Canad. J. Math. Vol. 77 (3), 2025 pp. 1041–1071
http://dx.doi.org/10.4153/S0008414X24000245
© The Author(s), 2024. Published by Cambridge University Press on behalf of Canadian Mathematical Society

Geometric structures in pseudo-random graphs

Thang Pham, Steven Senger, Michael Tait, and Vu Thi Huong Thu

Abstract. In this paper, we provide a general framework for counting geometric structures in pseudo-random graphs. As applications, our theorems recover and improve several results on the finite field analog of questions originally raised in the continuous setting. The results present interactions between discrete geometry, geometric measure theory, and graph theory.

1 Introduction

Let $\mathbb{F}_q$ be a finite field of order q where q is an odd prime power. The investigation of finite field analogs of problems originally raised in geometric measure theory has a long tradition, for instance, the Erdős–Falconer distance problem [6, 7, 18], sum-product estimates [4, 10], the Kakeya problem [8, 33], frame theory [16, 17], and restriction problems [15, 23, 24, 28, 29]. Studying these problems over finite fields is not only interesting by itself, but it also offers new ideas to attack the original questions. Some of these problems can be proved by using results from graph theory. For instance, in [18], Iosevich and Rudnev proved the following theorem on the distribution of distances in a given set.

Theorem 1.1 (Iosevich–Rudnev [18]) *Let E be a set in $\mathbb{F}_q^d$. Assume that $|E| \gg q^{\frac{d+1}{2}}$, then*

$$\Delta(E) := \{||x-y|| = (x_1-y_1)^2 + \cdots + (x_d-y_d)^2 : x, y \in E\} = \mathbb{F}_q.$$

It is well known that this theorem can be reproved by using the famous expander mixing lemma, which helps describe the behavior of (n, d, λ)-graphs. We say $\mathcal{G}$ is an (n, d, λ)-graph, if it is a regular graph with n vertices, each of degree d, and all eigenvalues of its adjacency matrix are bounded in absolute value by λ (with the exception of the largest eigenvalue, which will be d, as the graph is regular). Specifically, the expander mixing lemma states that for such a graph, the number of edges in a given vertex set U, denoted by $e(U)$, is bounded from both above and below

Received by the editors September 16, 2023; revised February 9, 2024; accepted February 19, 2024.
Published online on Cambridge Core March 15, 2024.
T. Pham was supported by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 101.99–2021.09. M. Tait was partially supported by National Science Foundation Grant DMS-2011553 and a Villanova University Summer Grant.
AMS subject classification: **52C10**, 11T06.
Keywords: Pseudo-random graphs, geometric structures, finite fields.

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

by the inequality

$$\left| e(U) - \frac{d|U|^2}{2n} \right| \leq \lambda |U|.$$

To derive Theorem 1.1 from this estimate, for $\alpha \in \mathbb{F}_q^*$, one just needs to define the distance graph $\mathcal{DG}_\alpha$ with the vertex set $\mathbb{F}_q^d$ and there is an edge between two vertices x and y if and only if $||x-y|| = \alpha$. It is not hard to check that $\mathcal{DG}_\alpha$ is a regular Cayley graph with q^d vertices, of degree $(1+o(1))q^{d-1}$, and the second eigenvalue is bounded by $2q^{\frac{d-1}{2}}$ by using Kloosterman sums [19, 32]. So, for any $\alpha \neq 0$, the expander mixing lemma implies directly that any vertex set U of size at least $2q^{\frac{d+1}{2}}$ spans at least one edge.

We observe that the argument above only made use of the pseudo-randomness properties of the graph, and once the eigenvalues were calculated ignored anything about $\mathbb{F}_q$ or the distance function. Because of this observation, this machinery provides a unified proof for a series of similar questions, for example, one can replace the distance function by bilinear forms [12], Minkowski distance function [13], or other functions [31].

From this observation, it is very natural to ask what kinds of finite field models can be extended to the graph setting? That is, what "geometric structures" can we guarantee in a general graph with some pseudo-randomness condition? The main purpose of this paper is to provide three such configurations, and the three topics we present here can be viewed as generalizations of the Erdős–Falconer distance conjectures, which have been studied intensively in the literature. Our theorems imply several results found previously as special cases. Moreover, as they rely on the pseudo-randomness of an underlying graph, they can be applied in a straightforward manner to other settings, such as modules over finite rings.

Throughout the paper, we say that G is an (n,d,λ)-colored graph with color set D if it is a graph edge-colored with $|D|$ colors such that the subgraph of any fixed color is an (n,d,λ)-graph.

1.1 Cartesian product structures

We first start with the following question about finding rectangles in $\mathbb{F}_q^d$.

Question 1.2 *Let E be a set in $\mathbb{F}_q^d$ and $\alpha, \beta \in \mathbb{F}_q^*$. How large does E need to be to guarantee that there are four points $w, x, y, z \in E$ such that they form a rectangle of side lengths α and β, i.e.,*

(1)
$$(w-x)\cdot(x-y) = 0,\ (x-y)\cdot(y-z) = 0,\ (y-z)\cdot(z-w) = 0,\ (z-w)(w-x) = 0,$$

and

(2)
$$||w-x|| = ||y-z|| = \alpha,\ ||x-y|| = ||z-w|| = \beta.$$

Lyall and Magyar [26] proved that for any $\delta \in (0,1)$, there exists an integer $q_0 = q_0(\delta)$ with the following property: if $q \geq q_0$ and $E \subset \mathbb{F}_q^{2d}$ with $|E| \geq \delta q^{2d}$, then E contains four points a, b, c, and d satisfying (1) and (2). This is the finite field model

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

of a result in the same paper which states that for any given rectangle $\mathcal{R}$ in $\mathbb{R}^{2d}$, if $S \subset \mathbb{R}^{2d}$ has positive Banach density, then there exists a threshold $\lambda_0 = \lambda_0(S, \mathcal{R})$ such that S contains an isometric copy of $\lambda\mathcal{R}$ for any $\lambda \geq \lambda_0$. Notice that the result in [26] was actually proved in a more general form, for d-dimensional rectangles, though we state it here for two-dimensional rectangles.

In the first theorem of this paper, we extend this result to a general graph setting.

For two graphs G and H, the *Cartesian product of G and H*, denoted by $G \square H$, is the graph where $V(G \square H) = V(G) \times V(H)$ and $(u_1, v_1) \sim (u_2, v_2)$ if and only if either $u_1 = u_2$ and $\{v_1, v_2\} \in E(H)$ or $v_1 = v_2$ and $\{u_1, u_2\} \in V(G)$. We use $S(x)$ to denote the indicator function of the set S.

Theorem 1.3 *Let G_i be (n_i, d_i, λ_i)-graphs with $1 \leq i \leq 2$. Set $G = G_1 \square G_2$. For any $0 < \delta' < \delta < 1$, there exists $\varepsilon > 0$ such that for any $S \subset V(G_1 \square G_2)$ with $|S| \geq \delta |V(G_1 \square G_2)|$, if $\max\left\{\frac{\lambda_1}{d_1}, \frac{\lambda_2}{d_2}\right\} < \varepsilon$, then*

$$N = \sum_{(u_1,u_2)\in E(G_1),(v_1,v_2)\in E(G_2)} S(u_1, v_1)S(u_1, v_2)S(u_2, v_1)S(u_2, v_2) > \delta' 4n_1n_2d_1d_2.$$

Theorem 1.3 recovers the theorem on rectangles in $\mathbb{F}_q^d$ by Lyall and Magyar (Proposition 2.1 in [26]) which we now state.

Proposition 1.4 (Proposition 2.1 in [26]) *For any $0 < \delta \leq 1$, there exists an integer $q_0 = q_0(\delta)$ with the following property: if $q \geq q_0$ and $t_1, \ldots, t_d \in \mathbb{F}_q^*$, then any $S \subseteq \mathbb{F}_q^{2d}$ with $|S| \geq \delta q^{2d}$ will contain points $\{x_{11}, x_{12}\} \times \cdots \times \{x_{d1}, x_{d2}\} \subseteq V_1 \times \cdots \times V_d$ with $|x_{j2} - x_{j1}|^2 = t_j$ for $1 \leq j \leq d$ where we have written $\mathbb{F}_q^{2d} = V_1 \times \cdots \times V_d$ with $V_j \cong \mathbb{F}_q^2$ pairwise orthogonal coordinate subspaces.*

To see this, let G_1 and G_2 be the graphs each with vertex set $\mathbb{F}_q^d$ where $a \sim b$ in G_1 if $||a - b|| = \alpha$ and $x \sim y$ in G_2 if $||a - b|| = \beta$. Then G_1 and G_2 are graphs with q^d vertices, degree asymptotically q^{d-1}. We next appeal to a result from Vinh in [31] that puts together work from [1, 22] in a form useful for us. Specifically:

Theorem 1.5 (Theorem 10.1 in [31]) *Let Q be a nondegenerate quadratic form on $\mathbb{F}_q^d$. For any $\alpha \in \mathbb{F}_q^*$, the graph whose vertices are $\mathbb{F}_q^d$ and whose edges are the pairs of distinct vertices $x, y \mathbb{F}_q^d$ satisfying $Q(x - y) = \alpha$, is a $(q^d, (1 + o(1))q^{d-1}, 2q^{(d-1)/2})$-graph.*

This gives us that $\lambda \leq 2q^{(d-1)/2}$ here. Note that if $||u_1 - u_2|| = \alpha$ and $||v_1 - v_2|| = \beta$, then letting $w = (u_1, v_1)$, $x = (u_1, v_2)$, $y = (u_2, v_1)$, and $z = (u_2, v_2)$, we have that w, x, y, and z form a rectangle in $\mathbb{F}_q^{2d}$ with side lengths α and β. Applying Theorem 1.3 to these specific graphs shows that for q large enough, any subset of $\mathbb{F}_q^{2d}$ of size at least δq^{2d} contains $\Omega(q^{4d-2})$ rectangles, giving a quantitative strengthening of Lyall and Magyar's result. Another application of Theorem 1.3 is on the number of rectangles in $\mathbb{F}_q^2$ with side lengths in a given multiplicative subgroup of $\mathbb{F}_q$; precisely, given a multiplicative subgroup A of $\mathbb{F}_q$, we define $G_1 = G_2$ being the graph with the vertex set $\mathbb{F}_q$ and there is an edge between x and y if $x - y \in A$. This is clear that this is a Cayley graph with q vertices, of degree $|A|$, and it is also well known that $\lambda \leq q^{1/2}$ (see [20, (1)] for computations). Applying Theorem 1.3, we recover Theorem 1.1 from [20].

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

Theorem 1.6 (Theorem 1.1 in [20]) *For any $0 < \delta < 1$, there exists an integer q_0 with the following property: if $q \geq q_0$ and A is a multiplicative subgroup of $\mathbb{F}_q^*$ with $|A| \geq q^{2/3}$, then any set $S \subset \mathbb{F}_q^2$ with $|S| \geq \delta q^2$ contains at least $\Omega\left(\frac{|S|^4|A|^2}{q^5}\right)$ rectangles with side lengths in A.*

1.2 Distribution of cycles

Our motivation of this section comes from the following question.

Question 1.7 *Let E be a set in $\mathbb{F}_q^d$, and let $m \geq 4$ be an integer. How large does E need to be to guarantee that the number of cycles of the form $(x_1, \dots, x_m)$ with $\|x_i - x_{i+1}\| = 1$ for all $1 \leq i \leq m-1$, and $\|x_m - x_1\| = 1$, is close to the expected number $|E|^m q^{-m}$?*

Iosevich, Jardine, and McDonald [14] proved that the number of cycles of length m, denoted by $C_m(E)$, is close to the expected number if E is sufficiently large. Here, we record their result, which is Theorem 1.2 in [14].

Theorem 1.8 (Theorem 1.2 in [14]) *Given $E \subseteq \mathbb{F}_q^d$, we have that*

$$C_m(E) = (1 + o(1))\frac{|E|^m}{q^m}, \tag{3}$$

whenever

$$|E| \geq \begin{cases} q^{\frac{1}{2}\left(d+2-\frac{k-2}{k-1}+\delta\right)} & : m = 2k, \text{ even} \\ q^{\frac{1}{2}\left(d+2-\frac{2k-3}{2k-1}+\delta\right)} & : m = 2k+1, \text{ odd} \end{cases}$$

where

$$0 < \delta < \frac{1}{2\left\lfloor \frac{m}{2} \right\rfloor^2}.$$

In the continuous setting, this is a difficult problem, and there are only a few partial results. For instance, as a consequence of a theorem due to Eswarathasan, Iosevich, and Taylor [9], we know that if the Hausdorff dimension of E, denoted by s, is at least $\frac{d+1}{2}$, then we know that the upper Minkowski dimension of the set of cycles in E is at most $2s - m$. If we consider the case of paths, then Bennett, Iosevich, and Taylor [3] showed that there exists an open interval I such that for any sequence $\{t_i\}_{i=1}^m$ of elements in I, we always can find paths of length $m+1$ with gaps $\{t_i\}_{i=1}^m$ between subsequent elements in E as long as the Hausdorff dimension of E is greater than $\frac{d+1}{2}$. We refer the reader to [25] for the recent study on this problem. It is worth noting that in the discrete setting, results on distribution of paths also play a crucial role in proving (3).

In the graph setting, we have the following extension. We note here that we are counting any sequence of m vertices $(v_1, \dots, v_m)$ with $v_i \sim v_{i+1}$ and $v_1 \sim v_m$ as a cycle of length m. That is, we are counting labeled cycles and we include degenerate cycles in the count. One could combine Theorem 1.9 for various values of m and lemmas used to prove it to obtain results about nondegenerate cycles as well, but we do not do this explicitly here. In what follows, when two quantities $X(n)$ and $Y(n)$ vary with respect to some controlling parameter n, we write $X(n) \ll Y(n)$ to indicate that $X(n) = O(Y(n))$.

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

Theorem 1.9 *Let $\mathcal{G}$ be an (n,d,λ)-graph, and let U be a vertex set with $\lambda \cdot \frac{n}{d} = o(|U|)$. Let $C_m(U)$ denote the number of (labeled, possibly degenerate) cycles of length m with vertices in U. Then we have*

$$\left| C_m(U) - \frac{|U|^m d^m}{n^m} \right| = O\left(\frac{\lambda |U|^{m-1} d^{m-1}}{n^{m-1}} + \frac{\lambda^{m-2} |U|^2 d}{n} \right).$$

The error term cannot be improved for $m = 4$. For instance, we define a graph with the vertex set $\mathbb{F}_q^2$ where two vertices (a,b) and (c,d) are adjacent if and only if $ac + bd = 1$. Using the geometric facts in $\mathbb{F}_q^2$ that any two lines intersect in at most one point and there is only one line passing through two given points, we can see that this graph contains no C_4, even though it is a $(q^2, q, \sqrt{q})$ graph (if one includes loops).

We also remark that as a corollary of the following result due to Alon, we know that the number of cycles in U is close to the expected number as long as $|U| \gg \lambda(n/d)^2$. Here, we state a version of this result given as [21, Theorem 4.10].

Theorem 1.10 *Let H be a fixed graph with r edges, s vertices, and maximum degree Δ, and let $G = (V,E)$ be an (n,d,λ)-graph, where $d \le 0.9n$. Let $m < n$ satisfy $m \gg \lambda \left(\frac{n}{d}\right)^{\Delta}$. Then, for every subset $V' \subseteq V$ of cardinality m, the number of (not necessarily induced) copies of H in V' is*

$$(1 + o(1)) \frac{m^s}{|\mathrm{Aut}(H)|} \left(\frac{d}{n}\right)^r.$$

The statement about cycles implied by this result is of course weaker than Theorem 1.9. We now discuss how Theorem 1.9 implies and improves previous, more specific results. In [14], counting results for cycles are proved in both the distance graph and the dot-product graph over $\mathbb{F}_q^d$. Formally, let G_t^{dist} and G_t^{prod} be the graphs on vertex set $\mathbb{F}_q^d$ where $u \sim v$ in G_t^{dist} if $\|u - v\| = t$ and $u \sim v$ in G_t^{prod} if $u \cdot v = t$. As each of these graphs is approximately q^{d-1} regular and with second eigenvalue bounded above by $2q^{(d-1)/2}$, Theorem 1.9 can be applied. In [14], the same quantitative results are proved for both graphs but with different methods, and the authors write the following:

"We note that in this paper, we obtain the same results for the distance graph and the dot- product graph. While the techniques are, at least superficially, somewhat different due to the lack of translation invariance in the dot-product setting, it is reasonable to ask whether a general formalism is possible."

Theorem 1.9 answers this question in a strong way, as it may be applied in a much more general setting than just distance or dot-product graphs. Furthermore, Theorem 1.9 implies the estimate (3) with an improved threshold on the size of the subset, namely we may remove the δ in the exponent that appears in Theorem 1.8, from [14]. The proof of Theorem 1.9 requires estimates on the number of paths in our graph, for example, Proposition 3.5. This is again done in a general way for (n,d,λ)-graphs. We also note that colorful versions of Theorem 1.9 and the lemmas required to prove it can be proved with only minor modifications to the proof. That is, given an (n,d,λ)-colored graph and a fixed coloring of a path or cycle, one can obtain the same estimates on the number of such colorful subgraphs that appear. For ease of exposition, we only prove an uncolored version of Theorem 1.9, but Theorems 1.15 and 1.16 (see below)

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

are stated and proved in a colorful way as proof of concept. It is possible through this general setup to recover Theorem 1.1 of [2] and Theorem 6 of [5], both of which we explicitly state below, for the sake of completeness.

Theorem 1.11 (Theorem 1.1 of [2]) *Let $E \subseteq \mathbb{F}_q^d$, where $d \geq 2$ and $|E| > \frac{2k}{\ln 2} q^{\frac{d+1}{2}}$. Suppose that $t_i \neq 0$, for $1 \leq i \leq k$. Then*

$$\left| |\{(x^1, \dots, x^{k+1}) \in E \times \cdots \times E : \|x^i - x^{i+1}\| = t_i, 1 \leq i \leq k\}| - \frac{|E|^{k+1}}{q^k} \right| \leq \frac{2k}{\ln 2} q^{\frac{d+1}{2}} \frac{|E|^k}{q^k}.$$

Theorem 1.12 (Theorem 6 of [5]) *Let $E \subseteq \mathbb{F}_q^d$, where $d \geq 2$, $\alpha, \beta \in \mathbb{F}_q^*$, and E is large enough so that for all $\varepsilon > 0$, there exists a positive constant C_ε so that $|E| \gg q^{\frac{d+1}{2}+\varepsilon}$. Then*

$$\left|\{(x^1, x^2, x^3) \in E \times E \times E : x^1 \cdot x^2 = \alpha, x^2 \cdot x^3 = \beta\}\right| = (1 + o(1)) \frac{|E|^3}{q^2}.$$

Finally, we prove Theorem 1.9 in two different ways. The second approach is quite specific to counting cycles, but more straightforward (it is also slightly weaker: we obtain the same quantitative results for $m \geq 5$ but for $m = 4$ only prove the result up to a multiplicative constant factor). The first approach passes the problem to counting structures in the tensor product of two (n, d, λ)-graphs. We note that the tensor product of two (n, d, λ)-graphs is itself an $(n^2, d^2, d\lambda)$ graph, and so one may try to use pseudo-randomness of this graph to count subgraphs. However, this is not good enough for our purpose, and we must prove a version of the expander mixing lemma that applies specifically to tensor products of graphs. This result (Proposition 3.2) is significantly stronger than directly applying the classical expander mixing lemma to the tensor product graph, and we believe it is of independent interest, as the second approach along with Proposition 3.2 could be used to count other structures in tensor products of pseudo-random graphs.

1.3 Distribution of disjoint trees

The last question we consider in this paper is the following.

Question 1.13 *Let E be a set in $\mathbb{F}_q^d$, and let T be a tree of m vertices. How large does E need to be to guarantee that the number of vertex disjoint copies of T in E is close to $|E|/m$?*

That is, we are asking for a threshold such that any set of large enough size has an almost spanning T-factor. We now to introduce the notion of the stringiness of a graph, T, denoted $\sigma(T)$, which is defined as $(d_1 + 1) \prod_{i=2}^n d_i$ where $d_1 \geq d_2 \geq \cdots \geq d_n$ is the degree sequence of T in nonincreasing order. Using this concept of stringiness, Soukup [30] proved a result about trees. While his result holds in more general settings, we state the relevant version, using our notation, here.

Theorem 1.14 (Theorem 6 in [30]) *For any tree T of m vertices with stringiness $\sigma(T)$, and for any $E \subset \mathbb{F}_q^d$, if $|E| \gg \sigma(T) q^{\frac{d+1}{2}}$, then the number of disjoint copies of T in E is at*

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

least

$$\frac{|U|}{\sigma(T)} - q^{\frac{d+1}{2}}.$$

In this section, we provide improvements of this result.

Theorem 1.15 *Let G be an (n,d,λ)-colored graph with the color set D. Let T be a tree with edges colored by D. For any $U \subset V(G)$, with $|U| = r \cdot \frac{\lambda n}{d}$, for some $r > 1$, the number of disjoint copies of H in U is at least*

$$\frac{|U|}{\sigma(T)} - \frac{\lambda n}{d},$$

where $\sigma(T)$ is the stringiness of T.

Theorem 1.15 directly generalizes Soukup's result in [30] to pseudo-random graphs. However, the stringiness of a tree may be exponential in the number of vertices. Using a different method, we prove a theorem which for most trees does much better.

Theorem 1.16 *Let G be an (n,d,λ)-colored graph with the color set D. Let T be a tree of m vertices with edges colored by D. For any $U \subset V(G)$ with $|U| \geq m(m-1) \cdot \frac{\lambda n}{d}$, the number of disjoint copies of T in U is at least*

$$\frac{|U|}{m} - \frac{\lambda n}{d}.$$

2 Proof of Theorem 1.3

Set $V_i = V(G_i)$ and $E_i = E(G_i)$ for $1 \leq i \leq 2$. If i satisfies $\frac{\lambda_i}{d_i} = \max\left\{\frac{\lambda_1}{d_1}, \frac{\lambda_2}{d_2}\right\}$, then throughout the proof we will use $\frac{\lambda}{d}$ to denote $\frac{\lambda_i}{d_i}$.

2.1 Square-norm

For functions $f_1, f_2, f_3, f_4 \colon V_1 \times V_2 \to [-1,1]$, we define

$$\begin{aligned} N(f_1,f_2,f_3,f_4) &:= \mathbb{E}_{\substack{a,b,c,d \\ (a,b)\in E_1,(c,d)\in E_2}} f_1(a,c)f_2(a,d)f_3(b,c)f_4(b,d) \\ &:= \frac{1}{|V_1|d_1|V_2|d_2} \sum_{\substack{a,b,c,d \\ (a,b)\in E_1,(c,d)\in E_2}} f_1(a,c)f_2(a,d)f_3(b,c)f_4(b,d) \end{aligned}$$

and

$$\begin{aligned} M(f_1,f_2,f_3,f_4) &:= \mathbb{E}_{a,b,c,d} f_1(a,c)f_2(a,d)f_3(b,c)f_4(b,d) \\ &:= \frac{1}{|V_1|^2|V_2|^2} \sum_{a,b,c,d} f_1(a,c)f_2(a,d)f_3(b,c)f_4(b,d). \end{aligned}$$

Let S be any subset of $V_1 \times V_2$. Recall that when context is clear, we use $S(\cdot)$ to denote the characteristic function χ_S on the set S. We now prove two simple but useful facts about M using Cauchy–Schwarz.

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

Proposition 2.1

$$M(S,S,S,S) \geq \left(\frac{|S|}{|V_1||V_2|}\right)^4.$$

Proof We write the definition of $|S|$ as a sum and apply Cauchy–Schwarz twice to get

$$\begin{aligned}
|S| &= \sum_{a\in V_1}\sum_{b\in V_2} S(a,b) \leq \left(\sum_{a\in V_1} 1^2\right)^{\frac{1}{2}} \left(\sum_{a\in V_1}\left(\sum_{b\in V_2} S(a,b)\right)^2\right)^{\frac{1}{2}} \\
&= |V_1|^{\frac{1}{2}} \left(\sum_{b\in V_2}\sum_{c\in V_2}\sum_{a\in V_1} S(a,b)S(a,c)\right)^{\frac{1}{2}} \\
&\leq |V_1|^{\frac{1}{2}} \left(\left(\sum_{b\in V_2}\sum_{c\in V_2} 1^2\right)^{\frac{1}{2}} \left(\sum_{b\in V_2}\sum_{c\in V_2}\left(\sum_{a\in V_1} S(a,b)S(a,c)\right)^2\right)^{\frac{1}{2}}\right)^{\frac{1}{2}} \\
&= |V_1|^{\frac{1}{2}} \left(|V_2| \left(\sum_{b\in V_2}\sum_{c\in V_2}\sum_{a\in V_1}\sum_{d\in V_1} S(a,b)S(a,c)S(d,b)S(d,c)\right)^{\frac{1}{2}}\right)^{\frac{1}{2}},
\end{aligned}$$

which, upon rearranging and renaming variables, becomes

$$|V_1|^{\frac{1}{2}}|V_2|^{\frac{1}{2}} \left(|V_1|^2|V_2|^2 M(S,S,S,S)\right)^{\frac{1}{4}}.$$

Comparing this to $|S|$ yields the desired result. ∎

For any function $f\colon V_1 \times V_2 \to [-1,1]$, we define

$$\|f\|_{\square(V_1\times V_2)} := M(f,f,f,f)^{1/4}.$$

Lemma 2.2 *For functions $f_1, f_2, f_3, f_4\colon V_1 \times V_2 \to [-1,1]$, we have*

$$M(f_1,f_2,f_3,f_4) \leq \min_i \|f_i\|_{\square(V_1\times V_2)}.$$

Proof We apply Cauchy–Schwarz to the definition of M to get

$$\begin{aligned}
M(f_1,f_2,f_3,f_4) &= \frac{1}{|V_1|^2|V_2|^2} \sum_{a,b\in V_1, c,d\in V_2} f_1(a,c)f_2(a,d)f_3(b,c)f_4(b,d) \\
&= \frac{1}{|V_1|^2|V_2|^2} \sum_{a,b\in V_1} \left(\sum_{c\in V_2} f_1(a,c)f_3(b,c)\right)\left(\sum_{d\in V_2} f_2(a,d)f_4(b,d)\right) \\
&\leq \frac{1}{|V_1|^2|V_2|^2} \left(\sum_{a,b\in V_1}\left(\sum_{c\in V_2} f_1(a,c)f_3(b,c)\right)^2\right)^{\frac{1}{2}}
\end{aligned}$$

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

$$\cdot \left(\sum_{a,b \in V_1} \left(\sum_{d \in V_2} f_2(a,d) f_4(b,d) \right)^2 \right)^{\frac{1}{2}}$$
$$= (M(f_1, f_1, f_3, f_3))^{\frac{1}{2}} \cdot (M(f_2, f_2, f_4, f_4))^{\frac{1}{2}}.$$

A similar calculation using Cauchy–Schwarz and reversing the roles of V_1 and V_2 gives that

$$M(f_1, f_2, f_3, f_4) \leq (M(f_1, f_2, f_1, f_2))^{1/2} \cdot (M(f_3, f_4, f_3, f_4))^{1/2}.$$

We finish by combining these inequalities and using the fact that $M(f_i, f_i, f_i, f_i) \leq 1$ for $i = 1, 2, 3, 4$. ∎

2.2 A weak hypergraph regularity lemma

Let $\mathcal{B}$ be a σ-algebra on V_1, and let $\mathcal{C}$ be a σ-algebra on V_2. We recall here that a σ-algebra on V_i is a collection of sets in V_i that contains V_i, $\emptyset$, and is closed under finite intersections, unions, and complements.

The complexity of a σ-algebra $\mathcal{B}$ is the smallest number of sets (atoms) needed to generate $\mathcal{B}$, and we denote by $\texttt{complexity}(\mathcal{B})$. Notice that $|\mathcal{B}| \leq 2^{\texttt{complexity}(\mathcal{B})}$. We denote the smallest σ-algebra on $V_1 \times V_2$ that contains both $\mathcal{B} \times V_2$ and $V_1 \times \mathcal{C}$ by $\mathcal{B} \vee \mathcal{C}$.

For a function $f: V_1 \times V_2 \to \mathbb{R}$, we define the conditional expectation $\mathbb{E}(f|\mathcal{B} \vee \mathcal{C}): V \to \mathbb{R}$ by the formula

$$\mathbb{E}(f|\mathcal{B} \vee \mathcal{C})(x) := \frac{1}{|(\mathcal{B} \vee \mathcal{C})(x)|} \sum_{y \in (\mathcal{B} \vee \mathcal{C})(x)} f(y),$$

where $(\mathcal{B} \vee \mathcal{C})(x)$ denotes the smallest element of $\mathcal{B} \vee \mathcal{C}$ that contains x. We note that an atom of $\mathcal{B} \vee \mathcal{C}$ has the form $U \times V$ where U and V are atoms of $\mathcal{B}$ and $\mathcal{C}$, respectively.

The following lemma is a special case of the weak hypergraph regularity lemma due to Lyall and Magyar (Lemma 2.2 in [26]). We refer the reader to [26] for a detailed proof. Rather than stating their lemma in its full generality, which would require a substantial amount of additional background, we state the following special case of their result, which will suffice for our purposes. As before, we let S denote a subset of $V_1 \times V_2$.

Lemma 2.3 (Lemma 2.2 in [26] (special case)) *For any $\varepsilon > 0$, there exist σ-algebras $\mathcal{B}$ on V_1 and $\mathcal{C}$ on V_2 such that each algebra is spanned by at most $O(\varepsilon^{-8})$ sets, and*

$$\|S - \mathbb{E}(S|\mathcal{B} \vee \mathcal{C})\|_{\square(V_1 \times V_2)} \leq \varepsilon.$$

We recall that

$$\mathbb{E}(S|\mathcal{B} \vee \mathcal{C})(x) = \frac{|S \cap (B \times C)|}{|B||C|},$$

where $B \times C$ is the atom of $\mathcal{B} \vee \mathcal{C}$ containing x.

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

2.3 A generalized von-Neumann-type estimate

Lemma 2.4 *For functions $f_1, f_2, f_3, f_4 : V_1 \times V_2 \to [-1,1]$, we have*

$$|N(f_1,f_2,f_3,f_4)| \le \min_j \|f_j\|_{\square(V_1\times V_2)} + O\left(\frac{\lambda^{1/4}}{d^{1/4}}\right).$$

To prove this lemma, we recall the following result, often called the expander mixing lemma, which was proved at least as early as 1980 by Haemers in his Ph.D. thesis [11, Theorem 3.1.1].

Lemma 2.5 *Let $G = (V,E)$ be an (n,d,λ)-graph, and let A be its adjacency matrix. For real $f, g \in L^2(V)$, we have*

$$|\langle f, Ag\rangle - d|V|\mathbb{E}(f)\mathbb{E}(g)| \le \lambda \|f\|_2 \|g\|_2,$$

where

$$\mathbb{E}(f) := \frac{1}{|V|}\sum_{v\in V} f(v),\ \|f\|_2^2 = \sum_{v\in V}|f(v)|^2.$$

Proof of Lemma 2.4 Set $\sigma_i(x,y) = |V_i|/d_i$ if $(x,y) \in E_i$ and 0 otherwise, and let $\mathbb{E}_{x,y\in V_i} := \frac{1}{|V_i|^2}\sum_{x,y\in V_i}$. Then, for $f, g : V_i \to [-1,1]$, by using the expander mixing lemma, one has

$$\sum_{x\sim y} f(x)g(y) = \langle f, Ag\rangle \le \frac{d_i}{|V_i|}\sum_{x,y} f(x)g(y) + \lambda_i \|f\|_2\|g\|_2.$$

Dividing both sides by $d_i|V_i|$ and using $\|f\|_2\|g\|_2 \le |V_i|$ gives

$$\mathbb{E}_{x,y\in V_i} f(x)g(y)\sigma_i(x,y) = \left(\frac{1}{|V_i|^2}\sum_{x,y} f(x)g(y)\right) + \frac{\lambda_i}{d_i} = \mathbb{E}_{x,y} f(x)f(y) + \frac{\lambda_i}{d_i}.$$

Thus,

$$\begin{aligned}|\mathbb{E}_{x,y} f(x)g(y)\sigma_i(x,y)|^2 &= (\mathbb{E}_{x,y,z,t} f(x)g(z)f(y)g(t)) + 2\frac{\lambda_i}{d_i}\mathbb{E}_{x,y} f(x)g(y) + \frac{\lambda_i^2}{d_i^2}\\ &\le \mathbb{E}_{x,y} f(x)f(y) + 3\frac{\lambda_i}{d_i},\end{aligned}$$

where we have used the fact that $\mathbb{E}_{z,t} g(z)g(t), \mathbb{E}_{x,y} f(x)g(y) \le 1$. In other words, for functions $f, g : V_i \to [-1,1]$, we have

$$|\mathbb{E}_{x,y} f(x)g(y)\sigma_i(x,y)|^2 \le \mathbb{E}_{x,y} f(x)f(y) + 3\frac{\lambda_i}{d_i}. \tag{4}$$

The same holds when we switch between f and g:

$$|\mathbb{E}_{x,y} f(x)g(y)\sigma_i(x,y)|^2 \le \mathbb{E}_{x,y} g(x)g(y) + 3\frac{\lambda_i}{d_i}. \tag{5}$$

In the next step, we want to show that

$$N(f_1,f_2,f_3,f_4) \le \|f_1\|_{\square(V_1\times V_2)} + O\left(\frac{\lambda}{d}\right).$$

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

Using the definitions, we have

$$N(f_1,f_2,f_3,f_4)=\mathbb{E}_{a,b,c,d}f_1(a,c)f_2(a,d)f_3(b,c)f_4(b,d)\sigma_1(a,b)\sigma_2(c,d).$$

For a fixed pair (c,d), set $f_{c,d}(a)=f_1(a,c)f_2(a,d)$ and $g_{c,d}(b)=f_3(b,c)f_4(b,d)$. Then we have

$$\begin{aligned}
|N(f_1,f_2,f_3,f_4)|^2 &= \left(\frac{1}{|V_1|^2|V_2|^2}\sum_{a,b,c,d} f_{c,d}(a)g_{c,d}(b)\sigma_1(a,b)\sigma_2(c,d)\right)^2\\
&\le \left(\frac{1}{|V_1|^2|V_2|^2}\sum_{c,d}\sqrt{\sigma_2(c,d)}\left|\sqrt{\sigma_2(c,d)}\sum_{a,b} f_{c,d}(a)g_{c,d}(b)\sigma_1(a,b)\right|\right)^2\\
&\le \frac{1}{|V_1|^4|V_2|^4}\left(\sum_{c,d}\sigma_2(c,d)\right)\left(\sum_{c,d}\sigma_2(c,d)\left|\sum_{a,b} f_{c,d}(a)g_{c,d}(b)\sigma_1(a,b)\right|^2\right)\\
&= \left(\mathbb{E}_{c,d}\sigma_2(c,d)\right)\left(\mathbb{E}_{c,d}\sigma_2(c,d)|\mathbb{E}_{a,b}f_{c,d}(a)g_{c,d}(b)\sigma_1(a,b)|^2\right)\\
&= \mathbb{E}_{c,d}\sigma(c,d)|\mathbb{E}_{a,b}f_{c,d}(a)g_{c,d}(b)\sigma(a,b)|^2,
\end{aligned}$$

where the first inequality uses the triangle inequality and rearranging, the second inequality is Cauchy–Schwarz, the next line is rearranging, and the last equality uses the fact that $\mathbb{E}_{c,d}\sigma_2(c,d)=\frac{1}{|V_2|^2}\cdot\frac{|V_2|}{d_2}\cdot|V_2|\cdot d_2=1$. Therefore, the inequality (4) implies

$$\begin{aligned}
(6)\quad |N(f_1,f_2,f_3,f_4)|^2 &\le \mathbb{E}_{c,d}\sigma_2(c,d)\left(\mathbb{E}_{a,b}f_{c,d}(a)f_{c,d}(b)+3\frac{\lambda_1}{d_1}\right)\\
&= \mathbb{E}_{a,b,c,d}\sigma_2(c,d)f_{c,d}(a)f_{c,d}(b)+3\left(\mathbb{E}_{c,d}\sigma_2(c,d)\right)\frac{\lambda_1}{d_1}\\
&= \left(\mathbb{E}_{a,b,c,d}f_1(a,c)f_1(b,c)f_2(a,d)f_2(b,d)\sigma_2(c,d)\right)+3\frac{\lambda_1}{d_1}.
\end{aligned}$$

By another similar argument with $\hat{f}_{a,b}(c)=f_1(a,c)f_1(b,c)$ and $\hat{g}_{a,b}(d)=f_2(a,d)f_2(b,d)$ for each fixed pair (a,b), we have

$$\begin{aligned}
|N(f_1,f_2,f_3,f_4)|^4 &\le \left(\mathbb{E}_{a,b,c,d}\hat{f}_{a,b}(c)\hat{g}_{a,b}(d)\sigma_2(c,d)+3\frac{\lambda_1}{d_1}\right)^2\\
&\le \left(\mathbb{E}_{a,b,c,d}\hat{f}_{a,b}(c)\hat{g}_{a,b}(d)\sigma_2(c,d)\right)^2+15\frac{\lambda_1}{d_1},
\end{aligned}$$

using that $\mathbb{E}_{a,b,c,d}\hat{f}_{a,b}(c)\hat{g}_{a,b}(d)\sigma_2(c,d)\le 1$ and $\lambda_1/d_1\le 1$. Now

$$\begin{aligned}
\left(\mathbb{E}_{a,b,c,d}\hat{f}_{a,b}(c)\hat{g}_{a,b}(d)\sigma_2(c,d)\right)^2 &= \frac{1}{|V_1|^4|V_2|^4}\left(\sum_{a,b}1\sum_{c,d}\hat{f}_{a,b}(c)\hat{g}_{a,b}(d)\sigma_2(c,d)\right)^2\\
&\le \frac{1}{|V_1|^2}\sum_{a,b}\left(\frac{1}{|V_2|^2}\sum_{c,d}\hat{f}_{a,b}(c)\hat{g}_{a,b}(d)\sigma_2(c,d)\right)^2\\
&\le \frac{1}{|V_1|^2}\sum_{a,b}\left(\mathbb{E}_{c,d}\hat{f}_{a,b}(c)\hat{f}_{a,b}(d)+3\frac{\lambda_2}{d_2}\right)\\
&= \mathbb{E}_{a,b}\left(\mathbb{E}_{c,d}f_1(a,c)f_1(b,c)f_1(a,d)f_1(b,d)+3\frac{\lambda_2}{d_2}\right),
\end{aligned}$$

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

by Cauchy–Schwarz and (4), respectively. As a consequence, we obtain

$$|N(f_1,f_2,f_3,f_4)|^4 \le \mathbb{E}_{a,b,c,d} f_1(a,c)f_1(b,c)f_1(a,d)f_1(b,d) + 15\frac{\lambda_1}{d_1} + 3\frac{\lambda_2}{d_2}$$
$$= M(f_1,f_1,f_1,f_1) + O\left(\frac{\lambda_1}{d_1} + \frac{\lambda_2}{d_2}\right).$$

Notice that the same holds when f_1 on the right-hand side is replaced by f_i for $2 \le i \le 4$. In short,

$$|N(f_1,f_2,f_3,f_4)| \le \min_j \|f_j\|_{\square(V_1\times V_2)} + O\left(\frac{\lambda^{1/4}}{d^{1/4}}\right).$$

This completes the proof. ∎

With Lemmas 2.3 and 2.4 in hand, we are ready to prove Theorem 1.3.

Proof of Theorem 1.3:

For any $\varepsilon > 0$, by Lemma 2.3, we can see that there exist σ-algebras $\mathcal{B}$ and $\mathcal{C}$ on V_1 and V_2, respectively, with complexity bounded above by $O\left(\varepsilon^{-8}\right)$, so that

$$\|S - \mathbb{E}(S|\mathcal{B}\vee\mathcal{C})\|_{\square(V_1\times V_2)} \le \varepsilon. \tag{7}$$

Let g denote $\mathbb{E}(S|\mathcal{B}\vee\mathcal{C})$, and define

$$h(x) := S(x) - g(x).$$

Therefore, (7) gives us that

$$\|h\|_{\square(V_1\times V_2)} \le \varepsilon. \tag{8}$$

Both g and h are functions from $V_1 \times V_2$ to the interval $[-1,1]$. Recalling the definition of N above, we see that

$$N(S,S,S,S) = N(g,g,g,g) + N(h,h,h,h) + R,$$

where R is a sum over all expressions of the form

$$N(f_1,f_2,f_3,f_4),$$

where the f_j in each term are either g or h, but not all the same. Specifically, set $\Omega := \{g,h\}^4\backslash\{(g,g,g,g),(h,h,h,h)\}$, denote a quadruple of functions by $F = (f_1,f_2,f_3,f_4) \in \Omega$, and write

$$R = \sum_{F\in\Omega} \mathbb{E}_{a,b,c,d} f_1(a,c)f_2(a,d)f_3(b,c)f_4(b,d)E_1(a,b)E_2(c,d),$$

where $E_i(x,y)$ is the indicator that $xy \in E(G_i)$. Combining Lemma 2.4 and (8) gives

$$|N(h,h,h,h)| \le \|h\|_{\square(V_1\times V_2)} + O\left(\frac{\lambda^{1/4}}{d^{1/4}}\right) \le \varepsilon + O\left(\frac{\lambda^{1/4}}{d^{1/4}}\right).$$

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

Similarly, for any other choice of $F \in \Omega$, we must have h in at least one entry, so we get

$$|N(F)| \leq \min_j \|f_j\|_{\square(V_1\times V_2)} + O\left(\frac{\lambda^{1/4}}{d^{1/4}}\right) \leq \|h\|_{\square V_1\times V_2} + O\left(\frac{\lambda^{1/4}}{d^{1/4}}\right) \leq \varepsilon + O\left(\frac{\lambda^{1/4}}{d^{1/4}}\right).$$

Putting these together, we get that

$$|N(S,S,S,S) - N(g,g,g,g)| = O\left(\varepsilon + \frac{\lambda^{1/4}}{d^{1/4}}\right). \tag{9}$$

Similarly, by Lemma 2.2, we know that

$$M(F) \leq \min_i \|f_i\|_{\square(V_1\times V_2)},$$

so we get that

$$|M(S,S,S,S) - M(g,g,g,g)| = O(\varepsilon). \tag{10}$$

By definition, g is a linear combination of indicator functions of atoms of the σ-algebra $\mathcal{B}\vee\mathcal{C}$. By Lemma 2.3, we know that there is some positive constant $c>0$ so that the number of terms in this linear combination is no more than $2^{c\varepsilon^{-8}}$. So we can write $N(g,g,g,g)$ as a linear combination of terms of the form

$$\begin{aligned}&N(B_1\times C_1, B_2\times C_2, B_3\times C_3, B_4\times C_4)\\ &= \mathbb{E}_{a,b,c,d}(B_1\times C_1)(a,c)\cdot(B_2\times C_2)(a,d)\cdot(B_3\times C_3)(b,c)\cdot(B_4\times C_4)(b,d)\sigma_1(a,b)\sigma_2(c,d),\end{aligned}$$

for some atoms $B_j\times C_j$ (and their indicator functions) in $\mathcal{B}\times\mathcal{C}$. Here as before we use $\sigma_i(x,y)$ equals $|V_i|/d_i$ if $\{x,y\}\in E_i$ and 0 otherwise. However, if we split this up by variables, we get that

$$\begin{aligned}&N(B_1\times C_1, B_2\times C_2, B_3\times C_3, B_4\times C_4)\\ &= \mathbb{E}_{a,b,c,d}(B_1\cap B_2)(a)\cdot(B_3\cap B_4)(b)\cdot(C_1\cap C_3)(c)\cdot(C_2\cap C_4)(d)\sigma_1(a,b)\sigma_2(c,d)\\ &= \left(\mathbb{E}_{a,b}(B_1\cap B_2)(a)\cdot(B_3\cap B_4)(b)\sigma_1(a,b)\right)\left(\mathbb{E}_{c,d}(C_1\cap C_3)(c)\cdot(C_2\cap C_4)(d)\sigma_2(c,d)\right).\end{aligned}$$

By applying the expander mixing lemma as in the proof of Lemma 2.4, we see

$$\begin{aligned}&N(B_1\times C_1, B_2\times C_2, B_3\times C_3, B_4\times C_4)\\ &= \left(\mathbb{E}_{a,b}(B_1\cap B_2)(a)\cdot(B_3\cap B_4)(b) + O\left(\frac{\lambda_1}{d_1}\right)\right)\left(\mathbb{E}_{c,d}(C_1\cap C_3)(c)\cdot(C_2\cap C_4)(d) + O\left(\frac{\lambda_2}{d_2}\right)\right)\\ &= M(B_1\times C_1, B_2\times C_2, B_3\times C_3, B_4\times C_4) + O\left(\frac{\lambda}{d}\right),\end{aligned}$$

where the last line uses the definition of M and that each expectation is at most 1. Since g is a linear combination of at most $2^{c\varepsilon^{-8}}$ terms, we see that

$$|N(g,g,g,g) - M(g,g,g,g)| = O\left(2^{c'\varepsilon^{-8}}\frac{\lambda}{d}\right),$$

for some positive constant c'. Using (9) followed by the previous estimate and (10), we get that for some constant $k>0$, we have

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

$$\begin{aligned} N(S,S,S,S) &\geq N(g,g,g,g) - k\varepsilon - k\frac{\lambda^{1/4}}{d^{1/4}} \\ &\geq M(g,g,g,g) - k2^{c'\varepsilon^{-8}}\frac{\lambda}{d} - k\varepsilon - k\frac{\lambda^{1/4}}{d^{1/4}} \\ &\geq M(S,S,S,S) - k\varepsilon - k2^{c\varepsilon^{-8}}\frac{\lambda}{d} - k\varepsilon - k\frac{\lambda^{1/4}}{d^{1/4}}. \end{aligned}$$

Now applying Proposition 2.1 to this estimate gives us

$$N(S,S,S,S) \geq \left(\frac{|S|}{|V_1||V_2|}\right)^4 - 2k\varepsilon - k2^{c\varepsilon^{-8}}\frac{\lambda}{d} - k\frac{\lambda^{1/4}}{d^{1/4}}. \tag{11}$$

Recall that by assumption, $|S| \geq \delta|V_1||V_2|$, so to guarantee that $N = N(S,S,S,S)$ is positive, we just need to pick ε so that the right-hand side of (11) is bigger than $\delta' 4$, or equivalently,

$$\delta^4 - \delta' 4 \geq 2k\varepsilon + k2^{c\varepsilon^{-8}}\frac{\lambda}{d} + k\frac{\lambda^{1/4}}{d^{1/4}}.$$

3 Proof of Theorem 1.9

To prove Theorem 1.9, we present two approaches based on two counting lemmas. While the second counting lemma is a direct consequence of the expander mixing lemma for a single graph, the first counting lemma is a stronger and more practical variant for tensor of two pseudo-random graphs, which is quite interesting on its own.

3.1 The first counting lemma for cycles

Let us briefly describe the ideas of counting cycles here. Assume that we want to count the number of cycles of length $2k$ for some integer $k \geq 2$. Given four vertices x, y, z, w, if x and y are connected by a path of length $k-1$, and the same happens for z and w, then we will have a cycle of length $2k$ of the form $x - yw - zx$ (Figure 1) when there are edges between x and z, and between y and w. Thus, the problem is reduced to counting the number of pairs of edges between the endpoints of pairs of paths of length $k-1$.

To this end, we make use of the notation of tensor of two pseudo-random graphs. For two graphs $\mathcal{G}_1 = (V_1, E_1)$ and $\mathcal{G}_2 = (V_2, E_2)$, the tensor product $\mathcal{G}_1 \otimes \mathcal{G}_2$ is a graph with vertex set $V(\mathcal{G}_1 \otimes \mathcal{G}_2) = V_1 \times V_2$, and there is an edge between (u,v) and (u',v') if and only if $(u,u') \in E_1$ and $(v,v') \in E_2$. Suppose that the adjacency matrices of $\mathcal{G}_1$ and $\mathcal{G}_2$ are A and B, respectively, then the adjacency matrix of $\mathcal{G}_1 \otimes \mathcal{G}_2$ is the tensor product of A and B. It is well known that if $\gamma_1, \dots, \gamma_n$ are eigenvalues of A and $\gamma'_1, \dots, \gamma'_m$ are eigenvalues of B, then the eigenvalues of $A \otimes B$ are $\gamma_i \gamma'_j$ with $1 \leq i \leq n$, $1 \leq j \leq m$ (see [27] for more details).

It is not hard to use the expander mixing lemma to get the following.

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

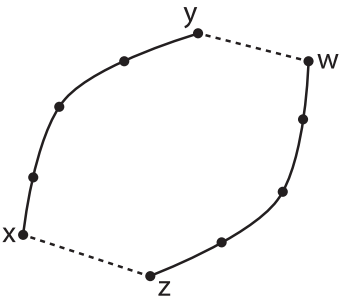


Figure 1: Counting pairs of edges xz and yw.

Proposition 3.1 *Let $\mathcal{G}$ be an (n,d,λ)-graph. For two nonnegative functions $f,g\colon V\times V\to\mathbb{R}$, we have*

$$\left|\sum_{(x,z)\in E,(y,w)\in E} f(x,y)g(z,w)-\frac{d^2}{n^2}\|f\|_1\|g\|_1\right|\le d\lambda\|f\|_2\|g\|_2.$$

Our first counting lemma offers better bounds as follows.

Proposition 3.2 *(First counting lemma) Let $\mathcal{G}$ be an (n,d,λ)-graph. For two nonnegative functions $f,g\colon V\times V\to\mathbb{R}$, we define $F(x)=\sum_y f(x,y)$, $G(z)=\sum_w g(z,w)$, $F'(y)=\sum_x f(x,y)$, and $G'(w)=\sum_z g(z,w)$. Then we have*

$$\left|\sum_{(x,z)\in E,(y,w)\in E} f(x,y)g(z,w)-\frac{d^2}{n^2}\|f\|_1\|g\|_1\right|\le\lambda^2\|f\|_2\|g\|_2 + \frac{d\lambda}{n}\left(\|F\|_2\|G\|_2+\|F'\|_2\|G'\|_2\right).$$

Proof Suppose $\mathcal{G}$ is a d-regular graph on vertex set V with $|V|=n$, and let A denote its adjacency matrix. For two real-valued functions $f,g\colon V\times V\to\mathbb{R}$, we define

$$\langle f,g\rangle=\sum_{(v_1,v_2)\in V\times V} f(v_1,v_2)g(v_1,v_2)$$

and

$$\|f\|_2^2=\langle f,f\rangle.$$

We denote the set of all real-valued functions on $V\times V$ by $L^2(V\times V)$. For the remainder of the proof, we will assume that $f,g\in L^2(V\times V)$ are nonnegative functions.

We define

$$A\otimes Af(v_1,v_2)=\sum_{(u_1,u_2):(u_1,v_1)\in E,(u_2,v_2)\in E} f(u_1,u_2).$$

That is, $A\otimes A$ is the adjacency matrix of $\mathcal{G}\otimes\mathcal{G}$. In the remainder, we denote $A\otimes A$ by B.

Let $\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_n$ be the eigenvalues of A corresponding to eigenfunctions $e_1,\dots,e_n$. Without loss of generality, assume that the e_i form an orthonormal basis of $\mathbb{R}^n$. Then the eigenfunctions of B are exactly $e_i\otimes e_j$ for all $1\le i,j\le n$ corresponding to eigenvalue $\lambda_{ij}:=\lambda_i\lambda_j$.

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

We observe that

$$f = \sum_{i,j} \langle f, e_i \otimes e_j \rangle e_i \otimes e_j.$$

So

$$Bg = \sum_{i,j} \langle Bg, e_i \otimes e_j \rangle e_i \otimes e_j = \sum_{e_i \otimes e_j} \lambda_{ij} \langle g, e_i \otimes e_j \rangle e_i \otimes e_j.$$

We note that A has a constant eigenfunction that will be denoted by e_1, i.e.,

$$e_1(v) = 1/\sqrt{n}, \ \forall v \in V.$$

This means that B also has constant eigenfunction defined by

$$e_1 \otimes e_1(u, v) = 1/n \ \ \forall (u, v) \in V \times V.$$

We have

$$\sum_{(x,z)\in E,(y,w)\in E} f(x, y)g(z, w) = \langle f, Bg \rangle = \sum_{i,j} \lambda_{ij} \langle g, e_i \otimes e_j \rangle \langle f, e_i \otimes e_j \rangle.$$

Define

$$\begin{aligned} S_1 &:= \lambda_{11} \langle g, e_1 \otimes e_1 \rangle \langle f, e_1 \otimes e_1 \rangle, \\ S_2 &:= \sum_{j=2}^{n} \lambda_{1j} \langle g, e_1 \otimes e_j \rangle \langle f, e_1 \otimes e_j \rangle, \\ S_3 &:= \sum_{i=2}^{n} \lambda_{i1} \langle g, e_i \otimes e_1 \rangle \langle f, e_i \otimes e_1 \rangle, \\ S_4 &:= \sum_{i,j=2}^{n} \lambda_{ij} \langle g, e_i \otimes e_j \rangle \langle f, e_i \otimes e_j \rangle. \end{aligned}$$

And so

$$\sum_{(x,z)\in E,(y,w)\in E} f(x, y)g(z, w) - S_1 = S_2 + S_3 + S_4.$$

We now estimate each S_i. Since $\lambda_1 = d$ and e_1 is constant, it is easy to see that

$$S_1 = \lambda_{11} \left\langle f, \frac{1}{n}\mathbf{1} \right\rangle \left\langle g, \frac{1}{n}\mathbf{1} \right\rangle = \frac{d^2}{n^2} \|f\|_1 \|g\|_1.$$

For S_4, if $i, j > 1$, we have that $\lambda_{ij} \le \lambda^2$ and hence

$$\begin{aligned} S_4 \le \lambda^2 \sum_{i,j=2}^{n} \langle g, e_i \otimes e_j \rangle \langle f, e_i \otimes e_j \rangle &\le \lambda^2 \left(\sum_{i,j=2}^{n} \langle g, e_i \otimes e_j \rangle^2 \right)^{1/2} \left(\sum_{i,j=2}^{n} \langle f, e_i \otimes e_j \rangle^2 \right)^{1/2} \\ &\le \lambda^2 \left(\sum_{i,j=1}^{n} \langle g, e_i \otimes e_j \rangle^2 \right)^{1/2} \left(\sum_{i,j=1}^{n} \langle f, e_i \otimes e_j \rangle^2 \right)^{1/2} \\ &= \lambda^2 \|f\|_2 \|g\|_2, \end{aligned}$$

where the second inequality follows by Cauchy–Schwarz.

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

To estimate S_2, note that $\lambda_{1j} \le \lambda d$, and

$$e_1 \otimes e_j(v_1, v_2) = \frac{1}{\sqrt{n}} e_j(v_2).$$

Using Cauchy–Schwarz, we have that

$$S_2 \le \lambda d \sum_{j=2}^{n} \langle g, e_1 \otimes e_j\rangle \langle f, e_1 \otimes e_j\rangle \le \lambda d \left(\sum_{j=2}^{n} \langle g, e_1 \otimes e_j\rangle^2\right)^{1/2} \left(\sum_{j=2}^{n} \langle f, e_1 \otimes e_j\rangle^2\right)^{1/2}$$
$$\le \lambda d \left(\sum_{j=1}^{n} \langle g, e_1 \otimes e_j\rangle^2\right)^{1/2} \left(\sum_{j=1}^{n} \langle f, e_1 \otimes e_j\rangle^2\right)^{1/2}.$$

To estimate this quantity, note that

$$\langle g, e_1 \otimes e_j\rangle = \sum_{u,v} g(u,v) e_1 \otimes e_j(u,v) = \frac{1}{\sqrt{n}} \sum_{u,v} g(u,v) e_j(v) = \frac{1}{\sqrt{n}} \sum_{v} G'(v) e_j(v),$$

and similarly $\langle f, e_1 \otimes e_j\rangle = \frac{1}{\sqrt{n}} \sum_v F'(v) e_j(v)$. Therefore, we have that

$$\sum_{j=1}^{n} \langle g, e_1 \otimes e_j\rangle^2 = \sum_{j=1}^{n} \frac{1}{n} \sum_{u,v} G'(u) G'(v) e_j(u) e_j(v) = \frac{1}{n} \sum_{u,v} \left(G'(u) G'(v) \sum_{j=1}^{n} e_j(u) e_j(v) \right).$$

Now notice that because the e_i form an orthonormal basis, we have that

$$\sum_{j=1}^{n} e_j(u) e_j(v) = \begin{cases} 1, & u = v, \\ 0, & u \ne v. \end{cases}$$

Hence, we have

$$\sum_{j=1}^{n} \langle g, e_1 \otimes e_j\rangle^2 = \frac{1}{n} \sum_{u=1}^{n} \left((G'(u))^2 \sum_{j=1}^{n} e_j(u)^2 \right) = \frac{1}{n} \sum_{u=1}^{n} (G'(u))^2 = \frac{1}{n} \|G'\|_2^2.$$

Similarly, $\sum_{j=1}^{n} \langle f, e_1 \otimes e_j\rangle^2 = \frac{1}{n} \|F'\|_2^2$. Combining everything, we have that

$$S_2 \le \frac{\lambda d}{n} \|G'\|_2 \|F'\|_2.$$

A symmetric proof shows that

$$S_3 \le \frac{\lambda d}{n} \|G\|_2 \|F\|_2.$$

■

3.2 The second counting lemma for cycles

Assume that we want to count the number of cycles of length $2k$ for some integer $k \ge 1$. Our second strategy for cycles can be explained as follows. Given three vertices x, y, and z, if x and y are connected by a path of length k, and x and z are connected by a path of length $k-1$, then we have a cycle of length $2k$ of the form $x - yz - x$ if and only if y and z are adjacent (Figure 2). So the problem is reduced to counting the number of edges between the endpoints of pairs of paths pinned at a vertex.

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

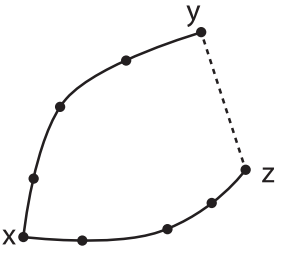


Figure 2: Counting edges yz.

Proposition 3.3 (Second counting lemma) *Let $\mathcal{G}$ be an (n,d,λ)-graph. Let U be a set of vertices in $\mathcal{G}$. For any two vertices x and y, let $p_k(x,y)$ be the number of paths of length k between x and y with vertices in between belonging to U. Then we have*

$$\left|C_{2k+1}(U) - \frac{d}{n}\sum_{x\in U}\left(\sum_{y\in U} p_k(x,y)\right)^2\right| \le \lambda \sum_{x,y\in U} p_k(x,y)^2$$

and

$$\left|C_{2k}(U) - \frac{d}{n}\sum_{x\in U}\left(\sum_{y\in U} p_k(x,y)\right)\cdot\left(\sum_{z\in U} p_{k-1}(x,z)\right)\right| \le \lambda \sum_{x\in U}\left(\sum_{y\in U} p_k(x,y)^2\right)^{1/2} \cdot\left(\sum_{z\in U} p_{k-1}(x,z)^2\right)^{1/2}.$$

Proof We first observe that the number of odd cycles of length $2k+1$ in U is equal to the sum

$$\sum_{x,y,z\in U^3,(y,z)\in E(G)} p_k(x,y)p_k(x,z).$$

Given $x \in U$, set $f(y) = U(y)p_k(x,y)$, then the above sum can be rewritten as

$$\sum_{x\in U}\sum_{(y,z)\in E(G)} f(y)f(z).$$

Applying Lemma 2.5, the first statement is proved.

For the second statement, as above, the number of even cycles of length $2k$ in U is equal to the sum

$$\sum_{x,y,z\in U^3,(y,z)\in E(G)} p_k(x,y)p_{k-1}(x,z).$$

Given $x \in U$, set $f(y) = U(y)p_k(x,y)$ and $g(z) = U(z)p_{k-1}(x,z)$, then the above sum can be rewritten as

$$\sum_{x\in U}\sum_{(y,z)\in E(G)} f(y)g(z).$$

Applying Lemma 2.5, the proposition is proved. ∎

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

Using the facts that

$$\sum_{x\in U}\left(\sum_{y\in U} p_k(x,y)\right)^2 = P_{2k}(U),$$

$$\sum_{x,y\in U} p_k(x,y)^2 = C_{2k}(U),$$

$$\sum_{x\in U}\left(\sum_{y\in U} p_k(x,y)\right)\cdot\left(\sum_{z\in U} p_{k-1}(x,z)\right) = P_{2k-1}(U),$$

and the following application of Cauchy–Schwarz,

$$\sum_{x\in U}\left(\sum_{y\in U} p_k(x,y)^2\right)^{1/2}\cdot\left(\sum_{z\in U} p_{k-1}(x,z)^2\right)^{1/2} \le \left(C_{2k}(U)C_{2k-2}(U)\right)^{1/2},$$

one derives the following corollary.

Corollary 3.4 *Let $\mathcal{G}$ be an (n,d,λ)-graph. Let U be a set of vertices in $\mathcal{G}$. Then*

$$\left|C_{2k+1}(U) - \frac{d}{n}P_{2k}(U)\right| \le \lambda C_{2k}(U)$$

and

$$\left|C_{2k}(U) - \frac{d}{n}P_{2k-1}(U)\right| \le \lambda\left(C_{2k}(U)C_{2k-2}(U)\right)^{1/2}.$$

3.3 Distribution of paths

We have seen that to apply the two counting lemmas, we need to have estimates on the paths of a given length in a vertex set. We now provide relevant results on paths.

Proposition 3.5 *Let $\mathcal{G}$ be an (n,d,λ)-graph, let $k\ge 1$ be an integer, and let U be a vertex set with $\lambda\cdot\frac{n}{d} = o(|U|)$. Let $P_k(U)$ denote the number of paths of length k in U. Then we have*

$$P_k(U) = (1+o(1))\,\frac{|U|^{k+1}d^k}{n^k}.$$

Proof We first prove the following two estimates:

$$\left|P_{2k+1}(U) - \frac{dP_k(U)^2}{n}\right| \le \lambda P_{2k}(U) \tag{12}$$

and

$$\left|P_{2k}(U) - \frac{dP_k(U)P_{k-1}(U)}{n}\right| \le \lambda\sqrt{P_{2k}(U)P_{2k-2}(U)}. \tag{13}$$

For $u\in U$, let $f(u)$ be the number of paths of length k of the form $(u_1,\dots,u_k,u)$ where $u_i\in U$. Similarly, for $v\in U$, let $g(v)$ be the number of paths of length k of the

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

form $(v_1, \dots, v_k, v)$ where $v_i \in U$. To use Lemma 2.5, we need to estimate the norms and the inner product. We have that the adjacency matrix A acts on g by the formula

$$Ag(u) = \sum_{(u,v)\in E(\mathcal{G})} g(v),$$

which is the number of paths of length $k+1$ of the form $(v_1, \dots, v_k, v, u)$. For the inner product, we have

$$\langle f, Ag\rangle = \sum_{u\in V(\mathcal{G})} f(u)Ag(u) = \sum_{u\in V(\mathcal{G})} f(u)Ag(u) = P_{2k+1}(U).$$

It is clear that

$$\mathbb{E}(f) = \mathbb{E}(g) = \frac{1}{|V|}\cdot P_k(U)$$

and

$$||f||_2^2 = ||g||_2^2 = P_{2k}(U).$$

Applying Lemma 2.5, we have that

$$\left|P_{2k+1}(U) - d|V|\left(\frac{1}{|V|}\cdot P_k(U)\right)^2\right| \le \lambda P_{2k}(U),$$

which is equivalent to (12). The estimate (13) also follows from a similar argument with the same f and $g(v)$ defined to be the number of paths of length $k-1$ of the form $(v_1, \dots, v_{k-1}, v)$.

We now proceed by induction on k. The case $k = 0$ is trivial, and the case $k = 1$ follows from Lemma 2.5 and the estimate (13).

Suppose that the statement holds for all $2k \ge 1$. We now show that it also holds for $2k+1$ and $2k+2$. Indeed, it follows from the estimate (12) and induction hypothesis that we have

$$\begin{aligned} P_{2k+1}(U) &\le \frac{d}{n}P_k(U)^2 + \lambda P_{2k}(U) \\ &\le \frac{d}{n}\frac{|U|^{2k+2}d^{2k}}{n^{2k}}\left(1+O\left(\frac{\lambda n}{d|U|}\right)\right)^2 + \lambda\frac{|U|^{2k+1}d^{2k}}{n^{2k}}\left(1+O\left(\frac{\lambda n}{d|U|}\right)\right) \\ &= \frac{|U|^{2k+2}d^{2k+1}}{n^{2k+1}}\left(1+O\left(\frac{\lambda n}{d|U|}\right)\right) \end{aligned}$$

whenever $|U| \gg \lambda\frac{n}{d}$. The lower bound follows in the same way.

For the case $2k+2$, it also follows from the estimate (13) that

$$P_{2k+2}(U) \le \frac{dP_k(U)P_{k+1}(U)}{n} + \lambda\sqrt{P_{2k}(U)P_{2k+2}(U)}.$$

Solving this inequality in $x = \sqrt{P_{2k+2}(U)}$, we obtain

$$P_{2k+2}(U) \le \left(\frac{-\lambda\sqrt{P_{2k}(U)} + \sqrt{\lambda^2 P_{2k}(U) + \frac{4dP_k(U)P_{k+1}(U)}{n}}}{2}\right)^2.$$

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

Using the induction hypothesis and that $\frac{\lambda n}{d} = o(|U|)$, we have that

$$\lambda^2 P_{2k}(U) = o\left(\frac{dP_k(U)P_{k+1}(U)}{n} \cdot \frac{\lambda n}{d|U|}\right)$$

and that

$$\lambda\sqrt{P_{2k}(U)}\sqrt{\frac{dP_k(U)P_{k+1}(U)}{n}} = O\left(\frac{|U|^{2k+3}d^{2k+2}}{n^{2k+2}} \cdot \frac{\lambda n}{d|U|}\right).$$

Hence, the entire expression is bounded above by

$$\frac{|U|^{2k+3}d^{2k+2}}{n^{2k+2}}\left[1 + O\left(\frac{\lambda n}{d|U|}\right)\right].$$

Using lower bounds of the estimates (12) and (13) and an identical argument also gives us

$$P_k(U) \geq \left[1 - O\left(\frac{\lambda n}{d|U|}\right)\right]|U|^{k+1}\left(\frac{d}{n}\right)^k,$$

under the condition $\lambda\frac{n}{d} = o(|U|)$. This completes the proof of the proposition. ∎

3.4 Proof of Theorem 1.9 using the first counting lemma

Proof of Theorem 1.9 We proceed by induction on m.

We first start with the base case $m = 4$.

Let $f, g : U \times U \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} 1, & \text{if } (x,y) \in E(\mathcal{G}), \\ 0, & \text{otherwise} \end{cases}$$

and

$$g(z,w) = \begin{cases} 1, & \text{if } (z,w) \in E(\mathcal{G}), \\ 0, & \text{otherwise.} \end{cases}$$

It is clear that $C_4(U) = \sum_{(x,z),(y,w)\in E(\mathcal{G})} f(x,y)g(z,w)$. To apply Proposition 3.2, we need to check the norms of functions f, g, F, G, F', and G'.

Using Proposition 3.5, we have

$$\|f\|_1 = \sum_{x,y\in U} f(x,y) = P_1(U) = (1+o(1))\,|U|^2\frac{d}{n},$$

$$\|f\|_2 = \left(\sum_{x,y\in U} f(x,y)^2\right)^{1/2} = \left(\sum_{x,y\in U} f(x,y)\right)^{1/2} = (P_1(U))^{1/2} = (1+o(1))\,|U|\sqrt{\frac{d}{n}},$$

using the Taylor series for $\sqrt{1+x}$ and the assumption that $\frac{\lambda n}{d} = o(|U|)$. Similarly,

$$\|g\|_1 = (1+o(1))\,|U|^2\frac{d}{n}, \text{ and } \|g\|_2 = (1+o(1))\,|U|\sqrt{\frac{d}{n}}.$$

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

For functions F, G, F', G' defined as in Proposition 3.2, we have that

$$||F||_2 = \left(\sum_{x\in U} F(x)^2\right)^{1/2} = (P_2(U))^{1/2} = \left(\frac{|U|^3 d^2}{n^2}(1+o(1))\right)^{1/2}.$$

Similarly,

$$||G||_2 = ||F'||_2 = ||G'||_2 = \left(\frac{|U|^3 d^2}{n^2}(1+o(1))\right)^{1/2}.$$

Substituting these estimates into Proposition 3.2, we have that

$$\left|C_4(U) - (1+o(1))\,|U|^4\frac{d^4}{n^4}\right| \le \frac{\lambda^2|U|^2 d}{n}\left(1+O\left(\frac{\lambda n}{d|U|}\right)\right) + \frac{2\lambda|U|^3 d^3}{n^3}\left(1+O\left(\frac{\lambda n}{d|U|}\right)\right).$$

Using the assumption that $\frac{\lambda n}{d} = o(|U|)$ completes this case.

Assume that the statement holds for any cycle of length smaller than $m-1$, we now show that it holds for cycles of length m.

We fall into two cases:

Case 1: $m = 2k+1$.

As above, for $x, y \in U$, we define

$$f(x,y) = \text{the number of paths of length } k \text{ between } x \text{ and } y$$

and

$$g(x,y) = \text{the number of paths of length } k-1 \text{ between } x \text{ and } y.$$

Then

$$||f||_1 = P_k(U) = (1+o(1))\,\frac{d^k}{n^k}|U|^{k+1},$$
$$||f||_2^2 = C_{2k}(U) = (1+o(1))\frac{|U|^{2k}d^{2k}}{n^{2k}} + O\left(\frac{\lambda^{2k-2}|U|^2 d}{n}\right),$$
$$||F||_2^2 = ||F'||_2^2 = P_{2k}(U) = (1+o(1))\frac{d^{2k}}{n^{2k}}|U|^{2k+1},$$

where we left an error term out of our estimate of $C_{2k}(U)$ because it is dominated by the main term by our assumption that $\frac{\lambda n}{d} = o(|U|)$. Similarly, we have

$$||g||_1 = P_{k-1}(U) = (1+o(1))\,\frac{d^{k-1}}{n^{k-1}}|U|^k,$$
$$||g||_2^2 = C_2(U) = (1+o(1))\frac{|U|^2 d}{n} \quad \text{if } m = 5,$$
$$||g||_2^2 = C_{2k-2}(U) = (1+o(1))\frac{|U|^{2k-2}d^{2k-2}}{n^{2k-2}} + O\left(\frac{\lambda^{2k-4}|U|^2 d}{n}\right) \quad \text{if } m \ge 7,$$
$$||G||_2^2 = ||G'||_2^2 = P_{2k-2}(U) = (1+o(1))\frac{d^{2k-2}}{n^{2k-2}}|U|^{2k-1}.$$

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

Applying Proposition 3.2,

$$\left|C_{2k+1}(U) - \frac{d^2}{n^2}P_k(U)P_{k-1}(U)\right| \le \lambda^2\sqrt{C_{2k}(U)C_{2k-2}(U)} + 2\frac{d\lambda}{n}\left(\sqrt{P_{2k}(U)P_{2k-2}(U)}\right).$$

When $m = 5$, we have

$$\left|C_5(U) - (1+o(1))\,\frac{|U|^5d^5}{n^5}\right| = O\left(\lambda^2\sqrt{\frac{|U|^6d^5}{n^5} + \frac{\lambda^2|U|^4d^2}{n^2}} + \frac{\lambda d^4|U|^4}{n^4}\right).$$

Note that $\frac{\lambda d^4|U|^4}{n^4} = o\left(\frac{|U|^5d^5}{n^5}\right)$ due to our assumption that $\frac{\lambda n}{d} = o(|U|)$. If $|U| \le \frac{\lambda n^{3/2}}{d^{3/2}}$, then the second term in the square root is bigger than the first, and hence

$$\lambda^2\sqrt{\frac{|U|^6d^5}{n^5} + \frac{\lambda^2|U|^4d^2}{n^2}} = O\left(\frac{\lambda^3|U|^2d}{n}\right).$$

If $|U| \ge \frac{\lambda n^{3/2}}{d^{3/2}}$, then the first term is bigger than the second and we have

$$\lambda^2\sqrt{\frac{|U|^6d^5}{n^5} + \frac{\lambda^2n^4d^2}{n^2}} = O\left(\frac{\lambda^2|U|^3d^{5/2}}{n^{5/2}}\right).$$

Using the assumption that $|U| \ge \frac{\lambda n^{3/2}}{d^{3/2}}$ gives that $\frac{\lambda^2|U|^3d^{5/2}}{n^{5/2}} \le \frac{\lambda|U|^4d^4}{n^4} = o\left(\frac{|U|^5d^5}{n^5}\right)$ by our assumption that $\frac{\lambda n}{d} = o(|U|)$. In either case, the inequality is satisfied. For $m \ge 7$, we have

$$\left|C_{2k+1}(U) - (1+o(1))\,\frac{|U|^{2k+1}d^{2k+1}}{n^{2k+1}}\right| =$$

$$O\left(\lambda^2\sqrt{\left[\frac{|U|^{2k}d^{2k}}{n^{2k}} + \frac{\lambda^{2k-2}|U|^2d}{n}\right]\left[\frac{|U|^{2k-2}d^{2k-2}}{n^{2k-2}} + \frac{\lambda^{2k-4}|U|^2d}{n}\right]} + \frac{\lambda|U|^{2k}d^{2k}}{n^{2k}}\right).$$

First, note that $\frac{\lambda|U|^{2k}d^{2k}}{n^{2k}} = o\left(\frac{|U|^{2k+1}d^{2k+1}}{n^{2k+1}}\right)$ by the assumption that $\frac{\lambda n}{d} = o(|U|)$, so we may ignore this term. Hence, if each of the four terms

$$\frac{|U|^{4k-2}d^{4k-2}}{n^{4k-2}}, \quad \frac{\lambda^{2k-2}|U|^{2k}d^{2k-1}}{n^{2k-1}}, \quad \frac{\lambda^{2k-4}|U|^{2k+2}d^{2k+1}}{n^{2k+1}}, \quad \frac{\lambda^{4k-6}|U|^4d^2}{n^2}$$

is either

$$O\left(\frac{|U|^{4k}d^{4k}}{\lambda^2n^{4k}}\right) \quad \text{or} \quad O\left(\frac{\lambda^{4k-6}|U|^4d^2}{n^2}\right),$$

then we are done. The fourth term trivially satisfies the inequality. The first term is $o\left(\frac{|U|^{4k}d^{4k}}{\lambda^2n^{4k}}\right)$ and

$$\frac{\lambda^{2k-2}|U|^{2k}d^{2k-1}}{n^{2k-1}} = o\left(\frac{\lambda^{2k-4}|U|^{2k+2}d^{2k+1}}{n^{2k+1}}\right)$$

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

by the assumption that $\frac{\lambda n}{d} = o(|U|)$. Finally, if $|U| \geq \lambda \left(\frac{n}{d}\right)^{(2k-1)/(2k-2)}$, then

$$\frac{\lambda^{2k-4}|U|^{2k+2}d^{2k+1}}{n^{2k+1}} \leq \frac{|U|^{4k}d^{4k}}{\lambda^2 n^{4k}}.$$

Otherwise,

$$\frac{\lambda^{2k-4}|U|^{2k+2}d^{2k+1}}{n^{2k+1}} \leq \frac{\lambda^{4k-6}|U|^4 d^2}{n^2}.$$

Case 2: $m = 2k$.

For this case, we want to apply Proposition 3.2 again, so we need to define suitable functions f and g, namely, for $x, y \in U$,

$$f(x,y) = g(x,y) = \text{the number of paths of length } k-1 \text{ between } x \text{ and } y.$$

Then, by inductive hypothesis and Proposition 3.5, one has

$$\|f\|_1 = \|g\|_1 = P_{k-1}(U) = (1+o(1))\frac{d^{k-1}}{n^{k-1}}|U|^k,$$
$$\|f\|_2^2 = \|g\|_2^2 = C_{2k-2}(U) = (1+o(1))\frac{|U|^{2k-2}d^{2k-2}}{n^{2k-2}} + \Theta\left(\lambda^{2k-4}\frac{d}{n}|U|^2\right),$$
$$\|F\|_2^2 = \|G\|_2^2 = \|F'\|_2^2 = \|G'\|_2^2 = P_{2k-2}(U) = (1+o(1))\frac{d^{2k-2}}{n^{2k-2}}|U|^{2k-1}.$$

By applying Proposition 3.2 and the estimates above, we get that

$$\left|C_{2k}(U) - \frac{d^2}{n^2}(P_{k-1}(U))^2\right| \leq \lambda^2 C_{2k-2}(U) + 2\frac{\lambda d}{n}P_{2k-2}(U),$$

and hence

$$\left|C_{2k}(U) - (1+o(1))\frac{|U|^{2k}d^{2k}}{n^{2k}}\right| = O\left(\frac{|U|^{2k-2}d^{2k-2}\lambda^2}{n^{2k-2}} + \frac{\lambda^{2k-2}|U|^2 d}{n} + \frac{\lambda|U|^{2k-1}d^{2k-1}}{n^{2k-1}}\right).$$

Since $\frac{\lambda|U|^{2k-1}d^{2k-1}}{n^{2k-1}}$ and $\frac{|U|^{2k-2}d^{2k-2}\lambda^2}{n^{2k-2}}$ are both $o\left(\frac{|U|^{2k}d^{2k}}{n^{2k}}\right)$ because of the assumption that $\frac{\lambda n}{d} = o(|U|)$, we are done. ∎

3.5 Proof of Theorem 1.9 using the second counting lemma

Proof of Theorem 1.9 Using the second counting lemma, we are able to prove Theorem 1.9 for all $m \geq 5$, i.e.,

$$C_m(U) = \frac{|U|^m d^m}{n^m} + \Theta\left(\frac{\lambda|U|^{m-1}d^{m-1}}{n^{m-1}} + \lambda^{m-2}\frac{d}{n}|U|^2\right),$$

but for $m = 4$, the result becomes slightly weaker, namely,

$$C_4(U) = O\left(\frac{|U|^4 d^4}{n^4} + \frac{\lambda^2|U|^2 d}{n}\right).$$

We proceed by induction.

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

Case 1: $m = 2k$.

For $m = 4$, by Corollary 3.4 and Proposition 3.5, we have that

$$\left|C_4(U) - \frac{|U|^4 d^4}{n^4}(1+o(1))\right| \le \sqrt{(1+o(1))\frac{|U|^2\lambda^2 d}{n}C_4(U)}, \tag{14}$$

using that $C_2(U) = P_2(U)$ and $P_2(U) = \frac{d|U|^2}{n}(1+o(1))$ by the assumption that $\frac{\lambda n}{d} = o(|U|)$. Hence, we may set up a quadratic in $\sqrt{C_4(U)}$ to obtain

$$\begin{aligned} C_4(U) &\le \left(\frac{\sqrt{(1+o(1))\frac{|U|^2\lambda^2 d}{n}} + \sqrt{(1+o(1))\frac{|U|^2\lambda^2 d}{n} + 4\frac{|U|^4 d^4}{n^4}(1+o(1))}}{2}\right)^2 \\ &= O\left(\frac{|U|^4 d^4}{n^4} + \frac{\lambda^2|U|^2 d}{n}\right). \end{aligned}$$

This gives the desired estimate for C_4. If one wishes to have the main term $\frac{|U|^4 d^4}{n^4}$ instead of $c\frac{|U|^4 d^4}{n^4}$, for some positive constant c, with this approach, then it can be pushed further as follows.

Using the above upper bound for C_4 and the estimate (14) gives us

$$\left|C_4(U) - \frac{|U|^4 d^4}{n^4}(1+o(1))\right| = O\left(\sqrt{\frac{|U|^4\lambda^4 d^2}{n^2} + \frac{|U|^6\lambda^2 d^5}{n^5}}\right).$$

This gives

$$C_4(U) = \frac{|U|^4 d^4}{n^4} + \Theta\left(\frac{\lambda^2 d|U|^2}{n} + \frac{\lambda|U|^3 d^3}{n^3} + \frac{|U|^3\lambda d^{5/2}}{n^{5/2}}\right).$$

Note that this gives the estimate (3) under the more restrictive condition that $\lambda\frac{n^{3/2}}{d^{3/2}} = o(|U|)$.

Assume that the upper bound holds for all cycles of length at most $m-1$. We now show that it also holds for cycles of length m. Indeed, if $m = 2k$, then we can apply Corollary 3.4 to have

$$C_{2k}(U) \le \frac{d}{n}P_{2k-1}(U) + \lambda\left(C_{2k}(U)C_{2k-2}(U)\right)^{1/2}.$$

Solving a quadratic in $\sqrt{C_{2k}(U)}$ gives

$$C_{2k}(U) \le \left(\frac{\lambda\sqrt{C_{2k-2}(U)} + \sqrt{\lambda^2 C_{2k-2}(U) + 4\frac{d}{n}P_{2k-1}(U)}}{2}\right)^2.$$

Using Proposition 3.5 gives that

$$C_{2k}(U) - \frac{|U|^{2k}d^{2k}}{n^{2k}}(1+o(1)) = O\left(\lambda^2 C_{2k-2}(U) + \lambda\sqrt{\lambda^2(C_{2k-2}(U))^2 + \frac{d}{n}C_{2k-2}(U)P_{2k-1}(U)}\right).$$

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

By the inductive hypothesis and the assumption that $\frac{\lambda n}{d} = o(|U|)$, we have that

$$\lambda^2 C_{2k-2}(U) = O\left(\frac{\lambda|U|^{2k-1}d^{2k-1}}{n^{2k-1}} + \frac{\lambda^{2k-2}|U|^2 d}{n}\right),$$

and hence we are done as long as

$$\lambda\sqrt{\frac{d}{n}C_{2k-2}(U)P_{2k-1}(U)} = O\left(\frac{\lambda|U|^{2k-1}d^{2k-1}}{n^{2k-1}} + \frac{\lambda^{2k-2}|U|^2 d}{n}\right).$$

By the inductive hypothesis and Proposition 3.5, we have

$$\lambda\sqrt{\frac{d}{n}C_{2k-2}(U)P_{2k-1}(U)} = O\left(\sqrt{\frac{\lambda^2|U|^{4k-2}d^{4k-2}}{n^{4k-2}} + \frac{\lambda^3|U|^{4k-3}d^{4k-3}}{n^{4k-3}} + \frac{\lambda^{2k-2}|U|^{2k+2}d^{2k+1}}{n^{2k+1}}}\right).$$

Since $\frac{\lambda n}{d} = o(|U|)$, we have that $\frac{\lambda^3|U|^{4k-3}d^{4k-3}}{n^{4k-3}} = o\left(\frac{\lambda^2|U|^{4k-2}d^{4k-2}}{n^{4k-2}}\right)$. Therefore, because

$$\sqrt{\frac{\lambda^2|U|^{4k-2}d^{4k-2}}{n^{4k-2}}} = \frac{\lambda|U|^{2k-1}d^{2k-1}}{n^{2k-1}},$$

we are done as long as $\frac{\lambda^{k-1}|U|^{k+1}d^{(2k+1)/2}}{n^{(2k+1)/2}}$ is small enough. If $|U| \geq \frac{\lambda n^{(2k-3)/(2k-4)}}{d^{(2k-3)/(2k-4)}}$, then

$$\frac{\lambda^{k-1}|U|^{k+1}d^{(2k+1)/2}}{n^{(2k+1)/2}} \leq \frac{\lambda|U|^{2k-1}d^{2k-1}}{n^{2k-1}}.$$

Otherwise, we have

$$\frac{\lambda^{k-1}|U|^{k+1}d^{(2k+1)/2}}{n^{(2k+1)/2}} \leq \frac{\lambda^{2k-2}|U|^2 d}{n},$$

and the upper bound is complete. An analogous calculation gives the corresponding lower bound, and we omit the details.

Case 2: $m = 2k+1$.
This case follows directly from Corollary 3.4 and the case $m = 2k$ above. ∎

4 Proofs of Theorems 1.15 and 1.16

4.1 Technical lemmas

To prove Theorems 1.15 and 1.16, we use the following results, which are direct consequences of the expander mixing lemma. The first result guarantees that vertex sets bigger than $\lambda n/d$ will have an edge of each color.

Lemma 4.1 *Let G be an (n, d, λ)-colored graph with color set D, and $A, B \subset V(G)$ with $|A| = |B| > \frac{\lambda n}{d}$. Then, for each color $c \in D$, there exists an edge uv of color c with $u \in A$ and $v \in B$. In other words, every vertex set of size greater than $\frac{\lambda n}{d}$ determines every color.*

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

Proof For each color c in D, let G_c be the induced graph on c, then G_c is an (n, d, λ)-graph. Applying Lemma 2.5 with

$$f(u) = \begin{cases} 1, & \text{if } u \in A, \\ 0, & \text{otherwise} \end{cases}$$

and

$$g(v) = \begin{cases} 1, & \text{if } v \in B, \\ 0, & \text{otherwise,} \end{cases}$$

we have

$$\langle f, Ag \rangle = e(A, B) \coloneqq \left| \{(a, b) \in A \times B : ab \in E(G_c)\} \right|.$$

It is clear that

$$\mathbb{E}(f) = \frac{|A|}{n}, \ \mathbb{E}(g) = \frac{|B|}{n}$$

and

$$\|f\|_2 = \sqrt{|A|}, \ \|g\|_2 = \sqrt{|B|}.$$

Then we have

$$\left| e(A, B) - \frac{d}{n}|A||B| \right| \leq \lambda \sqrt{|A||B|}.$$

So

$$e(A, B) \geq \frac{d}{n}|A||B| - \lambda \sqrt{|A||B|}.$$

Since $|A| = |B| > \frac{\lambda n}{d}$,

$$e(A, B) \geq \frac{d}{n}|A|^2 - \lambda |A| \geq |A| \left(\frac{d}{n} \cdot |A| - \lambda \right) > |A| \left(\frac{d}{n} \cdot \frac{\lambda n}{d} - \lambda \right) > 0.$$

Which means that there exists at least one edge of color c between A and B. ∎

The next technical lemma uses the previous result to give an upper bound on the number of vertices with small degree of a given edge color.

Lemma 4.2 *Let G be an (n, d, λ)-colored graph with color set D, and let $U \subset V(G)$, with $|U| = r \cdot \frac{\lambda n}{d}$, for some $r > 1$. Then, for any fixed color $c \in D$, $s \in \mathbb{N}$, there are at most $s \cdot \frac{\lambda n}{d}$ vertices of U for which each of them is incident with fewer than s edges colored by c.*

Proof Let H be the induced graph on color c. Consider the subgraph H^* of H generated by only those vertices of degree less than s, so H^* can be s-colorable. That is, we have a vertex partition into s independent sets. Using Lemma 4.1, an independent set in H (and thus in H^*) has size at most $\frac{\lambda n}{d}$. Otherwise, by Lemma 4.1, every vertex set of size greater than $\frac{\lambda n}{d}$ determines every color, which means there exists two vertices connected by a c-color edge, contradicting the independence. As a result, $|V(H^*)| \leq s \cdot \frac{\lambda n}{d}$, proving the lemma. ∎

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

The next lemma develops this further by giving lower bounds on the number of disjoint copies of star graphs.

Lemma 4.3 *Let G be an (n, d, λ)-colored graph with color set D, and let $U \subset V(G)$ with $|U| = r \cdot \frac{\lambda n}{d}$, for some $r > 1$. Then the number of vertex disjoint copies of the nonempty star graph $K_{1,m}$ with any fixed edge-coloring from D is at least $\frac{r-m}{m+1} \cdot \frac{\lambda n}{d}$.*

Proof Let T be the maximal set of copies of $K_{1,m}$ in U, and let H be the union of all copies in T. Then $U - H$ will have no copies of $K_{1,m}$.

Suppose the set of color of $K_{1,m}$ is $\{c_1, c_2, \ldots, c_t\}$ with multiplicities $\{m_1, m_2, \ldots, m_t\}$. Using Lemma 4.2, for each i, there are at most $m_i \cdot \frac{\lambda n}{d}$ vertices that are incident with fewer than m_i edges colored by c_i. Summing over i, we get that there are at most

$$\sum_{i=1}^{t} m_i \cdot \frac{\lambda n}{d} = m \cdot \frac{\lambda n}{d}$$

vertices of $U - H$ which are not colored c_i from at least m_i other vertices of $U - H$ for every i. If vertex $v \in U - H$ is incident with at least m_i edges color c_i for every i, then v is the singleton bipartition set of an instance of $K_{1,m}$. Thus, $|U - H| \leq m \cdot \frac{\lambda n}{d}$. By disjointness,

$$|T| = \frac{|H|}{m+1} \geq \frac{r \cdot \frac{\lambda n}{d} - m \cdot \frac{\lambda n}{d}}{m+1} = \frac{r-m}{m+1} \cdot \frac{\lambda n}{d},$$

as required. ∎

Our final technical lemma is a simple application of Lemma 4.2 that gives a lower bound on the number of disjoint edges of a given color in a vertex set.

Lemma 4.4 *Let G be an (n, d, λ)-colored graph with color set D, and let $U \subset V(G)$ with $|U| \geq \frac{2\lambda n}{d}$. Then, for each color $c \in D$, the number of disjoint c colored edges in U is at least $\frac{|U|}{2} - \frac{\lambda n}{d}$.*

Proof We partition the vertex set of U into two sets, A and B, such that $|A| = |B| = \frac{|U|}{2}$. Choose as large a matching of color c as possible between, say, $A' \subseteq A$ and $B' \subseteq B$. We have that the two sets $A \backslash A'$ and $B \backslash B'$ both have size at most $\frac{\lambda n}{d}$. Otherwise, by Lemma 4.1, we could increase the size of our matching. As a result, the number of disjoint c colored edges in U is at least

$$|A^{'}| = |B^{'}| \geq \frac{|U|}{2} - \frac{\lambda n}{d},$$

as required. ∎

4.2 Proof of Theorem 1.15

The proof proceeds by strong induction on the number of edges in T. If T contains no edges the theorem is clearly true; if T is a star graph $K_{1,m}$, then $\sigma(G) = m + 1$ and the theorem is Lemma 4.3.

Now assume T is not a star graph. Let $T^{'}$ be the graph produced by deleting all leaves of T. Since T is not a star graph, T' is a tree which has at least two leaves,

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

we can choose v be a leaf of T' such that there exists another leaf of T', say w, such that $\deg_T v \le \deg_T w$. Suppose the set of leaves of T connected to v is $\{v_1, v_2, \dots, v_y\}$. Define the graph T^* to be $T \backslash \{v_1, v_2, \dots, v_y\}$. By construction, T^* is a tree with fewer edges than T and $\sigma(T) = \sigma(T^*) \cdot (y+1)$. Define r by the relation $|U| = r \cdot \frac{\lambda n}{d}$. By the inductive hypothesis we have the number of disjoint copies of T^* in U denoted by C_{T^*} is at least $\left(\frac{r}{\sigma(T^*)} - 1\right) \cdot \frac{\lambda n}{d}$.

We are building our tree T out of stars instead of edges. Let W be the set of copies of v in U. By disjointness $|W| = |C_{T^*}|$. Let $K_{1,y}$ be the star graph generated by $\{v, v_1, v_2, \dots, v_y\}$ where the root is v. Using Lemma 4.3, there exist at least $\frac{|W|/\frac{\lambda n}{d} - y}{y+1} \cdot \frac{\lambda n}{d}$ disjoint copies of $K_{1,y}$ in W. For each copy of $K_{1,y}$, we can build our tree T by adding the copies of T^* that correspond to v. These are disjoint copies of T because of the disjointness of T^* and the disjointness of $K_{1,y}$. So there are at least

$$\begin{aligned}
\frac{|W|/\frac{\lambda n}{d} - y}{y+1} \cdot \frac{\lambda n}{d} &= \frac{|C_{T^*}|/\frac{\lambda n}{d} - y}{y+1} \cdot \frac{\lambda n}{d} \\
&\ge \frac{\left(\frac{r}{\sigma(T^*)} - 1\right) - y}{y+1} \cdot \frac{\lambda n}{d} \\
&= \left(\frac{r}{(y+1)\sigma(T^*)} - 1\right) \cdot \frac{\lambda n}{d} \\
&= \left(\frac{r}{\sigma(T)} - 1\right) \cdot \frac{\lambda n}{d}
\end{aligned}$$

disjoint copies of T as required.

4.3 Proof of Theorem 1.16

The proof proceeds by induction on the number of edges on T. If T contains no edges, the theorem is clearly true. If T is an edge, then $|V(T)| = 2$, the theorem is Lemma 4.4.

So assume T is a tree with m vertices. Consider the subgraph T^* of T produced by deleting one leaf on vertex x. Let us say the edge we are just removing has color c. By construction, T^* is a tree with fewer edges than T. By inductive hypothesis we have the collection of disjoint copies of T^* in U is at least $\frac{|U|}{m-1} - \frac{\lambda n}{d}$.

Choose $\frac{|U|}{m}$ copies of them arbitrarily, and let this set of vertices be called S. This is possible since $|U| \ge m(m-1)\frac{\lambda n}{d}$.

So S has size $(m-1) \cdot \frac{|U|}{m}$. Now, in these copies of T^*, denote by A the set copies of x to which we will be trying to add an edge of color c, so $|A| = \frac{|U|}{m}$. Let $B = U \backslash S$, so $|B| = |U| - (m-1) \cdot \frac{|U|}{m} = \frac{|U|}{m}$.

Choose as large of a matching color c as possible between, say, $A' \subseteq A$ and $B' \subseteq B$, each matching creates a copy of T. Let C and D be the sets of vertices in $A \backslash A'$ and $B \backslash B'$, respectively. Then we have that $|C| = |D| \le \frac{\lambda n}{d}$. Otherwise, using Lemma 4.1, we can find at least one c colored edge between C and D, which would increase the size

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

of our matching. So the number of disjoint copies of T is

$$|A'| = |A| - |C| \geq \frac{|U|}{m} - \frac{\lambda n}{d},$$

as required.

Acknowledgements T. Pham and S. Senger would like to thank the VIASM for the hospitality and for the excellent working conditions.

References

[1] E. Bannai, O. Shimabukuro, and H. Tanaka, *Finite Euclidean graphs and Ramanujan graphs*. Discret. Math. **309**(2009), no. 20, 6126–6134.
[2] 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. **53**(2016), 115–126.
[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.
[4] J. Bourgain, N. Katz, and T. Tao, *A sum-product estimate in finite fields, and applications*. Geom. Funct. Anal. **14**(2004), no. 1, 27–57.
[5] D. Covert and S. Senger, *Pairs of dot products in finite fields and rings*. In: Combinatorial and additive number theory II, Springer, Cham, 2015, pp. 129–138.
[6] X. Du, A. Iosevich, O. Yumeng, H. Wang, and R. Zhang, *An improved result for Falconer's distance set problem in even dimensions*. Math. Ann. **380**(2021), no. 3, 1215–1231.
[7] X. Du and R. Zhang, *Sharp* L^2 *estimates of the Schrödinger maximal function in higher dimensions*. Ann. Math. **189**(2019), no. 3, 837–861.
[8] Z. Dvir, *On the size of Kakeya sets in finite fields*. J. Amer. Math. Soc. **22**(2009), no. 4, 1093–1097.
[9] S. Eswarathasan, A. Iosevich, and K. Taylor, *Fourier integral operators, fractal sets, and the regular value theorem*. Adv. Math. **228**(2011), no. 4, 2385–2402.
[10] M. Z. Garaev, *An explicit sum-product estimate in* F_p. Int. Math. Res. Not. IMRN **2007**(2007), Article no. rnm035, 11 pp.
[11] W. Haemers, *Eigenvalue techniques in design and graph theory*. Ph.D. thesis, Eindhoven University of Technology, 1980.
[12] D. Hart, A. Iosevich, D. Koh, and M. Rudnev, *Averages over hyperplanes, sum-product theory in vector spaces over finite fields and the Erdős–Falconer distance conjecture*. Trans. Amer. Math. Soc. **363**(2011), no. 6, 3255–3275.
[13] D. Hart, A. Iosevich, and J. Solymosi, *Sum-product estimates in finite fields via Kloosterman sums*. Int. Math. Res. Not. IMRN **2007**(2007), Article no. rnm007. https://doi.org/10.1093/imrn/rnm007
[14] A. Iosevich, G. Jardine, and B. McDonald, *Cycles of arbitrary length in distance graphs on* F_q^d. Proc. Steklov Inst. Math. **314**(2021), no. 1, 27–43.
[15] A. Iosevich, D. Koh, and M. Lewko, *Finite field restriction estimates for the paraboloid in high even dimensions*. J. Funct. Anal. **278**(2020), no. 11, 108450.
[16] A. Iosevich, M. Kolountzakis, Y. Lyubarskii, A. Mayeli, and J. Pakianathan, *On Gabor orthonormal bases over finite prime fields*. Bull. Lond. Math. Soc. **53**(2021), no. 2, 380–391.
[17] A. Iosevich, A. Mayeli, and J. Pakianathan, *The Fuglede conjecture holds in* $\mathbb{Z}_p \times \mathbb{Z}_p$. Anal. PDE **10**(2017), no. 4, 757–764.
[18] A. Iosevich and M. Rudnev, *Erdős distance problem in vector spaces over finite fields*. Trans. Amer. Math. Soc. **359**(2007), no. 12, 6127–6142.
[19] H. Iwaniec and E. Kowalski, *Analytic number theory*, Colloquium Publications, 53, American Mathematical Society, Providence, RI, 2004.
[20] D. Koh, S. Lee, T. Pham, and C.-Y. Shen, *Configurations of rectangles in a set in* F_q^2. Finite Fields and Their Applications, **86**(2023), 102147.
[21] M. Krivelevich and B. Sudakov, *Pseudo-random graphs*. In: More sets, graphs and numbers, Springer, Berlin–Heidelberg, 2006, pp. 199–262.
[22] W. M. Kwok, *Character tables of association schemes of affine type*. Eur. J. Comb. **13**(1992), no. 3, 167–185.

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press

[23] M. Lewko, *New restriction estimates for the 3-d paraboloid over finite fields*. Adv. Math. **270**(2015), no. 1, 457–479.
[24] M. Lewko, *Finite field restriction estimates based on Kakeya maximal operator estimates*. J. Eur. Math. Soc. **21**(2019), no. 12, 3649–3707.
[25] B. Liu, *On Hausdorff dimension of radial projections*. Rev. Mat. Iberoam. **37**(2020), no. 4, 1307–1319.
[26] N. Lyall and Á. Magyar, *Weak hypergraph regularity and applications to geometric Ramsey theory*. Trans. Amer. Math. Soc. Ser. B **9**(2022), no. 5, 160–207.
[27] R. Merris, *Multilinear algebra*, CRC Press, Boca Raton, FL, 1997.
[28] G. Mockenhaupt and T. Tao, *Restriction and Kakeya phenomena for finite fields*. Duke Math J. **121**(2004), no. 1, 35–74.
[29] M. Rudnev and I. D. Shkredov, *On the restriction problem for discrete paraboloid in lower dimension*. Adv. Math. **339**(2018), 657–671.
[30] D. M. Soukup, *Embeddings of weighted graphs in Erdős-type settings*. Mosc. J. Combin. Number Theory **8**(2019), no. 2, 117–123.
[31] L. A. Vinh, *The solvability of norm, bilinear and quadratic equations over finite fields via spectra of graphs*. Forum Math. **26**(2014), no. 1, 141–175.
[32] A. Weil, *On some exponential sums*. Proc. Natl. Acad. Sci. USA **34**(1948), 204–207.
[33] T. Wolff, *Recent work connected with the Kakeya problem*. In: Prospects in mathematics (Princeton, NJ, 1996). Vol. 2, American Mathematical Society, Providence, RI, 1999, pp. 129–162.

University of Science, Vietnam National University, Hanoi, Vietnam
e-mail: thangpham.math@vnu.edu.vn

Department of Mathematics, Missouri State University, Springfield, MO, United States
e-mail: StevenSenger@MissouriState.edu

Department of Mathematics and Statistics, Villanova University, Villanova, PA, United States
e-mail: michael.tait@villanova.edu

University of Science, Vietnam National University, Hanoi, Vietnam
e-mail: VuThiHuongThu_T64@hus.edu.vn

https://doi.org/10.4153/S0008414X24000245 Published online by Cambridge University Press