

The Edge-Statistics Conjecture for Hypergraphs

Vishesh Jain¹, Matthew Kwan^{2,*}, Dhruv Mubayi¹, and Tuan Tran³

¹Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago, Chicago IL, 60607 USA

²Institute of Science and Technology Austria (ISTA), Am Campus 1, 3400 Klosterneuburg, Austria

³School of Mathematical Sciences, University of Science and Technology of China, Hefei, 230026 Anhui, China

*Correspondence to be sent to: e-mail: matthew.a.kwan@gmail.com

Communicated by Benny Sudakov

Let r, k, ℓ be integers such that $0 \leq \ell \leq \binom{k}{r}$. Given a large r -uniform hypergraph G , we consider the fraction of k -vertex subsets that span exactly ℓ edges. If ℓ is 0 or $\binom{k}{r}$, this fraction can be exactly 1 (by taking G to be empty or complete), but for all other values of ℓ , one might suspect that this fraction is always significantly smaller than 1.

In this paper we prove an essentially optimal result along these lines: if ℓ is not 0 or $\binom{k}{r}$, then this fraction is at most $(1/e) + \varepsilon$, assuming k is sufficiently large in terms of r and $\varepsilon > 0$, and G is sufficiently large in terms of k . Previously, this was only known for a very limited range of values of r, k, ℓ (due to Kwan–Sudakov–Tran, Fox–Saueremann, and Martinsson–Mousset–Noever–Trujić). Our result answers a question of Alon–Hefetz–Krivelevich–Tyomkyn, who suggested this as a hypergraph generalization of their *edge-statistics conjecture*. We also prove a much stronger bound when ℓ is far from 0 and $\binom{k}{r}$.

1 Introduction

Given a k -vertex graph H , what is the maximum possible number of k -vertex subsets of an n -vertex graph that induce a copy of H ? Denote this number by $N(n, H)$, so we have $0 \leq N(n, H) \leq \binom{n}{k}$. A simple averaging argument shows that $N(n, H)/\binom{n}{k}$ is nonincreasing in n (for $n \geq k$), so we can define

$$\text{ind}(H) = \lim_{n \rightarrow \infty} \frac{N(n, H)}{\binom{n}{k}}. \quad (1.1)$$

This quantity is called the *inducibility* or the *maximum induced density* of H . It was first considered in 1975 by Pippenger and Golumbic [19] and has been studied intensively in the intervening decades.

In general, it is very difficult to determine $\text{ind}(H)$, even for small graphs H (e.g., the inducibility of the 4-vertex path is still unknown). However, we do know the *minimum* of $\text{ind}(H)$, among all k -vertex graphs H (provided k is sufficiently large): indeed, Pippenger and Golumbic [19] showed that if H has k vertices, then

$$\text{ind}(H) \geq \frac{k!}{k^k - k}, \quad (1.2)$$

Received: May 14, 2025. Revised: August 7, 2025. Accepted: August 20, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

and it is known (see [2, 10]) that if $k = 5$ or if k is sufficiently large, there is a choice of H that attains this bound.

What about the *maximum* value of $\text{ind}(H)$? It is easy to see that $\text{ind}(H) = 1$ when H is a complete or empty graph; if we exclude these “trivial” examples then we arrive at the so-called *large inducibility conjecture* of Alon, Hefetz, Krivelevich, and Tyomkyn [1, Conjecture 1.2]. Namely, they identified several (nontrivial) infinite classes of graphs H with $\text{ind}(H) > 1/e$, and conjectured that the maximum of $\text{ind}(H)$ over all nontrivial k -vertex graphs H tends to $1/e$ as $k \rightarrow \infty$.

Alon, Hefetz, Krivelevich, and Tyomkyn also made a second much stronger conjecture called the *edge-statistics conjecture* [1, Conjecture 1.1], concerning a much looser variant of graph inducibility. Specifically, for $0 \leq \ell \leq \binom{k}{2}$, let $N_2(n, k, \ell)$ be the maximum possible number of k -vertex subsets of an n -vertex graph that induce exactly ℓ edges, and let

$$\text{ind}_2(k, \ell) = \lim_{n \rightarrow \infty} \frac{N_2(n, k, \ell)}{\binom{n}{k}}. \quad (1.3)$$

For each value of k , say that 0 and $\binom{k}{2}$ are the “trivial” values of ℓ ; the edge-statistics conjecture says that the maximum of $\text{ind}_2(k, \ell)$ over all nontrivial ℓ tends to $1/e$ as $k \rightarrow \infty$. This conjecture has significance beyond its consequences for graph inducibility: it can be interpreted as giving a limit on “how uniform” a graph can be, with respect to statistics of edges in small subsets.

It is equally natural to consider the large inducibility and edge-statistics conjectures for *hypergraphs*; these generalizations were actually explicitly suggested in the same paper of Alon, Hefetz, Krivelevich, and Tyomkyn, though they wrote “needless to say that we expect these questions to be difficult”. To be precise, for an r -uniform hypergraph H and for $0 \leq \ell \leq \binom{k}{r}$, we define $N(n, H)$ and $N_r(k, \ell)$ to be the maximum possible numbers of k -vertex subsets in an n -vertex r -uniform hypergraph, which induce a copy of H and which induce exactly ℓ edges, respectively. Then we can define $\text{ind}(H)$ and $\text{ind}_r(k, \ell)$ as in (1.1) and (1.3), and use these notions to generalize the large inducibility and edge-statistics conjectures in the obvious ways: namely, the maximum value of $\text{ind}(H)$, over all k -vertex r -uniform hypergraphs that are neither empty nor complete, and the maximum value of $\text{ind}_r(k, \ell)$ over all $0 < \ell < \binom{k}{r}$, both tend to $1/e$ as $k \rightarrow \infty$ (holding r fixed).

In a combination of papers by Kwan, Sudakov, and Tran [14], Fox and Sauermann [9], and Martinsson, Mousset, Noever, and Trujić [17], the edge-statistics conjecture for graphs (i.e., $r = 2$), and therefore the large inducibility conjecture for graphs have been resolved. These papers also provide evidence for the hypergraph edge-statistics conjecture, establishing it in the special cases where $\ell = o(k)$ and where $r = 3$ and $\ell = \Omega(k^3)$. Our first main theorem completely resolves the hypergraph edge statistics conjecture, and therefore the hypergraph large inducibility conjecture.

Theorem 1.1. Fix any $r \in \mathbb{N}$ and $\varepsilon > 0$. Suppose k is sufficiently large in terms of r, ε . If $\ell \notin \{0, \binom{k}{r}\}$, then

$$\text{ind}_r(k, \ell) \leq \frac{1}{e} + \varepsilon.$$

Consequently, for any k -vertex r -uniform hypergraph H that is neither empty nor complete,

$$\text{ind}(H) \leq \frac{1}{e} + \varepsilon.$$

Alon, Hefetz, Krivelevich, and Tyomkyn were also interested in the value of $\text{ind}_2(k, \ell)$ when ℓ is far from 0 and $\binom{k}{2}$. They made several conjectures in this direction [1, Conjecture 6.1 and 6.2], which have since been resolved by Kwan, Sudakov, and Tran [14] and Kwan and Sauermann [13]. We also prove a theorem along these lines for hypergraphs.

Theorem 1.2. Fix any $r \in \mathbb{N}$ and $\varepsilon > 0$. If $\alpha \binom{k}{r} \leq \ell \leq (1 - \alpha) \binom{k}{r}$ for some $\alpha \in (0, 1/2]$, and if αk is sufficiently large in terms of r, ε , then

$$\text{ind}_r(k, \ell) \leq \frac{1}{(\alpha k)^{1/2 - \varepsilon}}.$$

We note that Kwan, Sudakov, and Tran [14, Theorem 1.3] previously proved the $r = 3$ case of Theorem 1.2 in the “extremely dense” regime where α is a constant that may not vary with k (using the induced hypergraph removal lemma).

Theorems 1.1 and 1.2 are both essentially optimal, as follows. Write $N(G, k, \ell)$ for the number of k -vertex subsets with ℓ edges, in an r -uniform hypergraph G .

- It was observed by Fox and Sauermaann [9] that for every r the constant “ $1/e$ ” in Theorem 1.1 cannot be improved. To briefly explain why: for any $1 \leq s \leq r$, let F be a random s -uniform hypergraph on n vertices, in which every possible edge is present with probability $1/\binom{k}{s}$ independently. Then, let G be the r -uniform hypergraph on the same vertex set whose edges are the r -sets which are supersets of some edge of F . For $\ell = \binom{k-s}{r-s}$, it is easy to establish the convergence in probability

$$\frac{N(G, k, \ell)}{\binom{n}{k}} \xrightarrow{\mathbb{P}} \left(1 - \frac{1}{\binom{k}{s}}\right)^{\binom{k}{s}-1} > \frac{1}{e} \quad (1.4)$$

as $n \rightarrow \infty$, which implies that $\text{ind}_r(k, \ell) > 1/e$.

- It is easy to see that the exponent “ $1/2$ ” in Theorem 1.2 cannot be improved. Indeed, consider any $\alpha \in (0, 1/2)$ and $k \in \mathbb{N}$ such that αk is an integer, and consider any n divisible by k . Let G be an n -vertex r -uniform hypergraph with a distinguished set S of αn vertices, whose edges are the r -sets that intersect S in exactly one vertex. Then, with $\ell = \alpha k \binom{k-\alpha k}{r-1}$ (so ℓ has order of magnitude $\alpha \binom{k}{r}$), one can compute

$$\lim_{n \rightarrow \infty} \frac{N(G, k, \ell)}{\binom{n}{k}} = \binom{k}{\alpha k} \alpha^{\alpha k} (1-\alpha)^{(1-\alpha)k} \geq \frac{c}{(\alpha k)^{1/2}} \quad (1.5)$$

for some absolute constant $c > 0$, which implies that $\text{ind}_r(k, \ell) > c(\alpha k)^{-1/2}$.

The estimates in (1.4) and (1.5) can be confirmed via direct computation, but from a more conceptual point of view, they can also be intuitively understood in terms of two different approximations for the binomial distribution, which lead to two different *anticoncentration* bounds (i.e., upper bounds on point probabilities). On the one hand, a binomial distribution with a low success probability can be closely approximated by a Poisson distribution. The bound in (1.4) is related to the fact that a Poisson random variable with parameter 1 is equal to 1 with probability $1/e$, and this is the maximum possible probability for any Poisson random variable to take any particular nonzero value. On the other hand, a binomial distribution with k trials and success probability $\alpha \in (0, 1/2)$ can be approximated by a Gaussian distribution with standard deviation about $\sqrt{\alpha k}$. The point probabilities of such a binomial distribution are at most about $1/\sqrt{\alpha k}$; this corresponds to the fact that the corresponding Gaussian distribution has density at most about $1/\sqrt{\alpha k}$.

The proofs of Theorems 1.1 and 1.2 both crucially depend on anticoncentration inequalities that vastly generalize the above two observations about binomial distributions. Specifically, we need two general anticoncentration inequalities for low-degree polynomials of independent random variables: (a strengthened form of) a “Poisson-type” anticoncentration inequality due to Fox, Kwan and Sauermaann [8], and bounds on the so-called *polynomial Littlewood–Offord problem* due to Meka, Nguyen, and Vu [18].

Recall that $\text{ind}_r(k, \ell)$ can be defined in terms of quantities of the form $N(G, k, \ell)$, which can be understood as the probability that a random k -vertex subset of G has exactly ℓ edges. It is not hard to interpret the number of edges in a random k -vertex subset of G as a polynomial evaluated at a random vector (namely, at a random point on a “slice of the Boolean hypercube”). Unfortunately, since the entries of this random vector are not independent, one cannot directly apply the aforementioned polynomial anticoncentration inequalities (in fact, it is easy to see that the conclusions of these inequalities are in general false on a slice of the Boolean hypercube). The key new contributions in this paper are several different ways to “transfer” polynomial anticoncentration inequalities to slices of the Boolean hypercube. For the proof of Theorem 1.2, we generalize a coupling lemma due to Kwan, Sudakov, and Tran [14], and combine it with a result of Bollobás and Scott [3] related to discrepancy of hypergraphs. For the proof of Theorem 1.1, we additionally use a classical estimate on hypergeometric distributions due to Ehm [6], which allows us to transfer Poisson-type anticoncentration inequalities to functions of the slice that only depend on a few vertices. Then, we prove a delicate combinatorial lemma, which says

that every hypergraph has a small vertex subset Y that “only sees large matchings”. We use our Poisson-type anticoncentration inequality to understand “what happens inside Y ”, and then after conditioning on this information we apply polynomial Littlewood–Offord bounds (which are effective for hypergraphs with large matchings).

1.1 Further directions

There is nearly unlimited potential to ask more precise questions about the quantities $\text{ind}_r(k, \ell)$. Most obviously, there is the question of removing the “ ε ” in the exponent of Theorem 1.2: in the setting of Theorem 1.2 we conjecture that

$$\text{ind}_r(k, \ell) \leq \frac{C_r}{(\alpha k)^{1/2}} \quad (1.6)$$

for some constant C_r depending only on r (this also appears as [14, Conjecture 5]). In the case $r = 2$, this bound was recently proved by Kwan and Sauermaun [13], via new progress on the so-called *quadratic Littlewood–Offord problem*. Actually, our proof of Theorem 1.2 reduces the general- r case of (1.6) to a well-known conjecture in Littlewood–Offord theory. We introduce the polynomial Littlewood–Offord problem properly, and discuss these aspects further, in Section 2.

One could also ask about the “ ε ” in Theorem 1.1: for each r, k , what is the exact maximum value of $\text{ind}_r(k, \ell)$, among all $\ell \notin \{0, \binom{k}{r}\}$? We wonder if the maximum is always attained at $\ell = 1$. We remark that the exact value of $\text{ind}_2(k, 1)$, for all k , was recently found by Liu, Mubayi, and Reiher [16, Theorem 1.13] (see also [11] for earlier work in the case $k = 4$).

It is also interesting to consider the maximum possible value of $\text{ind}_r(k, \ell)$ among all pairs (k, ℓ) satisfying $\ell \notin \{0, \binom{k}{r}\}$. In the case $r = 2$, it was suggested by Alon, Hefetz, Krivelevich, and Tyomkyn [1] that this maximum value might be $\text{ind}_2(3, 1) = 3/4$.

Finally, it would be very interesting to investigate “stability” in the settings of Theorems 1.1 and 1.2. Although the constant “ $1/e$ ” in Theorem 1.1 is best-possible in general, we conjecture that it can be improved when $\min(\ell, \binom{k}{r} - \ell)$ is not of the form $\binom{k-s}{r-s}$ (a related theorem was very recently proved by Ueltzen [22] in the setting of graph inducibility). Similarly, although the exponent “ $1/2$ ” in Theorem 1.2 is best-possible in general, we conjecture that there is $\delta_r > 0$ such that $\text{ind}_r(k, \ell) \leq k^{-1/2-\delta_r}$ for “generic” ℓ (i.e., for a $1 - o(1)$ fraction of ℓ in the range $0 \leq \ell \leq \binom{k}{r}$), where asymptotics are as $k \rightarrow \infty$, holding r fixed). This seems to be related to a conjecture of Costello [4, Conjecture 3] on “stability” for the polynomial Littlewood–Offord problem.

1.2 Notation

We use standard graph theory notation throughout. For a hypergraph G , we write $e(G)$ for the number of edges in G , and for a vertex subset U , we write $G[U]$ to denote the subgraph of G induced by U .

We also use asymptotic notation throughout. For functions $f = f(n)$ and $g = g(n)$, we write $f = O(g)$ or $f \lesssim g$ to mean that there is a constant C such that $|f| \leq C|g|$, $f = \Omega(g)$ or $f \gtrsim g$ to mean that there is a constant $c > 0$ such that $f(n) \geq c|g(n)|$ for sufficiently large n , and $f = o(g)$ to mean that $f/g \rightarrow 0$ as $n \rightarrow \infty$. Subscripts on asymptotic notation indicate quantities that should be treated as constants.

For parameters $\alpha, \beta_1, \dots, \beta_q$, we write $\alpha \ll \beta_1, \dots, \beta_q$ to mean “ α is sufficiently small in terms of $\beta_1, \dots, \beta_q$ ” (i.e., it is shorthand for a statement of the form “ $\alpha \leq f(\beta_1, \dots, \beta_q)$ ”, for some function f that we do not wish to specify explicitly). Similarly, we write $\alpha \gg \beta_1, \dots, \beta_q$ to mean “ α is sufficiently large in terms of $\beta_1, \dots, \beta_q$ ”.

For a positive integer n and an integer $0 \leq d \leq n$, we write $[n] = \{1, \dots, n\}$ and denote the set of all size- d subsets of $[n]$ by $\binom{[n]}{d}$. For a real number x , the floor and ceiling functions are denoted $\lfloor x \rfloor = \max\{i \in \mathbb{Z} : i \leq x\}$ and $\lceil x \rceil = \min\{i \in \mathbb{Z} : i \geq x\}$. We will, however, sometimes omit floor and ceiling symbols and assume large numbers are integers, when divisibility considerations are not important. All logarithms in this paper without an explicit base are to base e , and the natural numbers $\mathbb{N}$ do not include zero.

For a vector $\vec{x} \in \mathbb{R}^n$, we write $x_1, \dots, x_n$ for its coordinates, and for $W \subseteq [n]$, we write $\vec{x}^W$ to denote the monomial $\prod_{i \in W} x_i$. For a multilinear polynomial $P \in \mathbb{R}[x_1, \dots, x_n]$, we denote the coefficient of the monomial $\vec{x}^W$ by $\hat{P}(W)$.

1.3 Organization of the paper

In Section 2, we introduce the polynomial Littlewood–Offord problem (on anticoncentration of polynomials of independent random variables) and describe the best-known bounds for this problem. These results will play a crucial role in the proofs of both Theorems 1.1 and 1.2, but in order to actually apply them we need a general coupling lemma for polynomials “on a slice of the Boolean hypercube”, which we present in Section 3. In Section 4 we prove Theorem 1.2, using the above tools and a result of Bollobás and Scott [3].

Then, in the rest of the paper, we focus on Theorem 1.1. First, in Section 5 we apply the above tools in a different way, to prove an anticoncentration bound for “sparse” polynomials with “large matchings”. In Section 6, we state and prove our Poisson-type anticoncentration inequality. In Section 7, we show how to use an estimate of Ehm [6] to compare certain functions “on a slice of the Boolean hypercube” with corresponding functions of product distributions. Then, in Section 8, we prove a lemma showing that every hypergraph has a small set of vertices that “only sees large matchings”. Finally, after some technical variance estimates in Section 9, we prove Theorem 1.1 in Section 10.

2 The Polynomial Littlewood–Offord Problem

In this section we introduce the *polynomial Littlewood–Offord problem*, concerning anticoncentration of polynomials of independent random variables. More specifically, let $P \in \mathbb{R}[x_1, \dots, x_k]$ be a k -variable polynomial and let $\xi_1, \dots, \xi_k$ be i.i.d. Rademacher random variables (i.e., $\mathbb{P}[\xi_i = -1] = \mathbb{P}[\xi_i = 1] = 1/2$). What upper bounds can be proved on the maximum point probability

$$\sup_{\ell \in \mathbb{R}} \mathbb{P}[P(\xi_1, \dots, \xi_k) = \ell],$$

in terms of simple combinatorial information about the polynomial P ? The most well-known theorem in this direction is due to Erdős [7]: improving a theorem of Littlewood and Offord [15], Erdős proved that if P is a linear form with at least m nonzero coefficients, then

$$\mathbb{P}[P(\xi_1, \dots, \xi_k) = \ell] \leq \frac{\binom{m}{\lfloor m/2 \rfloor}}{2^m} \lesssim \frac{1}{\sqrt{m}}.$$

By letting $P(x_1, \dots, x_k) = x_1 + \dots + x_m$ and $\ell = 2 \lfloor m/2 \rfloor - m$, we see that this bound is exactly best possible.

For higher-degree polynomials, one cannot hope for a comparable bound in terms of the number of nonzero coefficients: indeed, the multilinear polynomial $P(x_1, \dots, x_k) = (x_1 + x_2)(x_3 + \dots + x_k)$ has $2k - 4$ nonzero coefficients, but we have $\mathbb{P}[P(\xi_1, \dots, \xi_k) = 0] \geq \mathbb{P}[x_1 \neq x_2] = 1/2$. There are a number of different ways that rule out this kind of degenerate situation; in this paper, we will parameterize P by the *matching number* of a certain hypergraph associated with P .

Definition 2.1. For a multilinear polynomial $P \in \mathbb{R}[x_1, \dots, x_k]$ and $a \geq 0$, let $H_a^{(d)}(P)$ be the d -uniform hypergraph on the vertex set $[k]$ with an edge $\binom{[k]}{d}$ whenever the coefficient of the monomial x^I has absolute value strictly greater than a .

Definition 2.2. A *matching* in a hypergraph H is a collection of edges that are pairwise vertex-disjoint. Let $\nu(H)$ be the maximum number of edges in a matching in H .

It was first proved by Razborov and Viola [20] (building on work of Rosiński and Samorodnitsky [21] and Costello, Tao and Vu [5]) that if $\nu(H_0^{(d)}(P)) \geq m$, then

$$\sup_{\ell \in \mathbb{R}} \mathbb{P}[P(\xi_1, \dots, \xi_k) = \ell] \lesssim_d m^{-c_d} \quad (2.1)$$

for some $c_d > 0$ depending only on d . That is to say, if P has many degree- d terms with nonzero coefficients, featuring disjoint sets of variables, then $P(\xi_1, \dots, \xi_k)$ is anticoncentrated.

The bound in (2.1) has since been improved, but in general the best possible bound is still unknown. To ensure that the results in this paper are compatible with potential future improvements, we define a function to describe “the best possible bound for the polynomial Littlewood–Offord problem”, as follows:

Definition 2.3. For $d, m \geq 0$, let

$$\text{LO}_d(m) = \sup_{P, \ell, k} \mathbb{P}[P(\xi_1, \dots, \xi_k) = \ell],$$

where the supremum ranges over all $\ell \in \mathbb{R}$, all $k \in \mathbb{N}$, and all multilinear polynomials $P \in \mathbb{R}[x_1, \dots, x_k]$ of degree at most d with $\nu(H_0^{(d)}(P)) \geq m$, and we take $\xi_1, \dots, \xi_k$ to be i.i.d. Rademacher random variables.

By considering the multilinear polynomial obtained from $(x_1 + \dots + x_m)^d$ by substituting $x_i^2 = 1$ for all i , it is easy to see that

$$\text{LO}_d(m) \gtrsim \frac{1}{\sqrt{m}}.$$

It is widely believed (a conjecture along these lines seems to have been first posed by Nguyen and Vu [18, 20]) that the matching upper bound $\text{LO}_d(m) \lesssim_d 1/\sqrt{m}$ should also hold. In the case $d = 1$ this is a classical result of Littlewood–Offord and Erdős, and in the case $d = 2$ this was recently proved by Kwan and Sauerermann [13] (improving intermediate results by Costello [4]). For general d , the best available bound is due to Meka, Nguyen, and Vu [18] (via a theorem of Kane [12]), as follows.

Theorem 2.4. For any $d, m \in \mathbb{N}$,

$$\text{LO}_d(m) \lesssim_d \frac{(\log m)^{O_d(1)}}{\sqrt{m}}.$$

We remark that the “ ε ” in Theorem 1.2 is entirely due to the logarithmic factors in Theorem 2.4; if we knew that $\text{LO}_d(m) \lesssim_d 1/\sqrt{m}$, we would be able to obtain the optimal result (1.6). Also, we remark that while the proof of Theorem 2.4 is a little involved, Razborov and Viola’s proof of the bound $\text{LO}_d(m) \lesssim_d m^{-c_d}$ is very simple, and this weaker bound is enough for our proof of Theorem 1.1.

We will need a version of Theorem 2.4 that takes terms of all degrees into account (not just the degree- d terms), as follows.

Corollary 2.5. Consider a multilinear polynomial $P \in \mathbb{R}[x_1, \dots, x_k]$ of degree at most $d \geq 0$, and for each $f \in \{0, \dots, d\}$ let b_f be an upper bound on the absolute values of all degree- f coefficients. Suppose that for some $f \in \{0, \dots, d\}$ and $t \in \mathbb{N}$ we have $\nu(H_{a(f,t)}^{(f)}) \geq t$, where $a(f, t) = tb_{f+1} + t^2b_{f+2} + \dots + t^{d-f}b_d$. Then,

$$\sup_{\ell \in \mathbb{R}} \mathbb{P}[P(\xi_1, \dots, \xi_k) = \ell] \lesssim_d \sup_{f \leq d} \text{LO}_f(\Omega_d(t)).$$

Proof. We prove the desired statement by induction on d . The case $d = 0$ holds vacuously, since $\nu(H_a^{(0)}(P)) \leq 1$ for any P, a . Fix $d \geq 1$ and suppose that the desired bound is true for smaller d . Fix a degree- d multilinear polynomial $P \in \mathbb{R}[x_1, \dots, x_k]$, and suppose that for some $f \leq d$ we have $\nu(H_{a(f,t)}^{(f)}) \geq t$.

We split into cases. First, suppose that $\nu(H_0^{(d)}(P)) \geq t/(2d)$. In this case we have

$$\sup_{\ell \in \mathbb{R}} \mathbb{P}[P(\xi_1, \dots, \xi_k) = \ell] \leq \text{LO}_d(t/(2d)),$$

as desired.

Otherwise, we have $\nu(H_0^{(d)}(P)) < t/(2d)$. This means that $f \leq d - 1$. Consider a maximum matching M in $H_0^{(d)}(P)$, and let $I \subseteq [k]$ be the set of vertices in this matching, so $|I| = d|M| \leq t/2$. Let $\tilde{\xi}[I] := (\xi_i)_{i \in I}$, and consider any outcome $\tilde{\xi}[I] = (\xi_i)_{i \in I} \in \{-1, 1\}^I$. Let $P_{\tilde{\xi}[I]} \in \mathbb{R}[x_i : i \notin I]$ be the multilinear polynomial obtained from $P(x_1, \dots, x_k)$ by substituting $x_i = \xi_i$ for all $i \in I$. Since M is a maximum matching, $P_{\tilde{\xi}[I]}$ has degree at most $d - 1$. Also, note that the multilinear degree- f coefficients in $P_{\tilde{\xi}[I]}$ differ from the corresponding coefficients in P by at most

$$\binom{|I|}{1} b_{f+1} + \binom{|I|}{2} b_{f+2} + \dots + \binom{|I|}{d-f} b_d \leq a(f, t) - a(f, t/2),$$

so all the edges of the induced subgraph $H_{a(f, t)}^{(f)}(P)[[k] \setminus I]$ are also edges of $H_{a(f, t/2)}^{(f)}(P_{\tilde{\xi}[I]})$, and therefore,

$$\nu(H_{a(f, t/2)}^{(f)}(P_{\tilde{\xi}[I]})) \geq t - |I| \geq t/2.$$

The induction hypothesis then yields

$$\sup_{\ell \in \mathbb{R}} \mathbb{P}[P(\xi_1, \dots, \xi_k) = \ell] \leq \sup_{\tilde{\xi}[I]} \sup_{\ell \in \mathbb{R}} \mathbb{P}[P_{\tilde{\xi}[I]}(\xi_i : i \notin I) = \ell \mid \tilde{\xi}[I]] \lesssim_d \sup_{f \leq d-1} \text{LO}_f(\Omega_d(t)),$$

as desired. ■

3 Describing the Slice via a Product Measure

In this section, we record a general lemma describing polynomials of “slice” measures in terms of polynomials of independent Rademacher random variables. This is necessary to apply the results from Section 2 to study the parameters $\text{ind}_d(k, \ell)$, which can naturally be understood as being about polynomials evaluated on a slice measure.

Specifically, for any hypergraph G on the vertex set $[n]$, and any k -vertex subset U , we can express

$$e(G[U]) = \sum_{W \in E(G)} \bar{\sigma}^W,$$

where $E(G)$ is the set of edges of G , and $\bar{\sigma}$ is the vector in $\{0, 1\}^n$ with exactly k ones, defined by setting $\sigma_i = 1$ if $i \in U$ and $\sigma_i = 0$ otherwise. When U is a uniformly random k -vertex subset, the corresponding $\bar{\sigma}$ is a uniformly random vector in $\{0, 1\}^n$ with exactly k ones (this is called a “slice of the Boolean hypercube”).

Definition 3.1. For an r -uniform hypergraph G on the vertex set $[n]$, let $\lambda_G \in \mathbb{R}[x_1, \dots, x_n]$ denote the polynomial $\sum_{W \in E(G)} \bar{x}^W$. In other words, the coefficient $\hat{\lambda}_G(W)$ of $\bar{x}^W$ is 1 if W is an edge of G , and 0 if W is not an edge of G .

Definition 3.2. $\text{Slice}(n, k)$ denotes the subset of vectors in $\{0, 1\}^n$ with exactly k different 1-entries. We write $\bar{\sigma} \sim \text{Slice}(n, k)$ to mean that the random vector $\bar{\sigma}$ is uniformly distributed on $\text{Slice}(n, k)$.

Now, note that we can obtain $\text{Slice}(n, k)$ (for $n \geq 2k$) by first randomly choosing k pairs of disjoint vertices, then flipping an unbiased coin for each pair to decide which of the pair to actually take (as a 1-entry). We introduce notation for this, as follows:

Definition 3.3. Consider $k, n \in \mathbb{N}$. Let

$$\vec{v} = (v_1(-1), v_1(1), \dots, v_k(-1), v_k(1))$$

be a uniformly random sequence of $2k$ distinct elements of $[n]$, and let $V_{\vec{v}}$ be the set of elements in $\vec{v}$. Independently from $\vec{v}$, let $\tilde{\xi} = (\xi_1, \dots, \xi_k)$ be a sequence of k i.i.d. Rademacher random variables. Let $U_{\vec{v}, \tilde{\xi}} = \{v_1(\xi_1), \dots, v_k(\xi_k)\}$, so $U_{\vec{v}, \tilde{\xi}}$ is a uniformly random subset of k elements of $[n]$.

Define $\vec{\sigma}_{\vec{v}, \vec{\xi}} = (\sigma_1, \dots, \sigma_n) \in \{0, 1\}^n$ by

$$\sigma_v = \begin{cases} 1 & \text{if } v \in U_{\vec{v}, \vec{\xi}}, \\ 0 & \text{otherwise,} \end{cases}$$

so $\vec{\sigma}_{\vec{v}, \vec{\xi}} \sim \text{Slice}(n, k)$.

The setup in Definition 3.3 allows us to interpret a polynomial on the slice as a polynomial in Rademacher random variables, as follows (a similar slightly weaker result appears as [14, Lemma 2.8]).

Lemma 3.4. Recall the setup in Definition 3.3 and consider an n -variable multilinear polynomial $\lambda = \sum_{W \subseteq [n]} \hat{\lambda}(W) \vec{x}^W \in \mathbb{R}[x_1, \dots, x_n]$. Then, $\lambda(\vec{\sigma}_{\vec{v}, \vec{\xi}})$ is a multilinear polynomial of $\vec{\xi}$, with coefficients that depend on $\vec{v}$. Specifically, we have

$$\lambda(\vec{\sigma}_{\vec{v}, \vec{\xi}}) = \sum_{I \subseteq [k]} A_{\vec{v}}(I) \vec{\xi}^I,$$

for coefficients $A_{\vec{v}}(I)$ given explicitly by

$$A_{\vec{v}}(I) = \sum_{W \in \mathcal{W}_{\vec{v}}(I)} (-1)^{|W \cap \{v_i(-1): i \in I\}|} \hat{\lambda}(W) 2^{-|W|},$$

where $\mathcal{W}_{\vec{v}}(I)$ is the collection of all subsets $W \subseteq V_{\vec{v}}$ satisfying $|W \cap \{v_i(-1), v_i(1)\}| = 1$ for every $i \in I$.

Remark 3.5. In particular, if λ has degree at most d and all coefficients of λ have absolute value at most q , then for any $I \subseteq [k]$, we have

$$|A_{\vec{v}}(I)| \leq q \cdot |\mathcal{W}_{\vec{v}}(I) \cap \{W \subseteq [n] : |W| \leq d\}| \leq q 2^{|I|} n^{d-|I|}.$$

Additionally, if $|I| > d$, then $\mathcal{W}_{\vec{v}}(I) \cap \{W \subseteq [n] : |W| \leq d\} = \emptyset$, so that $A_{\vec{v}}(I) = 0$.

Proof. For convenience of notation, write $\vec{\sigma} = \vec{\sigma}_{\vec{v}, \vec{\xi}}$. Then, $\sigma_{v_i(1)} = (1 + \xi_i)/2$ and $\sigma_{v_i(-1)} = (1 - \xi_i)/2$ for all $i \in [k]$ and $\sigma_w = 0$ for all other w . Let $\mathcal{W}_{\vec{v}}$ denote the set of all subsets $W \subseteq V_{\vec{v}}$, which satisfy $|W \cap \{v_i(-1), v_i(1)\}| \leq 1$ for each i . Note that if $W \notin \mathcal{W}_{\vec{v}}$, then $\vec{\sigma}^W = 0$. Therefore,

$$\begin{aligned} \lambda(\vec{\sigma}) &= \sum_{W \subseteq [n]} \hat{\lambda}(W) \vec{\sigma}^W = \sum_{W \in \mathcal{W}_{\vec{v}}} \hat{\lambda}(W) \vec{\sigma}^W \\ &= \sum_{W \in \mathcal{W}_{\vec{v}}} \hat{\lambda}(W) 2^{-|W|} \prod_{i: v_i(1) \in W} (1 + \xi_i) \prod_{j: v_j(-1) \in W} (1 - \xi_j) \\ &= \sum_{I \subseteq [k]} \left(\sum_{W \in \mathcal{W}_{\vec{v}}(I)} (-1)^{|W \cap \{v_i(-1): i \in I\}|} \hat{\lambda}(W) 2^{-|W|} \right) \vec{\xi}^I = \sum_{I \subseteq [k]} A_{\vec{v}}(I) \vec{\xi}^I. \end{aligned}$$

■

4 A Strong Bound for Edge-Statistics “In the Bulk”

In this section, we prove Theorem 1.2. It will be a simple consequence of the following anticoncentration inequality for edge-statistics of dense hypergraphs.

Lemma 4.1. Let $n = 2k$, let $\beta \in (0, 1/2]$, and let G be an n -vertex r -uniform hypergraph with $\beta \binom{n}{r} \leq e(G) \leq (1 - \beta) \binom{n}{r}$. Then, for U a uniformly random subset of k vertices of G , we have

$$\sup_{\ell \in \mathbb{R}} \mathbb{P}[e(G[U]) = \ell] \lesssim_r \sup_{f \leq r} \text{LO}_f(\Omega_r(\beta n)).$$

Before proving Lemma 4.1, we see how to deduce Theorem 1.2.

Proof of Theorem 1.2 given Lemma 4.1. Recall that we are to prove that $\text{ind}_r(k, \ell) \leq (\alpha k)^{-1/2+\varepsilon}$, where $\text{ind}_r(k, \ell) = \lim_{n \rightarrow \infty} N_r(n, k, \ell) / \binom{n}{k}$ and $N_r(n, k, \ell)$ is the maximum possible number of k -vertex subsets of an n -vertex graph that induce exactly ℓ edges. Let $n = 2k$, let G be an n -vertex graph, and let U be a uniformly random subset of k vertices of G . Due to the fact that $N_r(n, k, \ell) / \binom{n}{k}$ is nonincreasing in n , it suffices to prove that

$$\mathbb{P}[e(G[U]) = \ell] \leq \frac{1}{(\alpha k)^{1/2-\varepsilon}},$$

when αk is sufficiently large in terms of ε, r .

We may assume that G has at least $\alpha \binom{k}{r}$ edges and at least $\alpha \binom{k}{r}$ non-edges; otherwise, it is impossible to have $e(G[U]) = \ell$. Recalling that $n = 2k$ (so $\binom{k}{r} \gtrsim \binom{n}{r}$), this means that the assumption in Lemma 4.1 is satisfied for some $\beta \gtrsim \alpha$, and Lemma 4.1 implies that

$$\mathbb{P}[e(G[U]) = \ell] \lesssim_r \sup_{f \leq r} \text{LO}_f(\Omega_r(\alpha k)).$$

The desired result then follows from Theorem 2.4. ■

We need a few key ingredients to prove Lemma 4.1. First, we need a lemma essentially due to Bollobás and Scott [3].

Definition 4.2. Consider an r -uniform hypergraph G on the vertex set $[n]$. For $W \in \binom{[n]}{r}$, let $\widehat{G}(W) = 1$ when W is an edge and let $\widehat{G}(W) = 0$ when W is not an edge. For $s \in [r]$, define

$$Q_s(G) = \sum_{\vec{x}} \left| \sum_W (-1)^{|W \cap \{x_1(-1), \dots, x_s(-1)\}|} \widehat{G}(W) \right|,$$

where the first sum is over sequences $\vec{x} = (x_1(-1), x_1(1), \dots, x_s(-1), x_s(1))$ of $2s$ distinct vertices, and the second sum is over all $W \in \binom{[n]}{r}$ that satisfy $|W \cap \{x_i(-1), x_i(1)\}| = 1$ for every $i \in [s]$.

Lemma 4.3. Consider r, n satisfying $n \geq 3r$, and consider an r -uniform hypergraph G on the vertex set $[n]$, satisfying $\beta \binom{n}{r} \leq e(G) \leq (1 - \beta) \binom{n}{r}$. Then, for some $s \in [r]$, we have

$$Q_s(G) \gtrsim_r \beta n^{r+s}.$$

Proof. Let $p := e(G) / \binom{n}{r}$ denote the density of G . Since $\beta \binom{n}{r} \leq e(G) \leq (1 - \beta) \binom{n}{r}$, we have $\beta \leq p \leq 1 - \beta$. Define the function $f : \binom{[n]}{r} \rightarrow \mathbb{R}$ by

$$f(W) = \begin{cases} 1 - p & \text{if } W \text{ is an edge of } G, \\ -p & \text{if } W \text{ is not an edge of } G. \end{cases}$$

Then, we have

$$\|f\|_1 = \left(e(G) \cdot (1 - p) + \left(\binom{n}{r} - e(G) \right) \cdot p \right) = 2p(1 - p) \cdot \binom{n}{r} \geq 2\beta(1 - \beta) \cdot \binom{n}{r} \gtrsim \beta n^r.$$

Let $(n)_{2s}$ denote the falling factorial $n(n-1)\dots(n-2s+1)$ (i.e., the number of ways to choose a sequence of $2s$ different vertices); in [3], Bollobás and Scott define (actually, they first define the W -vector in a slightly different way, and prove that the two definitions are equivalent in [3, Lemma 8]) the “ W -vector” of f to be the vector $(q_0(f), \dots, q_r(f))$ given by

$$q_s(f) = \frac{1}{(n)_{2s}} \binom{n-2s}{r-s}^{-1} \sum_{\vec{x}} \left| \sum_W (-1)^{|W \cap \{x_1(-1), \dots, x_s(-1)\}|} f(W) \right|,$$

where the first sum is over sequences $\vec{x} = (x_1(-1), x_1(1), \dots, x_s(-1), x_s(1))$ of $2s$ distinct vertices, and the second sum is over all $W \in \binom{[n]}{r}$ that satisfy $|W \cap \{x_i(-1), x_i(1)\}| = 1$ for every $i \in [s]$. Note that $q_s(f)$ is not affected by adding a constant function to f , so

$$q_s(f) = \frac{1}{(n)_{2s}} \binom{n-2s}{r-s}^{-1} Q_s(G) \lesssim_r n^{-r-s} Q_s(G)$$

and, in particular, $q_0(f) = 0$.

Then, [3, Lemma 9] says that

$$q_0(f) + \dots + q_r(f) \gtrsim_r n^{-r} \|f\|_1;$$

the desired result follows. ■

Next, the following technical lemma shows that random variables of a certain type are unlikely to be very small (this lemma is stated in slightly more general form than we need, as it will also be applied again later in the paper). It is proved by a simple application of Chebyshev’s inequality. Given $f \in \mathbb{N}$ and a set V , an *ordered f -subset* of V is a sequence of f distinct elements of V . We say that two ordered f -subsets $\vec{X}, \vec{Y}$ of V are *disjoint* if there is no element of V that appears in both $\vec{X}$ and $\vec{Y}$.

Lemma 4.4. Consider $f, m \in \mathbb{N}$ and a set V , with $2f \leq mf \leq |V|$. For each $i \in [m]$, let $\mathcal{F}_i$ be a collection of at least $\gamma|V|^f$ ordered f -subsets of V , and let $\vec{X}_1, \dots, \vec{X}_m$ be uniformly random pairwise disjoint ordered f -subsets of V . Let N be the number of i such that $\vec{X}_i \in \mathcal{F}_i$. Then,

$$\mathbb{P}[N < \gamma m/2] \lesssim_f \frac{1}{\gamma m}.$$

Proof. Let $\mathbb{1}_i$ be the indicator random variable for the event that $\vec{X}_i \in \mathcal{F}_i$, and write $n = |V|$. By linearity of expectation,

$$\mathbb{E}[N] = \sum_{i=1}^m \mathbb{E}[\mathbb{1}_i] = \sum_{i=1}^m \frac{|\mathcal{F}_i|}{n(n-1)\dots(n-f+1)} \geq \sum_{i=1}^m \frac{|\mathcal{F}_i|}{n^f} \geq \gamma m.$$

Then, for all $i, j \in [m]$ with $i \neq j$, we have

$$\begin{aligned} \text{Cov}[\mathbb{1}_i, \mathbb{1}_j] &\leq \frac{|\mathcal{F}_i| \cdot |\mathcal{F}_j|}{n(n-1)\dots(n-2f+1)} - \left(\frac{|\mathcal{F}_i|}{n(n-1)\dots(n-f+1)} \right) \left(\frac{|\mathcal{F}_j|}{n(n-1)\dots(n-f+1)} \right) \\ &\lesssim_f \frac{|\mathcal{F}_i|}{n^f} \cdot \frac{|\mathcal{F}_j|}{n^f} \cdot \frac{1}{n}, \\ \text{Var}[N] &\leq \sum_i \mathbb{E}[\mathbb{1}_i] + \sum_{i \neq j} \text{Cov}[\mathbb{1}_i, \mathbb{1}_j] \\ &\lesssim_f \sum_i \frac{|\mathcal{F}_i|}{n^f} + \frac{1}{n} \cdot \left(\sum_i \frac{|\mathcal{F}_i|}{n^f} \right)^2 \leq \mathbb{E}[N] + \frac{\mathbb{E}[N]^2}{n}. \end{aligned}$$

By Chebyshev's inequality, we deduce the desired result

$$\mathbb{P}[N < \gamma m/2] \leq \mathbb{P}[N < \mathbb{E}[N]/2] \leq \frac{\text{Var}[N]}{(\mathbb{E}[N]/2)^2} \lesssim_f \frac{1}{\mathbb{E}[N]} + \frac{1}{n} \lesssim \frac{1}{\gamma m}.$$

Given the conclusion of Lemma 4.3, we can apply Lemma 4.4 in the setting of Lemma 3.4, as follows.

Lemma 4.5. Consider $r, k \in \mathbb{N}$ satisfying $k \geq r$, and let $n = 2k$. Consider an r -uniform hypergraph G on the vertex set $[n]$, and its corresponding polynomial λ_G . Suppose $Q_s(G) \geq \beta n^{r+s}$ for some $\beta > 0$ and $s \in [r]$. Recall the coefficients $A_{\vec{v}}(I)$ from Lemma 3.4 (defined in terms of a random sequence $\vec{v}$).

Then, except with probability at most $O_r(1/(\beta n))$ over the randomness of $\vec{v}$, the following holds: there exist $t \gtrsim_r \beta n$ disjoint s -sets $I_1, \dots, I_t \subseteq [k]$ such that $|A_{\vec{v}}(I_j)| \gtrsim_r \beta n^{r-s}$ for each j .

Proof. For a sequence $\vec{x} = (x_1(-1), x_1(1), \dots, x_s(-1), x_s(1))$ of $2s$ distinct indices, let

$$a(\vec{x}) = \left| \sum_W (-1)^{|W \cap \{x_1(-1), \dots, x_s(-1)\}|} \widehat{G}(W) \right|,$$

where the sum is over all $W \in \binom{[n]}{r}$ that satisfy $|W \cap \{x_i(-1), x_i(1)\}| = 1$ for every $i \in [s]$. By definition, we have $Q_s(G) = \sum_{\vec{x}} a(\vec{x})$. Note that we always have $a(\vec{x}) \leq 2^s n^{r-s}$, so it must be the case that $a(\vec{x}) \gtrsim_s \beta n^{r-s}$ for at least $\Omega_s(\beta n^{2s})$ different $\vec{x}$. Denote the set of all such $\vec{x}$ by $\mathcal{F}$.

Now, let $m = \lfloor k/s \rfloor \gtrsim_s n$. Recall the random sequence $\vec{v} = (v_1(-1), v_1(1), \dots, v_k(-1), v_k(1))$ from Definition 3.3, and for each $j \in [m]$, let

$$I_j = \{s(j-1) + 1, s(j-1) + 2, \dots, sj\},$$

$$\vec{X}_j = (v_{s(j-1)+1}(-1), v_{s(j-1)+1}(1), \dots, v_{sj}(-1), v_{sj}(1)).$$

So, in the language of Lemma 4.4, $\vec{X}_1, \dots, \vec{X}_m$ are uniformly random disjoint ordered $2s$ -subsets of $[n]$. If $\vec{X}_j \in \mathcal{F}$, then $|A_{\vec{v}}(I_j)| = 2^{-r} a(\vec{X}_j) \gtrsim_r \beta n^{r-s}$. So, the desired result follows from Lemma 4.4, with $f = 2s$ and $\gamma \gtrsim_r \beta$ and $\mathcal{F}_j = \mathcal{F}$ for all $j \in [m]$. ■

Now, it is straightforward to deduce Lemma 4.1.

Proof of Lemma 4.1. By Lemma 4.3 and Lemma 4.5, there is $s \in [r]$ and $t \gtrsim_r \beta n$ such that, except with probability $O_r(1/(\beta n)) \lesssim \sup_{f \leq d} \text{LO}_f(\beta n)$ over the randomness of $\vec{v}$, the following holds: there are at least t disjoint s -sets $I_1, \dots, I_t \subseteq [k]$ such that $|A_{\vec{v}}(I_j)| > r 2^r t n^{r-s-1}$ for each j . Condition on such an outcome of $\vec{v}$.

Recalling Remark 3.5, we have $|A_{\vec{v}}(I)| \leq 2^r n^{r-f} =: b_f$ for all $I \in \binom{[k]}{f}$. So, in the notation of Corollary 2.5, we have

$$|A_{\vec{v}}(I_j)| > r 2^r t n^{r-s-1} \geq a(s, t)$$

for each $j \in [t]$. We can therefore apply Corollary 2.5 using the randomness of $\vec{\xi}$ to obtain the desired result. ■

5 A Littlewood–Offord-Type Inequality for “Sparse” Polynomials on the Slice

For the proof of Theorem 1.1, we will need one further application of polynomial Littlewood–Offord bounds. Here we consider a degree- d polynomial λ , which is “sparse” (almost all its top-degree coefficients are zero), and give a bound in terms of the matching number of the hypergraph of nonzero degree- d coefficients of λ . This matching number is *a priori* quite different to the matching number of

the polynomial obtained from Lemma 3.4; we take advantage of our sparseness assumption to relate the two (this is actually a somewhat delicate matter, and requires a careful minimality argument).

Recall that the notation $\alpha \gg \beta_1, \dots, \beta_q$ means “ α is sufficiently large in terms of $\beta_1, \dots, \beta_q$ ”.

Lemma 5.1. For $d, q \in \mathbb{N}$, there is $\delta > 0$ such that the following holds. Consider $R, k, m, n \in \mathbb{N}$ with $2k \leq n \leq Rk$ and let $\lambda \in \mathbb{R}[x_1, \dots, x_n]$ be an n -variable multilinear polynomial with degree at most d , whose coefficients all lie in $\{0, \dots, q\}$. Suppose $v(H_0^{(d)}(\lambda)) \geq m$ and suppose that λ has at most δn^d nonzero degree- d terms. Then, for $\bar{\sigma} \sim \text{Slice}(n, k)$, provided that $k \gg d, q, R$, we have

$$\sup_{\ell \in \mathbb{R}} \mathbb{P}[\lambda(\bar{\sigma}) = \ell] \lesssim_{d,R} \sup_{f \leq d} \text{LO}_f(\Omega_{R,d,q}(m)) \lesssim_{d,R,q} \frac{1}{m^{1/3}}.$$

For our proof of Lemma 5.1, we use the following simple lemma, showing that if a graph G has a large matching, then a random subset of its vertices also typically has a large matching.

Lemma 5.2. Let G be a d -uniform hypergraph whose number of vertices is between $2k$ and Rk and satisfies $v(G) \geq m$. Let U be a uniformly random k -vertex subset, for some $k \geq 2d$. Then, except with probability $O_{d,R}(1/m)$, we have $v(G[U]) \gtrsim_{d,R} m$.

Proof. Let n be the number of vertices in G . First, note that we have the trivial upper bound $m \leq n/d \lesssim_R k$. By shrinking m if necessary, note that we can assume $k/m \gg q, \delta, d, R$ (we would only shrink m to a value of the form $\Omega_{R,d,q}(m)$, so the form of the final bound would be unchanged).

Let $E_1, \dots, E_m$ be the edges of a matching in G . For each $i \in [m]$, let $\mathbb{1}_i$ be the indicator random variable for the event $E_i \subseteq U$, and let $N = \mathbb{1}_1 + \dots + \mathbb{1}_m$. Then, with similar calculations as in Lemma 4.4 we have (actually, with slightly more careful calculations one can see that $\text{Cov}[\mathbb{1}_i, \mathbb{1}_j] \leq 0$; this is essentially done in the proof of Proposition 9.1)

$$\mathbb{E}[\mathbb{1}_i] = \frac{\binom{n-d}{k-d}}{\binom{n}{k}} \gtrsim_{d,R} 1, \quad \text{Cov}[\mathbb{1}_i, \mathbb{1}_j] = \frac{\binom{n-2d}{k-2d}}{\binom{n}{k}} - \left(\frac{\binom{n-d}{k-d}}{\binom{n}{k}} \right)^2 \lesssim_{d,R} 1/k,$$

and therefore

$$\mathbb{E}[N] \gtrsim_{d,R} m, \quad \text{Var}[N] \lesssim_{d,R} m + m^2/k \lesssim_R m$$

(using that $m \leq (Rk)/d \lesssim_R k$). The desired result follow from Chebyshev's inequality. $\blacksquare$

Now, we prove Lemma 5.1.

Proof of Lemma 5.1. Recall the coefficients $A_{\bar{v}}(\ell)$ from Lemma 3.4, defined in terms of a random sequence $\bar{v} = (v_1(-1), v_1(1), \dots, v_k(-1), v_k(1))$. Let $\bar{v}(1) = (v_1(1), \dots, v_k(1))$. By Lemma 5.2, over the randomness of $\bar{v}(1)$, except with probability $O_{d,R}(1/m) \lesssim_{d,R} \text{LO}_d(m)$, the subgraph of $H_0^{(d)}(\lambda)$ induced by the vertices of $\bar{v}(1)$ has a matching of size $m' = \Omega_{R,d}(m)$. Condition on such an outcome of $\bar{v}(1)$, and let $E_1, \dots, E_{m'}$ be the edges of this matching. Given such a choice of $\bar{v}(1)$, condition on an arbitrary outcome of $V_{\bar{v}}$ such that $V_{\bar{v}} \supseteq \{v_1(1), \dots, v_k(1)\}$. Note that there is still some randomness remaining in $\bar{v}$: namely, $v_1(-1), \dots, v_k(-1)$ is a uniformly random ordering of the vertices in $U(-1) := \{v_1(-1), \dots, v_k(-1)\}$.

For a set $W \subseteq V_{\bar{v}}$, let $B(W)$ be the number of nonzero coefficients $\hat{\lambda}(Z)$ of λ , among all size- d sets Z satisfying $W \subseteq Z \subseteq V_{\bar{v}}$. By assumption,

$$B(\emptyset) \leq \delta n^d. \tag{5.1}$$

Consider $M \gg \delta, q, d$. For all $j \in [m']$, we have $B(E_j) = 1 = (n/M)^{d-|E_j|}$, so we may define $F_j \subseteq E_j$ to be a minimal subset of E_j satisfying $B(F_j) \geq (n/M)^{d-|F_j|}$. Recalling (5.1), and assuming $\delta < M^{-d}$, we have $F_j \neq \emptyset$. Assume without loss of generality that all the sets $F_1, \dots, F_{m'/d}$ have the same size $d' \in [d]$.

Next, for each $j \in [m'/d]$, let $I_j = \{i : v_i(1) \in F_j\} \subseteq [k]$. For each $j \in [m'/d]$, we have $|I_j| = |F_j| = d'$, since $F_j \subseteq E_j \subseteq \{v_1(1), \dots, v_k(1)\}$. Define the sequence $\bar{X}_j = (v_i(-1) : i \in I_j)$ and its underlying set $X_j = \{v_i(-1) : i \in I_j\}$.

Then, note that each coefficient $A_{\bar{v}}(I_j)$ depends only on the randomness of $\bar{X}_j$. Recall that since we have conditioned on a choice of the set $U(-1)$, the number of possibilities for $\bar{X}_j$ is $k(k-1) \cdots (k-d'+1) \leq k^{d'}$.

Claim. Let $M \gg q, d, R$ and $\alpha \ll q, M, R$. For each $j \in [m'/d]$, let $\mathcal{F}_j$ be the collection of all possible outcomes of $\bar{X}_j$ that satisfy $A_{\bar{v}}(I_j) \geq \alpha n^{d-d'}$. Then, $|\mathcal{F}_j| \gtrsim k^{d'}$.

Proof of claim. Consider an arbitrary $j \in [m'/d]$ and let $\varepsilon = 1/M^{d-d'}$, so that $B(F_j) \geq \varepsilon n^{d-d'}$. By the definition of F_j , for every proper subset $Y \subsetneq F_j$, we have $B(Y) < (n/M)^{d-|Y|} \leq (\varepsilon/M)n^{d-|Y|}$. The sum of $B(T)$ over all size- d' sets $T \supseteq Y$ satisfying $T \setminus Y \subseteq U(-1)$ is therefore at most

$$\binom{d-|Y|}{d'-|Y|} B(Y) \lesssim_{d,R} \frac{\varepsilon}{M} n^{d-d'} \binom{k-|Y|}{d'-|Y|}.$$

That is to say, if T is uniformly random subject to the constraints $T \supseteq Y$ and $T \setminus Y \subseteq U(-1)$, then $\mathbb{E}[B(T)] \lesssim_{d,R} (\varepsilon/M)n^{d-d'}$, so by Markov's inequality, provided M is sufficiently large in terms of d, R , we have

$$\mathbb{P}[B(T) > (\varepsilon/M^{1/3})n^{d-d'}] \leq M^{-1/3}. \quad (5.2)$$

Now, say that a set $X \subseteq U(-1)$ is good if $B(T) \leq (\varepsilon/M^{1/3})n^{d-d'}$ for every size- d' subset $T \subseteq F_j \cup X$ containing at least one vertex of X . If X_j is good, then

$$\begin{aligned} A_{\bar{v}}(I_j) &\geq \sum_{W \supseteq F_j} \hat{\lambda}(W) 2^{-|W|} - \sum_{W \not\supseteq F_j} \hat{\lambda}(W) 2^{-|W|} \\ &\geq \left(B(F_j) - d' \binom{2k-1}{d-d'-1} \right) \cdot 2^{-d} - q \cdot \max_T B(T) \\ &\gtrsim_d \varepsilon n^{d-d'} - (\varepsilon/M^{1/3})n^{d-d'} \\ &\gtrsim_{d,M} n^{d-d'}, \end{aligned}$$

where the sums on the first line are over all $W \subseteq V_0$, which satisfy $|W \cap \{v_i(-1), v_i(1)\}| = 1$ for every $i \in I_j$, and the “max” in the second line is over all size- d' subsets $T \subseteq F_j \cup X_j$ containing at least one vertex of X_j .

So, it suffices to show that X_j is good with probability $\Omega(1)$ (recall that the randomness comes from a uniformly random permutation of the elements of $U(-1)$). To this end, for each proper subset $J \subsetneq I_j$, let $Y_J = \{v_i(1) : i \in J\} \subseteq F_j$ and let T_J be the random size- d' set containing $v_i(1)$ for $i \in J$ and $v_i(-1)$ for $i \notin J$. If $B(T_J) \leq (\varepsilon/M^{1/3})n^{d-d'}$ for all $J \subsetneq I_j$, then X_j is good. But note that each T_J is a uniformly random set satisfying $T_J \supseteq Y$ and $T \setminus Y \subseteq U(-1)$, so by (5.2) and a union bound over all $J \subsetneq I_j$, we see that $\mathbb{P}[X_j \text{ is not good}] \leq 2^d \cdot M^{-1/3} \leq 1/2$. ■

Now, recall from Remark 3.5 that we always have $|A_{\bar{v}}(I)| \leq q2^f(2k)^{d-f} =: b_f$ for all $I \in \binom{[k]}{f}$. As in Corollary 2.5, let $a(d', t) = tb_{d'+1} + \cdots + t^{d-d'}b_d \lesssim_{q,d} mk^{d-d'-1}$.

In the notation of Lemma 4.4, note that $\bar{X}_1, \dots, \bar{X}_{m'/d}$ are uniform random ordered d' -subsets of $U(-1)$. So, by Lemma 4.4 (with $f = d'$ and $\gamma = \Omega(1)$) and the above claim, except with probability $O(1/m) \lesssim \text{LO}_d(m)$, we have $|A_{\bar{v}}(I_j)| \gtrsim_{q,M,d,R} k^{d-d'}$. Therefore, $|A_{\bar{v}}(I_j)| > a(d', t)$ for $t = \Omega(m'/d)$ different j (here we are using that $k/m \gg q, \delta, d, R$). Condition on an outcome of $\bar{v}$ such that this is the case.

We may then apply Corollary 2.5, using the randomness of $\bar{\xi}$, to obtain the desired conclusion.

6 A Poisson-Type Anticoncentration Inequality

Another key ingredient for the proof of Theorem 1.1 is a “Poisson-type” anticoncentration inequality for polynomials of independent random variables, strengthening a result of Fox, Kwan, and Sauermann [8, Theorem 1.8].

Theorem 6.1. Fix $\gamma > 0$, and suppose p is sufficiently small in terms of γ . Let $F \in \mathbb{R}[x_1, \dots, x_s]$ be a multilinear polynomial that has nonnegative coefficients, and zero constant coefficient. Let $\beta_1, \dots, \beta_s$ be i.i.d. **Bernoulli**(p) random variables. Then for any $t \in [0, \infty)$ and $\ell > 3^s t$, we have

$$\mathbb{P}[|F(\beta_1, \dots, \beta_s) - \ell| \leq t] \leq \frac{1}{e} + \gamma.$$

Proof. We prove by induction on s that whenever $\ell > 3^s t$ and $F \in \mathbb{R}[x_1, \dots, x_s]$ is a multilinear polynomial that has nonnegative coefficients and zero constant coefficient, we have

$$\mathbb{P}[|F(\vec{\beta}) - \ell| \leq t] \leq \max_{n \in \mathbb{N}} \mathbb{P}[X_{n,p} = 1],$$

where $X_{n,p} \sim \text{Bin}(n, p)$. This suffices, because elementary estimates show that the right-hand side of the above expression converges to $1/e$ as $p \rightarrow 0$ (see [8, Lemma 3.3]).

The desired statement is vacuously true for $s = 0$, so fix some $s \geq 1$ and assume the desired statement is true for smaller s . Write $\|\vec{\beta}\|_0$ for the number of nonzero entries of $\vec{\beta}$.

Let a_i be the coefficient of x_i in $F(x_1, \dots, x_s)$. First, if we have $a_i > (\ell + t)/2$ for all i , then we can only have $|F(\vec{\beta}) - \ell| \leq t$ if $\|\vec{\beta}\|_0 = 1$. Indeed, if we had $\|\vec{\beta}\|_0 = 0$ we would have $F(\vec{\beta}) = 0$, and if we had $\|\vec{\beta}\|_0 \geq 2$ we would have $F(\vec{\beta}) - \ell > 2(\ell + t)/2 - \ell = t$. The desired bound follows.

Otherwise, suppose without loss of generality that $a_s \leq (\ell + t)/2$. We can write $F(x_1, \dots, x_s) = G(x_1, \dots, x_{s-1}) + a_s x_s + x_s H(x_1, \dots, x_{s-1})$, where $G, H \in \mathbb{R}[x_1, \dots, x_{s-1}]$ are multilinear polynomials that have nonnegative coefficients and zero constant coefficient. We will show the desired bound conditional on both of the two possible outcomes of β_s .

Let $\vec{\beta}' = (\beta_1, \dots, \beta_{s-1})$. If $\beta_s = 0$, then $|F(\vec{\beta}) - \ell| \leq t$ if and only if $|G(\vec{\beta}') - \ell| \leq t$, and the desired result follows by induction. Otherwise, if $\beta_s = 1$, then $|F(\vec{\beta}) - \ell| \leq t$ if and only if $|(G + H)(\vec{\beta}') - (\ell - a_s)| \leq t$. Note that $\ell - a_s \geq (\ell - t)/2 \geq \ell/3 > 3^{s-1}t$, so the desired result again follows by induction. ■

7 Comparison Between the Slice and a Product Measure

Recall that in order to apply bounds on the polynomial Littlewood–Offord problem, we needed a lemma describing a slice measure in terms of a product measure (Lemma 3.4). In order to apply Theorem 6.1, we also need a comparison between a slice measure and a product measure, but since Theorem 6.1 only applies to polynomials with nonnegative coefficients, we will need a lemma of a rather different flavour, for functions on the slice that only depend on a few coordinates.

For probability distributions μ, ν with the same sigma-algebra of events, recall that the *total variation distance* $d_{\text{TV}}(\mu, \nu)$ is the supremum of $|\mu(A) - \nu(A)|$ over all events A .

Theorem 7.1. Let $k \leq n/2$. Consider any set S and function $F : \{0, 1\}^n \rightarrow S$, such that $F(x_1, \dots, x_n)$ only depends on $x_1, \dots, x_s$. Let $\vec{\sigma} \sim \text{Slice}(n, k)$ and let $\vec{\beta}$ be a vector of n independent **Bernoulli**(k/n) random variables. Then,

$$d_{\text{TV}}(F(\vec{\sigma}), F(\vec{\beta})) \leq \frac{\max(s, 2n/k) - 1}{n - 1}.$$

We will prove Theorem 7.1 using the following theorem of Ehm [6]. Write **Hyp**(n, k, t) for the hypergeometric distribution with k draws from a population of size n with t “featured” elements, and let **Bin**(n, p) be the binomial distribution with n trials and success probability p .

Theorem 7.2 ([6, Theorem 2]). If $(k/n)(1 - k/n)t \geq 1$, then

$$d_{\text{TV}}(\text{Hyp}(n, k, t), \text{Bin}(t, k/n)) \leq \frac{t - 1}{n - 1}.$$

Proof of Theorem 7.1. Let $t = \max(s, 2n/k)$, so $(k/n)(1 - k/n)t \geq 1$ and $F(x_1, \dots, x_n)$ only depends on $x_1, \dots, x_t$ (since $t \geq s$). For $\vec{x} \in \{0, 1\}^n$, let $\|\vec{x}\|_{0,t}$ be the number of $i \in [t]$ for which $x_i = 1$. Let $\vec{\sigma} \sim \text{Slice}(n, k)$, and let $\vec{\beta}$ be a vector of n independent **Bernoulli**(k/n) random variables. Then clearly $\|\vec{\sigma}\|_{0,t} \sim \text{Hyp}(n, k, t)$ and $\|\vec{\beta}\|_{0,t} \sim \text{Bin}(t, k/n)$. For all q in the support of both $\text{Hyp}(n, k, t)$ and $\text{Bin}(t, k/n)$, the conditional distribution of $F(\vec{\sigma})$ given $\|\vec{\sigma}\|_{0,t} = q$ is the same as the conditional distribution of $F(\vec{\beta})$ given $\|\vec{\beta}\|_{0,t} = q$. So,

$$d_{\text{TV}}(F(\vec{\sigma}), F(\vec{\beta})) \leq d_{\text{TV}}(\|\vec{\sigma}\|_{0,t}, \|\vec{\beta}\|_{0,t}) \leq \frac{t-1}{n-1},$$

where the last inequality follows from Theorem 7.2. ■

8 The Vertex Cover Lemma

In this section we prove a lemma showing that every hypergraph G has a small vertex set Y that “only sees large matchings”, in the sense that for any subset $S \subseteq Y$, if we remove all the edges intersecting S , and we remove all the vertices of $Y \setminus S$ from all the remaining edges, the resulting hypergraph either has a large matching or no edges at all.

Definition 8.1. Let G be an r -uniform hypergraph. For vertex subsets $X \subseteq Y$, write $G_Y(X) = \{e \setminus X : e \in G - (Y \setminus X), e \setminus X \neq \emptyset\}$. In other words, delete all edges that intersect $Y \setminus X$, and look at the portion of each edge outside X . This is a mixed-uniformity hypergraph, whose edges have sizes between 1 and r . For $d \in [r]$, let $G_Y^{(d)}(X)$ be the subhypergraph of edges that have size exactly d .

Lemma 8.2. Let $r, m \in \mathbb{N}$ and let G be an r -uniform hypergraph. We can find a vertex set Y of size $O_{r,m}(1)$ such that the following holds. Consider any $X \subseteq Y$ such that $G_Y(X)$ is nonempty, and let d be the maximum integer such that $G_Y^{(d)}(X)$ is nonempty. Then $G_Y^{(d)}(X)$ has a matching of size at least m . Moreover, if $Y \neq \emptyset$, then every edge of G has non-empty intersection with Y .

Proof. It will be more convenient to prove the following statement. Let V be the vertex set of G . For $S \subseteq Y \subseteq V$, let $\Gamma_Y(S) := \{e \setminus S : e \in G, e \cap Y = S\}$ (this is a hypergraph whose edges have size exactly $r - |S|$). Say that S is Y -relevant if $\Gamma_Y(S)$ is nonempty and $\Gamma_Y(S')$ is empty for all $S' \subsetneq S$. We say that a relevant set S is m -bad (or simply, bad), if the maximum matching in $\Gamma_Y(S)$ has size at most $m - 1$. We will show that we can choose a vertex set Y of size $O_{r,m}(1)$ such that there are no Y -relevant sets $S \subseteq Y$ that are bad. To see that this implies the conclusion of the lemma, for $X \subseteq Y$, let d be the maximum integer such that $G_Y^{(d)}(X)$ is nonempty. Since $G_Y^{(d)}(X)$ is nonempty, there exists $S \subseteq X$ of size $r - d$ such that $\Gamma_Y(S)$ is non-empty, and since d is the largest integer satisfying this property, it must be the case that $\Gamma_Y(S') = \emptyset$ for all $S' \subsetneq S$. In other words, S is Y -relevant. Therefore, $\Gamma_Y(S)$ has a matching of size at least m , and since $\Gamma_Y(S) \subseteq G_Y^{(d)}(X)$, our conclusion follows.

We will construct our desired set Y iteratively, by iterating a “greedy cover” map ϕ starting from the empty set. Specifically, fix an ordering of the vertex set, and for $Z \subseteq V$, we define $\phi(Z) \supseteq Z$ as follows:

- If there is some Z -relevant set $S \subseteq Z$ such that $\Gamma_Z(S)$ has no matching of size at least m , then consider such a set S with the smallest size (breaking ties lexicographically, according to our ordering of the vertex set) and a maximum matching M in $\Gamma_Z(S)$ (again, breaking ties lexicographically), let W be the vertex set of M , and set $\phi(Z) = Z \cup W$.
- Otherwise, set $\phi(Z) = Z$.

For $Z \subseteq V$, let $\phi^*(Z)$ be the result of repeatedly applying the map ϕ , starting with Z , until it stabilizes, and let $Y = \phi^*(\emptyset)$. We just need to show that $|Y| = O_{r,m}(1)$; this will be a simple inductive consequence of the following claim.

Claim. For $Z \subseteq V$ and $\ell \in \{0, \dots, r\}$, let $N_\ell(Z)$ be the number of Z -relevant sets $T \subseteq Z$ with $|T| = \ell$. If $\phi(Z) \neq Z$, then there exists $d \in \{0, \dots, r\}$ for which the following hold:

- (1) $N_d(\phi(Z)) < N_d(Z)$,
- (2) $N_f(\phi(Z)) \leq N_f(Z)$ for all $f < d$,
- (3) $|\phi(Z)| < |Z| + rm$, and
- (4) $N_f(\phi(Z)) \leq |\phi(Z)|^f < (|Z| + rm)^f$ for all $f > d$.

Proof of claim. First, parts (3) and (4) are immediate, because $\phi(Z)$ is obtained by adding fewer than rm vertices to Z .

To prove parts (1) and (2), suppose $\phi(Z) \neq Z$ and let $S \subseteq Z$ be the set appearing in the definition of $\phi(Z)$. Note that, by construction, $\Gamma_{\phi(Z)}(S) = \emptyset$. In particular, S is a Z -relevant set that is not $\phi(Z)$ -relevant. Let $S^* \subseteq Z$ be a smallest set that is Z -relevant but not $\phi(Z)$ -relevant and let $d = |S^*|$. This is the value of d for which we will prove parts (1) and (2). In order to do this, it suffices to show the following: for any $\phi(Z)$ -relevant set $T \subseteq \phi(Z)$ with $|T| \leq d$, it must be the case that $T \subseteq Z$ and in fact, that T is Z -relevant. This immediately implies (2), and since S^* is a Z -relevant set that is not $\phi(Z)$ -relevant, we also get (1).

Suppose $T \subseteq \phi(Z)$ is a $\phi(Z)$ -relevant set of size $|T| \leq d$. Suppose for contradiction that either (a) $T \not\subseteq Z$, or (b) $T \subseteq Z$ but T is not Z -relevant. In either case, since $\Gamma_{\phi(Z)}(T)$ is non-empty, it follows that $\Gamma_Z(T \cap Z)$ is non-empty as well, so that there exists $T' \subseteq T \cap Z$ such that T' is Z -relevant. In case (a), $|T \cap Z| < |T| \leq d$ and in case (b) $T' \subsetneq T$, so that in either case, $|T'| \leq d - 1$. Since S^* is the smallest Z -relevant set that is not $\phi(Z)$ -relevant, it must be the case that T' is $\phi(Z)$ -relevant. However, since $T' \subseteq T$, this contradicts that T is $\phi(Z)$ -relevant. ■

Recall that we wish to show that $|Y| = |\phi^*(\emptyset)| = O_{r,m}(1)$. For nonnegative integers $N_0, \dots, N_r, z$, define $F(N_0, \dots, N_r, z)$ to be the maximum of $|\phi^*(Z)|$, over all $Z \subseteq V$ with

$$|Z| = z, \quad N_0(Z) \leq N_0, \quad N_1(Z) \leq N_1, \quad \dots, \quad N_r(Z) \leq N_r,$$

and over all r -uniform hypergraphs G on any vertex set V . (A priori, it is possible that no such maximum exists, in which case we set $F(N_0, \dots, N_r, z) = \infty$.) Since $|Y| = |\phi^*(\emptyset)| \leq F(1, 0, \dots, 0)$, our goal is to show that $F(1, 0, \dots, 0) < \infty$. Since the input parameters to the function F are only r and m (via the definition of $\phi^*(Z)$), in this case it is clear that $F(1, 0, \dots, 0)$ depends only on r and m . In fact, writing $\mathbb{N}_{\geq 0}$ for the nonnegative integers, we will show that for any $(N_0, \dots, N_r, z) \in \mathbb{N}_{\geq 0}^{r+2}$, we have $F(N_0, \dots, N_r, z) < \infty$.

To see this, note that the above claim implies a recurrence for $F(N_0, \dots, N_r, z)$: for all $(N_0, \dots, N_r, z)$, either $F(N_0, \dots, N_r, z) = z$ or

$$F(N_0, \dots, N_r, z) \leq \max_{d \in \{0, \dots, r\}: N_d > 0} F(N_0, \dots, N_{d-1}, N_d - 1, (z + rm)^{d+1}, \dots, (z + rm)^r, z + rm) \quad (8.1)$$

(here we use the convention that the maximum of the empty set is $-\infty$; in other words, $F(0, \dots, 0, z) = z$, which is also easy to see directly). Now, the desired result follows by induction (most easily described in a “transfinite” way): since the lexicographic order on $\mathbb{N}_{\geq 0}^{r+2}$ is a well order, if $F(N_0, \dots, N_r, z) = \infty$ for some $(N_0, \dots, N_r, z)$, then there must exist a lexicographically minimal $(N_0^*, \dots, N_r^*, z^*)$ for which $F(N_0^*, \dots, N_r^*, z^*) = \infty$. But this is impossible: we would have $F(N_0^*, \dots, N_r^*, z^*) \neq z^*$, so (8.1) would contradict lexicographic minimality.

Finally, the “moreover” part is clear by construction.

9 A Variance Bound for Polynomials on the Slice

We need one more technical ingredient for the proof of Theorem 1.1, namely a bound on the variance of a polynomial on the slice.

Proposition 9.1. For any $n \geq k$, let $\lambda \in \mathbb{R}[x_1, \dots, x_n]$ be an n -variable multilinear polynomial with degree at most d whose coefficients all have absolute value at most q . Let $\vec{\sigma} \sim \text{Slice}(n, k)$. Then,

$$\text{Var}[\lambda(\vec{\sigma})] \lesssim_{d,q} n^{2d-1}.$$

Proof. Since $x_1 + \dots + x_n$ is constant on $\vec{x} \in \text{Slice}(n, k)$, it follows that $\text{Var}[\lambda(\vec{\sigma})] = \text{Var}[Q(\vec{\sigma})]$, where $Q(\vec{x}) = \lambda(\vec{x}) + q(x_1 + \dots + x_n)^d$. Since the coefficients of λ have absolute value at most q , it follows that the coefficients of Q are non-negative and have absolute value at most $O_d(q)$. The key point is the following: let $W, T \subseteq [n]$ be disjoint sets with $|W| = i$ and $|T| = j$. Then, writing $a \equiv b$ if a and b have the same sign,

we have

$$\begin{aligned}
\text{Cov}[\vec{\sigma}^W, \vec{\sigma}^T] &= \frac{\binom{n-i-j}{k-i-j}}{\binom{n}{k}} - \frac{\binom{n-i}{k-i}}{\binom{n}{k}} \cdot \frac{\binom{n-j}{k-j}}{\binom{n}{k}} \\
&\equiv \binom{n-i-j}{k-i-j} \cdot \binom{n}{k} - \binom{n-i}{k-i} \cdot \binom{n-j}{k-j} \\
&\equiv \binom{n}{i} \cdot \binom{k-j}{i} - \binom{n-j}{i} \cdot \binom{k}{i} \\
&\leq 0,
\end{aligned}$$

where the last inequality follows since

$$\frac{\binom{n}{i} \cdot \binom{k-j}{i}}{\binom{n-j}{i} \cdot \binom{k}{i}} = \prod_{s=0}^{i-1} \frac{(n-s)(k-j-s)}{(n-j-s)(k-s)} \leq 1$$

for $k \leq n$. Note also that for any $W, T \subseteq [n]$, we have $\text{Cov}[\vec{\sigma}^W, \vec{\sigma}^T] \leq \mathbb{E}[\vec{\sigma}^W \vec{\sigma}^T] \leq 1$.

Recall that the coefficients $\widehat{Q}(S)$ are non-negative of size at most $O_d(q)$. Moreover, $\widehat{Q}(S) = 0$ if $|S| > d$. Putting everything together, we have

$$\begin{aligned}
\text{Var}[Q(\vec{\sigma})] &= \sum_{W, T} \widehat{Q}(W) \widehat{Q}(T) \text{Cov}[\vec{\sigma}^W, \vec{\sigma}^T] \leq \sum_{\substack{W, T \subseteq [n] \\ W \cap T \neq \emptyset}} \widehat{Q}(W) \widehat{Q}(T) \text{Cov}[\vec{\sigma}^W, \vec{\sigma}^T] \\
&\lesssim_d \sum_{|W| \leq d} \sum_{\substack{|T| \leq d \\ T \cap W \neq \emptyset}} q^2 \lesssim_d n^d \cdot n^{d-1} \cdot q^2.
\end{aligned}$$

■

10 Completing the Proof of the Hypergraph Edge-Statistics Conjecture

Now, we combine all the ingredients collected so far to prove Theorem 1.1. Recall that the notation $\alpha \ll \beta_1, \dots, \beta_q$ (respectively, $\alpha \gg \beta_1, \dots, \beta_q$) means “ α is sufficiently small in terms of $\beta_1, \dots, \beta_q$ ” (respectively, “ α is sufficiently large in terms of $\beta_1, \dots, \beta_q$ ”).

Proof Proof of Theorem 1.1. Recall that we are to prove that if $k \gg r, \varepsilon$ and $\ell \notin \{0, \binom{k}{r}\}$, then $\text{ind}_r(k, \ell) \leq 1/e + \varepsilon$. Since $\text{ind}_r(k, \ell) = \text{ind}_r(k, \binom{k}{r} - \ell)$, it suffices to assume that $\ell \leq \lceil \binom{k}{r}/2 \rceil$. Further, it suffices to prove the statement only for (say) $\ell < k^{-1/2} \binom{k}{r}$, since in the complementary regime $k^{-1/2} \binom{k}{r} \leq \ell \leq \lceil \binom{k}{r}/2 \rceil$, a stronger statement follows from Theorem 1.2.

So, consider integers k, ℓ satisfying $0 < \ell < k^{-1/2} \binom{k}{r}$ and fix any $\varepsilon > 0$. The dependence of k on r, ε will be moderated by additional parameters R, m, q, δ , which will play a role later in the argument. Specifically, we first need $R \gg \varepsilon$, then $m \gg R$, then $q \gg m, r$, and then $\delta \ll q$. Finally, $k \gg m, q, r, R, \delta, \varepsilon$. To summarize, the relative sizes of various parameters should be thought of as

$$k \gg 1/\delta \gg q \gg m \gg R \gg 1/\varepsilon.$$

Let $n = Rk$, let G be an r -uniform hypergraph on the vertex set $[n]$, and let U be a random subset of k vertices of G . As $N_r(n, k, \ell)/\binom{n}{k}$ is nonincreasing in n , it suffices to prove that

$$\mathbb{P}[e(G[U]) = \ell] \leq 1/e + \varepsilon.$$

Let Y be the set obtained by applying Lemma 8.2 to G (with our value of m).

Case I: $Y \neq \emptyset$ Consider the random variable $\mathbb{E}[e(G[U]) \mid Y \cap U]$. This can be interpreted as a multilinear polynomial evaluated at $\vec{\sigma} \sim \text{Slice}(n, k)$ (where $\sigma_i = \mathbb{1}_{i \in U}$), that only depends on $|Y| = O_{m,r}(1)$ of its variables. Indeed, we have

$$\begin{aligned} \mathbb{E}[e(G[U]) \mid Y \cap U] &= \sum_{W \in E(G)} \mathbb{P}[W \in E(G[U]) \mid Y \cap U] = \sum_{W \in E(G)} \mathbb{P}[W \setminus Y \subseteq U] \mathbb{1}_{W \cap Y \subseteq U \cap Y} \\ &= \sum_{W \in E(G)} \mathbb{P}[W \setminus Y \subseteq U] \vec{\sigma}^{W \cap Y}, \end{aligned}$$

where $E(G)$ denotes the set of edges of G (cf. the expression for $e(G[U])$ at the start of Section 3). Note that this polynomial has nonnegative coefficients. Also, since we are assuming $Y \neq \emptyset$, every edge of G intersects Y , so the constant coefficient of this polynomial is zero. So, by Theorems 6.1 and 7.1 (with $s = |Y|$ and $\gamma = \varepsilon/2$ and $t = 3^{-|Y|}\ell$), since $R, k \gg \varepsilon$, it follows that except with probability $1/e + \varepsilon/2$,

$$\left| \mathbb{E}[e(G[U]) \mid Y \cap U] - \ell \right| > 3^{-|Y|}\ell \gtrsim_{r,m} \ell. \quad (10.1)$$

Condition on any outcome of $Y \cap U$ such that (10.1) holds. The remaining randomness is comprised of the random set $U \setminus Y$ (which is a uniformly random subset of $[n] \setminus Y$ of size $k - |Y \cap U|$).

Recall from Definition 8.1 that $G_Y(X) = \{e \setminus X : e \in G - (Y \setminus X), e \setminus X \neq \emptyset\}$. If $G_Y(Y \cap U) = \emptyset$, then we are done: given our conditioning, $e(G[U])$ would then take some value with probability 1, and this value cannot be equal to ℓ since we are working with an outcome of $Y \cap U$ for which (10.1) holds. Therefore, we can assume that $G_Y(Y \cap U)$ is non-empty. In this case, given our conditioning, we can write $e(G[U]) = \lambda(\vec{\sigma})$, where $\vec{\sigma} \sim \text{Slice}(n - |Y|, k - |Y \cap U|)$ and λ is a multilinear polynomial of some degree $d \in [r - 1]$. Note that the coefficients of λ all lie in the set $\{0, 1, \dots, q\}$ (here we are using that $q \gg m, r$, so $q \geq \binom{|Y|}{r}$). Also, by the definition of Y , we have $\nu(H_0^d(\lambda)) = \nu(G_Y(Y \cap U)) \geq m$. Our objective is to show that $\mathbb{P}[\lambda(\vec{\sigma}) = \ell] \leq \varepsilon/2$.

Recall that $\delta \ll r, q$ and $m \gg R \gg \varepsilon$ and $k \gg r, q, m$. By Lemma 5.1, at least one of the following holds:

- (1) λ has at least δn^d nonzero degree- d coefficients, or
- (2) we have

$$\sup_{\ell \in \mathbb{R}} \mathbb{P}[\lambda(\vec{\sigma}) = \ell] \lesssim_{d,q} \frac{1}{m^{1/3}} \leq \varepsilon/2.$$

In case (2) we are done. In case (1), we have $\mathbb{E}[\lambda(\vec{\sigma})] \gtrsim \delta k^d \gtrsim_\delta k^d$ while $\text{Var}[\lambda(\vec{\sigma})] \lesssim_{d,q} n^{2d-1} \lesssim_{d,q,R} k^{2d-1}$ by Proposition 9.1. So, by Chebyshev's inequality, except with probability at most $\varepsilon/2$, we have

$$\left| \lambda(\vec{\sigma}) - \mathbb{E}[\lambda(\vec{\sigma})] \right| \lesssim_{d,q,R,\delta,\varepsilon} k^{-1/2} \mathbb{E}[\lambda(\vec{\sigma})].$$

Recalling from (10.1) that $\mathbb{E}[\lambda(\vec{\sigma})]$ is excluded from a range of the form $(1 \pm \Omega_{r,m}(1))\ell$, the desired result follows for $k \gg m, q, r, R, \delta, \varepsilon$.

Case II: $Y = \emptyset$ If G is empty, then we are done, since $\ell \neq 0$ by assumption. Otherwise, we can write $e(G[U]) = \lambda_G(\vec{\sigma})$, where λ_G has degree r and $\vec{\sigma} \sim \text{Slice}(n, k)$. In this case we will be able to prove the stronger bound

$$\mathbb{P}[e(G[U]) = \ell] = \mathbb{P}[\lambda_G(\vec{\sigma}) = \ell] \leq \varepsilon/2 \leq 1/e + \varepsilon.$$

As in Case I above, we have $\nu(H_0^r(\lambda)) \geq m$, and similarly arguing via Lemma 5.1, we only need to consider the case that G has at least δn^r nonzero degree- r coefficients. By the same argument as above, except with probability at most $\varepsilon/2$ we have

$$\left| \lambda_G(\vec{\sigma}) - \mathbb{E}[\lambda_G(\vec{\sigma})] \right| \lesssim_{d,q,R,\delta,\varepsilon} k^{-1/2} \mathbb{E}[\lambda_G(\vec{\sigma})].$$

The desired conclusion now follows for $k \gg q, r, R, \delta, \varepsilon$, since $\mathbb{E}[\lambda_G(\vec{\sigma})] \gtrsim_{R,\delta} k^r$, whereas by assumption $\ell < k^{-1/2} \binom{k}{r} \leq k^{r-1/2}$. ■

Funding

This work was supported by NSF CAREER award DMS-2237646 [to V.J.], ERC Starting Grant “RAND-STRUCT” [no. 101076777 to M.K.], NSF grant DMS-2153576 [to D.M.], and the National Key Research and Development Program of China [2023YFA101020 to T.T.].

Acknowledgments

We would like to thank Lisa Sauermann for her helpful comments. We would also like to thank Alex Grebennikov for identifying an oversight in the application of Theorem 7.1 (in a previous version of this paper).

References

- Alon, N., D. Hefetz, M. Krivelevich, and M. Tyomkyn. “Edge-statistics on large graphs.” *Combin. Probab. Comput.* **29**, no. 2 (2020): 163–89. <https://doi.org/10.1017/S0963548319000294>.
- Balogh, J., H. Ping, B. Lidický, and F. Pfender. “Maximum density of induced 5-cycle is achieved by an iterated blow-up of 5-cycle.” *Eur. J. Combin.* **52** (2016): 47–58. <https://doi.org/10.1016/j.ejc.2015.08.006>.
- Bollobás, B. and A. Scott. “Intersections of hypergraphs.” *J. Combin. Theory Ser. B* **110** (2015): 180–208. <https://doi.org/10.1016/j.jctb.2014.08.002>.
- Costello, K. P. “Bilinear and quadratic variants on the Littlewood-Offord problem.” *Israel J. Math.* **194**, no. 1 (2013): 359–94. <https://doi.org/10.1007/s11856-012-0082-4>.
- Costello, K. P., T. Tao, and V. Van. “Random symmetric matrices are almost surely nonsingular.” *Duke Math. J.* **135**, no. 2 (2006): 395–413. <https://doi.org/10.1215/S0012-7094-06-13527-5>.
- Ehm, W. “Binomial approximation to the Poisson binomial distribution.” *Statist. Probab. Lett.* **11**, no. 1 (1991): 7–16. [https://doi.org/10.1016/0167-7152\(91\)90170-V](https://doi.org/10.1016/0167-7152(91)90170-V).
- Erdős, P. “On a lemma of Littlewood and Offord.” *Bull. Amer. Math. Soc.* **51** (1945): 898–902. <https://doi.org/10.1090/S0002-9904-1945-08454-7>.
- Fox, J., M. Kwan, and L. Sauermann. “Combinatorial anti-concentration inequalities, with applications.” *Math. Proc. Cambridge Philos. Soc.* **171**, no. 2 (2021): 227–48. <https://doi.org/10.1017/S0305004120000183>.
- Fox, J. and L. Sauermann. “A completion of the proof of the edge-statistics conjecture.” *Adv. Comb.* (2020): Paper No. 4, 52.
- Fox, J., L. Sauermann, and F. Wei. “On the inducibility problem for random Cayley graphs of abelian groups with a few deleted vertices.” *Random Struct. Algorithms* **59**, no. 4 (2021): 554–615.
- Hirst, J. “The inducibility of graphs on four vertices.” *J. Graph Theory* **75**, no. 3 (2014): 231–43.
- Kane, D. M. “The correct exponent for the Gotsman-Linial conjecture.” *Comput. Complexity* **23**, no. 2 (2014): 151–75.
- Kwan, M. and L. Sauermann. “Resolution of the quadratic Littlewood-Offord problem.” arXiv preprint [2312.13826](https://arxiv.org/abs/2312.13826).
- Kwan, M., B. Sudakov, and T. Tran. “Anticoncentration for subgraph statistics.” *J. Lond. Math. Soc. (2)* **99**, no. 3 (2019): 757–77.
- Littlewood, J. E. and A. C. Offord. “On the number of real roots of a random algebraic equation. III.” *Rec. Math. [Mat. Sbornik] N.S.* **12**, no. 54 (1943): 277–86.
- Liu, X., D. Mubayi, and C. Reiher. “The feasible region of induced graphs.” *J. Combin. Theory Ser. B* **158**, no. part 2 (2023): 105–35.
- Martinsson, A., F. Mousset, A. Noever, and M. Trujić. “The edge-statistics conjecture for $\ell \ll k^{6/5}$.” *Israel J. Math.* **234**, no. 2 (2019): 677–90.
- Meka, R., O. Nguyen, and V. Van. “Anti-concentration for polynomials of independent random variables.” *Theory Comput.* **12** (2016): Paper No. 11, 16.
- Pippenger, N. and M. C. Golumbic. “The inducibility of graphs.” *J. Combin. Theory Ser. B* **19**, no. 3 (1975): 189–203.
- Razborov, A. and E. Viola. “Real advantage.” *ACM Trans. Comput. Theory* **5**, no. 4 (2013): Art. 17, 8.
- Rosiński, J. and G. Samorodnitsky. “Symmetrization and concentration inequalities for multilinear forms with applications to zero-one laws for Lévy chaos.” *Ann. Probab.* **24**, no. 1 (1996): 422–37.
- Ueltzen, R. “Characterizing graphs with high inducibility.” arXiv preprint [2411.17362](https://arxiv.org/abs/2411.17362).