9.3 Technique for Constructing Caterpillar with Half its Multiplicities Specified by the Square of a Known Generating Function

In this section, we show a technique that allows one to specify a sequence of multiplicities and construct a graph such that the larger half of its distances attain (approximately) the specified multiplicities. The problem amounts to whether or not the sequence associated with the square root of the generating function for the specified sequence is known.

Let $P = (v_0, \ldots, v_{d-2})$ be a path such that for each $i \in [0, d-2]$, v_i has a_i leaves joined to it. Let $b = (b_0, b_1, \ldots, b_{\lfloor \frac{d-2}{2} \rfloor})$ be a sequence of reals greater than or equal to 1. Define D to be an integer multiset where for each $r \in [d - \lfloor \frac{d-2}{2} \rfloor, d]$, $\text{mult}_D(r) = b_{r-(d-\lfloor \frac{d-2}{2} \rfloor)}$. Then we show below that there exists a caterpillar graph T whose distance multiset $D(T)$ satisfies the property that for each $r \in [d - \lfloor \frac{d-2}{2} \rfloor, d]$, $\text{mult}_{D(T)}(r) \approx \text{mult}_D(r)$. Note the same cannot be said for the multiplicities of the smaller half of distances.

The idea is to first assume that $a_i = a_{d-2-i}$, which ensures that $\text{mult}(r) \approx \sum_{i=0}^{d-r} a_i a_{d-r-i}$ (this will be explained shortly) and also that there are $\lfloor \frac{d-2}{2} \rfloor + 1$ unknowns

$$a_0 = a_{d-2}, a_1 = a_{d-3}, \ldots, a_{\lfloor \frac{d-2}{2} \rfloor} = a_{d-2-\lfloor \frac{d-2}{2} \rfloor}.$$

Then we solve for these unknowns using the following system of $\lfloor \frac{d-2}{2} \rfloor + 1$ equations: for each $r \in [d - \lfloor \frac{d-2}{2} \rfloor, d]$, we have

$$\sum_{i=0}^{d-r} a_i a_{d-r-i} = b_{r-(d-\lfloor \frac{d-2}{2} \rfloor)}. \tag{1}$$

Note that $\text{mult}(r)$ also includes distances between leaves and vertices in P, and there are $\sum_{i=0}^{d-1-r} 2a_i$ of such distances for each $r \in [d - \lfloor \frac{d-2}{2} \rfloor, d-1]$, but these terms are all linear in a_i, and so contribute much less than the quadratic terms in the convolution sum above; so for this reason and simplicity, we do not emphasize these distances going forward. Also observe that there are $d-1-r$ occurrences of distance $r \in [3, d-2]$ within P. So in general,

$$\text{mult}(r) = \sum_{i=0}^{d-r} a_i a_{d-r-i} + \sum_{i=0}^{d-1-r} 2a_i + \sum_{i=0}^{d-1-r} 1.$$

Now back to developing the technique for working with the quadratic terms in the convolution. Since

$$F(x)^2 := \Big(\sum_{d-r \geq 0} a_{d-r} x^{d-r} \Big)^2 = \sum_{d-r \geq 0} \Big(\sum_{i=0}^{d-r} a_i a_{d-r-i} \Big) x^{d-r},$$

if b is given by a generating function $G(x)$, then $F(x) = \sqrt{G(x)}$. Thus if $\sqrt{G(x)}$ is known, then we know $F(x)$ and can construct a caterpillar whose larger half of distances closely attain the multiplicities specified by b. For example, if for some constant C, $b = (C, C, \ldots, C)$, then

$$\sum_{d-r\geq 0} \Big(\sum_{i=0}^{d-r} a_i a_{d-r-i} \Big) x^{d-r} = \sum_{d-r\geq 0} C x^{d-r}$$

$$\Leftrightarrow \sum_{d-r\geq 0} \Big(\sum_{i=0}^{d-r} \big(\frac{a_i}{\sqrt{C}}\big)\big(\frac{a_{d-r-i}}{\sqrt{C}}\big) \Big) x^{d-r} = \sum_{d-r\geq 0} x^{d-r}$$

$$\Leftrightarrow \sum_{i=0}^{d-r} \big(\frac{a_i}{\sqrt{C}}\big)\big(\frac{a_{d-r-i}}{\sqrt{C}}\big) = 1,$$

which implies that $\frac{a_{d-r}}{\sqrt{C}} = \binom{2(d-r)}{d-r} 4^{-(d-r)}$, since this is the unique sequence of reals (with $a_0/\sqrt{C} = 1$) that has convolutions always equalling 1. Thus setting a_{d-r} and a_{r-2} to $\lceil \binom{2(d-r)}{d-r} 4^{-(d-r)} \sqrt{C} \rceil$ for each $r \in [d - \lfloor \frac{d-2}{2} \rfloor, d]$ ensures that $\text{mult}_T(r) \approx \text{mult}_D(r) = b_{r-(d-\lfloor \frac{d-2}{2} \rfloor)} = C$. In fact, if we allow for real number multiplicities, where the distances from one of the appended leaves at each internal path vertex are possibly included only partially, then we can remove the ceiling function and get that $\text{mult}_T(r) = \text{mult}_D(r) + \sum_{i=0}^{d-1-r} 2a_i + \sum_{i=0}^{d-1-r} 1$.

Here is an example of such a caterpillar:

Example 4. Let $d = 9$ and $C = 78 = \binom{13}{2}$. Then

$$(a_0, a_1, \ldots, a_7) \approx (8.832, 4.416, 3.312, 2.760, 2.760, 3.312, 4.416, 8.832),$$

and if we denote the quadratic terms in $\text{mult}(r)$ as $\text{mult}_Q(r)$, then we have

$$\begin{aligned} \text{mult}_Q(6) = \text{mult}_Q(7) = \text{mult}_Q(8) = \text{mult}_Q(9) &= 78.004224 \\ \text{mult}_Q(5) &= 84.098304 \\ \text{mult}_Q(4) &= 101.161728 \\ \text{mult}_Q(3) &= 133.155648 \\ \text{mult}(2) &= 96.772224 \\ \text{mult}(1) &= 46.64. \end{aligned}$$

Remark 9.2. When pursuing this caterpillar idea, I was originally motivated by the following problem: Find a smallest order tree with diameter d whose distance multiset contains ***all*** distance multisets of trees of order k and diameter d (There's an analogous problem for finding a largest order tree whose distance multiset *is contained by* all trees of order k and diameter d). Looking at the previous example, since no tree of order 14 can have distance multiplicities larger than $\binom{13}{2}$ (extremal example is $K_{1,13}$ for $\text{mult}(2)$), this caterpillar is actually a (definitely not smallest order) solution to the extremal problem. By the way, when b is a constant sequence, it is relatively easy to show that $\text{mult}_Q(r) \leq \text{mult}_Q(r-1)$ for all $r \in [4, d]$, so it isn't a fluke that the smaller distance multiplicities in the example are also large.

We can represent the fractional multiplicities in a graph using the following terminology. A vertex v labelled with a number x indicates that the distances involving the vertex v should be considered with partial multiplicity; we call x the *multiplicity scalar* of v. If x and y are multiplicity scalars of v and u, then $d(v, u)$ is counted with multiplicity xy. If a multiplicity scalar is unspecified, then assume it is 1.

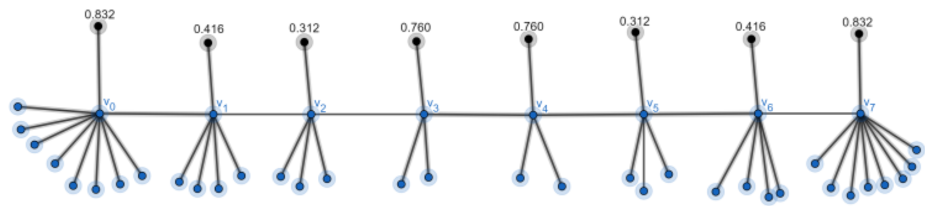

Figure 4: Caterpillar with larger multiplicities close to 78, where "close" means the quadratic terms sum to ≈ 78.

This technique can be applied so long as the choice of b enables the system of equations in (1) to be solved. That is, we need to be able to find a non-negative sequence of reals $a_0, a_1, \ldots, a_n$ such that

$$\sum_{n\geq 0} a_n x^n = \sqrt{\sum_{n\geq 0} b_n x^n}.$$

The trade off here is that the order of T could be very large, and it is not necessarily the case that it is minimal with respect to the property of its large half of distances having the specified multiplicities. Put another way, T could have a lot of extraneous small distances.

For another example, geometric progressions are also suitable choices for b. If $b_i = \alpha^i$, then a simple calculation shows that we should set $a_{d-r} = \binom{2(d-r)}{d-r} 4^{-(d-r)} \alpha^{d-r}$. Here's the calculation:

$$\begin{aligned}
\sum_{d-r\geq 0} \left(\sum_{i=0}^{d-r} a_i a_{d-r-i} \right) x^{d-r} &= \sum_{d-r\geq 0} \alpha^{d-r} x^{d-r} \\
\left(\sum_{d-r\geq 0} a_{d-r} x^{d-r} \right)^2 &= \frac{1}{1-\alpha x} \\
\sum_{d-r\geq 0} a_{d-r} x^{d-r} &= (1-\alpha x)^{-1/2} \\
&= \sum_{d-r\geq 0} \binom{-1/2}{d-r} (-1)^{d-r} \alpha^{d-r} x^{d-r} \\
&= \sum_{d-r\geq 0} \binom{2(d-r)}{d-r} 4^{-(d-r)} \alpha^{d-r} x^{d-r}.
\end{aligned}$$

I'm hoping this is an early version of a class of techniques that allows us to construct graphs such that ***all*** (or nearly all) of its distance multiplicities closely resemble those specified. The two main drawbacks to the above approach are (1) only half the multiplicities can be specified and (2) this can cause the order of the constructed graph to be quite large (though this varies depending on the choice of b), which means it might not be a minimal order construction (which would probably be a more cool thing to find).

What makes this technique work is that convolutions relate to squares of generating functions, and convolutions account for the majority of multiplicities in caterpillars with number of leaves symmetric about the center of the graph. However, many, if not most, of the distances are not accounted for by this caterpillar construction (though it's neat that nearly half of the distance multiplicities essentially are!). Ideally, there would be a way to get a system of maybe just less than d equations so that the same number of unknowns $a_0, a_1, a_2, \ldots, a_{d-2}$ (describing number of leaves (or some other "gadgets") to be appended to specified vertices) could be solved for, given a list of $\sim d$ specified multiplicities.

I tried playing with cycles for a bit to see if convolutions could be applied there to count multiplicities, but had no luck. I think the convolution = square observation can only be leveraged in cases where only half the distance multiplicities are specified. However, it might be possible to find a way to get a system of $\sim d$ equations using a caterpillar (*i.e.* where we don't assume $a_i = a_{d-2-i}$). In this case, what we would need to do is figure out how to solve the following: for each $r \in [3, d]$, we have the following system of equations:

$$\mathrm{mult}_Q(r) = \sum_{i=0}^{d-r} a_i a_{i+r-2} = b_{r-3}, \tag{2}$$

but I don't currently have any good ideas about how to work with this sum to find $a_0, a_1, \ldots, a_{d-2}$ (also, it looks like there's an extra unknown here). Anyway, I think this section has sufficiently established that it is worth looking into this type of problem!

9.4 Constructing Trees That Maximize Multiplicity for a Given Distance

Our goal here is to set up and state Conjecture 9.4 (which I think is mostly proven below) on a tight upper bound for the multiplicities of even distances in trees of order k and diameter d. That is, let $\mathcal{T}(k,d)$ denote the set of all trees of order k and diameter d, then we want to find $\max_{T\in\mathcal{T}(k,d)}(\mathrm{mult}_T(i))$, where $\mathrm{mult}_T(i)$ is the multiplicity of distance i in the tree T. We begin by finding the tight example for distance d (when d is even), and then we argue that this example can be slightly altered to maximize the multiplicity of any even distance. What follows is a technical lemma that is necessary to prove the optimality of the construction in the diameter multiplicity case.

Lemma 8. *Let $\{a_1, a_2, \ldots, a_n\}$ be a set of positive reals such that $\sum_{i=1}^n a_i = k - rn \geq n$, where $k, r \in \mathbb{Z}^+$ and $k \geq r/2$. Then the sum $\sum_{i<j\leq n} a_i a_j$ is maximized at $f(n,k,r) = \binom{n}{2}\frac{(k-rn)^2}{n^2}$ (with $a_i = \frac{k-rn}{n}$), when $n = \lceil n_0(k,r)\rceil$, where n_0 is the solution to the following cubic such that $n_0 \in [n_{\max}-1, n_{\max}]$, where $\frac{\partial}{\partial n} f(n_{\max}, k, r) = 0$ and $2 \leq n_{\max} < k - rn$:*

$$n^3(2r^2) + n^2(2r^2 - 2kr) + n(-2kr) + k^2 = 0.$$

Proof Sketch. Observe that $\sum_{1=i<j\leq n} a_i a_j$ is maximized when $a_1 = a_2 = \cdots = a_n = \frac{k-rn}{n}$. Thus

$$\sum_{i<j\leq n} a_i a_j \leq \binom{n}{2}\frac{(k-rn)^2}{n^2} = \frac{(n-1)(k-rn)^2}{2n},$$

which is the function $f(n,k,r)$ in the lemma statement. Now we find integers n that maximize $f(n,k,r)$. Note that $f(n,k,r)$ is maximized at $n=\frac{1}{4}\left(\sqrt{\frac{8k}{r}+1}+1\right)=n_{\max}$ when $k-rn\geq n$. To find nearest integer values that maximize $f(n,k,r)$, we need to consider $f(n,k,r)$ within the interval I within which f is concave and $n_{\max}\in I$. This interval I is $[1,b]$ where b is the unique real number satisfying $\frac{\partial^2}{\partial n^2}f(b,k,r)=0$, whereby the concavity of f fails when $n>b$. Since $\frac{\partial^2}{\partial n^2}f(n,k,r)=r^2-\frac{k^2}{n^3}$, $b=\frac{k^{2/3}}{r^{2/3}}$ and so $I=[1,(\frac{k}{r})^{2/3}]$. Thus within this interval I, $f(n,k,r)$ is concave. To find the integer(s) that maximize $f(n,k,r)$, we find n_0 for which $f(n_0,k,r)=f(n_0+1,k,r)$ (such an n_0 exists and is unique by continuity and concavity of f within I, respectively), and then the maximal integer solution(s) will be $\{\lceil n_0\rceil,\lfloor n_0+1\rfloor\}$. Note that $f(n,k,r)$ has only two local extrema in $\mathbb{R}^+$, one local maximum at $f(n_{\max},k,r)$ in I and one local minimum at $f(k/r)$ in $\mathbb{R}\setminus I$, so we solve the following equation for n and choose the root within I:

$$\begin{aligned}\frac{(n-1)(k-rn)^2}{2n}&=\frac{(n)(k-r(n+1))^2}{2(n+1)}\\ \Leftrightarrow 2(n+1)(n-1)(k-rn)^2&=2n^2(k-r(n+1))^2\\ \Leftrightarrow n^3(2r^2)+n^2(2r^2-2kr)+n(-2kr)+k^2&=0.\end{aligned}$$

We want the solution within the interval $[x_{\max}-1,x_{\max}]$; denote this number by $n_0(k,r)$. This $n_0(k,r)$ can be easily found for a given k and r using a computer solver like Sage. Thus $f(\lceil n_0(k,r)\rceil,k,r)$ is the maximum of $f(n,k,r)$ over I restricted to the integers.

In particular, we have the following expression for $n_0(k,r)$ [wow, simplification needed! The root is real, so it should be possible to simplify out the $\sqrt{-1}$s somehow]:

$$\begin{aligned}n_0(k,r)=&-\frac{1}{24}\cdot 4^{\frac{2}{3}}\left(i\sqrt{3}+1\right)\left(\frac{4\,(k-r)^3}{r^3}+\frac{18\,(k-r)k}{r^2}-\frac{27\,k^2}{r^2}+\frac{3\sqrt{3}k\sqrt{-\frac{8\,k^3-11\,k^2r-4\,kr^2-4\,r^3}{r}}}{r^2}\right)^{\frac{1}{3}}\\ &+\frac{4^{\frac{1}{3}}\left(i\sqrt{3}-1\right)\left(\frac{(k-r)^2}{r^2}+\frac{3\,k}{r}\right)}{6\left(\frac{4\,(k-r)^3}{r^3}+\frac{18\,(k-r)k}{r^2}-\frac{27\,k^2}{r^2}+\frac{3\sqrt{3}k\sqrt{-\frac{8\,k^3-11\,k^2r-4\,kr^2-4\,r^3}{r}}}{r^2}\right)^{\frac{1}{3}}}\\ &+\frac{k-r}{3\,r}.\end{aligned}$$

This is the end of the proof. [But like clearly I need to stare at the Sage output for $n_0(k,r)$ above for a while to figure out how to simplify it...] $\square$

Now back to the graph problem. We want to find a tree of order k and diameter d that maximizes the multiplicity of d across all such trees. We will use Lemma 8 with $r=d/2-1$ and $n_0(k-1,d/2-1)$. Since a distance d is counted as a sum of products of the form $(\deg(v_i)-1)(\deg(v_j)-1)$ where $\{v_i,v_j\}\in\binom{V(T)}{2}$ such that $d(v_i,v_j)=d-2$, we have for some $S\subseteq\binom{V(T)}{2}$ that $\operatorname{mult}(d)\leq\sum_{\{v_i,v_j\}\in S}(\deg(v_i)-1)(\deg(v_j)-1)$. By Lemma 8, the largest this sum can be is when there are vertices $v_1,v_2,\ldots,v_{\lceil n_0(k-1,d/2-1)\rceil}$ with balanced degree and degree sum $k-1-(d/2-1)\lceil n_0(k-1,d/2-1)\rceil$ each at distance $d-2$ from one another. We construct a graph with such structure presently, denote it using $G(k,d)$, and call it a *balanced-broom tree* of order k and even diameter d.

Here is the construction for a balanced-broom tree $G(k, d)$ with order k and even diameter d. Begin with a centre vertex v. Then append $\lceil n_0(k-1, d/2-1)\rceil$ length $d/2-1$ paths to v by identifying an end vertex of each path to v. Now for the other end vertices (equivalently, the leaves of the tree built so far), call them $v_1, v_2, \ldots, v_{\lceil n_0(k-1,d/2-1)\rceil}$, we append $k - 1 - (d/2 - 1)\lceil n_0(k - 1, d/2 - 1)\rceil$ leaves to these end vertices in a balanced way so that their degrees differ by at most one. Let $S = \{\{v_i, v_j\} : 1 \le i < j \le \lceil n_0(k - 1, d/2 - 1)\rceil\}$, and notice that each pair $\{v_i, v_j\} \in S$ satisfies $d(v_i, v_j) = d - 2$. Then by Lemma 8, we have

$$\sum_{\{v_i,v_j\}\in S} (\deg(v_i) - 1)(\deg(v_j) - 1) = \max_{\substack{T\in\mathcal{T}(k,d)\\ S'\subseteq\binom{V(T)}{2}\\ \{v_i,v_j\}\in S'\\ d(v_i,v_j)=d-2}} (\deg(v_i) - 1)(\deg(v_j) - 1) = \max_{T\in\mathcal{T}(k,d)} (\mathrm{mult}_T(d)).$$

That is, any tree that maximizes the multiplicity of distance d must, by Lemma 8, maximize the sum on the LHS; $G(k, d)$ does this, therefore $G(k, d)$ maximizes the multiplicity of d across ***all*** trees of order k and diameter d.

It is straightforward to count $\mathrm{mult}_{G(k,d)}(d)$, so we have the following proposition.

Proposition 9.3. *Let T be a tree of order k and diameter d such that d is even. Set $n_0 := n_0(k - 1, d/2 - 1)$ as defined above and set $\ell := k - 1 - (d/2 - 1)\lceil n_0\rceil$. Let $\ell = q\lceil n_0\rceil + s$, where $0 \le s < \lceil n_0\rceil$. Then the following bound holds with tight example $G(k, d)$:*

$$\mathrm{mult}_T(d) \le \binom{s}{2}\left\lceil\frac{\ell}{\lceil n_0\rceil}\right\rceil^2 + \binom{\lceil n_0\rceil - s}{2}\left\lfloor\frac{\ell}{\lceil n_0\rceil}\right\rfloor^2 + s(\lceil n_0\rceil - s)\left\lceil\frac{\ell}{\lceil n_0\rceil}\right\rceil\left\lfloor\frac{\ell}{\lceil n_0\rceil}\right\rfloor.$$

I am quite sure that a similar graph to $G(k, d)$ maximizes the multiplicity for all even distances $2 \le i \le d$. The construction is simply to create a $G(k - (d - i), i)$ and then identify an end-vertex of a path of length $d - i$ to any leaf of $G(k, i)$. This results in a graph with order k and diameter d, which we denote by $G(k, d, i)$. It is straightforward to see that $\mathrm{mult}_{G(k,d,i)}(i) = \mathrm{mult}_{G(k-(d-i),i)}(i) + d - i$. I think a proof that $\mathrm{mult}_{G(k,d,i)}(i) = \max_{T\in\mathcal{T}(k,d)}(\mathrm{mult}_T(i))$ should follow from the truth of $\mathrm{mult}_{G(k-(d-i),i)}(i) = \max_{T\in\mathcal{T}(k-(d-i),i)} \mathrm{mult}_T(i)$ and showing (probably by appealing to Lemma 8) that $G(k, d, i)$ is an optimal way to maximize the number of vertex pairs $\{v_i, v_j\}$ satisfying $d(v_i, v_j) = i - 2$ while also having order d. We just need to show that appending the path of length $d-i$ doesn't undermine optimality of $G(k, d, i)$. I'm pretty sure that this should be easy. Anyway, so I think the following generalization of Proposition 9.3 is true and our discussion thus far is at least a mostly complete proof sketch.

Conjecture 9.4. *Let T be a tree of order k and diameter d (d not necessarily even). Let $2 \le i \le d$ be an even integer. Set $n_0^{(i)} := n_0(k-(d-i)-1, i/2-1)$ and set $\ell^{(i)} := k-(d-i)-1-(i/2-1)\lceil n_0^{(i)}\rceil$. Let $\ell^{(i)} = q\lceil n_0^{(i)}\rceil + s^{(i)}$, where $0 \le s^{(i)} < \lceil n_0^{(i)}\rceil$. Then the following bound holds with tight example $G(k, d, i)$:*

$$\mathrm{mult}_T(i) \le \binom{s^{(i)}}{2}\left\lceil\frac{\ell^{(i)}}{\lceil n_0^{(i)}\rceil}\right\rceil^2 + \binom{\lceil n_0^{(i)}\rceil - s^{(i)}}{2}\left\lfloor\frac{\ell^{(i)}}{\lceil n_0^{(i)}\rceil}\right\rfloor^2 + s^{(i)}(\lceil n_0^{(i)}\rceil - s^{(i)})\left\lceil\frac{\ell^{(i)}}{\lceil n_0^{(i)}\rceil}\right\rceil\left\lfloor\frac{\ell^{(i)}}{\lceil n_0^{(i)}\rceil}\right\rfloor.$$

If we can prove this last optimality step (showing there is no $T \in \mathcal{T}(k, d)$ such that $\mathrm{mult}_T(i) > \mathrm{mult}_{G(k,d,i)}(i)$), then we have a tight upper bound on the multiplicities of even distances in trees, which would be pretty cool! I haven't thought much about odd distances yet – this case seems less nice because of the lack of symmetry in the branches; but perhaps it's not too bad given the even case.

Also, I believe the ***lower bound*** problem is approachable as well, and contrary to the upper bound tight examples, which are very starry, tight examples for the lower bound should be very pathy like in long legged spiders. Note that the lower bound tight example isn't trivially a path of length $k - 1$, because k can be arbitrarily larger than d. So, the construction needs to have branching; but where must this branching occur, and how many branches are there?