EMBEDDING DISTANCE GRAPHS IN FINITE FIELD VECTOR SPACES

Alex Iosevich and Hans Parshall

ABSTRACT. We show that large subsets of vector spaces over finite fields determine certain point configurations with prescribed distance structure. More specifically, we consider the complete graph with vertices as the points of $A \subseteq \mathbf{F}_q^d$ and edges assigned the algebraic distance between pairs of vertices. We prove nontrivial results on locating specified subgraphs of maximum vertex degree at most t in dimensions $d \geq 2t$.

1. Introduction

The famous Erdős and Falconer distance problems aim to quantify the extent to which large sets must determine many distances. A near optimal result on the planar Erdős distinct distance problem was established by Guth and Katz [10], who showed that every set of N points in $\mathbf{R}^2$ determines at least a constant multiple of $\frac{N}{\log(N)}$ distinct distances. However, even in the Euclidean plane, the Falconer distance problem continues to present a gap in our understanding of distance sets. Falconer [7] proved that when the Hausdorff dimension of compact $A \subseteq \mathbf{R}^d$ is large, specifically $\dim(A) > \frac{d+1}{2}$, the distance set $\{|x - y| : x, y \in A\} \subseteq \mathbf{R}$ must have positive Lebesgue measure. Wolff [19] in dimension $d = 2$ and Erdoğan [6] in dimensions $d \geq 3$ established the same conclusion with the weaker hypothesis of $\dim(A) > \frac{d}{2} + \frac{1}{3}$. Falconer conjectures that this can be relaxed further to $\dim(A) > \frac{d}{2}$, which would be optimal. Recent significant progress in the plane by Guth, the first listed author, Ou, and Wang [9] improves Wolff's result to allow for $\dim(A) > 5/4$.

More generally, many authors have considered the related problem of locating point configurations with prescribed metric structure within subsets $A \subseteq \mathbf{R}^d$ of large Hausdorff dimension. For instance, one result in this direction of Bennett, the first listed author and Taylor [3, Theorem 1.7] implies:

Theorem 1 ([3]). *For $d \geq 2$, let $A \subseteq \mathbf{R}^d$ be compact with $\dim(A) > \frac{d+1}{2}$. For any $k \in \mathbf{N}$, there exists an open interval $(a, b) \subset \mathbf{R}$ where for every $\lambda \in (a, b)$, there exist distinct $x_0, \ldots, x_k \in A$ with $|x_{j+1} - x_j| = \lambda$ for $0 \leq j < k$.*

One can interpret this statement in terms of equilateral Euclidean distance graphs $\mathcal{G}_\lambda(A)$, defined for $A \subseteq \mathbf{R}^d$ and $\lambda \in \mathbf{R}$, where we consider the points of A as vertices and connect every pair of points $x, y \in A$ with an edge exactly when $|x - y| = \lambda$. With this notion, Theorem 1 states that for every compact $A \subseteq \mathbf{R}^d$ with $\dim(A) > \frac{d+1}{2}$ and for arbitrarily large $k \in \mathbf{N}$, there exists an open interval of distances $\lambda \in \mathbf{R}$ for which $\mathcal{G}_\lambda(A)$ contains a path of length k. Under the stronger hypothesis $\dim(A) > \frac{d+3}{2}$, Greenleaf, the first listed author and Pramanik [8] demonstrate an open interval of distances $\lambda \in \mathbf{R}$ for which $\mathcal{G}_\lambda(A)$ contains cycles of length $2k$. Positive results for cycles of length 3 within $\mathcal{G}_\lambda(A)$ in $\mathbf{R}^4$, corresponding to the vertices of equilateral triangles, were recently obtained [11], but there seems to be a gap in the literature for the general case of arbitrarily long odd length cycles with equilateral edge lengths. Additional results on locating more general distance graph structure within sets of large Hausdorff dimension have been obtained by Chatzikonstantinou, the first listed author, Mkrtchyan and Pakianathan [5].

Our goal here is to illustrate a related approach for locating distance graph structure within large subsets of $\mathbf{F}_q^d$, the vector space of dimension d over a finite field with odd characteristic. For $v, w \in \mathbf{F}_q^d$, we consider their usual dot product

$$v \cdot w := \sum_{j=1}^{d} v_j w_j$$

and refer to $|v|^2 := v \cdot v$ as the *length* of v and $|v - w|^2$ as the *distance* between v and w. While this notion of distance is certainly not a metric, this setup provides a useful model setting for the Falconer distance problem and its variants. For instance, the first listed author and Rudnev [12] demonstrated that every $A \subseteq \mathbf{F}_q^d$ with $|A| > 2q^{\frac{d+1}{2}}$ determines all distances, in the sense that $\{|x - y|^2 : x, y \in A\} = \mathbf{F}_q$. An analogue to Wolff's result in this setting was first achieved by Chapman, Erdoğan, Hart, the first listed author and Koh [4], who showed that when $q \equiv 3 \mod 4$, every $A \subseteq \mathbf{F}_q^2$ with $|A| \geq q^{4/3}$ determines at least a constant proportion of q distances. This was shown to further hold when $q \equiv 1 \mod 4$ by Bennett, Hart, the first listed author, Pakianathan and Rudnev [2] through a group action method. Analogous problems for locating point configurations with prescribed distance structure have been studied by several authors. Many of these results show that large sets determine isometric copies of a positive proportion of a class of point configurations (for instance, [2, 4]), but here we will show that large sets determine isometric copies of *all* point configurations of a fixed class (as was done in [15, 16]).

We find it convenient to organize distance structure within subsets $A \subseteq \mathbf{F}_q^d$ in the language of graph theory. We call a graph $\mathcal{G} = (V, E)$ a *distance graph*

when to each edge $e \in E$ there is some associated nonzero length $\lambda_e \in \mathbf{F}_q^*$. We call $X \subseteq \mathbf{F}_q^d$ an *isometric copy* of $\mathcal{G}$ when there exists a distance preserving bijection $\varphi : V \to X$ where for every $v, w \in V$ connected by an edge $e \in E$, $|\varphi(v) - \varphi(w)|^2 = \lambda_e$. Our main result is the following.

Theorem 2. *Let $n, t \in \mathbf{N}$, and let $A \subseteq \mathbf{F}_q^d$ with $|A| \geq 12n^2 q^{\frac{d-1}{2}+t}$. Then A contains an isometric copy of every distance graph with n vertices and maximum vertex degree t.*

The main analytic input into Theorem 2 is Theorem 4 below. An interested reader can check that the "distance" function $|\cdot|^2$ can be replaced by any non-degenerate quadratic form using the results of [12] and [2]. We stick to the "Euclidean distance" in this context for the sake of notational clarity.

Theorem 2 allows us to locate arbitrarily large configurations of maximum degree t within large subsets of $\mathbf{F}_q^{2t}$ provided q is taken sufficiently large with respect to the desired number of vertices n. For instance, we recover nontrivial results for locating all cycle graphs of length n in dimensions $d \geq 4$, which is new. Previous results of this type typically require the dimension d of $\mathbf{F}_q^d$ to depend on the number of sought after vertices. Vinh [16] has proven that for a constant C depending only upon n, if $A \subseteq \mathbf{F}_q^d$ with $|A| > Cq^{\frac{d-3}{2}+n}$, then A contains an isometric copy of every complete distance graph on n vertices. While our methods differ, Theorem 2 recovers Vinh's exponent of $\frac{d-3}{2} + n$ for complete graphs on n vertices. Vinh's result yields a nontrivial exponent for every distance graph by considering an embedding into a complete graph. Theorem 2 yields an improvement over this exponent whenever the maximum degree t is strictly less than $n-1$. It is also worth comparing with another result of Vinh [17, Theorem 2.25], which immediately implies a version of Theorem 2 accounting for *almost all* distance graphs, i.e., a positive proportion approaching 1 as $q \to \infty$, with the same exponent. Our result improves this to recovering *all* distance graphs.

There are two cases where Theorem 2 does not even meet the quantitative information provided by previous similar results. Bennett et al. [1] showed that if $A \subseteq \mathbf{F}_q^d$ with $|A| > 4(n+1)q^{\frac{d+1}{2}}$, then A contains an isometric copy of every path with n vertices. This is nontrivial in dimensions $d \geq 2$, while applying Theorem 2 for paths is unfortunately only nontrivial for $d \geq 4$. A more careful application of our methods yields a result for paths in dimensions $d \geq 3$, but we do not pursue this. Additionally, the second listed author [15] has proven results for complete graphs on n vertices that are nontrivial in dimensions $d \geq n$ provided one is willing to impose technical conditions on the edge distances; we instead opt for a result uniform in nonzero edge distances.

Our strategy is to proceed by an edge deletion induction. This relies on a functional distance counting theorem, Theorem 4, which previously appeared in the work of Bennett et al. [1]. As this is our main tool, we record its proof after setting notation and recalling some facts about exponential sums over

finite fields. Before proving a more quantitative form of Theorem 2 in the final section, we demonstrate our method with the small case of equilateral triangles; this helps to highlight why our methods are constrained by the maximum vertex degree.

2. Preliminaries

Let $\chi : \mathbf{F}_q \to \mathbf{C}$ denote the canonical additive character. More precisely, if p is the characteristic of $\mathbf{F}_q$ and $\mathrm{Tr} : \mathbf{F}_q \to \mathbf{F}_p$ is the usual trace function, we define for $a \in \mathbf{F}_q$

$$\chi(a) := \exp\Big(\frac{2\pi i\, \mathrm{Tr}(a)}{p}\Big).$$

We find it convenient to use averaging notation for $f : \mathbf{F}_q^d \to \mathbf{C}$ of

$$\mathbf{E}_x f(x) := q^{-d} \sum_{x \in \mathbf{F}_q^d} f(x),$$

and we will condense multiple averages $\mathbf{E}_{x_1}\mathbf{E}_{x_2}\cdots\mathbf{E}_{x_n}$ as $\mathbf{E}_{x_1,x_2,\dots,x_n}$. With this notation, the familiar orthogonality relation for χ and $y \in \mathbf{F}_q^d$ takes the form

$$\mathbf{E}_x \chi(x \cdot y) = \begin{cases} 1 & \text{if } y = 0, \\ 0 & \text{if } y \in \mathbf{F}_q^d \setminus \{0\}. \end{cases}$$

For $\lambda \in \mathbf{F}_q$, $x \in \mathbf{F}_q^d$ we define the normalized indicator function

$$\sigma_\lambda(x) := \begin{cases} q & \text{if } |x|^2 = \lambda, \\ 0 & \text{otherwise.} \end{cases}$$

This plays the role of the surface measure of the sphere of radius λ. For convenience we perform the standard calculation that shows σ_λ is essentially L^1-normalized, with error tending to zero as $q \to \infty$ provided $d \geq 2$.

Lemma 3. *For any $\lambda \in \mathbf{F}_q^*$,*

(1) $$|\mathbf{E}_x \sigma_\lambda(x) - 1| \leq q^{-\frac{d-1}{2}}.$$

Proof. By orthogonality, we have

$$\mathbf{E}_x \sigma_\lambda(x) = \mathbf{E}_x \sum_{\ell \in \mathbf{F}_q} \chi(\ell|x|^2 - \ell\lambda)$$

$$= \sum_{\ell \in \mathbf{F}_q} \chi(-\ell\lambda)\mathbf{E}_x\chi(\ell|x|^2)$$

$$= 1 + \sum_{\ell \in \mathbf{F}_q^*} \chi(-\ell\lambda)\mathbf{E}_x\chi(\ell|x|^2).$$

Introducing the quadratic character $\eta : \mathbf{F}_q^* \to \{-1, 1\}$ defined for $a \in \mathbf{F}_q^*$ by

$$\eta(a) := \begin{cases} 1 & \text{if } a \text{ is a square in } \mathbf{F}_q^*, \\ -1 & \text{otherwise,} \end{cases}$$

and the Gaussian sum

$$G(\chi, \eta) := \sum_{a \in \mathbf{F}_q^*} \chi(a)\eta(a),$$

it follows from [14, Theorem 5.33] that we have

$$\mathbf{E}_x \chi(\ell|x|^2) = q^{-d} G(\chi, \eta)^d \eta(\ell)^d.$$

Together with our computation above,

$$\left| \mathbf{E}_x \sigma_\lambda(x) - 1 \right| \le q^{-d} |G(\chi, \eta)|^d \left| \sum_{\ell \in \mathbf{F}_q^*} \chi(-\ell\lambda)\eta(\ell)^d \right|.$$

As $\lambda \in \mathbf{F}_q^*$, we can reindex in ℓ for

$$\left| \mathbf{E}_x \sigma_\lambda(x) - 1 \right| \le q^{-d} |G(\chi, \eta)|^d \left| \sum_{\ell \in \mathbf{F}_q^*} \chi(\ell)\eta(\ell)^d \right|.$$

It is not hard [14, Theorem 5.15] to show $|G(\chi, \eta)| = q^{1/2}$. When d is even, the sum in ℓ amounts to -1 (by orthogonality of χ), and when d is odd, the sum in ℓ contributes another Gaussian sum. That is, we could slightly improve (1) in even dimensions; in any case we can conclude (1). $\square$

3. A functional distance theorem

The first listed author and Rudnev [12] showed not only that large sets determine all distances, but that distances within large subsets equidistribute. We will use a functional version of their result, which appears in the work of Bennett et al. [1, Theorem 2.1]. More precisely, with the L^2 norm of $f : \mathbf{F}_q^d \to \mathbf{C}$ normalized as

$$\|f\|_2 := \left(\mathbf{E}_x |f(x)|^2 \right)^{1/2},$$

our main distance counting tool is the following.

Theorem 4 ([1]). *For any $f, g : \mathbf{F}_q^d \to \mathbf{C}$ and $\lambda \in \mathbf{F}_q^*$,*

$$\left| \mathbf{E}_{x,y} f(x)g(y)\sigma_\lambda(x-y) - \mathbf{E}_x f(x) \mathbf{E}_y g(y) \right| \le 2q^{-\frac{d-1}{2}} \|f\|_2 \|g\|_2.$$

In order to include the proof of Theorem 4, we recall some Fourier analytic facts here. The *Fourier transform* $\widehat{f} : \mathbf{F}_q^d \to \mathbf{C}$ for $f : \mathbf{F}_q^d \to \mathbf{C}$ is defined for $\xi \in \mathbf{F}_q^d$ by

$$\widehat{f}(\xi) = \mathbf{E}_x f(x)\chi(x \cdot \xi).$$

The orthogonality of χ allows one to quickly conclude the Fourier inversion formula

$$f(x) = \sum_{\xi \in \mathbf{F}_q^d} \widehat{f}(\xi)\chi(-x \cdot \xi)$$

and the Plancherel identity

$$\mathbf{E}_x |f(x)|^2 = \sum_{\xi \in \mathbf{F}_q^d} |\widehat{f}(\xi)|^2.$$

While Lemma 3 relied only on Gaussian sums, Theorem 4 relies on more delicate cancellation. To be precise, for $a, b \in \mathbf{F}_q$ let us define the *Kloosterman sum*

$$K(a,b) := \sum_{\ell \in \mathbf{F}_q^*} \chi\left(as + \frac{b}{s}\right)$$

and the *Salié sum*

$$S(a,b) := \sum_{\ell \in \mathbf{F}_q^*} \chi\left(as + \frac{b}{s}\right)\eta(s).$$

If either a or b is nonzero, both bounds $|S(a,b)| \leq 2\sqrt{q}$ and $|K(a,b)| \leq 2\sqrt{q}$ are well-known; see, for instance, [13]. These bounds are somewhat more involved than showing $|G(\chi, \eta)| = q^{1/2}$. While the Salié sum can indeed be related back to a Gaussian sum, the Kloosterman sum bound relies on the fundamental work of Weil [18]. With these bounds, we are ready to prove the functional distance theorem.

Proof of Theorem 4. Proceeding as in Lemma 3, we apply orthogonality for

$$\mathbf{E}_{x,y}f(x)g(x+y)\sigma_\lambda(y)$$

$$= \mathbf{E}_{x,y}f(x)g(x+y)\sum_{\ell \in \mathbf{F}_q} \chi(\ell|y|^2)$$

$$= \mathbf{E}_x f(x)\mathbf{E}_y g(y) + \sum_{\ell \in \mathbf{F}_q^*} \chi(-\ell\lambda)\mathbf{E}_y \chi(\ell|y|^2)\mathbf{E}_x f(x)g(x+y).$$

Applying Fourier inversion,

$$\mathbf{E}_x f(x)g(x+y) = \sum_{\xi \in \mathbf{F}_q^d} \widehat{f}(\xi)\widehat{g}(-\xi)\chi(y \cdot \xi),$$

so we have shown

$$\mathbf{E}_{x,y}f(x)g(x+y)\sigma_\lambda(y) - \mathbf{E}_x f(x)\mathbf{E}_y g(y)$$

$$= \sum_{\xi \in \mathbf{F}_q^d} \widehat{f}(\xi)\widehat{g}(-\xi)\sum_{\ell \in \mathbf{F}_q^*}\chi(-\ell\lambda)\mathbf{E}_y\chi(\ell|y|^2 + y \cdot \xi).$$

By [14, Theorem 5.33],

$$\mathbf{E}_y\chi(\ell|y|^2 + y \cdot \xi) = q^{-d}G(\chi,\eta)^d\chi\left(-\frac{|\xi|^2}{4\ell}\right)\eta(\ell)^d$$

which we combine with our previous work to obtain

$$\left| \mathbf{E}_{x,y} f(x) g(x+y) \sigma_\lambda(y) - \mathbf{E}_x f(x) \mathbf{E}_y g(y) \right|$$

$$\leq q^{-d/2} \sum_{\xi \in \mathbf{F}_q^d} |\widehat{f}(\xi)| |\widehat{g}(-\xi)| \left| \sum_{\ell \in \mathbf{F}_q^*} \chi\left(-\ell\lambda - \frac{|\xi|^2}{4\ell} \right) \eta(\ell)^d \right|.$$

As $\lambda \neq 0$, we can recognize the sum in ℓ as either a Kloosterman sum (when d is even) or a Salié sum (when d is odd) and arrive at

$$\left| \mathbf{E}_{x,y} f(x) g(x+y) \sigma_\lambda(y) - \mathbf{E}_x f(x) \mathbf{E}_y g(y) \right| \leq 2q^{-\frac{d-1}{2}} \sum_{\xi \in \mathbf{F}_q^d} |\widehat{f}(\xi)| |\widehat{g}(-\xi)|.$$

Applying Cauchy-Schwarz and Plancherel,

$$\sum_{\xi \in \mathbf{F}_q^d} |\widehat{f}(\xi)| |\widehat{g}(-\xi)| \leq \|f\|_2 \|g\|_2,$$

completing the proof. □

4. Equilateral triangles

The rough plan is to argue that a properly normalized count of isometric copies of a particular graph $\mathcal{G}$ within sufficiently large subset $A \subseteq \mathbf{F}_q^d$ is roughly equal to a properly normalized count of isometric copies of the subgraph of $\mathcal{G}$ formed by deleting an edge. We will illustrate our approach through the small example of counting equilateral triangles, by which we mean complete graphs on 3 vertices that are connected by edges of the same length. This will amount to a total of three edge deletions, see the figure, and each edge deletion requires an application of Theorem 4, resulting in an error term. Provided we can keep the sum of these error terms under control, we will find ourselves counting points in A with no distance restrictions at all.

Proposition 5. *Let* $\lambda \in \mathbf{F}_q^*$, $A \subseteq \mathbf{F}_q^d$ *with* $|A| = \alpha q^d$ *and* $\alpha \geq 24q^{2-\frac{d+1}{2}}$. *Then*

$$(2) \qquad \mathbf{E}_{x,y,z} 1_A(x) 1_A(y) 1_A(z) \sigma_\lambda(x-y) \sigma_\lambda(x-z) \sigma_\lambda(y-z) \geq \frac{1}{2} \alpha^3.$$

Proof. Set $f_x(y) = 1_A(y) \sigma_\lambda(x-y)$. Applying Theorem 4 for a fixed x,

$$|\mathbf{E}_{y,z} f_x(y) f_x(z) \sigma_\lambda(y-z) - \mathbf{E}_y f_x(y) \mathbf{E}_z f_x(z)| \leq 2q^{\frac{1-d}{2}} \|f_x\|_2^2.$$

Note that f_x is not L^2-normalized, so we pick up an extra factor of q. That is,

$$\|f_x\|_2^2 = \mathbf{E}_y 1_A(y) \sigma_\lambda(x-y)^2$$
$$= q \mathbf{E}_y f_x(y).$$

Averaging over potential vertices x in A and applying the triangle inequality, we see

$$|\mathbf{E}_{x,y,z} 1_A(x) f_x(y) f_x(z) \sigma_\lambda(y-z) - \mathbf{E}_{x,y,z} 1_A(x) f_x(y) f_x(z)|$$
$$\leq 2q^{\frac{3-d}{2}} \mathbf{E}_x 1_A(x) \mathbf{E}_y f_x(y).$$

Recall $f_x(y) = 1_A(y)\sigma_\lambda(x - y)$. Since α is not too small, Theorem 4 tells us

$$\mathbf{E}_x 1_A(x)1_A(y)\sigma_\lambda(x - y) \leq 2\alpha^2.$$

Combining this with our assumption that α is not too small, we have shown

$$|\mathbf{E}_{x,y,z}1_A(x)f_x(y)f_x(z)\sigma_{\lambda_3}(y - z) - \mathbf{E}_{x,y,z}1_A(x)f_x(y)f_x(z)| \leq \frac{1}{6}\alpha^3.$$

What we have done so far is to show that the normalized count of equilateral triangles within A, $\{x, y, z\} \subseteq A$ each of distance λ from each other, is roughly equal to the normalized count of paths of length 2 within A, $\{x, y, z\} \subseteq A$ where both y and z are distant λ from x, where the error amounts to a small constant proportion of the expected count of α^3. In other words, by deleting an edge length from our desired configuration and counting the resulting configurations, we are still counting a positive proportion of the expected number of desired configurations.

We now aim to delete another edge to estimate $\mathbf{E}_{x,y,z}1_A(x)1_A(y)\sigma_\lambda(x - y)1_A(z)\sigma_\lambda(x - z)$. Continuing with the notation $f_z(x) = 1_A(x)\sigma_\lambda(x - z)$, we apply Theorem 4 for a fixed z to obtain

$$|\mathbf{E}_{x,y}f_z(x)1_A(y)\sigma_\lambda(x - y) - \alpha\mathbf{E}_x f_z(x)| \leq 2\alpha^{1/2}q^{\frac{1-d}{2}}\|f_z\|_2.$$

Averaging over potential vertices z in A and applying the triangle inequality,

$$|\mathbf{E}_{x,y,z}f_z(x)1_A(y)1_A(z)\sigma_\lambda(x-y)-\alpha\mathbf{E}_{x,z}f_z(x)1_A(z)| \leq 2\alpha^{1/2}q^{\frac{1-d}{2}}\mathbf{E}_z 1_A(z)\|f_z\|_2.$$

Applying Cauchy-Schwarz,

$$\begin{aligned}
\mathbf{E}_z 1_A(z)\|f_z\|_2 &\leq \alpha^{1/2}\mathbf{E}_z 1_A(z)\|f_z\|_2^2 \\
&= \alpha^{1/2}q^{1/2}(\mathbf{E}_{x,z}1_A(x)1_A(z)\sigma_\lambda(x - z))^{1/2} \\
&\leq 2\alpha^{3/2}q^{1/2},
\end{aligned}$$

where again we have used that α is not too small to productively apply Theorem 4. In total,

$$|\mathbf{E}_{x,y,z}f_z(x)1_A(y)1_A(z)\sigma_\lambda(x - y) - \alpha\mathbf{E}_{x,z}f_z(x)1_A(z)| \leq 4\alpha^2 q^{\frac{2-d}{2}},$$

where the error here is smaller than our previous error by a factor of q. In particular,

$$\begin{aligned}
|\mathbf{E}_{x,y,z}1_A(x)1_A(y)1_A(z)\sigma_\lambda(x - y)\sigma_\lambda(x - z)\sigma_\lambda(y - z) \\
- \alpha\mathbf{E}_{x,z}1_A(x)1_A(z)\sigma_\lambda(x - z)| \leq \frac{1}{3}\alpha^3.
\end{aligned}$$

We have now reduced the problem of counting equilateral triangles of side length λ within A to the problem of counting pairs of points of distance λ apart within A, which is exactly what Theorem 4 does for us. As α is not too small,

$$|\mathbf{E}_{x,z}1_A(x)1_A(z)\sigma_\lambda(x - z) - \alpha^2| \leq 2\alpha q^{\frac{1-d}{2}}$$

FIGURE 1. Each application of Theorem 4 deletes an edge to count simpler configurations.

which together with the previous reductions leads quickly to

$$|\mathbf{E}_{x,y,z}1_A(x)1_A(y)1_A(z)\sigma_\lambda(x-y)\sigma_\lambda(x-z)\sigma_\lambda(y-z) - \alpha^3| \leq \frac{1}{2}\alpha^3,$$

from which (2) follows immediately. □

In the argument above, deleting an edge via Theorem 4 allowed us to count simpler configurations at the price of including additional error resulting from our functions f_x not being L^2-normalized. This is the heart of our argument, and the rest is essentially bookkeeping. The quality of the error terms in each edge deletion above depended only upon the degrees of the vertices at the end of the chosen edge. This should give some hope that for counting cycles of length n, we may have to sum more error terms (an amount depending upon n), but each power of q appearing is at worst $q^{\frac{3-d}{2}}$. Roughly, as we proceed through the general argument in the next section, we will see that deleting an edge between two vertices of degree at most t amounts to increasing the $q^{\frac{1-d}{2}}$ factor appearing in the error from Theorem 4 to a factor of $q^{t+\frac{1-d}{2}}$. This accounts for the necessary size constraint imposed on α. It should be believable that the restriction to equilateral configurations, while conceptually convenient, is not particularly important since we delete one edge at a time.

It is reasonable to worry that Proposition 5 does not guarantee us three *distinct* points $\{x, y, z\} \subseteq A$ forming the desired triangle; this is required to make any conclusion about A containing an isometric copy of a complete graph on 3 vertices with 3 edge lengths of λ. It is not terribly surprising that coincidences between these vertices amount to a well-controlled error term, and we make this precise at the end of the next section in the general case.

5. Edge deletion induction

We now establish our main theorem. For the remainder we will consider distance graphs $\mathcal{G}$ with vertices labeled as $V = \{v_1, \ldots, v_n\}$ and when v_i is adjacent to v_j we label their common edge $e_{i,j}$. Recall that to each such edge we have a distance $\lambda_{e_{i,j}} \in \mathbf{F}_q^*$. As in the simple example in the previous section, we will apply Theorem 4 to delete a single distance constraint, an edge length, and consider the resulting subgraph. We find it convenient to introduce the

shorthand for $x \in \mathbf{F}_q^d$ of

$$\sigma_{i,j}(x) := \begin{cases} \sigma_{\lambda_{e_{i,j}}}(x) & \text{if } v_i \text{ and } v_j \text{ are adjacent,} \\ 1 & \text{otherwise,} \end{cases}$$

to compactly for $A \subseteq \mathbf{F}_q^d$ the count of

$$(3) \qquad \mathcal{N}_{\mathcal{G}}(A) := \mathbf{E}_{x_1,\dots,x_n} \prod_{i=1}^n 1_A(x_i) \prod_{j=i+1}^n \sigma_{i,j}(x_i - x_j).$$

This is *almost* a normalized count for the number of (ordered) isometric copies of $\mathcal{G}$ appearing within A. The only difference is the additional contribution from degenerate configurations that occur when some of the non-adjacent x_i coincide. We will account for this contribution at the end of the section. Our main asymptotic here is:

Theorem 6. *Let $\mathcal{G} = (V, E)$ be a distance graph with n vertices, m edges, and maximum degree t. Let $A \subseteq \mathbf{F}_q^d$ with $|A| = \alpha q^d$ and*

$$(4) \qquad\qquad\qquad \alpha \geq 4mq^{t - \frac{d+1}{2}}.$$

Then

$$(5) \qquad\qquad\qquad \left| \mathcal{N}_{\mathcal{G}}(A) - \alpha^n \right| \leq 4m\alpha^{n-1} q^{t - \frac{d+1}{2}}.$$

Proof. We proceed by induction on the number of edges, m. In the case $m = 1$, note the maximum degree is $t = 1$ and we may as well assume we only have $n = 2$ vertices, since each isolated vertex contributes a factor of α to $\mathcal{N}_{\mathcal{G}}(A)$. Then (5) is actually weaker by a constant factor than the content of Theorem 4 applied with $f = g = 1_A$, so we have one base case already established.

   We assume that the theorem has been established for graphs with at most $m - 1$ edges. It is obvious, but crucial for us, that the maximum degree of subgraphs of $\mathcal{G}$ remains at most t so that we can proceed to apply (5) to the subgraph of $\mathcal{G}$ formed by removing an edge. For notational convenience, there is no harm in assuming that v_1 and v_2 are adjacent. Set

$$f(x_1) := 1_A(x_1) \prod_{j=3}^n \sigma_{1,j}(x_1 - x_j),$$

and, for $2 \leq i \leq n$,

$$g_i(x_i) := 1_A(x_i) \prod_{j=i+1}^n \sigma_{i,j}(x_i - x_j).$$

Note that f and the g_i together account for all edges of $\mathcal{G}$ except for $e_{1,2}$, which we will be deleting. With this notation, we can rearrange our count (3) as

$$(6) \qquad \mathcal{N}_{\mathcal{G}}(A) = \mathbf{E}_{x_3,\dots,x_n} \prod_{i=3}^n g_i(x_i) \mathbf{E}_{x_1,x_2} f(x_1) g_2(x_2) \sigma_{1,2}(x_1 - x_2).$$

Applying Theorem 4 to the inner average,

$$(7) \qquad \mathbf{E}_{x_1,x_2} f(x_1) g_2(x_2) \sigma_{1,2}(x_1 - x_2) = \mathbf{E}_{x_1} f(x_1) \mathbf{E}_{x_2} g_2(x_2) + \mathcal{E},$$

where

$$|\mathcal{E}| \leq 2q^{\frac{1-d}{2}} \|f\|_2 \|g_2\|_2.$$

Since f and g_2 amount to normalized indicator functions, we can essentially replace these L^2 norms with simple averages provided we are careful about normalization. Since they each account for at most $t-1$ edges, we have

$$\|f\|_2^2 = \mathbf{E}_{x_1} f(x_1)^2$$
$$\leq q^{t-1} \mathbf{E}_{x_1} f(x_1)$$

and similarly $\|g_2\|_2^2 \leq q^{t-1} \mathbf{E}_{x_2} g(x_2)$. Hence,

$$|\mathcal{E}| \leq 2q^{t-\frac{d+1}{2}} \left(\mathbf{E}_{x_1} f(x_1) \right)^{1/2} \left(\mathbf{E}_{x_2} g(x_2) \right)^{1/2}.$$

Together with (6) and (7), we see

$$\left| \mathcal{N}_{\mathcal{G}}(A) - \mathbf{E}_{x_1,\dots,x_n} f(x_1) \prod_{i=2}^{n} g_i(x_i) \right| \leq 2q^{t-\frac{d+1}{2}} \mathbf{E}_{x_3,\dots,x_n} \prod_{i=3}^{n} g_i(x_i) \|f\|_1^{1/2} \|g_2\|_1^{1/2}.$$

This essentially says that, up to an error, the count for copies of $\mathcal{G}$ within A is the same as that for the subgraph of $\mathcal{G}$ formed by deleting the edge $e_{1,2}$. We invoke our induction hypothesis to observe

$$\left| \mathbf{E}_{x_1,\dots,x_n} f(x_1) \prod_{i=2}^{n} g_i(x_i) - \alpha^n \right| \leq 4(m-1)\alpha^{n-1} q^{t-\frac{d+1}{2}},$$

in which case we have shown

$$(8) \qquad \begin{aligned} & \left| \mathcal{N}_{\mathcal{G}}(A) - \alpha^n \right| \\ & \leq 4(m-1)\alpha^{n-1} q^{t-\frac{d+1}{2}} + 2q^{t-\frac{d+1}{2}} \mathbf{E}_{x_3,\dots,x_n} \prod_{i=3}^{n} g_i(x_i) \|f\|_1^{1/2} \|g_2\|_1^{1/2}. \end{aligned}$$

Now we need only show that we can bring these two error terms in line. Applying Cauchy-Schwarz,

$$\mathbf{E}_{x_3,\dots,x_n} \prod_{i=3}^{n} f_i(x_i) \|f\|_1^{1/2} \|g_2\|_1^{1/2}$$
$$\leq \left(\mathbf{E}_{x_3,\dots,x_n} \prod_{i=3}^{n} g_i(x_i) \right)^{1/2} \left(\mathbf{E}_{x_1,\dots,x_n} f(x_1) \prod_{i=2}^{n} g_i(x_i) \right)^{1/2}.$$

By our induction hypothesis, and our assumption that α is not too small, we have both

$$\mathbf{E}_{x_3,\dots,x_n} \prod_{i=3}^{n} g_i(x_i) \leq 2\alpha^{n-2}, \text{ and}$$

$$\mathbf{E}_{x_1,\ldots,x_n} f(x_1) \prod_{i=2}^{n} g_i(x_i) \leq 2\alpha^n.$$

In particular, together with (8), this allows us to conclude (5) as desired. $\square$

To extract Theorem 2, we must argue that at least one of the many configurations counted by $\mathcal{N}_{\mathcal{G}}(A)$ is genuine, in the sense that it is made up of distinct vertices. As one might expect, we will in fact show that when A is large enough, many of these configurations are genuine. Note that Theorem 2 follows immediately from the following.

Theorem 7. *Let $n, t \in \mathbf{N}$ and let $A \subseteq \mathbf{F}_q^d$ with $|A| = \alpha q^d$ with $\alpha \geq 12n^2 q^{t - \frac{d+1}{2}}$. Then A contains at least $\frac{1}{2}|A|^n q^{-m}$ isometric copies of every distance graph with n vertices, m edges, and maximum vertex degree t.*

Proof. Let $\mathcal{G}$ be a distance graph with n vertices, m edges, and maximum vertex degree t with the same labeling scheme as before. We introduce the restricted average

$$\mathbf{E}^*_{x_1,\ldots,x_n} := q^{-nd} \sum_{\substack{x_1,\ldots,x_n \in \mathbf{F}_q^d \\ x_1,\ldots,x_n \text{ distinct}}}$$

and consider

$$\mathcal{N}^*_{\mathcal{G}}(A) := \mathbf{E}^*_{x_1,\ldots,x_n} \prod_{i=1}^{n} 1_A(x_i) \prod_{j=i+1}^{n} \sigma_{i,j}(x_i - x_j),$$

which only counts genuine isometric copies of $\mathcal{G}$ appearing within A. If we are considering a configuration within A that is detected by $\mathcal{N}_{\mathcal{G}}$, but not by $\mathcal{N}^*_{\mathcal{G}}(A)$, then some vertices coincide. We will show that the contribution when x_n is equal to one of $x_1, \ldots, x_{n-1}$, the contribution is negligible. In particular, we will bound

$$\mathcal{E}_n := \mathbf{E}_{x_1,\ldots,x_{n-1}} \prod_{i=1}^{n-1} 1_A(x_i) \prod_{j=i+1}^{n-1} \sigma_{i,j}(x_i - x_j) q^{-d}$$

$$\sum_{x_n \in \{x_1,\ldots,x_{n-1}\}} 1_A(x_n) \prod_{i=1}^{n-1} \sigma_{i,n}(x_i - x_n).$$

Of course, we can express this as

$$\mathcal{E}_n = q^{-d} \mathbf{E}_{x_1,\ldots,x_{n-1}} \prod_{i=1}^{n-1} 1_A(x_i) \prod_{j=i+1}^{n-1} \sigma_{i,j}(x_i - x_j) \sum_{j=1}^{n-1} \prod_{i=1}^{n-1} \sigma_{i,n}(x_i - x_j).$$

When the innermost sum is nonzero, its contribution is at least bounded by $(n-1)q^t$, since v_n has degree at most t. As we are assuming α is not too small, we can invoke Theorem 6 to conclude

$$\mathcal{E}_n \leq 2(n-1)\alpha^{n-1} q^{t-d}.$$

Applying a symmetric argument for coincidences involving vertices other than x_n, we have shown

$$|\mathcal{N}_{\mathcal{G}}(A) - \mathcal{N}_{\mathcal{G}}^*(A)| \leq 2n^2\alpha^{n-1}q^{t-d}.$$

Combining this with Theorem 6 with our size requirement on α yields

$$|\mathcal{N}_{\mathcal{G}}^*(A) - \alpha^n| \leq 6n^2\alpha^{n-1}q^{t-\frac{d+1}{2}}$$

which implies that $\mathcal{N}_{\mathcal{G}}^*(A) \geq \frac{1}{2}\alpha^n$. The result follows by then clearing normalizations. $\square$

References

[1] M. Bennett, J. Chapman, D. Covert, D. Hart, A. Iosevich, and J. Pakianathan, *Long paths in the distance graph over large subsets of vector spaces over finite fields*, J. Korean Math. Soc. **53** (2016), no. 1, 115–126. `https://doi.org/10.4134/JKMS.2016.53.1.115`

[2] M. Bennett, D. Hart, A. Iosevich, J. Pakianathan, and M. Rudnev, *Group actions and geometric combinatorics in $\mathbb{F}_q^d$*, Forum Math. **29** (2017), no. 1, 91–110. `https://doi.org/10.1515/forum-2015-0251`

[3] M. Bennett, A. Iosevich, and K. Taylor, *Finite chains inside thin subsets of $\mathbb{R}^d$*, Anal. PDE **9** (2016), no. 3, 597–614. `https://doi.org/10.2140/apde.2016.9.597`

[4] J. Chapman, M. B. Erdoğan, 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. `https://doi.org/10.1007/s00209-011-0852-4`

[5] N. Chatzikonstantinou, A. Iosevich, S. Mkrtchyan, and J. Pakianathan, *Rigidity, graphs and Hausdorff dimension*, arXiv preprint arXiv:1708.05919 (2017).

[6] M. B. Erdoğan, *A bilinear Fourier extension theorem and applications to the distance set problem*, Int. Math. Res. Not. **2005** (2005), no. 23, 1411–1425. `https://doi.org/10.1155/IMRN.2005.1411`

[7] K. J. Falconer, *On the Hausdorff dimensions of distance sets*, Mathematika **32** (1985), no. 2, 206–212 (1986). `https://doi.org/10.1112/S0025579300010998`

[8] A. Greenleaf, A. Iosevich, and M. Pramanik, *On necklaces inside thin subsets of $\mathbb{R}^d$*, Math. Res. Lett. **24** (2017), no. 2, 347–362. `https://doi.org/10.4310/MRL.2017.v24.n2.a4`

[9] L. Guth, A. Iosevich, Y. Ou, and H. Wang, *On Falconer's distance set problem in the plane*, arXiv preprint arXiv:1808.09346 (2018).

[10] L. Guth and N. H. Katz, *On the Erdős distinct distances problem in the plane*, Ann. of Math. (2) **181** (2015), no. 1, 155–190. `https://doi.org/10.4007/annals.2015.181.1.2`

[11] A. Iosevich and B. Liu, *Equilateral triangles in subsets of $\mathbb{R}^d$ of large hausdorff dimension*, arXiv preprint arXiv:1603.01907 (2016).

[12] A. Iosevich and M. Rudnev, *Erdős distance problem in vector spaces over finite fields*, Trans. Amer. Math. Soc. **359** (2007), no. 12, 6127–6142. `https://doi.org/10.1090/S0002-9947-07-04265-1`

[13] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, American Mathematical Society Colloquium Publications, **53**, American Mathematical Society, Providence, RI, 2004. `https://doi.org/10.1090/coll/053`

[14] R. Lidl and H. Niederreiter, *Finite Fields*, second edition, Encyclopedia of Mathematics and its Applications, **20**, Cambridge University Press, Cambridge, 1997.

[15] H. Parshall, *Simplices over finite fields*, Proc. Amer. Math. Soc. **145** (2017), no. 6, 2323–2334. `https://doi.org/10.1090/proc/13493`

[16] L. A. Vinh, *On kaleidoscopic pseudo-randomness of finite Euclidean graphs*, Discuss. Math. Graph Theory **32** (2012), no. 2, 279–287. `https://doi.org/10.7151/dmgt.1597`

[17] _____, *The sovability of norm, bilinear and quadratic equations over finite fields via spectra of graphs*, Forum Math. **26** (2014), no. 1, 141–175. `https://doi.org/10.1515/form.2011.155`

[18] A. Weil, *On some exponential sums*, Proc. Nat. Acad. Sci. U. S. A. **34** (1948), 204–207. `https://doi.org/10.1073/pnas.34.5.204`

[19] T. Wolff, *Decay of circular means of Fourier transforms of measures*, Internat. Math. Res. Notices **1999** (1999), no. 10, 547–567. `https://doi.org/10.1155/S1073792899000288`

ALEX IOSEVICH
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ROCHESTER
ROCHESTER, NY 14627, USA
Email address: `iosevich@math.rochester.edu`

HANS PARSHALL
DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY
COLUMBUS, OH 43210, USA
Email address: `parshall.6@osu.edu`