

## LONG PATHS IN THE DISTANCE GRAPH OVER LARGE SUBSETS OF VECTOR SPACES OVER FINITE FIELDS

MICHAEL BENNETT, JEREMY CHAPMAN, DAVID COVERT, DERRICK HART,  
ALEX IOSEVICH, AND JONATHAN PAKIANATHAN

ABSTRACT. Let  $E \subset \mathbb{F}_q^d$ , the  $d$ -dimensional vector space over the finite field with  $q$  elements. Construct a graph, called the distance graph of  $E$ , by letting the vertices be the elements of  $E$  and connect a pair of vertices corresponding to vectors  $x, y \in E$  by an edge if  $\|x - y\| := (x_1 - y_1)^2 + \cdots + (x_d - y_d)^2 = 1$ . We shall prove that the non-overlapping chains of length  $k$ , with  $k$  in an appropriate range, are uniformly distributed in the sense that the number of these chains equals the statistically correct number,  $1 \cdot |E|^{k+1} q^{-k}$  plus a much smaller remainder.

### 1. Introduction

The classical Euclidean distance graph can be described as follows. Let the vertices of the graph be the points of the Euclidean plane. Connect two vertices by an edge if the Euclidean distance between the corresponding vectors is equal to 1. A very interesting open question, called the Hadwiger-Nelson problem, is to find the exact value of the chromatic number of this graph, the minimal number of colors required so that no two points of the same color are a distance 1 apart. It is known that the answer is at least four and at most seven.

In this paper, we continue the investigation of the corresponding distance graph in  $\mathbb{F}_q^d$ , the  $d$ -dimensional vector space over the finite field with  $q$  elements. Once again, the vertices of the graph are the points of  $\mathbb{F}_q^d$  and two vertices  $x, y$  are connected by an edge if  $\|x - y\| = 1$ , where

$$\|x\| = x_1^2 + x_2^2 + \cdots + x_d^2.$$

The distance graph of  $\mathbb{F}_q^d$  has been previously considered. For example, in [6] the authors show that the distance graph of  $\mathbb{F}_q^d$  is asymptotically Ramanujan for large  $q$ , meaning that the adjacency matrix of the distance graph of  $\mathbb{F}_q^d$  (which is a  $k$ -regular graph for some  $k$  depending on  $q$ ) satisfies the property that for every eigenvalue  $\lambda$  with  $|\lambda| \neq k$ , we have  $|\lambda| \leq 2\sqrt{k-1}$ . For more

---

Received August 29, 2014.

2010 *Mathematics Subject Classification.* 52C10, 11T23.

*Key words and phrases.* Erdos distance problem, finite fields, graph theory.

details on this result and for other results on the distance graph of  $\mathbb{F}_q^d$ , see [1], [3], [6] and [7].

In this paper we consider a much more complicated case when instead of taking all points in  $\mathbb{F}_q^d$  as the vertices of the distance graph, we merely consider points in a subset of  $\mathbb{F}_q^d$  of a sufficiently large size. To see that this formulation is meaningful, recall that the 5th listed author and Misha Rudnev proved in [4] that if  $E \subset \mathbb{F}_q^d$ ,  $d \geq 2$ , and  $t \neq 0$ , then

$$(1.1) \quad |\{(x, y) \in E \times E : \|x - y\| = t\}| = \frac{|E|^2}{q} + R(t),$$

where

$$|R(t)| \leq 2q^{\frac{d-1}{2}}|E|.$$

Here and throughout,  $|S|$  denotes the number of elements in a (finite) set  $S$ . In particular, if  $t = 1$  and  $|E| \geq 4q^{\frac{d+1}{2}}$ , then

$$|\{(x, y) \in E \times E : \|x - y\| = 1\}| \geq \frac{|E|^2}{2q}$$

or, in other words, the number of edges in the distance graph of  $E$  is at least  $\frac{|E|^2}{4q}$ .

Using this fact and elementary counting, one can show under this size assumption that the distance graph with vertices in  $E$  contains a chain of length at least  $C\frac{|E|}{q}$ . With a bit more work, we can even show that the number of chains of length  $k$  is comparable to a constant multiple of  $|E|^{k+1}q^{-k}$ . We are going to see below that something much stronger is true, namely that the  $k$ -chains are uniformly distributed in the sense that the number of  $k$ -chains equals the statistically correct number,  $1 \cdot |E|^{k+1}q^{-k}$  plus a much smaller remainder.

### 1.1. Main results

We know that if  $E$  is of sufficiently large cardinality, then the distance graph of  $E$  contains many edges, and an interesting question is whether the distance graph of  $E \subset \mathbb{F}_q^d$  must contain a long path. It turns out that the answer is affirmative. More precisely, we have the following result.

**Theorem 1.1.** *Let  $E \subset \mathbb{F}_q^d$ , where  $d \geq 2$  and  $|E| > \frac{2k}{\ln 2}q^{\frac{d+1}{2}}$ . Suppose that  $t_i \neq 0$ ,  $1 \leq i \leq k$ , and let  $\vec{t} = (t_1, \dots, t_k)$ . Define*

$$C_k(\vec{t}) = |\{(x^1, \dots, x^{k+1}) \in E \times \dots \times E : \|x^i - x^{i+1}\| = t_i, 1 \leq i \leq k\}|.$$

Then

$$C_k(\vec{t}) = \frac{|E|^{k+1}}{q^k} + \mathcal{D}_k(\vec{t}),$$

where

$$|\mathcal{D}_k(\vec{t})| \leq \frac{2k}{\ln 2}q^{\frac{d+1}{2}}\frac{|E|^k}{q^k}.$$

In particular, since  $|E| > \frac{2k}{\ln 2} q^{\frac{d+1}{2}}$ ,  $C_k(\vec{t}) > 0$ .

**Definition 1.2.** A path of length  $k$  in a simple graph  $G$  is a sequence of vertices  $v_1, \dots, v_{k+1} \in G$  such that each pair  $v_i, v_{i+1}$ ,  $1 \leq i \leq k$  is connected by an edge. We say that a path of length  $k$  is non-overlapping if all the  $v_j$ s in the definition are distinct.

**Corollary 1.3.** *With the notation above, suppose that  $|E| \geq \frac{4k}{\ln 2} q^{\frac{d+1}{2}}$ . Then the distance graph of  $E$  contains a non-overlapping chain of length  $k$  of every type.*

*Remark 1.4.* As we noted in the introduction above, this conclusion can be obtained by simpler methods, but Theorem 1.1 provides us with very precise information about the distribution of  $k$ -chains.

In addition to studying paths in a distance graph, we study the following closely related configuration. Given a set  $E \subset \mathbb{F}_q^d$ , fix a point  $x \in E$ , and count all of the vectors that are simultaneously some prescribed distance from  $x$  (we call such configurations  $k$ -stars). We show that if  $E$  is of sufficiently large cardinality, then we achieve the statistically correct number of  $k$ -stars. More precisely, we prove the following.

**Theorem 1.5.** *Let  $E \subset \mathbb{F}_q^d$ , and suppose  $t_i \neq 0$ ,  $1 \leq i \leq k$ . For  $\vec{t} = (t_1, \dots, t_k)$ , define*

$$\nu_k(\vec{t}) = |\{(x, x^1, \dots, x^k) \in E^{k+1} : \|x - x^i\| = t_i, x^i = x^j \iff i=j, 1 \leq i, j \leq k\}|.$$

*If  $|E| > 12q^{\frac{d+1}{2}}$ , then  $\nu_k(\vec{t}) > 0$  for any  $k < \frac{|E|}{12q^{\frac{d+1}{2}}}$ .*

*If  $|E| > 12q^{\frac{d+3}{2}}$ , then  $\nu_k(\vec{t}) > 0$  for any  $k < \frac{|E|}{12q}$ .*

## 1.2. Fourier analysis in $\mathbb{F}_q^d$

Let  $f : \mathbb{F}_q^d \rightarrow \mathbb{C}$  and  $\chi$  a nontrivial additive character of  $\mathbb{F}_q$ . Then,

$$\widehat{f}(m) = q^{-d} \sum_{x \in \mathbb{F}_q^d} \chi(-m \cdot x) f(x).$$

We have the following Plancherel and inversion formulas:

$$\sum_{m \in \mathbb{F}_q^d} |\widehat{f}(m)|^2 = q^{-d} \sum_{x \in \mathbb{F}_q^d} |f(x)|^2,$$

$$f(x) = \sum_{m \in \mathbb{F}_q^d} \widehat{f}(m) \chi(m \cdot x).$$

The proofs are straightforward. See, for example, [4] and the references contained therein.

**2. Proof of Theorem 1.1**

We shall need the following functional version of the arithmetic analog of Falconer’s result proved in [2].

**Theorem 2.1.** *Let  $f, g : \mathbb{F}_q^d \rightarrow \mathbb{R}^+$ . Let  $S_t = \{x \in \mathbb{F}_q^d : \|x\| = t\}$ , where  $\|x\| = x_1^2 + \dots + x_d^2$  and  $t \neq 0$ . Then*

$$\sum_{x, y \in \mathbb{F}_q^d} f(x)g(y)S_t(x - y) = \frac{1}{q} \cdot \|f\|_1 \cdot \|g\|_1 + D(f, g),$$

where

$$(2.1) \quad |D(f, g)| \leq 2q^{\frac{d-1}{2}} \|f\|_2 \|g\|_2.$$

To prove Theorem 2.1, we write

$$\begin{aligned} & \sum_{x, y \in \mathbb{F}_q^d} f(x)g(y)S_t(x - y) \\ &= \frac{1}{q} \sum_{x, y \in \mathbb{F}_q^d} f(x)g(x + y) \sum_{s \in \mathbb{F}_q} \chi(s(\|y\| - t)) \\ &= \frac{1}{q} \cdot \|f\|_1 \cdot \|g\|_1 + q^{d-1} \sum_{m \in \mathbb{F}_q^d} \widehat{f}(m)\widehat{g}(-m) \sum_{s \neq 0} \chi(-st) \sum_{y \in \mathbb{F}_q^d} \chi(s\|y\| - y \cdot m) \\ &= \frac{1}{q} \cdot \|f\|_1 \cdot \|g\|_1 + q^{d-1} \sum_{m \in \mathbb{F}_q^d} \widehat{f}(m)\widehat{g}(-m) \sum_{s \neq 0} \chi\left(-st - \frac{\|m\|}{4s}\right) \sum_{y \in \mathbb{F}_q^d} \chi\left(s\left\|y - \frac{m}{2s}\right\|\right). \end{aligned}$$

The rightmost sum is a Gauss sum, which shows that the sum is equal to

$$(2.2) \quad \frac{1}{q} \cdot \|f\|_1 \cdot \|g\|_1 + u_1 q^{\frac{3d-2}{2}} \sum_{m \in \mathbb{F}_q^d} \widehat{f}(m)\widehat{g}(-m) \sum_{s \neq 0} \left(\frac{s}{q}\right)^d \chi\left(-st - \frac{\|m\|}{4s}\right),$$

where  $u_1$  is a unimodular complex number and  $(\cdot)$  is the Legendre symbol (see [5] for more details regarding the Gauss sum).

We need the following fact about Kloosterman/Salié sums. See, for example, [5].

**Lemma 2.2.** *If  $a \neq 0$  or  $b \neq 0$ , then*

$$\left| \sum_{s \neq 0} \left(\frac{s}{q}\right)^d \chi(as + bs^{-1}) \right| \leq 2\sqrt{q}.$$

Applying the lemma shows that (2.2) is equal to

$$\frac{1}{q} \cdot \|f\|_1 \cdot \|g\|_1 + 2u_2 q^{\frac{3d-1}{2}} \sum_{m \in \mathbb{F}_q^d} |\widehat{f}(m)\widehat{g}(-m)|$$

where  $|u_2| \leq 1$ . We use Cauchy-Schwarz to see that the expression equals

$$\frac{1}{q} \cdot \|f\|_1 \cdot \|g\|_1 + 2u_2 q^{\frac{3d-1}{2}} \left( \sum_{m_1, m_2 \in \mathbb{F}_q^d} |\widehat{f}(m_1)|^2 |\widehat{g}(-m_2)|^2 \right)^{1/2}.$$

Applying Plancherel to the sum over  $m_1$  and the sum over  $m_2$  completes the proof of Theorem 2.1.

We shall prove Theorem 1.1 in the case  $t = (1, \dots, 1)$  for the sake of ease of exposition, but the reader can easily convince oneself that the argument extends to the general case. Let  $C_n = C_n(1, \dots, 1)$ . The basic mechanism of our proof is encapsulated in the following claim.

**Lemma 2.3.** *With the notation above, we have*

$$C_{2k+1} = q^{-1} C_k^2 + R_{2k+1},$$

$$C_{2k} = q^{-1} C_k C_{k-1} + R_{2k},$$

where

$$|R_{2k+1}| \leq 2q^{\frac{d-1}{2}} C_{2k}$$

and

$$|R_{2k}| \leq 2q^{\frac{d-1}{2}} \sqrt{C_{2k} C_{2k-2}}.$$

To prove the lemma, define

$$f_1(x) = (E * S)(x)E(x),$$

where  $S = S_1$ , and where

$$(f * g)(x) = \sum_{y \in \mathbb{F}_q^d} f(y)g(x-y).$$

Further, let

$$f_{k+1}(x) = (f_k * S)(x)E(x).$$

Unraveling the definition of  $C_{2k+1}$ , we see that it equals

$$\sum_{x, y} f_k(x) f_k(y) S(x-y),$$

which, by Theorem 2.1 is equal to

$$q^{-1} \left( \sum_x f_k(x) \right)^2 + R_{2k+1},$$

where

$$|R_{2k+1}| \leq 2q^{\frac{d-1}{2}} \|f_k\|_2^2.$$

Similarly,

$$C_{2k} = \sum_{x, y} f_k(x) f_{k-1}(y) S(x-y)$$

$$= q^{-1} \sum_x f_k(x) \cdot \sum_y f_{k-1}(y) + R_{2k},$$

where

$$|R_{2k}| \leq 2q^{\frac{d-1}{2}} \|f_k\|_2 \|f_{k-1}\|_2.$$

By a direct calculation,

$$\|f_k\|_1 = C_k$$

and

$$\|f_k\|_2^2 = C_{2k}.$$

Putting everything together we recover the conclusions of Lemma 2.3.

**Lemma 2.4.** *With the notation above, we have*

$$C_n \leq |E| \left( \frac{|E| + 2q^{\frac{d+1}{2}}}{q} \right)^n.$$

To prove Lemma 2.4, we first put  $X = \frac{|E| + 2q^{\frac{d+1}{2}}}{q}$ . We know that  $C_1 \leq |E|X$  by (1.1). Now we induct on the chain length. Suppose it holds for  $C_k$  when  $k < n$ . From the previous lemma,

$$\begin{aligned} C_{2k+1} &\leq \frac{C_k^2}{q} + 2q^{\frac{d-1}{2}} C_{2k} \\ &\leq \frac{(|E|X^k)^2}{q} + 2q^{\frac{d-1}{2}} |E|X^{2k} = |E|X^{2k+1}. \end{aligned}$$

Recall we have shown in the previous lemma that  $C_{2k} \leq q^{-1}C_k C_{k-1} + 2q^{\frac{d-1}{2}} \sqrt{C_{2k} C_{2k-2}}$ . Completing the square, we may then write

$$(\sqrt{C_{2k}} - q^{\frac{d-1}{2}} \sqrt{C_{2k-2}})^2 \leq q^{-1}C_k C_{k-1} + q^{d-1}C_{2k-2}.$$

Taking square roots yields  $\sqrt{C_{2k}} \leq q^{\frac{d-1}{2}} \sqrt{C_{2k-2}} + \sqrt{q^{-1}C_k C_{k-1} + q^{d-1}C_{2k-2}}$ . Finally squaring both sides yields:

$$\begin{aligned} C_{2k} &\leq \frac{C_k C_{k-1}}{q} + 2q^{d-1} C_{2k-2} + 2\sqrt{q^{2d-2} C_{2k-2}^2 + q^{d-2} C_{2k-2} C_k C_{k-1}} \\ &\leq \frac{|E|^2 X^{2k-1}}{q} + 2q^{d-1} |E| X^{2k-2} + 2\sqrt{q^{2d-2} |E|^2 X^{4k-4} + q^{d-2} |E|^3 X^{4k-3}} \\ &= \frac{|E|^2 X^{2k-1}}{q} + 2q^{d-1} |E| X^{2k-2} \left( 1 + \sqrt{1 + \frac{|E|X}{q^d}} \right) \\ &= \frac{|E|^2 X^{2k-1}}{q} + 2q^{d-1} |E| X^{2k-2} \left( 1 + \sqrt{\frac{q^{d+1} + |E|^2 + 2|E|q^{\frac{d+1}{2}}}{q^{d+1}}} \right) \\ &= \frac{|E|^2 X^{2k-1}}{q} + 2q^{d-1} |E| X^{2k-2} \left( 2 + \frac{|E|}{q^{\frac{d+1}{2}}} \right) \end{aligned}$$

$$= \frac{|E|^2 X^{2k-1}}{q} + 2q^{\frac{d-1}{2}} |E| X^{2k-1} = |E| X^{2k}.$$

This completes the proof of Lemma 2.4.

We are now ready to complete the proof of Theorem 1.1. By the previous lemma, we have

$$\begin{aligned} C_n &\leq |E| \left( \frac{|E| + 2q^{\frac{d+1}{2}}}{q} \right)^n \\ &= \frac{|E|}{q^n} \sum_{i=0}^n \binom{n}{i} |E|^{n-i} \left( 2q^{\frac{d+1}{2}} \right)^i \\ &= \frac{|E|^{n+1}}{q^n} + \frac{2q^{\frac{d+1}{2}} |E|}{q^n} \sum_{i=1}^n \binom{n}{i} |E|^{n-i} \left( 2q^{\frac{d+1}{2}} \right)^{i-1}. \end{aligned}$$

By assumption,  $q^{\frac{d+1}{2}} \leq \frac{|E| \ln 2}{2n}$ , so

$$\begin{aligned} C_n &\leq \frac{|E|^{n+1}}{q^n} + \frac{2q^{\frac{d+1}{2}} |E|^n}{q^n} \sum_{i=1}^n \binom{n}{i} \left( \frac{\ln 2}{n} \right)^{i-1} \\ &= \frac{|E|^{n+1}}{q^n} + \frac{2n}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^n}{q^n} \left( \left( 1 + \frac{\ln 2}{n} \right)^n - 1 \right) \\ &\leq \frac{|E|^{n+1}}{q^n} + \frac{2n}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^n}{q^n} (e^{\ln 2} - 1) = \frac{|E|^{n+1}}{q^n} + \frac{2n}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^n}{q^n}. \end{aligned}$$

For the lower bound, we use induction. We will first establish the inequality for the odd indices. Note that (1.1) yields

$$C_1 \geq \frac{|E|^2}{q} - 2q^{\frac{d+1}{2}} \frac{|E|}{q} \geq \frac{|E|^2}{q} - \frac{2}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|}{q}.$$

Next we deal with  $C_3$ . Lemma 2.3 implies that

$$(2.3) \quad \left| C_{2n+1} - \frac{C_n^2}{q} \right| \leq 2q^{\frac{d-1}{2}} C_{2n}.$$

Hence,

$$\begin{aligned} C_3 &\geq \frac{1}{q} C_1^2 - 2q^{\frac{d-1}{2}} C_2 \\ &= \frac{1}{q} \left( \frac{|E|^2}{q} - 2q^{\frac{d-1}{2}} |E| \right)^2 - 2q^{\frac{d-1}{2}} C_2 \\ &\geq \frac{|E|^4}{q^3} - \frac{6}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^3}{q^3}, \end{aligned}$$

whenever

$$C_2 \leq \left( \frac{3}{\ln 2} - 2 \right) \frac{|E|^3}{q^2} + 2q^{\frac{d-3}{2}} |E|^2.$$

Lemma 2.4 shows that we already have the inequality

$$C_2 \leq |E| \left( \frac{|E| + q^{\frac{d+1}{2}}}{q} \right)^2,$$

so we need only demonstrate that

$$|E| \left( \frac{|E| + q^{\frac{d+1}{2}}}{q} \right)^2 \leq \left( \frac{3}{\ln 2} - 2 \right) \frac{|E|^3}{q^2} + 2q^{\frac{d-3}{2}} |E|^2,$$

which holds for

$$|E| \geq q^{\frac{d+1}{2}}.$$

Since  $|E|$  is assumed to be in the range  $|E| \geq \frac{6}{\ln 2} q^{\frac{d+1}{2}}$ , we have established the lower bound for  $C_3$ .

For  $n \geq 2$ , suppose that for some  $k < 2n + 1$ , we have

$$C_k \geq \frac{|E|^{k+1}}{q^k} - \frac{2k}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^k}{q^k}.$$

Applying (2.3) yields

$$\left| C_{2n+1} - \frac{C_n^2}{q} \right| \leq 2q^{\frac{d-1}{2}} C_{2n}.$$

This implies

$$\begin{aligned} C_{2n+1} &\geq \frac{\left( \frac{|E|^{n+1}}{q^n} - \frac{2n}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^n}{q^n} \right)^2}{q} - 2q^{\frac{d-1}{2}} \left( \frac{|E|^{2n+1}}{q^{2n}} + \frac{4n}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^{2n}}{q^{2n}} \right) \\ &= \frac{|E|^{2n+2}}{q^{2n+1}} - q^{\frac{d+1}{2}} \frac{|E|^{2n+1}}{q^{2n+1}} \left( \frac{4n}{\ln 2} - q^{\frac{d+1}{2}} \frac{4n^2}{|E|(\ln 2)^2} + 2 + \frac{8n}{|E| \ln 2} q^{\frac{d+1}{2}} \right). \end{aligned}$$

Observe that

$$\frac{8n}{|E| \ln 2} q^{\frac{d+1}{2}} - q^{\frac{d+1}{2}} \frac{4n^2}{|E|(\ln 2)^2} < 0$$

for  $n \geq 2$  so that

$$C_{2n+1} \geq \frac{|E|^{2n+2}}{q^{2n+1}} - \frac{4n+2}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^{2n+1}}{q^{2n+1}}.$$

Since we have dealt separately with the case  $n = 1$ , it follows that the above inequality holds for all  $n \geq 1$ . A nearly identical argument gives us the analog for  $C_{2n}$ .



### 3. Proof of Corollary 1.3

In analogy with  $f_k$ , let  $g_k(x)$  be the number of non-overlapping paths of length  $k$  in  $E$  beginning at  $x$ . Then the total number of such paths is

$$G_k = \sum_x g_k(x).$$

Given a non-overlapping  $n$ -path  $(v_0, v_1, v_2, \dots, v_n)$ , we must be able to find at least  $-n + \sum_y E(y)S(v_0 - y)$  choices of  $x$  so that the path  $(x, v_0, v_1, \dots, v_n)$  is also non-overlapping. There may be some values of  $i$  for which  $S(v_i - v_0) = 1$ , which is why we must subtract  $n$ . Otherwise we may be including some overlapping paths in the count.

We then have the following recursive inequality:

$$G_{n+1} \geq \sum_x g_n(x) \left( -n + \sum_y E(y)S(x - y) \right) = -nG_n + \sum_{x,y} g_n(x)E(y)S(x-y).$$

By Theorem 2.1, we see that

$$\sum_{x,y} g_n(x)E(y)S(x-y) \geq \frac{|G_n||E|}{q} - 2q^{\frac{d-1}{2}}|E|^{1/2} \left( \sum_x (g_n(x))^2 \right)^{1/2}.$$

We also know that

$$\sum_x (g_n(x))^2 \leq \sum_x (f_n(x))^2 = C_{2n}.$$

By Theorem 1.1, we have

$$C_{2n} \leq \frac{|E|^{2n+1}}{q^{2n}} + \frac{4n}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^{2n}}{q^{2n}},$$

and by assumption,  $|E| \geq \frac{4n}{\ln 2} q^{\frac{d+1}{2}}$ . Thus

$$C_{2n} \leq 2 \frac{|E|^{2n+1}}{q^{2n}}.$$

Moreover

$$G_n \leq C_n \leq \frac{|E|^{n+1}}{q^n} + \frac{2n}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^n}{q^n}.$$

We will induct on the chain length and assume that

$$G_k \geq \frac{|E|^{k+1}}{q^k} - \frac{4k}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^k}{q^k}.$$

Putting everything together, we get

$$\begin{aligned} G_{k+1} &\geq \frac{G_k|E|}{q} - kG_k - 2q^{\frac{d-1}{2}} \sqrt{2 \frac{|E|^{2k+2}}{q^{2k}}} \\ &\geq \frac{|E|^{k+2}}{q^{k+1}} - \frac{4k}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^{k+1}}{q^{k+1}} - k \frac{|E|^{k+1}}{q^k} - \frac{4k^2}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^k}{q^k} - 2\sqrt{2} q^{\frac{d+1}{2}} \frac{|E|^{k+1}}{q^{k+1}} \end{aligned}$$

$$= \frac{|E|^{k+2}}{q^{k+1}} - q^{\frac{d+1}{2}} \frac{|E|^{k+1}}{q^{k+1}} \left( \frac{4k}{\ln 2} + \frac{k}{q^{\frac{d-1}{2}}} + \frac{4k^2 q}{|E| \ln 2} + 2\sqrt{2} \right).$$

By assumption,  $|E| \geq \frac{4k}{\ln 2} q^{\frac{d+1}{2}}$ , so

$$G_{k+1} \geq \frac{|E|^{k+2}}{q^{k+1}} - q^{\frac{d+1}{2}} \frac{|E|^{k+1}}{q^{k+1}} \left( \frac{4k}{\ln 2} + \frac{2k}{q^{\frac{d-1}{2}}} + 2\sqrt{2} \right).$$

Lastly, since  $|E| \leq q^d$ , we have  $k \leq \frac{\ln 2}{4} q^{\frac{d-1}{2}}$ , giving us

$$\begin{aligned} G_{k+1} &\geq \frac{|E|^{k+2}}{q^{k+1}} - q^{\frac{d+1}{2}} \frac{|E|^{k+1}}{q^{k+1}} \left( \frac{4k}{\ln 2} + \frac{\ln 2}{2} + 2\sqrt{2} \right) \\ &\geq \frac{|E|^{k+2}}{q^{k+1}} - \frac{4k+4}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^{k+1}}{q^{k+1}}. \end{aligned}$$

#### 4. Proof of Theorem 1.5

Suppose  $\vec{t} = (t_1, \dots, t_k)$  is  $k$ -dimensional with all nonzero indices. Notice that a  $k$ -star with edge-lengths given by  $\vec{t}$  can also be described by any permutation of the indices of  $\vec{t}$ . Hence we assume without loss of generality that for some value  $l \leq \min\{k, q-1\}$  the vector  $\vec{t}$  satisfies

$$t_1 = t_2 = \dots = t_{i_1} = s_1; t_{i_1+1} = \dots = t_{i_2} = s_2; \dots; t_{i_{l-1}+1} = \dots = t_{i_l} = s_l$$

for some pairwise distinct nonzero elements  $s_1, \dots, s_l$ , and where  $i_l = k$ . We define

$$h(x) = \#\{y \in E : \|x - y\| = 1\},$$

where the choice of 1 as the length is arbitrary. We may equivalently write

$$h(x) = \sum_y E(y)S(x-y).$$

We begin by estimating

$$H_n = \#\{x \in E : h(x) \geq n\}.$$

That is,  $H_n$  is the number of points of  $E$  from which an  $n$ -star with all edge lengths 1 can be made. We use Cauchy-Schwarz to get

$$\begin{aligned} \left( \sum_{x, h(x) \geq n} E(x)h(x) \right)^2 &\leq \left( \sum_{x, h(x) \geq n} E(x)(h(x))^2 \right) \left( \sum_{x, h(x) \geq n} E(x) \right) \\ &\leq H_n \sum_x E(x)(h(x))^2. \end{aligned}$$

Notice that the sum on the right-hand side is

$$\sum_{x, y, z} E(x)E(y)E(z)S(x-y)S(x-z).$$

By Theorem 1.1, this is less than  $\frac{|E|^3 + 6q^{\frac{d+1}{2}}|E|^2}{q^2}$ .

For the left-hand side, we notice first that

$$\sum_{x, h(x) \geq n} E(x)h(x) \geq \sum_x E(x)(h(x) - n) = -n|E| + \sum_x E(x)h(x).$$

Using Theorem 1.1 again, we see that the left side is bounded below by

$$\left( \frac{|E|^2 - 2q^{\frac{d+1}{2}}|E|}{q} - n|E| \right)^2.$$

Solving for  $H_n$  gives

$$H_n \geq |E| - 10q^{\frac{d+1}{2}} - 2qn.$$

We now return to our vector  $\vec{t}$ . We now know that there are at least  $H_{i_j - i_{j-1}}$  points in  $E$  with  $(i_j - i_{j-1})$ -stars having all edge lengths  $s_j$ . By pigeonholing, there must be at least  $|E| - \sum_{j=1}^l (|E| - H_{i_j - i_{j-1}})$  points in  $E$  from which emanate  $k$ -stars given by  $\vec{t}$ . This quantity is larger than zero simply when

$$|E| > \sum_{j=1}^l (|E| - H_{i_j - i_{j-1}}).$$

Recall that  $l \leq \min\{k, q-1\}$ ,  $i_0 = 0$ , and  $i_l = k$ . Since

$$\begin{aligned} \sum_{j=1}^l (|E| - H_{i_j - i_{j-1}}) &< \sum_{j=1}^l \left( 10q^{\frac{d+1}{2}} + 2q(i_j - i_{j-1}) \right) \\ &= 10lq^{\frac{d+1}{2}} + 2q(i_l - i_0) \leq \min\{k, q-1\}10q^{\frac{d+1}{2}} + 2qk, \end{aligned}$$

it suffices to have  $|E| > \min\{k, q-1\}10q^{\frac{d+1}{2}} + 2qk$ . The result follows.

## References

- [1] J. Chapman, M. B. Erdogan, D. Hart, A. Iosevich, and D. Koh, *Pinned distance sets,  $k$ -simplices, Wolff's exponent in finite fields and sum-product estimates*, Math Z. **271** (2012), no. 1-2, 63–93.
- [2] D. Covert, D. Hart, A. Iosevich, D. Koh, and M. Rudnev, *Generalized incidence theorems, homogeneous forms and sum-product estimates in finite fields*, European J. Combin. **31** (2010), no. 1, 306–319.
- [3] D. Covert, D. Hart, A. Iosevich, S. Senger, and I. Uriarte-Tuero, *A Furstenberg-Katznelson-Weiss type-theorem on  $(d+1)$ -point configurations in sets of positive density in finite field geometries*, Discrete Math. **311** (2011), no. 6, 423–430.
- [4] A. Iosevich and M. Rudnev *Erdős distance problem in vector spaces over finite fields*, Transactions of the AMS, 2007.
- [5] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, Colloquium Publications **53** (2004).
- [6] A. Medrano, P. Myers, H. Stark, and A. Terras, *Finite analogues of Euclidean space*, J. Comput. Appl. Math. **68** (1996), no. 1-2, 221–238.
- [7] L. Vinh, *The Szemerédi-Trotter type theorem and the sum-product estimate in finite fields*, European J. Combin. **32** (2011), no. 8, 1177–1181.

MICHAEL BENNETT  
SCHOOL OF MATHEMATICAL SCIENCES  
ROCHESTER INSTITUTE OF TECHNOLOGY  
ROCHESTER, NY 14623, USA  
*E-mail address:* `mbennet4@z.rochester.edu`

JEREMY CHAPMAN  
DEPARTMENT OF MATHEMATICS  
LYON COLLEGE  
BATESVILLE, AR 72501, USA  
*E-mail address:* `jeremy.chapman@lyon.edu`

DAVID COVERT  
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE  
UNIVERSITY OF MISSOURI-SAINTE LOUIS  
ST. LOUIS, MO 63121, USA  
*E-mail address:* `covert@umsl.edu`

DERRICK HART  
DEPARTMENT OF MATHEMATICS  
ROCKHURST UNIVERSITY  
KANSAS CITY, MO 64110, USA  
*E-mail address:* `Derrick.Hart@rockhurst.edu`

ALEX IOSEVICH  
DEPARTMENT OF MATHEMATICS  
UNIVERSITY OF ROCHESTER  
ROCHESTER, NY 14642, USA  
*E-mail address:* `aiosevic@z.rochester.edu`

JONATHAN PAKIANATHAN  
DEPARTMENT OF MATHEMATICS  
UNIVERSITY OF ROCHESTER  
ROCHESTER, NY 14642, USA  
*E-mail address:* `jonathan.pakianathan@rochester.edu`