# Ph.D. Working Notes

December 12, 2022

## 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 $M \in V(T)$, $i \in [1, n-1]$, $\sum_{j<k}^{\deg(M)} |D_j| \cdot |D_k| < n$. $\square$

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

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

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

*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 $1, 2, \ldots, \sum_{i=a-t+2}^{k} w_i$ occurs as a partial sum in $\mathbf{w}$. Call $w_{a-t+2}, w_{a-t+3}, \ldots, w_a$ the *unit segment* of $\mathbf{w}$ and $w_{a+1}, w_{a+2}, \ldots, w_k$ 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 $r = 1$, $w_{a+r} \in \{1, \ldots, t\}$, and since the unit segment has $t-1$ 1s, for each $j \in \{1, 2, \ldots, t-1\}$, 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 $1, 2, \ldots, \sum_{i=a-t+2}^{a+r-1} w_i$ occur at least once by inductive hypothesis. We have that $w_{a+r} \in \{1, 2, \ldots, t\}$ 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 $j \in \{1, 2, \ldots, t-1\}$. 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 $j \in \{1, 2, \ldots, t-1\}$. $\square$

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 $\{1, 2, \ldots, t\}$. Then for every path $P = (v_1v_2, v_2v_3, \ldots, v_{t-1}v_t)$ in $T$ such that $w(v_iv_{i+1}) = 1$ for $i \in \{1, 2, \ldots, t-1\}$, 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 $1, 2, \ldots, d(v_1, u')$ 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. $\square$

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,

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

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

**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

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

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 $p$. □

# 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* $3 \le k \le \mathrm{Diam}(G)$,

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

*Proof sketch.* The cases $k = 1$ and $k = 2$ are straightforward and no distance can be larger than the diameter of $G$. Suppose $3 \le k \le \mathrm{Diam}(G)$. 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 $m(k)$. □

## 3.1 Conjectures

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

**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 $i \in \{d, \cdots, 2\}$.

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 $n = 20$.

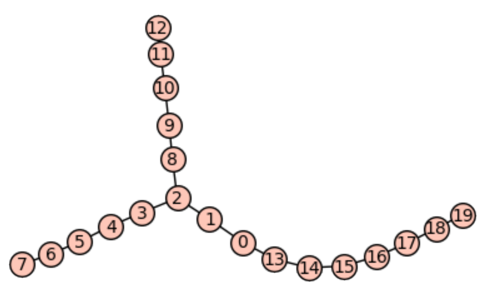



Figure 1: Extremal tree example minimizing $m(d)$.

**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 $d \in \{\lceil n/3 \rceil + 1, \lceil n/3 \rceil + 2\}$.

**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.

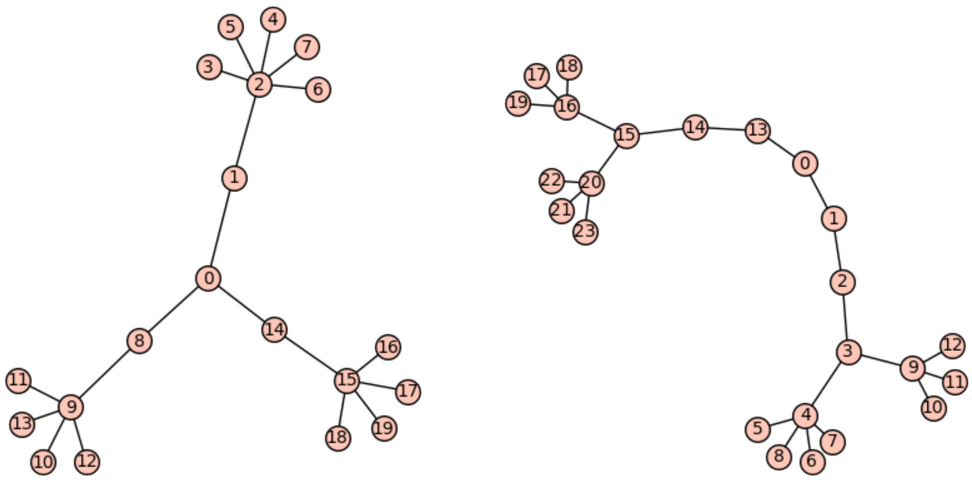
**Construction 1:** Refer to Figure 2a for an example. When $d \leq C_1\frac{n}{3}+C_2$ and even, do the following:

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 3.9.** 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.



(a) Construction 1 (b) Construction 2.

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

**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 $m(d) = 42$.

**Conjecture 3.10.** *Let $d$ be the largest distance with max multiplicity. Then $d \leq \lceil \frac{n}{3} \rceil + 2$.*

**Remark 3.11.** 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 $d \sim n/3$.

## References

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