

IMPROVED BOUNDS FOR EMBEDDING CERTAIN CONFIGURATIONS IN SUBSETS OF VECTOR SPACES OVER FINITE FIELDS

PAIGE BRIGHT, XINYU FANG, BARRETT HERITAGE, ALEX IOSEVICH,
AND MAXWELL SUN

ABSTRACT. The fourth listed author and Hans Parshall ([7]) proved that if $E \subset \mathbb{F}_q^d$, $d \geq 2$, and G is a connected graph on $k + 1$ vertices such that the largest degree of any vertex is m , then if $|E| \geq Cq^{m + \frac{d-1}{2}}$, for any $t \neq 0$, there exist $k + 1$ points x^1, \dots, x^{k+1} in E such that $\|x^i - x^j\| = t$ if the i 'th vertex is connected to the j 'th vertex by an edge in G . In this paper, we give several indications that the maximum degree is not always the right notion of complexity and prove several concrete results to obtain better exponents than the Iosevich-Parshall result affords. This can be viewed as a step towards understanding the right notion of complexity for graph embeddings in subsets of vector spaces over finite fields.

CONTENTS

1. Introduction	11
1.1. Hölder Extensions	13
1.2. Hölder Extensions of Chains	14
1.3. Chains of Simplices	14
1.4. Trees of Simplices	15
2. Preliminary Tools	16
3. Hölder Extension of m -Chains	20
4. Chains of m -Simplices	24
5. Trees of m -Simplices	27
5.1. Hölder Extension Method	28
References	30

Date: Received 28 March 2024. Accepted 4 July 2024.

2020 Mathematics Subject Classification. 42B05, 52C10.

Key words and phrases. Finite point configurations, Erdős distance problem, discrete Fourier transform, character sum estimates.

This work was supported, in part, by the National Science Foundation Grant NSF DMS 2241623 and NSF DMS 1947438. The fourth listed author is supported in part by the National Science Foundation grant NSF DMS 2154232.

1. INTRODUCTION

The Erdős distance problem is one of the fundamental questions in geometric combinatorics. In its original formulation, the question asks, what is the smallest number of distinct distances determined by a subset of \mathbb{R}^d , $d \geq 2$, of a given size. Erdős conjectured in 1946 ([3]) that if P is a finite subset of \mathbb{R}^d , $d \geq 2$, of size N , then for any $\epsilon > 0$ there exists $C_\epsilon > 0$ such that

$$\#\Delta(P) \geq C_\epsilon N^{\frac{2}{d}-\epsilon},$$

where

$$\Delta(P) = \{|x - y| : x, y \in P\},$$

with $|x| = \sqrt{x_1^2 + x_2^2 + \dots + x_d^2}$, the usual Euclidean distance. After more than half a century of efforts by many outstanding mathematicians, the conjecture was resolved in two dimensions by Larry Guth and Nets Katz ([4]). In higher dimensions, the conjecture is still open, with the best-known results due to József Solymosi and Van Vu ([8]).

The purpose of this paper is to study point configuration problems stemming from Erdős type distance problems in vector spaces over finite fields. Throughout this paper, we let q denote an odd prime power. For $d \geq 2$, \mathbb{F}_q^d is the d -dimensional vector space over the field \mathbb{F}_q with q elements. Given $x = (x_1, \dots, x_d) \in \mathbb{F}_q^d$, we let

$$\|x\| = x_1^2 + \dots + x_d^2.$$

In finite fields, the problem takes on its own unique characteristics. For example, suppose that q is a prime congruent to 1 modulo 4. Then $i = \sqrt{-1}$ is in \mathbb{F}_q , which leads to the following example that cannot occur in Euclidean space. Let

$$E = \{(t, it) : t \in \mathbb{F}_q\}.$$

It follows that if we define the distance set

$$\Delta(E) = \{\|x - y\| : x, y \in E\},$$

then $\Delta(E) = \{0\}$. In order to get around this issue, Jean Bourgain, Nets Katz and Terry Tao, who proved the initial result for distance sets in vector spaces over finite fields [2], assumed that q is a prime congruent to 3 modulo 4. Another way around this difficulty was devised by the fourth listed author of this paper and Misha Rudnev. They proved in [5] that if $E \subset \mathbb{F}_q^d$, $d \geq 2$, $t \neq 0$, then

$$|\{(x, y) \in E \times E : \|x - y\| = t\}| = |E|^2 q^{-1} + \text{error},$$

where

$$|\text{error}| \leq 2q^{\frac{d-1}{2}} |E|.$$

The “error” is smaller than the main term when $|E| > 2q^{\frac{d+1}{2}}$ and, in such a case, all non-zero distances are realized. Since the zero distance is realized by taking $x = y$, all distances are in fact realized. This result can be viewed as a statement

about two point configurations or, more precisely, about the embedding of complete graphs on two vertices, in this sense, in subsets of \mathbb{F}_q^d of a given size. This raises the question about more complicated point configurations and graphs that determine them, as well as the embeddings of those graphs in \mathbb{F}_q^d in the sense to be made precise below.

The question we particularly focus on is the following: how large does a subset of $E \subseteq \mathbb{F}_q^d$ need to be such that E necessarily contains a specified graph of points in \mathbb{F}_q^d with distances assigned between edges? In particular, we are interested in embedding *distance graphs* into finite field vector spaces. We call a graph $\mathcal{G} = (V, \mathcal{E})$ a *distance graph* when for each edge $e \in \mathcal{E}$, there is some associated nonzero edge length $\lambda_e \in \mathbb{F}_q^*$. Then, we call $X \subseteq \mathbb{F}_q^d$ an *isometric copy* of \mathcal{G} when there exists a distance preserving bijection $\varphi : V \rightarrow X$ where for each $v, w \in V$ with an edge e connecting v and w , we have $|\varphi(v) - \varphi(w)|^2 = \lambda_e$. We are interested in estimating the following values for various graphs \mathcal{G} :

$$\mathcal{N}_{\mathcal{G}}(E) := \#\{\text{embeddings of a distance graph } \mathcal{G} \text{ in } E\},$$

$$\mathcal{N}_{\mathcal{G}}^*(E) := \#\{\text{non-degenerate embeddings of a distance graph } \mathcal{G} \text{ in } E\}$$

$$\mathcal{D}_{\mathcal{G}}(E) := \#\{\text{degenerate embeddings of a distance graph } \mathcal{G} \text{ in } E\}$$

In recent work by Alex Iosevich and Hans Parshall (see [7]), they showed that given $n, s \in \mathbb{N}$ and $E \subseteq \mathbb{F}_q^d$ with $|E| \geq 12n^2q^{\frac{d-1}{2}+s}$, E contains an isometric copy of every distance graph with n vertices and maximum vertex degree s . The same exponent can be deduced, with some work, from the main result in [9]. This result does not bode well for graphs with high maximum vertex degree, which begs the question: given distance graphs with relatively high vertex degree, how large does a subset $E \subseteq \mathbb{F}_q^d$ need to be to contain that distance graph? In this paper, we study specific configurations of points to improve upon Iosevich and Parshall's results in certain cases.

The simplest example of a graph having high vertex degree but low complexity is a star with many rays. A k -star is a special case of a tree of size $k + 1$, which can be embedded easily as shown in the following theorem by Iosevich, Jardine and McDonald.

Theorem 1.1 ([6]). *Let T be a tree with $k + 1$ vertices and hence k edges. For $\epsilon > 0$, if $|E| > q^{\frac{d+1}{2}+\epsilon}$, then we have*

$$\mathcal{N}_T(E) \geq \frac{|E|^{k+1}}{q^k} \left(1 - 8q^{-\frac{2\epsilon}{k+1}}\right)$$

where $\mathcal{N}_T(E)$ is the number of embeddings of T into E .

Notice that the size of the tree only affects the constant but not the exponent for the size of E required to embed it. Therefore, this result implies that we can embed a star of any degree in a subset of size asymptotically $q^{\frac{d+1}{2}}$, with a constant

that grows with the degree. This greatly improves the bound given by Iosevich–Parshall, which becomes trivial whenever the degree is greater than or equal to $\frac{d+1}{2}$.

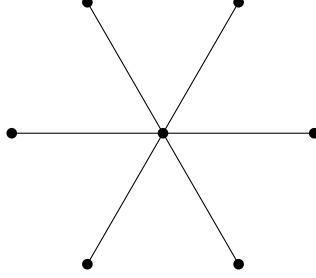


FIGURE 1. A 6-star

In this paper, we only consider distance graphs whose edge lengths are identical, and denote that by $t \in \mathbb{F}_q^*$. From now on, whenever we refer to an edge in a distance graph, it is assumed to have length t .

1.1. Hölder Extensions. We first define the Hölder extension of a graph.

Definition 1.2 (Hölder extension of a Graph). *Let G be a simple graph with vertex set V . Let e be its edge function defined by $e(u, v) = 1$ if $u, v \in V$ are connected and 0 otherwise. Let $S = \{v_1, \dots, v_n\}$ be a subset of V . Let $m \geq 2$ be an integer. Then the k -Hölder extension of this graph (with respect to the set S) is a new graph G' defined as follows. The vertex set of G' consists of V plus $k - 1$ copies of the vertices in S , labelled $\{v_j^i : 1 \leq j \leq n, 1 \leq i \leq k - 1\}$. The edge function e' of G' is defined as follows.*

$$e'(v_j^i, w) = \begin{cases} e(v_j, w), & w \in V \setminus S \\ e(v_j, v_j^i), & w = v_j^i \\ 0, & \text{otherwise,} \end{cases}$$

and $e' = e$ on $V \times V$.

We call this a Hölder extension because it is relatively easy to apply Hölder's inequality in order to lower bound the number of embeddings of these graphs. We have the following results.

Lemma 1.3. *Let $E \subseteq \mathbb{F}_q^d$, $d \geq 2$. Let G be a Hölder extension of H of degree $k \geq 2$, with $S = \{v_1, \dots, v_n\}$, $V \setminus S = \{w_1, \dots, w_l\}$, where V is the vertex set of H and S is the subset of vertices being duplicated in G . Then we have that*

$$\mathcal{N}_G(E) \geq \frac{(\mathcal{N}_H(E))^k}{(\mathcal{N}_{H \setminus S}(E))^{k-1}}.$$

Corollary 1.4. *Let $E \subseteq \mathbb{F}_q^d$, $d \geq 2$. Let G be a Hölder extension of H of degree $k \geq 1$, with $S = \{v_1, \dots, v_n\}$, $V \setminus S = \{w_1, \dots, w_l\}$. Then we have*

$$\mathcal{N}_G(E) \geq \frac{(\mathcal{N}_H(E))^k}{|E|^{(k-1)l}}.$$

Corollary 1.4 follows immediately from Lemma 1.3 since the number of embeddings of $H \setminus S$ in E is at most $|E|^l$. Section 2 gives a proof of Lemma 1.3 and also states more results regarding counting degenerate embeddings of Hölder extensions.

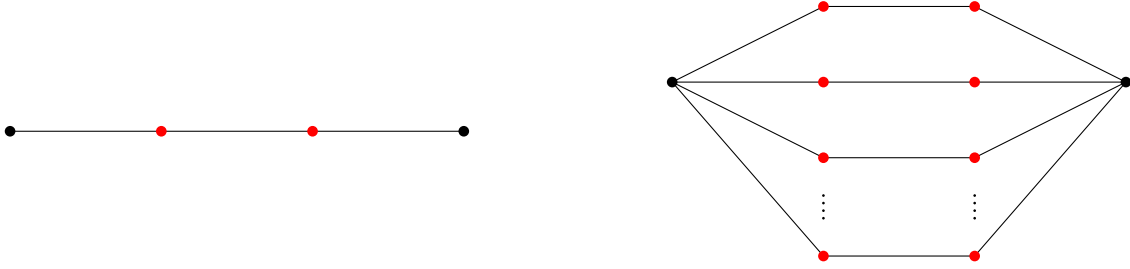


FIGURE 2. A 3-chain and the k -Hölder extension of it with respect to the middle vertices

1.2. Hölder Extensions of Chains. The first example we study is a Hölder extension of chains.

Definition 1.5 (m -chain). *An m -chain is a distance graph on $m+1$ vertices (x_1, \dots, x_{m+1}) such that x_i and x_{i+1} are connected by an edge for $1 \leq i \leq m$.*

We denote an m -chain by C_m . The number of non-degenerate embeddings of m -chains in a subset $E \subseteq \mathbb{F}_q^d$ were studied by Bennett et.al. in [1].

In Section 3, we consider the k -Hölder extension of an m -chain fixing the endpoints, denoted by $G_{k,m}$, as illustrated in Figure 2. Corollary 1.4 immediately implies the following given input on the count of embeddings of m -chains.

Theorem 1.6. *Given that $|E| > \frac{4k}{\ln 2} q^{\frac{d+1}{2}}$,*

$$\mathcal{N}_{G_{k,m}}(E) \geq \frac{|E|^{km-k+2}}{q^{km}} \left(1 - \frac{4mq^{\frac{d+1}{2}}}{\ln 2 \cdot |E|} \right)^k.$$

More importantly, we give further results on the number of *non-degenerate* embeddings of such configurations in Section 3.

1.3. Chains of Simplices. The second example we study are chains of m -simplices, discussed in detail in Section 4. We first recall the definition of an m -simplex.

Definition 1.7 (Regular m -simplex). *An m -simplex is a distance graph on $m+1$ vertices such that there is an edge between every pair of vertices.*

Now, we can form chains of simplices by linking copies of them up in a certain way. For example, Figure 3 shows a chain of 2-simplices.

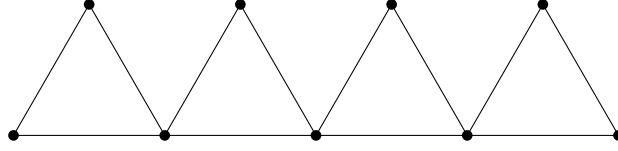


FIGURE 3. A 4-chain of 2-simplices

We define a k -chain of m -simplices precisely as follows.

Definition 1.8 (k -chains of m -simplices). *Let $k \geq 3$. A k -chain of m -simplices is a distance graph obtained by replacing each edge in a k -chain with an m -simplex, with two vertices in the simplex at the two endpoints of the original edge of the chain.*

Let T_k^m denote a k -chain of m -simplices. We have the following results regarding embeddings of T_k^m into subsets of \mathbb{F}_q^d .

Theorem 1.9. *Let ℓ be such that $2^\ell < k \leq 2^{\ell+1}$. Suppose $E \subset \mathbb{F}_q^d$ with $m < \frac{d+1}{2}$. Whenever $|E| > 2^{\ell+4} \cdot 3(m+1)^2 q^{\frac{d-1}{2}+m}$,*

$$\mathcal{N}_{T_k^m}(E) \geq \frac{|E|^{mk+1}}{q^{k\binom{m+1}{2}}} \cdot 2^{2k-2^{\ell+1}(m+2)-2}.$$

Theorem 1.10. *Let ℓ be such that $2^\ell < k \leq 2^{\ell+1}$. Suppose $E \subset \mathbb{F}_q^d$ with $m < \frac{d+1}{2}$. Whenever $|E| > 2^{k+2^{\ell+1}(m+2)+3} \cdot 3(m+1)^2 q^{\frac{d-1}{2}+m}$,*

$$\mathcal{N}_{T_k^m}^*(E) \geq \frac{|E|^{mk+1}}{q^{k\binom{m+1}{2}}} \cdot 2^{2k-2^{\ell+1}(m+2)-3}.$$

1.4. Trees of Simplices. Lastly, in Section 5, we study trees of m -simplices. These are a natural generalization of chains of m -simplices, defined as follows.

Definition 1.11 (Tree of m -simplices). *A tree of m -simplices is the graph obtained from a usual tree by replacing each of its edges with an m -simplex, with two vertices of the simplex being the two endpoints of the original edge. We will call the vertices coming from the original tree nodes, and the m -simplices connections. The size of such a tree is the number of simplices it contains.*

Theorem 1.12. *Let T_ℓ be a tree of m -simplices of size ℓ . Suppose*

$$|E| \geq 12(m+1)^2 \cdot 2^{(\ell-1)(m+5)} \cdot q^{m+\frac{d-1}{2}}.$$

Then

$$\mathcal{N}_{T_\ell}(E) \geq 2^{-\eta(m,\ell)} \frac{|E|^{m\ell+1}}{q^{\binom{m+1}{2} \cdot \ell}}$$

where

$$\eta(m, \ell) = (\ell - 1)(m + 5)(\ell m/2 + 1) - (m + 2)\ell.$$

Theorem 1.13. *Let T_ℓ be a tree of m -simplices of size ℓ . Suppose*

$$|E| \geq 2^{\eta(m, \ell) + 3\ell} \cdot 12(m + 1)^2 \cdot q^{m + \frac{d-1}{2}}.$$

Then

$$\mathcal{N}_{T_\ell}^*(E) \geq 2^{-\eta(m, \ell) + 3\ell + 1} \frac{|E|^{m\ell + 1}}{q^{\binom{m+1}{2} \cdot \ell}}$$

where η is as defined in Theorem 1.12.

The constants grow/diminish exponentially with ℓ and m , but if we are willing to keep them constant, then we still get the statistically correct lower bounds.

2. PRELIMINARY TOOLS

We first prove Lemma 1.3, which is a useful tool for obtaining a lower bound on the number of embeddings of a graph that is a Hölder extension of a simpler graph.

Proof of Lemma 1.3. Let $F(x_1, \dots, x_l) = F(x_I)$ be the indicator function for the configuration $H - S$; i.e. $F(x_I) = 1$ if $(w_1 = x_1, \dots, w_l = x_l)$ is an embedding of $H - S$ and $F(x_I) = 0$ otherwise. Let $G(x_1, \dots, x_l, y_1, \dots, y_k) = G(x_I, y_J)$ be the indicator function for the configuration G defined similarly, where $w_i = x_i$ and $v_j = y_j$.

Then

$$\mathcal{N}_H(E) = \sum_{x_1, \dots, x_l \in \mathbb{F}_q^d} \left(F(x_I) \sum_{y_1, \dots, y_k \in \mathbb{F}_q^d} G(x_I, y_J) \right).$$

Therefore,

$$\begin{aligned} \mathcal{N}_G(E) &= \sum_{x_1, \dots, x_l \in \mathbb{F}_q^d} \left(F(x_I) \sum_{y_1, \dots, y_k \in \mathbb{F}_q^d} G(x_I, y_J) \right)^m \\ &\geq \frac{\left(\sum_{x_1, \dots, x_l \in \mathbb{F}_q^d} F(x_I) \sum_{y_1, \dots, y_k \in \mathbb{F}_q^d} G(x_I, y_J) \right)^k}{\left(\sum_{x_1, \dots, x_l \in \mathbb{F}_q^d} F(x_I) \right)^{k-1}} && \text{(Hölder's inequality)} \\ &= \frac{(\mathcal{N}_H(E))^k}{(\mathcal{N}_{H \setminus S}(E))^{k-1}}. \end{aligned}$$

□

Below, we give a means for counting degeneracies for embeddings of connected graphs. This is helpful since Lemma 1.3 gives a lower bound on the number of total embeddings but allows for possibly many degenerate counts.

Theorem 2.1. Fix a subset L of \mathbb{F}_q . For $\epsilon > 0$ and a positive integer $n \geq 2$, if $|E| > q^{\frac{d+1}{2} + \epsilon}$, there exists a subset $E' \subseteq E$ such that

$$|E \setminus E'| \leq 2|L|q^{-\frac{2\epsilon}{n}}|E|$$

and the total number of embeddings of all connected distance graphs with n vertices with edge lengths all chosen from L in E' is bounded above as follows:

$$\mathcal{N}_{n,L}(E') \leq c_{n,|L|} \frac{|E|^n}{q^{n-1}} \quad (1)$$

where $c_{n,|L|} = 2^{\binom{n}{2} - n + 2} |L|^{n-1} n^{n-2}$.

In addition, if G is a connected graph with edge lengths chosen solely from L with $n+1$ vertices, then the number of degenerate embeddings of G in E' is bounded above by

$$n(n+1) \frac{|E|^n}{q^{n-1}}.$$

Proof. Let $\phi_t : \mathbb{F}_q^d \times \mathbb{F}_q^d \rightarrow \mathbb{R}$ be such that $\phi_t(x, y) = 1$ if $\|x - y\| = t$, and $\phi_t(x, y) = 0$ otherwise. Then, let

$$E_t = \left\{ x \in E : \sum_{y \in E} E(y) \phi_t(x, y) \leq \lambda \frac{|E|}{q} \right\}, \quad E' = \bigcap_{t \in L} E_t.$$

We have,

$$|E \setminus E_t| \leq \frac{q}{\lambda|E|} \sum_{x,y} E(x)E(y)\phi_t(x, y) \leq 2 \frac{|E|}{\lambda} \quad (2)$$

by [1], so

$$|E \setminus E'| \leq 2|L| \frac{|E|}{\lambda}. \quad (3)$$

We need the following result. It is analogous to Lemma 4.4 in [6] and proved in exactly the same way.

Lemma 2.2. Let T be a tree with $r+1$ vertices and r edges with edge lengths only from the set L . If n_T^* is the number of embeddings of T in E' , then

$$\left| n_T^* - \frac{|E|^{r+1}}{q^r} \right| \leq 4r \left(\lambda^{-1} + \lambda^{\frac{r-1}{2}} q^{\frac{d+1}{2}} \frac{|E|}{q} \right) \frac{|E|^{r+1}}{q^r}.$$

We can set $\lambda = q^{2\epsilon/(r+1)}$ and use the above lemma to obtain

$$\left| n_T^* - \frac{|E|^{r+1}}{q^r} \right| \leq 8 \cdot \frac{|E|^{r+1}}{q^r} \cdot q^{-\frac{2\epsilon}{r+1}}. \quad (4)$$

Now, note that all graphs with n vertices are spanned by a tree with at most n vertices. Given an assignment of lengths in L to edges, the number of embeddings of a graph is at most the number of embeddings of any spanning tree with analogous

edge assignment. Let N_T be the number of simple graphs with n vertices with T as a spanning tree. For a given T , we have at most $|L|^{n-1}$ ways to assign the edge lengths. Thus, summing over all trees T with n vertices, we have,

$$\mathcal{N}_{n,L}(E') \leq \sum_T N_T |L|^{n-1} n_T^* \quad (5)$$

$$= 2^{\binom{n}{2}-n+1} |L|^{n-1} \cdot n^{n-2} \cdot 2 \frac{|E|^n}{q^{n-1}} \quad (6)$$

$$\leq 2^{\binom{n}{2}-n+2} |L|^{n-1} \cdot n^{n-2} \cdot \frac{|E|^n}{q^{n-1}}. \quad (7)$$

The last part of the theorem follows from Eq. (4) and, noting that a degeneration involves setting two vertices equal, which results in a connected graph with $n - 1$ vertices. □

We introduce the following ‘‘Shaving Lemma’’, which allows us to find a subset of a given set $E \subseteq \mathbb{F}_q^d$ in which the vertex degrees are close to the statistically correct count, without losing too many points. Let

$$S(x) := \begin{cases} 1, & \text{if } \|x\| = t, \\ 0, & \text{otherwise.} \end{cases}$$

Then we have that the convolution

$$E * S(x) = \sum_{y \in \mathbb{F}_q^d} E(y) S(x - y) = \#\{\text{edges in } E \text{ incident on } x\}.$$

Theorem 2.3 (Shaving Lemma for One Edge). *Let $E \subseteq \mathbb{F}_q^d$ with $|E| \geq Cq^{\frac{d+1}{2}}$, where $C > 4$. Then for any $\lambda > \frac{2C}{C-4}$, there exists a subset $E^* \subseteq E$ with*

$$|E^*| \geq \frac{(1 - 2\lambda^{-1})}{2\lambda} |E|$$

such that for all $x \in E^*$

$$(1 - 2\lambda^{-1})\lambda^{-1} \frac{|E|}{q} \leq E * S(x) \leq \lambda \frac{|E|}{q}.$$

Proof. Let

$$E' := \left\{ x \in E : E * S(x) \leq \lambda \cdot \frac{|E|}{q} \right\}.$$

By counting the number of edges in E incident to vertices in $E \setminus E'$, we see that

$$|E \setminus E'| \cdot \lambda \frac{|E|}{q} \leq \sum_x E * S(x) \leq \frac{2|E|^2}{q},$$

where the last inequality holds since $|E| \geq 2q^{\frac{d+1}{2}}$. Thus, $|E \setminus E'| \leq \frac{2|E|}{\lambda}$, which implies

$$|E'| \geq |E|(1 - 2\lambda^{-1}) \geq 4q^{\frac{d+1}{2}}$$

by our choice of λ . Now let

$$E^* := \left\{ x \in E' : E' * S(x) \geq \lambda^{-1} \cdot \frac{|E'|}{q} \right\}.$$

Then,

$$|E' \setminus E^*| \cdot \lambda^{-1} \frac{|E'|}{q} + |E^*| \cdot \lambda \frac{|E'|}{q} \geq \sum_x E' * S(x) \geq \frac{|E'|^2}{2q}.$$

Upper bounding $|E' \setminus E^*|$ by $|E'|$ and rearranging, we get

$$\implies |E^*| \geq \left(\frac{1}{2} - \lambda^{-1} \right) \frac{|E'|^2}{\lambda \frac{|E'|}{q}} = \frac{1}{2\lambda} (1 - 2\lambda^{-1}) |E'|.$$

□

Note in the above proof that the constants λ and λ^{-1} in the upper and lower bounds can be made independent, as long as they are chosen in an appropriate range. In fact, this theorem can be stated much more generally for an arbitrary configuration with a given basepoint for which we have a lower bound on the size of a subset containing the statistically correct number of embeddings. The proof follows essentially the same idea as above.

Theorem 2.4 (General Shaving Lemma). *Let \mathcal{G} be a distance graph with n vertices and m edges, and a chosen vertex v (which we call the “basepoint”). Suppose that, for all $E \subseteq \mathbb{F}_q^d$ with $|E| > N$,*

$$\frac{|E|^n}{2q^m} \leq \mathcal{N}_{\mathcal{G}}(E) \leq \frac{2|E|^n}{q^m}.$$

For $x \in E$ let

$$f(x) := \#\{\text{embeddings of } \mathcal{G} \text{ in } E \text{ with } v \text{ at } x\}. \quad (8)$$

For $\lambda_2 \geq 4$ and $\lambda_1 < 2^{-n}$, we have that, given $|E| > 2N$,

$$\left| \left\{ x \in E : \lambda_1 \frac{|E|^{n-1}}{q^m} \leq f(x) \leq \lambda_2 \frac{|E|^{n-1}}{q^m} \right\} \right| \geq \left(\frac{1}{2} - 2^{n-1} \lambda_1 \right) (1 - 2\lambda_2^{-1})^n |E|. \quad (9)$$

Proof. Let

$$E' = \left\{ x \in E : f(x) \leq \lambda_2 \cdot \frac{|E|^{n-1}}{q^m} \right\},$$

We have,

$$|E \setminus E'| \leq \frac{q^m}{\lambda_2 |E|^{n-1}} \sum_{x \in E} f(x) \leq \frac{q^m}{\lambda_2 |E|^{n-1}} \frac{2|E|^n}{q^m} = \frac{2}{\lambda_2} |E|$$

which implies

$$|E'| \geq (1 - 2\lambda_2^{-1})|E| \geq \frac{|E|}{2}. \quad (10)$$

Now, let

$$g(x) := \#\{\text{embeddings of } \mathcal{G} \text{ in } E' \text{ with } v \text{ at } x\}. \quad (11)$$

and

$$E^* = \left\{ x \in E' : g(x) \geq \lambda_0 \frac{|E'|^{n-1}}{q^m} \right\}.$$

We obtain that

$$\begin{aligned} \frac{|E'|^n}{2q^m} &\leq \mathcal{N}_{\mathcal{G}}(E') = \sum_{x \in E'} g(x) \leq |E' \setminus E^*| \lambda_0 \frac{|E'|^{n-1}}{q^m} + |E^*| \lambda_2 \frac{|E'|^{n-1}}{q^m} \\ \implies |E^*| &\geq \left(\frac{1}{2} - \lambda_0 \right) \frac{|E'|^n}{q^m} / \left[\lambda_2 \frac{|E'|^{n-1}}{q^m} \right] \geq \left(\frac{1}{2} - \lambda_0 \right) (1 - 2\lambda_2^{-1})^n |E|. \end{aligned}$$

For any $x \in E^*$, we obtain that

$$f(x) \geq g(x) \geq \lambda_0 2^{1-n} \frac{|E'|^{n-1}}{q^m}.$$

Now, let $\lambda_1 = 2^{1-n} \lambda_0$ to obtain the result. □

3. HÖLDER EXTENSION OF m -CHAINS

In this section, we study Hölder extensions of m -chains. In particular, we find a bound on the total number of k -fold Hölder extensions of an m -chain given a restriction on $|E|$, then deduce a count for non-degenerate extensions. We do this by counting degeneracies.

Assume $k \geq 3$, since $k = 2$ implies a cycle. Recall that C_m denotes an m -chain and $G_{k,m}$ denotes the k -fold Hölder extension of C_m with two fixed endpoints.

We need the following input by Bennett et al. on the number of non-degenerate embeddings of chains.

Theorem 3.1 ([1]). *Suppose $|E| \geq \frac{4m}{\ln 2} \cdot q^{\frac{d+1}{2}}$. Then*

$$\left| \mathcal{N}_{C_m}^*(E) - \frac{|E|^{m+1}}{q^m} \right| \leq \frac{4m}{\ln 2} \cdot q^{\frac{d+1}{2}} \cdot \frac{|E|^m}{q^m}.$$

We immediately deduce the following.

Theorem 3.2. *Given that $|E| > \frac{4k}{\ln 2} q^{\frac{d+1}{2}}$,*

$$\mathcal{N}_{G_{k,m}}(E) \geq \frac{|E|^{km-k+2}}{q^{km}} \left(1 - \frac{4mq^{\frac{d+1}{2}}}{\ln 2 \cdot |E|} \right)^k.$$

Proof. By Corollary 1.4,

$$\mathcal{N}_{G_{k,m}}(E) \geq \frac{\mathcal{N}_{G_{1,m}}^*(E)^k}{|E|^{2(k-1)}} \geq \frac{|E|^{k(m-1)+2}}{q^{km}} \left(1 - \frac{4mq^{\frac{d+1}{2}}}{\ln 2 \cdot |E|}\right)^k.$$

□

Theorem 3.3. When $|E| > \max\left((km)^{km} q^{\frac{d+1}{2}}, 2(km)^2 q^k\right)$,

$$\mathcal{N}_{G_{k,m}}^*(E) \geq \frac{|E|^{k(m-1)+2}}{2q^{km}} \left(1 - \frac{1}{km} - \frac{2(km)^2 q^k}{|E|}\right),$$

which is positive when

$$|E| \geq \frac{2(km)^3 q^k}{km - 1}.$$

Alternatively, assume that $|E| > (\log q) \max\left(q^{\frac{d+1}{2}}, q^k\right)$. Then we obtain an asymptotic lower bound for sufficiently large q :

$$\frac{|E|^{k(m-1)+2}}{q^{km}} \left(1 - \frac{8mk}{\ln 2 \ln q} - 2km(\ln q)^{-\frac{2}{km}}\right).$$

Proof. We first make the general assumption that $q \geq f(q)q^{\frac{d+1}{2}}$. By Theorem 2.1, there exists a subset $E' \subseteq E$ such that

$$|E \setminus E'| \leq \lambda|E| \text{ in other words } (1 - \lambda)|E| \leq |E'|.$$

where $\lambda = 2f(q)^{-\frac{2}{k(m-1)+1}}$, so that

$$\mathcal{D}_{G_{m,k}}(E') \leq c_{k,m} \frac{|E|^{k(m-1)+1}}{q^{k(m-1)}}, \quad (12)$$

where $c_{k,m} = (k(m-1) + 2)(k(m-1) + 1)$. Now it suffices to choose a suitable $f(q)$ so that when $|E| > f(q)q^{\frac{d+1}{2}}$ we have that

$$|E'| > \frac{8km}{\ln 2} q^{\frac{d+1}{2}}$$

so that

$$\begin{aligned} \mathcal{N}_{G_{k,m}}(E') &\geq \frac{|E'|^{k(m-1)+2}}{q^{km}} \left(1 - \frac{4mq^{\frac{d+1}{2}}}{\ln 2 \cdot |E'|}\right)^k \\ &\geq \frac{|E|^{k(m-1)+2}}{q^{km}} \left(1 - \frac{4mkq^{\frac{d+1}{2}}}{\ln 2 \cdot |E'|}\right) (1 - \lambda)^{k(m-1)+2} \\ &\geq \frac{|E|^{k(m-1)+2}}{q^{km}} \left(1 - \frac{4mk}{\ln 2(1 - \lambda)f(q)}\right) (1 - (k(m-1) + 2)\lambda). \end{aligned} \quad (13)$$

Since $|E'| \geq (1 - \lambda)|E|$, it suffices to have

$$f(q)(1 - \lambda) = f(q) \left(1 - f(q)^{-\frac{2}{k(m-1)+1}}\right) > \frac{8km}{\ln 2}.$$

This is clearly true by the assumptions. We can then simplify (13) to obtain

$$\mathcal{N}_{G_{k,m}}(E') \geq \frac{|E|^{k(m-1)+2}}{2q^{km}} \left(1 - kmf(q)^{-\frac{2}{km}}\right) \geq \frac{|E|^{k(m-1)+2}}{2q^{km}} \left(1 - \frac{1}{km}\right).$$

From (12), we can get

$$\mathcal{D}_{G_{m,k}}(E') < (km)^2 \frac{q^k}{|E|} \cdot \frac{|E|^{k(m-1)+2}}{q^{km}}$$

so that

$$\mathcal{N}_{G_{k,m}}^*(E) \geq \frac{|E|^{k(m-1)+2}}{2q^{km}} \left(1 - \frac{1}{km} - \frac{2(km)^2 q^k}{|E|}\right)$$

Now, if we alternatively assume that $f(q) = \log q$, then $\lambda < \frac{1}{2}$ for sufficiently large q and (12) turns into

$$\mathcal{N}_{G_{k,m}}(E') \geq \frac{|E|^{k(m-1)+2}}{q^{km}} \left(1 - \frac{8mk}{\ln 2f(q)}\right) (1 - (k(m-1) + 2)\lambda) \quad (14)$$

$$\geq \frac{|E|^{k(m-1)+2}}{q^{km}} \left(1 - \frac{8mk}{\ln 2f(q)} - 2kmf(q)^{-\frac{2}{km}}\right). \quad (15)$$

□

Theorem 3.4. *Assume that $|E| > q^{\frac{1}{2}(d+2-\frac{m-2}{m-1}+\delta)}$ for some $0 < \delta < \frac{1}{2m^2}$. Then,*

$$\mathcal{N}_{G_{k,m}}^*(E') \geq \frac{|E|^{k(m-1)+2}}{q^{km}} \left(1 - \frac{2km}{\ln q} - \frac{4km}{\ln 2q^{\frac{1}{m-1}+\delta}} - \frac{2(\ln q)^{(m-2)(k-2)-1} q^{d(k-2)+1}}{|E|^{k-1}}\right).$$

The above is positive and tends to the statistically correct count when q is sufficiently large and

$$|E| > (\ln q)^{m-2} q^{\frac{d(k-2)+1}{k-1}}.$$

Proof. Suppose $\lambda \geq 4$ and let

$$E' = \left\{x \in E : \sum_{y \in E} S_t(x, y) \leq \lambda \frac{|E|}{q}\right\}.$$

Then we have

$$|E'| \geq \left(1 - \frac{2}{\lambda}\right)|E|.$$

The number of embeddings of a $2m$ -cycle in E' is at most

$$2 \frac{|E|^{2m}}{q^{2m}}$$

when q is sufficiently large and

$$|E'| \geq q^{\frac{1}{2}(d+2-\frac{m-2}{m-1}+\delta')}$$

where $0 < \delta' < \frac{1}{2m^2}$ (simply take δ' marginally smaller than δ since $|E'| \geq \frac{1}{2}|E|$). Since the “branches” involve $(m-2)(k-3) + (m-3) = (m-2)(k-2) - 1$ vertices (there are $k-3$ non-overlapping chains outside of the cycle, each having $m-2$ vertices on the branches, and there are $m-3$ vertices on the branches on the overlapping chain) and there are $k-2$ vertices remaining, we have that

$$\mathcal{D}_{G_{k,m}}(E') \leq 2 \frac{|E|^{2m}}{q^{2m}} \left(\frac{\lambda|E|}{q} \right)^{(m-2)(k-2)-1} \cdot q^{(d-2)(k-2)} = 2\lambda^{(m-2)(k-2)-1} \cdot \frac{|E|^{(m-2)k+3}}{q^{(m-d)k+2d-1}}.$$

Now when λ is sufficiently large

$$\mathcal{N}_{G_{k,m}}(E') \geq \left(1 - \frac{2}{\lambda}\right)^{k(m-1)+2} \cdot \frac{|E|^{k(m-1)+2}}{q^{km}} \left(1 - \frac{4mq^{\frac{d+1}{2}}}{\ln 2 \cdot |E'|}\right)^k.$$

Then, suppose that $\lambda \rightarrow \infty$ as $q \rightarrow \infty$. The above can then be written as

$$\mathcal{N}_{G_{k,m}}(E') \geq \frac{|E|^{k(m-1)+2}}{q^{km}} \left(1 - 2km\lambda^{-1} - \frac{4km}{\ln 2q^{\frac{1}{m-1}+\delta}}\right)$$

so that

$$\mathcal{N}_{G_{k,m}}(E') - \mathcal{D}_{G_{k,m}}(E') \geq \frac{|E|^{k(m-1)+2}}{q^{km}} \left(1 - 2km\lambda^{-1} - \frac{4km}{\ln 2q^{\frac{1}{m-1}+\delta}} - 2\lambda^{(m-2)(k-2)-1} \frac{q^{d(k-2)+1}}{|E|^{k-1}}\right).$$

Let $\lambda = \log q$ to get the result. □

The second bound is always tighter than the bound

$$|E| \gg q^{\frac{1}{2}(d+2-\frac{m-2}{m-1}+\delta)}$$

required by the cycles result, except when $k=3$ or $(k,d,m) = (4,2,3), (4,2,4), (5,2,3)$. We summarize the bounds we obtain as follows.

Theorem 3.5. *We have an asymptotic lower bound on $\mathcal{N}_{G_{k,m}}^*$ when*

$$|E| \geq \min \left\{ (\log q) \max \left(q^{\frac{d+1}{2}}, q^k \right), \max \left(q^{\frac{1}{2}(d+2-\frac{m-2}{m-1}+\delta)}, (\ln q)^{m-2} q^{\frac{d(k-2)+1}{k-1}} \right) \right\}.$$

When $k \geq 4$ and $(k,d,m) \neq (4,2,3), (4,2,4), (5,2,3)$, as discussed above, the condition in Theorem 3.5 becomes the following:

$$|E| \geq \min \left\{ (\ln q) q^{\max\{k, \frac{d+1}{2}\}+\epsilon}, (\ln q)^{m-2} q^{\frac{d(k-2)+1}{k-1}+\epsilon} \right\}.$$

In this case, the optimal bounds are as follows:

- when $k < \frac{d+1}{2}$: $|E| \geq (\ln q) q^{\frac{d+1}{2}}$;

- when $\frac{d+1}{2} \leq k < d - 1 - \frac{1}{k-2}$: $|E| \geq (\ln q)q^k$;
- when $k \geq d - 1 - \frac{1}{k-2}$: $|E| \geq (\ln q)^{m-2}q^{\frac{d(k-2)+1}{k-1}}$.

For $k \geq 4$, $\frac{1}{k-2} < 1$, so we have simplified ranges for the optimal bounds that do not involve k :

- when $k < \frac{d+1}{2}$: $|E| \geq (\ln q)q^{\frac{d+1}{2}}$;
- when $\frac{d+1}{2} \leq k \leq d - 2$: $|E| \geq (\ln q)q^k$;
- when $k \geq d - 1$: $|E| \geq (\ln q)^{m-2}q^{\frac{d(k-2)+1}{k-1}}$.

4. CHAINS OF m -SIMPLICES

Throughout this section, let $m \geq 2$ and $k \geq 1$ be integers.

In this section, we prove a lower bound on the size of a subset E such that it contains the statistically correct number of copies of a k -chain of m -simplices. We also show that under the same hypothesis, a *non-degenerate* embedding exists.

Recall that T_k^m denotes a k -chain of m -simplices, sometimes also referred to as a chain of m -simplices of length k . Note that the number of edges and vertices in T_k^m are given by

$$\#\text{edges} = \frac{(m+1)mk}{2}, \quad \#\text{vertices} = mk + 1.$$

Proof of Theorem 1.9. Theorem 7 of [7] implies the following asymptotic count of *non-degenerate* embeddings of T_1^m in a subset $E \subset \mathbb{F}_q^d$.

Lemma 4.1. *For $|E| \geq 12(m+1)^2q^{m+\frac{d-1}{2}}$, we have*

$$\left| \mathcal{N}_{T_1^m}^*(E) - \frac{|E|^{m+1}}{q^{\binom{m+1}{2}}} \right| \leq 6(m+1)^2|E|^m q^{\frac{d-1}{2} - \frac{m(m-1)}{2}}.$$

In particular, we have the following lower bound.

$$\mathcal{N}_{T_1^m}^*(E) \geq \frac{|E|^{m+1}}{q^{\binom{m+1}{2}}} \left(1 - \frac{6(m+1)^2q^{\frac{d-1}{2}+m}}{|E|} \right).$$

Note that this result is non-trivial only when $m < \frac{d+1}{2}$, so we assume that throughout.

Let v be a degree m vertex of the first m -simplex in T_k^m . For $x \in \mathbb{F}_q^d$, let

$$f_{k,m}(x) := \#\{\text{embeddings of } T_k^m \text{ in } E \text{ with } v = x\}.$$

We prove by induction that whenever

$$|E| \geq 12(m+1)^2q^{m+\frac{d-1}{2}},$$

the number of embeddings of chains of length $k = 2^\ell$ is lower bounded by

$$\frac{|E|^{2^\ell m+1}}{q^{2^\ell \binom{m+1}{2}}} \left(1 - \frac{2^{\ell+1} \cdot 3(m+1)^2 q^{\frac{d-1}{2}+m}}{|E|} \right)$$

using Cauchy-Schwarz. The base case $\ell = 0$ was Lemma 4.1. The inductive step follows from

$$\begin{aligned} \mathcal{N}_{T_{2^{\ell+1}}^m}(E) &= \sum_{x \in E} f_{2^\ell, m}(x)^2 \\ &\geq |E|^{-1} \left(\sum_{x \in E} f_{2^\ell, m}(x) \right)^2 \\ &\geq |E|^{-1} \mathcal{N}_{T_{2^\ell}^m}(E)^2 \\ &\geq \frac{|E|^{2^{\ell+1} m+1}}{q^{2^{\ell+1} \binom{m+1}{2}}} \left(1 - \frac{2^{\ell+2} \cdot 3(m+1)^2 q^{\frac{d-1}{2}+m}}{|E|} \right). \end{aligned}$$

Now suppose $2^\ell < k \leq 2^{\ell+1}$. We induct downwards $2^{\ell+1} - k$ steps to prove that whenever

$$|E| \geq \left(1 - \frac{2}{\lambda} \right)^{-1} 12(m+1)^2 q^{m+\frac{d-1}{2}},$$

we have

$$\mathcal{N}_{T_k^m}(E) \geq \frac{|E|^{mk+1}}{q^{k \binom{m+1}{2}}} \cdot \frac{\left(1 - \frac{2}{\lambda} \right)^{2^{\ell+1} m+1}}{\lambda^{2^{\ell+1}-k}} \left(1 - \frac{2^{\ell+2} \cdot 3(m+1)^2 q^{\frac{d-1}{2}+m}}{\left(1 - \frac{2}{\lambda} \right) |E|} \right).$$

The base case $k = 2^{\ell+1}$ was done earlier. For the inductive step, observe that

$$\mathcal{N}_{T_{k+1}^m}(E) = \sum_{x \in E} f_{k,m}(x) f_{1,m}(x) \leq \left(\max_{x \in E} f_{1,m}(x) \right) \sum_{x \in E} f_{k,m}(x) = \left(\max_{x \in E} f_{1,m}(x) \right) \mathcal{N}_{T_k^m}(E).$$

This implies that

$$\mathcal{N}_{T_k^m}(E) \geq \frac{\mathcal{N}_{T_{k+1}^m}(E)}{\max_{x \in E} f_{1,m}(x)}.$$

Now consider

$$E' := \left\{ x \in E : f_{1,m}(x) \leq \lambda \cdot \frac{|E|^m}{q^{\binom{m+1}{2}}} \right\}.$$

We have,

$$|E \setminus E'| \leq \frac{q^{\binom{m+1}{2}}}{\lambda |E|^m} \sum_{x \in E} f_{1,m}(x) \leq \frac{2|E|}{\lambda},$$

so

$$|E'| \geq |E| \left(1 - \frac{2}{\lambda} \right) \geq 12(m+1)^2 q^{m+\frac{d-1}{2}}.$$

Then

$$\begin{aligned} \mathcal{N}_{T_k}(E') &\geq \mathcal{N}_{T_{k+1}^m}(E') \cdot \frac{q^{\binom{m+1}{2}}}{\lambda |E|^m} \\ &\geq \mathcal{N}_{T_{2^{\ell+1}}^m}(E) \cdot \left(\frac{q^{\binom{m+1}{2}}}{\lambda |E|^m} \right)^{2^{\ell+1}-k} \\ &\geq \frac{|E|^{mk+1}}{q^{k\binom{m+1}{2}}} \frac{(1 - \frac{2}{\lambda})^{2^{\ell+1}m+1}}{\lambda^{2^{\ell+1}-k}} \left(1 - \frac{2^{\ell+2} \cdot 3(m+1)^2 q^{\frac{d-1}{2}+m}}{(1 - \frac{2}{\lambda})|E|} \right). \end{aligned}$$

Setting $\lambda = 4$ yields the desired result. □

Proof of Theorem 1.10. Let

$$E^* = \left\{ x \in E : f_{1,m}(x) \leq \lambda' \cdot \frac{|E|^m}{q^{\frac{m(m+1)}{2}}} \right\}.$$

Then,

$$|E \setminus E^*| \leq \frac{q^{\frac{m(m+1)}{2}}}{\lambda' |E|^m} \sum_{x \in E} f_{1,m}(x) \leq \frac{q^{\frac{m(m+1)}{2}}}{\lambda' |E|^m} \frac{2|E|^{m+1}}{q^{\frac{m(m+1)}{2}}} = \frac{2|E|}{\lambda'}.$$

Let

$$E'' = \left\{ x \in E : f_{1,m-1}(x) \leq \lambda' \cdot \frac{|E|^{m-1}}{q^{\frac{m(m-1)}{2}}} \right\}.$$

Then, similarly,

$$|E \setminus E''| \leq \frac{q^{\frac{m(m-1)}{2}}}{\lambda' \cdot |E|^{m-1}} \sum_{x \in E} f_{1,m-1}(x) \leq \frac{q^{\frac{m(m-1)}{2}}}{\lambda' |E|^{m-1}} \cdot \frac{2|E|^m}{q^{\frac{m(m-1)}{2}}} = \frac{2|E|}{\lambda'}.$$

Thus, if we let $E_0 = E^* \cap E''$, then $|E \setminus E_0| \leq 4|E|/\lambda'$. We count the number of degenerate embeddings of T_k^m in E_0 . We start with a point and multiply by $f_{1,m}$ to add a m -simplex. When we reach a repeated vertex we multiply by $f_{1,m-1}$. This yields an upper bound of

$$|E| \left(\lambda' \frac{|E|^m}{q^{\frac{m(m+1)}{2}}} \right)^{k-1} \cdot \lambda' \cdot \frac{|E|^{m-1}}{q^{\frac{m(m-1)}{2}}} = (\lambda')^k \frac{|E|^{mk}}{q^{k\frac{m(m+1)}{2}-m}}.$$

This is much less than

$$c_{m,k} \cdot \frac{|E|^{mk+1}}{q^{k\frac{m(m+1)}{2}}}$$

when $|E| \gg q^m$, which is always. □

5. TREES OF m -SIMPLICES

In this section, we prove the lower bounds on the number of (non-degenerate) embeddings of trees of simplices given in Theorem 1.12 and Theorem 1.13. Throughout, assume $m \geq 2$.

We prove the following proposition, where λ and μ are parameters we will define in the proof.

Proposition 5.1. *Let T_ℓ be a tree of m -simplices of size ℓ . Suppose $|E| \geq 12(m+1)^2 \mu^{1-\ell} q^{m+\frac{d-1}{2}}$. Then*

$$\mathcal{N}_{T_\ell}(E) \geq \frac{1}{2} \lambda^{1-\ell} (1 - 2\lambda^{-1})^{m\ell+1} \mu^{(\ell-1)(\ell m/2+1)} \frac{|E|^{m\ell+1}}{q^{\binom{m+1}{2} \cdot \ell}}.$$

If we take $\lambda = 4$, then $\mu = 2^{-(m+5)}$, we obtain Theorem 1.12. Then using the same arguments from Section 4 for counting degeneracies for chains of simplices, we obtain that the number of degenerate embeddings of T_ℓ is upper bounded by

$$|E| \left(\lambda' \frac{|E|^m}{q^{\binom{m+1}{2}}} \right)^{\ell-1} \cdot \lambda' \frac{|E|^{m-1}}{q^{\frac{m(m-1)}{2}}} = (\lambda')^\ell \frac{|E|^{m\ell}}{q^{\ell \binom{m+1}{2} - m}}.$$

Upon taking $\lambda' = 8$ and making the upper bound for degenerate embeddings half of the lower bound on $\mathcal{N}_{T_\ell}(E)$, we get Theorem 1.13.

Proof of Proposition 5.1. To count the number of embeddings of a tree of simplices, we simply induct on the size of the tree. The main estimate is the size of the following subset, E^* . Recall from the previous section that

$$E' = \left\{ x \in E : f_{1,m}(x) \leq \lambda \cdot \frac{|E|^m}{q^{\binom{m+1}{2}}} \right\},$$

where

$$f_{k,m}(x) = \#\{\text{embeddings of } T_k^m \text{ into } E \text{ with } x \text{ as the base point}\}.$$

Define

$$E^* = \{x \in E' : g_{1,m}(x) \geq \lambda^{-1} \frac{|E'|^m}{q^{\binom{m+1}{2}}}\},$$

where λ is to be determined later. Let

$$g_{k,m}(x) = \#\{\text{embeddings of } T_k^m \text{ into } E' \text{ with } x \text{ as the base point}\}.$$

Then,

$$\begin{aligned} \frac{|E'|^{m+1}}{2q^{\binom{m+1}{2}}} &\leq \mathcal{N}_{T_1^m}(E') = \sum_{x \in E'} g_{1,m}(x) \leq |E' \setminus E^*| \lambda^{-1} \frac{|E'|^m}{q^{\binom{m+1}{2}}} + |E^*| \lambda \frac{|E'|^m}{q^{\binom{m+1}{2}}}. \\ \implies |E^*| &\geq \left(\frac{1}{2} - \frac{1}{\lambda} \right) \frac{|E'|^{m+1}}{q^{\binom{m+1}{2}}} / \left[\lambda \frac{|E'|^m}{q^{\binom{m+1}{2}}} \right] = \frac{1}{2\lambda} (1 - 2\lambda^{-1})^{m+2} |E|. \end{aligned}$$

Let

$$\mu := \frac{(1 - 2\lambda^{-1})^{m+2}}{2\lambda}.$$

Then the above becomes $|E^*| \geq \mu|E|$. We prove that the number of embeddings of a tree containing ℓ connections into E is at least

$$\frac{1}{2}\lambda^{1-\ell}(1 - 2\lambda^{-1})^{m\ell+1}\mu^{(\ell-1)(\ell m/2+1)}\frac{|E|^{m\ell+1}}{q^{\binom{m+1}{2}\cdot\ell}}$$

when

$$|E| \geq 12(m+1)^2\mu^{1-\ell}q^{m+\frac{d-1}{2}}.$$

The base case $\ell = 1$ is done in Lemma 4.1:

$$\mathcal{N}_{T_1^m}(E) \geq \frac{|E|^{m+1}}{2q^{\binom{m+1}{2}}}.$$

Let T_ℓ denote a tree of size ℓ of m -simplices. Fix an ending m -simplex in T_ℓ and let $T_{\ell-1}$ denote the tree by replacing that simplex in T_ℓ with one vertex. Now by our assumption,

$$|E^*| \geq \mu|E| \geq 12(m+1)^2\mu^{1-(\ell-1)}q^{m+\frac{d-1}{2}},$$

we can apply the induction hypothesis on E^* to get

$$\begin{aligned} \mathcal{N}_{T_{\ell-1}}(E^*) &\geq \frac{1}{2}\lambda^{2-\ell}(1 - 2\lambda^{-1})^{m(\ell-1)+1}\mu^{(\ell-2)((\ell-1)m/2+1)}\frac{|E^*|^{m(\ell-1)+1}}{q^{\binom{m+1}{2}\cdot(\ell-1)}} \\ &\geq \frac{1}{2}\lambda^{2-\ell}(1 - 2\lambda^{-1})^{m(\ell-1)+1}\mu^{(\ell-1)(\ell m/2+1)}\frac{|E|^{m(\ell-1)+1}}{q^{\binom{m+1}{2}\cdot(\ell-1)}}. \end{aligned}$$

Then we have

$$\begin{aligned} \mathcal{N}_{T_\ell}(E) &\geq \mathcal{N}_{T_\ell}(E') \geq \mathcal{N}_{T_{\ell-1}}(E^*) \cdot \lambda^{-1}\frac{|E'|^m}{q^{\binom{m+1}{2}}} \\ &\geq \frac{1}{2}\lambda^{1-\ell}(1 - 2\lambda^{-1})^{m\ell+1}\mu^{(\ell-1)(\ell m/2+1)}\frac{|E|^{m\ell+1}}{q^{\binom{m+1}{2}\cdot\ell}}, \end{aligned}$$

as desired. □

5.1. Hölder Extension Method. This is another method to count the number of embeddings of trees of m -simplices, but the constants are not explicitly determined.

Base case: Suppose that T is a chain. We actually have that

$$\mathcal{N}_T(E') \geq \frac{|E|^{mk+1}}{q^{k\binom{m+1}{2}}} \frac{(1 - \frac{2}{\lambda})^{2km+1}}{\lambda^k} \left(1 - \frac{4k \cdot 3(m+1)^2 q^{\frac{d-1}{2}+m}}{(1 - \frac{2}{\lambda})|E|} \right) \geq \frac{|E|^{mk+1}}{q^{k\binom{m+1}{2}}} \frac{(1 - \frac{2}{\lambda})^{2km+1}}{2\lambda^k}$$

provided that $|E|$ is large enough. Let $C_T = \frac{(1 - \frac{2}{\lambda})^{2km+1}}{2\lambda^k}$ in this case.

At some node v , there are $r \geq 2$ chains attached to the rest of the tree. Let T' be the tree obtained by making all of these chains the same length (the maximum length of the chains). Let this length be ℓ (i.e., ℓ simplices). Let T_1 be the tree obtained by eliminating $r - 1$ of the chains from T' and let T_0 be the tree obtain by eliminating all of them. Then, T' is an r -Hölder extension of T_1 , so

$$\mathcal{N}_{T'} \geq \frac{\mathcal{N}_{T_1}^r}{\mathcal{N}_{T_0}^{r-1}}$$

by Lemma 1.3. Again define E' so that $f_{1,m}(x) \leq \lambda \frac{|E|^m}{q^{\binom{m+1}{2}}}$ for all $x \in E'$, then we have $|E \setminus E'| \leq \frac{2|E|}{\lambda}$. Let T_0 have v_0 nodes (so it contains $v_0 - 1$ simplices). We then have that

$$\mathcal{N}_{T_0} \leq \lambda^{v_0-1} \frac{|E|^{mv_0-m+1}}{q^{(v_0-1)\binom{m+1}{2}}}$$

and by the inductive hypothesis that

$$\mathcal{N}_{T_1}(E') \geq C_{T_1} |E| \left(\frac{|E|^m}{q^{\binom{m+1}{2}}} \right)^{\ell+v_0-1},$$

so that

$$\mathcal{N}_{T'}(E') \geq C_{T_1}^r \left[|E| \left(\frac{|E|^m}{q^{\binom{m+1}{2}}} \right)^{\ell+v_0-1} \right]^r \cdot \left(\lambda^{v_0-1} \frac{|E|^{mv_0-m+1}}{q^{(v_0-1)\binom{m+1}{2}}} \right)^{1-r} \quad (16)$$

$$\geq C_{T_1}^r \frac{|E|^{r(m(\ell+v_0-1)+r)}}{q^{r(\ell+v_0-1)\binom{m+1}{2}}} \lambda^{(v_0-1)(1-r)} \frac{|E|^{(1-r)(mv_0-m+1)}}{q^{(1-r)(v_0-1)\binom{m+1}{2}}} \quad (17)$$

$$= C_{T_1}^r \lambda^{(v_0-1)(1-r)} \frac{|E|^{m(r\ell+v_0-1)+1}}{q^{(r\ell+v_0-1)\binom{m+1}{2}}}. \quad (18)$$

In particular, we always have that $\mathcal{N}_{T'}(E') \geq C_{T'} \frac{|E|^{m(r\ell+v_0-1)+1}}{q^{(r\ell+v_0-1)\binom{m+1}{2}}}$ for some constant $C_{T'}$ depending on T' . We can then use the fact that $\mathcal{N}_{T'}(E') \leq \lambda \frac{|E|^m}{q^{\binom{m+1}{2}}} \mathcal{N}_{T'_{-1}}(E')$ where T'_{-1} is T' with one vertex removed. We use this fact repeatedly to obtain a lower bound on $\mathcal{N}_T(E')$. In particular, it is not hard to see that this must be done at most $(\ell - 1)(r - 1)$ times, so that we obtain the bound

$$\mathcal{N}_T(E') \geq C_{T_1}^r \lambda^{(v_0+\ell-2)(1-r)} \frac{|E|^{mV-m+1}}{q^{E\binom{m+1}{2}}}.$$

where V and E are the number of nodes and connections in T respectively. This completes the induction.

REFERENCES

- [1] Michael Bennett, Jeremy Chapman, David Covert, Derrick Hart, Alex Iosevich, and Jonathan Pakianathan. Long paths in the distance graph over large subsets of vector spaces over finite fields. *J. Korean Math. Soc.*, 53(1):115–126, 2016.
- [2] Jean Bourgain, Nets Katz, and Terence Tao. A sum-product estimate in finite fields, and applications. 2006.
- [3] Paul Erdős. On sets of distances of n points. *American Mathematical Monthly*, 53(5):248–250, 1946.
- [4] Larry Guth and Nets Hawk Katz. On the erdős distinct distances problem in the plane. *Annals of Mathematics*, pages 155–190, 2015.
- [5] A. Iosevich and M. Rudnev. Erdős distance problem in vector spaces over finite fields. *Trans. Amer. Math. Soc.*, 359(12):6127–6142, 2007.
- [6] Alex Iosevich, Gail Jardine, and Brian McDonald. Cycles of arbitrary length in distance graphs on \mathbb{F}_q^d . *Tr. Mat. Inst. Steklova*, 314:31–48, 2021. English version published in *Proc. Steklov Inst. Math.* **314** (2021), no. 1, 27–43.
- [7] Alex Iosevich and Hans Parshall. Embedding distance graphs in finite field vector spaces. *J. Korean Math. Soc.*, 56(6):1515–1528, 2019.
- [8] József Solymosi and Van H. Vu. Near optimal bounds for the Erdős distinct distances problem in high dimensions. *Combinatorica*, 28:113–125, 2008.
- [9] Le Anh Vinh. The solvability of norm, bilinear, and quadratic equations over finite fields via spectra of graphs. *Forum Math.*, 26(1):141–175, 2014.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Email address: paigeb@mit.edu

HARVARD UNIVERSITY

Email address: fxinyu@umich.edu

WILLIAMS COLLEGE

Email address: bth4@williams.edu

UNIVERSITY OF ROCHESTER

Email address: iosevich@gmail.com

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Email address: maxsun2022@gmail.com