A TRANSFERENCE PRINCIPLE AND A COUNTING LEMMA FOR SPARSE HYPERGRAPHS

PETER ALLEN, JULIA BÖTTCHER, JOANNA LADA, AND DOMENICO MERGONI CECCHELLI

Abstract. The last two decades have witnessed a growing trend towards proving sparse random analogues of combinatorial theorems. One unified approach to proving such theorems, formalised by Conlon and Gowers, involves establishing a 'transference principle' which allows one to translate between robust properties in the dense setting to the sparse random setting, provided the density is not too small. Our results extend the previous literature by generalising a transference theorem of Conlon and Gowers to counting (hyper)graphs which are not necessarily strictly balanced; we obtain an asymptotically optimal probability of success and use our result to prove a counting lemma for hypergraphs in the sparse regime.

1. Introduction

Extremal combinatorics is the branch of discrete mathematics dealing with finding conditions that force the existence of specified substructures. A rich area of study focuses on finding density conditions that guarantee that a subset of a complete set contains a specific structure.

Our goal in this work is to formalise and extend the Transference Principle, which is a method that can be traced back to the seminal paper of Green and Tao [10] and which was then further developed by Conlon and Gowers [4].

Intuitively, the transference principle is a method that allows one to translate "robust" counting results that are known in the dense regime to a random sparse regime. Let us see an example.

For graphs H and G, let us denote by $c(H,G)$ the number of copies of H in G. Moreover, let $m_2(H) = \max_{H' \subseteq H, |H'| \geq 3} \frac{e(H')-1}{v(H')-2}$ be the two-density of H. For any graph H, Erdős-Stone-Simonovits' Theorem [8, 7] guarantees that for any $\varepsilon > 0$ there is N large enough such that any subgraph F of K_N with at least $\big(1-\frac{1}{\chi(H)-1}+o(1)\big)\binom{N}{2}$ edges contains a copy of H. This result is "robust", by which we mean that any subgraph of K_N with at least $\big(1-\frac{1}{\chi(H)-1}+\varepsilon\big)\binom{N}{2}$ contains $\Omega(N^{v(H)})$ copies of H, as proved by Erdős and Simonovits [9] (see also [13] for a survey on the topic).

Now that we have an example of a robust counting result that is known in the dense regime, let us see how we can translate it to a sparse random regime. In this variation, we are interested in finding copies of H in subgraphs of $G_{N,p}$, the random graph over N vertices where each edge is selected independently with probability $p = p(N)$.

What the transference principle allows us to do is to reduce a counting in the sparse random regime to a counting in the dense regime, which we know how to do. It enables us to count the copies of H in a subgraph G of $\Gamma = G_{N,p}$ having sufficient relative density, by counting the copies of H in a dense model G' of G, which looks 'similar' to G under the appropriate point of view.

The formal translation from the sparse random regime to the dense regime (which is a special case of our general transference principle) is as follows.

Theorem 1. *Let H be a fixed graph and $\varepsilon > 0$. Then there exists a constant $C > 0$ such that the following holds. Suppose that $p > CN^{-1/m_2(H)}$, and let η_N be the probability that the number of copies of H in $\Gamma = G_{N,p}$ exceeds $(1 + \frac{\varepsilon}{2})p^{e(H)}N^{v(H)}$. Then with probability at least $1 - \eta_N$, for every subgraph $G \subseteq \Gamma$ there exists a graph G' on $V(G)$ that satisfies:*

$$e(G)p^{-1} = e(G') \pm \varepsilon N^2 \quad \text{and} \quad c(H, G)p^{-e(H)} = c(H, G') \pm \varepsilon N^{v(H)}.$$

We call such a G' a *good* model of G. This result allows us to count copies of the fixed graph H in G in the following way. Given $p > CN^{-1/m_2(H)}$ and G a subset of $\Gamma = G_{N,p}$ with at least $\left(1 - \frac{1}{\chi(H)-1} + \varepsilon\right)p\binom{N}{2}$ edges. By Theorem 1, applied with the right parameter $\varepsilon = \varepsilon'$, we can find a good model G' of G. Which means we can find a subgraph G' of K_N with at least $\left(1 - \frac{1}{\chi(H)-1} + \varepsilon'\right)\binom{N}{2}$ edges, and such that the number of copies of H in G is (up to rescaling) the number of copies of H in G'. In particular, Erdős and Simonovits' result [9] gives us that G' contains $\Omega(N^{v(H)})$ copies of H, which guarantees that G contains $\Omega(N^{v(H)}p_N^{e(H)})$ copies of H, which is a positive proportion of the expected number of copies of H in $G_{N,p}$.

The reader will note that the threshold

We also have that our "success probability" $1 - \eta_N$ is optimal, but we postpone to the proof of the theorem to see the details.

2. A General Transference Principle and its applications

We mentioned that our interest is to generalise and extend the transference principle. We start by seeing how we can translate Theorem 1 in a more abstract setting. We are then going to state and prove an extended version of this translation.

The main idea here is that we can formulate Theorem 1 as a statement about the set of edges of K_N. That is, we can consider $n = \binom{N}{2}$, and arbitrarily define a bijection between $[n]$ and the set of edges of K_N. Once we have done that, we can define the hypergraph S of all copies of H in $[n]$. By construction, S is a subset of $\binom{n}{k}$ (we have $k = 4$ as H has 4 edges), and has size of the order of $N^{v(H)}$. Counting copies of H in G is the same as counting elements of S contained in $[n]_{p_N}$.

Notice a similar procedure can also be done for counting copies of an r-uniform hypergraph (we would just need to consider $n = \binom{N}{r}$).

Given a set $[n]$, a uniform hypergraph S, and a subset Y of the random set $[n]_{p_N}$ (for appropriate values of p_N), our general transference principle allows us to find a dense model Z of Y such that the number of elements of S in Y is (up to scaling) close to the number of elements of S in Z.

Actually, our transference principle allows for a further layer of generality, for which we need additional notation. We now introduce the necessary notation and state the general version of our transference principle.

Given positive integers $n, k \geq 2$, a *k-uniform ordered hypergraph* S of size n is a k-uniform hypergraph on $[n]$ with an order associated to each of its edges. That is, each edge of a k-uniform ordered hypergraph is an (ordered) sequence of length k of elements of $[n]$. Given $x \in [n]$ and $i \in [k]$, we write $S_i(x)$ for the subset of S consisting of all edges whose i-th entry is x. Given such an hypergraph S, and a sequence $\mathbf{x}$ of length k of elements of $[n] \cup \{*\}$, we write $\deg_S(\mathbf{x})$ for the number of edges of S which agree with $\mathbf{x}$ at all positions which do not equal $*$. That is, those entries equalling $*$ are allowed to vary, while the others are fixed to the value they have in $\mathbf{x}$. For ℓ a positive integer, we write $\Delta_\ell(S)$ for the maximum value of $\deg_S(\mathbf{x})$ over all sequences $\mathbf{x}$ with exactly ℓ entries not equal to $*$. This is the standard codegree in the ordered hypergraph setting, where ℓ vertices are fixed and the number of edges containing them is counted.

We call a function $\sigma : [n] \to [0,1]$ over the set of vertices a *similarity function*. We call a function $\omega : S \to [0,1]$ over the set of edges a *subcount*. We abuse notation by denoting with $\mathbf{1}$ any function that takes value 1 on its domain (whatever that might be). We write $\mathbb{1}$ to denote the indicator function of a proposition, which is, we write for example $\mathbb{1}(y \in Y)$ to be the function that has value 1 when '$y \in Y$' is true, and value 0 when it is false (the domain is always clear from the context). For any real numbers x, y, z, we also write $x = y \pm z$ to indicate $y - z \le x \le y + z$.

Very importantly, we now introduce a general setting that accompanies us for the rest of this work. That is, we fix now the following quantities, and refer back to them frequently in the following.

Setting 2. *Let $k, n \ge 2$ be fixed integers, let $c, p > 0$ be real numbers with $p \in (0,1)$. Let S be a k-uniform ordered hypergraph on $[n]$, and let Σ and Ω be sets of respectively similarity functions on $[n]$ and subcounts of S. Let both Σ and Ω contain the $\mathbf{1}$ function that takes value 1 everywhere in their respective domains, and let Σ contain each of the n functions $f(x) \equiv \mathbb{1}(x = i)$.*

We point out that this setting contains no conditions on any of these objects, which is why we need the following definition.

Definition (C-conditions)**.** *Let us be in Setting 2. For $C \ge 0$ a real number, we say that the C-conditions are satisfied if all the following inequalities are respected. We first ask $p \ge C(\log^{2k} n)n^{-1}$, and that for all $1 \le \ell \le k$, we have*

$$\Delta_\ell(S) \le cC^{1-\ell}p^{\ell-1}\tfrac{e(S)}{n} \, .$$

Where $e(S)$ is the number of edges of S. We also ask that Σ and Ω have at most $\exp\left(\frac{pn}{C}\right)$ elements.

This setting and definition allow us to set up statements as follows. "Let us be in Setting 2. For every $\varepsilon > 0$ there is $C > 0$ such that, if the C-conditions are satisfied, then …".

We need one more set of definitions.

Definition. *Let us be in Setting 2. We say that $Z \subseteq [n]$ is an ε-good dense model for $Y \subseteq [n]$ if it satisfies the following:*

(1) For each $\sigma \in \Sigma$, we have $\displaystyle\sum_{y\in[n]} p^{-1}\mathbb{1}(y \in Y)\sigma(y) = \sum_{z\in[n]} \mathbb{1}(z \in Z)\sigma(z) \pm \varepsilon n$, and

(2) For each $\omega \in \Omega$, we have $\displaystyle\sum_{s\in S} p^{-k}\mathbb{1}(s \subseteq Y)\omega(s) = \sum_{s\in S} \mathbb{1}(s \subseteq Z)\omega(s) \pm \varepsilon e(S)$.

Notice that whether Z is an ε-good dense model of Y depends on Ω and Σ even if this is not explicit from the notation. We say that Z is an ε-good dense lower model if the second equality of the definition is just a lower bound, i.e. if it satifies 1 and

$$\forall \omega \in \Omega, \ \sum_{s\in S} p^{-k}\mathbb{1}(s \subseteq Y)\omega(s) \ge \sum_{s\in S} \mathbb{1}(s \subseteq Z)\omega(s) - \varepsilon e(S) \, .$$

We are now ready to introduce our general transference principle.

Theorem 3. *Let us be in Setting 2. For every $\varepsilon > 0$ there exists a constant $C > 0$ such that, if the C-conditions are satisfied, the following holds.*

(1) ***Lower bound:*** *With probability at least* $1-\exp\left(-\frac{pn}{C}\right)$*, every subset* Y *of the binomial random set* $X=[n]_p$ *has an* ε*-good dense lower model* $Z\subseteq[n]$*.*

(2) ***Upper bound:*** *Let*

$$\eta_n := \mathbb{P}\left(|\{s\in S: s\subseteq [n]_p\}| \geq (1+\tfrac{\varepsilon}{2})\cdot \mathbb{E}(|\{s\in S: s\subseteq [n]_p\}|)\right) + \exp(-\tfrac{pn}{C})\,.$$

With probability at least $1-\eta_n$*, every subset* Y *of the binomial random set* $X=[n]_p$ *has an* ε*-good dense model* $Z\subseteq[n]$*.*

(3) ***Dense model with deletion:*** *With probability at least* $1-\exp\left(-\frac{pn}{C}\right)$*, there exists a subset* $\tilde{X}$ *with at least* $(1-\varepsilon)pn$ *elements of the binomial random set* $X=[n]_p$ *such that for every subset* $Y\subseteq\tilde{X}$*, there is an* ε*-good dense model* $Z\subseteq[n]$ *for* Y*.*

Notice that the probabilities mentioned above are asymptotically optimal. Indeed, the failure probability has the same order of magnitude of the probability that X contains no element of S at all for cases 1 and 3. Moreover, for case 2, η_n corresponds to the probability that X contains many more elements of S than expected, plus an error term of the order of magnitude of the probability that $[n]_p$ contains no element of S at all. In this case, taking $Y=X$ would show that we cannot ask for the existence of a good dense model for all subsets of X.

2.1. **A further note about graphs.** We now see that Theorem 1 follows from Theorem 3. Indeed, if we take Σ, and Ω to be minimal (as required by Setting 2), we obtain a counting result that is exactly Theorem 1. This is because for a subset Y of $[n]_p$, Theorem 3 gives us an ε-good dense model $Z\subseteq[n]$ such that $|Y|p^{-1}=|Z|\pm\varepsilon n$ and $\sum_{s\in S}\mathbb{1}(s\subseteq Y)p^{-k}=\sum_{s\in S}\mathbb{1}(s\subseteq Z)\pm\varepsilon n$. The first equality says that Z has the appropriate size, and the second allows us to know the number of copies of H in Y provided we can count the copies of H in Z.

2.2. **Counting lemma for sparse hypergraphs.** We provide a further application of our transference principle, which is a counting lemma for sparse hypergraphs. However, because much more notation is needed to state such a theorem, we postpone its statement to Section 12, where it is presented as Theorem 30. Because Section 12 is completely separated from the preceding sections, besides for the use of our transference principle, the interested reader can explore Section 12 independently from the rest of this work.

Theorem 30 is a strong counting result for hypergraphs in the sparse random regime. Indeed, it provides a more precise counting statement than the one obtained by Balogh, Morris, and Samotij [1], and by Saxton, and Thomason [16] with the container method. Similarly, with their version of the transference principle, Conlon, Gowers, Samotij, and Schacht [5] also obtained weaker lower bounds, and were able to obtain an upper bound only in the case of strictly-balanced graphs (while their work can probably be generalised to hypergraphs, no such generalisation has been completed).

2.3. **The Deletion Version of our Transference Principle.** Before Section 11, we focus on item 3 of Theorem 3, which is the deletion version of our Counting Lemma. In Section 11 we show how to obtain the rest of Theorem 3 from item 3. We restate now item 3 of Theorem 3 as an independent theorem and make explicit the notation.

Definition (ε-deletion)**.** *Let* X *be a sample of the binomial random set* $[n]_p$*. Given* $\varepsilon>0$*, we say that* $\tilde{X}$ *is an* ε*-deletion of* X *if* $\tilde{X}$ *is a subset of* X *with at least* $(1-\varepsilon)pn$ *elements.*

Theorem 4 (Case 3 of Theorem 3). *Let us be in Setting 2. For every $\varepsilon > 0$ there exists $C > 0$ such that, if the C-conditions are satisfied, then with probability at least $1 - \exp\big(-\frac{pn}{C}\big)$, the binomial random set $X = [n]_p$ admits an ε-deletion $\tilde{X}$ such that for each $Y \subseteq \tilde{X}$, there is an ε-good dense model $Z \subseteq [n]$ for Y.*

3. Tools

3.1. **Concentration Inequalities.** We start with some standard concentration inequalities. Theorem 5, Lemma 16 and Theorem 7 can be found in [18], respectively in Section 2.3, Section 2.8, and Section 2.9.

Theorem 5 (Chernoff's inequality). *Let $X_1, \dots, X_n$ be independent Bernoulli random variables, let $Y = \sum_{i=1}^n X_i$, and let $\delta \in (0,1)$. Then we have*

$$\mathbb{P}[Y \geq (1+\delta)\mathbb{E}[Y]], \mathbb{P}[Y \leq (1-\delta)\mathbb{E}[Y]] \leq \exp\big(-\tfrac{\delta^2}{3}\mathbb{E}[Y]\big).$$

The following result is known as Bernstein's inequality.

Lemma 6 (Bernstein's inequality). *Let $Y_1, \dots, Y_n$ be independent random variables taking values in $[-M, M]$; let $S = Y_1 + \ldots + Y_n$. For $\lambda \geq 0$ we have*

$$\mathbb{P}\big[\big|S - \mathbb{E}[S]\big| \geq \lambda\big] \leq 2\exp\left(\frac{-\lambda^2/2}{\frac{M\lambda}{3} + \sum_i \mathrm{Var}(Y_i)}\right).$$

The following result is due to McDiarmid.

Theorem 7 (McDiarmid's inequality). *Let $X_1, \dots, X_n$ be independent real-valued random variables, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. Assume that the value of $f(x)$ can change by at most $c_i > 0$ under an arbitrary change[1] of the i-th coordinate of $x \in \mathbb{R}^n$. Then, for every $\varepsilon > 0$ we have*

$$\mathbb{P}\left[\big|f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)]\big| \geq \varepsilon\right] \leq 2\exp\Big(-\tfrac{2\varepsilon^2}{\sum_{j=1}^n c_j^2}\Big).$$

We also require a further concentration inequality, due to Kim and Vu [12] which provides a concentration result for a multi-variable polynomial over independent Bernoulli random variables as follows. Let F be an hypergraph with $V(F) = \{1, \dots, n\}$ and edge set $E(F)$. Let us assume each edge e is associated to a weight $w(e) > 0$ and that each edge of F contains at most d vertices. Moreover, for any $A \subseteq V(F)$, let F_A denote the A-truncated sub-hypergraph of F, which is the hypergraph with vertex set $V(F) \setminus A$ and edge set $E(F_A) = \{e' \subseteq V(F_A) : e' \cup A \in E(F)\}$. Note that w extends in a unique way from $E(F)$ to $E(F_A)$, therefore we abuse notation and use w to denote either function.

Suppose now $t_1, \dots, t_n$ are independent random variables, such that for each $i \in [n]$ there is $p_i \in [0,1]$ such that t_i is either a Bernoulli $\{0,1\}$ random variable with $\mathbb{E}(t_i) = p_i$, or the constant random variable $t_i \equiv p_i$. The following polynomial is a well-defined random variable

$$Y_F = \sum_{e \in E(F)} w(e) \prod_{t_i \in e} t_i\,.$$

Analogously we can define Y_{F_A}, where by convention $\prod_{t_i \in \emptyset} t_i = 1$.

In order to provide a concentration statement for Y_F, we need to introduce a language to describe its deviations. For $i \in \{0, \dots, d\}$ let $\mathbb{E}_i(Y_F) = \max_{A \subseteq V(F) : |A| = i} \mathbb{E}(Y_{F_A})$. Note

[1] This means that for any index i and any $x_1, \dots, x_n, x_i'$ we have $|f(\dots, x_i, \dots) - f(\dots, x_i', \dots)| \leq c_i$.

$\mathbb{E}_0(Y_F) = \mathbb{E}(Y_F)$ is just the expectation of Y_F. Let $\mathbb{E}'(Y_F) = \max_i \mathbb{E}_i(Y_F)$ and $\mathbb{E}''(Y_F) = \max_{i\geq 1} \mathbb{E}_i(Y_F)$.

Theorem 8 (Kim-Vu's inequality)**.** *Let F, w, d, and $\{t_1, \dots, t_n\}$ be as above. For $\lambda > 1$ and $a_d := 8^d d!^{1/2}$, we have*

$$P\big[|Y_F - \mathbb{E}(Y_F)| > a_d(\mathbb{E}'(Y_F)\mathbb{E}''(Y_F))^{1/2}\lambda^d\big] = O(\exp(-\lambda + (d-1)\log n))\,.$$

The moral of this theorem is that if the average effect of any group of at most d random variables is considerable smaller than the expectation of Y_F, then Y_F is strongly concentrated.

3.2. **Optimisation tools.** Here and in the following, by *polytope* we mean a convex polytope, i.e. the convex hull of a finite set of points in a finite-dimensional Euclidean space. Given a polytope Φ, the vertex set of Φ is the[2] minimal set V of points whose convex hull equals Φ. The reader should not confuse the vertex set of a polytope with the vertex set of a graph or hypergraph.

Lemma 9. *Consider $f : \mathbb{R}^n \to \mathbb{R}$ a polynomial in n variables that can be written in the form $f(x_1, \dots, x_n) = \sum_{i=1}^n a_i x_i^d$, where d is either 1 or any positive even integer, and where $a_1, \dots, a_n \geq 0$. Let Φ be a polytope in $\mathbb{R}^n$ with vertex set V. Then f attains its maximum over Φ at a vertex of Φ, which is:*

$$\max_{x\in\Phi} f(x) = \max_{v\in V} f(v)\,.$$

Proof. We show that if a maximiser is in the interior of a line segment in Φ, then all points on the line segment are also maximisers.

For distinct $(Y_{\mathbf{1}}, \dots, y_n)$ and $(z_1, \dots, z_n)$ in Φ, let us denote by $(x_1, \dots, x_n)$ their middle point $\frac{1}{2}(Y_{\mathbf{1}} + z_1, \dots, y_n + z_n)$. If $\sum_{i=1}^n a_i x_i^d$ is at least $\sum_{i=1}^n a_i y_i^d$ and strictly larger than $\sum_{i=1}^n a_i z_i^d$, then it is also larger than $\sum_{i=1}^n \frac{1}{2} a_i(y_i^d + z_i^d)$. By averaging, there exists i such that $a_i x_i^d > \frac{1}{2} a_i(y_i^d + z_i^d)$, so $x_i^d > \frac{1}{2}(y_i^d + z_i^d)$. But the function $x \to x^d$ is convex, a contradiction.

If x is in the interior of a face of Φ of some dimension D, by picking a line through x in this face, we see that there is a maximiser in a boundary face of dimension $D-1$, and iterating we reach a vertex which is a maximiser. ☐

A *functional* is a function that has $\mathbb{R}$ as codomain. Given a functional $h : X \to \mathbb{R}_{\geq 0}$, we say a functional $f : X \to \mathbb{R}_{\geq 0}$ is *h-bounded* if $0 \leq f(x) \leq h(x)$ for all $x \in X$. More generally, given a collection H of functionals from X to $\mathbb{R}_{\geq 0}$, we say that a functional $f : X \to \mathbb{R}_{\geq 0}$ is *H-bounded* if there exists a functional $h \in H$ such that f is h-bounded. Suppose that f is H-bounded. We say that f is *H-extreme* if there is $h \in H$ such that for every $x \in X$ we have either $f(x) = 0$ or $f(x) = h(x)$.

A celebrated theorem about the existence of functionals is the Hahn-Banach theorem.

Theorem 10 (Hahn-Banach)**.** *Let K be a closed convex set in $\mathbb{R}^n$ and let f be a vector that does not belong to K. Then there is a linear functional ψ on $\mathbb{R}^n$ such that $\psi(f) > 1$ and such that $\psi(g) \leq 1$ for every $g \in K$.*

Another celebrated result is the Stone-Weierstrass Theorem, which we present in its original form, proved by Weierstrass. We refer to Theorem 7.26 of [15].

Theorem 11 (Weierstrass Approximation)**.** *If f is a continuous real function on $[a, b]$. For every $\varepsilon > 0$ there exists a polynomial P with real coefficients such that for every $x \in [a, b]$ we have $|P(x) - f(x)| \leq \varepsilon$.*

[2] A proof of uniqueness follows by greedy selection.

4. Main technical theorem

The focus of this section is to rewrite our setting in the language of *functionals*, and create a parallelism between sets, functionals, and vectors. This is done following the example of Green and Tao [10], and Conlon and Gowers [4] after them.

4.1. Sets, functionals, and vectors.

It is fundamental for understanding the rest of this work the idea that we can represent subsets of a specific set as functionals, and functionals as vectors, and that tools used in one of these scenarios often have a useful translation in one of the others.

We start by introducing the equivalence between subsets of $[n]$ and functionals from $[n]$ to $\mathbb{R}_{\geq 0}$. The statement of Theorem 4 is about random subsets of a given set $[n]$. Given a sample $X = [n]_p$ we write $\mu = \mu(X)$ for the scaled indicator function $x \to p^{-1}\mathbb{1}(x \in X)$. This functional μ is our representation of X in the space of functionals $[n] \to \mathbb{R}$. Strictly speaking, we should not say this, since $\mu(X)$ depends not just on X but on the value of p used when X was chosen, but this is always clear from context. In this language, we think of a weighted subset of X as being a functional $f : [n] \to \mathbb{R}$ satisfying $0 \leq f(x) \leq \mu(x)$ for all $x \in [n]$. Also, the unweighted subset $Y \subseteq X$ corresponds to the scaled indicator function $p^{-1}\mathbb{1}(x \in Y)$ which takes value p^{-1} on Y and 0 elsewhere.

Often we also want to think of a functional $f : [n] \to \mathbb{R}$ as a vector of $\mathbb{R}^n$, in order to define more easily operations and norms over the space of such functionals. While quite standard, we give an explicit example of how this allows us to define an inner product on the set of said functionals by

$$\langle f, g \rangle := \tfrac{1}{n} \sum_{x \in [n]} f(x)g(x)\,. \tag{1}$$

It is quite important in the following that this operation is indeed an inner-product, and therefore is, in particular, linear in each component and symmetric. We often apply real operations and operators to vectors, by which we always intend to apply them pointwise. For example, the product of two vectors, written fg, is the vector for which the component x has value the pointwise product $f(x)g(x)$. We also define, maybe in a less standard way, the operator $\cdot^{+}$, which we apply to functionals and vectors alike, as follows. For f a functional,

$$f^{+}(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}.$$

Also, for a positive integer d, the notation f^d indicates the pointwise product, so this indicates the functional $f^d(x)$, or equivalently the vector with d-powers at every component.

4.2. Further definitions towards our goal.

The objective of this section is to define a norm over the set of functionals $[n] \to \mathbb{R}$. We want to use this norm to rewrite Theorem 4 in the language of functionals. More precisely, we want to define a norm such that, if the functional f representing Y and the functional g representing Z are close with respect to this norm, then Z is a good dense model for Y. We now proceed with introducing the necessary definitions, before moving to rewriting Theorem 4.

A fundamental operation for our work is the *convolution*, which is an operation on functionals that is dependent of S.

Definition (Convolution)**.** *Given our k-uniform ordered hypergraph S on $[n]$, let $i \in [k]$ be an index, let $f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k$ be non-negative functionals from $[n]$ to $\mathbb{R}_{\geq 0}$, and let $\omega : S \to [0,1]$ be a subcount. The* convolution *$*_{i,S,\omega}(f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k)$ is the functional $[n] \to \mathbb{R}$ defined as follows. For $x \in [n]$,*

$$*_{i,S,\omega}(f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k)(x) := \frac{n}{e(S)} \sum_{s \in S_i(x)} \omega(s) \prod_{j \neq i} f_j(s_j)\,.$$

In the following, we only write $*_{i,\omega}(f_1, \dots, f_k)$, as S is fixed. Moreover, we would write "let $f_1, \dots, f_k$ be non-negative functionals" instead of "let $f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k$" when we only need $k-1$ functionals, for ease of indexing.

Note that the convolution operator is multilinear, i.e. it is linear in each of the $f_1, \dots, f_k$.

The reason we need such a definition is the following. Consider the expression

$$\langle f_i, *_{i,\omega}(f_1, \dots, f_k)\rangle = \frac{1}{|S|} \sum_{s \in S} \omega(s) \prod_{j=1}^{k} f_j(s_j)\,. \tag{2}$$

If for each $j \in [k]$ we select f_j to be the scaled indicator function of the sparse subset Y, i.e. $f_j(x) = p^{-1}\mathbb{1}(x \in Y)$, then we obtain that equation (2) becomes the left-hand side of point 2 of Theorem 4 (without the error term). On the other hand, if we select as g_j the indicator function of the dense model Z, i.e. $g_j(x) = \mathbb{1}(x \in Z)$, then equation (2) becomes the right-hand side of point 2 of Theorem 4 (without the error term).

In order to say that the right and left side of point 2 of Theorem 4 are close to each others we bound a telescoping sum. Which is, we prove that the quantity

$$\langle f_i - g_i, *_{i,\omega}(f_1, \dots, f_{i-1}, g_{i+1}, \dots, g_k)\rangle \tag{3}$$

is small whenever g_i is a dense model of f_i. The reader should see this as a further example of taking advantage of the parallelism between the sets, functionals, and vectors formalisms.

We now go one step forward, and define a polytope Φ. The goal of this polytope is contain some *witness functionals* so that if $\langle f - g, \phi\rangle$ is small for all ϕ in the polytope, then the various (3) are also small.

Following Conlon and Gowers [4], we give a simplified version of their definition as follows.

Definition. *Let us be in Setting 2, and let H be a set of functionals $[n] \to \mathbb{R}$.*

*A functional ϕ is said to be H-*anti-uniform *if it is in Σ, or if it can be written in the form $*_{i,\omega}(f_1, \dots, f_k)$ for some H-bounded functionals $f_1, \dots, f_k$, some $i \in [k]$, and some $\omega \in \Omega$. The polytope $\Phi(H, \Sigma, \Omega)$ of H-anti-uniform functionals is the polytope in the space of functionals $\mathbb{R}^{[n]}$ defined by convex hull of the set containing all the H-anti-uniform functions ϕ and their inverses $-\phi$.*

Because we use this definition only under Setting 2, we usually write $\Phi(H)$ instead of $\Phi(H, \Sigma, \Omega)$. Also, when we enumerate H we write $\Phi(h_1, \dots, h_s)$ instead of $\Phi(\{h_1, \dots, h_s\})$. E.g. we often write $\Phi(\tilde{\mu})$ or $\Phi(\mathbf{1})$ for $\Phi(\{\tilde{\mu}\})$ and $\Phi(\{\mathbf{1}\})$.

Because the convolution operator is multilinear, every vertex of $\Phi(H)$ is either in Σ, or it is a convolution $*_{i,\omega}(f_1, \dots, f_k)$ where each f_i is H-extreme. Moreover, if we have $f_1, \dots, f_k$ non-negative functionals, respectively bounded by H-elements $h_1, \dots, h_k$, then we also have that for all $x \in [n]$ it holds

$$0 \leq *_{i,\omega}(f_1, \dots, f_k)(x) \leq *_{i,\mathbf{1}}(h_1, \dots, h_k)(x)\,.$$

This justifies the following definition.

Definition. *For $h_1, \ldots, h_k$ in H, we say that an element of $\Phi(H)$ that can be written as $*_{i,\mathbf{1}}(h_1, \ldots, h_k)$ is an* H-largest anti-uniform functional of $\Phi(H)$ *(or a largest anti-uniform functional of $\Phi(H)$).*

Moreover, we call H-extreme anti-uniform functional *a functional that is in Σ or of the form $*_{i,\omega}(h_1, \ldots, h_k)$, where $h_1, \ldots, h_k$ are H-extreme.*

Remark 12. We consider a few properties of H-anti-uniform functionals and of $\Phi(H)$.

- $\Phi(H)$ is by definition centrally symmetric.
- Because Σ contains by definition all the standard basis vectors, which is all the functions of the form $f(x) \equiv \mathbb{1}(x = i)$, we have that $\Phi(H)$ is a full-dimensional polytope in $\mathbb{R}^{[n]}$.
- All H-largest anti-uniform functionals are H-extreme anti-uniform functionals, but not vice-versa.
- For any set H, and for any $v \in \Phi(H)$ a vertex of the polytope, there exist a H-largest anti-uniform functional $*_{i,\mathbf{1}}(h_1, \ldots, h_k)$ such that we have $0 \leq v \leq *_{i,\mathbf{1}}(h_1, \ldots, h_k)$ pointwise. This follows, as mentioned above, from multilinearity of the convolution operator.

As mentioned, the objective of this section is to rewrite in the language of functionals and vectors the statement of Theorem 4. The last technical step needed is the definition of a norm over $\mathbb{R}^{[n]}$. Our candidate is the following:

$$\|f\|_{\Phi(H)} := \max_{\phi \in \Phi(H)} \langle f, \phi \rangle .$$

To see that this is indeed a norm, we can consider that $\Phi(H)$ is a centrally symmetric polytope of dimension n, thus $\max_{\phi \in \Phi(H)} \langle f, \phi \rangle$ is zero if and only if $f = 0$ by the hyperplane separation Theorem (section 2.3 of [3]). Moreover, absolute homogeneity comes from equation (1). Finally we leave triangle inequality as an exercise for the reader. In the following, we write $\| \cdot \|$ when $\Phi(H)$ is clear from the context.

A useful bit of notation is as follows.

Notation 13. *In Setting 2, given $\tilde{X}$ a subset of $[n]$, we denote by $\tilde{\mu}$ the functional $\tilde{\mu}(x) = p^{-1}(x \in \tilde{X})$ with domain $[n]$ and codomain $\{0, p^{-1}\} \subseteq \mathbb{R}$. We often denote with Φ the polytope* $\Phi(\tilde{\mu}, \mathbf{1})$.

We now have the language to state our main technical theorem, which is a functional version of Theorem 4.

Theorem 14. *Let us be in Setting 2. For every $\varepsilon > 0$ there exists $C > 0$ such that, if the C-conditions are satisfied, then with probability at least $1 - \exp\left(-\frac{pn}{C}\right)$ the random set $X = [n]_p$ admits an ε-deletion $\tilde{X}$ such that —using Notation 13— for every $\tilde{\mu}$-bounded functional f there exists a $\mathbf{1}$-bounded functional g such that $\|f - g\|_{\Phi(\tilde{\mu}, \mathbf{1})} \leq \varepsilon$.*

Something to note is that we have allowed f to be a general $\tilde{\mu}$-bounded function (not just a scaled indicator function of a subset of $\tilde{X}$, which would be the exact translation of Theorem 4) but we also relaxed our conclusion to let the dense model g be a $\mathbf{1}$-bounded function, not necessarily $\{0, 1\}$-valued. To prove Theorem 4, we need to return to integer-valued dense models, which is the subject of the next section.

5. Integer dense models

5.1. **Integer dense models suffice.** The following result says that we can approximate the dense model g given by Theorem 14 by an integer-valued model.

Theorem 15. *Let us be in Setting 2. For every $\varepsilon > 0$ there is $C > 0$ such that, if the C-conditions are satisfied, then for any functional $g : [n] \to [0,1]$, there is a functional $g^* : [n] \to \{0,1\}$ such that $\|g - g^*\|_{\Phi(\mathbf{1})} \le \varepsilon$.*

The proof of Theorem 4 from Theorem 14 and Theorem 15 is an exercise in functional analysis. We write the statement as functionals, then replace the sparse functional f representing $\tilde{X}$ with its fractional dense model g by a telescoping sum, then the fractional dense model with its integer dense model g^* by another telescoping sum. This proof contains the type or argument needed when converting a statement to the functional setting.

Proof of Theorem 4. We are in Setting 2. Given $\varepsilon > 0$ we can take C such that both Theorem 14 and Theorem 15 hold in Setting 2 with $\frac{1}{2k}\varepsilon$ (instead of ε) if the the C-conditions are satisfied.

Suppose now that the likely event of Theorem 14 occurs for $X = [n]_p$, and let $\tilde{X}$ be the set that this event provides. Now, for any given $Y \subseteq \tilde{X}$, let $f(y) = p^{-1}\mathbb{1}(y \in Y)$. By definition, f is $\tilde{\mu}$-bounded, so by Theorem 14 there is a $\mathbf{1}$-bounded g such that

$$\|f - g\|_{\Phi(\tilde{\mu},\mathbf{1})} \le \tfrac{1}{2k}\varepsilon\,. \tag{4}$$

By Theorem 15, there is an integer $\mathbf{1}$-bounded function g^* such that

$$\|g - g^*\|_{\Phi(\mathbf{1})} \le \tfrac{1}{2k}\varepsilon\,. \tag{5}$$

Let $Z = \{z \in [n] : g^*(z) = 1\}$. Given $\sigma \in \Sigma$, since σ and $-\sigma$ are in $\Phi(\tilde{\mu},\mathbf{1})$ and in its subset $\Phi(\mathbf{1})$, the inequalities (4) and (5) give us

$$\langle f, \sigma\rangle = \langle g, \sigma\rangle \pm \tfrac{\varepsilon}{2k} = \langle g^*, \sigma\rangle \pm \tfrac{\varepsilon}{k}$$

which, multiplying by n and filling in the definitions of inner product, f and g^*, gives 1.

Given now $\omega \in \Omega$, we have the telescoping expression

$$\begin{aligned}\langle f, *_{1,\omega}(f,\dots,f)\rangle &= \langle g, *_{1,\omega}(f,\dots,f)\rangle \pm \tfrac{\varepsilon}{2k} = \langle f, *_{2,\omega}(g,f,\dots,f)\rangle \pm \tfrac{\varepsilon}{2k}\\ &= \dots = \langle g, *_{k,\omega}(g,\dots,g)\rangle \pm \tfrac{1}{2}\varepsilon\,,\end{aligned}$$

where we have in total k replacements of an f with a g, in each case using that the corresponding convolution and its negative are in $\Phi(\tilde{\mu},\mathbf{1})$; and k rearrangements of terms, where the value does not change but the inner product is rewritten.

Repeating the same telescoping argument, but this time replacing each occurrence of g with g^*, and using that the corresponding convolutions are in $\Phi(\mathbf{1})$, we get

$$\begin{aligned}\langle g, *_{k,\omega}(g,\dots,g)\rangle &= \langle g, *_{1,\omega}(g,\dots,g)\rangle = \langle g^*, *_{1,\omega}(g,g,\dots,g)\rangle \pm \tfrac{\varepsilon}{2k}\\ &= \dots = \langle g^*, *_{k,\omega}(g^*,\dots,g^*)\rangle \pm \tfrac{1}{2}\varepsilon\,,\end{aligned}$$

Putting these two expressions together we have

$$\langle f, *_{1,\omega}(f,\dots,f)\rangle = \langle g^*, *_{k,\omega}(g^*,\dots,g^*)\rangle \pm \varepsilon\,,$$

which filling in the definitions of f, g^*, inner product and convolution, and multiplying by $n \cdot \frac{e(S)}{n}$, is 2. $\square$

5.2. **Random splitting: a useful technique.** In this section we prove Theorem 15. We start by giving a sketch of the approach, as some of the ideas reappear later. In particular, we use a refinement of similar techniques to prove Theorem 18.

We start by defining g^* via *randomised rounding*. That is, independently for each x, we generate $g^*(x)$ by choosing 1 with probability $g(x)$ and 0 otherwise. We then argue that the required closeness in norm is likely.

A first intuitive approach would be to try leverage our optimization Lemma 9 and say that the extremal value is attained at a vertex. This would allow us to argue that for any given vertex ϕ of $\Phi(\mathbf{1})$, with high probability we have $\langle g - g^*, \phi\rangle < \varepsilon$ and then take a union bound over the choices of ϕ. The reason to believe this might work is that $g(x) - g^*(x)$ is, for each $x \in [n]$, a random variable in $[-1, 1]$ with mean zero, while ϕ is a fixed vector, so the inner product is a sum of independent mean zero random variables. Unfortunately, this fails by a technical detail: there are too many choices of vertex for the required union bound. To get around this, we now define a polytope which contains $\Phi(\mathbf{1})$ but has fewer vertices.

Definition (Random split). *Let L be a positive integer, and let $\chi : [n] \to [L]$ be a sample of the uniform random function. For $i \in [L]$ we then denote by ν_i the function on $[n]$ such that $\nu_i(x) = L$ if $\chi(x) = i$, and $\nu_i(x) = 0$ otherwise. We have $\mathbf{1} = \frac{1}{L}\sum_{i=1}^{L} \nu_i$. We call this a* random split *of* $\mathbf{1}$.

By linearity, every vertex of $\Phi(\mathbf{1})$ is a convex combination of vertices of $\Phi(\nu_1, \ldots, \nu_L)$, so it suffices to show $\langle g - g^*, \phi\rangle < \varepsilon$ holds for all vertices $\phi \in \Phi(\nu_1, \ldots, \nu_L)$.

It follows from the $\Delta_1(S)$ bounds of the C-conditions and from the definition of $\Phi(\mathbf{1})$ that any $\phi \in \Phi(\mathbf{1})$ only attains values with absolute value at most c. Unfortunately, no such bound holds for functionals in $\Phi(\nu_1, \ldots, \nu_L)$, which can attain values as large as $L^{k-1}c$. Such large values spoil the concentration we require of the random variable $\langle g - g^*, \phi\rangle$. We deal with this by splitting up ϕ in two components ϕ^{small} and ϕ^{big}: we define $\phi^{\text{small}}(x) = \phi(x)\mathbb{1}(|\phi(x)| \le 2c)$, and $\phi^{\text{big}} = \phi - \phi^{\text{small}}$.

We can now write $\langle g - g^*, \phi\rangle = \langle g - g^*, \phi^{\text{small}}\rangle + \langle g - g^*, \phi^{\text{big}}\rangle$. The point of this is that the random variable $\langle g - g^*, \phi^{\text{small}}\rangle$ does concentrate well, while we can use a high moment argument to show that $\langle g - g^*, \phi^{\text{big}}\rangle$ is tiny. Importantly, while our concentration argument needs to take a union bound over all vertices of $\Phi(\nu_1, \ldots, \nu_L)$ (i.e. all $\{\nu_1, \ldots, \nu_L\}$-extreme functions and their negatives), we only need to bound high moments of the $\{\nu_1, \ldots, \nu_L\}$-largest anti-uniform functions.

We start with a technical lemma that has apparently nothing to do with the proof we want to show. We present this lemma separately because we use it also in a later section.

Definition. *Let us be in Setting 2. Let d be a positive integer, $x \in [n]$ and $i_1, \ldots, i_d \in [k]$. A* configuration *with spine x and index tuple $(i_1, \ldots, i_d)$ is a tuple $(s^1, \ldots, s^d)$ of edges of S such that $s^j_{i_j} = x$. If $i_1 = \ldots = i_d = i$, we call this a d-book with spine x.*

For $\mathbf{i} = (i_1, \ldots, i_d)$ and t a positive integer, we denote by $\alpha(\mathbf{i}, t, x)$ the number of configurations $(s^1, \ldots, s^d)$ with spine x and index tuple $\mathbf{i}$ such that $\left|\cup_i s^i \setminus \{x\}\right| = t$.

Lemma 16. *Let us be in Setting 2. Let $C > 0$ be a positive real number and let d and t be positive integers, with $k - 1 \le t \le d(k - 1)$. If the C-conditions are satisfied, then for any $x \in [n]$ and any $\mathbf{i} = (i_1, \ldots, i_d)$ we have:*

$$\alpha(\mathbf{i}, t, x) \le t^d \cdot (2^{dk}k!)^d \cdot c^d C^{d+t-kd} p^{kd-d-t} e(S)^d n^{-d}\,.$$

Moreover, for $t = d(k-1)$ we have:

$$\alpha(\mathbf{i}, t, x) \le c^d e(S)^d n^{-d}$$

Proof. Let us fix x, t, d and $\mathbf{i}$. We now describe a process that can generate any configuration with spine x, index tuple $(i_1, \ldots, i_d)$, and covering t vertices besides x. By counting the number of choices we make until a specific configuration is selected, we can upper bound $\alpha(\mathbf{i}, t, x)$. We start by picking non-negative integers $m_1, \ldots, m_d \le k-1$ with $m_1 = k-1$. We choose s^1 to be an edge of S whose i_1-th element is x. We then pick $k - m_2$ elements, including x, among the k elements of s^1 that are also to be contained in S_2. We then fix a position of these elements in S_2, which is an injection from these $k - m_2$ elements to $[k]$, making sure that x is assigned position i_2 in s^2. We then select an element s^2 of S that satisfy these constraints. We repeat a similar procedure, fixing $k - m_3$ elements of S_2 to generate S_3 (fixing x in i_3 for s^3), and repeat the procedure until we get s^d.

In this procedure, the main contribution to the number of books constructed comes from choosing the $m_1 = k-1$ new elements of s^1, the m_2 new elements of s^2, and so on; the number of ways to do the i-th step is a constant —that counts the number of ways we have to fix elements of the previous edges into the new one, and can be upper-bounded by $2^k k!$, a constant— multiplied by the codigree of S of the right magnitude $\Delta_{k-m_i}(S)$ for which we have by hypothesis the upper-bound $cC^{1+m_i-k}p^{l-1-m_i}\frac{e(S)}{n}$. The total number of elements of $[n] \setminus \{x\}$ our constructed book covers is at most $\sum_{i=1}^d m_i$ (we do not enforce that the 'new' elements are really distinct from the previously chosen ones). We can therefore ignore books which cover too few elements of $[n]$ and assume $\sum_{i=1}^d m_i = t$. This means that the product of codegrees we get is $c^d C^{d+t-kd} p^{kd-d-t} e(S)^d n^{-d}$. This gives an upper bound on $\alpha(\mathbf{i}, t, x)$ of

$$\alpha(\mathbf{i}, t, x) \le t^d \cdot (2^{dk}k!)^d \cdot c^d C^{d+t-kd} p^{kd-d-t} e(S)^d n^{-d} .$$

Indeed, the t^d counts the ways to choose $m_1, \ldots, m_d$, the factor $2^{dk}k!$ corresponds to picking a subset of used elements and an injection to $[k]$, and the final product is the product of codegrees. Note that in one special case we can do better: when $t = d(k-1)$, we have $m_1 = \ldots = m_d = k-1$, and we do not have to pick any used elements (we must pick x and no other element every time) nor injection (x must be the i_j-th vertex of each S_j, and no other elements are repeated) and we get the upper bound $\alpha(\mathbf{i}, t, x) \le c^d e(S)^d n^{-d}$ on the number of these books. □

We are now ready for the proof of Theorem 15.

Proof of Theorem 15. We are in Setting 2. Given $\varepsilon > 0$, let $d \ge 4$ be an integer such that $2^{4-d}c^2 \le \frac{1}{2}\varepsilon$. Let $L = \lceil 1000c^2 d\varepsilon^{-2} \rceil$ be another integer, and set

$$C = 100(dk)^{d+1}(2^{dk}k!)^d L .$$

We can assume now that the C-conditions are satisfied in our setting. Let $\nu_1, \ldots, \nu_L$ be a random split of $\mathbf{1}$. In the following claim, recall that when ϕ is a vector, ϕ^d denotes the pointwise power.

Claim 17. *With high probability*[3]*, the following properties are satisfied. For each $i \in [L]$, we have the inequality $\langle \mathbf{1}, \nu_i \rangle \le 2$; and in addition, for each $j \in [k]$ and $i_1, \ldots, i_k \in [L]$, we have*

$$\langle \mathbf{1}, \left(*_{j,\mathbf{1}} (\nu_{i_1}, \ldots, \nu_{i_k}) \right)^d \rangle \le 2c^d .^4$$

[3]With probability tending to 0 as n tends to ∞.

This claim is our bound on high moments of the $\{\nu_1, \dots, \nu_L\}$-extreme functions.

Proof. For the first statement, fix i. As $\langle \mathbf{1}, \nu_i \rangle = \frac{1}{n} \sum_{x=1}^{n} \nu_i(x)$, we are asking for the probability that ν_i has more than $2n/L$ entries equal to L. If we consider $\sum_x \mathbb{1}(\nu_i(x) = L)$, this is a binomial random variable with mean n/L, so by Chernoff's inequality (Theorem 5) the probability that it exceeds $2n/L$ is at most $\exp\big(-\frac{1}{3} n/L\big)$. Considering an union bound over i, the probability of failure of the first statement is $o(1)$.

For the second statement, fix j and $i_1, \dots, i_k$. Let $Z = \langle \mathbf{1}, \big(*_{j,\mathbf{1}} (\nu_{i_1}, \dots, \nu_{i_k})\big)^d \rangle$. We first argue that $\mathbb{E}[Z] \le \frac{3}{2} c^d$. We have

$$
\begin{aligned}
Z &= \frac{1}{n} \sum_{x \in [n]} \big(*_{j,\mathbf{1}} (\nu_{i_1}, \dots, \nu_{i_k})\big)^d (x) = \sum_{x \in [n]} \frac{1}{n} \frac{n^d}{e(S)^d} \left(\sum_{s \in S_j(x)} \prod_{t \neq j} \nu_{i_t}(s_t) \right)^d \\
&= \sum_{x \in [n]} \frac{1}{n} \frac{n^d}{e(S)^d} L^{d(k-1)} \left(\sum_{s \in S_j(x)} \prod_{t \neq j} \mathbb{1}(\chi(s_t) = i_t) \right)^d \\
&= \sum_{x \in [n]} \sum_{s^1, \dots, s^d \in S_j(x)} \frac{1}{n} \frac{n^d}{e(S)^d} L^{d(k-1)} \cdot \prod_{t \neq j} \prod_{h=1}^{d} \mathbb{1}(\chi(s_t^h) = i_t)\,.
\end{aligned}
$$

Notice that the internal sum of our last equation is a sum of d-books with spine x. Each term of said sum takes value either zero or $\frac{1}{n} \cdot \frac{n^d}{e(S)^d} \cdot L^{(k-1)d}$. Let us fix a d-book $s^1, \dots, s^d$, and let us ask what is the probability that the internal sum takes the larger value. If we let $Q = \cup_{t=1}^{d} s^t \setminus \{x\}$ and $q = |Q|$, the probability depends only on q. Indeed, notice that for each element of Q, the random variable χ needs to attain a specific value, otherwise the whole term is set to zero. Therefore, given $s^1, \dots, s^d$, the probability that the corresponding element of the sum takes value $\frac{1}{n} \cdot \frac{n^d}{e(S)^d} \cdot L^{(k-1)d}$ is at most L^{-q} (it can be that the probability is zero, for example if we have $\nu_{i_1}(y)\nu_{i_2}(y)$ as a term in our sum). Notice that lower values of q imply larger probability that the corresponding element of the sum samples the higher value. We can use Lemma 16 to count the number of books as follows.

Fix $\mathbf{i} = (j, \dots, j)$ a d-tuple with all entries equal to j. For the calculation of the expectation of Z, we need the following to bound the main term.

$$
\sum_{x \in [n]} \sum_{s^1, \dots, s^d \in S_j(x)} \prod_{t \neq j} \prod_{h=1}^{d} \mathbb{1}(\chi(s_t^h) = i_t) \le \sum_{x \in [n]} \sum_{q=k-1}^{d(k-1)} \alpha(\mathbf{i}, q, x) L^{-q}\,.
$$

If we insert the bounds of Lemma 16 in the calculations we obtain:

$$
\mathbb{E}[Z] \le c^d + \tfrac{1}{2} c^d\,.
$$

where the first term c^d is the $q = d(k-1)$, and by choice of C each other term in the sum contributes at most $\frac{1}{2}(dk)^{-1} c^d$.

We next want to apply McDiarmid's inequality (Theorem 7) to Z. We therefore need to argue that Z does not vary a lot when just one component of the colouring χ is changed. For any fixed $y \in [n]$, consider that changing the colouring at y affects only the terms of the sum Z where y is in at least one edge of the book $s^1, \dots, s^d$. As before, we upper-bound the

[4] The two $\mathbf{1}$ in this statement are functionals over different domains.

number of these terms by showing a procedure that can generate any such book containing y, and keeping track of the choices we made. We start by picking $i \in [d]$ and $i' \in [k]$ such that y is vertex number i' of s^i. Because the C-conditions are satisfied, $\Delta_1(S) \leq c\frac{e(S)}{n}$, and therefore there are at most $c\frac{e(S)}{n}$ choices of s^i containing y in position i'. For the same reason, the remaining $d-1$ elements of the book (which all contain x at position j) can be chosen in at most $c^{d-1}\frac{e(S)^{d-1}}{n^{d-1}}$ ways. This means that the chance of value of χ at y can influence at most $dkc^d e(S)^d n^{-d}$ terms (we multiplied by d to take into account the choice of i and by k to take into account the choice of i'). Since each term takes value either 0 or $\frac{1}{n}\frac{n^d}{e(S)^d}L^{d(k-1)}$, the effect of changing the colouring at y is at most

$$\tfrac{1}{n}\tfrac{n^d}{e(S)^d}L^{d(k-1)} \cdot dkc^d e(S)^d n^{-d} = \tfrac{1}{n}L^{d(k-1)}dkc^d\,.$$

Applying McDiarmid's inequality, the probability that Z exceeds its expectation by $\frac{1}{2}c^d$ is at most

$$\exp\big(-\tfrac{2\cdot\frac{1}{4}c^{2d}}{n\cdot(L^{d(k-1)}dkc^d n^{-1})^2}\big)\,,$$

which tends to zero exponentially in n.

Taking the union bound over the at most kL^{d-1} choices of j and $i_1,\dots,i_{k-1}$, the failure probability for the second statement is $o(1)$. □

Let g be a $\mathbf{1}$-bounded functional from $[n]$ to $[0,1]$. Our aim is to prove that there exists a functional $g^*:[n]\to\{0,1\}$ such that $\|g-g^*\|_{\Phi(\mathbf{1})} \leq \varepsilon$. We take g^* to be a random rounding of g, which means that for each $x\in[n]$ we sample $g^*(x)$ independently at random to take the value 1 with probability $g(x)$.

By Claim 17, there exists $\nu_1,\dots,\nu_L$ a random split such that the likely event of Claim 17 holds (otherwise it wouldn't hold with high probability). Fix such $\nu_1,\dots,\nu_L$. In order to prove that $\|g-g^*\|_{\Phi(\mathbf{1})} \leq \varepsilon$ we first show that for any ϕ an arbitrary vertex of $\Phi(\nu_1,\dots,\nu_L)$, we have $\langle g-g^*,\phi\rangle \leq \varepsilon$. We then show that this is enough because $\Phi(\mathbf{1}) \subseteq \Phi(\nu_1,\dots,\nu_L)$ and because linear functions attain their maximum over a polytope at a vertex (Lemma 9).

For a vertex ϕ of $\Phi(\nu_1,\dots,\nu_L)$, we write $\phi^{\text{small}}(x) := \phi(x)\mathbb{1}\big(|\phi(x)| \leq 2c\big)$ and $\phi^{\text{big}} = \phi - \phi^{\text{small}}$. We first prove that for each vertex ϕ of $\Phi(\nu_1,\dots,\nu_L)$ we have $\langle g-g^*,\phi^{\text{small}}\rangle \leq \frac{\varepsilon}{2}$, and then we prove a similar statement for ϕ^{big}.

For ϕ^{small}, we do this by union-bounding, for the choice of ϕ, the probability that we selected a g^* that is too far from g with respect to ϕ^{small}. To apply the union bound, we start by considering that the number of vertices of $\Phi(\nu_1,\dots,\nu_L)$ is at most

$$|\Sigma| + k\cdot|\Omega|\cdot L^{k-1}2^{(k-1)2n/L}\,.$$

Indeed, every element of $\Phi(\nu_1,\dots,\nu_L)$ can be seen as the convex combination of elements of Σ (at most $|\Sigma|$ many) and $\nu_1,\dots,\nu_L$-extreme anti-uniform functionals (by Remark 12). The number of these latter functionals can be bounded by the fact that each of them is determined by being of the form $*_{i,\omega}(f_1,\dots,f_k)$, where there are k choices for the value i; there are $|\Omega|$ choices for ω; and each of the $k-1$-many f_j comes from the selection of one of L-many elements of $\{\nu_1,\dots,\nu_L\}$ and a subset of the at most $2n/L$-many (by Claim 17) non-zero entries of the selected element of $\{\nu_1,\dots,\nu_L\}$.

We now observe that, considering g^* as a random variable with $\mathbb{E}[g^*(x)] = g(x)$, we have that $\langle g-g^*,\phi^{\text{small}}\rangle$ is a sum of n-many 0-mean random variables, each with range at most

$2cn^{-1}$ and so variance at most c^2n^{-2} [2]. Applying Bernstein's inequality, we have

$$\mathbb{P}\big[\langle g-g^*,\phi^{\mathrm{small}}\rangle > \tfrac{1}{2}\varepsilon\big] \le \exp\Big(-\frac{\varepsilon^2/4}{\frac{4}{3}\cdot\frac{1}{n}\cdot\frac{1}{2}\varepsilon+n\cdot c^2n^{-2}}\Big) \le \exp\Big(-\frac{\varepsilon^2 n}{32c^2}\Big)\,.$$

By choice of L, taking the union we obtain that with high probability we have $\langle g-g^*,\phi^{\mathrm{small}}\rangle \le \frac{1}{2}\varepsilon$ for every vertex ϕ of $\Phi(\nu_1,\dots,\nu_L)$. Therefore, there must exist a g^* for which this condition holds. Fix such a g^*.

Let us now prove that $\langle g-g^*,\phi^{\mathrm{big}}\rangle \le \frac{\varepsilon}{2}$ for all ϕ in $\Phi(\nu_1,\dots,\nu_L)$. Take such a ϕ and let ψ be a $\{\nu_1,\dots,\nu_L\}$-largest anti-uniform functional such that $\phi\le\psi$ pointwise (which exists, as discussed in Remark 12). We have

$$\begin{aligned}|\langle g-g^*,\phi^{\mathrm{big}}\rangle| &\le |\langle g,\phi^{\mathrm{big}}\rangle| + |\langle g^*,\phi^{\mathrm{big}}\rangle|\\ &\le \langle g,|\phi^{\mathrm{big}}|\rangle + \langle g^*,|\phi^{\mathrm{big}}|\rangle \le 2\langle \mathbf{1},|\phi^{\mathrm{big}}|\rangle\\ &\le \langle \mathbf{1},(\phi^{\mathrm{big}})^2\rangle \le 2\cdot(2c)^{2-d}\langle \mathbf{1},(\phi^{\mathrm{big}})^d\rangle\\ &\le 2\cdot(2c)^{2-d}\langle\mathbf{1},\psi^d\rangle \le 4c^d(2c)^{2-d} \le \frac{\varepsilon}{2}\,,\end{aligned}$$

where the first line holds by triangle inequality, the second line holds by non-negativity of g, g^* and $|\phi^{\mathrm{big}}|$, the third line holds because $|\phi^{\mathrm{big}}|$ is bounded pointwise by $(\phi^{\mathrm{big}})^2$ and because all these entries are either zero or at least $2c$. The final line follows since $\phi^{\mathrm{big}}\le\phi\le\psi$ pointwise, and then uses Claim 17. By choice of d, this final number is at most $\frac{1}{2}\varepsilon$.

Putting these two estimates together, we have for every vertex ϕ of $\Phi(\nu_1,\dots,\nu_L)$ the bound

$$\langle g-g^*,\phi\rangle \le \tfrac{1}{2}\varepsilon+\tfrac{1}{2}\varepsilon=\varepsilon\,.$$

Since linear functionals over a polytope are maximised at vertices, we conclude the same bound holds for every $\phi\in\Phi(\nu_1,\dots,\nu_L)$.

To complete the proof, we now show $\Phi(\mathbf{1})\subseteq\Phi(\nu_1,\dots,\nu_L)$. Because both sets are polytopes, it is enough to show that all vertices of $\Phi(\mathbf{1})$ are in $\Phi(\nu_1,\dots,\nu_L)$. Given a vertex ϕ of $\Phi(\mathbf{1})$, either $\phi\in\Sigma$ —in which case $\phi\in\Phi(\nu_1,\dots,\nu_L)$ and we are done—, or $\phi=*_{i,\omega}(f_1,\dots,f_k)$ for some $\mathbf{1}$-bounded functions $f_1,\dots,f_k$. For each $j\in[k]$ and $t\in[L]$, let $f_{j,t}(x):=f_j(x)\nu_t(x)$, which is ν_t-bounded. By definition of random split, $f_j=\frac{1}{L}\sum_{t\in[L]}f_{j,t}$. Therefore, we have by linearity

$$*_{i,\omega}(f_1,\dots,f_k)=L^{1-k}\sum_{t_1,\dots,t_k\in[L]}*_{i,\omega}(f_{1,t_1},\dots,f_{k,t_k})\,,$$

which is a convex combination of elements of $\Phi(\nu_1,\dots,\nu_L)$. □

6. Reduction to anti-correlation

We now show that an anti-correlation statement implies Theorem 14. Which is, we reduce Theorem 14 to the following.

Theorem 18. *Let us be in Setting 2. For every $\varepsilon>0$ there exists $C>0$ such that, if the C-conditions are satisfied, then with probability at least $1-\exp\big(-\frac{pn}{C}\big)$ the random set $X=[n]_p$ admits an ε-deletion $\tilde{X}$ such that —using Notation 13— for every $\phi\in\Phi(\tilde{\mu},\mathbf{1})$ we have $\langle\tilde{\mu}-\mathbf{1},\phi^+\rangle<\varepsilon$ and $|\langle\tilde{\mu},\phi\rangle|,|\langle\mathbf{1},\phi\rangle|\le 2c$. In addition we have $\|\tilde{\mu}-\mathbf{1}\|_{\Phi(\tilde{\mu},\mathbf{1})}<\varepsilon$.*

We now show, following closely the proof of Lemma 2.5 of Conlon and Gowers' paper [4], that Theorem 18 implies Theorem 14.

Proof of Theorem 14. We are in Setting 2. Given $\varepsilon > 0$, let $\delta = \frac{1}{10c}\varepsilon^2$. Take C such that Theorem 18 holds in Setting 2 with δ (in place of ε) if the C-conditions are satisfied. Suppose that the likely event of Theorem 18 occurs; that is, we are given $\tilde{\mu}$ such that $\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle < \delta$ and $|\langle \mathbf{1}, \phi \rangle| \leq 2$ for every $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$. For the rest of the proof, we only consider the polytope $\Phi(\tilde{\mu}, \mathbf{1})$ and simply denote it by Φ.

Suppose now that f is some $\tilde{\mu}$-bounded function which contradicts the conclusion of Theorem 14. That is, we cannot write $f = g + h$ where g is $\mathbf{1}$-bounded and $\|h\|_\Phi < \varepsilon$.

We first show that we can write $\frac{1}{1+2\varepsilon^{-1}\delta} f = g + h$ where g is $\mathbf{1}$-bounded and $\|h\|_\Phi \leq \frac{1}{2}\varepsilon$. Suppose for a contradiction that this is impossible. The set K of functions of the form $g + h$ where g is $\mathbf{1}$-bounded and $\|h\|_\Phi \leq \frac{1}{2}\varepsilon$ is a convex set containing the zero function, since the $\mathbf{1}$-bounded functions form a hypercube (which is convex) containing zero, and norm-balls are convex and contain zero. By the Hahn-Banach Theorem (Theorem 10), if $\frac{1}{1+2\varepsilon^{-1}\delta} f$ is not in K there is a hyperplane separation. Because linear functionals can be represented as scalar products, this means that there is $\psi \in \mathbb{R}^{[n]}$ such that $\langle \frac{1}{1+2\varepsilon^{-1}\delta} f, \psi \rangle > 1$ but $\langle g + h, \psi \rangle \leq 1$ for all $\mathbf{1}$-bounded g and $\|h\|_\Phi \leq \frac{1}{2}\varepsilon$.

A functional analysis argument shows that $\phi = \frac{1}{2}\varepsilon\psi$ is in Φ. To see this, we consider that $\max_{h' \in \Phi} \langle h', \psi \rangle \leq 2\varepsilon^{-1}$ due to linearity of the product and from the Hahn-Banach Theorem (and that 0 is a $\mathbf{1}$-bounded function). From this, we obtain that the dual norm (see [14, Ch. 4]) $\|\psi\|^*$ is at most $2\varepsilon^{-1}$, which is sufficient to conclude, considering that Φ is a full-dimensional polytope containing zero. The fact that $\phi \in \Phi$ gives us, because of Theorem 18, that $\langle \tilde{\mu} - \mathbf{1}, \psi^+ \rangle < 2\varepsilon^{-1}\delta$. If we let $\bar{g}(x) = \mathbb{1}(\psi(x) \geq 0)$, we can write

$$1 + 2\varepsilon^{-1}\delta < \langle f, \psi \rangle \leq \langle f, \psi^+ \rangle \leq \langle \tilde{\mu}, \psi^+ \rangle < \langle \mathbf{1}, \psi^+ \rangle + 2\varepsilon^{-1}\delta = \langle \bar{g}, \psi \rangle + 2\varepsilon^{-1}\delta \,. \tag{6}$$

Where the first inequality comes from Hahn-Banach, the second from considering that f is non-negative, the third from the fact that f is $\tilde{\mu}$-bounded, the fourth we just proved, and the last equality follows by definition of $\bar{g}$. Since $\bar{g}$ is $\mathbf{1}$-bounded, we have $\langle \bar{g}, \psi \rangle \leq 1$. But (6) now reads $1 + 2\varepsilon^{-1}\delta < 1 + 2\varepsilon^{-1}\delta$, a contradiction.

We can therefore write $f = g + 2\varepsilon^{-1}\delta g + h$, where g is $\mathbf{1}$-bounded and $\|h\|_\Phi < \frac{1}{2}\varepsilon$. By triangle inequality, to complete the proof it suffices to show $\|2\varepsilon^{-1}\delta g\|_\Phi \leq \frac{1}{2}\varepsilon$. But this is equivalent to showing that for every element ϕ of Φ, we have $\langle g, \phi \rangle \leq \frac{1}{4}\varepsilon^2\delta^{-1}$.

As Φ is the convex hull of non-negative elements and their negatives, and as g is non-negative, we can assume that ϕ is non-negative as well. We can thus write

$$\langle g, \phi \rangle \leq \langle \mathbf{1}, \phi \rangle \leq 2c$$

where the first inequality holds because $g \leq \mathbf{1}$ and ϕ is non-negative, and the second is by Theorem 18. By choice of δ, this proves $\|2\varepsilon^{-1}\delta g\|_\Phi \leq \frac{1}{2}\varepsilon$. □

In this proof we did not use the conclusion $\|\tilde{\mu} - \mathbf{1}\|_{\Phi(\tilde{\mu},\mathbf{1})} < \varepsilon$ of Theorem 18, however this is a convenient fact to record.

The rest of this work is concerned with proving Theorem 18. The proof of this result follows the same broad lines that we followed in proving Theorem 15. There are however some important differences. Before entering in details in the next sections, we give a broad informal outlook of these differences.

First, in Theorem 18 we need to optimize for ϕ^+, which is not linear in ϕ. This means that that we cannot assume $\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle$ is maximised at a vertex of Φ as we did in Theorem 15. In the following Section 7 we show how to reduce this problem to a linear (and thus maximised at a vertex) optimisation problem over a different polytope.

Second, in Theorem 15 we were approximating a $[0,1]$-valued functional via a random rounding. In Theorem 18 we have to obtain concentration inequalities for $\tilde{\mu}$, which is a sparse random function. Therefore the kind of concentration we can hope for is much weaker. However, we still need an optimisation over $\Phi(\tilde{\mu},\mathbf{1})$, which has $2^{\Omega(n)}$ vertices. Thus, the same union bounds that we used in Theorem 15 would simply not work here.

Third, $\Phi(\tilde{\mu},\mathbf{1})$ itself depends on the randomness in $\tilde{\mu}$. Therefore, one cannot fix a vertex ϕ of Φ before revealing $\tilde{\mu}$.

It turns out that a concept similar to the previously-defined 'random splitting' deals with both these second and third problems; we describe the random splitting in Section 8 and prove it does the job in Section 9.

Finally, entries of ϕ^+ can be as large as $\log n$, which makes bounding inner products more difficult. However, the same idea that worked for Theorem 15 —applying moment bounds to control exceptionally large entries— works just as well here. We prove the required moment bounds hold with high probability in Section 10, and in Section 11 we show that this high probability can (at the cost of some deletion) be upgraded to exponentially high probability.

7. A linear approximation

Part of proving Theorem 18 is to show that for every $\phi \in \Phi(\tilde{\mu},\mathbf{1})$ we have $\langle \tilde{\mu}-\mathbf{1},\phi^+\rangle < \varepsilon$. The difficulty in proving this statement is that the function $\phi \to \phi^+$ makes this a non-linear optimisation problem over $\Phi(\tilde{\mu},\mathbf{1})$. Thus, we cannot use, out-of-the-box, that $\langle \tilde{\mu}-\mathbf{1},\phi^+\rangle$ is maximised —as a function of ϕ in $\Phi(\tilde{\mu},\mathbf{1})$— at a vertex of $\Phi(\tilde{\mu},\mathbf{1})$. We show in this section that we can get around this by using the Weierstrass Approximation Theorem (Theorem 11) to approximate $\phi \to \phi^+$ with a polynomial. As we now see, this translates our optimisation problem to a linear one over the *product polytope* $\Phi^d := \{\prod_{i=1}^d \phi_i : \phi_i \in \Phi(\tilde{\mu},\mathbf{1})\}$, with $d \in \mathbb{N}$ determined by Weierstrass' Approximation Theorem. Since the constant $\mathbf{1}$ function is in Σ, and therefore in $\Phi(\tilde{\mu},\mathbf{1})$, any product of at most d elements of $\Phi(\tilde{\mu},\mathbf{1})$ is in Φ^d. Any vertex of Φ^d is a product of d vertices of $\Phi(\tilde{\mu},\mathbf{1})$.

We need to be careful because the Weierstrass Approximation Theorem allows us to approximate well the function $x \to x^+$ only within a closed and bounded interval: we use the interval $[-2c,2c]$. We show using high moment bounds that the contribution to the inner product where ϕ lies outside of this interval is almost surely negligible. This argument is broadly similar to the one used in the proof of Theorem 15.

To this end, for any function ϕ on $[n]$, write ϕ^{big} for the function which takes the value $\phi(x)$ on $x \in [n]$ whenever $|\phi(x)| > 2c$, and 0 otherwise, and $\phi^{\mathrm{small}} = \phi - \phi^{\mathrm{big}}$. That is,

$$\phi^{\mathrm{small}}(x) = \begin{cases} \phi(x) & \text{if } \phi(x) \in [-2c,2c] \\ 0 & \text{otherwise} \end{cases} \qquad \phi^{\mathrm{big}}(x) = \begin{cases} 0 & \text{if } \phi(x) \in [-2c,2c] \\ \phi(x) & \text{otherwise} \end{cases} .$$

Note that ϕ^{big} and ϕ^{small} have disjoint support. The aim of this section is to prove the following deterministic reduction of Theorem 18, which tells us that the above sketched approach works.

Lemma 19. *Let us be in Setting 2 and let $\tilde{X}$ be a subset of $[n]$. Let us use Notation 13. For any $\varepsilon' > 0$ there exist $\varepsilon > 0$ and d, d', with d' even, such that if the following holds:*

(1) For all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})^d$, we have $|\langle \tilde{\mu} - \mathbf{1}, \phi \rangle| < \varepsilon$,
(2) For all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$ we have $|\langle \tilde{\mu}, \phi^{d'} \rangle|, \langle \mathbf{1}, \phi^{d'} \rangle| \leq 2c^{d'}$,

then for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$ we have $\left|\langle \tilde{\mu} - \mathbf{1}, \phi^{+} \rangle\right| < \varepsilon'$.

Proof. Recall that by ϕ^+ we mean applying the operator $\cdot^+$ on each component of ϕ. Therefore, in particular for any $x \in [n]$ we have $\phi^+(x) = \phi(x)^+$.

Consider the functional $\cdot^+ : [-2c, 2c] \to \mathbb{R}^+$ (which we remind the reader can defined as $x^+ := x \cdot \mathbb{1}(x \geq 0)$). This is a continuous function from a closed interval of $\mathbb{R}$ to $\mathbb{R}^+$. By Weierstrass Approximation Theorem (Theorem 11), for any given $\varepsilon' > 0$, we can find a polynomial $P(x) = a_d x^d + \ldots + a_1 x + a_0$ of maximum degree d such that for any $x \in [-2c, 2c]$ we have $|P(x) - x^+| < \frac{\varepsilon'}{12}$ (note that without loss of generality we can assume $d \geq 2$). Define now $M = \max_{i \in \{0,\ldots,d\}} |a_i|$ and set $\varepsilon = \frac{\varepsilon'}{2M(d+1)}$. Moreover, set d' to be the smallest positive even integer such that $2^{1-d'}(2c)^{2d} \leq \frac{\varepsilon'}{8M(d+1)+8}$.

What we want to do is to upper bound $|\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle|$ given an arbitrary $\phi \in \Phi$. We use the linearity of the inner product and triangle inequality to obtain the following inequality.

$$|\langle \tilde{\mu} - \mathbf{1}, \phi^+ \rangle| \leq |\langle \tilde{\mu} - \mathbf{1}, P(\phi) \rangle| + |\langle \tilde{\mu} - \mathbf{1}, P(\phi) - \phi^+ \rangle| . \tag{7}$$

We remind the reader that every operator here and in the following is defined componentwise. Therefore, $P(\phi)$ is defined as the functional such that $P(\phi)(x) = P(\phi(x))$. We now upper bound each of the right hand side terms with $\frac{\varepsilon'}{2}$.

To upper bound the first term $|\langle \tilde{\mu} - \mathbf{1}, P(\phi) \rangle|$, we expand the polynomial into its terms. Using again linearity of the inner product and triangle inequality, we obtain

$$|\langle \tilde{\mu} - \mathbf{1}, P(\phi) \rangle| = |\langle \tilde{\mu} - \mathbf{1}, \sum_{i=0}^{d} \phi^i \rangle| \leq M \sum_{i=0}^{d} |\langle \tilde{\mu} - \mathbf{1}, \phi^i \rangle| .$$

For any $i \in \{0, \ldots, d\}$ and $\phi \in \Phi$, we have that $\phi^i \in \Phi^d$. Indeed, we have that $\mathbf{1} \in \Phi$, and therefore we can make up for the missing $d - i$ terms by multiplying $\phi^i \cdot \mathbf{1}^{d-i} = \phi^i$. Therefore, by 1 of Lemma 19 we have $|\langle \tilde{\mu} - \mathbf{1}, \phi^i \rangle| \leq \varepsilon = \frac{\varepsilon'}{2M(d+1)}$. Summing over the various terms, we obtain

$$|\langle \tilde{\mu} - \mathbf{1}, P(\phi) \rangle| \leq M(d+1)\varepsilon \leq \tfrac{\varepsilon'}{2} .$$

We now turn to the second term of (7), for which we apply the splitting of ϕ into the two functionals ϕ^{big} and ϕ^{small}. As before, we have $\phi^{\text{small}}(x) = x\mathbb{1}(|x| \leq 2x)$ and $\phi^{\text{big}}(x) = \phi(x) - \phi^{\text{small}}(x)$. In general, neither of these is in the polytope Φ. Since $\phi^{\text{big}}, \phi^{\text{small}}$ have disjoint support, and all operations are done pointwise, we have

$$P\big(\phi^{\text{small}} + \phi^{\text{big}}\big) - \big(\phi^{\text{small}} + \phi^{\text{big}}\big)^+ = P\big(\phi^{\text{small}}\big) - \big(\phi^{\text{small}}\big)^+ + P\big(\phi^{\text{big}}\big) - \big(\phi^{\text{big}}\big)^+ .$$

In order to complete the proof, by linearity of inner product, it suffices to show

$$|\langle \tilde{\mu}, P(\phi^{\text{small}}) - (\phi^{\text{small}})^+ \rangle|, |\langle \mathbf{1}, P(\phi^{\text{small}}) - (\phi^{\text{small}})^+ \rangle| \leq \tfrac{\varepsilon'}{8} \qquad \text{and} \tag{8}$$

$$|\langle \tilde{\mu}, P(\phi^{\text{big}}) - (\phi^{\text{big}})^+ \rangle|, |\langle \mathbf{1}, P(\phi^{\text{big}}) - (\phi^{\text{big}})^+ \rangle| \leq \tfrac{\varepsilon'}{8} . \tag{9}$$

Of these, we address (8) first. Consider first that by definition of ϕ^{small}, we have that $\phi^{\text{small}}(x)$ is always in $[-2c, 2c]$. Moreover, for every x, we have that by definition of P it holds $\left|P(\phi^{\text{small}})(x) - (\phi^{\text{small}})^+(x)\right| \le \frac{\varepsilon'}{12}$. The upper bound $|\langle \mathbf{1}, P(\phi^{\text{small}}) - (\phi^{\text{small}})^+\rangle| \le \frac{\varepsilon'}{8}$ follows by triangle inequality as the inner product with $\mathbf{1}$ can be upper bounded by

$$\frac{1}{n}\sum_x \left|P(\phi^{\text{small}})(x) - (\phi^{\text{small}})^+(x)\right| \le \frac{\varepsilon'}{12} < \frac{\varepsilon'}{8}$$

as we just saw. For the inner product with $\tilde{\mu}$, observe that by 1 of Lemma 19, we have

$$|\langle \tilde{\mu} - \mathbf{1}, \mathbf{1}\rangle| = |\langle \tilde{\mu}, \mathbf{1}\rangle - \langle \mathbf{1}, \mathbf{1}\rangle| = \tfrac{1}{n}|p^{-1}|\tilde{X}| - n| < \varepsilon\,,$$

so $\tilde{\mu}$ takes the value p^{-1} on at most $(1+\varepsilon)np < \frac{3}{2}pn$ entries, and it is zero elsewhere. Thus, by these considerations, triangular inequality, and definition of P, we get that the inner product with $\tilde{\mu}$ is bounded by

$$|\langle \tilde{\mu}, P(\phi^{\text{small}}) - (\phi^{\text{small}})^+\rangle| \le \frac{1}{n}\sum_x \mathbb{1}(\tilde{\mu}(x) \ne 0)\cdot p^{-1}\cdot \frac{\varepsilon'}{12} \le \frac{3}{2}pn\cdot p^{-1}\cdot\frac{\varepsilon'}{12}\cdot n^{-1} = \frac{1}{8}\varepsilon'\,.$$

It remains to deal with (9). Here we use 2 of Lemma 19. Since every entry of ϕ^{big} is either equal to zero or has absolute value larger than $2c > 1$, we have pointwise $(\phi^{\text{big}})^+ \le \phi^2$. For the same reason, we have pointwise

$$\forall i,j \ge 0,\ (\phi^{\text{big}})^i \le (\phi^{\text{big}})^{2i} \le (2c)^{-2j}\phi^{2i+2j}\,. \tag{10}$$

In particular, for any fixed $1 \le i \le d$, let $j = \frac{1}{2}(d' - 2i)$. Then we have by 2 of Lemma 19

$$\langle \tilde{\mu}, \phi^i\rangle, \langle \mathbf{1}, \phi^i\rangle \le (2c)^{-d'+2i}(2c^{d'}) \le 2^{1-d'}(2c)^{2d} \le \frac{\varepsilon'}{8M(d+1)+8}\,, \tag{11}$$

where the final inequality is by choice of d'.

We now use the triangle inequality and (10) to write

$$|\langle \tilde{\mu}, P\big(\phi^{\text{big}}\big) - \big(\phi^{\text{big}}\big)^+\rangle| \le \sum_{i=0}^{d}|a_i|\langle \tilde{\mu}, \phi^{2i}\rangle + \langle \tilde{\mu}, \phi^2\rangle\,,$$

and from (11) we get

$$|\langle \tilde{\mu}, P\big(\phi^{\text{big}}\big) - \big(\phi^{\text{big}}\big)^+\rangle| \le M(d+1)\frac{\varepsilon'}{8M(d+1)+8} + \frac{\varepsilon'}{8M(d+1)+8} = \frac{\varepsilon'}{8}\,.$$

An identical argument replacing $\tilde{\mu}$ with $\mathbf{1}$ completes (9) and hence completes the proof. □

8. More independence, less vertices

We introduced Lemma 19 to be of use in the proof of Theorem 18. When proving Theorem 18 we start by showing that $\tilde{\mu} - \mathbf{1}$ is unlikely to correlate with any $\phi \in \Phi(\tilde{\mu}, \mathbf{1})^d$, for some large fixed d given to us by Weierstrass Approximation Theorem. Much as in Section 5, a problem we encounter when doing so is that Φ^d has too many vertices, and therefore we cannot directly apply a union bound. As in Section 5, the solution to this problem is to randomly split $\tilde{\mu}$ and $\mathbf{1}$. An additional problem that exists in this section, which was not present in Section 5, is that in order to write down a vertex ϕ of $\Phi(\tilde{\mu}, \mathbf{1})^d$ we need to know $\tilde{\mu}$. Therefore, we cannot then ask for $\tilde{\mu} - \mathbf{1}$ to be independent of ϕ, if ϕ is a vertex of $\Phi(\tilde{\mu}, \mathbf{1})^d$. It turns out that random splitting deals with this problem as well.

We now introduce a finer notation for dealing with random splitting, and then prove that anti-correlation over $\Phi(\tilde{\mu}, \mathbf{1})^d$ is implied by anti-correlation over a new polytope with fewer vertices.

Notation 20. *Let us be in Setting 2. We assume we are using Notation 13 throughout whenever needed.*

If it is given a function $\chi_{\mathbf{1}} : [n] \to \{1, \dots, \lceil Lp^{-1} \rceil\}$ —called **1**-colouring*—, we denote by ν_i (for $i \in [\lceil Lp^{-1} \rceil]$) the functional:*

$$\nu_i(x) = \begin{cases} \lceil Lp^{-1} \rceil & \text{if } \chi_{\mathbf{1}}(x) = i \\ 0 & \text{else} \end{cases} .$$

If it is given a function $\chi_\mu : [n] \to \{1, \dots, L\}$ —called μ-colouring*— and a subset X of $[n]$, we denote by μ_i (for $i \in [L]$) the functional:*

$$\mu_i(x) = \begin{cases} Lp^{-1} & \text{if } x \in X \text{ and } \chi_\mu(x) = i \\ 0 & \text{else} \end{cases} .$$

We call each pre-image $\chi_\mu^{-1}(x)$ a part *of the μ-colouring (similarly for $\chi_{\mathbf{1}}$), and we call* colours *the codomains of $\chi_{\mathbf{1}}$ and χ_μ.*

If, in addition to χ_μ, it is given $\tilde{X}$ a subset of $[n]$, we denote by $\tilde{\mu}_i$ (for $i \in [L]$) the functional:

$$\tilde{\mu}_i(x) = \begin{cases} Lp^{-1} & \text{if } x \in \tilde{X} \text{ and } \chi_\mu(x) = i \\ 0 & \text{else} \end{cases} .$$

If, in addition to χ_μ, $\chi_{\mathbf{1}}$, and $\tilde{X}$, it is given a positive integer d, we denote by Φ' the polytope $\Phi' := \Phi\big(\tilde{\mu}_1, \dots, \tilde{\mu}_L, \nu_1, \dots, \nu_{\lceil Lp^{-1} \rceil}\big)^d$.

Also, for any ϕ a vertex of Φ', let $Q_\phi \subseteq [L]$ be the minimum set of μ-colours such that we can write ϕ as a product of at most d functions which are $\{\tilde{\mu}_j : j \in Q_\phi\} \cup \{\nu_1, \dots, \nu_{\lceil Lp^{-1} \rceil}\}$-anti-uniform. We denote by $Y(\phi)$ the revealed part *of ϕ, i.e. the set $\{x \in [n] : \chi_\mu(x) \in Q_\phi\}$.*

Finally, for $\phi \in \Phi(\tilde{\mu}, \mathbf{1})^d$, we write $\phi^{\text{small}}(x) := \phi(x)\mathbb{1}(|\phi(x)| \le 2c^d)$ and $\phi^{\text{big}}(x) = \phi(x) - \phi^{\text{small}}(x)$.

We now prove that, with the above notation, Φ' contains Φ^d in the following deterministic lemma.

Lemma 21. *Let us be in Setting 2. Let $\chi_{\mathbf{1}} : [n] \to \{1, \dots, \lceil Lp^{-1} \rceil\}$ and $\chi_\mu : [n] \to \{1, \dots, L\}$ be an arbitrary* **1**- *and μ-colouring respectively, and let $\tilde{X}$ be an arbitrary subset of $[n]$. Let us use Notation 20. For all $d \ge 1$, we have the set inclusion $\Phi(\tilde{\mu}, \mathbf{1})^d \subseteq \Phi'$.*

Proof. It is enough to show that the vertices of Φ^d are in Φ'. By definition of Φ^d and Φ' (as the power of polytopes generated by anti-uniform functionals and their negatives), we can consider a vertex $\phi \in \Phi^d$ which is a product of at most d of the $\{\tilde{\mu}, \mathbf{1}\}$-anti-uniform functions. We say at most d and not exactly d because $\mathbf{1} \in \Phi$.

By definition, every vertex of Φ is either in Σ or is of the form $*_{i,\omega}(f_1, \dots, f_k)$, where $i \in [k], \omega \in \Omega$, and $f_1, \dots, f_k$ are $\{\tilde{\mu}, \mathbf{1}\}$-bounded, or is the negative of such a vertex. Therefore, we can write our vertex ϕ as

$$\phi = \prod_{j=1}^{\ell} *_{i_j,\omega_j}(f_1^{(j)}, \dots, f_k^{(j)}) \prod_{j=1}^{\ell'} \sigma_j$$

where $\ell+\ell' \le d$ and where each σ_j is in Σ. Consider a specific $j \in [\ell]$ and $j' \in [k] \setminus \{i_j\}$. Then $f_{j'}^{(j)}$ is bounded either by $\tilde{\mu}$ or by $\mathbf{1}$.

If $f_{j'}^{(j)}$ is bounded by $\tilde{\mu}$, then we can write

$$f_{j'}^{(j)} = \frac{1}{L} \sum_{j'' \in [L]} f_{j'}^{(j)} \frac{\tilde{\mu}_{j''}}{\tilde{\mu}}.$$

Where the fraction $\frac{\tilde{\mu}_{j''}}{\tilde{\mu}}$ is to be interpreted pointwise and if $\tilde{\mu}(x)=0$ (so $\tilde{\mu}_{j''}(x)=0$ too) then we define the result to be 0.

On the other hand, if $f_{j'}^{(j)}$ is bounded by $\mathbf{1}$, then we can write

$$f_{j'}^{(j)} = \frac{1}{\lceil Lp^{-1} \rceil} \sum_{j'' \in [\lceil Lp^{-1} \rceil]} f_{j'}^{(j)} \nu_{j''}.$$

Recall that $*_{i_j,\omega_j}(f_1^{(j)}, \ldots, f_k^{(j)})$ is linear in each argument. Therefore, substituting the two equations above into the definition of ϕ, and pulling the sums and coefficients $\frac{1}{L}$ and $\frac{1}{\lceil Lp^{-1} \rceil}$ out by linearity, we have written ϕ as a weighted sum of vertices of Φ'. The sum has $L^q \lceil Lp^{-1} \rceil^{q'}$ terms, where q is the number of functions bounded by $\tilde{\mu}$ and q' the number bounded by $\mathbf{1}$. Each term in the sum has the same coefficient $L^{-q} \lceil Lp^{-1} \rceil^{-q'}$, so that this weighted sum is a convex combination and we proved $\phi \in \Phi'$. □

9. The final probabilistic estimate

In this section, we finally show that, assuming some moment bounds, it is likely that $\big|\langle \tilde{\mu} - \mathbf{1}, \phi \rangle\big| < \varepsilon$ for all $\phi \in \Phi'$.

This proof looks quite similar to the corresponding statement from the proof of Theorem 15 in Section 5. As before, it is enough to prove anti-correlation for vertices of Φ'. And as before, we split the anti-correlation into anti-correlation with $\phi^{\text{small}}(x)$ and the remaining ϕ^{big}.

Much as in Section 5, we can show that $\langle \tilde{\mu} - \mathbf{1}, \phi^{\text{big}} \rangle$ is small by applying some moment bounds. However, proving $\langle \tilde{\mu} - \mathbf{1}, \phi^{\text{small}} \rangle$ is small requires some new ideas. There are two reasons for this: first, the entries of $\tilde{\mu}$ are not independent random variables, and second, in order to describe a vertex of ϕ we first need to reveal some entries of $\tilde{\mu}$.

In the case when $\tilde{X}$ is an ε-deletion of X, we have that μ and $\tilde{\mu}$ are equal in most components. Therefore, we show that it suffices to prove $\langle \mu - \mathbf{1}, \phi^{\text{small}} \rangle$ is small. We then show that this holds as this inner product is a sum of independent mean zero random variables. It turns out we do not need to reveal many entries of μ in order to describe ϕ. We give more details of why this is later, but the idea is as follows.

Given a vertex ϕ of Φ', recall that ϕ is a product of some at most d functions which are $\{\tilde{\mu}_1, \ldots, \tilde{\mu}_L, \nu_1, \ldots, \nu_{\lceil Lp^{-1} \rceil}\}$-anti-uniform. Therefore, with Notation 20, we have that $|Q_\phi| \le d(k-1)$, and hence that $Y(\phi)$ is a small subset of $[n]$. The idea is to split the inner product $\langle \mu - \mathbf{1}, \phi^{\text{small}} \rangle$ into the contribution from $Y(\phi)$, which we can bound using moment bounds, and the contribution from the remainder, which we can bound using Bernstein's inequality. We do this latter bound in the next lemma. That is, we now show how to apply Bernstein's inequality to the contributions not from $Y(\phi)$.

Lemma 22. *Let us be in Setting 2. Given d be a positive integer, and $\delta > 0$, let $C = 100c^{2d}dk\delta^{-2}$. If $L \ge 16C$ and the C-conditions are satisfied in Setting 2, then with probability* *at least* $1-\exp\big(-\frac{1}{10}\delta^2 pn\big)$ *over the uniform and independent choices of* $X=[n]_p$, *and* $\chi_\mu:[n]\to\{1,\dots,L\}$, *and* $\chi_{\mathbf{1}}:[n]\to\{1,\dots,\lceil Lp^{-1}\rceil\}$, *the following holds. For any given* $\tilde{X}\subseteq X$, *let us use Notation 20. For each vertex* ϕ *of* Φ', *we have*

$$\big|\langle \mu-\mathbf{1},\phi^{\mathrm{small}}\cdot\mathbb{1}\big([n]\setminus Y(\phi)\big)\rangle\big|<\delta\,. \tag{12}$$

Something we need to be a little careful about in the above statement is that in order to know any vertices of Φ', we need to reveal all of μ (because Φ' depends on $\tilde{\mu}$). We actually show the above bound for vertices of $\Phi(\mu_1,\dots,\mu_L,\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil})^d$, and deduce the required statement for Φ'.

Proof. By Chernoff's inequality and by doing a union bound, we get that with probability at least $1-2Lp^{-1}\exp\big(-\frac{1}{300}pn\big)$ each part of the μ-colouring of $[n]$ has size at most $\frac{1.1n}{L}$, and each part of the $\mathbf{1}$-colouring has size at most $\frac{2pn}{L}$. Suppose this likely event occurs, and reveal the μ- and $\mathbf{1}$-colourings χ_μ and $\chi_{\mathbf{1}}$.

Without revealing X, we know that every vertex ϕ of Φ' has a revealed part $Y(\phi)$ which is the union of some at most $d(k-1)$ parts of the μ-colouring. We can therefore prove the lemma by a union bound over the possible choices of Y; which is, over the choices of at most $d(k-1)$-many μ-colours of $[L]$.

Let Q be a set of at most $d(k-1)$ colours in $[L]$, and let Y be the union of the parts of the μ-colouring that are mapped to Q. We can now consider the random variable $X\cap Y$ (where Y is fixed and X needs to be sampled). By Chernoff's inequality and by doing a union bound, we obtain that with probability at least $1-|Q|\exp\big(-\frac{1}{8}pn\big)$, for each $q\in Q$ the number of elements of X with μ-colour q is at most $\frac{2pn}{L}$. Suppose that this likely event occurs, and reveal $X\cap Y$. We now can define the set $\Psi(Q)$ of $\{\mu_j : j\in Q\}\cup\{\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil}\}$-extreme anti-uniform functions. For this proof only, let us denote with $H_Q=\{\mu_j:j\in Q\}\cup\{\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil}\}$.

We can upper bound $|\Psi(Q)|$ as follows. First consider that by definition every vertex of $\Psi(Q)$ is either in Σ, or of the form $*_{i,\omega}(f_1,\dots,f_k)$ where $f_1,\dots,f_k$ are H_Q-extreme. By definition, f_j is H_Q-extreme if there is $h_j\in H_Q$ such that for every $x\in[n]$ we have either $f(x)=0$ or $f(x)=h(x)$. Therefore, to upper bound $|\Psi(Q)|$ we can consider that every H-extreme anti-uniform function can be in Σ, or of the form $*_{i,\omega}(f_1,\dots,f_k)$ obtained as follows. We first select $\omega\in\Omega$ and a sequence of $k-1$ bounding functions $h_1,\dots,h_k$ from H; we then choose for each bounding function h_j, from the at most $\frac{2pn}{L}$ non-zero entries, the non-zero entries of f_j (which by definition of 'extreme' are equal to the corresponding entries of h_j). The total number of elements of $\Psi(Q)$ is therefore is at most

$$|\Sigma|+|\Omega|(2Lp^{-1})^{k-1}\cdot 2^{\frac{2pn}{L}(k-1)}\,.$$

We now select a function ϕ which is a product of at most d elements of $\Psi(Q)$. The number of possible choices for ϕ is at most

$$d\Big(|\Omega|(2Lp^{-1})^{k-1}\cdot 2^{\frac{2pn}{L}(k-1)}+|\Sigma|\Big)^d\le d(2Lp^{-1})^{d(k-1)}2^{\delta^2pn/16}2^{\frac{4pn}{L}d(k-1)}\,.$$

By definition, the entries of $\phi^{\mathrm{small}}\cdot\mathbb{1}\big([n]\setminus Y(\phi)\big)$ are in $[-2c^d,2c^d]$, and only the entries outside $Y(\phi)$ can be non-zero. Thus, the quantity

$$\langle\mu-1,\phi^{\mathrm{small}}\cdot\mathbb{1}\big([n]\setminus Y(\phi)\big)\rangle$$

is a sum of $n-|Y(\phi)|\le n$ independent random variables

$$\tfrac{1}{n}\big(\mathbb{1}(x\in X)p^{-1}-1\big)c_x$$

where the number $c_x = \phi^{\mathrm{small}}(x) \cdot \mathbb{1}\big(x \in [n] \setminus Y(\phi)\big)$ is in $[-2c^d, 2c^d]$. Since the probability of $x \in X$ is p, these random variables all have mean zero, and are bounded between $\frac{-2c^d}{n}$ and $\frac{2c^d}{n}p^{-1}$. It remains to calculate the variance. We have

$$\mathrm{Var}(\mathbb{1}(x \in X)p^{-1} - 1) = p(p^{-1}-1)^2 + (1-p)(-1)^2 = p^{-1} - 1 \le p^{-1}\,,$$

so that the variance of each of our random variables is at most $\frac{4c^{2d}}{n^2}p^{-1}$.

By Bernstein's inequality (Lemma 6), the probability that when we reveal $X \setminus Y(\phi)$ we get

$$\big|\langle \mu - 1, \phi^{\mathrm{small}} \cdot \mathbb{1}\big([n] \setminus Y(\phi)\big)\rangle\big| \ge \delta$$

is at most

$$2\exp\Big(-\frac{\delta^2/2}{2p^{-1}\delta/(3n) + n\frac{4c^{2d}}{n^2}p^{-1}}\Big) \le 2\exp\big(-\tfrac{1}{16c^{2d}}\delta^2 pn\big)\,.$$

Taking a union bound over the choices of ϕ, the probability that there exists any product ϕ of at most d elements of $\Psi(Q)$ with

$$\big|\langle \mu - 1, \phi^{\mathrm{small}} \cdot \mathbb{1}\big([n] \setminus Y(\phi)\big)\rangle\big| \ge \delta$$

is at most

$$d(2Lp^{-1})^{d(k-1)} \cdot 2^{\frac{4pn}{L}d(k-1) + \delta^2 pn/16} \cdot 2\exp\big(-\tfrac{1}{16c^{2d}}\delta^2 pn\big) + d(k-1)\exp(-\tfrac{1}{8}pn)\,.$$

Finally, taking a union bound over the choices of Q, the probability that there exists Q and a product ϕ of at most d elements of $\Psi(Q)$ such that

$$\big|\langle \mu - 1, \phi^{\mathrm{small}} \cdot \mathbb{1}\big([n] \setminus Y(\phi)\big)\rangle\big| \ge \delta$$

is at most

$$2^L d(2Lp^{-1})^{d(k-1)} \cdot 2^{\frac{4pn}{L}d(k-1)} 2^{\delta^2 pn/16} \cdot 4\exp\big(-\tfrac{1}{16c^{2d}}\delta^2 pn\big) \le \exp\big(-\tfrac{1}{100c^2}\delta^2 pn\big)\,,$$

where the final inequality is by choice of L and since $pn \ge 100c^{2d}\delta^{-2}dk\log n$.

Suppose now that X is such that this unlikely event does not occur. Given $\tilde{X}$, we can now calculate the polytope Φ'. Let ϕ be a vertex of this polytope: then ϕ is a product of at most d extreme restricted anti-uniform functions. Letting Q be the set of μ-colours bounding ϕ, we see ϕ is a product of at most d members of $\Psi(Q)$, because for each j the function $\tilde{\mu}_j$ is pointwise either equal to μ_j or equal to zero. The lemma statement follows. □

There is a last anti-correlation lemma we need. But before introducing that, we state a moment bound lemma (Lemma 23) which is needed in its proof. The proof of Lemma 23 is left for a later section.

Lemma 23. *Given $\delta > 0$, and d' an even positive integer, there exists L_0 such that, if $L \ge L_0$, then there exist C such that, if the C-conditions are satisfied in Setting 2, then with probability at least $1 - 3\exp(-\frac{1}{8}\delta^2 pn)$ over the choice of $X = [n]_p$, the following happens. With probability at least 0.9 over the choice of $\chi_\mu : [n] \to [L]$ and $\chi_{\mathbf{1}} : [n] \to [\lceil Lp^{-1}\rceil]$ independent and uniform at random, there is a δ-deletion $\tilde{X}$ of X such that the following happens. Let us use Notation 20. For any $1 \le \ell \le d'$ and ψ an largest anti-uniform functional in either $\Phi(\tilde{\mu}, \mathbf{1})^\ell$ or $\Phi(\tilde{\mu}_1, \dots, \tilde{\mu}_L, \nu_1, \dots, \nu_{\lceil Lp^{-1}\rceil})^\ell$,*

$$\langle \tilde{\mu}, \psi\rangle \le 2c^\ell \qquad \textit{and} \qquad \langle \mathbf{1}, \psi\rangle \le 2c^\ell\,.$$

In addition, if ψ is any largest anti-uniform functional in $\Phi(\tilde{\mu}_1,\dots,\tilde{\mu}_L,\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil})^\ell$, and $1\le j\le L$ and $1\le j'\le\lceil Lp^{-1}\rceil$ then we have

$$\langle\tilde{\mu}_j,\psi\rangle\le 2c^\ell\qquad\text{and}\qquad\langle\nu_{j'},\psi\rangle\le 2c^\ell\,.$$

We are now in a position to state and prove the final anti-correlation lemma we need: Lemma 24. The main probabilistic inputs to this lemma are the above Lemma 22 and the moment bounds Lemma 23, which we prove in a following section.

Lemma 24. *Let us be in Setting 2. Given d,d' positive integers with d' even, given $\varepsilon>0$, there exist L_0 such that, if $L\ge L_0$, there exists C such that, if the C-conditions are satisfied in Setting 2, then with probability at least $1-\exp\big(-\frac{pn}{C}\big)$ over the choice of $X=[n]_p$, there is an ε-deletion $\tilde{X}$ of X such that the following happens. There exist functions $\chi_\mu:[n]\to[L]$ and $\chi_{\mathbf{1}}:[n]\to\lceil Lp^{-1}\rceil$ such that, using Notation 20, we have $\langle\tilde{\mu}-\mathbf{1},\phi\rangle<\varepsilon$ for all $\phi\in\Phi(\tilde{\mu}_1,\dots,\tilde{\mu}_L,\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil})^d$. In addition, for all $\phi\in\Phi(\tilde{\mu},\mathbf{1})$, we have $|\langle\tilde{\mu},\phi\rangle|,|\langle\mathbf{1},\phi\rangle|\le 2c$ and $\langle\tilde{\mu},\phi^{d'}\rangle,\langle\mathbf{1},\phi^{d'}\rangle\le 2c^{d'}$.*

Proof. In this proof, let us use the notation $\tilde{H}=\{\tilde{\mu}_1,\dots,\tilde{\mu}_L,\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil}\}$ and $H=\{\mu_1,\dots,\mu_L,\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil}\}$.

Given $d,d',\varepsilon>0$ with d' even, we set $\delta=\frac{\varepsilon}{12c^d}$ and $d''=\max(d',d\big(1+\lceil\log_2\frac{2c^d}{\delta}\rceil\big)$. Let $L_0=1600c^{2d}dk\delta^{-2}$, which guarantees that if $L\ge L_0$, then it satisfies the conditions for Lemma 22 with input δ. Let C be large enough for Lemma 23 and Lemma 22. Without loss of generality we assume $C\ge 16d\delta^{-2}$ and that the C-conditions are satisfied.

Chernoff's inequality tells us that with probability at least $1-\exp\big(-\frac13 pn\big)$, the set $X=[n]_p$ has at most $2pn$ elements. Moreover, Lemma 22, with input δ, tells us that with probability (over the product probability space of $[n]_p$ and the μ- and $\mathbf{1}$-colourings) at least $1-\exp\big(\frac{1}{10}\delta^2pn\big)$ we have, for each vertex $\phi\in\Phi'$, the following inequality holds:

$$\big|\langle\mu-\mathbf{1},\phi^{\text{small}}\cdot\mathbb{1}\big([n]\setminus Y(\phi)\big)\rangle\big|<\delta\,. \tag{13}$$

In particular, with probability at least $1-\exp\big(\frac{1}{20}\delta^2pn\big)$ over $[n]_p$, the probability of the μ- and $\mathbf{1}$-colourings having this property is at least 0.9.

In addition, because of the conditions on L and d'' and C, Lemma 23 tells us that with probability at least $1-\exp\big(-\frac18\delta^2pn\big)$ (over the choice of $[n]_p$), the set $X=[n]_p$ has the following property. There exists a δ-deletion $\tilde{X}$ of X such that we have, with probability at least 0.9 (over the random choice of χ_μ and $\chi_{\mathbf{1}}$) that for any $1\le\ell\le d''$ and ψ' a $\tilde{H}$-largest anti-uniform functional in $\Phi(\tilde{H})^\ell$, and $1\le j\le L$, it holds

$$\langle\tilde{\mu}_j,\psi'\rangle,\langle\tilde{\mu},\psi'\rangle,\langle\mathbf{1},\psi'\rangle\le 2c^\ell\,. \tag{14}$$

Suppose now that X is such that all three likely events occur, which by the union bound has probability at least

$$1-\exp\big(-\tfrac13 pn\big)-\exp\big(\tfrac{1}{20}\delta^2pn\big)-\exp\big(\tfrac18\delta^2pn\big)\ge 1-\exp\big(\tfrac{1}{30}\delta^2pn\big)\,.$$

Fix $\tilde{X}$ a δ-deletion witnessing the likely event occurring. The probability that the μ- and $\mathbf{1}$-colourings are such that their likely events occur is by the union bound at least 0.8. Suppose this likely event occurs: this gives us that there exist χ_μ and $\chi_{\mathbf{1}}$ as in the lemma statement.

We next establish the anti-correlation claimed in the lemma.

Claim 25. *For each $\tilde{H}$-largest anti-uniform functional $\psi \in \Phi'$ and each $j \in [L]$, we have*

$$\langle \tilde{\mu}_j, \psi \rangle \leq 2c^d \,, \tag{15}$$

$$\langle \tilde{\mu}, \psi^{\mathrm{big}} \rangle \leq \delta \,, \tag{16}$$

$$\langle \mathbf{1}, \psi^{\mathrm{big}} \rangle \leq \delta \,. \tag{17}$$

Proof. Equation (15) is immediate from (14) taking $\psi' = \psi$ with $\ell \leq d$ and using $c \geq 1$.

For the remaining two equations, choose ℓ minimal such that ψ is a $\tilde{H}$-largest anti-uniform functional in $\Phi(\tilde{H})^\ell$, and note $\ell \leq d$. Let $a = \lceil \log_2 \frac{2c^d}{\delta} \rceil$, and note $(1+a)\ell \leq (1+a)d \leq d'$.

For (16), observe that by definition of $\tilde{H}$-largest anti-uniform functional (with this specific $\tilde{H}$), if $\psi \neq 0$ for some x then $\psi(x) > 2c^d$. It follows that

$$\langle \tilde{\mu}, \psi^{\mathrm{big}} \rangle \cdot (2c^d)^a \leq \langle \tilde{\mu}, (\psi^{\mathrm{big}})^{1+a} \rangle \leq \langle \tilde{\mu}, \psi^{1+a} \rangle \leq 2c^{(1+a)\ell} \,,$$

where the final inequality is by (14) with $\psi' = \psi^{1+a}$. By choice of a, we have the upper bound $2c^{(1+a)\ell}(2c^d)^{-a} \leq \delta$, giving (16). Swapping $\mathbf{1}$ for $\tilde{\mu}$ in the above calculation establishes (17). □

By Lemma 9, the maximum $\max_{\phi \in \Phi'} \big|\langle \tilde{\mu} - \mathbf{1}, \phi \rangle\big|$ is attained in one of the vertices of Φ'. By central symmetry in the definition of Φ' and linearity of the inner product, the maximum value of $\big|\langle \tilde{\mu} - \mathbf{1}, \phi \rangle\big|$ over Φ' is also an extremal value of $|\langle \tilde{\mu} - \mathbf{1}, \phi \rangle|$ over the vertices of Φ' which are products of d restricted anti-uniform functions (and not their opposites).

Let us therefore fix such a vertex ϕ in Φ', and let $Y = Y(\phi)$. Our goal is to show that $|\langle \tilde{\mu} - \mathbf{1}, \phi \rangle| < \varepsilon$. We use linearity of the inner product and the triangle inequality to split this up. Write $\tilde{\mu}' = \tilde{\mu}\mathbb{1}([n] \setminus Y)$ and $\tilde{\mu}'' = \tilde{\mu}\mathbb{1}(Y)$; define similarly μ', μ'' and $\mathbf{1}'$ and $\mathbf{1}''$. We obtain

$$\big|\langle \tilde{\mu} - \mathbf{1}, \phi \rangle\big| \leq \big|\langle \tilde{\mu}' - \mathbf{1}', \phi \rangle\big| + \big|\langle \tilde{\mu}'' - \mathbf{1}'', \phi \rangle\big| \,.$$

We can further split the first term

$$\big|\langle \tilde{\mu}' - \mathbf{1}', \phi \rangle\big| \leq \big|\langle \mu' - \mathbf{1}', \phi^{\mathrm{small}} \rangle\big| + \big|\langle \mu' - \tilde{\mu}', \phi^{\mathrm{small}} \rangle\big| + \big|\langle \tilde{\mu}' - \mathbf{1}', \phi^{\mathrm{big}} \rangle\big| \,.$$

Of these terms, (13) tells us that the first term is bounded by δ. Since $\tilde{\mu}$ and μ differ in at most δpn places, $\mu' - \tilde{\mu}'$ is equal to p^{-1} in at most δpn places and otherwise equal to zero, while $|\phi^{\mathrm{small}}|$ is bounded by $2c^d$, so the second term is at most $\frac{1}{n} \cdot p^{-1} \cdot \delta pn \cdot 2c^d = 2c^d\delta$. Splitting the third term

$$\big|\langle \tilde{\mu}' - \mathbf{1}', \phi^{\mathrm{big}} \rangle\big| \leq \langle \tilde{\mu}', \phi^{\mathrm{big}} \rangle + \langle \mathbf{1}', \phi^{\mathrm{big}} \rangle \leq \langle \tilde{\mu}, \phi^{\mathrm{big}} \rangle + \langle \mathbf{1}, \phi^{\mathrm{big}} \rangle \,,$$

where in the final two inner products, all terms are non-negative.

Returning to split

$$\big|\langle \tilde{\mu}'' - \mathbf{1}'', \phi \rangle\big| \leq \langle \tilde{\mu}'', \phi \rangle + \langle \mathbf{1}'', \phi \rangle \,,$$

again all the terms in the inner products are non-negative. In particular, if ψ is any function which is pointwise greater than or equal to ϕ, replacing ϕ with ψ gives an upper bound on all these non-negative inner products. Let ψ then be a largest restricted anti-uniform function which is pointwise at least ϕ. By (16), (17), we have $\langle \tilde{\mu}, \psi^{\mathrm{big}} \rangle, \langle \mathbf{1}, \psi^{\mathrm{big}} \rangle < \delta$.

We apply (15) to obtain $\langle \tilde{\mu}_j, \psi \rangle \leq 2c^d$ where $\tilde{\mu}_j$ is revealed by ϕ, that is, $\chi_\mu^{-1}(j) \subseteq Y$. Recall that the normalisation of $\tilde{\mu}_j$ is $p^{-1}L$, so that $\tilde{\mu}'' = \frac{1}{L}\sum_j \tilde{\mu}_j$, where the sum ranges over j with $\chi_\mu^{-1}(j) \subseteq Y$. This gives

$$\langle \tilde{\mu}', \phi \rangle \leq \langle \tilde{\mu}'', \psi \rangle = \frac{1}{L}\sum_j \langle \tilde{\mu}_j, \psi \rangle \leq \frac{2c^d d(k-1)}{L} \,.$$

Finally, we come to $\langle \mathbf{1}'', \psi\rangle$. Here we split $\psi = \psi^{\text{small}} + \psi^{\text{big}}$, and write

$$\langle \mathbf{1}'', \psi\rangle = \langle \mathbf{1}'', \psi^{\text{small}}\rangle + \langle \mathbf{1}'', \psi^{\text{big}}\rangle \le \langle \mathbf{1}'', \psi^{\text{small}}\rangle + \langle \mathbf{1}, \psi^{\text{big}}\rangle .$$

To deal with the first term of this, observe that $\mathbf{1}''$ takes the value 1 in at most $\frac{2d(k-1)n}{L}$ places, and zero elsewhere, while ψ^{small} is bounded by $2c^d$, so that the first inner product is at most $\frac{1}{n} \cdot \frac{2d(k-1)n}{L} \cdot 2c^d = \frac{4c^d d(k-1)}{L}$. The second inner product is one we have already bounded, using (17), by δ.

Putting the pieces together, we have

$$\big|\langle \tilde{\mu} - \mathbf{1}, \phi\rangle\big| \le \delta + 2c^d\delta + \delta + \delta + \tfrac{2c^d d(k-1)}{L} + \tfrac{4c^d d(k-1)}{L} + \delta \le \varepsilon ,$$

as required.

Finally, we need to prove the moment bounds required in the lemma. By Lemma 21, we have $\Phi(\tilde{\mu}, \mathbf{1}) \subseteq \Phi(\mu_1, \dots, \mu_L, \nu_1, \dots, \nu_{\lceil Lp^{-1}\rceil})$, so it suffices to prove the required moment bounds hold for all ϕ in the latter polytope.

Consider the optimisation problem $\max_\phi \langle \tilde{\mu}, \phi^{d'}\rangle$, over $\phi \in \Phi(H)$. By Fact 9, the maximum is attained at a vertex of $\Phi(H)$. Since $\tilde{\mu}$ is a non-negative vector, the vertex in question is a H-anti-uniform function ψ (and not a negation). If $\psi \in \Sigma$, then since $0 \le \psi \le 1$ we have $\langle \tilde{\mu}, \psi\rangle \le \langle \tilde{\mu}, \mathbf{1}\rangle \le 2$ since X has at most $2pn$ elements, which is sufficient. So we may assume ψ is not in Σ. Again since $\tilde{\mu}$ is non-negative, we may assume this anti-uniform function is pointwise maximised, in other words it is an H-largest anti-uniform functional in $\Phi(H)$, and therefore $\psi^{d'}$ is an H-largest anti-uniform functional in $\Phi(H)^{d'}$. Applying (14), we have $\langle \tilde{\mu}, \psi^{d'}\rangle \le 2c^{d'}$ as required.

A similar argument applies to the optimisation problem $\max_\Phi |\langle \tilde{\mu}, \phi\rangle|$. Since Φ is centrally symmetric, the maximum is the same as for the linear problem $\max_\Phi \langle \tilde{\mu}, \phi\rangle$; as above, this is attained for ϕ an H-largest anti-uniform functional in $\Phi(H)$, and (14) gives $\langle \tilde{\mu}, \psi\rangle \le 2c$ for such functionals.

The same argument, replacing $\tilde{\mu}$ with $\mathbf{1}$, gives the other required moment bounds. ☐

Finally, we are in a position to prove Theorem 18: at this stage, this simply amounts to putting together the lemmas we showed in the last two sections.

Proof of Theorem 18. Given $\varepsilon > 0$, let $\varepsilon_1 > 0$ and d, d' be returned by Lemma 19 for input ε. Without loss of generality, we may assume $\varepsilon_1 \le \varepsilon$. Note that d' is guaranteed to be even. We now input d, d', and ε_1 to Lemma 24, which returns L_0, and, provided $L \ge L_0$ also C.

Now, assume that our setting satisfies the C-conditions. In particular, the conditions of Lemma 24 are satisfied, so with probability at least $1 - \exp\left(-\frac{pn}{C}\right)$, the set $X = [n]_p$ has an ε_1-deletion $\tilde{X}$ such that there exist $\chi_\mu, \chi_{\mathbf{1}}$ for which the following hold, with Notation 20. For all $\phi \in \Phi'$, we have $\langle \tilde{\mu} - \mathbf{1}, \phi\rangle < \varepsilon_1$, and in addition for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$ we have $|\langle \tilde{\mu}, \phi\rangle|, |\langle \mathbf{1}, \phi\rangle| \le 2c$ and $\langle \tilde{\mu}, \phi^{d'}\rangle, \langle \mathbf{1}, \phi^{d'}\rangle \le 2c^{d'}$. Suppose X satisfies the likely event, and fix $\tilde{X}$ and $\chi_\mu, \chi_{\mathbf{1}}$ witnessing this.

The inequalities $|\langle \tilde{\mu}, \phi\rangle|, |\langle \mathbf{1}, \phi\rangle| \le 2c$ for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$ are as required for Theorem 18, while the inequalities $\langle \tilde{\mu}, \phi^{d'}\rangle, \langle \mathbf{1}, \phi^{d'}\rangle \le 2c^{d'}$ for $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$ verify 2 of Lemma 19.

Applying Lemma 21, we have $\Phi(\tilde{\mu}, \mathbf{1})^d \subseteq \Phi'$, so in particular we obtain $\langle \tilde{\mu} - \mathbf{1}, \phi\rangle < \varepsilon_1$ for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})^d$. This verifies 1 of Lemma 19, and hence we obtain the conclusion that $|\langle \tilde{\mu} - \mathbf{1}, \phi^+\rangle| < \varepsilon$ for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$. In addition, since $\Phi(\tilde{\mu}, \mathbf{1}) \subseteq \Phi(\tilde{\mu}, \mathbf{1})^d$, we have $\langle \tilde{\mu} - \mathbf{1}, \phi\rangle < \varepsilon_1 \le \varepsilon$ for all $\phi \in \Phi(\tilde{\mu}, \mathbf{1})$, which is the same as $\|\tilde{\mu} - \mathbf{1}\|_{\Phi(\tilde{\mu}, \mathbf{1})} < \varepsilon$, completing the proof of Theorem 18. ☐

It remains to prove Lemma 23.

10. Moment estimates

Ultimately, as per the reduction in Theorem 18, we seek that our random subset X contains, outside of an event with exponentially small probability, a large subset $\tilde{X}$ whose corresponding functional $\tilde{\mu}$ satisfies certain anti-correlation and moment bound properties with the functions in the polytope Φ'. In this section we show that these properties hold with a reasonably high probability for X itself; we use this to prove that a subset with these properties $\tilde{X} \subseteq X$ exists with the required exponential probability in the next section.

To state the precise lemma, we need the following definition.

Definition ((q,d)-special product)**.** *Let us be in Setting 2. A (q,d)-special product is a random functional $\psi : [n] \to \mathbb{R}$ obtained as the product of at most d convolution functions $*_{i,\mathbf{1}}(f_1, \dots, f_k)$, in which each of the f_j is either equal to the $\mathbf{1}$ function, or it is a scaled copy of the random set $[n]_q$ (having entries valued 0 or q^{-1}). Moreover the copies of $[n]_q$ in the product comprising ψ have the property that each is either identical to, or completely independent from, any of the other copies of $[n]_q$ used in ψ.*

The technical lemma we require is as follows.

Lemma 26. *Let us be in Setting 2. Given $d' \in \mathbb{N}$ and $\alpha \geq 0$, there exists a C such that the following holds if the C-conditions are satisfied. Let q be at least $C \log^{2k} n^{-1}$. Then with probability at least $1 - \frac{1}{n^{\alpha k}}$ over the sample of a (q,d')-special product ψ, and over the choice of X as a copy of $[n]_q$ which is either identical to a copy of $[n]_q$ in ψ, or completely independent from all copies, we have the following:*

$$\langle \mu, \psi \rangle \leq 2c^{d'} \qquad \textit{and} \qquad \langle \mathbf{1}, \psi \rangle \leq 2c^{d'} .$$

Proof. Let $C = (1+\alpha)^{2k} d'^{d'+1} k^{3k+d'} 2^{kd'^2+8k}$. As ψ is a (q,d')-special product we have for some $d \leq d'$ that $\psi(x) = \prod_{j \in [d]} *_{i_j,\mathbf{1}}(f_1^{(j)}, \dots, f_k^{(j)})$ for some $i_1, \dots, i_d \in [k]$ and some $f_\ell^{(j)}$ that are either equal to the $\mathbf{1}$ function, or to a scaled copy of independent random sets $[n]_q$ (with possibility of two functionals being the same, but all different samples taken independently). For such i_j and $f_\ell^{(j)}$ we can therefore write explicitly

$$\psi(x) = \Big(\frac{n}{e(s)}\Big)^d \prod_{j=1}^{d} \sum_{s \in S_{i_j}(x)} \prod_{\ell \neq i_j} f_\ell^{(j)}(s_\ell) .$$

It is helpful to refer to each of the terms in this summation individually. To this end we use the following notation

$$\hat{\psi}(x; s^{(1)}, \dots, s^{(d)}) = \prod_{j=1}^{d} \prod_{\ell \neq i_j} f_\ell^{(j)}(s_\ell^{(j)}) .$$

We require that with high probability $\langle \mathbf{1}, \psi \rangle \leq 2c^{d'}$ and $\langle \mu, \psi \rangle \leq 2c^{d'}$. Since we may assume $c \geq 1$, it is enough to show that $\langle \mathbf{1}, \psi \rangle \leq 2c^{d}$ and $\langle \mu, \psi \rangle \leq 2c^{d}$. To do so, we prove the concentration of the following polynomials around their expectations. We have:

$$Y_{\mathbf{1}} = \langle \mathbf{1}, \psi \rangle = \frac{1}{n} \left(\frac{n}{e(S)}\right)^d \sum_{x \in [n]} \sum_{\substack{s^{(1)} \in S_{i_1}(x) \\ \cdots \\ s^{(d)} \in S_{i_d}(x)}} \hat{\psi}(x; s^{(1)}, \dots, s^{(d)})$$

$$Y_\mu = \langle \mu, \psi \rangle = \frac{1}{n}\left(\frac{n}{e(S)}\right)^d \sum_{x\in[n]} \sum_{\substack{s^{(1)}\in S_{i_1}(x) \\ \cdots \\ s^{(d)}\in S_{i_d}(x)}} \mu(x)\hat{\psi}(x; s^{(1)}, \dots, s^{(d)})\,.$$

It is extremely important to notice that all of these terms make use of the same set of functionals $f_\ell^{(j)}$ (evaluated in different points of different edges). Thus, the difference between terms is not given by a difference in functionals, which are always the same, but a different in indices and hyperedges. This justifies the following notation: we denote by l the number of the $d(k-1)$ functions $f_\ell^{(j)}$ comprising ψ (and thus each of the $\hat{\psi}$) that are copies of $[n]_q$, and $l' = d(k-1) - l$ are copies of $\mathbf{1}$. Moreover, recalling that any copies of $[n]_q$ must be either identical or completely independent from each other, we denote by $w \le l$ the number of independent copies of $[n]_q$ in ψ. As mentioned above, it is important to keep in mind that l, l' and w hold term-by term, as the functionals do not change in between terms.

Our plan now is to first calculate the expectation of $Y_{\mathbf{1}}$ and Y_μ. We then use Kim-Vu's inequality (Theorem 8) to prove the concentration. This result applies since we may form new polynomials $\tilde{Y}_1, \tilde{Y}_\mu$, having the same value as $Y_{\mathbf{1}}, Y_\mu$, but consisting of independent Bernoulli random variables (as required by Theorem 8) by factoring out q^{-1} from each $\{0, q^{-1}\}$ valued Bernoulli variable into a collective weight, and dropping any repeat copies of the now $\{0,1\}$ valued Bernoulli variables within a configuration, we obtain the polynomials $\tilde{Y}_1, \tilde{Y}_\mu$ as required. The details are as follows.

Observe that each term in $Y_{\mathbf{1}}$ and in Y_μ corresponds to a tuple $(s^{(1)}, \dots, s^{(d)})$ of d hyperedges of S for which the $i_1, \dots, i_d$-th vertices within the corresponding hyperedge are the same element $x \in [n]$. We thus refer to the terms within these polynomials as (linked hyper-edge) configurations. Each configuration is completely determined by $(s^{(1)}, \dots, s^{(d)})$.

Notice that each of the $f_j^{(r)}(s_j^{(r)})$ with $r \in [d], j \in [k] \setminus \{i_r\}$, is a random variable taking a value of 1 if $f^{(r)} = \mathbf{1}$ is the constant one function, or else q^{-1} with a probability q, and 0 with probability $1-q$, if $f^{(r)}$ is a copy of $[n]_q$. Since ψ is a (q, d')-special product, any two of the random variables, $f_{j_1}^{(i_1)}(s_{j_1}^{(i_1)})$ and $f_{j_2}^{(i_2)}(s_{j_2}^{(i_2)})$, are dependent if and only if $f_{j_1}^{(i_1)}, f_{j_2}^{(i_2)}$ are the same copy of $[n]_q$ (if one of them is $\mathbf{1}$, or if they're two distinct copies of $[n]_q$, they'd be independent) and $s_{j_1}^{(i_1)} = s_{j_2}^{(i_2)}$ (every entry of $[n]_q$ is selected independently). In this case, they are identical. The second condition $s_{j_1}^{(i_1)} = s_{j_2}^{(i_2)}$, corresponds to the hyperedges $s^{(i_1)}, s^{(i_2)}$ overlapping on their j_1-th and j_2-th elements respectively. We again point out that the number of independent variables in each configuration is not given by any choice of functionals (which are always the same), but rather by how much the corresponding hyperedges of the configuration intersect one another (thus possibly allowing for two identical functionals to be evaluated at the same value).

We first calculate the quantities $\mathbb{E}(Y_{\mathbf{1}}), \mathbb{E}(Y_\mu)$ whose concentration we wish to establish. Let us consider the polynomial $Y_{\mathbf{1}}$. We want to calculate $\mathbb{E}(Y_{\mathbf{1}})$ applying the linearity of expectation and summing the contribution from each edge configuration. As mentioned above, every term of $Y_{\mathbf{1}}$ makes use of always the same functions $f_\ell^{(j)}$ which never change. Thus, in $Y_{\mathbf{1}}$, the number of independent variables $f_j^{(r)}(s_j^{(r)})$ within a configuration is at least $\max(k -$

$1, l'+w)$ and at most $d(k-1)$. The lower bound $k-1$ holds since each hyperedge $s^{(r)}$ contains k distinct elements of $[n]$, so for any fixed $r \in [d]$, the set $\{f_j^{(r)}(s_j^{(r)})\}, j \in [k] \setminus \{i_u\}$ is a set of mutually independent (possibly constant) random variables. The lower bound $l'+w$, holds because, as mentioned above, each term comprises of $l'+w$ independent random functionals (evaluated at some of their points, possibly the same). That is, if $\{f^{(u_1)}, \dots, f^{(u_{w+l'})}\}$ is a maximal independent set of functionals —containing w independent copies of $[n]_q$ and l' copies of constant $\mathbf{1}$— for ψ (and thus for the specific term we are considering) then any set of random variables formed taking one argument from each of these, is a set of mutually independent variables. Note for $\mathbb{E}(Y_\mu)$ the corresponding bounds are $\max(k-1, l'+w)+1$ and $d(k-1)+1$ since $\mu(x)$ contributes one variable independent from all the others.

For $\mathbb{E}(Y_{\mathbf{1}})$, suppose that, in a given configuration the edges are such that t of the variables are mutually independent with $\max(k-1, l'+w) \le t \le d(k-1)$. The number of repeat $\{0, q^{-1}\}$-valued Bernoulli variables is then $d(k-1)-t$, each of which contributes an extra factor of q^{-1} to the expectation of this configuration. To see this, consider the product of s such identical variables $Z = x_1 \dots x_s$. We have that Z takes value q^{-s} with probability q and 0 otherwise, so $\mathbb{E}(Z) = q^{-(s-1)}$, whereas the expectation of each of the x_i is 1.

To calculate the expectation of $Y_{\mathbf{1}}$, it is therefore enough to enumerate the configurations having the same number t of independent variables. To this end, let $S(\psi, t, x) \subseteq S_{i_1}(x) \times S_{i_2}(x) \times \dots \times S_{i_d}(x)$ be those d-tuples $(s^{(1)}, s^{(2)}, \dots, s^{(d)})$ in which exactly t of the $\{f_j^{(i)}(s_j^{(i)})\}$ are mutually independent (counting also those for which $f_j^{(i)}$ is the constant functional $\mathbf{1}$), and let $\alpha(\psi, t, x) = |S(\psi, t, x)|$. For the sake of notational brevity, let $t' = d(k-1)$ and $a = \max(k-1, l'+w)$. We have

$$\mathbb{E}(Y_{\mathbf{1}}) = \frac{1}{n}\Big(\frac{n}{e(S)}\Big)^d \sum_{x \in [n]} \sum_{a \le t \le t'} (q^{-1})^{(t'-t)} \alpha(\psi, t, x)\,.$$

Note that $\mathbb{E}(Y_{\mathbf{1}}) = \mathbb{E}(Y_\mu)$ as $\mu(x)$ is independent from all other variables within a given configuration, since the k elements in each edge s of S are distinct; therefore the argument of μ does not occur as the argument of any other $f_j^{(i)}$ within this configuration, and for any $x \in [n]$ we have that $\mu(x)$ contributes an expectation factor of $\mathbb{E}(\mu(x)) = 1$.

To obtain an upper bound for $\alpha(\psi, t, x)$ with $a \le t \le t'$, first note that when $t = t' = d(k-1)$, we may take the crude upper bound fixing only x in each $S_i(x)$, thus $\alpha(\psi, d(k-1), x) \le \prod_{j \in [d]} |S_{i_j}(x)| \le (\Delta_1)^d \le \Big(c\frac{e(S)}{n}\Big)^d$. If the number of independent variables t is less than $d(k-1)$ in a configuration, we have at least two random variables that are identical, say $f_{j_1}^{(i_1)}(s_{j_1}^{(i_1)}) = f_{j_2}^{(i_2)}(s_{j_2}^{(i_2)})$. Note that this can only occur if $s_{j_1}^{(i_1)} = s_{j_2}^{(i_2)}$. Thus, in order to have precisely t independent variables in the configuration $(s^{(1)}, s^{(2)}, \dots, s^{(d)}) \in S(\psi, t, x)$, it must be that the union of underlying hyperedges, $\cup_{i \in [d]} s^{(i)}$, together covers at most t vertices other than x. But this calculation is exactly what is given by Lemma 16 setting $\mathbf{i} = (i_1, \dots, i_d)$ and forgiving the horrible notation $\alpha(\psi, t, x) \le \alpha(\mathbf{i}, t, x)$ (which is only needed here). We therefore obtain:

$$\begin{aligned}
\mathbb{E}(Y_{\mathbf{1}}) &= \Big(\frac{n}{e(S)}\Big)^d \frac{1}{n} \sum_{x\in[n]} \sum_{a\le t\le t'} (q^{-1})^{(t'-t)} \alpha(\psi,t,x) \\
&\le \Big(\frac{n}{e(S)}\Big)^d \frac{1}{n} \sum_{x\in[n]} \alpha(\mathbf{i},t',x) + \Big(\frac{n}{e(S)}\Big)^d \frac{1}{n} \sum_{x\in[n]} \sum_{a\le t\le t'-1} (q^{-1})^{(t'-t)} \alpha(\mathbf{i},t,x) \\
&\le c^d + \sum_{k-1\le t\le t'-1} (q^{-1})^{(t'-t)} 2^{kd^2} c^d t^d C^{-(t'-t)} q^{t'-t} \\
&= c^d + 2^{kd^2} c^d \sum_{a\le t\le t'-1} t^d C^{-(t'-t)} \\
&= c^d + 2^{kd^2} c^d \sum_{1\le s\le t'-a} (t'-t)^d C^{-s} \\
&\le c^d + 2^{kd^2} c^d (kd)^{(d+1)} C^{-1} \le c^d(1+k^{1-3k}) \le 3c^d/2\,.
\end{aligned}$$

Where the last line follows because of our lower bound on C.

We now advise the reader to familiarise themselves with the notation of Kim-Vu's inequality (Theorem 8), which we now want to apply. It is not difficult to see, as claimed above, that $Y_{\mathbf{1}}$ is a polynomial in random variables exactly as the one studied by Kim-Vu's inequality (up to a scaling factor). We use here the notation introduced for Kim-Vu's inequality.

Consider now the calculation for $\mathbb{E}_i(Y_{\mathbf{1}})$ with $i \ge 1$. Suppose we fix the variables $A \subseteq \{f_j^{(u)}(x) : u \in [d], j \in [k] \setminus \{i_u\}, x \in [n]\}$. Note that if we have $f_j^{(u)}(a), f_j^{(u)}(b) \in A$ with $a \ne b$ then the expectation is 0 since no term in the sum contains both. Thus, also $\mathbb{E}_i(Y_{\mathbf{1}}) = 0$ whenever $i > d(k-1)$, and we may equivalently describe any subset of variables A for which $\mathbb{E}(Y_{\mathbf{1}_A})$ is non-vanishing, by specifying the elements $s_j^{(i)}$ held fixed in the corresponding functions $f_j^{(i)}$ in ψ. To this end, we introduce the following notation. Write $\mathbf{m_i}$ for the vector of length k, and whose entries may be empty, for the elements $s_j^{(i)}$ held fixed in $s^{(i)}$. Let m_i be the number of non-empty elements in $\mathbf{m_i}$ and $M = \sum m_i \le l$ (the total number of elements held fixed). For instance, suppose $\psi = *_{i_1,\mathbf{1}}(f_1^{(1)},\dots,f_k^{(1)}) \dots *_{i_d,\mathbf{1}}(f_1^{(d)},\dots,f_k^{(d)})$ and we fix $f_1^{(1)}(s_1^{(1)}), f_k^{(1)}(s_k^{(1)}), f_1^{(2)}(s_1^{(2)})$, with all other variables allowed to vary. Then letting $\star$ denote the empty element, we have $\mathbf{m_1} = (s_1^{(1)},\star,\dots,\star,s_k^{(1)})$, $\mathbf{m_2} = (s_1^{(2)},\star,\dots,\star)$ and for all other $3 \le j \le l$, $\mathbf{m_j}$ is the empty vector of length k. Write $\mathbf{M} = (\mathbf{m_1},\dots,\mathbf{m_d})$ for the collection of $\mathbf{m_i}$ and $H_{\mathbf{M}}$ for the truncated polynomial retaining those configurations fixing $\mathbf{M}$. We use the notation $S(\mathbf{m})$ for the collection of hyperedges fixing $\mathbf{m}$, and $S_i(x;\mathbf{m})$ for the set of hyperedges with x in the i-th position, and the elements of the $\mathbf{m}$ in the positions in which they occur. Note that wherever we use this notation, we have $\mathbf{m}$ empty in the i-th position, so there is no potential conflict here. Recalling that l is the number of copies of $[n]_q$ in ψ, let $\tilde{l} = \tilde{l}(\mathbf{M})$ be the number of these whose entries are fixed in $\mathbf{M}$. In this notation, the truncated polynomial $H_{1,\mathbf{M}}$ fixing $\mathbf{m_1},\dots,\mathbf{m_d}$ in the edges $s^{(1)},\dots,s^{(d)}$ takes the form

$$H_{1,\mathbf{M}} = \frac{q^{-\tilde{l}}}{n}\Big(\frac{n}{e(S)}\Big)^d \sum_{x\in[n]} \sum_{\substack{s^{(1)}\in S_1(x;\mathbf{m_1}) \\ \cdots \\ s^{(d)}\in S_d(x;\mathbf{m_d})}} g_1(\cdot)\dots g_{l-M}(\cdot)\,. \tag{18}$$

where we have omitted explicitly writing any variables corresponding to constant one functions since it does not change the value of the polynomial. $g_1(\cdot)\dots g_{l-M}(\cdot)$ denote only the $l-M$ unfixed variables $f_j^{(i)}(\cdot)$ where $f_j^{(i)}$ is a copy of $[n]_q$ and $\cdot$ the appropriate element $x \in [n]$ at which it is evaluated within the configuration.

In this form, it is clear that fixing additional constant-one valued variables decreases the polynomial expectation since it only amounts to dropping terms, each having a strictly positive expectation. Specifically given any $\mathbf{M}$, if $\mathbf{M'}$ fixes all the elements of $\mathbf{M}$, along with elements $s_{j_1}^{(i_1)},\dots,s_{j_a}^{(i_a)}$ for which the corresponding $f_{j_1}^{(i_1)},\dots,f_{j_a}^{(i_a)}$ are all constant-one valued, then the polynomial $H_{\mathbf{M'}}$ retains only those configurations (if any) of $H_{\mathbf{M}}$ which fix also $s_{j_1}^{(i_1)},\dots,s_{j_a}^{(i_a)}$ in $s^{(i_1)},\dots,s^{(i_a)}$ respectively and so $\mathbb{E}(H_{\mathbf{M}}) \geq \mathbb{E}(H_{\mathbf{M'}})$. We may therefore assume that no constant-one variables are fixed for the purpose of maximising $\mathbb{E}_i(Y_{\mathbf{1}})$ to apply Theorem 8. Recalling that ψ has l of its $d(k-1)$ comprising functions being copies of $[n]_q$ and the rest being constant one, we are interested only in $\mathbb{E}_i(Y_{\mathbf{1}})$ with $1 \leq i \leq l$. In general, the greater the number i of $0, q^{-1}$ valued variables being fixed, the larger the premultiplying coefficient q^{-i}. However, fixing these variables also reduces the number of contributing configurations by a factor of order $(C^{-1}q)^i < q^i$ (arising from the maximum co-degree condition), meaning an overall reduction in $\mathbb{E}_i(Y_{\mathbf{1}})$ with greater i. Moreover, where a fixed variable corresponds to an $f_j^{(i)}$ which is identical to an unfixed copy of $[n]_q$, there is a further reduction from the interdependence. Although in almost all configurations, the arguments for these indicator functions differs and thus the corresponding variables are independent. In this way one expects that $\mathbb{E}_i(Y_{\mathbf{1}})$ is greatest for $i=1$, decreasing by a factor of about C^{-1} per variable fixed.

Recalling $1 \leq M \leq l$. When $M = l$,

$$
\begin{aligned}
E(H_{1,\mathbf{M}}) &= \frac{q^{-l}}{n}\sum_{x\in[n]}\left(\frac{n}{e(S)}\right)^d \sum_{\substack{s^{(1)}\in S_{i_1}(x;\mathbf{m_1})\\ \dots \\ s^{(d)}\in S_{i_d}(x;\mathbf{m_d})}} \leq \frac{q^{-l}}{n}\cdot\left(\frac{n}{e(S)}\right)^d \sum_{\substack{s^{(1)}\in S(\mathbf{m_1})\\ s^{(2)}\in S_{i_2}(s^{(1)}_{i_1};\mathbf{m_2})\\ \dots \\ s^{(d)}\in S_{i_d}(s^{(1)}_{i_1};\mathbf{m_d})}} 1\\
&\leq \frac{q^{-l}}{n}\cdot\left(\frac{n}{e(S)}\right)^d \Delta_{m_1}\Delta_{m_2+1}\dots\Delta_{m_d+1}\\
&\leq \frac{q^{-l}}{n}\cdot\left(\frac{n}{e(S)}\right)^d c^d C^{d-(M+d-1)}q^{(M+d-1)-d}\left(\frac{e(S)}{n}\right)^d \leq \frac{q^{-1}c^dC^{1-l}}{n}\\
&\leq \frac{c^dC^{-l}}{log^{2k}(n)}.
\end{aligned}
$$

Thus, $\mathbb{E}_l(Y_{\mathbf{1}}) \leq \frac{c^dC^{-l}}{log^{2k}(n)}$. Otherwise, $M < l$ and within each configuration in $H_{\mathbf{M}}$ there are $l-M$ Bernoulli-$\{0,q^{-1}\}$ variables, which may or may not be independent. As with the calculation for $\mathbb{E}_0(Y_{\mathbf{1}})$, we allow that any configuration fixing $\mathbf{m_1},\dots,\mathbf{m_2}$ in $s^{(1)},\dots,s^{(d)}$ may result in precisely $t \in [l-M]$ of the $g_i(\cdot)$ being mutually independent. We follow a similar approach to that for $\mathbb{E}_0(Y_{\mathbf{1}})$, counting configurations which have the same number of mutually independent $g_i(\cdot)$. Note first that

$$\begin{aligned}\mathbb{E}(H_{1,\mathbf{M}}) &= \frac{q^{-M}}{n}\left(\frac{n}{e(S)}\right)^{d}\sum_{x\in[n]}\sum_{\substack{s^{(1)}\in S_{i_1}(x;\mathbf{m_1})\\ \cdots\\ s^{(d)}\in S_{i_d}(x;\mathbf{m_d})}}\mathbb{E}g_1(\cdot)\dots g_{M-l}(\cdot)\\ &= \frac{q^{-M}}{n}\left(\frac{n}{e(S)}\right)^{d}\sum_{s^{(1)}\in S(\mathbf{m_1})}\sum_{\substack{s^{(2)}\in S_{i_2}(s^{(1)}_{i_1});\mathbf{m_2})\\ \cdots\\ s^{(d)}\in S_{i_d}(s^{(1)}_{i_1};\mathbf{m_d})}}\mathbb{E}g_1(\cdot)\dots g_{M-l}(\cdot)\,.\end{aligned}$$

For a given $s^{(1)}\in S(\mathbf{m_1})$, let $\alpha(\psi,\mathbf{M},t,s^{(1)})$ be the size of the set $S(\psi,\mathbf{M},t,s^{(1)})\subseteq\{s^{(1)}\}\times S_{i_2}(s^{(1)};\mathbf{m_2})\times\ldots\times S_{i_d}(s^{(1)};\mathbf{m_d})$ for which t of the $g_i(\cdot)$ are mutually independent. The expectation of the product $g_1(\cdot)\dots g_{l-M}(\cdot)$ for such a configuration is $q^{-(l-M-t)}$. We have

$$\begin{aligned}\alpha(\psi,\mathbf{M},t,s^{(1)})\le&\sum_{\substack{r_2+\ldots+r_d\\=l-M-t}}\binom{(k-1)}{r_2}\Delta_{m_2+1+r_2}\binom{2(k-1)-r_2}{r_3}\Delta_{m_3+1+r_3}\cdot\ldots\\ &\ldots\cdot\binom{(d-2)(k-1)-\sum_{i\in[d-1]\setminus\{1\}}r_i}{r_d}\Delta_{m_d+1-r_d}\\ &\le 2^{kd^2}(l-M-t)^{d-1}c^{d-1}C^{-l+t+m_1}q^{l-t-m_1}\left(\frac{e(S)}{n}\right)^{d-1}.\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}(H_{1,\mathbf{M}}) &\le \frac{q^{-M}}{n}\left(\frac{n}{e(S)}\right)^{d}\sum_{s^{(1)}\in S(\mathbf{m_1})}\sum_{t\in[l-M]}q^{-(l-M-t)}2^{kd^2}(l-M-t)^{d-1}c^{d-1}\\ &\qquad\cdot C^{-l+t+m_1}q^{l-t-m_1}\left(\frac{e(S)}{n}\right)^{d-1}\\ &\le\frac{q^{-1}}{n}2^{kd^2}(l-M)^dc^dC^{1-M}\\ &\le\frac{c^d2^{kd^2}(l-M)^dC^{-M}}{log^{2k}(n)}\,.\end{aligned}$$

Hence, for $M\ge 1$, $\mathbb{E}(H_{\mathbf{M}})$ is maximised when $M=1$. Using the notation as in Theorem 8 we have $\mathbb{E}'(Y_{\mathbf{1}})\le\frac{c^d2^{kd^2}(l-1)^dC^{-1}}{log^{2k}(n)}\le\frac{c^d}{k^{3k}log^{2k}(n)}$ and using the lower bound for q and C. Clearly $\mathbb{E}'(Y_{\mathbf{1}})\ll 1<\mathbb{E}_0(Y_{\mathbf{1}})=\mathbb{E}(Y_{\mathbf{1}})$.

Using (Theorem 8) we have

$$P\left[|Y_{\mathbf{1}}-\mathbb{E}_0(Y_{\mathbf{1}})|>(8^k\cdot k!^{1/2})(\mathbb{E}(Y_{\mathbf{1}})\mathbb{E}'(Y_{\mathbf{1}}))^{1/2}\lambda^k\right]=O(e^{(-\lambda+(k-1)\log(n))})\,.$$

To achieve the concentration we require, take $\lambda=k(1+\alpha)\log(n)$. It then follows that we have $e^{(-\lambda+(k-1)\log(n))}\le e^{-(\alpha k+1)\log(n)}\le\frac{1}{n^{\alpha k}}$. We then have

$$\begin{aligned}(8^k \cdot k!^{1/2})(\mathbb{E}(Y_{\mathbf{1}})\mathbb{E}'(Y_{\mathbf{1}}))^{1/2}\lambda^k &\le (8^k \cdot k^{k/2})\Big(\frac{3c^{2d}}{2k^{3k}log^{2k}(n)}\Big)^{1/2}(k(1+\alpha)\log(n))^k \\ &\le 2c^d 8^k(1+\alpha)^k\,.\end{aligned}$$

We now turn to the concentration of $Y_\mu = \langle \mu, \psi\rangle$. Observe first that since the argument of μ is distinct from the argument of any other $f_j^{(i)}$ within a configuration, then $\mu(x)$ is independent of all other variables in the configuration. Since $\mathbb{E}(\mu(x)) = 1$ for any x, we have $\mathbb{E}_0(Y_\mu) = \mathbb{E}_0(Y_{\mathbf{1}}) \le 3/2$.

For any subset A of variables that we fix in $Y_{\mathbf{1}}$ to calculate $\mathbb{E}_A(Y_{\mathbf{1}})$, we have $\mathbb{E}(Y_{\mathbf{1}_A}) = \mathbb{E}_A(Y_{\mu_A})$ and for a fixed $x \in [n]$ and $B = A \cup \{\mu(x)\}$, we have $\mathbb{E}(Y_{\mu_B}) \le \mathbb{E}(Y_{\mu_A})$ since we merely sum the same expectations over fewer configurations. Thus, $\mathbb{E}_i(Y_\mu)$ is maximised for $i = 1$ as for $\mathbb{E}_i(Y_{\mathbf{1}})$. Note that if $A = \{f_j^{(i)}(s_j^{(i)})\}$ then $\mathbb{E}(Y_{\mathbf{1}_A}) = \mathbb{E}(Y_{\mu_A})$. If $A = \{\mu(x)\}$ for some fixed $x \in [n]$ then the calculation $\mathbb{E}(Y_{\mathbf{1}_A})$ proceeds just as for $\mathbb{E}_0(Y_{\mathbf{1}})$ except with the summation over $x \in [n]$ dropped. That is, for $A = \{\mu(x)\}$, we have $\mathbb{E}_A(Y_{\mu_A}) \le 3/2n \le \frac{q^{-l}}{n}2^{kd^2}l^d c^d$, thus the upper bound for $\mathbb{E}'(Y_{\mathbf{1}})$ holds for $\mathbb{E}'(Y_\mu)$ also, and the concentration obtained from the Kim-Vu inequality applies.

□

We deduce the following corollary.

Corollary 27. *Given d positive integer, there exist C and L_0 such that, if the C-conditions are satisfied in Setting 2 and $L \ge L_0$, then with high probability over the choice of $X = [n]_p$, and independently $\chi_\mu : [n] \to \{1,\dots,L\}$ uniformly at random, and independently $\chi_{\mathbf{1}} : [n] \to \{1,\dots,\lceil Lp^{-1}\rceil\}$ uniformly at random, the following holds. Let us use Notation 20. For any $1 \le \ell \le d$ and ψ a largest anti-uniform functional in $\Phi(\mu,\mathbf{1})^\ell$ or $\Phi(\mu_1,\dots,\mu_L,\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil})^\ell$.*

$$\langle \mu, \psi\rangle \le 2c^\ell \qquad \text{and} \qquad \langle \mathbf{1}, \psi\rangle \le 2c^\ell\,.$$

In addition, if ψ is any largest anti-uniform functional in $\Phi(\mu_1,\dots,\mu_L,\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil})^\ell$, and $1 \le j \le L$ and $1 \le j' \le \lceil Lp^{-1}\rceil$ then we have

$$\langle \mu_j, \psi\rangle \le 2c^\ell \qquad \text{and} \qquad \langle \nu_{j'}, \psi\rangle \le 2c^\ell\,.$$

The idea of the proof is to take a union bound over choices of ψ and j, of which there are only polynomially many, and use Lemma 26 to obtain the correlation bounds.

The only place where we need to be a bit careful is that the μ_j are not independent; and similarly the $\nu_{j'}$. We find some related functions $\hat{\mu}_j$ and $\hat{\nu}_{j'}$ which are independent and to which we apply Lemma 26, and deduce the required correlation bounds from these.

Proof. Let $H = \{\mu_1,\dots,\mu_L,\nu_1,\dots,\nu_{\lceil Lp^{-1}\rceil}\}$, let α be such that both $dk^d \cdot 2^{(k-1)d}\frac{1}{n^{\alpha k}}$ and $dk^d(2Lp^{-1})^{d(k-1)+1}\frac{1}{n^{\alpha k}}$ are $o(1)$. Let C be large enough so that Lemma 26 works for the choice $q = p$, α, and $d' = d$. Assume the C-conditions are satisfied.

We first establish bounds on $\langle \mu, \psi\rangle$ and $\langle \mathbf{1}, \psi\rangle$ for $\psi \in \Phi(\mu,\mathbf{1})^\ell$. Given $1 \le \ell \le d$, if ψ is a largest anti-uniform functional in $\Phi(\mu,\mathbf{1})^\ell$, then $\psi = \prod_{j=1}^{\ell} *_{i_j,\mathbf{1}}(f_{j,1},\dots,f_{j,k-1})$, where $1 \le i_j \le k$ for each j and each $f_{j,j'}$ is either μ or $\mathbf{1}$. For any such function, the probability of

$$\langle \mu, \psi\rangle > 2c^\ell \qquad \text{or} \qquad \langle \mathbf{1}, \psi\rangle > 2c^\ell$$

is, by Lemma 26, at most $\frac{1}{n^{\alpha k}}$. Taking the union bound over the at most $dk^d \cdot 2^{(k-1)d}$ choices of ℓ, i_j and $f_{j,j'}$, we see that the probability of any of these events occurring is at most $dk^d \cdot 2^{(k-1)d} \frac{1}{n^{\alpha k}}$, which is $o(1)$ because of our choice of α.

We now establish corresponding bounds on $\langle \mu_j, \psi \rangle$ and $\langle \nu_j, \psi \rangle$. Observe that, as before, we can describe any largest anti-uniform functional ψ in $\Phi(H)^\ell$ as follows. We choose $i_1, \ldots, i_\ell$, and for each of the $\ell(k-1)$ functions in the product, we must choose one of H. Finally, to describe the entire inner product, we must choose the left term in the inner product (either μ_j or ν_j) from H. In total, the number of choices is at most $dk^d(2Lp^{-1})^{d(k-1)+1}$.

Fix now one such set of choices. Let T denote a collection of $d(k-1)+1$ indices in $[L]$ such that μ_t is one of the chosen functions for each $t \in T$, and T' a subset of $[\lceil Lp^{-1} \rceil]$ of size $d(k-1)$ such that ν_t is chosen for each $t \in T'$.

Consider the following random experiment. For each $t \in T$, we first generate independent binomial random subsets $Z_t = [n]_q$, with $0 < q < 1$ chosen such that $(1-q)^{|T|} = 1 - t\frac{p}{L}$. We now obtain sets Z'_t for $t \in T$ as follows. For each $x \in \bigcup_{t \in T} Z_t$ independently, pick t uniformly at random from the set $\{t : x \in Z_t\}$, and let $x \in Z'_t$.

By definition of q, for a given $x \in [n]$ the probability that $x \in \bigcup_{t \in T} Z_t$ is $\frac{tp}{L}$, and conditioning on this event occurring, the events $x \in Z'_t$ are disjoint over $t \in T$, and x is equally likely to appear in any given Z'_t for $t \in T$, so that probability of $x \in Z'_t$ is $\frac{p}{L}$. Observe that this is the same probability as the event that $x \in X$ and $\chi_\mu(x) = t$, which are also disjoint events over $t \in T$. It follows that the distribution of $(Z_t)_{t \in T}$ is the same as the distribution of $(X \cap \{x : \chi_\mu(x) = t\})_{t \in T}$, so we can consider the coupling in which the latter sets are generated according to the above random experiment.

Let $\hat{\mu}_t(x) = q^{-1}\mathbb{1}(x \in Z_t)$. By construction, we have $0 \le \mu_t(x) \le Lp^{-1}q\hat{\mu}_t(x)$.

We now perform a similar, independent, random experiment. For each $t \in T'$, we generate independently $W_t = [n]_q$ where q is as defined above. Letting now $0 < q' < 1$ solve $(1-q')^{|T'|} = 1 - t\frac{1}{\lceil Lp^{-1} \rceil}$, we observe $q' \le q$. We generate W''_t by sampling the elements of W_t independently with probability $\frac{q'}{q}$, so that the W''_t are independent copies of $[n]_{q'}$. Finally, we generate W'_t by, as above, picking t from $\{t : x \in W'_t\}$ independently and uniformly and letting $x \in W'_t$.

As before, the distribution of $(W'_t)_{t \in T'}$ is identical to the distribution of $(\{x : \chi_{\mathbf{1}}(x) = t\})_{t \in T'}$ and we consider the coupling in which the latter sets are generated by the above random experiment. Letting $\hat{\nu}_t(x) = q^{-1}\mathbb{1}(x \in W_t)$, we have $0 \le \nu_t(x) \le \lceil Lp^{-1} \rceil q\hat{\nu}_t(x)$.

Let $\hat{\psi}$ denote the function obtained by replacing each μ_t with $\hat{\mu}_t$ for $t \in T$, and each ν_t with $\hat{\nu}_t$ for $t \in T'$, in the product defining ψ. Then we have

$$\langle \mu_j, \psi \rangle \le (\lceil Lp^{-1} \rceil q)^{d(k-1)+1} \langle \hat{\mu}_j, \hat{\psi} \rangle \quad \text{and} \quad \langle \nu_j, \psi \rangle \le (\lceil Lp^{-1} \rceil q)^{d(k-1)+1} \langle \hat{\nu}_j, \hat{\psi} \rangle .$$

Now, $\hat{\psi}$ is a (q,d)-special product. Assume C is also large enough that so that Lemma 26 holds for $d' = d$, our α, and $q = q$.

$$\langle \hat{\mu}_j, \hat{\psi} \rangle > \tfrac{7}{4}c^\ell \qquad \text{and} \qquad \langle \hat{\nu}_j, \hat{\psi} \rangle > \tfrac{7}{4}c^\ell$$

each have probability at most $\frac{1}{n^{\alpha k}}$ by Lemma 26. Since $\frac{7}{4}(\lceil Lp^{-1} \rceil q)^{d(k-1)+1} < 2$, the same probability bounds hold on the events

$$\langle \mu_j, \psi \rangle > 2c^\ell \qquad \text{and} \qquad \langle \nu_j, \psi \rangle > 2c^\ell .$$

Finally taking the union bound, the probability that any one of these events fails is $o(1)$ by our choice of α.

Suppose that none of the above bad events occur. We deduce, deterministically, the remaining bounds of Corollary 27. We begin with $1 \le \ell \le d$ and $\psi \in \Phi(H)^\ell$, for which we have

$$\langle \mu, \psi \rangle = \tfrac{1}{L}\sum_{i=1}^{L} \langle \mu_i, \psi \rangle \le \tfrac{1}{L}\sum_{i=1}^{L} 2c^\ell = 2c^\ell\,.$$

Similarly, we have

$$\langle \mathbf{1}, \psi \rangle = \tfrac{1}{\lceil Lp^{-1}\rceil}\sum_{i=1}^{\lceil Lp^{-1}\rceil} \langle \nu_i, \psi \rangle \le \tfrac{1}{\lceil Lp^{-1}\rceil}\sum_{i=1}^{\lceil Lp^{-1}\rceil} 2c^\ell = 2c^\ell\,. \qquad \square$$

11. Deletion method

11.1. **A general deletion method.** In this section we prove that the required $\tilde{X}$ satisfying moment estimates exists with exponentially small failure probability. This follows from the Harris inequality. Recall that a subset $\mathcal{D}$ of $\mathcal{P}\big([n]\big)$ is called *decreasing* if whenever $S' \subseteq S \in \mathcal{D}$ we have $S' \in \mathcal{D}$, and *increasing* if the same statement holds with $\subseteq$ replaced by $\supseteq$.

Theorem 28 (Harris [11]). *For any $p \in [0,1]$ and n, let $\mathcal{A}$ and $\mathcal{B}$ be two subsets of $\mathcal{P}\big([n]\big)$, which are both decreasing. Then*

$$\mathbb{P}([n]_p \in \mathcal{A} \cap \mathcal{B}) \ge \mathbb{P}([n]_p \in \mathcal{A})\mathbb{P}([n]_p \in \mathcal{B})\,.$$

Spöhel, Steger and Warnke [17] deduced the following theorem. They state their result for the specific case $[n] = \binom{[m]}{2}$ (i.e. for the random graph), but their proof works verbatim in the more general situation. For completeness, we give the details.

Theorem 29 ([17, Theorem 4]). *Let $\mathcal{D}$ be a decreasing subset of $\mathcal{P}\big([n]\big)$. Given $\alpha, \delta \in (0,1]$, let $p \in (0,1]$ be such that $\mathbb{P}\big([n]_p \in \mathcal{D}\big) \ge \delta$. Then with probability at least $1 - \delta^{-1}\exp\big(-\frac{1}{2}\alpha^2 pn\big)$, there is a subset of $[n]_p$ with at least $(1-\alpha)pn$ elements which is in $\mathcal{D}$.*

Proof. Let $\mathcal{I}$ be the subset of $\mathcal{P}([n])$ consisting of sets with at least $(1-\alpha)pn$ elements. Let $\mathcal{S}$ be the subset of sets $S \in \mathcal{P}([n])$ such that S has a subset in $\mathcal{I} \cap \mathcal{D}$, which is clearly increasing, so $\overline{\mathcal{S}}$ is decreasing. By Theorem 28, we have $\mathbb{P}\big([n]_p \in \overline{\mathcal{S}}\big)\mathbb{P}\big([n]_p \in \mathcal{D}\big) \le \mathbb{P}\big([n]_p \in \overline{\mathcal{S}} \cap \mathcal{D}\big)$. Rearranging, and observing $\overline{\mathcal{S}} \cap \mathcal{D} \subseteq \overline{\mathcal{I}}$, we get

$$\mathbb{P}\big([n]_p \in \overline{\mathcal{S}}\big) \le \frac{\mathbb{P}\big([n]_p \in \overline{\mathcal{S}} \cap \mathcal{D}\big)}{\mathbb{P}\big([n]_p \in \mathcal{D}\big)} \le \delta^{-1}\mathbb{P}\big([n]_p \in \overline{\mathcal{I}}\big)\,.$$

Chernoff's inequality now gives $\mathbb{P}\big([n]_p \in \overline{\mathcal{I}}\big) \le \exp\big(-\frac{1}{2}\alpha^2 pn\big)$, which gives the required probability bound. $\square$

We now have the tools to prove the last remaining Lemma, i.e. Lemma 23.

Proof of Lemma 23. We are in Setting 2. Let C and L be large enough so that Corollary 27 works for our choice of $d = d'$. Assume the C-conditions are satisfied.

Let $\mathcal{D} \subseteq \mathcal{P}([n])$ be the set of subsets $Y \subseteq [n]$ satisfying the following. Letting $\mu(x) = p^{-1}\mathbb{1}(x \in Y)$, for uniform random choices of χ_μ and $\chi_{\mathbf{1}}$, with probability at least 0.9, for all $1 \le \ell \le d$ and all largest anti-uniform functionals ψ either in the set $\Phi(\mu, \mathbf{1})^\ell$ or in the set $\Phi(\mu_1, \dots, \mu_L, \nu_1, \dots, \nu_{\lceil Lp^{-1}\rceil})^\ell$, we have

$$\langle \mu, \psi \rangle \le 2c^\ell \qquad \text{and} \qquad \langle \mathbf{1}, \psi \rangle \le 2c^\ell\,.$$

In addition, if ψ is any largest anti-uniform functional in $\Phi(H)^\ell$, and $1 \le j \le L$ and $1 \le j' \le \lceil Lp^{-1} \rceil$ then we have

$$\langle \mu_j, \psi \rangle \le 2c^\ell \qquad \text{and} \qquad \langle \nu_{j'}, \psi \rangle \le 2c^\ell \,.$$

Observe that since all the left hand sides of these conditions are increasing in X, the event $\mathcal{D}$ is a decreasing event. Furthermore, Corollary 27 states that $\mathbb{P}(\mathcal{D}) = 1 - o(1) \ge \frac{1}{2}$.

We now apply Theorem 29 with this $\mathcal{D}$, with $\alpha = \frac{1}{2}\delta$, and with $\mathbb{P}(\mathcal{D}) \ge \frac{1}{2}$, to deduce that with probability at least $1 - 2\exp\big(-\frac{1}{8}\delta^2 pn\big)$ there is a subset $\tilde{X}$ of X which is in $\mathcal{D}$ and which has at least $\big(1 - \frac{1}{2}\delta\big)pn$ elements. Additionally, the probability that $[n]_p$ has more than $\big(1 + \frac{1}{2}\delta\big)pn$ elements is by Theorem 5 at most $\exp\big(-\frac{1}{8}\delta^2 pn\big)$. Suppose that X satisfies both conditions, which occurs with probability at least $1 - 3\exp\big(-\frac{1}{8}\delta^2 pn\big)$ by the union bound. Then $|X \setminus \tilde{X}| \le \delta pn$ as required. □

11.2. **Transference Principle without Deletion.** We are finally ready to prove items 1 and 2 of Theorem 3.

1 and 2 of Theorem 3. We are in Setting 2. We have that 1 follows immediately from 3 as an ε-good dense model for $\tilde{X}$ provides an ε-good lower dense model for X.

Let us now show how to get 2 from 3 in Theorem 3. First, we may assume without loss of generality that $\bar{\Omega} = \{\bar{\omega} = \mathbf{1} - \omega : \omega \in \Omega\}$ is contained in Ω. This is because this assumption at most doubles the size of Ω, and therefore doesn't affect the order of magnitude of its size. Let C be large enough to guarantee that 3 works for $\varepsilon = \frac{\varepsilon}{2^{k+2}}$. Assume the C conditions are satisfied. Let X be a sample of $[n]_p$ such that $|\{s \in S : s \subseteq X\}| \le (1+\frac{\varepsilon}{2})\mathbb{E}[|\{s \in S : s \subseteq [n]_p\}|]$ and such that X admits an $\frac{\varepsilon}{2^{k+2}}$-deletion $\tilde{X}$ such that all subsets of $\tilde{X}$ have an $\frac{\varepsilon}{2^{k+2}}$-good dense model. This happens with probability at most $1 - \eta_n$ on the choice of $X = [n]_p$. Notice that we have $|\{s \in S : s \subseteq X \setminus \tilde{X}\}| \le \frac{\varepsilon}{2}\mathbb{E}[|\{s \in S : s \subseteq [n]_p\}|]$ because of our upper bound on $|\{s \in S : s \subseteq X\}|$ and because by Theorem 18 we have that $\mathbf{1}$ is a dense model of $\tilde{X}$. Let μ and $\tilde{\mu}$ be the p^{-1} scaled indicator functions of X and $\tilde{X}$ respectively.

Let us now consider a subset Y of X. Let $\tilde{Y} = Y \cap \tilde{X}$, and let $\bar{Y} = \tilde{X} \setminus Y$. Let f be the scaled indicator function of $\tilde{Y}$. We use $\bar{f}$ for the complements in $\tilde{X}$.

Fix an arbitrary $\omega \in \Omega$, and let $\bar{\omega} = \mathbf{1} - \omega$. We have the following.

$$\langle \tilde{\mu}, *_{i,\mathbf{1}}(\tilde{\mu}, \dots, \tilde{\mu}) \rangle = \langle \tilde{\mu}, *_{i,\omega}(\tilde{\mu}, \dots, \tilde{\mu}) \rangle + \langle \tilde{\mu}, *_{i,\bar{\omega}}(\tilde{\mu}, \dots, \tilde{\mu}) \rangle$$

Consider now that we can split the set of edges of S contained in $\tilde{X}$ by grouping together edges depending on what are the indices corresponding to elements of Y and which to element of $\bar{Y}$.

$$\langle \tilde{\mu}, *_{i,\omega}(\tilde{\mu}, \dots, \tilde{\mu}) \rangle = \sum_{\mathbf{f} \in \{f, \bar{f}\}^k} \langle \mathbf{f}_i, *_{i,\omega}(\mathbf{f}_1, \dots, \mathbf{f}_k) \rangle$$

Because $\tilde{Y}$ is a subset of $\tilde{X}$, we can ask for an $\frac{\varepsilon}{2^{k+2}}$-good dense model $Z_{\tilde{Y}}$ of $\tilde{Y}$. Let g be the scaled indicator function of its model $Z_{\tilde{Y}}$ and define $\bar{g}$ as $\mathbf{1} - g$. Because $Z_{\tilde{Y}}$ is a good model of $\tilde{Y}$ we have $\|f - g\|_{\Phi(\mathbf{1})} \le \frac{\varepsilon}{2^{k+2}}$. We therefore have:

$$\sum_{\mathbf{f}\in\{f,\bar{f}\}^k} \langle \mathbf{f}_i, *_{i,\omega}(\mathbf{f}_1,\dots,\mathbf{f}_k)\rangle = \sum_{\mathbf{g}\in\{g,\bar{g}\}^k} \langle \mathbf{g}_i, *_{i,\omega}(\mathbf{g}_1,\dots,\mathbf{g}_k)\rangle \pm \frac{\varepsilon}{4}\,.$$

We can substitute this to obtain the following:

$$\begin{aligned}\langle\tilde{\mu}, *_{i,\mathbf{1}}(\tilde{\mu},\dots,\tilde{\mu})\rangle \geq& \langle\tilde{\mu}, *_{i,\bar{\omega}}(\tilde{\mu},\dots,\tilde{\mu})\rangle + \langle f, *_{i,\omega}(f,\dots,f)\rangle - \langle g, *_{i,\omega}(g,\dots,g)\rangle\\ &+ \sum_{\mathbf{g}\in\{g,\bar{g}\}^k} \langle \mathbf{g}_i, *_{i,\omega}(\mathbf{g}_1,\dots,\mathbf{g}_k)\rangle - \frac{\varepsilon}{4}\,.\end{aligned}$$

Considering now that $g+\bar{g}=\mathbf{1}$, and that, by Theorem 18 we have $\|\tilde{\mu}-\mathbf{1}\|_{\Phi(\tilde{\mu},\mathbf{1})} < \frac{\varepsilon}{2^{k+2}}$, we obtain:

$$\begin{aligned}\langle\tilde{\mu}, *_{i,\mathbf{1}}(\tilde{\mu},\dots,\tilde{\mu})\rangle \geq& \langle\tilde{\mu}, *_{i,\bar{\omega}}(\tilde{\mu},\dots,\tilde{\mu})\rangle + \langle f, *_{i,\omega}(f,\dots,f)\rangle - \langle g, *_{i,\omega}(g,\dots,g)\rangle\\ &+\langle\tilde{\mu}, *_{i,\omega}(\tilde{\mu},\dots,\tilde{\mu})\rangle - \frac{\varepsilon}{2}\,.\end{aligned}$$

By cancelling out the terms (which we can do as $\mathbf{1}=\omega+\bar{\omega}$), we obtain:

$$\langle f, *_{i,\omega}(f,\dots,f)\rangle \leq \langle g, *_{i,\omega}(g,\dots,g)\rangle + \frac{\varepsilon}{2}\,.$$

Returning to the definition of inner product (i.e. noticing that by definition we have that $\langle f, *_{i,\omega}(f,\dots,f)\rangle = \frac{p^k}{e(S)}\sum_{s\in S}\omega(s)\mathbb{1}(s\subseteq Y)$ and similarly for $\tilde{\mu}$ and $\tilde{X}$, f and $\tilde{Y}$, and X and μ) we conclude. □

12. A sparse counting lemma

In this section we prove Theorem 30. This turns out to be an application of Theorem 3, together with a standard counting lemma for hypergraphs; most of what follows is simply dealing with the somewhat complicated hypergraph regularity setup.

Let k be a positive integer. A *k-complex* is a down-closed hypergraph in which all edges have size at most k. Given a k-complex H with at least $k+1$ vertices, we define its *k-density* as $d_k(H) := \frac{e_k(H)-1}{v(H)-k}$, where $e_k(H)$ is the number of edges of size k in H, and $v(H)$ denotes the number of vertices of H. We also define $m_k(H) := \max_{H'\subseteq H} d_k(H')$, where the maximum is taken over all sub-k-complexes H' of H with at least $k+1$ vertices.

Given a vertex set $[N]$, a *k-partition with ℓ clusters $\mathcal{V}$* consists of a family of disjoint subsets $V_{\{1\}},\dots,V_{\{\ell\}}\subseteq[N]$ called *clusters*, together with, for each integer $2\leq i\leq k$ and each subset $E\subseteq[\ell]$ of size i, a collection V_E of subsets of $[N]$ of size i, called *i-edges*. These V_E must satisfy the following compatibility condition: for every $e\in V_E$ and every $j\in E$, the set e intersects the cluster $V_{\{j\}}$ in exactly one element, and the remaining $i-1$ elements of e form an $(i-1)$-edge in $V_{E\setminus\{j\}}$. The *supporting $(i-1)$-graph* of V_E is the $(i-1)$-uniform hypergraph consisting of all $(i-1)$-sets that arise in this way from some edge of V_E.

Let $E\subseteq[\ell]$ with $|E|=i\geq 2$, and suppose V_E is given along with its supporting $(i-1)$-graphs $W_1,\dots,W_i$. For any subsets $Q_1\subseteq W_1,\dots,Q_i\subseteq W_i$, define $R(Q_1,\dots,Q_i)$ to be the collection of i-element subsets of $[N]$ that contain one element from each Q_j. In particular, $R(W_1,\dots,W_i)$ contains V_E. If $R(W_1,\dots,W_i)$ is nonempty, and given $p\in(0,1]$, we define respectively the *relative density* of V_E and the *relative p-density* of V_E as follows:

$$d^*(V_E) := \frac{|V_E|}{|R(W_1, \ldots, W_i)|} \qquad \text{and} \qquad d_p^*(V_E) := \frac{|V_E|}{p \cdot |R(W_1, \ldots, W_i)|}.$$

Finally, for singleton sets, we define $d^*(V_{\{i\}}) := |V_{\{i\}}|N^{-1}$.

Let $E \subseteq [\ell]$ be a set of size i, with $2 \leq i \leq k$, and let $p \in (0, 1]$. Consider V_E and let $W_1, \ldots, W_i$ denote the supporting $(i-1)$-graphs of V_E. We say that V_E is *(ε, r, p)-regular with respect to its supporting $(i-1)$-graphs* if the following holds. For any set R^* of the form $R^* = \bigcup_{j=1}^{r} R(Q_1^{(j)}, \ldots, Q_i^{(j)})$ where $Q_i^{(j)} \subseteq W_i$, we have that if $|R^*| \geq \varepsilon |R(W_1, \ldots, W_i)|$, then

$$\frac{|V_E \cap R^*|}{p|R^*|} = d_p^*(V_E) \pm \varepsilon.$$

If any of the parameters r, p, or both are omitted, they are understood to be equal to 1.

A k-partition is said to be *$(\varepsilon_k, \varepsilon, d_1, \ldots, d_k, r, p)$-regular* if the following conditions hold:

- For each $i \in [\ell]$, we have $|V_{\{i\}}| \geq d_1 N$;
- For every $E \subseteq [\ell]$ with $2 \leq |E| \leq k-1$, the set V_E is ε-regular with respect to its supporting $(|E| - 1)$-graphs, and its relative density satisfies $d^*(V_E) \geq d_{|E|}$;
- For every $E \subseteq [\ell]$ with $|E| = k$, the set V_E is (ε_k, r, p)-regular with respect to its supporting $(k-1)$-graphs, and its relative p-density satisfies $d_p^*(V_E) \geq d_k$.

Let H be a k-complex. An injective map $\phi : V(H) \to [\ell]$ is called a *k-complex homomorphism* if for every edge $e \in E(H)$, the image $\phi(e)$ has size $|e|$. That is, ϕ maps the vertices of each edge to distinct cluster indices. Given a k-partition $\mathcal{V}$ with ℓ clusters over the vertex set $[N]$, a map $\psi : V(H) \to [N]$ is said to be a *ϕ-partite copy of H in $\mathcal{V}$* if ψ is injective and for every edge $e \in E(H)$, the image $\psi(e)$ is an element of $V_{\phi(e)}$.

We are finally ready to introduce the hypergraph counting result.

Theorem 30 (Counting lemma for sparse hypergraphs). *Given $k \geq 2$, a fixed k-complex H, and $\delta > 0$, there exists $\varepsilon_k > 0$ such that for any $d_2, \ldots, d_k > 0$ (with $1/d_i \in \mathbb{N}$)[5] there exist $\varepsilon > 0$ and $r \in \mathbb{N}$ such that for any $d_1 > 0$ there exists C^* with the following property. Suppose that N is sufficiently large, and $p \geq \max\left(C^* N^{-1}, C^* N^{-1/m_k(H)}\right)$. With high probability, the random k-uniform hypergraph $\Gamma = G^{(k)}(N, p)$ has the following property.*

Given any $k \leq \ell \leq v(H)$ and $(\varepsilon_k, \varepsilon, d_1, \ldots, d_k, r, p)$-regular k-partition $\mathcal{V}$ with ℓ clusters on $[N]$, such that for each $E \subseteq [\ell]$ with $|E| = k$ we have $V_E \subseteq \Gamma$, and given any k-complex homomorphism $\phi : v(H) \to [\ell]$, the number of ϕ-partite copies of H in $\mathcal{V}$ is

$$(1 \pm \delta) N^{v(H)} \prod_{e \in E(H)} d^*\left(V_{\phi(e)}\right).$$

In the above theorem, we do allow for the possibility that some edges of H of uniformity smaller than k are not contained in any k-edges of H; that is, H need not be just the down-closure of a k-uniform hypergraph. This turns out to be required in some applications for $k \geq 3$; for $k = 2$ this extra generality is not interesting.

The proof of Theorem 30 is conceptually divided in four steps. The first one, deals with the special case where $p = 1$, H is the complete k-graph $K_{v(H)}^{(k)}$, and $\ell = v(H)$ is exactly [6, Lemma

[5]This condition is not necessary for any reason besides formality. We insert this for completeness as we use in the proof [6, Lemma 4]. Any similar result without this condition would extend to our setting.

4]. The second step is to drop the assumption that H is the complete k-graph $K^{(k)}_{v(H)}$ (keeping the assumptions $p = 1$, and $\ell = v(H)$). This step requires only a few lines of explanation, which we now provide. Indeed, what [6, Lemma 4] allows us to do is to count ϕ-partite copies of the complete k-graph over $v(H)$ in any given k-partition. Imagine now we want to count ϕ-partite copies of H in the k-partition $\mathcal{V}$ for some H that is not the complete graph. What we can do, is to form a new partition $\mathcal{V}'$ by adding all possible supported edges to $V_{\phi(e)}$ for each $e \notin E(H)$. That is, for each such e, we let $V_{\phi(e)}$ in $\mathcal{V}'$ consist of all k-sets supported by the relevant lower-level graphs. Under this modification, the number of ϕ-partite copies of H in $\mathcal{V}$ becomes equal to the number of ϕ-partite copies of the complete k-graph $K^{(k)}_{v(H)}$ in $\mathcal{V}'$, which is counted precisely by [6, Lemma 4]. The next step is to drop the condition $\ell = v(H)$, which requires a bit more care. The final step, dropping the condition $p = 1$, is where we actually make use of our transference principle.

Proof of Theorem 30, $p = 1$. Let $\varepsilon_k > 0$ be small enough for the $\ell = v(H)$ case of Theorem 30 with input $\frac{1}{2}\delta$ (which is given by our previous step and [6, Lemma 4]). Given $d_2, \dots, d_k$ (such that $1/d_i \in \mathbb{N}$), let $\varepsilon > 0$ and $r \in \mathbb{N}$ be returned by the $\ell = v(H)$ case for the same input. Suppose $v(H)N$ is sufficiently large for this case with a final input $\frac{1}{v(H)} d_1$.

Given ℓ and $\mathcal{V}$ as in the statement of Theorem 30, let $\mathcal{V}'$ on vertex set $[v(H)N]$ be obtained from $\mathcal{V}$ by, for each $i \in [\ell]$, taking $\left|\phi^{-1}(i)\right|$ copies of $V_{\{i\}}$ and adding all incident edges between them. Note that the increased size of the vertex set is sufficient to contain all these copies. Letting the clusters of $\mathcal{V}'$ be indexed by $[v(H)]$, let $\phi' : V(H) \to [v(H)]$ be an injective map sending each $x \in V(H)$ to a copy of $V_{\{\phi(x)\}}$.

Now, the ϕ-partite copies of H in $\mathcal{V}$ and ϕ'-partite copies of H in $\mathcal{V}'$ are almost in one-to-one correspondence: the difference is that some ϕ'-partite copies of H in $\mathcal{V}'$ do not correspond to injective maps to $\mathcal{V}$. However, there can be at most $\binom{v(H)}{2}(v(H)N)^{v(H)-1}$ such copies, so applying the known case of Theorem 30 we conclude that the number of ϕ-partite copies of H in $\mathcal{V}$ is

$$\begin{aligned}\left(1 \pm \tfrac{1}{2}\delta\right) N^{v(H)} \prod_{e \in E(H)} d^*\left(V'_{\phi' e)}\right) \pm \binom{v(H)}{2}(v(H)N)^{v(H)-1} \\ = (1 \pm \delta) N^{v(H)} \prod_{e \in E(H)} d^*\left(V_{\phi(e)}\right).\end{aligned}$$

as required, where the equality uses the fact that N is sufficiently large. The fact that the vertex set of $\mathcal{V}'$ has size $v(H)N$ is exactly cancelled by the corresponding decrease by a factor $v(H)$ in each $d^*\left(\{i\}\right)$. $\square$

Finally we use Theorem 3 to deduce the general case.

Proof of Theorem 30. The case $e_k(H) = 0$ of Theorem 30 is precisely the $p = 1$ case viewing H as a $(k-1)$-complex.

The case $e_k(H) = 1$ is standard and does not require Theorem 3. We give only a sketch. Letting H' be the $(k-1)$-complex H with the one k-edge removed. An application of the Extension Lemma [6, Lemma 5] shows that all but a tiny fraction of k-sets supported by any given $\mathcal{V}'$ a $(k-1)$-partition are in roughly the same number of ϕ-partite copies of H', and that the exceptional k-sets account for only a tiny fraction of all ϕ-partite copies of H'. A standard application of Chernoff's inequality shows that with very high probability, when $G^{(k)}_N$ is revealed, there are very few edges on these exceptional k-sets and the number of ϕ-partite

H-copies they generate is tiny compared to those on typical k-sets. Critically, this 'very high probability' is sufficient for a union bound over choices of $\mathcal{V}'$ and ϕ. Supposing now this likely event occurs, given any regular $\mathcal{V}$, letting $\mathcal{V}'$ the the $(k-1)$-partition obtained by removing the k layer, we see that (using the fact that ε_k is much smaller than d_k) most of the k-edges of $\mathcal{V}$ are on typical k-sets and a short calculation gives the desired count of ϕ-partite H-copies.

Given H with $e_k(H) \geq 2$ and $\delta > 0$, let $2\varepsilon_k > 0$ be small enough for the $p = 1$ case of Theorem 30 with input $\frac{1}{2}\delta$. Given $d_2, \ldots, d_k > 0$, let ε and r be given by the $p = 1$ case of Theorem 30 for input $d_2, \ldots, d_{k-1}, \frac{1}{2}d_k$. Let finally $d_1 > 0$ be given.

We set $c = 2v(H)!$, and apply Theorem 3 with input $k = e_k(H)^6$, c and error parameter

$$\varepsilon^* = \tfrac{\delta d_k \varepsilon_k^2}{10v(H)!} \prod_{e \in E(H)} d_{|e|} \,.$$

Let C be the constant returned by Theorem 3. Order arbitrarily the k-edges of H. Let $n = \binom{N}{k}$ enumerate the edges of $K_N^{(k)}$, and let S consist of the ordered subsets of $[n]$ corresponding to $e_k(H)$-sets in $[N]$ which form isomorphic copies of the k-uniform edges of H, in the chosen order.

Let $C^* = 10rkCk!$. We now verify the maximum degree condition on S holds for $p \geq C^* n^{-1/m_k(H)}$. To begin with, we estimate $e(S)$. Let $q(H)$ be the number of vertices of H which are not in any k-uniform edge of H. There are $(1+o(1))N^{v(H)-q(H)}$ injective maps from the vertices of H which are in k-edges to $[N]$, each of which gives one element of S, so $e(S) = (1+o(1))N^{v(H)-q(H)}$.

Given $1 \leq \ell \leq e_k(H)$, let $\mathbf{x}$ be a sequence of length $e_k(H)$ from $[n] \cup \{*\}$ with exactly ℓ entries not equal to $*$. For $\ell = 1$, by symmetry we have $\deg_S(\mathbf{x}) = \frac{e(S)}{n}$, which is as required. We now assume $\ell \geq 2$. Let $W \subseteq [N]$ be the vertices of $K_N^{(k)}$ which are contained in some edge in $\mathbf{x}$. By definition, if $\mathbf{x}$ has two identical non-$*$-entries, then $\deg_S(\mathbf{x}) = 0$, so we can assume that $\mathbf{x}$ has at least two distinct non-$*$ entries, and hence $|W| \geq k+1$. By definition of $m_k(H)$, we have

$$\tfrac{\ell-1}{|W|-k} \leq m_k(H)\,, \qquad \text{so} \qquad |W| \geq \tfrac{\ell-1}{m_k(H)} + k\,.$$

To obtain a member of S which agrees with $\mathbf{x}$ at the non-$*$ coordinates, we can at most pick a further $v(H) - |W| - q(H)$ vertices in k-edges of H and one of the at most $v(H)!$ maps from the vertices of H in k-edges to the picked vertices together with W. Thus, we have

$$\begin{aligned}
\deg_S(\mathbf{x}) &\leq N^{v(H)-|W|-q(H)} v(H)! \\
&\leq 2v(H)! e(S) N^{-|W|} \\
&\leq 2v(H)! e(S) N^{-\frac{\ell-1}{m_k(H)}-k} \\
&= 2v(H)! \tfrac{e(S)}{N^k} \big(N^{-1/m_k(H)}\big)^{\ell-1} \\
&\leq 2v(H)! \tfrac{e(S)}{n} \big(p/C^*\big)^{\ell-1}\,,
\end{aligned}$$

which is the required bound.

By construction, there can be at most $2^{kN^{k-1}}$ possible sets $R(Q_1, \ldots, Q_k)$ with the condition that $Q_1, \ldots, Q_k$ are disjoint subsets of $\binom{[N]}{k-1}$. Let Σ consist of the indicator functions of the

[6] This is bad notation, but k is only used in this proof as in the statement of Theorem 30, we have it here $k = e_k(H)$ because k also has a meaning in Theorem 3.

unions of any up to r sets of the form $R(Q_1,\dots,Q_k)$. Then we have

$$|\Sigma| \le 2\cdot 2^{rkN^{k-1}} \le \exp\left(\tfrac{pn}{C}\right),$$

where the inequality uses $p \ge C^*N^{-1}$ and the choice of C^*.

For each $k \le \ell \le v(H)$, consider each choice of a k-partition $\mathcal{V}$ with ℓ clusters whose k level is complete (i.e. each V_E with $|E| = k$ is equal to $R(W_1,\dots,W_k)$ where $W_1,\dots,W_k$ are the supporting $(k-1)$-graphs), and each $\phi : v(H) \to [\ell]$. For each such $(\ell,\mathcal{V},\phi)$ we construct a subcount ω as follows. For each member s of S, we count the number $w(s)$ of ϕ-partite copies ψ of H in $\mathcal{V}$ such that the i-th edge of H is mapped to the i-th member of s, for each $1 \le i \le e_k(H)$. Observe that necessarily $0 \le w(s) \le (v(H))!N^{q(H)}$, where $q(H)$ is the number of vertices of H not in any k-edge of H. We define $\omega(s) = \frac{1}{(v(H))!N^{q(H)}}w(s)$, which is therefore in $[0,1]$. We say this is the subcount corresponding to $(\ell,\mathcal{V}'',\phi)$ for any choice $\mathcal{V}''$ of a k-partition which is identical to $\mathcal{V}$ on any level except perhaps the k-th. We now upper bound the size $|\Omega|$ of the set of all such subcounts. There are $v(H)$ choices of ℓ, and at most $v(H)^\ell \le v(H)^{v(H)}$ choices of ϕ. What remains is to estimate the number of choices of $\mathcal{V}$. Observe that $\mathcal{V}$ is defined by the choices of V_E for $1 \le |E| \le k-1$. There are at most $N^{v(H)+1}$ ways to choose the clusters, since the clusters are disjoint. Again since the clusters are disjoint, to define V_E for each $2 \le |E| \le k-1$ it suffices to choose a subset of each of $\binom{[N]}{2}$ through $\binom{[N]}{k-1}$, which can be done in at most $2^{N^2}\cdots 2^{N^{k-1}}$ ways. We conclude

$$|\Omega| \le v(H)^{v(H)+1}N^{v(H)+1}2^{kN^{k-1}} \le \exp\left(\tfrac{pn}{C}\right),$$

where as before the inequality uses $p \ge C^*N^{-1}$ and the choice of C^*, and this time also that N is sufficiently large.

Suppose now that $X = [n]_p$ satisfies the likely event of Theorem 3 for this ε^*, S, Σ and Ω. Let Γ be the corresponding instance of $G^{(k)}(N,p)$.

Given $k \le \ell \le v(H)$ and an $(\varepsilon_k,\varepsilon,d_1,\dots,d_k,r,p)$-regular k-partition $\mathcal{V}$ with ℓ clusters on N, such that for each V_E with $|E| = k$ we have $V_E \subseteq \Gamma$, let Y be the subset of X consisting of elements in any V_E with $|E| = k$. Let Z be the dense model guaranteed by the likely event of Theorem 3, and let $\mathcal{V}'$ be the k-partition with ℓ clusters on N obtained by replacing each V_E where $|E| = k$ with V'_E corresponding to the elements of Z that are supported on the $(k-1)$-graphs supporting V_E.

We claim that $\mathcal{V}'$ is $(2\varepsilon_k,\varepsilon,d_1,\dots,d_{k-1},\frac{1}{2}d_k,r,1)$-regular and that the relative densities of the top level are close to the relative p-densities of $\mathcal{V}$. The regularity of the levels from 1 to $k-1$ follows from the regularity of $\mathcal{V}$, and what needs to be proved is that each V'_E with $|E| = k$ is $(2\varepsilon_k,r,1)$-regular with density $d^*(V'_E) = \left(1 \pm \frac{\delta}{10e_k(H)}\right)d^*_p(V_E) \ge \frac{1}{2}d_k$.

To see this holds, fix E and let $W_1,\dots,W_k$ be the supporting $(k-1)$-graphs of V_E (so also of V'_E). Let R^* be a union of at most r sets of the form $R(Q_1,\dots,Q_k)$ (as defined where we described the set Σ of similarity functions), with the extra condition $Q_i \subseteq W_i$ for each $1 \le i \le k$. Abusing notation slightly, we think of R^* as both a subset of $\binom{[N]}{k}$ and of $[n]$. Because Z is a dense model of Y, using the similarity function σ which takes the value 1 precisely on R^*, we have

$$p^{-1}|R^* \cap Y| = |R^* \cap Z| \pm \varepsilon^* n\,.$$

Taking the particular case that R^* is all k-sets supported by $W_1,\dots,W_k$, this immediately says that

$$d^*(V'_E)=d^*_p(V_e)\pm\varepsilon^* nN^{-k}\prod_{i=1}^{k-1}d_i^{-\binom{k}{i}}=\big(1\pm\tfrac{\delta}{10e_k(H)}\big)d^*_p(V_e)\geq\tfrac{d_k}{2}\,,$$

where the final equality is by choice of ε^*. Suppose now $|R^*|$ contains at least an ε_k-fraction of all k-edges supported by $W_1,\dots,W_k$. Because $\mathcal{V}$ is regular, we have

$$|R^*\cap Y|=|R^*\cap V_E|=(1\pm\varepsilon_k)d^*(V_E)|R^*|=(1\pm\varepsilon_k)pd^*_p(V_E)|R^*|\,.$$

Putting these bits together, we have

$$\begin{aligned}|R^*\cap Z|&=(1\pm\varepsilon_k)d^*_p(V_E)|R^*|\pm\varepsilon^* n\\&=(1\pm\varepsilon_k)\big(1\pm\tfrac{\delta}{10e_k(H)}\big)d^*(V'_E)|R^*|\pm\varepsilon^* n\\&=(1\pm2\varepsilon_k)d^*(V'_E)\end{aligned}$$

which verifies $(2\varepsilon_k,r,1)$-regularity of V'_E. Here again the final inequality is by choice of ε^*.

Applying the $p=1$ case of Theorem 3 to $\mathcal{V}$', we see that the number of ϕ-partite copies of H in $\mathcal{V}'$ is

$$\big(1\pm\tfrac12\delta\big)N^{v(H)}\prod_{e\in E(H)}d^*\big(V'_{\phi(e)}\big)\,.$$

Letting ω be the subcount corresponding to $(\ell,\mathcal{V},\phi)$, we have by definition of ω

$$\begin{aligned}\sum_{s\in S}\mathbb{1}(s\subseteq Z)\omega(s)(v(H))!N^{q(H)}&=\big(1\pm\tfrac12\delta\big)N^{v(H)}\prod_{e\in E(H)}d^*\big(V'_{\phi(e)}\big)\\&=\big(1\pm\tfrac34\delta\big)N^{v(H)}p^{-e(H)}\prod_{e\in E(H)}d^*\big(V_{\phi(e)}\big)\,.\end{aligned}$$

Where the final equality uses that $d^*_p(V_E)=p^{-1}d(V_E)=\big(1\pm\frac{\delta}{8e_k(H)}\big)d^*(V'_E)$ whenever $|E|=k$.

Since Z is a dense model of Y, we have

$$p^{-e_k(H)}\sum_{s\in S}\mathbb{1}(s\subseteq Y)\omega(s)=\sum_{s\in S}\mathbb{1}(s\subseteq Z)\omega(s)\pm\varepsilon^* e(S)\,.$$

We therefore get

$$\begin{aligned}&\big(1\pm\tfrac34\delta\big)N^{v(H)}p^{-e_k(H)}\prod_{e\in E(H)}d^*\big(V_{\phi(e)}\big)\\&\quad=p^{-e_k(H)}\sum_{s\in S}\mathbb{1}(s\subseteq Y)\omega(s)(v(H))!N^{q(H)}\pm\varepsilon^* e(S)(v(H))!N^{q(H)}\\&\quad=p^{-e_k(H)}\sum_{s\in S}\mathbb{1}(s\subseteq Y)\omega(s)(v(H))!N^{q(H)}\pm\varepsilon^* N^{v(H)}(v(H))!\,,\end{aligned}$$

and so

$$\begin{aligned}\sum_{s\in S}\mathbb{1}(s\subseteq Y)\omega(s)(v(H))!N^{q(H)}&=\big(1\pm\tfrac34\delta\big)N^{v(H)}\prod_{e\in E(H)}d^*\big(V_{\phi(e)}\big)\\&\qquad\pm\varepsilon^*(v(H))!N^{v(H)}p^{e_k(H)}\\&=\big(1\pm\delta\big)N^{v(H)}\prod_{e\in E(H)}d^*\big(V_{\phi(e)}\big)\end{aligned}$$

by choice of ε^*. Since the left-hand side of this is, by definition of ω, the number of ϕ-partite copies of H in $\mathcal{V}$, this completes the proof. $\square$

References

[1] József Balogh, Robert Morris, and Wojciech Samotij. "Independent sets in hypergraphs". In: *J. Amer. Math. Soc.* 28.3 (2015), pp. 669–709.

[2] Rajendra Bhatia and Chandler Davis. "A better bound on the variance". In: *Amer. Math. Monthly* 107.4 (2000), pp. 353–357.

[3] Theo Bühler and Dietmar A. Salamon. *Functional analysis.* Vol. 191. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2018, pp. xiv+466.

[4] D. Conlon and W. T. Gowers. "Combinatorial theorems in sparse random sets". In: *Ann. of Math. (2)* 184.2 (2016), pp. 367–454.

[5] D. Conlon et al. "On the KŁR conjecture in random graphs". In: *Israel J. Math.* 203.1 (2014), pp. 535–580.

[6] Oliver Cooley et al. "Embeddings and Ramsey numbers of sparse k-uniform hypergraphs". In: *Combinatorica* 29.3 (2009), pp. 263–297.

[7] P. Erdős and M. Simonovits. "A limit theorem in graph theory". In: *Studia Sci. Math. Hungar.* 1 (1966), pp. 51–57.

[8] P. Erdős and A. H. Stone. "On the structure of linear graphs". In: *Bull. Amer. Math. Soc.* 52 (1946), pp. 1087–1091.

[9] Paul Erdős and Miklós Simonovits. "Supersaturated graphs and hypergraphs". In: *Combinatorica* 3.2 (1983), pp. 181–192.

[10] Ben Green and Terence Tao. "The primes contain arbitrarily long arithmetic progressions". In: *Ann. of Math. (2)* 167.2 (2008), pp. 481–547.

[11] T. E. Harris. "A lower bound for the critical probability in a certain percolation process". In: *Proc. Cambridge Philos. Soc.* 56 (1960), pp. 13–20.

[12] Jeong Han Kim and Van H. Vu. "Concentration of multivariate polynomials and its applications". In: *Combinatorica* 20.3 (2000), pp. 417–434.

[13] Oleg Pikhurko and Zelealem B. Yilma. "Supersaturation problem for color-critical graphs". In: *J. Combin. Theory Ser. B* 123 (2017), pp. 148–185.

[14] Walter Rudin. *Functional analysis.* Second. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, 1991, pp. xviii+424.

[15] Walter Rudin. *Principles of mathematical analysis.* Third. International Series in Pure and Applied Mathematics. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976, pp. x+342.

[16] David Saxton and Andrew Thomason. "Hypergraph containers". In: *Invent. Math.* 201.3 (2015), pp. 925–992.

[17] Reto Spöhel, Angelika Steger, and Lutz Warnke. "General deletion lemmas via the Harris inequality". In: *J. Comb.* 4.2 (2013), pp. 251–271.

[18] Roman Vershynin. *High-dimensional probability.* Vol. 47. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2018, pp. xiv+284.

(PA) London School of Economics, Department of Mathematics, Houghton Street, London WC2A 2AE, UK
Email address: p.d.allen@lse.ac.uk

(JB) London School of Economics, Department of Mathematics, Houghton Street, London WC2A 2AE, UK
Email address: j.boettcher@lse.ac.uk

(JL) London School of Economics, Department of Mathematics, Houghton Street, London WC2A 2AE, UK
Email address: j.m.lada@lse.ac.uk

(DMC) London School of Economics, Department of Mathematics, Houghton Street, London WC2A 2AE, UK
Email address: d.mergoni@lse.ac.uk