf{w}$.

Proof. Suppose without loss of generality that $\sum_{i=1}^{a} w_i \leq \sum_{i=a-t+2}^{k} w_i$. Then it is sufficient to show that every value $1, 2, \dots, \sum_{i=a-t+2}^{k} w_i$ occurs as a partial sum in $\mathbf{w}$. Call $w_{a-t+2}, w_{a-t+3}, \dots, w_a$ the *unit segment* of $\mathbf{w}$ and $w_{a+1}, w_{a+2}, \dots, w_k$ the *non-unit segment* of $\mathbf{w}$. We proceed by induction on the number of terms r in the non-unit segment of $\mathbf{w}$. When $r = 1$, $w_{a+r} \in \{1, \dots, t\}$, and since the unit segment has $t-1$ 1s, for each $j \in \{1, 2, \dots, t-1\}$, we have the partial sums $j = \sum_{i=0}^{j-1} w_{a-i}$. Then the values between w_{a+r} and $\sum_{i=a-t+2}^{a+r} w_i$ are of the form $w_{a+r} + \sum_{i=0}^{j-1} w_{a-i}$. For the inductive step, the values $1, 2, \dots, \sum_{i=a-t+2}^{a+r-1} w_i$ occur at least once by inductive hypothesis. We have that $w_{a+r} \in \{1, 2, \dots, t\}$ and the values between $\sum_{i=a+1}^{a+r-1} w_i$ and $\sum_{i=a+1}^{a+r} w_i$ can be obtained from $\sum_{i=a+1-j}^{a+r-1} w_i$ for each $j \in \{1, 2, \dots, t-1\}$. Then similarly the values between $\sum_{i=a+1}^{a+r} w_i$ and $\sum_{i=a-t+2}^{a+r} w_i$ are $\sum_{i=a+1-j}^{a+r} w_i$ for $j \in \{1, 2, \dots, t-1\}$. □

We now apply this arithmetic lemma to crescent labelled trees to show that when there are many consecutive 1s on a path, the path cannot be too long with many large weight edges.

Proposition 1.2. *Let $L(T)$ be a crescent labelling of a tree T with edge weights in $\{1, 2, \dots, t\}$. Then for every path $P = (v_1v_2, v_2v_3, \dots, v_{t-1}v_t)$ in T such that $w(v_iv_{i+1}) = 1$ for $i \in \{1, 2, \dots, t-1\}$, it follows that $\max\{d(v_1, u) : u \in V(T)\} < n$ and $\max\{d(v_t, u) : u \in V(T)\} < n$.*

Proof. Let T be a tree with a path P specified in the proposition statement and $L(T)$ a crescent labelling. It is sufficient to show that $\max\{d(v_1, u) : u \in V(T)\} < n$ since the case for v_t is similar. Let $u' \in V(T)$ such that $d(v_1, u') = \max\{d(v_1, u) : u \in V(T)\}$. By Lemma 3, every distance $1, 2, \dots, d(v_1, u')$ occurs at least once. Since $L(T)$ is a crescent labelling, there can be at most $n-1$ distinct distances, so $d(v_1, u') < n$ as desired. □

The implication for when $t = 2$ is quite strong since this imposes a max distance condition on vertices incident to edges with weight 1.

Corollary 1.3. *Let $L(T)$ be a crescent labelling of a tree T. If $t = 2$, then every vertex incident to an edge with weight 1 has max distance at most $n - 1$.*

What follows is a basic lemma about trees that may turn out to be useful in case parameterizing by number of leaves becomes sensible.

Lemma 4 (From Chartrand and Lesniak's text "Graphs and Digraphs" 4th edition)**.** *Let T be a tree with n_i vertices with degree i, where $i \in \{1, 2, \ldots, \Delta(T)\}$. Then $n_1 = n_3 + 2n_4 + 3n_5 + \cdots + (\Delta(T) - 2)n_{\Delta(T)} + 2$.*

Proof. Note that $n = \sum_{i=1}^{\Delta(T)} n_i$. Since T is a tree,

$$\sum_{i=1}^{\Delta(T)} in_i = \sum_{v \in V(T)} \deg(v) = 2(n-1) = 2\Bigg(\sum_{i=1}^{\Delta(T)} n_i\Bigg) - 2.$$

Rearranging gives $2 + \sum_{i=1}^{\Delta(T)} (i-2)n_i = 0$. □

Corollary 1.4. *If T is a tree, then $\sum_{i=3}^{\Delta(T)} (i-2)n_i < n_1$*

2 Polynomial Method

Let $L(G)$ be a crescent labelling of a graph G with corresponding distance multiset $\{d_1^1, d_2^2, \ldots, d_{n-1}^{n-1}\}$. Consider the bipartite multigraph $\mathcal{M} := \mathcal{M}(G)$ where $V(\mathcal{M}) = X \cup Y$, where X consists of the distinct distances $d_1, d_2, \ldots, d_{n-1}$ and Y consists of the vertices of G, $v_1, v_2, \ldots, v_n$. For $d_k \in X$ and $v_i \in Y$, an edge $d_k v_i \in E(\mathcal{M})$ is included for every $j \in [n]$ such that $d(v_i, v_j) = d_k$. Note that since $L(G)$ is a crescent labelling, for each $k \in [n-1]$, $\deg(d_k) = 2k$. Observe also that the multiset neighbourhood of v_i is the multiset of the $n-1$ distances between v_i and the other vertices in G.

We show a variation of a result from Alon (see proof of Theorem 6.1 in [1]) about the existence of p-regular subgraphs of a multigraph whose average degree is very close to its max degree. If we relax this strong average degree condition, we can still obtain a rather powerful result whereby a subgraph U of $\mathcal{M}$ has vertex degrees in $\{p, 2p, 3p, 4p, 5p, 6p, 7p\} \cap [2(n-1)]$, where $p \in [\frac{n}{4}, \frac{n}{2}]$, which is significant because $\mathcal{M}$ is bipartite and so the structure of U can tell us some things about how the distances relate to the vertices in G. This subgraph U likely can't be too small, since then there would be a vertex $v \in G$ with too many other vertices at some distance d from v.

Remark 2.1. Relating the size of this U to the structure of G might be a fruitful way to proceed. For instance, paths require $|U|$ to be quite large (no vertex is at distance d with more than 2 other vertices for each d). I think stars might be similar in that they require $|U|$ to be rather large. Perhaps if $p \sim n/4$, or even asymptotically when $p \sim n/2 - (n/2)^{0.525}$ or so, U being large with min degree p forces convergence of crescent labelled trees to paths and stars. But I admit, I'm not really sure right now what to do when $|U|$ is big.

The proof applies Alon's combinatorial nullstellensatz [1]. The corollary of the nullstellensatz that we use is as follows:

Lemma 5 (Combinatorial Nullstellensatz)**.** *Let $\mathbb{F}$ be a field and let $f \in \mathbb{F}[x_1, x_2, \ldots, x_n]$ be a polynomial such that $\deg(f) = \sum_{i=1}^{n} t_i$ and the coefficient of $\prod_{i=1}^{n} x_i^{t_i}$ is non-zero. Let $S_1, S_2, \ldots, S_n$ be subsets of $\mathbb{F}$ such that $|S_i| > t_i$ for all $i \in [n]$. Then there exists $(s_1, s_2, \ldots, s_n) \in S_1 \times S_2 \times \cdots \times S_n$ such that $f(s_1, s_2, \ldots, s_n) \neq 0$.*

Proposition 2.2 (Variation of Theorem 6.1 in [1])**.** *Let p be a prime number in $[\frac{n}{4}, \frac{n}{2}]$. Then $\mathcal{M}(G)$ contains a subgraph U such that for every $u \in V(U)$,* $\deg(u) \in \{p, 2p, 3p, 4p, 5p, 6p, 7p\} \cap [2(n-1)]$.

Proof. We define a polynomial f with degree $|E(\mathcal{M})|$ over $\mathbb{F}_2$, and using the fact that $a^{p-1} \pmod p$ $\equiv 1$ for all $a \not\equiv 0 \pmod p$, we show the existence of the desired subgraph using the nullstellensatz directly.

Define the polynomial

$$f(x_e : e \in E(\mathcal{M})) = \prod_{v \in V(\mathcal{M})} \left[1 - \Big(\sum_{\substack{e \in E(\mathcal{M}) \\ v \in e}} x_e\Big)^{p-1}\right] - \prod_{e \in E(\mathcal{M})} (1 - x_e).$$

The degree of f is $|E(\mathcal{M})|$ because

$$|V(\mathcal{M})|(p-1) = (2n-1)(p-1) \leq (2n-1)(\tfrac{n}{2} - 1) = n(n - \tfrac{5}{2}) + 1 < n(n-1) = |E(\mathcal{M})|.$$

Note that the max degree term, $(-1)^{|E(\mathcal{M})|} \prod_{e \in E(\mathcal{M})} x_e$ has a non-zero coefficient. To apply the nullstellensatz, we consider solutions to f of the form $(s_1, s_2, \ldots, s_{|E(\mathcal{M})|}) \in \{0,1\}^{|E(\mathcal{M})|}$ (where $t_i = 1$ for all $i \in [|E(\mathcal{M})|]$). Thus by Lemma 5, there exists a vector, call it $r = (r_e : e \in E(\mathcal{M}))$, such that $f(r) \neq 0$. By the definition of f, $r \neq 0$ because $f(0) = 0$, so some of its entries are 1. This means that the latter product in f vanishes when evaluated at r. The former product in f can be non-zero only when $\Big(\sum_{\substack{e \in E(\mathcal{M}) \\ v \in e}} r_e\Big)^{p-1} \equiv 0 \pmod p$. It follows that r corresponds to a subgraph U of $\mathcal{M}(G)$ whose vertex degrees are congruent to $0 \pmod p$. Since $\Delta(\mathcal{M}) = 2(n-1)$ and $r \neq 0$, there exists a vertex $u \in U$ such that $\deg(u) \in \{p, 2p, 3p, 4p, 5p, 6p, 7p\} \cap [2(n-1)]$. Note that since the degrees of the vertices in the neighbourhood of u are all at least 1, U contains at least one vertex in each part of $\mathcal{M}$ with degree at least p. $\square$

3 Distance Multiplicities in Unweighted Graphs

Let G be a tree. Define $T(G)$ to be G without its leaves, and on each vertex v of $T(G)$ assign it a weight equal to the degree of v in G, $\deg(v)$. Define $m(k)$ to be the multiplicity of distance k in a graph G.

The following expresses $m(k)$ in terms of the degrees of the vertices of G, or equivalently, the vertex weights in $T(G)$.

Lemma 6 (Characterizing Distance Multiplicities in Terms of Vertex Degrees)**.** *It holds that* $m(1) = |E(G)|$, $m(2) = \sum_{v \in V(G)} \binom{\deg(v)}{2}$*, and when* $3 \leq k \leq \mathrm{Diam}(G)$,

$$m(k) = \sum_{\substack{\{x,y\} \subset T(G) \\ d(x,y) = k-2}} (\deg(x) - 1)(\deg(y) - 1).$$

Proof sketch. The cases $k = 1$ and $k = 2$ are straightforward and no distance can be larger than the diameter of G. Suppose $3 \leq k \leq \mathrm{Diam}(G)$. Let $x, y \in T(G)$ where $d(x,y) = k-2$ and let $P(x,y)$ be the unique path of length $k-2$ between x and y. There are $\deg(x) - 1$ and $\deg(y) - 1$ neighbours of x and y in G that are not in $P(x,y)$. Let w be such a neighbour of x and z such a neighbour of y. Then the unique (w,z)-path contains $P(x,y)$ and has length k. Thus $d(w,z) = k$ and there are

$(\deg(x)-1)(\deg(y)-1)$ such pairs. So, each pair $x, y \in T(G)$ satisfying $d(x,y)=k-2$ contributes a multiplicity for k in G of $(\deg(x)-1)(\deg(y)-1)$.

It is because G is a tree that this method counts all instances of distance k; if G has a cycle, then some distances can be over counted and this sum is an upper bound for $m(k)$. $\square$

3.1 Conjectures

Conjecture 3.1. *Let d be the largest distance that attains maximum multiplicity in a tree T. Then for every $i \in \{d, d+1, \dots, \mathrm{Diam}(T)-1\}$, $m(i) \geq m(i+1)$.*

Remark 3.2. I suspect it is possible to prove this by induction on the path lengths k. That is, every path of length $k+1$ corresponds to at least 1 distinct path of length k. But I think things get a bit tricky because somehow the maximality of d needs to come into play.

Remark 3.3. There are counter-examples to the related claim that $m(i) \geq m(i-1)$ for all $i \in \{d, d-1, \dots, 2\}$.

The remaining conjectures are all about upper bounding $m(d)$. The following proposition handles the lower bound.

Proposition 3.4. *Let d be the largest distance with max multiplicity in a tree T. If $1 \leq d \leq \lceil \frac{n}{3} \rceil$, then $m(d) \geq n-1$.*

Proof. It is sufficient to construct a tree T such that $m(d) = |E(T)| = n-1$. Let u be a root vertex. Append two paths X and Y of length $d-1$ to u. Then for the remaining $n-2(d-1)-1$ vertices, append them as a length $n-2(d-1)-1$ path to u. There are $3(d-1)$ distinct paths of length d with endpoints in $X \cup Y$. There are $n-2(d-1)-d$ paths of length d with endpoints in $V(T) \setminus (X \cup Y)$.

Altogether, there are $n-2d+2-d+3d-3 = n-1$ paths of length d in T. Note that since $d \leq \lceil n/3 \rceil$,

$$n-2(d-1)-d \geq n-3\lceil n/3 \rceil + 2 = \begin{cases} 0, & \text{if } n \equiv 1 \pmod 3 \\ 1, & \text{if } n \equiv 2 \pmod 3), \end{cases}$$

and in either case, $m(d) = n-1$. Observe that in fact $m(1) = m(2) = \cdots = m(d) = n-1$. $\square$

Example 1. Figure 1 shows a tree with maximum multiplicity $m(d) = n-1$ where $d=6$ is the largest distance with max multiplicity and $n=20$.

Remark 3.5. Below I conjecture that $d \leq \lceil n/3 \rceil + 2$. I have not yet looked for extremal trees that minimize $m(d)$ when $d \in \{\lceil n/3 \rceil + 1, \lceil n/3 \rceil + 2\}$.

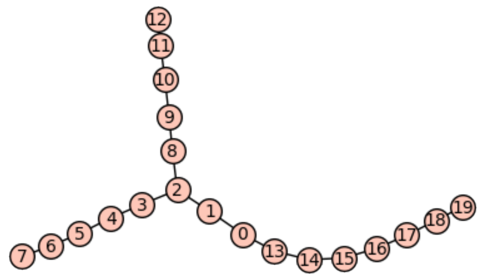


Figure 1: Extremal tree example minimizing $m(d)$.

Conjecture 3.6. *Let d be the largest distance with max multiplicity in a tree T.*

1. *If $d \leq C_1\frac{n}{3} + C_2$ and even, then $m(d) \leq (3-a-b)\lceil\frac{r}{3}\rceil\lfloor\frac{r}{3}\rfloor + \lfloor\frac{r}{3}\rfloor^{2a}\lceil\frac{r}{3}\rceil^{2b}$, where $r = n - \frac{3}{2}d + 2$ and*
$$(a,b) = \begin{cases} (1,0), & \text{if } r \equiv 1 \pmod 3 \\ (0,1), & \text{if } r \equiv 2 \pmod 3 \\ (0,0), & \text{otherwise.} \end{cases}$$

2. *If $C_1\frac{n}{3} + C_2 < d \leq \lceil\frac{n}{3}\rceil + 2$, then $m(d) \leq a\lfloor\frac{r'}{4}\rfloor^2 + (2-a)\lceil\frac{r'}{4}\rceil^2 + 2\lfloor\frac{r'}{4}\rfloor\lceil\frac{r'}{4}\rceil$, where $r' = n-d-1$ and*
$$a = \begin{cases} 2, & \text{if } r' \equiv 1 \pmod 4 \\ 1, & \text{if } r' \equiv 2 \pmod 4 \\ 0, & \text{otherwise.} \end{cases}$$

Remark 3.7. My experimentation suggests that $C_1 \sim 1$ and $-1 \leq C_2 \leq 1$; however, I have not yet examined these values carefully.

Remark 3.8. I believe there are 3 extremal trees that maximize $m(d)$; one unknown when $d \leq C_1\frac{n}{3} + C_2$ and odd, and the other two are described below.

Proposition 3.9 (Construction 1)**.** *Refer to Figure 2a for an example. When $d \leq C_1\frac{n}{3} + C_2$ and even, do the following:*

1. *First we use $3(\frac{d}{2}-1)+1$ vertices by making 3 branch paths with length $\frac{d}{2}-1$ from a root vertex u.*

2. *Let v, w, x be the vertices at the ends of each branch.*

3. *For the remaining $n - 3(\frac{d}{2}-1) - 1$ vertices, append them to v, w, x so that the number of leaf neighbours of v, w, and x differ from one another by at most 1.*

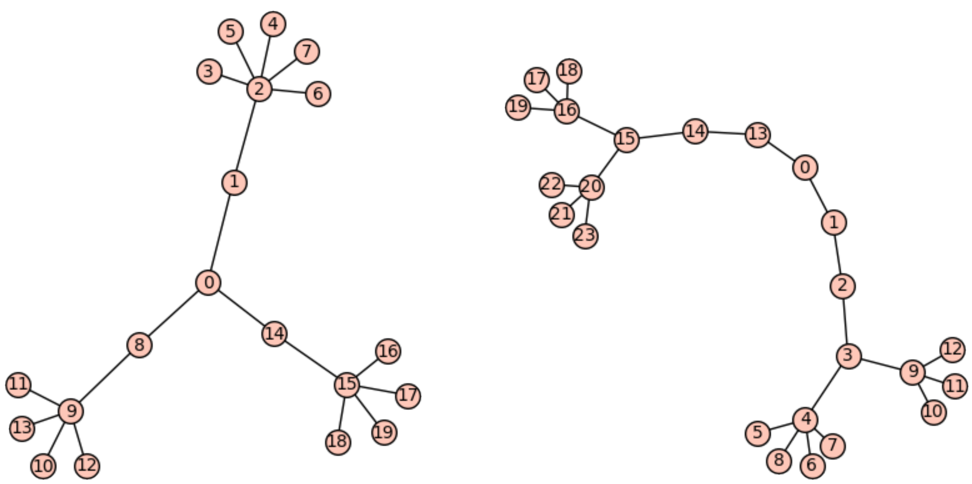
Remark 3.10. Trees with large $m(d)$ when $d \leq C_1\frac{n}{3} + C_2$ often tend to have a triple branching structure. The structure of T becomes much more constrained the larger d gets, and I think this is probably because it is most common for $d = 2$. When $d > 2$, then for $m(2) \leq m(d)$ to hold, **(1)** the degrees of the vertices of T cannot be too high, and **(2)** there needs to be enough branching in T to ensure enough distinct length d paths. Somehow the triple branching pattern in Construction

1 satisfies **(1)** and **(2)** while also maximizing $m(d)$; but I doubt that this extremal structure is fragile. That is, I think even when d is odd and $d \leq C_1 \frac{n}{3} + C_2$, an extremal tree has a similar triple branching structure.

Construction 2: Refer to Figure 2b for an example. When $d > C_1 \frac{n}{3} + C_2$, do the following:

1. Form a path of length $d-4$ and call its leaves x and y.
2. Append two vertices x_1 and x_2 to x and similarly y_1 and y_2 to y.
3. Append the remaining $n-d-1$ vertices to x_1, y_1, x_2, and y_2 so that the number of leaf neighbours on each differ from one another by at most 1. If $r' \equiv 2 \pmod 4$, then ensure that both x_1 and y_1 are each adjacent to $\lceil r'/4 \rceil$ leaves.

Example 2. Figure 2 shows examples from Constructions 1 and 2, which are mentioned above. In Figure 2a, $n = 20$, $d = 6$, and $m(6) = 56$. In Figure 2b, $n = 24$, $d = \lceil \frac{n}{3} \rceil + 2 = 10$, and $m(d) = 42$.



(a) Construction 1 (b) Construction 2.

Figure 2: Extremal tree examples that maximize $m(d)$.

Conjecture 3.11. *Let d be the largest distance with max multiplicity. Then $d \leq \lceil \frac{n}{3} \rceil + 2$.*

Remark 3.12. I have not yet found a counter-example to this conjecture. Please let me know if you find one! I have searched $n \leq 25$ without finding a CE, but it may well be that $d \leq \lceil n/3 \rceil + C\sqrt{n}$ or something. If so, then there would probably still be a sensible case division at $d \sim n/3$.

4 Construction with Large Gap Between Max Multiplicity and mult(2)

Let G be a tree of order n and max degree Δ. Let d be an even distance in $\{2, 4, \ldots, n-1\}$ with maximum multiplicity in G. Given Δ and d, the following construction attempts to maximize mult(d) while minimizing mult(2). The idea is to create a tree G where all vertices either have degree 1 or Δ and each path of length d has end-vertices with degree 1 in G. Note that these conditions determine a value of $n = n(d, \Delta)$. Recall from Proposition 3.4 and corresponding Figure 1 that we can minimize mult(d) at mult$(2) = n-1$. Similarly, in the construction from Proposition 3.9, we had mult$(2) = \mathcal{O}((\frac{n}{3} - \frac{3d}{2})^2)$ with mult$(d) = \mathcal{O}(n^2)$, which are asymptotically equivalent when d is not a function of n. The purpose of the following construction is to show that mult(d) need not be close to mult(2); in fact, we will see that it is possible for mult$(d) = \mathcal{O}(n^2)$ while mult$(2) = \mathcal{O}(n\Delta)$.

Construction. Let d and Δ be positive integers such that d is even. We construct a graph $G(d, \Delta)$ as follows:

1. Let r be a vertex with neighbours $v_1, v_2, \ldots, v_\Delta$.
2. For each $j \in [\Delta]$, identify v_j with the root of a $(\Delta - 1)$-ary tree with depth $\frac{d}{2} - 1$.

Example 3. Let $(d, \Delta) = (6, 4)$. Then the following tree of order 53 results from the above construction:

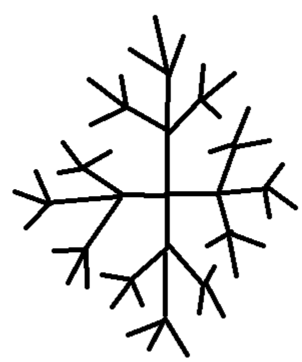


Figure 3: Example of a tree satisfying $(n, d, \Delta) = (53, 6, 4)$ where mult$(2) = 102$ and mult$(d) = 486$.

Let n_1 and n_Δ denote the numbers of vertices with degrees 1 and Δ, respectively. We can express both n_1 and n_Δ as functions of Δ and d, and since $n = n_1 + n_\Delta$, we can do the same for n. Note that the eccentricity of the root vertex is $\frac{d}{2}$. We can count the number of vertices by summing the number of vertices at distance $i \in \{0, 1, \ldots, \frac{d}{2}\}$ from the root. We have:

$$\begin{aligned} n &= 1 + \Delta + \Delta(\Delta - 1) + \cdots + \Delta(\Delta - 1)^{\frac{d}{2}-1} \\ &= 1 + \Delta \sum_{i=0}^{\frac{d}{2}-1} (\Delta - 1)^i. \end{aligned}$$

Note that from this we have $n_1 = \Delta(\Delta - 1)^{\frac{d}{2}-1} = \mathcal{O}(\Delta^{\frac{d}{2}})$, and since $n_1 > n_\Delta$, we have $n = n_1 + n_\Delta = \mathcal{O}(\Delta^{\frac{d}{2}})$ as well. The paths with length d always have end-vertices being leaves of G in the Δ distinct branches from the root of G. There are $(\Delta-1)^{\frac{d}{2}-1}$ leaves in each of these Δ branches, so $\mathrm{mult}(d) = \binom{\Delta}{2}(\Delta-1)^{d-2} = \mathcal{O}(\Delta^d)$. For $\mathrm{mult}(2)$, we sum the $\binom{\Delta}{2}$ distinct paths of length 2 at each of the $1+\Delta\sum_{i=0}^{\frac{d}{2}-2}(\Delta-1)^i$ non-leaf vertices. So, $\mathrm{mult}(2) = \binom{\Delta}{2}(1+\Delta\sum_{i=0}^{\frac{d}{2}-2}(\Delta-1)^i) = \mathcal{O}(\Delta^{\frac{d}{2}+1})$, and putting all these in terms of n, we have $\mathrm{mult}(d) = \mathcal{O}(n^2)$ and $\mathrm{mult}(2) = \mathcal{O}(n\Delta)$, which is quite a large distance. Thus it is possible to construct trees where the largest distance with maximum multiplicity has multiplicity much larger than the multiplicity of 2.

The following conjecture reflects the intuition that if distances 2 and $d = \mathrm{Diam}(G)$ have similar multiplicities, then there are few vertices with high degree and little branching.

Conjecture 4.1. *Let G be a tree with even diameter d where d is the largest distance with maximum multiplicity. If $\mathrm{mult}(2) \sim \mathrm{mult}(d)$, then there are at most 3 vertex disjoint paths between the centre and periphery of G.*

5 Distance Multisets of Specific Graphs

Observation 1. Paths of order n have distance multiset $\{1^{n-1}, 2^{n-2}, \ldots, (n-1)^1\}$.

Observation 2. Odd cycles of order n have distance multiset $\{1^n, 2^n, \ldots, \lfloor \frac{n}{2} \rfloor^n\}$. Even cycles of order n have distance multiset $\{1^n, 2^n, \ldots, (\frac{n}{2}-1)^n, (\frac{n}{2})^{\frac{n}{2}}\}$.

Proposition 5.1. *Let a and b be positive integers such that $a \geq b$. Then the multiplicity of each distance x in the a by b grid graph is given by*

$$\operatorname{mult}(x) = \begin{cases} (2bx - x^2)a + 2\binom{x+1}{3} - bx^2, & \text{if } x \in [1, b-1]; \\ b^2(a-x) + 2\binom{b+1}{3}, & \text{if } x \in [b, a-1]; \\ 2\binom{b+1-(x-a)}{3}, & \text{if } x \in [a, a-1+b-1]. \end{cases}$$

Proof. We first find a recurrence relation by calculating the distances between $G_{a,b-1}$ and a path of length $a-1$, which we call P_a. We call such distances "external distances", and the distances between vertices within $G_{a,b-1}$ and within P_a "internal distances". Throughout this proof, for a distance $x \in [1, a-1+b-1]$, we often distinguish between the internal and external multiplicities of x. It is straightforward to see that the distances of $G_{a,b}$ are the union of external and internal distances between $G_{a,b-1}$ and P_a. Once we know how to find the external distances, we will use this relation in a proof by induction to obtain a closed form expression for the distance multiset of $G_{a,b}$.

Let $v_1, v_2, \ldots, v_a$ be the vertices of P_a, and let $P(j)$ denote the path of length $b-1$ with v_j as an end-vertex in $G_{a,b}$. Note that there are a^2 ways to pair the vertices of P_a with the paths $P(1), P(2), \ldots, P(a)$.

For a distance $x \in [b-1]$, its external multiplicity is given by $a + \sum_{d=1}^{x-1} 2(a-d)$, where the $x-1$ in the upper limit ensures we are ignoring distances within P_a. There are $2(a-d)$ ordered pairs $(i,j) \in [a]^2$ such that $|i-j| = d$, and each of these corresponds to an external distance x. The a term comes from the a pairs (i,j) such that $|i-j| = 0$.

When $x \in [b, a-1]$, the external multiplicity is given by $\sum_{d=x-(b-1)}^{x-1} 2(a-d)$. In this case, we have to ignore smaller differences (*i.e.*, less than $x-(b-1)$) in P_a since $x \geq b-1$. When $x \in [a, a-1+b-1]$, the external multiplicity of x is $\sum_{d=x-(b-1)}^{a-1} 2(a-d)$. We can simplify these expressions as follows:

1. If $x \in [b-1]$:

$$\begin{aligned} a + \sum_{d=1}^{x-1} 2(a-d) &= a + 2a(x-1) - x(x-1) \\ &= a + (x-1)(2a-x) \end{aligned}$$

2. If $x \in [b, a-1]$:

$$\begin{aligned} \sum_{d=x-(b-1)}^{x-1} 2(a-d) &= 2a(b-1) - 2(b-1)(x-(b-1)) - (b-1)(b-2) \\ &= (b-1)(2a - 2(x-(b-1)) - (b-2)) \\ &= (b-1)(2a - 2x + 2b - 2 - b + 2) \\ &= (b-1)(2a - 2x + b) \end{aligned}$$

3. If $x \in [a, a-1+b-1]$:

$$\begin{aligned}\sum_{d=x-(b-1)}^{a-1} 2(a-d) &= 2a(a-x+b-1) - 2(a-x+b-1)(x-(b-1)) - (a-x+b-1)(a-x+b-2)\\ &= (a-x+b-1)(2a-2(x-b+1)-a+x-b+2)\\ &= (a+b-1-x)(2a-2x+2b-2-a+x-b+2)\\ &= (a+b-x-1)(a+b-x)\end{aligned}$$

Altogether, the external distances between $G_{a,b-1}$ and P_a are:

$$\mathrm{mult}(x)_{Ext} = \begin{cases} a+(x-1)(2a-x), & \text{if } x \in [1, b-1];\\ (b-1)(2a-2x+b), & \text{if } x \in [b, a-1];\\ 2\binom{a+b-x}{2}, & \text{if } x \in [a, a-1+b-1]. \end{cases}$$

Now we use this to find the distances in $G_{a,b}$. Below are the distance multisets of $G_{a,2}$, $G_{a,3}$, $G_{a,4}$, and $G_{a,5}$.

$G_{a,2} \Rightarrow \{1^{3a-2}\} \cup \{x^{4(a-x)+2} : x \in [2, a-1]\} \cup \{a^2\}$.
$G_{a,3} \Rightarrow \{1^{5a-3}, 2^{8a-10}\} \cup \{x^{9(a-x)+8} : x \in [3, a-1]\} \cup \{a^8, (a+1)^2\}$.
$G_{a,4} \Rightarrow \{1^{7a-4}, 2^{12a-14}, 3^{15a-28}\} \cup \{x^{16(a-x)+20} : x \in \{4, a-1\}\} \cup \{a^{20}, (a+1)^8, (a+2)^2\}$.
$G_{a,5} \Rightarrow \{1^{9a-5}, 2^{16a-18}, 3^{21a-37}, 4^{24a-60}\} \cup \{x^{25(a-x)+40} : x \in [5, a-1]\} \cup \{a^{40}, (a+1)^{20}, (a+2)^8, (a+3)^2\}$.

We claim that the distance multiset of $G_{a,b}$ is given by

$$\mathrm{mult}(x) = \begin{cases} (2bx-x^2)a + 2\binom{x+1}{3} - bx^2, & \text{if } x \in [1, b-1];\\ b^2(a-x) + 2\binom{b+1}{3}, & \text{if } x \in [b, a-1];\\ 2\binom{b+1-(x-a)}{3}, & \text{if } x \in [a, a-1+b-1]. \end{cases}$$

This is our inductive hypothesis. Note that the base cases are satisfied by our claimed solution.

We now begin the inductive step of the proof. By the inductive hypothesis, the cumulative internal distance multiplicities of $G_{a,b-1}$ and P_a for distance x are

$$\mathrm{mult}(x)_{Int} = \begin{cases} (2(b-1)x-x^2)a + 2\binom{x+1}{3} - (b-1)x^2 + (a-x), & \text{if } x \in [1, b-2];\\ (b-1)^2(a-x) + 2\binom{b}{3} + (a-x), & \text{if } x \in [b-1, a-1];\\ 2\binom{b-(x-a)}{3}, & \text{if } x \in [a, a-1+b-2]. \end{cases}$$

Note that we simply add $(a-x)$ (from P_a) when $x \in [1, a-1]$. We showed earlier that the external distances between $G_{a,b-1}$ and P_a are

$$\mathrm{mult}(x)_{Ext} = \begin{cases} a+(x-1)(2a-x), & \text{if } x \in [1, b-1];\\ (b-1)(2a-2x+b), & \text{if } x \in [b, a-1];\\ 2\binom{a+b-x}{2}, & \text{if } x \in [a, a-1+b-1]. \end{cases}$$

Now all we do is add these multiplicities together. When $x \in [a, a-1+b-1]$, we have

$$\begin{aligned}
\operatorname{mult}(x) &= 2\binom{b-(x-a)}{3} + 2\binom{a+b-x}{2} \\
&= \frac{(b-x+a)(b-x+a-1)(b-x+a-2)}{3} + (b-x+a)(b-x+a-1) \\
&= (b-x+a)(b-x+a-1)(\frac{b-x+a-2}{3}+1) \\
&= (b-x+a)(b-x+a-1)(\frac{b-x+a+1}{3}) \\
&= 2\binom{b+1-(x-a)}{3},
\end{aligned}$$

as desired.

When $x \in [b, a-1]$, we have

$$\begin{aligned}
\operatorname{mult}(x) &= (b-1)^2(a-x) + 2\binom{b}{3} + (a-x) + (b-1)(2a-2x+b) \\
&= (b-1)(a-x)(b-1+2) + b(b-1) + (a-x) + 2\binom{b}{3} \\
&= (b-1)(a-x)(b+1) + b(b-1) + (a-x) + \frac{b(b-1)(b-2)}{3} \\
&= (a-x)(1+(b-1)(b+1))b(b-1) + \frac{b(b-1)(b-2)}{3} \\
&= b^2(a-x) + b(b-1)(1+\frac{b-2}{3}) \\
&= b^2(a-x) + \frac{(b+1)b(b-1)}{3} \\
&= b^2(a-x) + 2\binom{b+1}{3},
\end{aligned}$$

as desired.

When $x = b - 1$, we have

$$\begin{aligned}
\mathrm{mult}(x) &= (b-1)^2(a-x) + 2\binom{b}{3} + (a-x) + a + (x-1)(2a-x) \\
&= (b-1)^2(a-x) + 2\binom{b}{3} + a - x + a + 2ax - x^2 - 2a + x \\
&= (b-1)^2(a-x) + 2\binom{b}{3} + 2ax - x^2 \\
&= x^2(a-x) + 2\binom{x+1}{3} + 2ax - x^2 \\
&= a(x^2+2x) + 2\binom{x+1}{3} - x^2(x+1) \\
&= (x^2+2x)a + 2\binom{x+1}{3} - bx^2 \\
&= (2(x+1)x - x^2) + 2\binom{x+1}{3} - bx^2 \\
&= (2bx - x^2)a + 2\binom{x+1}{3} - bx^2,
\end{aligned}$$

as desired.

When $x \in [1, b-2]$, we have

$$\begin{aligned}
\mathrm{mult}(x) &= (2(b-1)x - x^2)a + 2\binom{x+1}{3} - (b-1)x^2 + a + (x-1)(2a-x) + (a-x) \\
&= a(2(b-1)x - x^2) + 2\binom{x+1}{3} - bx^2 + x^2 + a + 2ax - x^2 - 2a + x + a - x \\
&= a(2(b-1)x - x^2) + 2\binom{x+1}{3} - bx^2 + 2ax \\
&= a(2(b-1)x - x^2 + 2x) + 2\binom{x+1}{3} - bx^2 \\
&= a(2bx - x^2) + 2\binom{x+1}{3} - bx^2,
\end{aligned}$$

as desired. Thus the proposition holds. □

Proposition 5.2. *Let T be a tree of order n that is neither K_2 nor $K_{1,3}$. Then T does not have uniform distance multiplicities.*

Proof. Let m be the multiplicity of each graph distance in T and d its diameter. Then $\binom{n}{2} = md$. Since T is a tree, $\mathrm{mult}(1) = n - 1$, so $m = n - 1$ and thus $d = \frac{n}{2}$. Let $T_0 = (v_1, v_2, \ldots, v_d)$ be a diametrical path. Order the vertices of $V(T) \setminus T_0 = (u_1, u_2, \ldots, u_{n-d})$ in a way so that for all $i \in \{1, 2, \ldots, n-d\}$, each graph T_i induced by $T_0 \cup \{u_1, u_2, \ldots, u_i\}$ is connected. Since T_0 is a diametrical path, u_1 in T_1 has to be adjacent to a vertex in T_0 with degree 2; thus both distances 1 and 2 in T_1 have multiplicity $\frac{n}{2}$. Note that u_i has a unique neighbour in T_i. Since T_{i-1} is connected, the neighbour of u_i in T_i has degree at least 1 in T_{i-1}, so the multiplicity of distance 2 in T_i increases by at least 1 compared to T_{i-1}. Observe that if for some $i \in \{2, 3, \ldots, n-d\}$, the neighbour of u_i in T_{i-1} has degree at least 2, then $\text{mult}(2) > \text{mult}(1)$ in T. So the only way for $\text{mult}(2) = \text{mult}(1)$ in T is if every added vertex gets joined with a leaf such that the resulting tree has diameter d. But in this case T has exactly 3 leaves, which means T has at most 3 diametrical paths and so $n - 1 = m \leq 3$. The only solutions are K_2 and $K_{1,3}$. □

Proposition 5.3. *The distance multiset of any tree of order n with diameter $d \leq 3$ can be expressed in terms of its degree sequence.*

Proof. Let T be a tree with diameter $d \leq 4$. If $d = 2$, then since T is a tree, $\text{mult}(1) = n - 1$ and so $\text{mult}(2) = \binom{n}{2} - (n-1) = (n-1)(\frac{n}{2} - 1)$. Note that T has exactly one vertex v with degree at least 2, so $\text{mult}(2) = \sum_{u \in V(T)} \deg(u) = \deg(v)$. If $d = 3$, then $\text{mult}(2) = \sum_{u \in V(T)} \deg(u)$ and $\text{mult}(3) = \binom{n}{2} - \text{mult}(2) - \text{mult}(1) = \binom{n}{2} - \sum_{u \in V(T)} \deg(u) - (n-1)$. □

Observation 3. Note that the edges of T must be known to determine the distance multiset when $d \geq 4$, because

$$\text{mult}(3) = \sum_{u_1 u_2 \in E(T)} (\deg(u_1) - 1)(\deg(u_2) - 1).$$

So for the case when $d = 4$, we have $\text{mult}(3) = \sum_{u_1 u_2 \in E(T)} (\deg(u_1) - 1)(\deg(u_2) - 1)$, and so

$$\text{mult}(4) = \binom{n}{2} - \sum_{u_1 u_2 \in E(T)} (\deg(u_1) - 1)(\deg(u_2) - 1) - \sum_{u \in V(T)} \deg(u) - (n-1).$$

Proposition 5.4. *Let T be a tree with m distinct multiplicities $m_1, m_2, \ldots, m_m$ such that for $i \in [m]$, there are x_i distances with multiplicity m_i. Then the diameter of T is $d = \ldots.$*

Proof. Note that $\binom{n}{2} = \sum_{i=1}^{m} m_i x_i$. Suppose wlog that $m_1 < m_2 < \cdots < m_m$. Note that $d = \sum_{i=1}^{m} x_i \Leftrightarrow x_m = d - \sum_{i=1}^{m-1} x_i$. Let $x = \sum_{i=1}^{m-1} m_i x_i$ and set $M = m_m$. Then $\binom{n}{2} = x + M(d - \sum_{i=1}^{m-1} x_i)$, so

$$\binom{n}{2} = x + Md - M \sum_{i=1}^{m-1} x_i$$

$$\Leftrightarrow d = \frac{1}{M} \left[\binom{n}{2} - \sum_{i=1}^{m-1} m_i x_i \right] + \sum_{i=1}^{m-1} x_i$$

□

6 Graph Distance Multiplicity Vector Space

In this section, we propose the study of a subset of vectors on a bounded surface in $\mathbb{R}^d$, in which the vectors are the multiplicity lists for graphs with diameter d. The "corners" of this surface are multiplicity vectors for special standard graphs whose multiplicities are well understood, like cycles, paths, grid graphs, and so on. One goal is to characterize general graph classes relative to the special graph "corners" in the following sense: if the multiplicity vector of a graph G is a linear combination of a pair of standard graph multiplicity vectors, then we say that G is similar to these two standard graphs. Subsequently, if G is similar to a pair of standard graphs H_1 and H_2, then using the inner products of the multiplicity vectors, we can determine whether G is closer (wrt to angle and/or norm) to H_1 or H_2 in $\mathbb{R}^d$. So, the goal is to try to classify graphs based on their distance multiplicities relative to standard graphs. There are two phases to this endeavor: (1) find and argue for good candidate standard graphs, and (2) try to build up a large and comprehensive surface in $\mathbb{R}^d$ in which a wide variety of graphs live.

Let P be a path of length d, then we say that $m(P) = (d, d-1, \ldots, 1)$ is the multiplicity vector of P. Let G be a graph of order n with multiplicity vector $m(G) = (m_1, m_2, \ldots, m_d)$, where m_i is the multiplicity of distance i in G. Note that $\sum_{i=1}^{d} m_i = \binom{n}{2}$ and $\sum_{i=1}^{d} im_i = W(G)$, where $W(G)$ is the sum of all distances in G (also known as the Weiner index of G). Then

$$\begin{aligned}
\langle m(P), m(G)\rangle &= \sum_{i=1}^{d}(d+1-i)\cdot m_i \\
&= (d+1)\sum_{i=1}^{d} m_i - \sum_{i=1}^{d} im_i \\
&= (d+1)\binom{n}{2} - W(G).
\end{aligned}$$

By the definition of the inner product in $\mathbb{R}^d$,

$$\langle m(P), m(G)\rangle = \|m(P)\|\,\|m(G)\|\cos(\theta),$$

where θ is the angle between $m(P)$ and $m(G)$ in $\mathbb{R}^d$. Note that

$$\|m(P)\| = \sqrt{\sum_{i=1}^{d}(d+1-i)^2} = \sqrt{\sum_{i=1}^{d} i^2} = \sqrt{\frac{d(d+1)(2d+1)}{6}}.$$

So altogether we have

$$(d+1)\binom{n}{2} - W(G) = \sqrt{\frac{d(d+1)(2d+1)}{6}}\cdot\|m(G)\|\cdot\cos(\theta).$$

Why is this interesting? Note n, $W(G)$, and $\|m(G)\|$ are parameters that depend only on G, so they determine the angle θ, which can be seen as a measure of how similar G and P are. For example, let $G = K_{1,n-1}$; since G has diameter 2, we assume $d = 2$. Then it is straightforward to see that $\|m(K_{1,n-1})\| = \sqrt{\binom{n-1}{2}^2 + (n-1)^2}$ and $W(K_{1,n-1}) = 2\binom{n-1}{2} + (n-1)$. Thus we have

$$
\begin{aligned}
3\binom{n}{2}-\left(2\binom{n-1}{2}+(n-1)\right) &= \sqrt{5}\cdot\sqrt{\binom{n-1}{2}^2+(n-1)^2}\cdot\cos(\theta)\\
\Leftrightarrow 2\binom{n}{2}-\binom{n-1}{2} &= \sqrt{5\left(\binom{n-1}{2}^2+(n-1)^2\right)}\cdot\cos(\theta)\\
\Leftrightarrow \cos(\theta) &= \frac{\binom{n+1}{2}}{\sqrt{5(\binom{n-1}{2}^2+(n-1)^2)}}
\end{aligned}
$$

Note that as $n\to\infty$, $\cos(\theta)$ tends to $\frac{1}{\sqrt{5}}$. This means that $\lim_{n\to\infty}\theta(n)\sim 1.1071$, which is about 63.43 degrees. While we might expect the angle to be closer to $\frac{\pi}{2}$ (since paths and stars have very different, almost oppositional, multiplicity structure in trees), it is still the case that $m(P)$ and $m(K_{1,n-1})$ are very dissimilar in both angle and (especially) norm.

Note that while all trees with diameter 2 are stars, there are many diameter two graphs containing cycles that do not have a spanning tree with the same distance multiset. **A follow-up question from this example:** does there exist a graph G with a cycle whose multiplicity vector has larger angle from P than $K_{1,n-1}$ and from $K_{1,n-1}$ than P? If not, then perhaps $m(K_{1,2})$ and $m(K_{1,n-1})$ form boundaries for multiplicity vectors of graphs with diameter 2.

On the Weiner Index and Arithmetic Progressions Recall the Weiner index is $W(G)=\sum_{i=1}^{d} im_i$. So if we are comparing G to some special standard graph H and $W(G)$ shows up in $\langle m(G),m(H)\rangle$, then this means that at least part of $m(H)$ has a distribution that scales linearly with the distance. That is if $m(H)=(\ell_1,\ell_2,\ldots,\ell_d)$, where ℓ_i is the multiplicity of distance i in H, then for $W(G)$ to show up in $\langle m(G),m(H)\rangle$, it must hold that $\ell_i=f(\ell_i)+ig+a$ for some $g,a\in\mathbb{Z}$ and some function $f:\mathbb{Z}\mapsto\mathbb{Z}$. This is because if $W(G)$ shows up in this way, we have

$$
\begin{aligned}
\langle m(G),m(H)\rangle &= \sum_{i=1}^{d} m_i\cdot\ell_i\\
&= \sum_{i=1}^{d} m_i\cdot f(\ell_i)+gW(G)+a\sum_{i=1}^{d} m_i\\
&= \sum_{i=1}^{d} m_i\cdot f(\ell_i)+g\sum_{i=1}^{d} im_i+a\sum_{i=1}^{d} m_i\\
&= \sum_{i=1}^{d} m_i(f(\ell_i)+gi+a).
\end{aligned}
$$

Note that if f vanishes and $g\neq 0$, then $m(H)$ is an arithmetic progression. Also if $a\neq 0$, then $\sum_{i=1}^{d} m_i=\binom{n(G)}{2}$ shows up in the inner product, which is also convenient.

Question: are all graphs with arithmetic progression multiplicities are similar to one another? That is, θ in their inner products are similar. Let's explore this...

Let $m(G)=(15,11,7,3)$, so $n=9$. Set aside the fact that no graph exists with these multiplicities, since we may generalize this approach to study distance multiplicities for any class of metrical objects. Let's compare $m(G)$ to $m(P) = (4, 3, 2, 1)$, the multiplicities of a path with diameter 4. Then $\langle m(P), m(G)\rangle = \sum_{i=1}^{4}(5-i)m_i = 60+33+14+3$, but note that $m_i = 4(d-i)+3 = 17-4i$. So....

References

[1] Alon, N. (1999). Combinatorial Nullstellensatz. Combinatorics, Probability and Computing, 8(1-2), 7-29. doi:10.1017/S0963548398003411