AN EFFICIENT CONTAINER LEMMA

JÓZSEF BALOGH AND WOJCIECH SAMOTIJ

ABSTRACT. We prove a new, efficient version of the hypergraph container theorems that is suited for hypergraphs with large uniformities. The main novelty is a refined approach to constructing containers that employs simple ideas from high-dimensional convex geometry. The existence of smaller families of containers for independent sets in such hypergraphs, which is guaranteed by the new theorem, allows us to improve upon the best currently known bounds for several problems in extremal graph theory, discrete geometry, and Ramsey theory. One of the applications of our efficient container lemma, a structural characterisation of n-vertex graphs with clique number $o(\log n/\log\log n)$, suggests that the new lemma is nearly best-possible for general hypergraphs of large uniformities.

1. INTRODUCTION

The hypergraph container theorems, proved several years ago by Balogh, Morris, and Samotij [6] and, independently, by Saxton and Thomason [42], state that the family of independent sets of any uniform hypergraph whose edges are distributed somewhat evenly may be covered by (the families of subsets of) a small collection of sets, called containers, each of which is nearly independent. The original motivation for these results were several specific questions concerning enumeration of graphs avoiding a given subgraph and of sets of integers contining no arithmetic progressions of a given length. (The idea of considering these problems in the general context of independent sets in hypergraphs had been successfully pursued earlier in the breakthrough works of Conlon–Gowers [14] and Schacht [44] on extremal properties of random graphs and random sets of integers.) However, over the years, the scope of applicability of the container theorems has grown quite substantially (see, for example, the survey [7] and references therein). The two major reasons for this are the general form of the theorems (many problems can be cast in the language of independent sets in auxiliary hypergraphs) and the explicit, optimal dependence between the various parameters disguised under the vague phrases 'small collection', 'evenly distributed', and 'nearly independent' above.

The vast majority of applications of the container theorems concern sequences of hypergraphs of fixed uniformity and growing order and size. As a result, the explicit dependence of the various parameters involved in the statements of the container theorems on the uniformity of the hypergraph is merely a minor detail in all these works. Recently, however, the container theorems have been used to analyse sequences of hypergraphs whose uniformities grow with the numbers of vertices and edges. In these applications of the container method to questions in Ramsey theory [12, 40], discrete

---

geometry [9], and extremal graph theory [5, 33], the explicit dependence between uniformity and other parameters turned out to lie at the heart of the matter, obstructing the way to obtaining optimal bounds for several well studied functions. It is only fair to note here that this dependence is more favourable in the version of the theorem proved by Saxton and Thomason [42]. Having said that, the two constructions of containers presented in [6] and [42] are essentially equivalent and the differences between the final results reflect merely the differences in their analyses. This analysis was performed more carefully, and with a wiser choice of parameters, by the authors of [42].

The basic container lemma, which is the building block of both proofs that really lies at the heart of the matter, is a statement that asserts the existence of a small family $\mathcal{C}$ of containers for independent sets of an s-uniform hypergraph $\mathcal{H}$ that satisfies $|C| \leqslant (1 - \delta)v(\mathcal{H})$ for every $C \in \mathcal{C}$ and some positive constant δ; see [6, Proposition 3.1] and [42, Theorem 3.4]. The stronger form of the theorem described in the first paragraph is then derived by recursively applying this basic lemma to the subhypergraphs of $\mathcal{H}$ induced by the sets $C \in \mathcal{C}$ as long as C still contains many edges of $\mathcal{H}$. The caveat here is that the proof methods used in both [6] and [42] necessarily yield $\delta \leqslant 1/s!$. (The short, non-algorithmic proof of the basic container lemma given recently by Bernshteyn, Delcourt, Towsner, and Tserunyan [10] seems to yield δ that is doubly-exponentially small in s.) Since one typically requires the ratio $|C|/v(\mathcal{H})$ to be bounded away from one for each final container C, at least $\exp\big(\Omega(s \log s)\big)$ iterations are required; this substantially blows up the final number of containers when s is no longer a fixed constant. Finally, we remark that a different method of building containers for independent sets in hypergraphs was proposed and analysed by Saxton and Thomason [43]. Even though the parameter δ in the basic container lemma proved in [43] is only polynomially small in the uniformity, the upper bound on the number of containers is far from optimal. Moreover, the lemma applies only to simple hypergraphs (i.e., hypergraphs whose every pair of vertices is contained in at most one edge) whereas the hypergraphs considered in most applications of the container method are far from being simple.

The main result of this work is a new, more efficient version of the basic container lemma in which the parameter δ is only polynomially small in the uniformity. We postpone stating the strongest form of our new lemma until Section 2 and state here only its corollary that can be easily compared with [6, Proposition 3.1]. Following the notational convention of [6], given a nonempty s-uniform hypergraph $\mathcal{H}$, we shall denote the numbers of its vertices and edges by $v(\mathcal{H})$ and $e(\mathcal{H})$, respectively. Moreover, for every $T \subseteq V(\mathcal{H})$, we define

$$\deg_{\mathcal{H}} T = |\{A \in E(\mathcal{H}) : T \subseteq A\}|$$

and, for every $t \in \{1, \ldots, s\}$, we let

$$\Delta_t(\mathcal{H}) = \max\big\{ \deg_{\mathcal{H}} T : T \subseteq V(\mathcal{H}) \text{ and } |T| = t \big\}.$$

The following theorem, an efficient basic container lemma, is a simplified version of our main technical result, Theorem 2.1 below.

Theorem 1.1. *Let s be a positive integer and let $\mathcal{H}$ be a nonempty s-uniform hypergraph. Suppose that $q \in (0, 1)$ and $K > 0$ are such that $q \cdot v(\mathcal{H}) \geqslant 10^8 s^6 K$ and, for every*

$t \in \{1, \ldots, s\}$,

$$\Delta_t(\mathcal{H}) \leqslant K \cdot \left(\frac{q}{10^6 s^5}\right)^{t-1} \cdot \frac{e(\mathcal{H})}{v(\mathcal{H})}. \tag{1}$$

Then, there exist a family $\mathcal{S} \subseteq \binom{V(\mathcal{H})}{\leqslant q \cdot v(\mathcal{H})}$ and functions $f \colon \mathcal{S} \to \mathcal{P}(V(\mathcal{H}))$ and $g \colon \mathcal{I}(\mathcal{H}) \to \mathcal{S}$ such that, for every $I \in \mathcal{I}(\mathcal{H})$,

$$g(I) \subseteq I \subseteq g(I) \cup f(g(I)) \qquad and \qquad |f(g(I))| \leqslant (1 - \delta) \cdot v(\mathcal{H}),$$

where $\delta = (10^3 s^4 K)^{-1}$. Moreover, if $g(I) \subseteq I'$ and $g(I') \subseteq I$ for some $I, I' \in \mathcal{I}(\mathcal{H})$, then $g(I) = g(I')$.

Qualitatively, Theorem 1.1 is identical to the original basic container lemmas [6, Proposition 3.1] and [42, Theorem 3.4]. Quantitively, however, it is a significant improvement of these results. In order to demonstrate this, we shall present four applications of our new theorem to problems in extremal graph theory, discrete geometry, and Ramsey theory that had previously been attacked using the original container theorems, obtaining an essential improvement of the state-of-the-art result in each case. We discuss these applications and the relevant background in detail in the next three sections. We just point out here that one of these applications, an efficient version of the classical theorem of Kolaitis, Prömel, and Rothschild [28], Theorem 1.2 below, strongly suggests that Theorem 1.1 is, up to lower-order terms, optimal for general hypergraphs of large uniformities (see the discussion below the statement of Theorem 1.2).

1.1. **The typical structure of graphs with no large cliques.** Given graphs G and H, we say that G is H-free if G does not contain H as a (not necessarily induced) subgraph. The study of typical properties of H-free graphs goes back to the seminal work of Erdős, Kleitman, and Rothschild [20], who proved that almost all triangle-free graphs are bipartite.[1] We shall consider here almost exclusively the case where H is a clique; for a wealth of information regarding other graphs, we refer the reader to [4]. The result of [20] was generalised by Kolaitis, Prömel, and Rothschild [28], who showed that, for every fixed $r \geqslant 2$, almost all K_{r+1}-free graphs are r-partite.

Let us point out that this 'structural' characterisation of a typical K_{r+1}-free graph is in fact an enumeration result in disguise. Since every r-partite graph is K_{r+1}-free, the main result of [28] is that the number of K_{r+1}-free subgraphs of K_n is asymptotically equal to the number of r-partite subgraphs of K_n. Taking this point of view, one can say that the result of [28] was anticipated by the aforementioned work of Erdős, Kleitman, and Rothschild [20], who also showed that, for every fixed r, there are $2^{\text{ex}(n,K_{r+1})+o(n^2)}$ many K_{r+1}-free subgraphs of K_n.[2] This estimate was generalised by Erdős, Frankl, and Rödl [17], who proved that, for every fixed graph H, there are $2^{\text{ex}(n,H)+o(n^2)}$ many H-free subgraphs of K_n.

In the above discussion, we have tried to stress that the forbidden graph H is fixed whereas n, the number of vertices in the host graphs, tends to infinity. Much less is known if one allows the order/size of H to grow with n. This more general question was considered only fairly recently by Bollobás and Nikiforov [11]; however, the results

---

[1]That is, the probability that a uniformly random K_3-free subgraph of K_n is bipartite tends to one as n tends to infinity.

[2]We write $\text{ex}(n, H)$ to denote the Turán number of a graph H, that is, the largest number of edges in an H-free graph with n vertices.

of [11] are only meaningful when the chromatic number of H stays bounded. A few years later, Mousset, Nenadov, and Steger [33], extending the Erdős–Kleitman–Rothschild bound, proved that there are $2^{\mathrm{ex}(n,K_{r+1})+o(n^2/r)}$ many K_{r+1}-free subgraphs of K_n as long as $r \leqslant (\log n)^{1/4}/2$.[3] Somewhat later, Balogh, Bushaw, Collares, Liu, Morris, and Sharifzadeh [5] strengthened this result considerably by showing that, under the slightly weaker assumption $2 \leqslant r \leqslant (\log n)^{1/4}$, almost all K_{r+1}-free subgraphs of K_n are r-partite. (Both [33] and [5] relied on the original hypergraph container theorems.) Our first application of the new, efficient container lemma is the following strengthening of this result.

Theorem 1.2. *If a function* $r \colon \mathbb{N} \to \mathbb{N}$ *satisfies* $2 \leqslant r(n) \leqslant \log n/(121 \log \log n)$, *then almost all* K_{r+1}-*free subgraphs of* K_n *are* r-*partite.*

We point out that the assumption on the growth rate of r in Theorem 1.2 is nearly optimal. Indeed, a standard first-moment calculation shows that, for every positive constant ε, a uniformly random subgraph $G \subseteq K_n$ contains no clique with $\lfloor (2+\varepsilon) \log_2 n \rfloor$ vertices whereas $\chi(G) \geqslant \Omega(n/\log n)$. On the other hand, it may well be that the assertion of the theorem remains true as long as $r(n) \leqslant (2-\varepsilon) \log_2 n$ for some positive constant ε. However, even removing the doubly-logarithmic term from the denominator in the assumed upper bound on $r(n)$ will likely require significantly new ideas.

1.2. **Lower bounds on ε-nets.** Suppose that X is a finite set and let $\mathcal{R}$ be an arbitrary collection of subsets of X. For a positive number ε, an ε-*net* in $\mathcal{R}$ is any set $N \subseteq X$ that intersects every element of $\mathcal{R}$ with cardinality at least $\varepsilon|X|$. In other words, N is an ε-net if $N \cap A \neq \emptyset$ for every $A \in \mathcal{R}$ with $|A| \geqslant \varepsilon|X|$.

One is usually interested in finding a small ε-net. However, this is not always possible. For example, if $\mathcal{R}$ comprises all subsets of X, then every ε-net in $\mathcal{R}$ must have more than $(1-\varepsilon)|X|$ elements. One can rule out such 'pathological' examples by imposing a natural assumption on a measure of complexity of the family $\mathcal{R}$ called the VC dimension. We say that a set S is *shattered* by a family $\mathcal{R}$ if $\{A \cap S : A \in \mathcal{R}\}$ contains all $2^{|S|}$ subsets of S. The *VC dimension* (a shorthand for *Vapnik–Chervonenkis dimension*) of $\mathcal{R}$ is the largest cardinality of a set that $\mathcal{R}$ shatters. A seminal result of Haussler and Welzl [26] states that every family of subsets whose VC dimension is at most d admits an ε-net with at most $\lceil (8d/\varepsilon) \log(8d/\varepsilon) \rceil$ elements, for every $\varepsilon > 0$. Komlós, Pach, and Woeginger [29] improved this upper bound on the smallest size of an ε-net to $\big(d + o(1)\big) \cdot (1/\varepsilon) \log(1/\varepsilon)$, where $o(1)$ denotes some function tending to zero with ε. Moreover, they constructed, for every $d \geqslant 2$, (random) families with VC dimension d that have no ε-net smaller than $\big(d - 2 + 2/(d+1) - o(1)\big) \cdot (1/\varepsilon) \log(1/\varepsilon)$.

On the other hand, it was proved that various set families arising in geometry admit ε-nets of cardinality merely $O(1/\varepsilon)$, see [29, 31]. In view of this, many researchers believed that in 'geometric scenarios' (with bounded VC dimension), there always exists an ε-net of size $O(1/\varepsilon)$. This belief was shown to be wrong by Alon [1], who proved that, for arbitrary small ε, there are finite sets X of points in the plane such that every ε-net for the family comprising the intersections of X with straight lines (the *range space* of lines on X) must have at least $(1/\varepsilon) \cdot \omega(1/\varepsilon)$ points, for some (very slowly growing) function ω

---

[3]The $o(n^2/r)$ error term is natural here as $\mathrm{ex}(n, K_{r+1}) = \binom{n}{2} - \Theta(n^2/r)$.

with $\lim_{x\to\infty} \omega(x) = \infty$. Alon speculated that there are planar sets of points X for which the factor $\omega(1/\varepsilon)$ in the above statement could be replaced by $\Omega\big(\log(1/\varepsilon)\big)$.

In a paper that served as the main motivation for this work, Balogh and Solymosi [9] showed that, for arbitrarily small $\varepsilon > 0$, there are sets $X \subseteq \mathbb{R}^2$ such that the range space of lines on X does not have ε-nets with fewer than $(1/\varepsilon)\big(\log(1/\varepsilon)\big)^{1/3-o(1)}$ points; their proof relied on the hypergraph container theorems. We review the construction of Balogh and Solymosi [9] and, using our new, efficient container lemma, we further improve their lower bound, replacing the constant $1/3$ in the exponent with $1/2$.

Theorem 1.3. *The following holds for every $\varepsilon_0 > 0$. There exists an $\varepsilon \in (0, \varepsilon_0)$ and a finite set $X \subseteq \mathbb{R}^2$ such that the smallest size of an ε-net for the family of intersections of straight lines with X is at least*

$$\frac{1}{80\varepsilon} \cdot \sqrt{\frac{\log(1/\varepsilon)}{\log\log(1/\varepsilon)}}.$$

We should mention here that, several years prior to [9], Pach and Tardos [36] showed that the families defined by intersections of finite point sets with axis parallel rectangles (in $\mathbb{R}^2$) and axis-parallel boxes in $\mathbb{R}^4$ may require ε-nets of sizes $\Omega\big((1/\varepsilon)\log\log(1/\varepsilon)\big)$ and $\Omega\big((1/\varepsilon)\log(1/\varepsilon)\big)$, respectively; both these lower bounds are tight up to multiplicative constants, see [3].

1.3. **Upper bounds on Ramsey numbers.** Given graphs G and H and a positive integer k, we write $G \to (H)_k$, and say that G *is Ramsey for H in k colours*, if every k-colouring of the edges of G contains a monochromatic copy of H. In other words, $G \to (H)_k$ if, for every $c\colon E(G) \to [\![k]\!]$, there is some $i \in [\![k]\!]$ such that the graph $c^{-1}(i)$ contains H as a subgraph.[4] The famous theorem of Ramsey [38] states that, for all positive integers n and k, there is an integer N such that $K_N \to (K_n)_k$; we shall denote the smallest such N, the *k-colour Ramsey number of K_n*, by $R(n; k)$. It is well-known that $R(n; k) \leqslant (kn)!/(n!)^k \leqslant k^{kn}$, see [19, 25].

Fifty years ago, Folkman [21] proved that, for every n, there exists a graph G such that $G \not\supseteq K_{n+1}$ but, nevertheless, $G \to (K_n)_2$ and Nešetřil and Rödl [35] generalised this result to an arbitrary number of colours. Define the *k-colour Folkman number for K_n* by

$$F(n; k) = \min\big\{N \in \mathbb{N} : G \to (K_n)_k \text{ for some } K_{n+1}\text{-free } G \subseteq K_N\big\}.$$

The constructions given in [21, 35] yielded upper bounds on $F(n; k)$ that are tower functions of height polynomial in n and k. On the other hand, the strongest lower bound on $F(n; k)$, due to Lefmann [30], is only exponential in kn. In the recent years, the transference theorems of Conlon–Gowers [14] and Schacht [44] (see also [23]) were used by Rödl, Ruciński, and Schacht [41] and by Conlon and Gowers (unpublished) to give improved upper bounds on $F(n; k)$ that were merely doubly-exponential in n and k. Soon afterwards, the first of these two groups of authors [40] used the hypergraph container theorems to give the first exponential bound $F(n; k) \leqslant \exp\big(O(n^4 \log n + n^3 k \log k)\big)$. Our next application of the efficient container lemma is the following improvement of this result.

---

[4]Throughout the paper, we write $[\![k]\!]$ as a shorthand for $\{1, \ldots k\}$.

Theorem 1.4. *There exists a constant C such that, for all positive integers n and k,*

$$F(n;k) \leqslant \big(CknR(n;k)\big)^{21n^2} \leqslant \exp\big(Ckn^3 \log k\big).$$

Another well studied variation of the classical Ramsey numbers are induced Ramsey numbers. Given graphs G and H and a positive integer k, we write $G \to_{\mathrm{ind}} (H)_k$, and say that *$G$ is induced-Ramsey for H in k colours*, if every k-colouring of the edges of G contains a monochromatic induced copy of H. In other words, $G \to_{\mathrm{ind}} (H)_k$ if, for every $c \colon E(G) \to [\![k]\!]$, there are an $i \in [\![k]\!]$ and an injection $\varphi \colon V(H) \to V(G)$ such that $\varphi\big(E(H)\big) \subseteq c^{-1}(i)$ and $\varphi\big(E(H)^c\big) \cap E(G) = \emptyset$. The existence of induced-Ramsey graphs for every H and any number of colours k was established, independently, by Deuber [15], by Erdős, Hajnal, and Pósa [18], and by Rödl [39]. We may thus define the *k-colour Ramsey number of H* by

$$R_{\mathrm{ind}}(H;k) = \min\big\{N \in \mathbb{N} : G \to_{\mathrm{ind}} (H)_k \text{ for some } G \subseteq K_N\big\}.$$

The upper bounds on $R_{\mathrm{ind}}(H;k)$ implied by the constructions of [15, 18, 39] were enormous. In spite of that, Erdős [16] conjectured that, for every n-vertex graph H, the 2-colour induced Ramsey number $R_{\mathrm{ind}}(H;2)$ is only exponential in n. The best-known result to date was obtained by Conlon, Fox, and Sudakov [13], who proved that $R_{\mathrm{ind}}(H;2) \leqslant \exp\big(O(n \log n)\big)$ for every n-vertex graph H. However, the method of [13] does not work when the number of colours is larger than two. The strongest general upper bound for k-colour induced Ramsey numbers in the case $k > 2$ that can be found in the literature is due to Fox and Sudakov [22], who showed that $R_{\mathrm{ind}}(H;k) \leqslant \exp(C_k n^3)$ for every n-vertex H, where C_k depends only on k. However, Fox (private communication) informed us that the methods of [22], which were optimised for sparse graphs H, may be used to prove that $R_{\mathrm{ind}}(H;k) \leqslant \exp(Ckn^2 \log k)$. Our final application of the efficient container lemma is a short derivation of this bound.

Theorem 1.5. *There exists a constant C such that, for every positive integer k and every n-vertex graph H,*

$$R_{\mathrm{ind}}(H;k) \leqslant \big(Cn^2 k R(n;k)\big)^{7n} \leqslant \exp\big(Ckn^2 \log k\big).$$

Finally, let us mention that Conlon, Dellamonica Jr., La Fleur, Rödl, and Schacht [12] used the original container theorems to prove strong bounds on the induced Ramsey numbers of uniform hypergraphs.

1.4. **Packaged statement.** Each of the four illustrations of Theorem 1.1 presented in this paper requires iterative/recursive applications of the theorem. In order to save ourselves (and the reader) from repeating similar, routine arguments and calculations several times, it will be convenient for us to work with the following 'packaged' version of the theorem that is analogous to [6, Theorem 2.2] and [42, Corollary 3.6].

Theorem 1.6. *Let s be a positive integer and let $\mathcal{H}$ be a nonempty s-uniform hypergraph. Suppose that $\alpha, \beta, q \in (0,1)$ and $E \geqslant v(\mathcal{H})$ are such that $\alpha\beta q \cdot v(\mathcal{H}) \geqslant 10^9 s^7$ and $10^4 s^5 q \leqslant \beta$ and, for every $t \in \{2, \ldots, s\}$,*

$$\Delta_t(\mathcal{H}) \leqslant \Big(\frac{q}{10^6 s^5}\Big)^{t-1} \cdot \frac{E}{v(\mathcal{H})}.$$

Then there is a family $\mathcal{C} \subseteq \mathcal{P}(V(\mathcal{H}))$ of at most $\exp\big(10^4 s^5 \beta^{-1} \log(e/\alpha) \cdot q \log(e/q) \cdot v(\mathcal{H})\big)$ sets such that:

(i) *For every $I \in \mathcal{I}(\mathcal{H})$, there is a $C \in \mathcal{C}$ such that $I \subseteq C$.*

(ii) *For every $C \in \mathcal{C}$, either $|C| \leqslant \alpha v(\mathcal{H})$ or there is a subset $W \subseteq C$ with $|W| \geqslant (1 - \beta)|C|$ such that $e(\mathcal{H}[W]) < E$.*

The derivation of Theorem 1.6 from Theorem 1.1 is presented in Section 2.4

1.5. **Organisation of the paper.** The remainder of this paper is organised as follows. In Section 2, we introduce the crucial concept of degree measures, state our main technical result, Theorem 2.1, and derive from it Theorems 1.1 and 1.6. Section 3 is devoted to establishing key properties of degree measures; these properties are used in the subsequent Section 4, which contains the proof of Theorem 2.1. Two probabilistic inequalities needed for the four applications of our new container lemma are stated in Section 5. Finally, Section 6 is devoted to the proof of Theorem 1.2, Section 7 gives the proof of Theorem 1.3, and Section 8 contains proofs of Theorems 1.4 and 1.5.

1.6. **Acknowledgement.** First of all, we are indebted to Rob Morris, David Saxton, and Andrew Thomason for sharing their numerous insights about the container theorems that had a strong bearing on this work. The notion of degree measures, which is central to our approach here, as well as the important idea of allowing hypergraphs to have multiple edges were first introduced by Andrew Thomason and David Saxton [42]. Additionally, we would like to thank Noga Alon for his comments and suggestions regarding lower bounds on ε-nets. We are also indebted to the anonymous referee for their extremely careful reading of the earlier version of this paper and their helpful comments and suggestions, which saved us from having several embarrassing mistakes in the final version of this work.

The second named author thanks Jacob Fox, Frank Mousset, and Bhargav Narayanan for inspiring discussions about upper-bounding induced Ramsey numbers. Last but not least, the second named author owes his deepest gratitude to Lev Buhovski for an inspiring discussion about high-dimensional convex geometry that laid foundations for Lemma 4.12, which lies at the very heart of the proof of Theorem 2.1.

2. The main technical result

2.1. **A word of motivation.** The key idea behind the proof of the container lemma due to Morris and the authors [6] is to, given an $(r + 1)$-uniform hypergraph $\mathcal{H}$ and an independent set $I \in \mathcal{I}(\mathcal{H})$, consider a sequence of vertices of $\mathcal{H}$ for inclusion in a small 'signature' set S and construct an r-uniform hypergraph $\mathcal{G}$ from the neighbourhoods (link hypergraphs) of those among the considered vertices that belong to I. Crucially, each element of this sequence is allowed to depend only on the intersection of I with the set of its predecessors; this guarantees that $\mathcal{G}$ depends solely on S. Since $\mathcal{G}$ comprises only neighbourhoods of vertices in I, we have $I \in \mathcal{I}(\mathcal{G})$. This facilitates induction on the uniformity of the hypergraph.

Whereas there is essentially one way to define containers for independent sets in a 1-uniform hypergraph, the general description of the inductive step given above leaves plenty of room for manoeuvre. The approach taken in [6] was, roughly speaking, to cap the degrees of all vertices of $\mathcal{G}$ at some predefined value Δ and, at the same time, make sure that $e(\mathcal{G}) \geqslant \beta \Delta v(\mathcal{G})$ for some constant β; this way, the ratio of the maximum and the average degrees of the constructed hypergraph $\mathcal{G}$ remained bounded by a constant.

The advantage of this approach was its relative simplicity. However, this simplicity came at a price; the gap between the maximum and the average degrees was forced to grow by a factor of at least $r + 1$, at each step of the induction (reducing uniformity from $r + 1$ to r).[5] As a result, the crucial parameter δ in the basic container lemma could not exceed $1/s!$, where s is the uniformity of the original hypergraph.

Here, we use a similar high-level inductive strategy. However, we take a refined approach to choosing a sequence of vertices of the $(r + 1)$-uniform $\mathcal{H}$ while constructing the r-uniform $\mathcal{G}$; this yields a much more favourable dependence of the parameter δ on the uniformity s. The key new idea is to abandon the wish to control the maximum degree of $\mathcal{G}$ and instead focus on the ℓ^2-norm of its degree sequence. In other words, we measure hypergraphs with ℓ^2-norms, rather than ℓ^∞-norms, of their degree sequences. Viewing hypergraphs as vectors in high-dimensional Euclidean spaces allows us to reduce the problem of constructing a sequence of vertices to be considered for inclusion in the 'signature' to an elementary problem in convex geometry.

2.2. **Degree measures.** We begin by extending the notion of the *degree measure* of a hypergraph, which was introduced by Saxton and Thomason [42]. For a non-empty r-uniform hypergraph $\mathcal{H}$ with vertex set V and a $t \in [\![r]\!]$, we define the *t-degree measure* of $\mathcal{H}$, denoted by $\sigma_{\mathcal{H}}^{(t)}$, to be the probability distribution on $\binom{V}{t}$, the family of all t-element subsets of V, given by

$$\sigma_{\mathcal{H}}^{(t)}(T) = \deg_{\mathcal{H}} T \cdot \left(\sum_{U \in \binom{V}{t}} \deg_{\mathcal{H}} U \right)^{-1} = \frac{\deg_{\mathcal{H}} T}{\binom{r}{t} \cdot e(\mathcal{H})}.$$

In other words, $\sigma_{\mathcal{H}}^{(t)}$ is the probability distribution induced by the following random experiment. Select an edge A of $\mathcal{H}$ uniformly at random and output a t-element subset $T \subseteq A$ chosen uniformly at random from $\binom{A}{t}$.

Throughout this paper, we shall identify (as we already did in the above definition) the measure $\sigma_{\mathcal{H}}^{(t)}$ with its density (with respect to the counting measure), which we shall view as an element of the $\binom{|V|}{t}$-dimensional vector space of $\mathbb{R}$-valued functions on $\binom{V}{t}$. Since the 1-degree measure will be of particularly high importance, we shall refer to it simply as the degree measure and often suppress the superscript $^{(1)}$ from the notation, denoting it by $\sigma_{\mathcal{H}}$. Given a positive integer d and a vector $\xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d$, we denote by $\|\xi\|$ its ℓ^2-norm, so that

$$\|\xi\|^2 = \sum_{i=1}^{d} \xi_i^2.$$

2.3. **The main technical result.** We are now ready to state the main technical result of this paper, Theorem 2.1 below. We postpone the proof of the theorem to Section 4; the proof will use several simple properties of degree measures that will be derived in Section 3.

---

[5]For those readers who are somewhat familiar with the proof in [6], the essence of the above shortcoming was the following. Only one vertex of degree Δ in $\mathcal{G}$ forced us to remove an edge from $\mathcal{H}$, but while counting the edges of $\mathcal{G}$ that contain some vertex with degree Δ, we accounted for the possibility that every edge of $\mathcal{G}$ contains r vertices of degree Δ.

Theorem 2.1. *Let $s \in \mathbb{N}$ and suppose that a nonempty s-uniform hypergraph $\mathcal{H}$ and reals $p, \delta \in (0,1)$ satisfy*

$$300s^4 \cdot \sum_{t=1}^{s} \binom{s-1}{t-1} \left(\frac{5000s^3}{p} \right)^{t-1} \|\sigma_{\mathcal{H}}^{(t)}\|^2 \leqslant \frac{1}{\delta \cdot v(\mathcal{H})} \leqslant \frac{p}{500}. \tag{2}$$

Then, there exist a family $\mathcal{S} \subseteq \binom{V(\mathcal{H})}{\leqslant 30s^2 p \cdot v(\mathcal{H})}$ and functions $f \colon \mathcal{S} \to \mathcal{P}(V(\mathcal{H}))$ and $g \colon \mathcal{I}(\mathcal{H}) \to \mathcal{S}$ such that, for every $I \in \mathcal{I}(\mathcal{H})$,

$$g(I) \subseteq I \subseteq g(I) \cup f(g(I)) \qquad and \qquad |f(g(I))| \leqslant (1-\delta) \cdot v(\mathcal{H}).$$

Moreover, if $g(I) \subseteq I'$ and $g(I') \subseteq I$ for some $I, I' \in \mathcal{I}(\mathcal{H})$, then $g(I) = g(I')$.

2.4. **The simple and packaged versions.** In this section, we derive Theorems 1.1 and 1.6 from our main technical result, Theorem 2.1 above. We start with a short proof of Theorem 1.1.

Derivation of Theorem 1.1 from Theorem 2.1. Let $p = q/(30s^2)$ and let $\delta = (10^3 s^4 K)^{-1}$. It suffices to verify that $\mathcal{H}$, p, and δ satisfy the assumptions of Theorem 2.1, which will give us the claimed family $\mathcal{S}$ and functions g and f. To this end, note that

$$\|\sigma_{\mathcal{H}}^{(t)}\|^2 = \sum_{T \in \binom{V(\mathcal{H})}{t}} \left(\frac{\deg_{\mathcal{H}} T}{\binom{s}{t} \cdot e(\mathcal{H})} \right)^2 \leqslant \frac{\Delta_t(\mathcal{H})}{\binom{s}{t} \cdot e(\mathcal{H})} \cdot \sum_{T \in \binom{V(\mathcal{H})}{t}} \frac{\deg_{\mathcal{H}} T}{\binom{s}{t} \cdot e(\mathcal{H})} = \frac{\Delta_t(\mathcal{H})}{\binom{s}{t} \cdot e(\mathcal{H})}$$

for every $t \in [\![s]\!]$ and thus the assumptions of the theorem imply that

$$\sum_{t=1}^{s} \binom{s-1}{t-1} \left(\frac{5000s^3}{p} \right)^{t-1} \|\sigma_{\mathcal{H}}^{(t)}\|^2 \leqslant \sum_{t=1}^{s} \binom{s-1}{t-1} \left(\frac{150000s^5}{q} \right)^{t-1} \cdot \frac{\Delta_t(\mathcal{H})}{\binom{s}{t} \cdot e(\mathcal{H})}$$

$$\leqslant \frac{K}{v(\mathcal{H})} \cdot \sum_{t=1}^{s} \frac{t}{s} \cdot \left(\frac{150000s^5}{10^6 \cdot s^5} \right)^{t-1} \leqslant \frac{2K}{v(\mathcal{H})}.$$

Moreover, $300s^4 \cdot 2K \leqslant 1/\delta$ and $p \cdot \delta \cdot v(\mathcal{H}) = q \cdot (30s^2 \cdot 10^3 s^4 K)^{-1} \cdot v(\mathcal{H}) \geqslant 500$. $\qquad \square$

We now turn to the proof of Theorem 1.6. The key ingredient here is the following lemma, which, roughly speaking, states that a hypergraph that is 'robustly dense' contains a large subhypergraph whose maximum degree is not much larger than its average degree. The statement and the proof of the lemma are inspired by the work of Morris and Saxton [32].

Lemma 2.2. *Let $\mathcal{H}$ be an s-uniform hypergraph and suppose that, for some positive β and M, every set $W \subseteq V(\mathcal{H})$ with $|W| \geqslant (1-\beta)v(\mathcal{H})$ satisfies $e(\mathcal{H}[W]) \geqslant M$. Then, there is a subhypergraph $\mathcal{H}' \subseteq \mathcal{H}$ with at least M edges that satisfies*

$$\Delta_1(\mathcal{H}') \leqslant \left\lceil \frac{s}{\beta} \cdot \frac{e(\mathcal{H}')}{v(\mathcal{H})} \right\rceil.$$

Proof. Let $\mathcal{H}'$ be a largest (in terms of the number of edges) subhypergraph of $\mathcal{H}$ satisfying $\Delta_1(\mathcal{H}') \leqslant \lceil \frac{sM}{\beta v(\mathcal{H})} \rceil$ and let $X \subseteq V(\mathcal{H})$ be the set of vertices of $\mathcal{H}'$ whose degree achieves the bound $\lceil \frac{sM}{\beta v(\mathcal{H})} \rceil$. Observe that every edge of $\mathcal{H}$ that is disjoint from X must

belong to $\mathcal{H}'$ and, consequently, $e(\mathcal{H}') \geqslant e(\mathcal{H} - X)$. If $|X| \leqslant \beta v(\mathcal{H})$, then $e(\mathcal{H}') \geqslant M$ by our assumption on $\mathcal{H}$. Otherwise, if $|X| > \beta v(\mathcal{H})$,

$$e(\mathcal{H}') = \frac{1}{s} \sum_{v \in V(\mathcal{H})} \deg_{\mathcal{H}'} v \geqslant \frac{|X|}{s} \cdot \left\lceil \frac{sM}{\beta v(\mathcal{H})} \right\rceil > M.$$

This completes the proof of the lemma. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem 1.6. We shall say that a set $C \subseteq V(\mathcal{H})$ is a *good container* if either $|C| \leqslant \alpha v(\mathcal{H})$ or if there is a subset $W \subseteq C$ with $|W| \geqslant (1-\beta)|C|$ such that $e(\mathcal{H}[W]) < E$. We will construct a rooted tree $\mathcal{T}$ whose vertices are subsets of $V(\mathcal{H})$ that has the following properties:

(i) The root of $\mathcal{T}$ is $V(\mathcal{H})$.
(ii) If an independent set $I \in \mathcal{I}(\mathcal{H})$ is contained in a non-leaf vertex of $\mathcal{T}$, then I is contained in some child of this vertex in $\mathcal{T}$.
(iii) Every leaf of $\mathcal{T}$ is a good container.
(iv) Every non-leaf vertex of $\mathcal{T}$ has at most $(e/q)^{q \cdot v(\mathcal{H})}$ children.
(v) The height of $\mathcal{T}$ is at most $10^4 s^5 \beta^{-1} \cdot \log(e/\alpha)$.

The set of leaves of $\mathcal{T}$ will then form a collection of containers for the independent sets of $\mathcal{H}$ that has the desired properties.

We build such a tree starting from the root $V(\mathcal{H})$ by iteratively applying Theorem 1.1 to (a carefully chosen subhypergraph of) the subhypergraph of $\mathcal{H}$ induced by a leaf C of $\mathcal{T}$ that is not yet a good container and attaching the resulting family of containers (for the independent sets of $\mathcal{H}[C]$, that is, the independent sets of $\mathcal{H}$ that are contained in C) as children of C (which, as a result, ceases to be a leaf of $\mathcal{T}$) until no such leaves are left. This way, properties (i)–(iii) are clearly satisfied. However, we still need to show that the final tree has properties (iv) and (v).

To this end, suppose that $C \subseteq V(\mathcal{H})$ is not a good container, that is, $|C| > \alpha v(\mathcal{H})$ and every $W \subseteq C$ with $|W| \geqslant (1-\beta)|C|$ satisfies $e(\mathcal{H}[W]) \geqslant E$. Lemma 2.2 invoked with $\mathcal{H} \leftarrow \mathcal{H}[C]$ supplies a subhypergraph $\mathcal{H}' \subseteq \mathcal{H}[C]$ with at least E edges that satisfies

$$\Delta_1(\mathcal{H}') \leqslant \left\lceil \frac{s}{\beta} \cdot \frac{e(\mathcal{H}')}{|C|} \right\rceil \leqslant \frac{2s}{\beta} \cdot \frac{e(\mathcal{H}')}{|C|} = \frac{2s}{\beta} \cdot \frac{e(\mathcal{H}')}{v(\mathcal{H}')},$$

where the second inequality follows from our assumption that $e(\mathcal{H}') \geqslant E \geqslant v(\mathcal{H}) \geqslant |C|$. Since, for every $t \in \{2, \ldots, s\}$,

$$\Delta_t(\mathcal{H}') \leqslant \Delta_t(\mathcal{H}) \leqslant \left(\frac{q}{10^6 s^5} \right)^{t-1} \cdot \frac{E}{v(\mathcal{H})} \leqslant \left(\frac{q}{10^6 s^5} \right)^{t-1} \cdot \frac{e(\mathcal{H}')}{v(\mathcal{H}')},$$

Theorem 1.1 invoked with $\mathcal{H} \leftarrow \mathcal{H}'$ and $K \leftarrow 2s/\beta$ supplies sets $\mathcal{S} \subseteq \binom{C}{\leqslant q \cdot |C|}$ and functions $f \colon \mathcal{S} \to \mathcal{P}(C)$ and $g \colon \mathcal{I}(\mathcal{H}') \to \mathcal{S}$ such that, for every $I \in \mathcal{I}(\mathcal{H}')$,

$$I \subseteq g(I) \cup f(g(I)) \qquad \text{and} \qquad \frac{|f(g(I))|}{|C|} \leqslant 1 - \frac{\beta}{2 \cdot 10^3 s^5}.$$

Since $\mathcal{H}' \subseteq \mathcal{H}[C]$, we have $\mathcal{I}(\mathcal{H}[C]) \subseteq \mathcal{I}(\mathcal{H}')$ and we may define

$$\mathcal{C}_C = \big\{ g(I) \cup f(g(I)) : I \in \mathcal{I}(\mathcal{H}[C]) \big\}.$$

By construction, the family $\mathcal{C}_C$ is a family of containers for the independent sets of $\mathcal{H}[C]$ and

$$|\mathcal{C}_C| \leqslant |\mathcal{S}| \leqslant \sum_{i=0}^{q \cdot |C|} \binom{|C|}{i} \leqslant \left(\frac{e}{q}\right)^{q \cdot |C|} \leqslant \left(\frac{e}{q}\right)^{q \cdot v(\mathcal{H})},$$

establishing (iv). Finally, for every $D \in \mathcal{C}_C$,

$$\frac{|D|}{|C|} \leqslant q + 1 - \frac{\beta}{2s} \cdot \frac{1}{10^3 s^4} \leqslant 1 - \frac{\beta}{10^4 s^5},$$

since we assumed that $10^4 s^5 q \leqslant \beta$. In particular, if a set C is a non-leaf vertex of the final tree $\mathcal{T}$ that lies at distance d from the root, then

$$\alpha \leqslant \frac{|C|}{v(\mathcal{H})} \leqslant \left(1 - \frac{\beta}{10^4 s^5}\right)^d \leqslant \exp\left(-\frac{\beta d}{10^4 s^5}\right).$$

This implies that

$$\text{height}(\mathcal{T}) \leqslant \frac{10^4 s^5}{\beta} \cdot \log\left(\frac{1}{\alpha}\right) + 1 \leqslant \frac{10^4 s^5}{\beta} \cdot \log\left(\frac{e}{\alpha}\right),$$

establishing (v). □

3. Properties of degree measures

3.1. **Norms of degree measures.** As we shall be estimating the ℓ^2-norms of t-degree measures of various uniform hypergraphs, we collect here several useful properties of this quantity. We first give general lower and upper bounds on the ℓ^2-norm of the t-degree measure of a hypergraph in terms of the numbers of its vertices and edges and its maximum t-degree. Throughout this section, r is a positive integer. We stress here that all of our hypergraphs are allowed to have multiple edges, that is, every edge can have an arbitrary positive multiplicity. (This idea was first introduced by Saxton and Thomason [42].) Moreover, when computing $\deg_{\mathcal{H}}$ and $e(\mathcal{H})$, we always count edges with multiplicities.

Fact 3.1. *Suppose that $\mathcal{H}$ is a nonempty r-uniform hypergraph. For every $t \in [\![r]\!]$,*

$$\max\left\{\frac{1}{\binom{v(\mathcal{H})}{t}}, \frac{1}{\binom{r}{t} \cdot e(\mathcal{H})}\right\} \leqslant \|\sigma_{\mathcal{H}}^{(t)}\|^2 \leqslant \frac{\Delta_t(\mathcal{H})}{\binom{r}{t} \cdot e(\mathcal{H})}.$$

Proof. The upper bound is straightforward:

$$\|\sigma_{\mathcal{H}}^{(t)}\|^2 = \sum_{T \in \binom{V(\mathcal{H})}{t}} \left(\frac{\deg_{\mathcal{H}} T}{\binom{r}{t} \cdot e(\mathcal{H})}\right)^2 \leqslant \frac{\Delta_t(\mathcal{H})}{\binom{r}{t} \cdot e(\mathcal{H})} \cdot \sum_{T \in \binom{V(\mathcal{H})}{t}} \frac{\deg_{\mathcal{H}} T}{\binom{r}{t} \cdot e(\mathcal{H})} = \frac{\Delta_t(\mathcal{H})}{\binom{r}{t} \cdot e(\mathcal{H})}.$$

For the lower bound, let

$$\mathcal{T} = \left\{T \in \binom{V(\mathcal{H})}{t} : \deg_{\mathcal{H}} T > 0\right\}$$

and observe that

$$|\mathcal{T}| \leqslant \min\left\{\binom{v(\mathcal{H})}{t}, \binom{r}{t} \cdot e(\mathcal{H})\right\}.$$

It follows form the Cauchy–Schwarz inequality that

$$\|\sigma_{\mathcal{H}}^{(t)}\|^2 = \sum_{T \in \mathcal{T}} \left(\frac{\deg_{\mathcal{H}} T}{\binom{r}{t} \cdot e(\mathcal{H})} \right)^2 \geqslant \frac{1}{|\mathcal{T}|} \cdot \left(\sum_{T \in \mathcal{T}} \frac{\deg_{\mathcal{H}} T}{\binom{r}{t} \cdot e(\mathcal{H})} \right)^2 = \frac{1}{|\mathcal{T}|},$$

implying the lower bound. $\square$

Our second observation states that the ℓ^2-norm of a t-degree measure of a hypergraph cannot increase much when one deletes from it a small proportion of its edges.

Fact 3.2. *If $\mathcal{H}'$ is a nonempty subhypergraph of an r-uniform hypergraph $\mathcal{H}$, then, for every $t \in [\![r]\!]$,*

$$\|\sigma_{\mathcal{H}'}^{(t)}\| \leqslant \frac{e(\mathcal{H})}{e(\mathcal{H}')} \cdot \|\sigma_{\mathcal{H}}^{(t)}\|.$$

Proof. The assertion follows simply because $\deg_{\mathcal{H}'} T \leqslant \deg_{\mathcal{H}} T$ for every $T \subseteq V(\mathcal{H}')$ and hence

$$\sigma_{\mathcal{H}'}^{(t)}(T) \leqslant \frac{e(\mathcal{H})}{e(\mathcal{H}')} \cdot \sigma_{\mathcal{H}}^{(t)}(T). \qquad \square$$

Our final lemma relates the ℓ^2-norm of the degree measure of a uniform hypergraph to a simple property of its edge distribution.

Lemma 3.3. *Suppose that $\mathcal{H}$ is a nonempty r-uniform hypergraph. If a set $D \subseteq V(\mathcal{H})$ satisfies $e(\mathcal{H} - D) \leqslant (1 - \varepsilon) \cdot e(\mathcal{H})$ for some $\varepsilon > 0$, then*

$$|D| \geqslant \left(\frac{\varepsilon}{r} \right)^2 \cdot \|\sigma_{\mathcal{H}}\|^{-2}.$$

Proof. Let $\mathbf{1}_D \in \mathbb{R}^{V(\mathcal{H})}$ be the characteristic vector of D. It follows from the Cauchy–Schwarz inequality that

$$\langle \mathbf{1}_D, \sigma_{\mathcal{H}} \rangle^2 \leqslant \|\mathbf{1}_D\|^2 \cdot \|\sigma_{\mathcal{H}}\|^2 = |D| \cdot \|\sigma_{\mathcal{H}}\|^2.$$

Since at least an ε-proportion of edges of $\mathcal{H}$ contain at least one vertex of D,

$$\langle \mathbf{1}_D, \sigma_{\mathcal{H}} \rangle = \frac{1}{r \cdot e(\mathcal{H})} \cdot \sum_{v \in D} \deg_{\mathcal{H}} v \geqslant \frac{\varepsilon}{r},$$

giving the desired lower bound on $|D|$. $\square$

3.2. **Degree measures and link hypergraphs.** Suppose that $\mathcal{H}$ is an $(r+1)$-uniform hypergraph with vertex set V. Given a $v \in V$, we shall denote by $\mathcal{H}_v$ the *link hypergraph* of v (the *neighbourhood* of v in $\mathcal{H}$), that is, the r-uniform hypergraph with vertex set V whose edges are all the r-element sets A such that $\{v\} \cup A$ is an edge of $\mathcal{H}$. A property of crucial importance for us is that, for each $t \in [\![r]\!]$, the t-degree measure of $\mathcal{H}$ is a convex combination of the t-degree measures of the link hypergraphs of its vertices. Moreover, each of these convex combinations has the same coefficients – the coordinates of the 1-degree measure vector $\sigma_{\mathcal{H}}$.

Remark. Even though $\sigma_{\mathcal{H}}^{(t)}$ was defined only for nonempty hypergraphs $\mathcal{H}$, for the sake of brevity, we shall often write $0 \cdot \sigma_{\mathcal{H}}^{(t)}$ even if $\mathcal{H}$ has no edges. In this case, $0 \cdot \sigma_{\mathcal{H}}^{(t)}$ should be interpreted as the zero vector of appropriate dimension.

Fact 3.4. *Suppose that $\mathcal{H}$ is a nonempty $(r+1)$-uniform hypergraph with vertex set V. For every $t \in [\![r]\!]$,*

$$\sum_{v \in V} \sigma_{\mathcal{H}}(v) \cdot \sigma_{\mathcal{H}_v}^{(t)} = \sigma_{\mathcal{H}}^{(t)}.$$

Proof. It follows from our definition of a link hypergraph that $e(\mathcal{H}_v) = \deg_{\mathcal{H}} v$ for each $v \in V$ and, more generally, for every $T \in \binom{V}{t}$,

$$\deg_{\mathcal{H}_v} T = \begin{cases} \deg_{\mathcal{H}}(T \cup \{v\}) & \text{if } v \notin T, \\ 0 & \text{if } v \in T. \end{cases}$$

Consequently,

$$\begin{aligned}
\sum_{v \in V} \sigma_{\mathcal{H}}(v) \cdot \sigma_{\mathcal{H}_v}^{(t)}(T) &= \sum_{v \in V} \frac{\deg_{\mathcal{H}} v}{(r+1) \cdot e(\mathcal{H})} \cdot \frac{\deg_{\mathcal{H}_v} T}{\binom{r}{t} \cdot e(\mathcal{H}_v)} \\
&= \frac{1}{(r+1)\binom{r}{t} \cdot e(\mathcal{H})} \cdot \sum_{v \in V \setminus T} \deg_{\mathcal{H}}(T \cup \{v\}) \\
&= \frac{(r+1-t) \cdot \deg_{\mathcal{H}} T}{(r+1)\binom{r}{t} \cdot e(\mathcal{H})} = \frac{\deg_{\mathcal{H}} T}{\binom{r+1}{t} \cdot e(\mathcal{H})} = \sigma_{\mathcal{H}}^{(t)}(T),
\end{aligned}$$

where we used the identity $(r+1)\binom{r}{t} = \binom{r+1}{t}(r+1-t)$. $\qquad\square$

In our arguments, we shall employ the following relation between the ℓ^2-norm of the $(t+1)$-degree measure of a hypergraph and the ℓ^2-norms of the t-degree measures of the link hypergraphs of its vertices.

Fact 3.5. *Suppose that $\mathcal{H}$ is a nonempty $(r+1)$-uniform hypergraph with vertex set V. For every $t \in [\![r]\!]$,*

$$\sum_{v \in V} \sigma_{\mathcal{H}}(v)^2 \cdot \|\sigma_{\mathcal{H}_v}^{(t)}\|^2 = \frac{\|\sigma_{\mathcal{H}}^{(t+1)}\|^2}{t+1}.$$

Proof. A quick way to verify the claimed identity is to observe that both the left- and the right-hand sides of the claimed equality express the probability of obtaining the same outcome in two independent executions of the following random process: Pick an edge A of $\mathcal{H}$ uniformly at random, choose a $(t+1)$-element subset S of A uniformly at random, mark a vertex $v \in S$ chosen uniformly at random, and return the pair (v, S).

More explicitly, using the identities $\deg_{\mathcal{H}} v = e(\mathcal{H}_v)$, valid for every $v \in V$, and $\deg_{\mathcal{H}_v} T = \deg_{\mathcal{H}}(T \cup \{v\})$, valid for each $T \in \binom{V}{t}$ and $v \in V \setminus T$, and $(r+1)\binom{r}{t} =$

$(t + 1)\binom{r+1}{t+1}$, we get

$$\sum_{v \in V} \sigma_{\mathcal{H}}(v)^2 \cdot \|\sigma_{\mathcal{H}_v}^{(t)}\|^2 = \sum_{v \in V} \left(\frac{\deg_{\mathcal{H}} v}{(r + 1) \cdot e(\mathcal{H})} \right)^2 \cdot \sum_{T \in \binom{V}{t}} \left(\frac{\deg_{\mathcal{H}_v} T}{\binom{r}{t} \cdot e(\mathcal{H}_v)} \right)^2$$

$$= \left(\frac{1}{(r + 1)\binom{r}{t} \cdot e(\mathcal{H})} \right)^2 \cdot \sum_{T \in \binom{V}{t}} \sum_{v \in V \setminus T} \left(\deg_{\mathcal{H}}(T \cup \{v\}) \right)^2$$

$$= \left(\frac{1}{(t + 1)\binom{r+1}{t+1} \cdot e(\mathcal{H})} \right)^2 \cdot \sum_{S \in \binom{V}{t+1}} \sum_{v \in S} (\deg_{\mathcal{H}} S)^2$$

$$= \frac{1}{t + 1} \cdot \sum_{S \in \binom{V}{t+1}} \left(\frac{\deg_{\mathcal{H}} S}{\binom{r+1}{t+1} \cdot e(\mathcal{H})} \right)^2 = \frac{\|\sigma_{\mathcal{H}}^{(t+1)}\|^2}{t + 1}. \qquad \square$$

3.3. **Linear combinations of degree measures.** It will be convenient to introduce another piece of notation. Given a vector $\alpha \in \mathbb{R}^r$ with nonnegative coordinates and a nonempty hypergraph $\mathcal{K}$ with uniformity at least r, we define

$$\sigma_\alpha(\mathcal{K}) = \left(\alpha_1^{1/2} \cdot \sigma_{\mathcal{K}}^{(1)}, \dots, \alpha_r^{1/2} \cdot \sigma_{\mathcal{K}}^{(r)} \right) \in \mathbb{R}^{\sum_{t=1}^{r} \binom{v(\mathcal{K})}{t}},$$

so that

$$\|\sigma_\alpha(\mathcal{K})\|^2 = \sum_{t=1}^{r} \alpha_t \cdot \|\sigma_{\mathcal{K}}^{(t)}\|^2.$$

The following generalisation of Fact 3.4 holds.

Fact 3.6. *Suppose that $\mathcal{K}$ is a nonempty hypergraph with uniformity at least $r + 1$. For every $\alpha \in \mathbb{R}^r$ with nonnegative coordinates,*

$$\sigma_\alpha(\mathcal{K}) = \sum_{v \in V(\mathcal{K})} \sigma_{\mathcal{K}}(v) \cdot \sigma_\alpha(\mathcal{K}_v).$$

4. Proof

4.1. **Outline.** Our proof of Theorem 2.1 follows the general strategy of [6]. We construct functions $g, f^* \colon \mathcal{I}(\mathcal{H}) \to \mathcal{P}(V(\mathcal{H}))$ that satisfy the following three conditions for every $I \in \mathcal{I}(\mathcal{H})$:

(a) $g(I) \subseteq I \subseteq g(I) \cup f^*(I)$;
(b) $|g(I)| \leqslant 30s^2 \cdot p \cdot v(\mathcal{H})$ and $|f^*(I)| \leqslant (1 - \delta) \cdot v(\mathcal{H})$;
(c) if $g(I) \subseteq I'$ and $g(I') \subseteq I$ for some $I' \in \mathcal{I}(\mathcal{H})$, then $g(I) = g(I')$ and $f^*(I) = f^*(I')$.

The existence of such functions easily implies the assertion of the theorem. Indeed, condition (c) guarantees that there is an implicit decomposition $f^* = f \circ g$.

Given an independent set I, the sets $g(I)$ and $f^*(I)$ are constructed by an algorithm that operates in a sequence of at most $s - 1$ rounds, which are indexed by $r = s - 1, \dots, 1$. At the start of round r, the algorithm receives as input an $(r + 1)$-uniform hypergraph $\mathcal{H}^{(r+1)}$ satisfying $I \in \mathcal{I}(\mathcal{H}^{(r+1)})$; we initialise the first round, indexed by $r = s - 1$, with $\mathcal{H}^{(s)} = \mathcal{H}$. During the round, the algorithm tries to construct an r-uniform hypergraph $\mathcal{H}^{(r)}$ with $I \in \mathcal{I}(\mathcal{H}^{(r)})$ and some additional desirable properties. The

definition of 'desirable' is where we significantly depart from the previous approaches. In [6], as well as in [42], this desirability property was defined in terms of a lower bound on the number of edges of $\mathcal{H}^{(r)}$ and upper bounds on the maximum degrees $\Delta_t(\mathcal{H}^{(r)})$, for all $t \in [\![r]\!]$. Here, we aim to control the ℓ^2-norms of the t-degree measures of $\mathcal{H}^{(r)}$. More precisely, 'desirable' means that a carefully chosen linear combination of $\|\sigma_{\mathcal{H}^{(r)}}^{(t)}\|^2$, where t ranges over $[\![r]\!]$, is small. As in [6], in the event that such a hypergraph $\mathcal{H}^{(r)}$ cannot be constructed for our input set I, the algorithm is able to define the required set $f^*(I)$ already at the end of round r. Crucially, the amount of information about the set I that is needed to describe $\mathcal{H}^{(r)}$, or the set $f^*(I)$, is rather small. More precisely, it naturally corresponds to a set of $O(p \cdot s \cdot v(\mathcal{H}))$ elements of I, which we shall denote here by $S^{(r)}$. In particular, if we let $g(I)$ be the union of the sets $S^{(r)}$ from all of the (at most $s-1$) rounds of the algorithm, the knowledge of $g(I)$ alone (without any additional knowledge of the set I other than the fact that it is independent) is sufficient to recreate the entire execution of the algorithm, and thus also the final set $f^*(I)$.

In order to construct the r-uniform $\mathcal{H}^{(r)}$ given the $(r+1)$-uniform $\mathcal{H}^{(r+1)}$ and the set I, our algorithm considers a sequence of queries 'Does v belong to I?' for some carefully chosen sequence of vertices $v \in V(\mathcal{H})$. We record all the positive answers by placing the respective vertices v in the (initially empty) set $S^{(r)}$. Each queried vertex v is clearly not in the set $I \setminus S^{(r)}$ and hence it may be omitted from the set $f^*(I)$. In particular, the algorithm will produce the desired set $f^*(I)$ if at least $\delta \cdot v(\mathcal{H})$ queries are made. In case a queried vertex v does belong to I, we add its r-uniform link hypergraph $\mathcal{H}_v^{(r+1)}$ to the (initially empty) hypergraph $\mathcal{H}^{(r)}$. Note that this guarantees that $I \in \mathcal{I}(\mathcal{H}^{(r)})$.

Recall that our aim is to produce a hypergraph $\mathcal{H}^{(r)}$ whose t-degree measures have small ℓ^2-norms. The crux of the matter is the choice of the next vertex v to be queried for membership in I. Indeed, if v happens to belong to I, then $\mathcal{H}_v^{(r+1)}$ will be added to $\mathcal{H}^{(r)}$ and, as a result of this operation, the t-degree measures of $\mathcal{H}^{(r)}$ will change – the new $\sigma_{\mathcal{H}^{(r)}}^{(t)}$ will be a convex combination of the old $\sigma_{\mathcal{H}^{(r)}}^{(t)}$ and of $\sigma_{\mathcal{H}_v^{(r+1)}}^{(t)}$, with appropriate coefficients (our hypergraphs are allowed to have multiple edges). It turns out that choosing the 'right' candidate vertex v is an optimisation problem that admits a rather simple geometric description. The solution to this geometric problem, presented as Lemma 4.12 and expressed in the language of degree measures of hypergraphs in Proposition 4.11, lies at the heart of our argument.

The bottom line is that there is a way to choose a sequence of vertices to be queried for membership in I such that, if at least $\Omega(p \cdot s \cdot v(\mathcal{H}))$ out of the first $\delta \cdot v(\mathcal{H})$ queried vertices belong to the set I, some linear combination of $\|\sigma_{\mathcal{H}^{(r)}}^{(t)}\|^2$, where t ranges over $[\![r]\!]$, will be at most $1 + O(1/s)$ times larger than a respective linear combination of $\|\sigma_{\mathcal{H}^{(r+1)}}^{(t)}\|^2$, where t ranges over $[\![r+1]\!]$. Consequently, either one of the $s-1$ rounds of the algorithm will output a desired set $f^*(I)$ of size $\delta \cdot c(\mathcal{H})$ or the algorithm will eventually produce a 1-uniform hypergraph $\mathcal{H}^{(1)}$ such that $I \in \mathcal{I}(\mathcal{H}^{(1)})$ and

$$\|\sigma_{\mathcal{H}^{(1)}}\| \leqslant \left(1 + O(1/s)\right)^{s-1} \cdot \|\sigma_{\mathcal{H}}\| \leqslant O(1) \cdot \|\sigma_{\mathcal{H}}\|.$$

In case the latter happens, we may simply let $f^*(I)$ comprise all vertices v such that $\{v\} \notin \mathcal{H}^{(1)}$. The upper bound on $\|\sigma_{\mathcal{H}^{(1)}}\|$ implies that there are at most $(1-\delta) \cdot v(\mathcal{H})$ such vertices, as shown in Lemma 3.3.

4.2. **The key lemma.** The following lemma summarises a single round of our new, refined algorithm for constructing containers. We denote by $\text{Hyp}_r(V)$ the family of r-uniform hypergraphs with vertex set V; recall again that we allow our hypergraphs to have multiple edges.

Lemma 4.1. *Let $\mathcal{G}$ be an $(r+1)$-uniform hypergraph with vertex set V. Suppose that $\varepsilon \in \big(0, (9(r+1))^{-1}\big)$ and $p \in (0,1)$ satisfy*

$$\|\sigma_\mathcal{G}\|^2 \leqslant \frac{\varepsilon^3 p}{50(r+1)}.$$

Let $\alpha \in \mathbb{R}^r$ be a vector with nonnegative coordinates and define

$$\alpha^* = (1+\varepsilon)^{10} \cdot (\alpha, 0) + \frac{50(r+1)}{\varepsilon^2 p} \cdot (0, \alpha) \in \mathbb{R}^{r+1} \qquad and \qquad b = \left\lceil \frac{2p}{\varepsilon} \cdot |V| \right\rceil. \qquad (3)$$

Then, there exist (disjoint) families $\mathcal{S}', \mathcal{S}'' \subseteq \binom{V}{\leqslant b}$ and functions $S \colon \mathcal{I}(\mathcal{G}) \to \mathcal{S}' \cup \mathcal{S}''$, $C \colon \mathcal{S}' \to \mathcal{P}(V)$, and $\mathcal{F} \colon \mathcal{S}'' \to \text{Hyp}_r(V)$ such that, for every $I \in \mathcal{I}(\mathcal{G})$, we have $S_I \subseteq I$. Moreover:

(1) If $S_I \in \mathcal{S}'$, then $I \setminus S_I \subseteq C(S_I)$ and $|V| - |C(S_I)| \geqslant \frac{\varepsilon^2}{(r+1)^2} \cdot \|\sigma_\mathcal{G}\|^{-2}$.

(2) If $S_I \in \mathcal{S}''$, then $I \in \mathcal{I}(\mathcal{F}(S_I))$ and $\|\sigma_\alpha(\mathcal{F}(S_I))\| \leqslant \|\sigma_{\alpha^*}(\mathcal{G})\|$.

Finally, if $S_I \subseteq I'$ and $S_{I'} \subseteq I$ for some $I, I' \in \mathcal{I}(\mathcal{G})$, then $S_I = S_{I'}$.

Before we embark on the proof of Lemma 4.1, we shall first show, in the next subsection, how it implies Theorem 2.1. The remainder of this section, Subsections 4.4–4.9 will be devoted to the proof of the lemma.

4.3. **Derivation of Theorem 2.1.** Let $\mathcal{H}$ be a nonempty s-uniform hypergraph with vertex set V and suppose that $\delta, p \in (0,1)$ satisfy (2), that is,

$$300s^4 \cdot \sum_{t=1}^{s} \binom{s-1}{t-1} \left(\frac{5000s^3}{p} \right)^{t-1} \|\sigma_\mathcal{H}^{(t)}\|^2 \leqslant \frac{1}{\delta \cdot v(\mathcal{H})} \leqslant \frac{p}{500}. \qquad (2)$$

Define

$$\varepsilon = \frac{1}{10s} \qquad and \qquad \Gamma = \frac{50s}{\varepsilon^2}$$

and let $\alpha^{(1)} \in \mathbb{R}^1, \ldots, \alpha^{(s)} \in \mathbb{R}^s$ be vectors defined by

$$\alpha_t^{(r)} = \binom{r-1}{t-1} \cdot (1+\varepsilon)^{10(r-t)} \cdot \left(\frac{\Gamma}{p} \right)^{t-1},$$

for every $r \in [\![s]\!]$ and each $t \in [\![r]\!]$. Given an independent set I of $\mathcal{H}$, we construct the sets $g(I)$ and $f^*(I)$ using the following procedure.

Construction of the container. Let $\mathcal{H}^{(s)} = \mathcal{H}$. Do the following for $r = s-1, \ldots, 1$:

(C1) Invoke Lemma 4.1 with $\mathcal{G} \leftarrow \mathcal{H}^{(r+1)}$ and $\alpha \leftarrow \alpha^{(r)}$ to obtain families $\mathcal{S}'$ and $\mathcal{S}''$ and functions S, C, and $\mathcal{F}$, as in the statement of the lemma.[6]

(C2) Let $S^{(r)} \leftarrow S_I$.

(C3) If $S_I \in \mathcal{S}'$, then let $g(I) = S^{(s-1)} \cup \cdots \cup S^{(r)}$ and $f^*(I) = C(S_I)$ and STOP.

(C4) Otherwise, if $S_I \in \mathcal{S}''$, we let $\mathcal{H}^{(r)} \leftarrow \mathcal{F}(S_I)$ and CONTINUE.

---

[6]In order to do so, we have to make sure that $\|\sigma_{\mathcal{H}^{(r+1)}}\| \leqslant \frac{\varepsilon^3 p}{50(r+1)}$. In the analysis of the procedure, below, we will verify that this is always the case.

If STOP has not been called, then $r = 1$ and $\mathcal{H}^{(1)}$ has been defined. Let $g(I) = S^{(s-1)} \cup \cdots \cup S^{(1)}$ and $f^*(I) = \{v \in V : \{v\} \notin \mathcal{H}^{(1)}\}$.

In the remainder of this section, we shall show that, for every $I \in \mathcal{I}(\mathcal{H})$, the above procedure indeed constructs sets $g(I)$ and $f^*(I)$ that have the desired properties; in particular, we shall show that f^* decomposes as $f^* = f \circ g$.

Claim 4.2. *For each $r \in [\![s]\!]$, the hypergraph $\mathcal{H}^{(r)}$, if it was defined, satisfies:*

(i) $I \in \mathcal{I}(\mathcal{H}^{(r)})$ *and*
(ii) $\|\sigma_{\alpha^{(r)}}(\mathcal{H}^{(r)})\|^2 \leqslant \frac{\varepsilon^2}{s^2 \delta |V|}$.

The proof of Claim 4.2 requires the following simple fact, which justifies our definition of the vectors $\alpha^{(1)}, \ldots, \alpha^{(s)}$.

Fact 4.3. *Suppose that $r \in [\![s-1]\!]$. Let $\alpha \in \mathbb{R}^r$ and let $\alpha^* \in \mathbb{R}^{r+1}$ be defined as in (3). If $\alpha_t \leqslant \alpha_t^{(r)}$ for all $t \in [\![r]\!]$, then $\alpha_t^* \leqslant \alpha_t^{(r+1)}$ for all $t \in [\![r+1]\!]$.*

Proof. Note first that

$$\alpha_1^* = (1+\varepsilon)^{10} \cdot \alpha_1 \leqslant (1+\varepsilon)^{10} \cdot \alpha_1^{(r)} = (1+\varepsilon)^{10r} = \alpha_1^{(r+1)}$$

and, since $50(r+1) \leqslant 50s = \Gamma\varepsilon^2$,

$$\alpha_{r+1}^* = \frac{50(r+1)}{\varepsilon^2 p} \cdot \alpha_r \leqslant \frac{\Gamma}{p} \cdot \alpha_r^{(r)} = \left(\frac{\Gamma}{p}\right)^r = \alpha_{r+1}^{(r+1)}.$$

Finally, if $1 < t \leqslant r$, then

$$\alpha_t^* = (1+\varepsilon)^{10} \cdot \alpha_t + \frac{50(r+1)}{\varepsilon^2 p} \cdot \alpha_{t-1} \leqslant (1+\varepsilon)^{10} \cdot \alpha_t^{(r)} + \frac{\Gamma}{p} \cdot \alpha_{t-1}^{(r)}$$

$$= \binom{r-1}{t-1}(1+\varepsilon)^{10(r-t+1)}\left(\frac{\Gamma}{p}\right)^{t-1} + \binom{r-1}{t-2}(1+\varepsilon)^{10(r-(t-1))}\left(\frac{\Gamma}{p}\right)^{t-2+1}$$

$$= \binom{r}{t-1}(1+\varepsilon)^{10(r+1-t)}\left(\frac{\Gamma}{p}\right)^{t-1} = \alpha_t^{(r+1)},$$

where the second to last equality is Pascal's formula. $\qquad \square$

Proof of Claim 4.2. We prove the claim by induction on $s-r$. The basis of the induction is the case $r = s$. Property (i) is satisfied, as $\mathcal{H}^{(s)} = \mathcal{H}$ and $I \in \mathcal{I}(\mathcal{H})$. In order to see that property (ii) holds as well, note first that the first inequality in the main assumption (2) of Theorem 2.1 gives

$$\|\sigma_{\alpha^{(s)}}(\mathcal{H}^{(s)})\|^2 = \|\sigma_{\alpha^{(s)}}(\mathcal{H})\|^2 = \sum_{t=1}^{s} \binom{s-1}{t-1}(1+\varepsilon)^{10(s-t)}\left(\frac{\Gamma}{p}\right)^{t-1} \|\sigma_{\mathcal{H}}^{(t)}\|^2$$

$$\leqslant (1+\varepsilon)^{10s} \cdot \sum_{t=1}^{s} \binom{s-1}{t-1}\left(\frac{\Gamma}{p}\right)^{t-1} \|\sigma_{\mathcal{H}}^{(t)}\|^2$$

$$\leqslant (1+\varepsilon)^{10s} \cdot \frac{1}{300s^4 \cdot \delta \cdot |V|} \leqslant \frac{\varepsilon^2}{s^2 \cdot \delta \cdot |V|},$$

where the second inequality holds as $\Gamma = 5000s^3$ and the last inequality holds since $\varepsilon = 1/(10s)$ and, consequently, $(1+\varepsilon)^{10s} \leqslant e^{10\varepsilon s} = e \leqslant 3$.

Suppose now that $r \in [\![s-1]\!]$ and that the hypergraph $\mathcal{H}^{(r+1)}$ was defined. We first argue that we are allowed to invoke Lemma 4.1 in step (C1) above. Indeed, since the second inequality in (2) implies that

$$\frac{1}{s \cdot \delta \cdot |V|} = \frac{10\varepsilon}{\delta \cdot |V|} \leqslant \frac{\varepsilon p}{50}$$

and $\alpha_1^{(r+1)} = (1+\varepsilon)^{10r} \geqslant 1$, our inductive assumption implies that

$$\|\sigma_{\mathcal{H}^{(r+1)}}\|^2 \leqslant \alpha_1^{(r+1)} \cdot \|\sigma_{\mathcal{H}^{(r+1)}}\|^2 \leqslant \|\sigma_{\alpha^{(r+1)}}(\mathcal{H}^{(r+1)})\|^2 \leqslant \frac{\varepsilon^2}{s^2 \delta |V|} \leqslant \frac{\varepsilon^3 p}{50(r+1)},$$

as needed (in order to apply Lemma 4.1 with $\mathcal{G} \leftarrow \mathcal{H}^{(r+1)}$). If the hypergraph $\mathcal{H}^{(r)}$ was defined, in step (C4), then Lemma 4.1 guarantees that

$$\|\sigma_{\alpha^{(r)}}(\mathcal{H}^{(r)})\|^2 \leqslant \|\sigma_{\alpha^*}(\mathcal{H}^{(r+1)})\|^2,$$

where α^* is defined as in (3), with $\alpha \leftarrow \alpha^{(r)}$. However, Fact 4.3 states that $\alpha^* \leqslant \alpha^{(r+1)}$ coordinate-wise and, therefore, we may conclude that

$$\|\sigma_{\alpha^{(r)}}(\mathcal{H}^{(r)})\|^2 \leqslant \|\sigma_{\alpha^*}(\mathcal{H}^{(r+1)})\|^2 \leqslant \|\sigma_{\alpha^{(r+1)}}(\mathcal{H}^{(r+1)})\|^2,$$

as needed. $\square$

We now verify that $g(I)$ and $f^*(I)$ have the desired properties and that f^* decomposes as $f^* = f \circ g$. Since $g(I) = S^{(s-1)} \cup \cdots \cup S^{(r)}$, for some $r \in [\![s-1]\!]$, the fact that $g(I) \subseteq I$ is an immediate consequence of the definitions of $S^{(s-1)}, \ldots, S^{(r)}$, made in step (C2), and the fact that the respective sets S_I are all contained in I, as guaranteed by Lemma 4.1. Moreover, since each of these sets S_I has at most $\lceil 2p|V|/\varepsilon \rceil$ elements, see Lemma 4.1, we have

$$|g(I)| \leqslant |S^{(s-1)}| + \cdots + |S^{(r)}| \leqslant (s-1) \cdot \left\lceil \frac{2p \cdot |V|}{\varepsilon} \right\rceil \leqslant \frac{2sp}{\varepsilon} \cdot |V| = 20s^2 p \cdot |V|,$$

where the last inequality holds because the main assumption (2) and Fact 3.1 imply that $p \geqslant s^4 \cdot \|\sigma_{\mathcal{H}}^{(1)}\|^2 \geqslant s^4/|V|$.

The set $f^*(I)$ is defined either in step (C3), for some $r \in [\![s-1]\!]$, or at the end of the procedure, if the 1-uniform hypergraph $\mathcal{H}^{(1)}$ is constructed. In the former case, $f^*(I) = C(S_I)$ for functions S and C obtained from Lemma 4.1. Note that $I \setminus g(I) \subseteq I \setminus S_I \subseteq C(S_i) = f^*(I)$, as $g(I) \supseteq S^{(r)} = S_I$. Moreover, $|f^*(I)| \leqslant (1-\delta)|V|$, since, on the one hand, Lemma 4.1 guarantees that

$$|V| - |f^*(I)| \geqslant \frac{\varepsilon^2}{(r+1)^2} \cdot \|\sigma_{\mathcal{H}^{(r+1)}}\|^{-2}$$

and, on the other hand, by Claim 4.2, as $\alpha_1^{(r+1)} = (1+\varepsilon)^{10r} \geqslant 1$,

$$\|\sigma_{\mathcal{H}^{(r+1)}}\|^2 \leqslant \alpha_1^{(r+1)} \cdot \|\sigma_{\mathcal{H}^{(r+1)}}\|^2 \leqslant \|\sigma_{\alpha^{(r+1)}}(\mathcal{H}^{(r+1)})\|^2 \leqslant \frac{\varepsilon^2}{s^2 \delta |V|} \leqslant \frac{\varepsilon^2}{(r+1)^2 \delta |V|}.$$

In the latter case, $f^*(I) = \{ v \in V : \{v\} \notin \mathcal{H}^{(1)} \}$. In particular, we must have $I \subseteq f^*(I)$, since otherwise I would not be an independent set in $\mathcal{H}^{(1)}$, which would contradict property (i) in Claim 4.2. Moreover, Lemma 3.3, invoked with $\mathcal{H} \leftarrow \mathcal{H}^{(1)}$, $D \leftarrow V \setminus f^*(I)$,

and $\varepsilon \leftarrow 1$, gives $|V \setminus f^*(I)| \geqslant \|\sigma_{\mathcal{H}^{(1)}}\|^{-2}$. Since $\alpha_1^{(1)} = 1$, we have $\sigma_{\mathcal{H}^{(1)}} = \sigma_{\alpha^{(1)}}(\mathcal{H}^{(1)})$ and property (ii) in Claim 4.2 allows us to conclude that

$$|V| - |f^*(I)| \geqslant \|\sigma_{\alpha^{(1)}}(\mathcal{H}^{(1)})\|^{-2} \geqslant \frac{s^2 \delta |V|}{\varepsilon^2} \geqslant \delta |V|.$$

Finally, we show that f^* decomposes as $f^* = f \circ g$. To this end, it suffices to show that if $g(I) = g(I')$, for some $I, I' \in \mathcal{I}(\mathcal{H})$, then $f^*(I) = f^*(I')$. In fact, we shall prove the following stronger statement.

Claim 4.4. *If $g(I) \subseteq I'$ and $g(I') \subseteq I$ for some $I, I' \in \mathcal{I}(\mathcal{H})$, then $g(I) = g(I')$ and $f^*(I) = f^*(I')$.*

Note that Claim 4.4 implies the desired property of f^*. Indeed, assume that $g(I) = g(I')$. Since $g(I) \subseteq I$ and $g(I') \subseteq I'$, as shown above, we have $g(I) = g(I') \subseteq I \cap I'$ and the claim yields $f^*(I) = f^*(I')$.

Proof of Claim 4.4. The claim is an easy consequence of the respective property of the function S from the statement of Lemma 4.1. Indeed, it suffices to show that the container-constructing procedure described above defines the same sets $S^{(r)}$ and the same hypergraphs $\mathcal{H}^{(r)}$ when applied to both I and I'; this is because $g(I)$ and $g(I')$ are unions of the respective sets $S^{(r)}$ and the sets $f^*(I)$ and $f^*(I')$ depend only on the sets $S^{(r)}$ and the hypergraphs $\mathcal{H}^{(r)}$. One may prove this assertion by induction on $s - r$. For the induction step, note that while the procedure performs step (C1), the respective hypergraphs $\mathcal{H}^{(r+1)}$ are identical (by the inductive assumption) and therefore so are the functions S, C, and $\mathcal{F}$. Moreover, since $S_I \subseteq g(I)$ and $S_{I'} \subseteq g(I')$, by our definition of $g(I)$ and $g(I')$, we may conclude that $S_I \subseteq I'$ and $S_{I'} \subseteq I$ and thus, the final assertion of Lemma 4.1 gives us the equality $S_I = S_{I'}$. $\qquad\square$

4.4. Pruning hypergraphs.
In order to streamline the analysis of our algorithm that constructs the r-uniform hypergraph $\mathcal{H}^{(r)}$ from the $(r+1)$-uniform $\mathcal{H}^{(r+1)}$, we will first prune the latter hypergraph by removing from it vertices with unusually high degree. More precisely, define, for a nonempty r-uniform hypergraph $\mathcal{H}$ and $t \in [\![r]\!]$,

$$\hat{\Delta}_t(\mathcal{H}) = \frac{r}{t} \cdot e(\mathcal{H}) \cdot \|\sigma_{\mathcal{H}}^{(t)}\|^2; \tag{4}$$

one should think of $\hat{\Delta}_t(\mathcal{H})$ as a robust analogue of the maximum degree $\Delta_t(\mathcal{H})$. In particular, Fact 3.1 implies that $\hat{\Delta}_1(\mathcal{H}) \leqslant \Delta_1(\mathcal{H})$; even though equality sometimes holds (when $\mathcal{H}$ is regular), in general the ratio $\Delta_1(\mathcal{H})/\hat{\Delta}_1(\mathcal{H})$ can be arbitrarily large. Our next lemma shows that this inequality becomes nearly tight, up to a multiplicative factor of $O(r)$, after we delete a small proportion of the edges of $\mathcal{H}$.

Lemma 4.5. *Suppose that $\mathcal{H}$ is a nonempty r-uniform hypergraph. Then, for every $R \geqslant r$, there is an $\mathcal{H}' \subseteq \mathcal{H}$ with $e(\mathcal{H}') > (1 - \frac{r}{R}) \cdot e(\mathcal{H})$ such that $\Delta_1(\mathcal{H}') \leqslant R \cdot \hat{\Delta}_1(\mathcal{H})$.*

Proof. Given a nonempty r-uniform hypergraph $\mathcal{H}$ and $R \geqslant r$, define

$$X = \left\{ v \in V(\mathcal{H}) : \deg_{\mathcal{H}} v > R \cdot \hat{\Delta}_1(\mathcal{H}) \right\}.$$

By the definition of X and $\hat{\Delta}_1(\mathcal{H})$, we have

$$\|\sigma_{\mathcal{H}}\|^2 \geqslant \sum_{v \in X} \left(\frac{\deg_{\mathcal{H}} v}{r \cdot e(\mathcal{H})} \right)^2 > \frac{R \cdot \hat{\Delta}_1(\mathcal{H})}{r \cdot e(\mathcal{H})} \cdot \sum_{v \in X} \frac{\deg_{\mathcal{H}} v}{r \cdot e(\mathcal{H})} = R \cdot \|\sigma_{\mathcal{H}}\|^2 \cdot \sum_{v \in X} \frac{\deg_{\mathcal{H}} v}{r \cdot e(\mathcal{H})},$$

which implies that

$$\sum_{v \in X} \deg_{\mathcal{H}} v < \frac{r}{R} \cdot e(\mathcal{H}).$$

In particular, deleting from $\mathcal{H}$ all edges containing at least one vertex of X yields a hypergraph $\mathcal{H}'$ satisfying the assertion of this lemma. $\qquad\square$

Since the degree measures $\sigma^{(t)}(\mathcal{H})$ are not defined when $\mathcal{H}$ is an empty hypergraph, in order to streamline our analysis, we will start building the hypergraph $\mathcal{H}^{(r)}$ by seeding it with a fixed well-behaved r-uniform hypergraph. In order to guarantee that, at the end of the algorithm, this initial seed constitutes only a negligible proportion of $\mathcal{H}^{(r)}$, we need to make sure that the link hypergraphs $\mathcal{H}_v^{(r+1)}$ that the algorithm adds to $\mathcal{H}^{(r)}$ are somewhat large. We will achieve this by (temporarily) removing vertices of very small degree from various subhypergraphs of $\mathcal{H}^{(r+1)}$.

Fact 4.6. *Suppose that $\mathcal{H}$ is a nonempty hypergraph with vertex set V. Then, for every $\beta > 0$, there is a spanning $\mathcal{H}' \subseteq \mathcal{H}$ such that $e(\mathcal{H}') \geqslant (1 - \beta) \cdot e(\mathcal{H})$ and, for each $v \in V$, $\deg_{\mathcal{H}'} v$ is either zero or at least $\beta \cdot e(\mathcal{H})/|V|$.*

Proof. Form a spanning subgraph $\mathcal{H}'$ of $\mathcal{H}$ by iteratively deleting all edges containing some vertex with degree smaller than $\beta \cdot e(\mathcal{H})/|V|$. Clearly, each edge of $\mathcal{H}'$ contains only vertices with degrees at least $\beta \cdot e(\mathcal{H})/|V|$. Moreover, the number of edges deleted from $\mathcal{H}$ while forming $\mathcal{H}'$ cannot exceed $\beta \cdot e(\mathcal{H})$. $\qquad\square$

4.5. **The algorithm.** We are ready to present the algorithm that underlies the proof of Lemma 4.1.

Input. Let $\mathcal{G}$ be an $(r+1)$-uniform hypergraph with vertex set V. Let $\varepsilon \in \left(0, (9(r+1))^{-1}\right)$ and $p \in (0, 1)$, define

$$a = \frac{25}{\varepsilon^2},$$

and suppose that

$$\|\sigma_{\mathcal{G}}\|^2 \leqslant \frac{\varepsilon^3 p}{50(r + 1)} = \frac{\varepsilon p}{2a(r + 1)}. \tag{5}$$

Observe that uniformly scaling the multiplicities of all edges of $\mathcal{G}$ by a positive integer factor k does not affect the t-degree measure $\sigma_{\mathcal{G}}^{(t)}$, for any $t \in [\![r]\!]$, nor does it change the family $\mathcal{I}(\mathcal{G})$ of independent sets of $\mathcal{G}$. It does, however, increase the value of $e(\mathcal{G})$, and thus also the value of $\hat{\Delta}_t(\mathcal{G})$, for each $t \in [\![r]\!]$, by the same multiplicative factor k. Consequently, we may assume, without loss of generality, that there is a (large) positive integer m such that

$$\frac{m}{2a} \cdot \binom{|V|}{r} \leqslant \hat{\Delta}_1(\mathcal{G}) \leqslant \frac{m}{a} \cdot \binom{|V|}{r}. \tag{6}$$

Finally, let $\alpha \in \mathbb{R}^r$ be a vector with nonnegative coordinates and let I be an independent set of $\mathcal{G}$.

Setup. Let L be the empty set and let $\mathcal{G}_*^{(0)}$ be the hypergraph obtained from the complete r-uniform hypergraph with vertex set V by changing the multiplicities of all of its edges to m, so that $e(\mathcal{G}_*^{(0)}) = m\binom{|V|}{r}$. Further, apply Lemma 4.5, with $R \leftarrow \frac{r+1}{\varepsilon}$, to find an $\mathcal{A}^{(0)} \subseteq \mathcal{G}$ satisfying

$$e(\mathcal{A}^{(0)}) \geqslant (1-\varepsilon) \cdot e(\mathcal{G}) \qquad \text{and} \qquad \Delta_1(\mathcal{A}^{(0)}) \leqslant \frac{r+1}{\varepsilon} \cdot \hat{\Delta}_1(\mathcal{G}). \tag{7}$$

Finally, let

$$b = \left\lceil \frac{2p}{\varepsilon} \cdot |V| \right\rceil \qquad \text{and} \qquad \sigma = (1-3\varepsilon)^{-1} \cdot \|\sigma_\alpha(\mathcal{G})\|.$$

Main loop. Do the following for $j = 0, 1, \ldots$:

(S1) If $|L| = b$ or $e(\mathcal{A}^{(j)}) < (1-2\varepsilon) \cdot e(\mathcal{G})$, then let $J = j$ and STOP.

(S2) Let $\hat{\mathcal{A}}^{(j)}$ be a canonically chosen spanning subgraph of $\mathcal{A}^{(j)}$ satisfying the assertion of Fact 4.6 with $\beta \leftarrow \varepsilon$. For each $v \in V$, let $\mathcal{G}_*^{(j,v)} = \mathcal{G}_*^{(j)} \cup \hat{\mathcal{A}}_v^{(j)}$ and let v_j be a canonically chosen vertex that minimises the quantity

$$e(\mathcal{G}_*^{(j,v)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j,v)})\|^2 - (1+\varepsilon) \cdot \sigma^2 \right)$$

over all $v \in V$ whose degree in $\hat{\mathcal{A}}^{(j)}$ is nonzero.

(S3) If $v_j \in I$, then add j to the set L and let

$$\mathcal{G}_*^{(j+1)} = \mathcal{G}_*^{(j,v_j)} = \mathcal{G}_*^{(j)} \cup \hat{\mathcal{A}}_{v_j}^{(j)}.$$

Otherwise, let $\mathcal{G}_*^{(j+1)} = \mathcal{G}_*^{(j)}$.

(S4) Let $\mathcal{A}^{(j+1)}$ be the hypergraph obtained from $\mathcal{A}^{(j)}$ by removing all edges containing v_j.

Output. After STOP is called in the main loop, let $\mathcal{A} = \mathcal{A}^{(J)}$ and $\mathcal{G}_* = \mathcal{G}_*^{(J)} \setminus \mathcal{G}_*^{(0)}$, that is, $\mathcal{G}_*$ is the hypergraph satisfying $\mathcal{G}_*^{(J)} = \mathcal{G}_* \cup \mathcal{G}_*^{(0)}$. (Recall once more that all our hypergraphs contain edges with multiplicities.)

4.6. **Basic properties of the algorithm and the key dichotomy.** In this section, we establish several basic properties of the algorithm and state its key 'dichotomy' property, which we shall derive in later sections. Moreover, we explain how to use the algorithm to prove Lemma 4.1. We start by showing that the algorithm terminates on every input and that the output hypergraph $\mathcal{G}_*$ and the final set L retain important information about the input set I.

Observation 4.7. *For every $I \in \mathcal{I}(\mathcal{G})$, the map $\{0, \ldots, J-1\} \ni j \mapsto v_j \in V$ is injective and hence the algorithm terminates. Moreover, $I \in \mathcal{I}(\mathcal{G}_*)$ and $L = \{ j \in \{0, \ldots, J-1\} : v_j \in I \}$.*

Proof. The first assertion holds because in step (S4), all edges containing v_j are removed, and hence its degree remains zero in each $\mathcal{A}^{(j')}$ with $j' > j$. Therefore, the algorithm stops after at most $|V|$ iterations of the main loop. The second assertion holds because $\mathcal{G}_*$ comprises only edges of the link hypergraphs $\mathcal{G}_v$ for which $v \in I$ and because $j \in L$ if and only if $v_j \in I$. $\qquad\square$

We next observe that the set L contains all the information about the input set I that is needed to reconstruct the execution of the algorithm.

Observation 4.8. *If the algorithm produces the same set L for two inputs $I, I' \in \mathcal{I}(\mathcal{G})$, then also the hypergraphs $\mathcal{G}_*$, the numbers J, and the sequences $v_0, \ldots, v_{J-1}$ are the same for both inputs.*

Proof. The only decisions that depend on the input set I are taken in step (S3) of the algorithm. Each time this step is executed, the decision taken is encoded in the set L by placing, or not placing, the index j in L. $\qquad\square$

The function $S\colon \mathcal{I}(\mathcal{G}) \to \mathcal{P}(V)$ whose existence is asserted by Lemma 4.1 will be defined as follows:

$$S_I = \{v_j \colon j \in L\} = \{v_j \colon 0 \leqslant j < J \text{ and } v_j \in I\}. \tag{8}$$

In other words, S_I comprises precisely those among the queried vertices $v_0, \ldots, v_{J-1}$ that belong to the input set I. Note that $|S_I| = |L| \leqslant b$, since the algorithm terminates in step (S1) as soon as L has b elements. We shall now show that the knowledge of the set S_I is enough to reconstruct the final set L and hence, as stated in Observation 4.8, the entire execution of the algorithm. In fact, the following stronger statement is true.

Lemma 4.9. *Suppose that, for two inputs $I, I' \in \mathcal{I}(\mathcal{G})$, we have $S_I \subseteq I'$ and $S_{I'} \subseteq I$. Then, for both these inputs, the algorithm outputs the same set L.*

Suppose that $S_I = S_{I'}$ for some $I, I' \in \mathcal{I}(\mathcal{G})$. As $S_I \subseteq I$ and $S_{I'} \subseteq I'$, by construction, Lemma 4.9 implies that the output set L must be the same for both I and I'.

Proof of Lemma 4.9. Suppose that two inputs I and I' yield sets L and L', respectively, with $L \neq L'$. Let j be the smallest index such that $j \in (L \setminus L') \cup (L' \setminus L)$; without loss of generality, we may assume that $j \in L \setminus L'$. Since $L \cap \{0, \ldots, j-1\} = L' \cap \{0, \ldots, j-1\}$, the algorithm produces the same sequences $v_0, \ldots, v_j$ while working with inputs I and I'. Since $j \in L \setminus L'$, we must have $v_j \in S_I$ and $v_j \notin I'$. In particular, $S_I \nsubseteq I'$. $\qquad\square$

Finally, define the vector $\alpha^* \in \mathbb{R}^{r+1}$ as in (3):

$$\alpha^* = (1+\varepsilon)^{10} \cdot (\alpha, 0) + \frac{50(r+1)}{\varepsilon^2 p} \cdot (0, \alpha).$$

The key dichotomy property, stated in our next lemma, is that either the algorithm inspects many vertices of the hypergraph (before encountering the bth vertex of I) or the final hypergraph $\mathcal{G}_*$ is a good 'model' of $\mathcal{G}$, in the sense that the ℓ^2-norm of $\sigma_\alpha(\mathcal{G}_*)$ does not exceed the ℓ^2-norm of $\sigma_{\alpha^*}(\mathcal{G})$.

Lemma 4.10. *At least one of the following holds:*
  (1) $J \geqslant \frac{\varepsilon^2}{(r+1)^2} \cdot \|\sigma_\mathcal{G}\|^{-2}$,
  (2) $\|\sigma_\alpha(\mathcal{G}_*)\| \leqslant \|\sigma_{\alpha^*}(\mathcal{G})\|$.

We shall prove Lemma 4.10, which lies at the heart of the matter, in the next two sections. We finish the current section with a short derivation of Lemma 4.1, which is now straightforward. Given an $(r+1)$-uniform hypergraph $\mathcal{G}$ and numbers ε and p as in the statement of the lemma, we may define the function $S\colon I(\mathcal{G}) \to \mathcal{P}(V)$ as in (8), by running the algorithm on each input $I \in \mathcal{I}(\mathcal{G})$. If $J \geqslant \frac{\varepsilon^2}{(r+1)^2} \cdot \|\sigma_\mathcal{G}\|^{-2}$, we place S_I in the family $\mathcal{S}'$ and let $C(S_I) = V \setminus \{v_j : 0 \leqslant j < J\}$; note that $I \setminus S_I \subseteq C(S_I)$ as $S_I = I \cap (V \setminus C(S_I))$ by (8). Otherwise, we place S_I in the family $\mathcal{S}''$ and let $\mathcal{F}(S_I) = \mathcal{G}_*$; Lemma 4.10 implies the desired property of each such hypergraph $\mathcal{F}(S_I)$. Lemma 4.9

and Observation 4.8 guarantee that the set $C(S_I)$ or the hypergraph $\mathcal{F}(S_I)$ depend only on the set S_I, and not on I itself, and that the function S has the claimed consistency property.

4.7. **The geometric lemma.** The most important elementary operation performed by the algorithm described in Section 4.5 is to choose some $v \in V$ and add the r-uniform link hypergraph $\hat{\mathcal{A}}_v^{(j)}$ to the hypergraph $\mathcal{G}_*^{(j)}$, obtaining a new hypergraph $\mathcal{G}_*^{(j,v)} = \mathcal{G}_*^{(j)} \cup \hat{\mathcal{A}}_v^{(j)}$. Since we want the final hypergraph $\mathcal{G}_*$ to have small ℓ^2-norm of $\sigma_\alpha(\mathcal{G}_*)$, in step (S2) of the algorithm, we consider a vertex v that, essentially, minimises the ℓ^2-norm of $\sigma_\alpha(\mathcal{G}_*^{(j,v)})$ over all eligible $v \in V$. The following proposition, which is the core of the proof of Theorem 2.1, bounds the minimum of $\|\sigma_\alpha(\mathcal{G}_*^{(j,v)})\|$ from above. Given an $(r+1)$-uniform hypergraph $\mathcal{A}$ and a vector $\alpha \in \mathbb{R}^r$, we define

$$\hat{\Delta}_\alpha(\mathcal{A}) = \sum_{t=1}^{r} \alpha_t \cdot \hat{\Delta}_{t+1}(\mathcal{A}).$$

Proposition 4.11. *Suppose that $\mathcal{A}$ is an $(r+1)$-uniform hypergraph with vertex set V and that $\mathcal{G}_*$ is an r-uniform hypergraph with the same vertex set. Suppose that $\alpha \in \mathbb{R}^r$ has nonnegative coordinates. Then, there exists a vertex $v \in V$ with nonzero degree in $\mathcal{A}$ such that the hypergraph $\mathcal{G}_*^v = \mathcal{G}_* \cup \mathcal{A}_v$ satisfies*

$$\|\sigma_\alpha(\mathcal{G}_*^v)\|^2 \leqslant \|\sigma_\alpha(\mathcal{G}_*)\|^2$$
$$+ \frac{\deg_{\mathcal{A}} v}{e(\mathcal{G}_*^v)} \cdot \left(\left(2 \cdot \frac{\|\sigma_\alpha(\mathcal{A})\|}{\|\sigma_\alpha(\mathcal{G}_*)\|} - 2 + \frac{\hat{\Delta}_1(\mathcal{A})}{e(\mathcal{G}_*)} \right) \cdot \|\sigma_\alpha(\mathcal{G}_*)\|^2 + \frac{\hat{\Delta}_\alpha(\mathcal{A})}{e(\mathcal{G}_*)} \right). \quad (9)$$

Since the right-hand side of (9) is rather complicated, let us explain the underlying intuition. The two terms $\hat{\Delta}_1(\mathcal{A})/e(\mathcal{G}_*)$ and $\hat{\Delta}_\alpha(\mathcal{A})/e(\mathcal{G}_*)$ should be viewed as 'error terms'. If we assumed that they are both zero, inequality (9) would simplify to

$$\|\sigma_\alpha(\mathcal{G}_*^v)\|^2 \leqslant \|\sigma_\alpha(\mathcal{G}_*)\|^2 + \frac{2 \deg_{\mathcal{A}} v}{e(\mathcal{G}_*^v)} \cdot \left(\|\sigma_\alpha(\mathcal{A})\| - \|\sigma_\alpha(\mathcal{G}_*)\| \right) \cdot \|\sigma_\alpha(\mathcal{G}_*)\|. \quad (10)$$

This simplified inequality (10) states that, as long as the ℓ^2-norm of $\sigma_\alpha(\mathcal{G}_*)$ exceeds that of $\sigma_\alpha(\mathcal{A})$, there is a vertex $v \in V$ such that the ℓ^2-norm of $\sigma_\alpha(\mathcal{G}_*^v)$ is strictly smaller than that of $\sigma_\alpha(\mathcal{G}_*)$. Moreover, the difference $\|\sigma_\alpha(\mathcal{G}_*)\| - \|\sigma_\alpha(\mathcal{G}_*^v)\|$ is proportional to the difference $\|\sigma_\alpha(\mathcal{A})\| - \|\sigma_\alpha(\mathcal{G}_*)\|$. Proposition 4.11 will allow us to show that, as we repeatedly update $\mathcal{G}_* \leftarrow \mathcal{G}_*^v$ in step (S3) of the algorithm, the value $\|\sigma_\alpha(\mathcal{G}_*)\|$ drifts, rather quickly, towards $\|\sigma_\alpha(\mathcal{A})\|$.

The reason why Proposition 4.11 is true stems from Fact 3.6, which states that the vector $\sigma_\alpha(\mathcal{A})$ is a convex combination of the vectors $\sigma_\alpha(\mathcal{A}_v)$, where v ranges over V, and the coefficient of each $\sigma_\alpha(\mathcal{A}_v)$ in this combination is proportional to $\deg_{\mathcal{A}} v$. This basic property of the degree measures enables us to express the problem of minimising $\|\sigma_\alpha(\mathcal{G}_*^v)\|$, solved by the proposition, in a simple, abstract way, as we now do in the next lemma.

Lemma 4.12. *Suppose that $\nu_1, \ldots, \nu_k \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}^k$ all have nonnegative coordinates and $\|\lambda\|_1 = \lambda_1 + \cdots + \lambda_k = 1$. Define*

$$\nu = \sum_{i=1}^{k} \lambda_i \cdot \nu_i.$$

For every positive x, every $\mu \in \mathbb{R}^d$ with nonnegative coordinates, and all $x_1, \ldots, x_k \in (0, x]$, there exists an $i \in [\![k]\!]$ such that $\lambda_i > 0$ and the vector μ_i defined by

$$\mu_i = (1 - x_i \lambda_i) \cdot \mu + x_i \lambda_i \cdot \nu_i$$

satisfies

$$\|\mu_i\|^2 \leqslant \|\mu\|^2 + \lambda_i x_i \cdot \left(\left(2 \cdot \frac{\|\nu\|}{\|\mu\|} - 2 + x \cdot \|\lambda\|^2 \right) \cdot \|\mu\|^2 + x \cdot \sum_{j=1}^{k} \lambda_j^2 \|\nu_j\|^2 \right). \qquad (11)$$

Proof. Note first that, for every $i \in [\![k]\!]$,

$$\|\mu_i\|^2 = \|\mu + x_i \lambda_i \cdot (\nu_i - \mu)\|^2 = \|\mu\|^2 + 2x_i \lambda_i \cdot \langle \nu_i - \mu, \mu \rangle + x_i^2 \lambda_i^2 \cdot \|\nu_i - \mu\|^2. \qquad (12)$$

Since $\langle \nu_i, \mu \rangle \geqslant 0$, by our assumption on non-negativity of the coordinates, we have

$$\|\nu_i - \mu\|^2 = \|\nu_i\|^2 - 2 \langle \nu_i, \mu \rangle + \|\mu\|^2 \leqslant \|\mu\|^2 + \|\nu_i\|^2.$$

Substituting this inequality into (12), dividing both sides by x_i, and summing over $i \in [\![k]\!]$ yields

$$\sum_{i=1}^{k} \frac{\|\mu_i\|^2 - \|\mu\|^2}{x_i} \leqslant \sum_{i=1}^{k} 2\lambda_i \cdot \langle \nu_i - \mu, \mu \rangle + \max_i x_i \cdot \left(\|\lambda\|^2 \|\mu\|^2 + \sum_{i=1}^{k} \lambda_i^2 \|\nu_i\|^2 \right). \qquad (13)$$

The definition of ν, the assumption $\|\lambda\|_1 = 1$, and the Cauchy–Schwarz inequality give

$$\sum_{i=1}^{k} \lambda_i \cdot \langle \nu_i - \mu, \mu \rangle = \langle \nu - \mu, \mu \rangle = \langle \nu, \mu \rangle - \|\mu\|^2 \leqslant \|\nu\| \cdot \|\mu\| - \|\mu\|^2.$$

Substituting this inequality into (13), recalling the assumption that $\max_i x_i \leqslant x$, yields

$$\sum_{i=1}^{k} \frac{\|\mu_i\|^2 - \|\mu\|^2}{x_i} \leqslant 2 \left(\|\nu\| \cdot \|\mu\| - \|\mu\|^2 \right) + x \cdot \left(\|\lambda\|^2 \|\mu\|^2 + \sum_{i=1}^{k} \lambda_i^2 \|\nu_i\|^2 \right). \qquad (14)$$

Finally, as $\|\lambda\|_1 = 1$, there must exist an $i \in [\![k]\!]$ such that $\lambda_i > 0$ and the ith summand in the left-hand side of (14) is at most λ_i times the right-hand side of (14). This gives

$$\|\mu_i\|^2 - \|\mu\|^2 \leqslant \lambda_i x_i \cdot \left(2 \left(\|\nu\| \cdot \|\mu\| - \|\mu\|^2 \right) + x \left(\|\lambda\|^2 \|\mu\|^2 + \sum_{j=1}^{k} \lambda_j^2 \|\nu_j\|^2 \right) \right),$$

which is easily seen to be equivalent to the desired inequality (11). $\qquad \square$

Proof of Proposition 4.11. For every $v \in V$, let $\mathcal{G}_*^v = \mathcal{G}_* \cup \mathcal{A}_v$. We claim that, for each $t \in [\![r]\!]$, the t-degree measure of $\mathcal{G}_*^v$ is a convex combination of the t-degree measures of $\mathcal{G}_*$ and $\mathcal{A}_v$ and the coefficients in this convex combination are proportional to $e(\mathcal{G}_*)$ and $\deg_{\mathcal{A}} v$, respectively. Indeed, since for every $T \subseteq V$, the degree of T in $\mathcal{G}_*^v$ is simply the sum of the degrees of T in $\mathcal{G}_*$ and $\mathcal{A}_v$, we have

$$e(\mathcal{G}_*^v) \cdot \sigma_{\mathcal{G}_*^v}^{(t)} = e(\mathcal{G}_*) \cdot \sigma_{\mathcal{G}_*}^{(t)} + e(\mathcal{A}_v) \cdot \sigma_{\mathcal{A}_v}^{(t)} = e(\mathcal{G}_*) \cdot \sigma_{\mathcal{G}_*}^{(t)} + \deg_{\mathcal{A}} v \cdot \sigma_{\mathcal{A}_v}^{(t)}.$$

Dividing the above equality through by $e(\mathcal{G}_*^v) = e(\mathcal{G}_*) + \deg_{\mathcal{A}} v$, we obtain

$$\sigma_{\mathcal{G}_*^v}^{(t)} = \left(1 - \frac{\deg_{\mathcal{A}} v}{e(\mathcal{G}_*^v)} \right) \cdot \sigma_{\mathcal{G}_*}^{(t)} + \frac{\deg_{\mathcal{A}} v}{e(\mathcal{G}_*^v)} \cdot \sigma_{\mathcal{A}_v}^{(t)}.$$

Define, for each $v \in V$,

$$x_v = \frac{(r+1) \cdot e(\mathcal{A})}{e(\mathcal{G}_*^v)} \leqslant \frac{(r+1) \cdot e(\mathcal{A})}{e(\mathcal{G}_*)}. \tag{15}$$

Since $\mathcal{A}$ is $(r+1)$-uniform, we have, for each $v \in V$,

$$x_v \sigma_{\mathcal{A}}(v) = \frac{\deg_{\mathcal{A}} v}{e(\mathcal{G}_*^v)}$$

and, consequently, for each $t \in [\![r]\!]$,

$$\sigma_{\mathcal{G}_*^v}^{(t)} = \left(1 - x_v \sigma_{\mathcal{A}}(v)\right) \cdot \sigma_{\mathcal{G}_*}^{(t)} + x_v \sigma_{\mathcal{A}}(v) \cdot \sigma_{\mathcal{A}_v}^{(t)}. \tag{16}$$

We now invoke Lemma 4.12 with $k = |V|$, the vectors $\nu_1, \dots, \nu_k$ replaced by $\{\sigma_\alpha(\mathcal{A}_v) : v \in V\}$, the vector λ replaced by $\sigma_{\mathcal{A}}$, the vector μ replaced by $\sigma_\alpha(\mathcal{G}_*)$, the numbers $x_1, \dots, x_k$ replaced by $\{x_v : v \in V\}$, and x replaced by $(r+1) \cdot e(\mathcal{A})/e(\mathcal{G}_*)$; note that $x_1, \dots, x_k \leqslant x$, see (15). Fact 3.6 implies that

$$\nu = \sum_{i=1}^{k} \lambda_i \cdot \nu_i = \sum_{v \in V} \sigma_{\mathcal{A}}(v) \cdot \sigma_\alpha(\mathcal{A}_v) = \sigma_\alpha(\mathcal{A}),$$

and, if $i \in [\![k]\!]$ corresponds to $v \in V$, then $\lambda_i x_i = x_v \sigma_{\mathcal{A}}(v) = \deg_{\mathcal{A}} v/e(\mathcal{G}_*^v)$ and thus $\mu_i = \sigma_\alpha(\mathcal{G}_*^v)$. Recalling the definition of $\hat{\Delta}_1(\cdot)$ from (4), we further have

$$x \cdot \|\lambda\|^2 = \frac{(r+1) \cdot e(\mathcal{A})}{e(\mathcal{G}_*)} \cdot \|\sigma_{\mathcal{A}}\|^2 = \frac{\hat{\Delta}_1(\mathcal{A})}{e(\mathcal{G}_*)}.$$

Finally, Fact 3.5 implies that

$$\begin{aligned}
x \cdot \sum_{i=1}^{k} \lambda_i^2 \|\nu_i\|^2 &= \frac{(r+1) \cdot e(\mathcal{A})}{e(\mathcal{G}_*)} \cdot \sum_{v \in V} \sigma_{\mathcal{A}}(v)^2 \cdot \|\sigma_\alpha(\mathcal{A}_v)\|^2 \\
&= \frac{(r+1) \cdot e(\mathcal{A})}{e(\mathcal{G}_*)} \cdot \sum_{t=1}^{r} \alpha_t \cdot \sum_{v \in V} \sigma_{\mathcal{A}}(v)^2 \cdot \|\sigma_{\mathcal{A}_v}^{(t)}\|^2 \\
&= \frac{(r+1) \cdot e(\mathcal{A})}{e(\mathcal{G}_*)} \cdot \sum_{t=1}^{r} \alpha_t \cdot \frac{\|\sigma_{\mathcal{A}}^{(t+1)}\|^2}{t+1} \\
&= \frac{1}{e(\mathcal{G}_*)} \cdot \sum_{t=1}^{r} \alpha_t \cdot \hat{\Delta}_{t+1}(\mathcal{A}) = \frac{\hat{\Delta}_\alpha(\mathcal{A})}{e(\mathcal{G}_*)}.
\end{aligned}$$

It is now straightforward to verify that Lemma 4.12 implies the existence of a vertex $v \in V$ satisfying the assertion of the proposition. $\square$

4.8. **Proof of the key dichotomy property.** In this section, we use Proposition 4.11 to bound the expression from step (S2) in the description of our algorithm. This is the most technically demanding part of the proof. Throughout this section, we use the notation introduced in Section 4.5. We start with an easy dichotomy.

Lemma 4.13. *If $e(\mathcal{A}) < (1 - 2\varepsilon) \cdot e(\mathcal{G})$, then*

$$J \geqslant \frac{\varepsilon^2}{(r+1)^2} \cdot \|\sigma_{\mathcal{G}}\|^{-2}.$$

Proof. Since $\mathcal{A}$ is obtained from $\mathcal{A}^{(0)}$ by removing all edges that contain at least one of the vertices $v_0, \ldots, v_{J-1}$, we have

$$e(\mathcal{A}^{(0)}) - e(\mathcal{A}) \leqslant \sum_{j=0}^{J-1} \deg_{\mathcal{A}^{(0)}} v_J \leqslant J \cdot \Delta_1(\mathcal{A}^{(0)}).$$

Consequently, it follows from (7) and our upper bound on $e(\mathcal{A})$ that

$$J \geqslant \frac{e(\mathcal{A}^{(0)}) - e(\mathcal{A})}{\Delta_1(\mathcal{A}^{(0)})} \geqslant \frac{\varepsilon^2 \cdot e(\mathcal{G})}{(r+1) \cdot \hat{\Delta}_1(\mathcal{G})} = \frac{\varepsilon^2}{(r+1)^2} \cdot \|\sigma_{\mathcal{G}}\|^{-2}. \qquad \square$$

Lemma 4.14. *If $e(\mathcal{A}) \geqslant (1 - 2\varepsilon) \cdot e(\mathcal{G})$, then $e(\mathcal{G}_*) \geqslant p \cdot e(\mathcal{G})$.*

Proof. By construction,

$$e(\mathcal{G}_*) = e(\mathcal{G}_*^{(J)}) - e(\mathcal{G}_*^{(0)}) = \sum_{j=0}^{J-1} \left(e(\mathcal{G}_*^{(j+1)}) - e(\mathcal{G}_*^{(j)}) \right) = \sum_{j \in L} \deg_{\hat{\mathcal{A}}^{(j)}} v_j \geqslant \sum_{j \in L} \frac{\varepsilon \cdot e(\mathcal{A}^{(j)})}{|V|}.$$

Our assumption $e(\mathcal{A}) \geqslant (1 - 2\varepsilon) \cdot e(\mathcal{G})$ implies that the algorithm terminated with $|L| = b \geqslant 2p/\varepsilon \cdot |V|$, see step (S1). Consequently, recalling that $\varepsilon < 1/4$,

$$e(\mathcal{G}_*) \geqslant |L| \cdot \frac{\varepsilon \cdot e(\mathcal{A})}{|V|} \geqslant \frac{2p \cdot |V|}{\varepsilon} \cdot \frac{\varepsilon \cdot (1 - 2\varepsilon) \cdot e(\mathcal{G})}{|V|} \geqslant p \cdot e(\mathcal{G}). \qquad \square$$

Our next lemma lies at the heart of the matter. For brevity, define

$$\hat{\Delta} := 3 \cdot \hat{\Delta}_\alpha(\mathcal{G}).$$

Lemma 4.15. *For every $j \in \{0, \ldots, J\}$,*

$$\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 \leqslant (1 + \varepsilon) \cdot \sigma^2 + \frac{a \cdot \hat{\Delta}}{e(\mathcal{G}_*^{(j)})}. \tag{17}$$

Proof. We prove (17) by induction on j. Since $\mathcal{G}_*^{(0)}$ is (an integer multiple of) the complete r-uniform hypergraph, $\sigma_{\mathcal{G}_*^{(0)}}^{(t)}$ is the uniform probability measure on $\binom{V}{t}$ and, consequently,

$$\|\sigma_{\mathcal{G}_*^{(0)}}^{(t)}\|^2 = \binom{|V|}{t}^{-1}$$

for every $t \in [\![r]\!]$. On the other hand, Fact 3.5 implies that, for every $t \in [\![r]\!]$,

$$\hat{\Delta}_{t+1}(\mathcal{G}) = \frac{r+1}{t+1} \cdot e(\mathcal{G}) \cdot \|\sigma_{\mathcal{G}}^{(t+1)}\|^2 = (r+1) \cdot e(\mathcal{G}) \cdot \sum_{v \in V} \sigma_{\mathcal{G}}(v)^2 \cdot \|\sigma_{\mathcal{G}_v}^{(t)}\|^2.$$

Fact 3.1 implies that, for every $v \in V$ such that $\mathcal{G}_v$ is nonempty,

$$\|\sigma_{\mathcal{G}_v}^{(t)}\|^2 \geqslant \binom{|V|}{t}^{-1} = \|\sigma_{\mathcal{G}_*^{(0)}}^{(t)}\|^2$$

and therefore

$$\hat{\Delta}_{t+1}(\mathcal{G}) \geqslant (r+1) \cdot e(\mathcal{G}) \cdot \sum_{v \in V} \sigma_{\mathcal{G}}(v)^2 \cdot \|\sigma_{\mathcal{G}_*^{(0)}}^{(t)}\|^2 = \hat{\Delta}_1(\mathcal{G}) \cdot \|\sigma_{\mathcal{G}_*^{(0)}}^{(t)}\|^2.$$

Recall from (6) that we have chosen m so that

$$\hat{\Delta}_1(\mathcal{G}) \geqslant \frac{m}{2a} \cdot \binom{|V|}{r} = \frac{e(\mathcal{G}_*^{(0)})}{2a},$$

which, substituted into the previous inequality, implies that

$$\|\sigma_{\mathcal{G}_*^{(0)}}^{(t)}\|^2 \leqslant \frac{2a}{e(\mathcal{G}_*^{(0)})} \cdot \hat{\Delta}_{t+1}(\mathcal{G}). \tag{18}$$

After we multiply both sides of (18) by α_t and sum the resulting inequalities over all $t \in [\![r]\!]$, we obtain

$$\|\sigma_\alpha(\mathcal{G}_*^{(0)})\|^2 \leqslant 2a \cdot \frac{\hat{\Delta}_\alpha(\mathcal{G})}{e(\mathcal{G}_*^{(0)})},$$

which implies (17) when $j = 0$.

Suppose now that $j \geqslant 0$ and assume that (17) holds; we shall show that this inequality remains true after we replace j with $j + 1$. We may assume that $\mathcal{G}_*^{(j+1)} = \mathcal{G}_*^{(j)} \cup \hat{\mathcal{A}}_{v_j}^{(j)}$, as otherwise $\mathcal{G}_*^{(j+1)} = \mathcal{G}_*^{(j)}$ and there is nothing to prove. Let v be a vertex satisfying the assertion of Proposition 4.11 with $\mathcal{A} \leftarrow \hat{\mathcal{A}}^{(j)}$ and $\mathcal{G}_* \leftarrow \mathcal{G}_*^{(j)}$. The vertex v_j was chosen in step (S2) so that

$$e(\mathcal{G}_*^{(j+1)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j+1)})\|^2 - (1 + \varepsilon) \cdot \sigma^2 \right) \leqslant e(\mathcal{G}_*^{(j,v)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j,v)})\|^2 - (1 + \varepsilon) \cdot \sigma^2 \right), \tag{19}$$

so it suffices to bound the right-hand side of (19) from above by $a \cdot \hat{\Delta}$.

The assertion of Proposition 4.11, inequality (9), is equivalent to the inequality

$$e(\mathcal{G}_*^{(j,v)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j,v)})\|^2 - \sigma^2 \right) \leqslant e(\mathcal{G}_*^{(j,v)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 - \sigma^2 \right)$$
$$+ \deg_{\hat{\mathcal{A}}^{(j)}} v \cdot \left(\left(\frac{\hat{\Delta}_1(\hat{\mathcal{A}}^{(j)})}{e(\mathcal{G}_*^{(j)})} + 2 \cdot \frac{\|\sigma_\alpha(\hat{\mathcal{A}}^{(j)})\|}{\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|} - 2 \right) \cdot \|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 + \frac{\hat{\Delta}_\alpha(\hat{\mathcal{A}}^{(j)})}{e(\mathcal{G}_*^{(j)})} \right). \tag{20}$$

Since $e(\mathcal{G}_*^{(j,v)}) = e(\mathcal{G}_*^{(j)}) + \deg_{\hat{\mathcal{A}}^{(j)}} v$, inequality (20) may be rewritten as

$$e(\mathcal{G}_*^{(j,v)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j,v)})\|^2 - \sigma^2 \right) \leqslant e(\mathcal{G}_*^{(j)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 - \sigma^2 \right)$$
$$+ \deg_{\hat{\mathcal{A}}^{(j)}} v \cdot \left(\left(\frac{\hat{\Delta}_1(\hat{\mathcal{A}}^{(j)})}{e(\mathcal{G}_*^{(j)})} + 2 \cdot \frac{\|\sigma_\alpha(\hat{\mathcal{A}}^{(j)})\|}{\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|} - 1 \right) \cdot \|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 - \sigma^2 + \frac{\hat{\Delta}_\alpha(\hat{\mathcal{A}}^{(j)})}{e(\mathcal{G}_*^{(j)})} \right). \tag{21}$$

We shall now simplify the right-hand side of (21) somewhat. To this end, observe first that, as the algorithm did not terminate in step (S1), we must have

$$e(\hat{\mathcal{A}}^{(j)}) \geqslant (1 - \varepsilon) \cdot e(\mathcal{A}^{(j)}) \geqslant (1 - \varepsilon) \cdot (1 - 2\varepsilon) \cdot e(\mathcal{G}) \geqslant (1 - 3\varepsilon) \cdot e(\mathcal{G}).$$

Consequently, Fact 3.2 implies that, for every $t \in [\![r + 1]\!]$,

$$\|\sigma_{\hat{\mathcal{A}}^{(j)}}^{(t)}\| \leqslant (1 - 3\varepsilon)^{-1} \cdot \|\sigma_{\mathcal{G}}^{(t)}\| \tag{22}$$

and, since clearly $e(\hat{\mathcal{A}}^{(j)}) \leqslant e(\mathcal{G})$,

$$\hat{\Delta}_t(\hat{\mathcal{A}}^{(j)}) \leqslant (1 - 3\varepsilon)^{-2} \cdot \hat{\Delta}_t(\mathcal{G}) \leqslant 3 \cdot \hat{\Delta}_t(\mathcal{G}). \tag{23}$$

Summing (22), with both sides squared, and (23) over all t, with appropriate weights, yields

$$\|\sigma_\alpha(\hat{\mathcal{A}}^{(j)})\| \leqslant (1 - 3\varepsilon)^{-1} \cdot \|\sigma_\alpha(\mathcal{G})\| = \sigma \qquad \text{and} \qquad \hat{\Delta}_\alpha(\hat{\mathcal{A}}^{(j)}) \leqslant \hat{\Delta}. \tag{24}$$

Furthermore, recall from (6) that we have chosen m so that

$$\hat{\Delta}_1(\mathcal{G}) \leqslant \frac{m}{a} \cdot \binom{|V|}{r} = \frac{e(\mathcal{G}_*^{(0)})}{a} \leqslant \frac{e(\mathcal{G}_*^{(j)})}{a}.$$

Consequently

$$\frac{\hat{\Delta}_1(\hat{\mathcal{A}}^{(j)})}{e(\mathcal{G}_*^{(j)})} \leqslant \frac{(1-3\varepsilon)^{-2}}{a} \leqslant \frac{3}{a} \qquad (25)$$

and, by (7),

$$\Delta_1(\hat{\mathcal{A}}^{(j)}) \leqslant \Delta_1(\mathcal{A}^{(j)}) \leqslant \Delta_1(\mathcal{A}^{(0)}) \leqslant \frac{r+1}{\varepsilon} \cdot \hat{\Delta}_1(\mathcal{G})$$

$$\leqslant \frac{r+1}{\varepsilon} \cdot \frac{e(\mathcal{G}_*^{(j)})}{a} \leqslant \frac{a}{2} \cdot e(\mathcal{G}_*^{(j)}), \qquad (26)$$

where the last inequality holds as $a^2 = 625/\varepsilon^4 \geqslant 2(r+1)/\varepsilon$. We may now substitute (24) and (25) into (21) and rearrange the terms to obtain the following inequality:

$$e(\mathcal{G}_*^{(j,v)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j,v)})\|^2 - \sigma^2 \right) \leqslant e(\mathcal{G}_*^{(j)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 - \sigma^2 \right)$$

$$+ \deg_{\hat{\mathcal{A}}^{(j)}} v \cdot \left(\left(\left(\frac{3}{a} - \left(\frac{\sigma}{\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|} - 1 \right)^2 \right) \cdot \|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 + \frac{\hat{\Delta}}{e(\mathcal{G}_*^{(j)})} \right). \quad (27)$$

We now consider two cases, depending on how large $\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2$ is.

Case 1. $\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 \leqslant (1+\varepsilon) \cdot \sigma^2 + a/2 \cdot \hat{\Delta}/e(\mathcal{G}_*^{(j)})$.

We first claim that

$$\left(\frac{3}{a} - \left(\frac{\sigma}{\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|} - 1 \right)^2 \right) \cdot \|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 \leqslant \frac{12\sigma^2}{a} \leqslant \varepsilon \cdot \sigma^2. \qquad (28)$$

To see this, note that the left-hand side of (28) is negative when $\|\sigma_\alpha(\mathcal{G}_*^{(j)})\| > 2\sigma$, as $a \geqslant 12$. Otherwise, the first factor in the left-hand side is at most $3/a$ and the second factor is at most $4\sigma^2$. Substituting (28) into (27), using the assumed upper bound on $\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2$, yields

$$e(\mathcal{G}_*^{(j,v)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j,v)})\|^2 - \sigma^2 \right) \leqslant e(\mathcal{G}_*^{(j)}) \cdot \varepsilon \cdot \sigma^2 + \frac{a}{2} \cdot \hat{\Delta} + \deg_{\hat{\mathcal{A}}^{(j)}} v \cdot \left(\varepsilon \cdot \sigma^2 + \frac{\hat{\Delta}}{e(\mathcal{G}_*^{(j)})} \right)$$

$$\leqslant \left(e(\mathcal{G}_*^{(j)}) + \deg_{\hat{\mathcal{A}}^{(j)}} v \right) \cdot \varepsilon \cdot \sigma^2 + \left(\frac{a}{2} + \frac{a}{2} \right) \cdot \hat{\Delta},$$

where the second inequality holds because $\deg_{\hat{\mathcal{A}}^{(j)}} v \leqslant \Delta_1(\hat{\mathcal{A}}^{(j)}) \leqslant (a/2) \cdot e(\mathcal{G}_*^{(j)})$, see (26). Finally, since $e(\mathcal{G}_*^{(j)}) + \deg_{\hat{\mathcal{A}}^{(j)}} v = e(\mathcal{G}_*^{(j,v)})$, we may conclude that

$$e(\mathcal{G}_*^{(j,v)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j,v)})\|^2 - (1+\varepsilon) \cdot \sigma^2 \right) \leqslant a \cdot \hat{\Delta}. \qquad (29)$$

By (19), this proves the desired estimate (inequality (17) with j replaced by $j+1$).

Case 2. $(1+\varepsilon) \cdot \sigma^2 + a/2 \cdot \hat{\Delta}/e(\mathcal{G}_*^{(j)}) < \|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 \leqslant (1+\varepsilon) \cdot \sigma^2 + a \cdot \hat{\Delta}/e(\mathcal{G}_*^{(j)})$.

We will show that the second term in the right-hand side of (27) is nonpositive, which will give

$$e(\mathcal{G}_*^{(j,v)}) \cdot \|\sigma_\alpha(\mathcal{G}_*^{(j,v)})\|^2 \leqslant e(\mathcal{G}_*^{(j)}) \cdot \left(\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 - \sigma^2\right) + e(\mathcal{G}_*^{(j,v)}) \cdot \sigma^2$$

$$\leqslant e(\mathcal{G}_*^{(j)}) \cdot \varepsilon \cdot \sigma^2 + a \cdot \hat{\Delta} + e(\mathcal{G}_*^{(j,v)}) \cdot \sigma^2,$$

which in turn implies the desired estimate (inequality (29)), as $e(\mathcal{G}_*^{(j)}) \leqslant e(\mathcal{G}_*^{(j,v)})$. The lower bound on $\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2$ assumed in Case 2 implies that

$$\frac{\hat{\Delta}}{e(\mathcal{G}_*^{(j)})} < \frac{2}{a} \cdot \|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2$$

and hence

$$\left(\frac{3}{a} - \left(\frac{\sigma}{\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|} - 1\right)^2\right) \cdot \|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 + \frac{\hat{\Delta}}{e(\mathcal{G}_*^{(j)})}$$

$$\leqslant \left(\frac{5}{a} - \left(\frac{\sigma}{\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|} - 1\right)^2\right) \cdot \|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2.$$

Moreover, since $\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|^2 \geqslant (1+\varepsilon) \cdot \sigma^2$, we have

$$\frac{5}{a} - \left(\frac{\sigma}{\|\sigma_\alpha(\mathcal{G}_*^{(j)})\|} - 1\right)^2 \leqslant \frac{5}{a} - \left(\frac{1}{\sqrt{1+\varepsilon}} - 1\right)^2 = \frac{5}{a} - \frac{\left(\sqrt{1+\varepsilon} - 1\right)^2}{1+\varepsilon}$$

$$= \frac{\varepsilon^2}{5} - \frac{\varepsilon^2}{\left(1 + \sqrt{1+\varepsilon}\right)^2 \cdot (1+\varepsilon)} \leqslant 0,$$

where the last inequality holds due to our assumption that $\varepsilon < (9(r+1))^{-1} \leqslant 1/18$. $\quad\square$

4.9. **Proof of the key lemma.** After having made all the preparations, we are finally ready to prove Lemma 4.10.

Proof of Lemma 4.10. If $e(\mathcal{A}) < (1 - 2\varepsilon) \cdot e(\mathcal{G})$, then Lemma 4.13 immediately gives (1). We may therefore assume that $e(\mathcal{A}) \geqslant (1 - 2\varepsilon) \cdot e(\mathcal{G})$. Recall from (6) that we have chosen m so that

$$e(\mathcal{G}_*^{(0)}) = m \cdot \binom{|V|}{r} \leqslant 2a \cdot \hat{\Delta}_1(\mathcal{G}) = 2a \cdot (r+1) \cdot e(\mathcal{G}) \cdot \|\sigma_\mathcal{G}\|^2.$$

It thus follows from (5) and Lemma 4.14 that

$$e(\mathcal{G}_*^{(0)}) \leqslant 2a \cdot (r+1) \cdot \frac{e(\mathcal{G}_*)}{p} \cdot \frac{\varepsilon p}{2a(r+1)} \leqslant \varepsilon \cdot e(\mathcal{G}_*).$$

Consequently, Fact 3.2 implies that

$$\|\sigma_\alpha(\mathcal{G}_*)\| \leqslant \frac{e(\mathcal{G}_*^{(J)})}{e(\mathcal{G}_*)} \cdot \|\sigma_\alpha(\mathcal{G}_*^{(J)})\| = \frac{e(\mathcal{G}_*) + e(\mathcal{G}_*^{(0)})}{e(\mathcal{G}_*)} \cdot \|\sigma_\alpha(\mathcal{G}_*^{(J)})\| \leqslant (1+\varepsilon) \cdot \|\sigma_\alpha(\mathcal{G}_*^{(J)})\|.$$

It now follows from Lemma 4.15 that

$$\|\sigma_\alpha(\mathcal{G}_*)\|^2 \leqslant (1+\varepsilon)^3 \cdot \sigma^2 + (1+\varepsilon)^2 \cdot \frac{a \cdot \hat{\Delta}}{e(\mathcal{G}_*^{(J)})}.$$

Since $e(\mathcal{G}_*^{(J)}) \geqslant e(\mathcal{G}_*) \geqslant p \cdot e(\mathcal{G})$, by Lemma 4.14 and $(1+\varepsilon)^7(1-3\varepsilon)^2 \geqslant 1$, as $\varepsilon < (9(r+1))^{-1} \leqslant 1/18$, we further have

$$
\begin{aligned}
\|\sigma_\alpha(\mathcal{G}_*)\|^2 &\leqslant (1+\varepsilon)^3 \cdot \sigma^2 + (1+\varepsilon)^2 \cdot \frac{a \cdot \hat{\Delta}}{p \cdot e(\mathcal{G})} \\
&\leqslant \frac{(1+\varepsilon)^3}{(1-3\varepsilon)^2} \cdot \|\sigma_\alpha(\mathcal{G})\|^2 + (1+\varepsilon)^2 \cdot \frac{3a \cdot \hat{\Delta}_\alpha(\mathcal{G})}{p \cdot e(\mathcal{G})} \\
&\leqslant (1+\varepsilon)^{10} \cdot \|\sigma_\alpha(\mathcal{G})\|^2 + \frac{4a \cdot \hat{\Delta}_\alpha(\mathcal{G})}{p \cdot e(\mathcal{G})} \\
&= \sum_{t=1}^{r} \alpha_t \cdot \left((1+\varepsilon)^{10} \cdot \|\sigma_\mathcal{G}^{(t)}\|^2 + \frac{4a}{p} \cdot \frac{r+1}{t+1} \cdot \|\sigma_\mathcal{G}^{(t+1)}\|^2 \right).
\end{aligned}
$$

Recall the definition of α^* given in (3). Since $4a/(t+1) \leqslant 50/\varepsilon^2$ for every $t \in [\![r]\!]$, we may conclude that

$$
\|\sigma_\alpha(\mathcal{G}_*)\|^2 \leqslant \|\sigma_{\alpha^*}(\mathcal{G})\|^2,
$$

as claimed. $\hfill\square$

5. Probabilistic inequalities

The proofs of Theorems 1.2, 1.3, and 1.4 make use of well-known probabilistic inequalities. The first of them are standard tail bounds for binomial distributions.

Lemma 5.1. *Let n be a positive integer, let $p \in [0,1]$, and suppose that $X \sim \mathrm{Bin}(n,p)$. Then, for every $\delta \in [0,1]$,*

$$
\Pr\left(X \leqslant (1-\delta)np \right) \leqslant \exp(-\delta^2 np/2) \quad and \quad \Pr\left(X \geqslant (1+\delta)np \right) \leqslant \exp(-\delta^2 np/3).
$$

We shall also need the following version of Janson's inequality [27], which can be easily deduced from the statements found in [2, Chapter 8].

Theorem 5.2 (Janson's inequality)**.** *Suppose that Ω is a finite set and let $B_1, \dots, B_k$ be arbitrary subsets of Ω. Form a random subset $R \subseteq \Omega$ by independently keeping each $\omega \in \Omega$ with probability $p_\omega \in [0,1]$. For each $i \in [\![k]\!]$, let X_i be the indicator of the event that $B_i \subseteq R$ and define*

$$
\mu = \sum_{i=1}^{k} \mathbb{E}[X_i] \qquad and \qquad \Delta = \sum_{\substack{1 \leqslant i < j \leqslant k \\ B_i \cap B_j \neq \emptyset}} \mathbb{E}[X_i X_j].
$$

Then,

$$
\Pr\left(B_i \not\subseteq R \text{ for all } i \in [\![k]\!] \right) \leqslant \exp\left(-\min\left\{ \frac{\mu}{2}, \frac{\mu^2}{4\Delta} \right\} \right).
$$

6. The typical structure of K_{r+1}-free graphs

6.1. **Outline.** The first, key part of the proof of Theorem 1.2 is showing that, for sufficiently small δ, the number of K_{r+1}-free subgraphs of K_n that are not δn^2-close to being r-partite is much smaller than $2^{\mathrm{ex}(n, K_{r+1})}$, which is a trivial lower bound on the number of K_{r+1}-free graphs. This statement is derived from a container lemma for K_{r+1}-free graphs (Proposition 6.1 below), which is obtained by applying Theorem 1.6 to the $\binom{r+1}{2}$-uniform hypergraph that encodes copies of K_{r+1} in K_n, and the 'supersaturated' version

of the stability theorem of Erdős and Simonovits proved in [5] and stated as Lemma 6.3 below. Proposition 6.1, which is the main result of this section, supplies a covering of all K_{r+1}-free subgraphs of K_n with few containers, each of which is a subgraph of K_n with either fewer than $n^2/8$ edges or fewer than $n^{r+1/2}$ copies of K_{r+1} (after we delete from it some $n^{2-1/(8r)}$ edges), whereas Lemma 6.3 is used to show that all containers with nearly $\mathrm{ex}(n, K_{r+1})$ edges must be close to being r-partite.

The remainder of the proof is showing that all but an $2^{-n/(10r)^4}$-proportion of K_{r+1}-free subgraphs of K_n that are δn^2-close to being r-partite are in fact r-partite. Our three-step argument is loosely based on the methods of [8]. First, we show that all but a tiny fraction of graphs in our collection admit an optimal, balanced r-partition with at most δn^2 monochromatic edges (i.e., edges whose both endpoints belong to the same part of the partition); an r-partition is optimal if it minimises the number of monochromatic edges and it is balanced if each partite set comprises at least $n/(2r)$ vertices. Second, we bound from above the number of remaining graphs whose associated r-partition induces a monochromatic copy of $K_{1,D}$ in one of the parts, where $D = \lfloor n/(2^{14}r^5 \log n) \rfloor$. Third, we bound from above the number of remaining graphs whose associated r-partition induces a monochromatic matching with a given number of edges in one of the parts. The second and third steps complement one another as every graph with t edges contains either a copy of $K_{1,D}$ or a matching with at least t/D edges.

6.2. An efficient container lemma for K_{r+1}-free graphs.

The following statement, which is the main technical result of this section, is an efficient container lemma for K_{r+1}-free subgraphs of K_n. It is obtained by applying Theorem 1.6 to the $\binom{r+1}{2}$-uniform hypergraph that encodes copies of K_{r+1} in K_n.

Proposition 6.1. *For almost all n and every r satisfying $2 \leqslant r \leqslant \log n/(121 \log \log n)$, there exists a collection $\mathcal{G}$ of at most $\exp\left(n^{2-1/(8r)}\right)$ subgraphs of K_n such that:*

(i) *Each K_{r+1}-free subgraph of K_n is contained in some member of $\mathcal{G}$.*

(ii) *Each $G \in \mathcal{G}$ either has fewer than $n^2/8$ edges or it contains a subgraph G' with $e(G') \geqslant e(G) - n^{2-1/(8r)}$ that has fewer than $n^{r+1/2}$ copies of K_{r+1}.*

Proof. Let n be a large integer and suppose that r satisfies $2 \leqslant r \leqslant \log n/(121 \log \log n)$. Let $\gamma = 1/(8r)$ and observe that

$$n^{\gamma} = \exp\left(\frac{\log n}{8r}\right) \geqslant \exp(12 \log \log n) = (\log n)^{12}. \tag{30}$$

Let $\mathcal{H}$ be the $\binom{r+1}{2}$-uniform hypergraph with vertex set $E(K_n)$ whose edges are the edge sets of all copies of K_{r+1} in K_n and set

$$s = \binom{r+1}{2}, \qquad q = n^{-3\gamma}, \qquad \text{and} \qquad E = n^{r+1/2}.$$

We now verify that we may apply Theorem 1.6, with $\alpha \leftarrow 1/4$ and $\beta \leftarrow n^{-\gamma}$, to the hypergraph $\mathcal{H}$. First, as $s \leqslant r^2 \leqslant (\log n)^2$, we have

$$\alpha\beta q \cdot v(\mathcal{H}) = \frac{n^{-4\gamma}}{4} \cdot \binom{n}{2} \geqslant n \geqslant 10^9 s^7$$

and, using (30),

$$10^4 s^5 q = 10^4 s^5 n^{-3\gamma} \leqslant 10^4 (\log n)^{10} n^{-3\gamma} \leqslant n^{-\gamma} = \beta,$$

provided that n is sufficiently large. Second, suppose that $t \in \{2, \ldots, s\}$ and let $\ell \in \{3, \ldots, r + 1\}$ be the unique integer satisfying $\binom{\ell-1}{2} < t \leqslant \binom{\ell}{2}$, so that

$$t - 1 \leqslant \binom{\ell}{2} - 1 = \frac{(\ell+1)(\ell-2)}{2} \leqslant r(\ell - 2).$$

Since a graph with t edges must have at least ℓ vertices, we have

$$\Delta_t(\mathcal{H}) = \binom{n - \ell}{r + 1 - \ell} \leqslant n^{r+1-\ell}$$

and, consequently,

$$\left(\frac{q}{10^6 s^5}\right)^{t-1} \cdot \frac{E}{v(\mathcal{H})} \cdot \frac{1}{\Delta_t(\mathcal{H})} \geqslant \left(\frac{q}{10^6 s^5}\right)^{r(\ell-2)} \cdot \frac{n^{r+1/2}}{n^2} \cdot \frac{1}{n^{r+1-\ell}}$$

$$= \left(\frac{n^{-3\gamma}}{10^6 s^5}\right)^{r(\ell-2)} \cdot n^{\ell-5/2} \geqslant \left(\frac{n^{1/(2r)-3\gamma}}{10^6 s^5}\right)^{r(\ell-2)}$$

$$\geqslant \left(\frac{n^\gamma}{10^6 (\log n)^{10}}\right)^{r(\ell-2)} \geqslant 1,$$

where the last inequality follows from (30). Theorem 1.6 supplies a collection $\mathcal{G}$ of containers for the independent sets of $\mathcal{H}$ (that is, K_{r+1}-free subgraphs of K_n) satisfying

$$|\mathcal{G}| \leqslant \exp\left(10^4 s^5 \beta^{-1} \log(e/\alpha) \cdot q \log(e/q) \cdot v(\mathcal{H})\right)$$

$$\leqslant \exp\left(10^5 (\log n)^{10} n^\gamma \cdot n^{-3\gamma} \log n \cdot n^2\right) \leqslant \exp\left(n^{2-\gamma}\right),$$

where the last inequality follows from (30), such that, for every $G \in \mathcal{G}$, either $e(G) \leqslant \alpha \cdot v(\mathcal{H}) \leqslant n^2/8$ or there is a subgraph $G' \subseteq G$ with $e(G') \geqslant (1 - \beta)e(G) \geqslant e(G) - n^{2-\gamma}$ and $e(\mathcal{H}[G']) < E$ (that is, G' contains fewer than E copies of K_{r+1}). $\square$

6.3. **Almost all K_{r+1}-free graphs are almost r-partite.** The following theorem, which is a rather straightforward consequence of our container lemma for K_{r+1}-free graphs (Proposition 6.1) and the 'supersaturated' version of the stability theorem of Erdős and Simonovits (Lemma 6.3 below) proved by Balogh, Bushaw, Collares, Liu, Morris, and Sharifzadeh [5], may be viewed as an approximate version of Theorem 1.2. It states that, under the assumptions of Theorem 1.2, almost all K_{r+1}-free subgraphs of K_n are almost r-partite. To make this notion precise, given nonnegative integers r and t with $r \geqslant 2$, we shall say that a graph G is t-close to being r-partite if G can be made r-partite by removing from it at most t edges. In other words, G is t-close to being r-partite if G contains an r-partite subgraph G' with $e(G') \geqslant e(G) - t$. Conversely, we shall say that a graph G is t-far from being r-partite if G is not t-close to being r-partite or, in other words, if $\chi(G') > r$ for every $G' \subseteq G$ with $e(G') \geqslant e(G) - t$.

Theorem 6.2. *The following holds for sufficiently large n and all r satisfying $2 \leqslant r \leqslant \log n/(121 \log \log n)$. Let $\mathcal{F}$ denote the family of K_{r+1}-free subgraphs of K_n that are $(8 \log n)^{-15} n^2$-far from being r-partite. Then*

$$|\mathcal{F}| \leqslant 2^{\mathrm{ex}(n, K_{r+1}) - n}.$$

Lemma 6.3 ([5]). *Suppose that n, r, and t are positive integers. Every n-vertex graph G that is t-far from being r-partite contains at least*

$$\frac{n^{r-1}}{e^{2r} \cdot r!} \left(e(G) + t - \left(1 - \frac{1}{r} \right) \frac{n^2}{2} \right)$$

copies of K_{r+1}.

Proof of Theorem 6.2. Let $\delta = (8 \log n)^{-15}$ so that $\mathcal{F}$ is the family of K_{r+1}-free subgraphs of K_n that are δn^2-far from being r-partite. Let $\mathcal{G}$ be the family of containers for K_{r+1}-free graphs supplied by Proposition 6.1. We partition $\mathcal{G}$ into two parts as follows:

$$\mathcal{G}_1 = \left\{ G \in \mathcal{G} : e(G) \geqslant \left(1 - \frac{1}{r} \right) \frac{n^2}{2} - 2n^{2-1/(8r)} \right\} \quad \text{and} \quad \mathcal{G}_2 = \mathcal{G} \setminus \mathcal{G}_1.$$

Fix an arbitrary $G \in \mathcal{G}_1$. Since $e(G) > n^2/8$, it must be the case that G contains a subgraph G' with

$$e(G') \geqslant e(G) - n^{2-1/(8r)} \geqslant \left(1 - \frac{1}{r} \right) \frac{n^2}{2} - 3n^{2-1/(8r)} \tag{31}$$

that contains fewer than $n^{r+1/2}$ copies of K_{r+1}. Let G' be any such subgraph of G and let t' be the smallest number of edges one can delete from G' to make it r-partite. Lemma 6.3 implies that

$$\frac{n^{r-1}}{e^{2r} \cdot r!} \left(e(G') + t' - 1 - \left(1 - \frac{1}{r} \right) \frac{n^2}{2} \right) < n^{r+1/2}$$

and hence, by (31),

$$t' - 3n^{2-1/(8r)} \leqslant e^{2r} \cdot r! \cdot n^{3/2} + 1 \leqslant r^{4r} \cdot n^{3/2} \leqslant n^{7/4}.$$

Since G' is t'-close to being r-partite, G is t-close from being r-partite for all $t \geqslant 5n^{2-1/(8r)}$, since

$$5n^{2-1/(8r)} > n^{7/4} + 4n^{2-1/(8r)} \geqslant t' + e(G) - e(G').$$

In particular, since

$$n^{-1/(8r)} = \exp \left(-\frac{\log n}{8r} \right) = \exp \left(-\frac{121}{8} \cdot \log \log n \right) \ll \frac{1}{(\log n)^{15}},$$

neither G nor any of its subgraphs can be δn^2-far from being r-partite.

Thus, every graph in $\mathcal{F}$ must be contained in some element of $\mathcal{G}_2$. Since

$$\mathrm{ex}(n, K_{r+1}) \geqslant \left(1 - \frac{1}{r} \right) \binom{n}{2} \geqslant \left(1 - \frac{1}{r} \right) \frac{n^2}{2} - n,$$

we have

$$\frac{|\mathcal{F}|}{2^{\mathrm{ex}(n, K_{r+1})}} \leqslant \sum_{G \in \mathcal{G}_2} 2^{e(G) - \left(1 - \frac{1}{r} \right) \frac{n^2}{2} + n} \leqslant |\mathcal{G}| \cdot 2^{n - 2n^{2-1/(8r)}}$$

$$\leqslant \exp \left(n^{2-1/(8r)} + 2 \log 2 \cdot \left(n - n^{2-1/(8r)} \right) \right) \leqslant 2^{-n},$$

as claimed. $\qquad\qquad\square$

6.4. **Balanced and unbalanced r-partitions.** Let Π be an arbitrary r-partition of $[\![n]\!]$. We shall say that Π is *balanced* if $\min_{P\in\Pi}|P| \geqslant \frac{n}{2r}$ and that it is *unbalanced* otherwise. In the sequel, we denote by K_Π the complete r-partite graph whose colour classes are the parts of Π.

Fact 6.4. *Suppose that $r \geqslant 2$ and let Π be an unbalanced r-partition of $[\![n]\!]$. Then*

$$e(K_\Pi) \leqslant \mathrm{ex}(n, K_{r+1}) - \frac{n^2}{16r^2} + n.$$

Proof. Let $P \in \Pi$ be an arbitrary part satisfying $|P| < \frac{n}{2r}$ and let $Q \in \Pi$ be an arbitrary part satisfying $|Q| \geqslant \frac{n}{r}$. Set

$$m = \left\lfloor \frac{|Q| - |P|}{2} \right\rfloor,$$

let Π' be the partition obtained from Π by moving some m vertices from Q to P, and observe that

$$e(K_{\Pi'}) - e(K_\Pi) = (|P| + m)(|Q| - m) - |P||Q| = (|Q| - |P|) \cdot m - m^2 \geqslant m^2.$$

Since $|Q| - |P| > \frac{n}{2r}$, then $m \geqslant \lfloor \frac{n}{4r} \rfloor$ and, consequently, $m^2 \geqslant \frac{n^2}{16r^2} - n$. Finally, since $K_{\Pi'}$ is an r-partite graph, we have $e(K_{\Pi'}) \leqslant \mathrm{ex}(n, K_{r+1})$ and the claimed upper bound on $e(K_\Pi)$ follows. $\square$

Our next lemma bounds from above the number of subgraphs of K_n that admit an unbalanced r-partition with few monochromatic edges.

Lemma 6.5. *The following holds for all sufficiently large n and all r satisfying $2 \leqslant r \leqslant \log n$. Let $\mathcal{F}$ denote the family of all $G \subseteq K_n$ that satisfy $e(G \setminus K_\Pi) \leqslant \frac{n^2}{(r\log n)^2}$ for some unbalanced r-partition Π. Then*

$$|\mathcal{F}| \leqslant 2^{\mathrm{ex}(n, K_{r+1}) - n}.$$

Proof. Denote by $\mathcal{P}_u$ the family of all unbalanced r-partitions of $[\![n]\!]$. For every $\Pi \in \mathcal{P}_u$, let $\mathcal{F}_\Pi$ denote the family of all graphs $G \subseteq K_n$ that satisfy $e(G \setminus K_\Pi) \leqslant \frac{n^2}{(r\log n)^2}$. We have

$$|\mathcal{F}_\Pi| \leqslant \sum_{t=0}^{n^2/(r\log n)^2} \binom{\binom{n}{2}}{t} \cdot 2^{e(K_\Pi)} \leqslant \left(n^2\right)^{\frac{n^2}{(r\log n)^2}} \cdot 2^{e(K_\Pi)} \leqslant 2^{e(K_\Pi) + \frac{4n^2}{r^2\log n}}.$$

Since $|\mathcal{P}_u| \leqslant r^n \leqslant 2^{2n\log n}$, Fact 6.4 gives

$$|\mathcal{F}| \leqslant \sum_{\Pi\in\mathcal{P}_u} |\mathcal{F}_\Pi| \leqslant 2^{\mathrm{ex}(n, K_{r+1}) - \frac{n^2}{16r^2} + \frac{4n^2}{r^2\log n} + n + 2n\log n} \leqslant 2^{\mathrm{ex}(n, K_{r+1}) - n},$$

provided that n is sufficiently large. $\square$

Let $\mathrm{Col}_r(n)$ denote the family of all r-partite subgraphs of K_n. Even though some graphs in $\mathrm{Col}_r(n)$ admit many different proper r-colourings, our next lemma, which is implicit in the work of Prömel and Steger [37], shows that the average number of proper r-colourings of a graph in $\mathrm{Col}_r(n)$ is only slightly larger than one. Our proof of the lemma is an adaptation of the argument underlying the proof of [8, Proposition 5.5].

Lemma 6.6. *The following holds for all sufficiently large n and all r satisfying $2 \leqslant r \leqslant$ $\log n$. Denoting by $\mathcal{P}$ the family of all r-partitions of $[\![n]\!]$, we have*

$$\sum_{\Pi \in \mathcal{P}} 2^{e(K_\Pi)} \leqslant \left(1 + 2^{-\frac{n}{5r^4}}\right) \cdot |\mathrm{Col}_r(n)|.$$

Proof. Denote by $\mathcal{P}_b$ the family of all balanced r-partitions of $[\![n]\!]$ and let $\mathcal{P}_u = \mathcal{P} \setminus \mathcal{P}_b$. Since $|\mathcal{P}_u| \leqslant r^n \leqslant 2^{2n\log n}$, Fact 6.4 gives

$$\sum_{\Pi \in \mathcal{P}_u} 2^{e(K_\Pi)} \leqslant 2^{\mathrm{ex}(n, K_{r+1}) - \frac{n^2}{16r^2} + n + 2n\log n} \leqslant 2^{\mathrm{ex}(n, K_{r+1}) - n},$$

provided that n is sufficiently large. (This bound is also a consequence of Lemma 6.5.) As $\mathrm{ex}(n, K_{r+1})$ is the (largest) number of edges of an n-vertex r-colourable graph, we have

$$\sum_{\Pi \in \mathcal{P}_u} 2^{e(K_\Pi)} \leqslant 2^{-n} \cdot |\mathrm{Col}_r(n)|.$$

Since, for every pair $\Pi, \Pi' \in \mathcal{P}$, there are exactly $2^{e(K_\Pi \cap K_{\Pi'})}$ subgraphs of K_n that are properly r-coloured by both Π and Π', Bonferroni's inequality (the inclusion-exclusion principle) gives

$$|\mathrm{Col}_r(n)| \geqslant \sum_{\Pi \in \mathcal{P}_b} 2^{e(K_\Pi)} - \sum_{\{\Pi, \Pi'\} \in \binom{\mathcal{P}_b}{2}} 2^{e(K_\Pi \cap K_{\Pi'})}.$$

The claimed inequality will thus follow once we establish the following claim.

Claim 6.7. *For every $\Pi \in \mathcal{P}_b$,*

$$\sum_{\Pi' \in \mathcal{P}_b \setminus \{\Pi\}} 2^{e(K_\Pi \cap K_{\Pi'})} \leqslant 2^{e(K_\Pi) - \frac{n}{4r^4}}.$$

Fix distinct $\Pi, \Pi' \in \mathcal{P}_b$. Suppose that $\Pi = \{P_1, \ldots, P_r\}$ and $\Pi' = \{P'_1, \ldots, P'_r\}$ and, for all $i, j \in [\![r]\!]$, let $P_{i,j} = P_i \cap P'_j$. We will say that the vertices in $P_{i,j}$ are *moved* from P_i to P'_j. For every $i \in [\![r]\!]$, define L_i and S_i as the largest and the second largest subclasses of P_i, respectively (with ties broken arbitrarily). Note that $|P_i| \geqslant \frac{n}{2r}$ implies that $|L_i| \geqslant \frac{n}{2r^2}$. Set $s = \max_{j \in [\![r]\!]} |S_j|$ and let $S = S_j$ for the smallest j for which the maximum in the definition of s is achieved. Note that $1 \leqslant s \leqslant n/2$, as $s = 0$ would imply that $(P'_1, \ldots, P'_r)$ is a permutation of $(P_1, \ldots, P_r)$, and therefore $\Pi = \Pi'$.

By the pigeonhole principle, either some pair $\{L_i, L_j\}$ of largest subclasses or some largest subclass L_i and S, where $S \not\subseteq P_i$, are moved to the same vertex class P'_k. Since P'_k is an independent set in $K_{\Pi'}$, it follows that $K_\Pi \cap K_{\Pi'}$ has no edges between the sets L_i and L_j or L_i and S. Since,

$$\min\{|L_i| \cdot |L_j|, |L_i| \cdot |S|\} \geqslant \min\left\{\left(\frac{n}{2r^2}\right)^2, \frac{n}{2r^2} \cdot s\right\} \geqslant \frac{sn}{2r^4},$$

we have $e(K_\Pi \cap K_{\Pi'}) \leqslant e(K_\Pi) - \frac{sn}{2r^4}$.

Observe that, given a $\Pi \in \mathcal{P}$, we can describe any $\Pi' \in \mathcal{P} \setminus \{\Pi\}$ by first picking the (ordered) partitions $(P_{i,j})_{j \in [\![r]\!]}$ for every i and then setting $P'_j = \bigcup_{i \in [\![r]\!]} P_{i,j}$. We claim that, for every s, the number of ways to choose all $P_{i,j}$ in such a way that $\max_{i \in [\![r]\!]} |S_i| = s$ is at most $n^{r^2} \cdot n^{sr^2}$. Indeed, one may first specify the sequence $(|P_{i,j}|)_{i,j \in [\![r]\!]}$ and then specify, for each $i \in [\![r]\!]$, the elements of each $P_{i,j}$ with $j \in [\![r]\!]$, apart from L_i (which will comprise all the remaining, unspecified elements of P_i).

We may thus conclude that

$$\sum_{\Pi' \in \mathcal{P}_b \setminus \{\Pi\}} 2^{e(K_\Pi \cap K_{\Pi'})} \leqslant \sum_{s \geqslant 1} n^{(s+1)r^2} \cdot 2^{e(K_\Pi) - \frac{sn}{2r^4}}$$

$$\leqslant 2^{e(K_\Pi)} \cdot \sum_{s \geqslant 1} 2^{4sr^2 \log n - \frac{sn}{2r^4}} \leqslant 2^{e(K_\Pi) - \frac{n}{4r^4}},$$

as claimed. $\qquad \square$

6.5. The number of K_{r+1}-free graphs with a monochromatic star.

The following lemma will be used to bound from above the number of K_{r+1}-free graphs whose optimal r-partition induces a monochromatic copy of $K_{1,D}$ in one of the parts.

Lemma 6.8. *Let D be an integer satisfying $D \geqslant 2^r r$. Suppose that Π is an r-partition of $[\![n]\!]$ and that S is a copy of $K_{1,D}$ with $V(K_{1,D}) \subseteq P$ for some $P \in \Pi$. If $v \in P$ is the centre vertex of S, then*

$$\left| \left\{ G \subseteq K_\Pi : G \cup S \not\supseteq K_{r+1} \text{ and } \deg_G(v, Q) \geqslant D \text{ for all } Q \in \Pi \setminus \{P\} \right\} \right| \leqslant 2^{e(K_\Pi) - \frac{D^2}{8r^2}}.$$

Proof. Let G be a uniformly chosen random subgraph of K_Π. Expose G on all the edges of K_Π that have an endpoint in P and condition on $\deg_G(v, Q) \geqslant D$ for all $Q \in \Pi \setminus \{P\}$. It suffices to show that, for every such conditioning, $\Pr(G \cup S \not\supseteq K_{r+1}) \leqslant 2^{-\frac{D^2}{8r^2}}$. For each $Q \in \Pi \setminus \{P\}$, choose an arbitrary set of D neighbours of v in Q and let $\mathcal{K}$ be the family of all D^r copies of K_r in K_Π whose vertices belong to the chosen D-element sets or to $V(K_{1,D}) \setminus \{v\} \subseteq P$. Since, under our conditioning, v is adjacent (in $G \cup S$) to all the vertices of each $K \in \mathcal{K}$, we have $\Pr(G \cup S \not\supseteq K_{r+1}) \leqslant \Pr(K \not\subseteq G \text{ for all } K \in \mathcal{K})$. We may bound the latter probability from above using Janson's inequality. Define, as in the statement of Theorem 5.2,

$$\mu = \sum_{K \in \mathcal{K}} \left(\frac{1}{2} \right)^{e(K)} \qquad \text{and} \qquad \Delta = \sum_{\substack{\{K, K'\} \in \binom{\mathcal{K}}{2} \\ E(K) \cap E(K') \neq \emptyset}} \left(\frac{1}{2} \right)^{e(K \cup K')}.$$

Observe that the function $\{2, \dots, r\} \ni \ell \mapsto D^\ell \cdot 2^{-\binom{\ell}{2}}$ is increasing. Indeed, our assumption implies that, for each $\ell \leqslant r - 1$, we have $D \cdot 2^{\binom{\ell}{2} - \binom{\ell+1}{2}} \geqslant D \cdot 2^{-r} \geqslant 1$. It follows that

$$\frac{\mu}{2} = \frac{D^r}{2} \cdot \left(\frac{1}{2} \right)^{\binom{r}{2}} \geqslant \frac{D^2}{4}.$$

We now estimate Δ. To this end, fix some $K \in \mathcal{K}$ and observe that, for each $k \in \{2, \dots, r-1\}$, there are at most $\binom{r}{k} \cdot D^{r-k}$ many $K' \in \mathcal{K}$ that share exactly k vertices with K and that $e(K \cup K') = 2\binom{r}{2} - \binom{k}{2}$ for each such K'. We conclude that

$$2\Delta \leqslant \mu \cdot \sum_{k=2}^{r-1} \binom{r}{k} \cdot D^{r-k} \cdot \left(\frac{1}{2} \right)^{\binom{r}{2} - \binom{k}{2}} \leqslant \mu^2 \cdot \sum_{k=2}^{r-1} \frac{2^{\binom{k}{2}} r^k}{D^k}.$$

Our assumption implies that, for every $k \in \{2, \dots, r-2\}$,

$$\frac{2^{\binom{k+1}{2}} r^{k+1}}{D^{k+1}} = \frac{2^k r}{D} \cdot \frac{2^{\binom{k}{2}} r^k}{D^k} \leqslant \frac{1}{2} \cdot \frac{2^{\binom{k}{2}} r^k}{D^k}$$

and hence,

$$\frac{2\Delta}{\mu^2} \leqslant 2 \cdot \frac{2r^2}{D^2}.$$

We conclude that

$$\Pr(G \cup S \nsupseteq K_{r+1}) \leqslant \exp\left(-\min\left\{\frac{\mu}{2}, \frac{\mu^2}{4\Delta}\right\}\right) \leqslant \exp\left(-\frac{D^2}{8r^2}\right),$$

as claimed. $\qquad\square$

6.6. **The number of K_{r+1}-free graphs with a monochromatic matching.** The following lemma will be used to bound from above the number of K_{r+1}-free graphs whose optimal r-partition induces a monochromatic matching with a given number of edges in one of the parts.

Lemma 6.9. *Suppose that Π is a balanced r-partition of $[\![n]\!]$ and that M is a matching with m edges such that $V(M) \subseteq P$ for some $P \in \Pi$. If $r^2 \cdot 2^{r+3} \leqslant n$, then*

$$\left|\left\{G \subseteq K_\Pi : G \cup M \nsupseteq K_{r+1}\right\}\right| \leqslant 2^{e(K_\Pi) - \frac{mn}{2^{10}r^4}}.$$

Proof. Let G be a uniformly chosen random subgraph of K_Π, so that the assertion of the lemma becomes equivalent to the inequality $\Pr(G \cup M \nsupseteq K_{r+1}) \leqslant 2^{-\frac{mn}{2^{10}r^4}}$. Let $N = \prod_{Q \in \Pi \setminus \{P\}} |Q|$ and note that the assumption that Π is balanced implies that $N \geqslant \left(\frac{n}{2r}\right)^{r-1}$. Denote by K_{r+1}^- the graph obtained from K_{r+1} by removing from it a single edge and let $\mathcal{K}$ be the collection of all copies of K_{r+1}^- in K_Π that form a K_{r+1} with an edge of M. Note that $|\mathcal{K}| = mN$ and that $\Pr(G \cup M \nsupseteq K_{r+1}) = \Pr(K \nsubseteq G$ for all $K \in \mathcal{K})$. We may thus bound this probability from above using Janson's inequality. Define, as in the statement of Theorem 5.2,

$$\mu = \sum_{K \in \mathcal{K}} \left(\frac{1}{2}\right)^{e(K)} \qquad \text{and} \qquad \Delta = \sum_{\substack{\{K,K'\} \in \binom{\mathcal{K}}{2} \\ E(K) \cap E(K') \neq \emptyset}} \left(\frac{1}{2}\right)^{e(K \cup K')}.$$

Observe that the function $\{2, \ldots, r\} \ni \ell \mapsto \left(\frac{n}{2r}\right)^{\ell-1} \cdot 2^{-\binom{\ell+1}{2}}$ is increasing. Indeed, our assumption on r implies that, for each $\ell \leqslant r - 1$, we have $\frac{n}{2r} \cdot 2^{\binom{\ell}{2} - \binom{\ell+1}{2}} \geqslant \frac{n}{2r} \cdot 2^{-r} \geqslant 1$. It follows that

$$\frac{\mu}{2} = \frac{mN}{2} \cdot \left(\frac{1}{2}\right)^{\binom{r+1}{2}-1} \geqslant m \cdot \left(\frac{n}{2r}\right)^{r-1} \cdot \left(\frac{1}{2}\right)^{\binom{r+1}{2}} \geqslant \frac{mn}{16r}.$$

We now estimate Δ. To this end, fix some $K \in \mathcal{K}$ and observe that:

(a) For each $k \in \{1, \ldots, r-2\}$, there are at most $\binom{r-1}{k} \cdot N \cdot \left(\frac{n}{2r}\right)^{-k}$ many $K' \in \mathcal{K}$ that share with K the two vertices in P and exactly k other vertices; we have $e(K \cup K') = 2\binom{r+1}{2} - \binom{k+2}{2} - 1$ for each such K'.

(b) For each $k \in \{2, \ldots, r-1\}$, there are at most $m \cdot \binom{r-1}{k} \cdot N \cdot \left(\frac{n}{2r}\right)^{-k}$ many $K' \in \mathcal{K}$ that share with K only some k vertices outside of P; we have $e(K \cup K') = 2\binom{r+1}{2} - \binom{k}{2} - 2$ for each such K'.

We conclude that

$$\frac{2\Delta}{\mu^2} \leqslant \frac{1}{m} \cdot \sum_{k=1}^{r-2} \binom{r-1}{k} \cdot \left(\frac{n}{2r}\right)^{-k} \cdot 2^{\binom{k+2}{2}-1} + \sum_{k=2}^{r-1} \binom{r-1}{k} \cdot \left(\frac{n}{2r}\right)^{-k} \cdot 2^{\binom{k}{2}}$$

$$\leqslant \frac{1}{m} \cdot \sum_{k=1}^{r-2} \frac{(2r^2)^k \cdot 2^{\binom{k+2}{2}-1}}{n^k} + \sum_{k=2}^{r-1} \frac{(2r^2)^k \cdot 2^{\binom{k}{2}}}{n^k}$$

$$\leqslant \frac{8r^2}{mn} + \sum_{k=2}^{r} \frac{(2r^2)^k \cdot 2^{\binom{k+2}{2}}}{n^k}.$$

Our assumption on r implies that, for every $k \in \{2, \ldots, r-1\}$,

$$\frac{(2r^2)^{k+1} 2^{\binom{k+3}{2}}}{n^{k+1}} = \frac{r^2 2^{k+3}}{n} \cdot \frac{(2r^2)^k 2^{\binom{k+2}{2}}}{n^k} \leqslant \frac{1}{2} \cdot \frac{(2r^2)^k 2^{\binom{k+2}{2}}}{n^k}$$

and hence, as $m \leqslant n/2$ and $r \geqslant 2$,

$$\frac{2\Delta}{\mu^2} \leqslant \frac{8r^2}{mn} + 2 \cdot \frac{4r^4 \cdot 2^6}{n^2} \leqslant \frac{258r^4}{mn}.$$

We conclude that

$$\Pr(G \cup M \nsubseteq K_{r+1}) \leqslant \exp\left(-\min\left\{\frac{\mu}{2}, \frac{\mu^2}{4\Delta}\right\}\right) \leqslant \exp\left(-\frac{mn}{2^{10}r^4}\right),$$

as claimed. □

6.7. **Proof of Theorem 1.2.** Suppose that positive integers n and r satisfy $2 \leqslant r \leqslant \log n/(121 \log \log n)$ and let $\mathrm{Col}_r(n)$ and $\mathcal{F}$ denote the families of all r-partite and all K_{r+1}-free subgraphs of K_n, respectively. Since $\mathrm{Col}_r(n) \subseteq \mathcal{F}$, it suffices to show that

$$|\mathcal{F} \setminus \mathrm{Col}_r(n)| \leqslant 2^{-\frac{n}{(10r)^4}} \cdot |\mathrm{Col}_r(n)|. \tag{32}$$

Let $\delta = (8 \log n)^{-15}$ and let $\mathcal{P}$ be the family of all r-partitions of $[\![n]\!]$. Define, for every graph $G \in \mathcal{F}$,

$$t(G) = \min\{e(G \setminus K_\Pi) : \Pi \in \mathcal{P}\}$$

and let

$$\mathcal{F}_{\mathrm{close}} = \{G \in \mathcal{F} : 1 \leqslant t(G) \leqslant \delta n^2\} \qquad \text{and} \qquad \mathcal{F}_{\mathrm{far}} = \{G \in \mathcal{F} : t(G) > \delta n^2\},$$

so that $\mathcal{F}_{\mathrm{close}} \cup \mathcal{F}_{\mathrm{far}} = \mathcal{F} \setminus \mathrm{Col}_r(n)$. Furthermore, for every $G \in \mathcal{F}_{\mathrm{close}}$, let $\Pi(G)$ be an arbitrary r-partition that achieves the minimum in the definition of $t(G)$. Let $\mathcal{F}^b_{\mathrm{close}}$ comprise these G in $\mathcal{F}_{\mathrm{close}}$ for which $\Pi(G)$ is a balanced partition and let $\mathcal{F}^u_{\mathrm{close}} = \mathcal{F}_{\mathrm{close}} \setminus \mathcal{F}^b_{\mathrm{close}}$. Finally, for every balanced partition $\Pi \in \mathcal{P}$ and every integer t satisfying $1 \leqslant t \leqslant \delta n^2$, define

$$\mathcal{F}_{t,\Pi} = \{G \in \mathcal{F}^b_{\mathrm{close}} : t(G) = t \text{ and } \Pi(G) = \Pi\}.$$

Letting $\mathcal{P}_b$ denote the set of balanced r-partitions of $[\![n]\!]$, we thus have

$$|\mathcal{F} \setminus \mathrm{Col}_r(n)| \leqslant |\mathcal{F}_{\mathrm{far}}| + |\mathcal{F}^u_{\mathrm{close}}| + \sum_{\Pi \in \mathcal{P}_b} \sum_{t=1}^{\delta n^2} |\mathcal{F}_{t,\Pi}|. \tag{33}$$

It follows from Theorem 6.2 and Lemma 6.5 that the first and the second terms in the right-hand side of (33) are at most $2^{\mathrm{ex}(n,K_{r+1})-n}$ each. To bound the final term, we shall derive the following estimate.

Claim 6.10. *For every integer t satisfying $1 \leqslant t \leqslant \delta n^2$ and each $\Pi \in \mathcal{P}_b$,*

$$|\mathcal{F}_{t,\Pi}| \leqslant 2^{e(K_\Pi) - \frac{n}{(8r)^4} - t}.$$

Let us first argue that inequality (33) and Claim 6.10 imply (32). Indeed, assuming Claim 6.10, we have

$$|\mathcal{F} \setminus \mathrm{Col}_r(n)| \leqslant 2 \cdot 2^{\mathrm{ex}(n,K_{r+1})-n} + \sum_{\Pi \in \mathcal{P}_b} \sum_{t=1}^{\delta n^2} 2^{e(K_\Pi) - \frac{n}{(8r)^4} - t}$$

$$\leqslant 2^{\mathrm{ex}(n,K_{r+1})-n+1} + 2^{-\frac{n}{(8r)^r}} \cdot \sum_{\Pi \in \mathcal{P}} 2^{e(K_\Pi)}.$$

Finally, since $|\mathrm{Col}_r(n)| \geqslant 2^{\mathrm{ex}(n,K_{r+1})}$ and $\sum_{\Pi \in \mathcal{P}} 2^{e(K_\Pi)} \leqslant 2|\mathrm{Col}_r(n)|$, by Lemma 6.6, we conclude that

$$|\mathcal{F} \setminus \mathrm{Col}_r(n)| \leqslant \left(2^{-n+1} + 2^{-\frac{n}{(8r)^4}+1} \right) \cdot |\mathrm{Col}_r(n)|,$$

which yields (32). It thus suffices to prove Claim 6.10.

Proof of Claim 6.10. Let $D = \left\lfloor \frac{n}{2^{14} r^5 \log n} \right\rfloor$ and define

$$\mathcal{F}_{t,\Pi}^S = \left\{ G \in \mathcal{F}_{t,\Pi} : G[P] \supseteq K_{1,D} \text{ for some } P \in \Pi \right\},$$
$$\mathcal{F}_{t,\Pi}^M = \left\{ G \in \mathcal{F}_{t,\Pi} : G[P] \text{ has a matching of size } \lceil t/(Dr) \rceil \text{ for some } P \in \Pi \right\}.$$

Since every graph with t edges contains either a vertex with degree at least D or a matching with at least t/D edges, we have $\mathcal{F}_{t,\Pi} = \mathcal{F}_{t,\Pi}^S \cup \mathcal{F}_{t,\Pi}^M$ and we may bound $|\mathcal{F}_{t,\Pi}|$ from above in two steps.

First, we claim that if $G \in \mathcal{F}_{t,\Pi}^S$ and $v \in P \in \Pi$ is the centre vertex of a copy of $K_{1,D}$ in $G[P]$, then $\deg_G(v, Q) \geqslant D$ for all $Q \in \Pi$. Indeed, if this were not true, then moving v from P to Q would yield a partition Π' such that

$$e(G \setminus K_{\Pi'}) = e(G \setminus K_\Pi) + \deg_G(v, Q) - \deg_G(v, P) < e(G \setminus K_\Pi),$$

which would contradict our assumption that $\Pi = \Pi(G)$. It thus follows from Lemma 6.8 (which we may apply as $D \geqslant n^{1/2} \geqslant 2^r r$ when n is sufficiently large) that

$$|\mathcal{F}_{t,\Pi}^S| \leqslant \left(n^2 \right)^t \cdot 2^{e(K_\Pi) - \frac{D^2}{8r^2}} \leqslant 2^{e(K_\Pi) - \frac{D^2}{8r^2} + 4t \log n} \leqslant 2^{e(K_\Pi)-n-t},$$

where the last inequality holds because, by our choice of δ and D,

$$\frac{D^2}{8r^2} \geqslant \frac{n^2}{2^{32} r^{12} (\log n)^2} \geqslant \frac{n^2}{(8 \log n)^{14}} \geqslant 8 \delta n^2 \log n \geqslant 4t \log n + n + t.$$

Second, it follows from Lemma 6.9 that

$$|\mathcal{F}_{t,\Pi}^M| \leqslant \left(n^2 \right)^t \cdot 2^{e(K_\Pi) - \frac{\lceil t/(Dr) \rceil n}{2^{10} r^4}} \leqslant 2^{e(K_\Pi) - \frac{\lceil t/(Dr) \rceil n}{2^{10} r^4} + 4t \log n} \leqslant 2^{e(K_\Pi) - \frac{n}{2^{11} r^4} - t},$$

where the last inequality holds because

$$\frac{\lceil t/(Dr) \rceil n}{2^{10} r^4} - 4t \log n - t \geqslant \min_{\tau \in \{1,2,\dots\}} \left\{ \frac{\tau n}{2^{10} r^4} - \tau Dr \cdot (4 \log n + 1) \right\} \geqslant \frac{n}{2^{11} r^4},$$

as the definition of D assures that $n \geqslant 2^{14} D r^5 \log n$. Since $|\mathcal{F}_{t,\Pi}| \leqslant |\mathcal{F}_{t,\Pi}^S| + |\mathcal{F}_{t,\Pi}^M|$, combining the two bounds above gives the assertion of the claim. $\qquad\square$

7. Lower bounds for ε-nets

7.1. **Outline.** Our (randomised) construction of planar point sets X without a small ε-net for the range space of lines on X is a slight simplification of the construction of Balogh and Solymosi [9]. The high-level idea of both constructions, which can be traced back to the work of Alon [1], may be summarised as follows. We find an integer s, a finite set $X \subseteq \mathbb{R}^2$, and a sub-collection $\mathcal{L}$ of all lines in $\mathbb{R}^2$ with the following property: Let $\mathcal{H}$ be the s-uniform hypergraph with vertex set X whose edges are all intersections of the lines in $\mathcal{L}$ with X that have exactly s points. The independence number of $\mathcal{H}$ is at most $(1-c)|X|$, for some constant $c > 0$.

Given such s, X, and $\mathcal{L}$, we set $\varepsilon = s/|X|$ and observe that the complement of every ε-net N for the range space of lines on X is an independent set of $\mathcal{H}$. Indeed, every such N intersects every line that contains at least s points of X; in particular, N must intersect every line in $\mathcal{L}$ that contains exactly s points of X. This means that $|N| \geqslant c|X| = cs/\varepsilon$, which improves upon the trivial bound $|N| \geqslant \Omega(1/\varepsilon)$ if s can be made arbitrarily large. The challenge is to make s as large as possible, as a function of $|X|$.

In the construction of Alon [1], the set X is a generic projection of the d-dimensional grid $[\![s]\!]^d$ to $\mathbb{R}^2$ and $\mathcal{L}$ is the image of all combinatorial lines in $[\![s]\!]^d$ via this projection; the key property is guaranteed, for large enough d, by the density version of the Hales–Jewett theorem proved by Furstenberg and Katznelson [24]. In Balogh and Solymosi's [9] construction, X was a generic projection of a random subset of a larger, high-dimensional integer grid, trimmed appropriately (so that each line in $\mathcal{L}$ contains no more than s points of X), and the key property of $\mathcal{H}$ was established, for a careful choice of $\mathcal{L}$, with the use of the hypergraph container theorem of Saxton and Thomason [42]. Here, we take X to be a random subset of $[\![n]\!]^2$, trimmed appropriately (as in [9]), and establish the key property of $\mathcal{H}$, for a careful choice of $\mathcal{L}$, using our efficient container lemma, Theorem 1.6.

7.2. **Proof of Theorem 1.3.** Let s be a positive integer, let $m = 10s$, and let M be a prime number satisfying $m^{m^2-1} \leqslant M \leqslant 2m^{m^2-1}$. Set $n = mM$, so that

$$n \geqslant m^{m^2} \qquad \text{and} \qquad m \geqslant \sqrt{\frac{\log n}{\log \log n}}. \tag{34}$$

We shall find an $\varepsilon \in (0, 1/n)$ and a set $X \subseteq \mathbb{R}^2$ without a small ε-net among the subsets of the integer grid $[\![n]\!]^2$, which we shall from now on denote by P. We will be able to prove the claimed lower bound on the smallest size of an ε-net of X for the range space of all lines in $\mathbb{R}^2$ by considering only a fairly small family $\mathcal{L}$ of lines that we now specify.

Given an integer $h \in [\![M-1]\!]$ and a point $(x_0, y_0) \in [\![n]\!] \times [\![M]\!]$, we let $\ell(x_0, y_0; h)$ be the line passing through (x_0, y_0) whose slope is M/h, that is,

$$\ell(x_0, y_0; h) = \big\{(x_0, y_0) + t \cdot (h, M) : t \in \mathbb{R}\big\}.$$

Since M is prime, and thus co-prime with h, the vector $t \cdot (h, M)$ has integer coordinates if and only if $t \in \mathbb{Z}$. Moreover, if t is an integer, then $y_0 + tM \in [\![n]\!]$ if and only if $t \in \{0, \ldots, m-1\}$. In particular, $\ell(x_0, y_0; h)$ intersects P in at most m points; it intersects P in exactly m points if and only if $x_0 + (m-1)h \leqslant n$. Now, for every $h \in [\![M-1]\!]$, let

$$\mathcal{L}_h = \big\{\ell(x_0, y_0; h) : (x_0, y_0) \in [\![n]\!] \times [\![M]\!] \text{ and } x_0 + (m-1)h \leqslant n\big\},$$

so that every line in $\mathcal{L}_h$ intersects P in exactly m points. Since the lines in $\mathcal{L}_h$ are pairwise disjoint (as they are parallel), we have

$$\left| \left(\bigcup \mathcal{L}_h \right) \cap P \right| = m \cdot |\mathcal{L}_h| = m \cdot \left(n - (m-1)h \right) \cdot M = n^2 \cdot \left(1 - \frac{(m-1)h}{n} \right).$$

Let $h_{\max} = \lfloor n/(10m) \rfloor$ so that $\bigcup \mathcal{L}_h$ has at least $9n^2/10$ points of P for every $h \in [\![h_{\max}]\!]$. Finally, define

$$\mathcal{L} = \bigcup_{h=1}^{h_{\max}} \mathcal{L}_h$$

and note that

$$\frac{n^3}{12m^2} \leqslant \frac{9n^2}{10m} \cdot h_{\max} \leqslant |\mathcal{L} \cap P| \leqslant \frac{n^2}{m} \cdot h_{\max} \leqslant \frac{n^3}{10m^2}. \tag{35}$$

We shall say that a set $A \subseteq P$ is $\mathcal{L}$-*collinear* if A is contained in some line in $\mathcal{L}$. As every line in $\mathcal{L}$ contains exactly m points of P, the number of a-element $\mathcal{L}$-collinear subsets of P is precisely $|\mathcal{L}| \cdot \binom{m}{a}$ for every $a \in \{2, \ldots, m\}$.

Suppose that p satisfies

$$K \cdot m^{10} \cdot n^{-1/(s-1)} \cdot \log n \leqslant p \leqslant m^{-1} \cdot n^{-1/s} \tag{36}$$

for some large absolute constant K; such a number does indeed exist as

$$n^{1/(s-1)-1/s} \geqslant n^{1/s^2} \geqslant m^{m^2/s^2} = m^{100} \geqslant K \cdot m^{10} \cdot \log n,$$

provided that m is sufficiently large. Let R be a p-random subset of P and let $X \subseteq R$ be a largest subset of R that contains no $\mathcal{L}$-collinear subset of $s+1$ points. By maximality of X, every point of $R \setminus X$ forms an $\mathcal{L}$-collinear $(s+1)$-element set with some s points of X. In particular, $|R \setminus X|$ is at most the number of $\mathcal{L}$-collinear $(s+1)$-element subsets of R. It follows that

$$\mathbb{E}\big[|R \setminus X|\big] \leqslant |\mathcal{L}| \cdot \binom{m}{s+1} \cdot p^{s+1} \leqslant \frac{n^3}{10m^2} \cdot m^{s+1} \cdot p^{s+1} \leqslant \frac{n^2 p}{10m},$$

by the second inequality in (36), and consequently, by Markov's inequality,

$$\Pr\left(|R \setminus X| \geqslant n^2 p/10 \right) \leqslant \frac{\mathbb{E}\big[|R \setminus X|\big]}{n^2 p/10} \leqslant \frac{1}{m} \leqslant \frac{1}{10}.$$

On the other hand, standard estimates for lower tails of binomial distributions (Lemma 5.1) yield

$$\Pr\left(|R| \leqslant 9n^2 p/10 \right) \leqslant \exp\left(-n^2 p/200 \right) \leqslant \exp(-n/200).$$

It follows that

$$\Pr\left(|X| \geqslant 4n^2 p/5 \right) \geqslant 4/5,$$

provided that n is sufficiently large.

Claim 7.1. *With probability at least $1/2$, every set $I \subseteq R$ with $|I| \geqslant 3n^2 p/5$ contains an $\mathcal{L}$-collinear subset of s points.*

Together with the above calculations, Claim 7.1 implies that there exists a set $X \subseteq P$ of at least $4n^2 p/5$ points that has the following two properties:

(a) X has no $\mathcal{L}$-collinear subset with $s+1$ elements;
(b) every set of $3n^2 p/5$ elements of X contains an $\mathcal{L}$-collinear s-element subset.

Suppose that X is such a set, let $\varepsilon = s/|X|$, and assume that $N \subseteq X$ is an ε-net for the range space of lines. In particular, N intersects every $\mathcal{L}$-collinear subset of X that has at least $s = \varepsilon|X|$ elements. Since X contains no $\mathcal{L}$-collinear set with more than s points, $X \setminus N$ contains no $\mathcal{L}$-collinear subset of s points, and thus

$$|N| = |X| - |X \setminus N| > |X| - 3n^2 p/5 \geqslant |X|/4 = s/(4\varepsilon).$$

Finally, since $1/\varepsilon = |X|/s \leqslant |P| = n^2$, we have, using (34),

$$\frac{s}{4} = \frac{m}{40} \geqslant \frac{1}{40} \cdot \sqrt{\frac{\log n}{\log \log n}} \geqslant \frac{1}{80} \cdot \sqrt{\frac{\log(1/\varepsilon)}{\log \log(1/\varepsilon)}}.$$

This gives the assertion of the theorem.

We now prove Claim 7.1. Let $\mathcal{H}$ be the s-uniform hypergraph with vertex set P whose edges are all $\mathcal{L}$-collinear s-element subsets of P. The assertion of the claim is that, with probability at least $1/2$, the random set R contains no independent set of $\mathcal{H}$ that has at least $3n^2 p/5$ elements. This is a simple consequence of the following lemma, which lies at the heart of the matter.

Lemma 7.2. *There is a family $\mathcal{C}$ of at most $\exp(pn^2/300)$ containers for the independent sets of $\mathcal{H}$ such that $|C| \leqslant n^2/2$ for every $C \in \mathcal{C}$.*

We first show how Lemma 7.2 implies the assertion of Claim 7.1. Let $\mathcal{C}$ be a family of containers for the independent sets of $\mathcal{H}$ supplied by the lemma and let $\mathcal{B}$ be the event that R contains an independent set of $\mathcal{H}$ with at least $3n^2 p/5$ elements. Since every independent set of $\mathcal{H}$ is contained in some member of $\mathcal{C}$, each of which has at most $n^2/2$ elements, we have

$$\Pr(\mathcal{B}) \leqslant \sum_{C \in \mathcal{C}} \Pr\left(|R \cap C| \geqslant 3n^2 p/5\right) \leqslant |\mathcal{C}| \cdot \Pr\left(\text{Bin}(n^2/2, p) \geqslant 3n^2 p/5\right).$$

Standard estimates for upper tails of binomial distributions (Lemma 5.1) yield

$$\Pr\left(\text{Bin}(n^2/2, p) \geqslant 3n^2 p/5\right) \leqslant \exp\left(-n^2 p/150\right)$$

and, consequently,

$$\Pr(\mathcal{B}) \leqslant |\mathcal{C}| \cdot \exp\left(-n^2 p/150\right) \leqslant \exp\left(-n^2 p/300\right) \leqslant \exp(-n/300) \leqslant 1/2,$$

provided that n is sufficiently large.

Finally, we prove Lemma 7.2 by combining our 'packaged' hypergraph container lemma, Theorem 1.6, with the following supersaturation statement for the hypergraph $\mathcal{H}$ of $\mathcal{L}$-collinear s-tuples.

Lemma 7.3. *If $Q \subseteq P$ has at least $n^2/3$ points, then $e(\mathcal{H}[Q]) \geqslant |\mathcal{L}|$.*

Proof. Define $b_s \colon \mathbb{R} \to \mathbb{R}$ by

$$b_s(x) = \begin{cases} \binom{x}{s} & \text{if } x \geqslant s - 1, \\ 0 & \text{if } x \leqslant s - 1, \end{cases}$$

so that b_s is convex and $b_s(a) = \binom{a}{s}$ whenever a is a nonnegative integer. Jensen's inequality gives

$$e(\mathcal{H}[Q]) = \sum_{\ell \in \mathcal{L}} \binom{|\ell \cap Q|}{s} = \sum_{\ell \in \mathcal{L}} b_s(|\ell \cap Q|) \geqslant |\mathcal{L}| \cdot b_s\left(\frac{1}{|\mathcal{L}|} \cdot \sum_{\ell \in \mathcal{L}} |\ell \cap Q|\right).$$

Recall that, for every $h \in \llbracket h_{\max} \rrbracket$, the lines in $\mathcal{L}_h$ cover all but at most $n^2/10$ points of P. In particular, for every such h,

$$\sum_{\ell \in \mathcal{L}_h} |\ell \cap Q| \geqslant |Q| - \frac{n^2}{10} \geqslant \frac{n^2}{5}$$

and thus

$$\frac{1}{|\mathcal{L}|} \cdot \sum_{\ell \in \mathcal{L}} |\ell \cap Q| = \frac{1}{|\mathcal{L}|} \cdot \sum_{h=1}^{h_{\max}} \sum_{\ell \in \mathcal{L}_h} |\ell \cap Q| \geqslant \frac{h_{\max}}{|\mathcal{L}|} \cdot \frac{n^2}{5} \geqslant \frac{m}{5}.$$

Consequently,

$$e\big(\mathcal{H}[Q]\big) \geqslant |\mathcal{L}| \cdot b_s\big(m/5\big) = |\mathcal{L}| \cdot \binom{m/5}{s} \geqslant |\mathcal{L}|,$$

as claimed. $\qquad\square$

Proof of Lemma 7.2. Set

$$q = \frac{p}{300m^5 \log n} \qquad \text{and} \qquad E = |\mathcal{L}|.$$

We now verify that we may apply Theorem 1.6, with $\alpha \leftarrow 1/2$ and $\beta \leftarrow 1/3$, to the hypergraph $\mathcal{H}$. First, we have

$$\alpha\beta q \cdot v(\mathcal{H}) = \frac{pn^2}{1800m^5 \log n} \geqslant 10^9 s^7 \qquad \text{and} \qquad 10^4 s^5 q = \frac{10^4 s^5 p}{300m^5 \log n} \leqslant \frac{1}{\log n} \leqslant \beta,$$

provided that n is sufficiently large. Second, for every $t \in \{2, \ldots, s\}$, by (35) and (36),

$$\left(\frac{q}{10^6 s^5}\right)^{t-1} \cdot \frac{E}{v(\mathcal{H})} = \left(\frac{p}{3 \cdot 10^8 s^5 m^5 \log n}\right)^{t-1} \cdot \frac{|\mathcal{L}|}{n^2} \geqslant \left(\frac{K \cdot m^5 \cdot n^{-1/(s-1)}}{3 \cdot 10^8 s^5}\right)^{s-1} \cdot \frac{n}{12m^2}$$

$$= \left(\frac{K}{3 \cdot 10^3}\right)^{s-1} \cdot \frac{1}{12m^2} \geqslant 2^m \geqslant \binom{m-t}{s-t} = \Delta_t(\mathcal{H}),$$

provided that K is sufficiently large. The theorem supplies a collection $\mathcal{C}$ of containers for the independent sets of $\mathcal{H}$ such that

$$|\mathcal{C}| \leqslant \exp\big(10^4 s^5 \beta^{-1} \log(e/\alpha) \cdot q \log(e/q) \cdot v(\mathcal{H})\big)$$

$$\leqslant \exp\big(m^5 \cdot q \log n \cdot n^2\big) \leqslant \exp\big(pn^2/300\big),$$

where we used the inequality $q \geqslant e/n$, which holds if K is sufficiently large, and, for every $C \in \mathcal{C}$, either $|C| \leqslant \alpha \cdot v(\mathcal{H}) = n^2/2$ or there is a subset $W \subseteq C$ with $|W| \geqslant (1-\beta)|C| = 2|C|/3$ such that $e(\mathcal{H}[W]) < E = |\mathcal{L}|$. We claim that, in fact, $|C| \leqslant n^2/2$ for every $C \in \mathcal{C}$. Indeed, if $|C| > n^2/2$ and $W \subseteq C$ satisfies $|W| \geqslant 2|C|/3 > n^2/3$, then $e(\mathcal{H}[W]) \geqslant |\mathcal{L}|$, by Lemma 7.3. $\qquad\square$

8. Upper bounds on Ramsey numbers

In this section, we derive the upper bounds on Folkman numbers and on induced Ramsey numbers stated in Theorems 1.4 and 1.5. We shall do this by building containers for non-Ramsey colourings of subgraphs of a large complete graph K_N and examining how a random subgraph of K_N, drawn with an appropriately chosen distribution for each of the two theorems, intersects these containers. This approach to studying Ramsey properties of random graphs was introduced in the work of Nenadov and Steger [34]. The

only Ramsey-theoretic ingredient in our proof is the following supersaturated version of Ramsey's theorem, which is a refinement of [34, Corollary 2.2].

Lemma 8.1. *Suppose that n and k are positive integers and let $R = R(n; k)$. If $N \geqslant R$, then every colouring $c \colon E(K_N) \to [\![k+1]\!]$ either assigns the colour $k+1$ to at least $(1/2) \cdot (N/R)^2$ edges or it contains at least $(1/2) \cdot (N/R)^n$ monochromatic copies of K_n in colours $1, \ldots, k$.*

Proof. The choice of R guarantees that the edge-colouring induced by every subset of R vertices of K_N contains either an edge coloured $k+1$ or a monochromatic copy of K_n in one of the remaining k colours. On the other hand, each edge and each copy of K_n are contained in, respectively, $\binom{N-2}{R-2}$ and $\binom{N-n}{R-n}$ such subsets. Denoting by M the total number of monochromatic copies of K_n in colours $1, \ldots, k$, we thus have

$$\binom{N}{R} \leqslant |c^{-1}(k+1)| \cdot \binom{N-2}{R-2} + M \cdot \binom{N-n}{R-n}. \tag{37}$$

In particular, since, for every $\ell \in \{2, n\}$,

$$\binom{N}{R} \cdot \binom{N-\ell}{R-\ell}^{-1} = \binom{N}{\ell} \cdot \binom{R}{\ell}^{-1} \geqslant \left(\frac{N}{R}\right)^{\ell}$$

inequality (37) implies that either $|c^{-1}(k+1)| \geqslant (1/2) \cdot (N/R)^2$ or $M \geqslant (1/2) \cdot (N/R)^n$. $\square$

8.1. **Folkman numbers (proof of Theorem 1.4).** Let k and n be positive integers, let $R = R(n; k)$ and suppose that an integer N satisfies

$$N \geqslant (\Gamma k n R)^{21n^2} \tag{38}$$

for some large constant Γ. We shall give a randomised construction of a K_{n+1}-free subgraph of K_N that satisfies $G \to (K_n)_k$, proving that $F(n; k) \leqslant N$. We shall from now on assume that $k \geqslant 2$ and $n \geqslant 3$, as otherwise the assertion of the theorem is trivial.

Suppose that $G \subseteq K_N$. We shall identify a k-colouring $c \colon E(G) \to [\![k]\!]$ of the edges of G with the set

$$\{(e, c_e) : e \in E(G)\} \subseteq E(K_N) \times [\![k]\!].$$

Let $\mathcal{H}$ be the hypergraph with vertex set $E(K_N) \times [\![k]\!]$ whose edges are all sets of the form

$$\varphi\big(E(K_n)\big) \times \{i\},$$

where $\varphi \colon V(K_n) \to V(K_N)$ is an arbitrary injection and $i \in [\![k]\!]$. If a graph $G \subseteq K_N$ admits a colouring $c \colon E(G) \to [\![k]\!]$ with no monochromatic copy of K_n, then c, when viewed as a subset of $E(K_N) \times [\![k]\!]$, is an independent set of $\mathcal{H}$.

We shall say that a graph $G \subseteq E(K_N)$ is *compatible* with a set $C \subseteq E(K_N) \times [\![k]\!]$ if there exists a colouring $c \colon E(G) \to [\![k]\!]$ that is contained in C. Equivalently, G is compatible with C if and only if $\big(\{e\} \times [\![k]\!]\big) \cap C \neq \emptyset$ for every $e \in E(G)$. In other words, defining

$$X(C) = \big\{e \in E(K_N) : \big(e \times [\![k]\!]\big) \cap C = \emptyset\big\},$$

G is compatible with C if and only if $X(C) \cap E(G) = \emptyset$.

Suppose that p satisfies

$$D \cdot (knR)^{20} \cdot N^{-2/(n+1)} \log N \leqslant p \leqslant \frac{N^{-2/(n+2)}}{DR} \tag{39}$$

for some large constant D; such a number does indeed exist as (38) implies that

$$\frac{N^{2/(n+1)-2/(n+2)}}{\log N} \geqslant N^{1/n^2} \geqslant D^2 \cdot (knR)^{21},$$

provided that Γ is sufficiently large. The following lemma is key.

Lemma 8.2. *There is a family $\mathcal{C}$ of at most $\exp\left(\frac{pN^2}{256R^2}\right)$ containers for the independent sets of $\mathcal{H}$ such that $|X(C)| \geqslant \left(\frac{N}{4R}\right)^2$ for every $C \in \mathcal{C}$.*

We first show how Lemma 8.2 implies the assertion of the theorem. To this end, suppose that $G \sim G_{N,p}$ and denote by Z the number of copies of K_{n+1} in G. The upper bound in (39) implies that

$$\mathbb{E}[Z] \leqslant p^{\binom{n+1}{2}} N^{n+1} = pN^2 \cdot \left(p^{(n+2)/2} N\right)^{n-1} \leqslant \frac{pN^2}{(DR)^{(n+2)(n-1)/2}} \leqslant \frac{pN^2}{64R^2},$$

provided that D is sufficiently large. Let G' be the subgraph obtained from G by deleting an arbitrary edge from every copy of K_{n+1}; observe that $K_{n+1} \not\subseteq G'$ and $e(G') \geqslant e(G) - Z$.

Suppose that $G' \not\rightarrow (K_n)_k$. This means that there is a colouring $c\colon E(G') \to [\![k]\!]$ that is an independent set of $\mathcal{H}$. Therefore, G' must be compatible with some container from $\mathcal{C}$. In other words, there is some $C \in \mathcal{C}$ such that $X(C) \cap E(G') = \emptyset$ and, consequently,

$$|X(C) \cap E(G)| \leqslant e(G) - e(G') \leqslant Z.$$

We may conclude that

$$\Pr\left(G' \not\rightarrow (K_n)_k\right) \leqslant \Pr\left(Z > 2\mathbb{E}[Z]\right) + \sum_{C \in \mathcal{C}} \Pr\left(|X(C) \cap E(G)| \leqslant 2\mathbb{E}[Z]\right). \qquad (40)$$

Fix an arbitrary $C \in \mathcal{C}$. Since $|X(C)| \geqslant \frac{N^2}{16R^2}$, standard estimates on the lower tails of binomial distributions (Lemma 5.1) yield

$$\Pr\left(|X(C) \cap E(G)| \leqslant 2\mathbb{E}[Z]\right) \leqslant \Pr\left(\mathrm{Bin}\left(\frac{N^2}{16R^2}, p\right) \leqslant \frac{pN^2}{32R^2}\right) \leqslant \exp\left(-\frac{pN^2}{128R^2}\right).$$

Substituting this estimate and the inequality $\Pr(Z > 2\mathbb{E}[Z]) < 1/2$ into (40) yields

$$\Pr\left(G' \not\rightarrow (K_n)_k\right) \leqslant \frac{1}{2} + \exp\left(-\frac{pN^2}{256R^2}\right) \leqslant \frac{1}{2} + e^{-N} \leqslant \frac{3}{4}.$$

In particular, there is a graph $G' \subseteq K_N$ such that $G' \not\supseteq K_{n+1}$ and $G' \rightarrow (K_n)_k$, as claimed.

Proof of Lemma 8.2. Set

$$s = \binom{n}{2}, \qquad q = \frac{p}{(10knR)^{10}\log N}, \qquad \beta = \frac{1}{16kR^2}, \qquad \text{and} \qquad E = \left(\frac{N}{2R}\right)^n.$$

We now verify that we may apply Theorem 1.6, with $\alpha \leftarrow \frac{1}{2k}$, to the hypergraph $\mathcal{H}$. First, as $s \leqslant n^2$, we have, by (38) and (39),

$$\alpha\beta q \cdot v(\mathcal{H}) \geqslant \frac{D \cdot (knR)^{10} \cdot N^{-2/(n+1)}}{10^{10}} \cdot \binom{N}{2} \geqslant N \geqslant (\Gamma n)^{21} \geqslant 10^9 s^7$$

and

$$10^4 s^5 q \leqslant 10^4 s^5 \cdot N^{-2/(n+2)} \leqslant \frac{10^4 n^{10}}{(knR)^{21n}} \leqslant \frac{1}{16kR^2} = \beta,$$

provided that n is sufficiently large. Second, suppose that $t \in \{2, \ldots, s\}$ and let $\ell \in \{3, \ldots, n\}$ be the unique integer satisfying $\binom{\ell-1}{2} < t \leqslant \binom{\ell}{2}$, so that

$$t - 1 \leqslant \binom{\ell}{2} - 1 = \frac{(\ell+1)(\ell-2)}{2} \leqslant \frac{(n+1)(\ell-2)}{2}.$$

Since a graph with t edges must have at least ℓ vertices, we have

$$\Delta_t(\mathcal{H}) = k \cdot \binom{N-\ell}{n-\ell} \leqslant kN^{n-\ell}$$

and, consequently,

$$\left(\frac{q}{10^6 s^5}\right)^{t-1} \cdot \frac{E}{v(\mathcal{H})} \cdot \frac{1}{\Delta_t(\mathcal{H})} \geqslant \left(\frac{q}{10^6 s^5}\right)^{\frac{(n+1)(\ell-2)}{2}} \cdot \frac{\left(N/(2R)\right)^n}{kN^2} \cdot \frac{1}{kN^{n-\ell}}$$

$$\geqslant \left(\frac{p}{10^{16} n^{20} (kR)^{10} \log N}\right)^{\frac{(n+1)(\ell-2)}{2}} \cdot \frac{N^{\ell-2}}{k^2 (2R)^n}$$

$$\geqslant \left(\frac{pN^{2/(n+1)}}{(10knR)^{20} \log N}\right)^{\frac{(n+1)(\ell-2)}{2}} \geqslant 1,$$

where the last inequality follows from (39). The theorem supplies a collection $\mathcal{C}$ of containers for the independent sets of $\mathcal{H}$ satisfying

$$|\mathcal{C}| \leqslant \exp\left(10^4 s^5 \beta^{-1} \log(e/\alpha) \cdot q \log(e/q) \cdot v(\mathcal{H})\right) \tag{41}$$

such that, for every $C \in \mathcal{C}$, either $|C| \leqslant \alpha \cdot v(\mathcal{H})$ or there is a subset $W \subseteq C$ with $|W| \geqslant (1-\beta)|C| \geqslant |C| - \beta k \binom{N}{2}$ and $e(\mathcal{H}[W]) < E$. We now turn to bounding the right-hand side of (41) from above. To this end, observe first that $v(\mathcal{H}) = k\binom{N}{2} \leqslant kN^2$ and that $s^5 \beta^{-1} \log(e/\alpha) \leqslant n^{10} R^2 k \log(2ek)$. Further, it follows from (39) that, if D is sufficiently large,

$$\frac{e}{q} = \frac{e(10knR)^{10} \log N}{p} \leqslant N^{2/(n+1)} \leqslant N$$

and thus $q \log(e/q) \leqslant p/(10knR)^{10}$. Substituting these three estimates into (41) yields

$$|\mathcal{C}| \leqslant \exp\left(10^4 \cdot n^{10} R^2 k \log(2ek) \cdot \frac{p}{(10knR)^{10}} \cdot kN^2\right) \leqslant \exp\left(\frac{pN^2}{256R^2}\right),$$

as desired.

It remains to show that $|X(C)| \geqslant \left(\frac{N}{4R}\right)^2$ for every $C \in \mathcal{C}$. Suppose that this were not true. If $|C| \leqslant \alpha v(\mathcal{H}) = \frac{k}{2k}\binom{N}{2} = \frac{1}{2}\binom{N}{2}$, then

$$|X(C)| \geqslant \binom{N}{2} - |C| \geqslant \frac{1}{2}\binom{N}{2} \geqslant \left(\frac{N}{4R}\right)^2,$$

so we may assume that $|C| > \alpha v(\mathcal{H})$. In particular, there must be a subset $W \subseteq C$ with $|C \setminus W| \leqslant \beta k \binom{N}{2}$ such that $e(\mathcal{H}[W]) < E$. Observe that

$$|X(W)| \leqslant |X(C)| + |C \setminus W| \leqslant \left(\frac{N}{4R}\right)^2 + \beta k \binom{N}{2} \leqslant \frac{N^2}{8R^2} \tag{42}$$

and let $c \colon E(K_N) \to [\![k+1]\!]$ be an arbitrary colouring such that $(e, c_e) \in W$ for every $e \notin X(W)$ and $c_e = k+1$ otherwise. By Lemma 8.1, either $|X(W)| = |c^{-1}(k+1)| \geqslant (1/2) \cdot (N/R)^2$ or the colouring c has at least $(1/2) \cdot (N/R)^n$ monochromatic copies of K_n in colours $1, \ldots, k$. However, the former inequality contradicts (42) and thus the latter

must hold. Finally, note that, if K is an arbitrary copy of K_n in K_N that c colours with some $i \in [\![k]\!]$, then $E(K) \times \{i\} \subseteq W$. This implies that $e(\mathcal{H}[W]) \geqslant (1/2) \cdot (N/R)^n \geqslant E$, contradicting our assumption. $\qquad\square$

8.2. **Induced Ramsey numbers (proof of Theorem 1.5).** Let k be a positive integer, let H be an arbitrary n-vertex graph, let $R = R(n; k)$, and suppose that an integer N satisfies

$$N \geqslant (10n^2 kR)^{7n}. \tag{43}$$

We shall prove that, with probability very close to one, the uniformly chosen random subgraph of K_N is induced-Ramsey for H in k colours, proving that $R_{\mathrm{ind}}(H; k) \leqslant N$. We shall from now on assume that $k \geqslant 2$ and $n \geqslant 3$, as otherwise the assertion of the theorem is trivial.

Suppose that $G \subseteq K_N$. We shall identify a k-colouring $c \colon E(G) \to [\![k]\!]$ of the edges of G with the set

$$\big\{(e, c_e) : e \in E(G)\big\} \cup \big\{(e, 0) : e \in E(K_N) \setminus E(G)\big\} \subseteq E(K_N) \times \{0, \ldots, k\}.$$

(That is, we extend c to a colouring of $E(K_N)$ by colouring all edges of $K_N \setminus G$ zero.) Let $\mathcal{H}$ be the hypergraph with vertex set $E(K_N) \times \{0, \ldots, k\}$ whose edges are all sets of the form

$$\big(\varphi(E(H)) \times \{i\}\big) \cup \big(\varphi(E(K_n) \setminus E(H)) \times \{0\}\big),$$

where $\varphi \colon V(H) \to V(K_N)$ is an arbitrary injection and $i \in [\![k]\!]$. If a graph $G \subseteq K_N$ admits a colouring $c \colon E(G) \to [\![k]\!]$ such that $c^{-1}(i)$ does not contain a copy of H that is induced in G for any $i \in [\![k]\!]$, then c, when viewed as a subset of $E(K_N) \times \{0, \ldots, k\}$, is an independent set of $\mathcal{H}$.

We shall say that a graph $G \subseteq E(K_N)$ is *compatible* with a set $C \subseteq E(K_N) \times \{0, \ldots, k\}$ if there exists a colouring $c \colon E(G) \to [\![k]\!]$ that is contained in C. Equivalently, G is compatible with C if and only if $(e, 0) \in C$ for every $e \in E(K_N) \setminus E(G)$ and $(\{e\} \times [\![k]\!]) \cap C \neq \emptyset$ for every $e \in E(G)$. In other words, defining

$$X_1(C) = \big\{e \in E(K_N) : (e, 0) \notin C\big\} \quad \text{and}$$
$$X_2(C) = \big\{e \in E(K_N) : (e \times [\![k]\!]) \cap C = \emptyset\big\},$$

G is compatible with C if and only if $X_1(C) \subseteq E(G)$ and $X_2(C) \cap E(G) = \emptyset$. The following lemma is key.

Lemma 8.3. *There is a family $\mathcal{C}$ of at most $\exp\left(\frac{N^2}{100R^2}\right)$ containers for the independent sets of $\mathcal{H}$ such that $|X_1(C) \cup X_2(C)| \geqslant \left(\frac{N}{4R}\right)^2$ for every $C \in \mathcal{C}$.*

Assuming that Lemma 8.3 is true, suppose that $G \sim G_{N, 1/2}$. If $G \not\to_{\mathrm{ind}} (H)_k$, then G must be compatible with some container from $\mathcal{C}$. It follows that

$$\Pr\big(G \not\to_{\mathrm{ind}} (H)_k\big) \leqslant \sum_{C \in \mathcal{C}} \Pr\big(X_1(C) \subseteq E(G) \text{ and } X_2(C) \cap E(G) = \emptyset\big)$$

$$\leqslant \sum_{C \in \mathcal{C}} 2^{-|X_1(C) \cup X_2(C)|} \leqslant |\mathcal{C}| \cdot 2^{-(N/(4R))^2}$$

$$\leqslant \exp\left(\frac{N^2}{100R^2} - \frac{N^2 \log 2}{16R^2}\right) \leqslant \exp(-N).$$

Proof of Lemma 8.3. Set

$$s = \binom{n}{2}, \qquad q = \frac{1}{(kR)^5}, \qquad \beta = \frac{1}{16kR^2}, \qquad \text{and} \qquad E = \left(\frac{N}{2R}\right)^n.$$

We now verify that we may apply Theorem 1.6, with $\alpha \leftarrow \frac{1}{3k}$, to the hypergraph $\mathcal{H}$. First, as $s \leqslant n^2$, we have

$$\alpha\beta q \cdot v(\mathcal{H}) = \frac{1}{48(kR)^7} \cdot (k+1)\binom{N}{2} \geqslant N \geqslant 2^{n/2} \geqslant 10^9 s^7$$

and

$$10^4 s^5 q = \frac{10^4 s^5}{(kR)^5} \leqslant \frac{10^4 n^{10}}{(kR)^5} \leqslant \frac{10^4 n^{10}}{(kR)^2 \cdot 2^n} \leqslant \frac{1}{16kR^2} = \beta,$$

provided that n is sufficiently large. Second, suppose that $t \in \{2, \ldots, s\}$ and let $\ell \in \{3, \ldots, n\}$ be the unique integer satisfying $\binom{\ell-1}{2} < t \leqslant \binom{\ell}{2}$, so that

$$t - 1 \leqslant \binom{\ell}{2} - 1 = \frac{(\ell+1)(\ell-2)}{2} \leqslant n(\ell - 2).$$

Since a graph with t edges must have at least ℓ vertices, we have

$$\Delta_t(\mathcal{H}) = k \cdot \binom{N - \ell}{n - \ell} \leqslant kN^{n-\ell}$$

and, consequently,

$$\left(\frac{q}{10^6 s^5}\right)^{t-1} \cdot \frac{E}{v(\mathcal{H})} \cdot \frac{1}{\Delta_t(\mathcal{H})} \geqslant \left(\frac{q}{10^6 s^5}\right)^{n(\ell-2)} \cdot \frac{(N/(2R))^n}{kN^2} \cdot \frac{1}{kN^{n-\ell}}$$

$$= \left(\frac{1}{10^6 (skR)^5}\right)^{n(\ell-2)} \cdot \frac{N^{\ell-2}}{k^2(2R)^n}$$

$$\geqslant \left(\frac{N^{1/n}}{10^7 (skR)^6}\right)^{n(\ell-2)} \geqslant 1,$$

where the last inequality follows from (43). The theorem supplies a collection $\mathcal{C}$ of containers for the independent sets of $\mathcal{H}$ satisfying

$$|\mathcal{C}| \leqslant \exp\left(10^4 s^5 \beta^{-1} \log(e/\alpha) \cdot q \log(e/q) \cdot v(\mathcal{H})\right)$$

$$\leqslant \exp\left(10^7 n^{10} R^2 k \log k \cdot \frac{5\log(ekR)}{(kR)^5} \cdot (k+1)\binom{N}{2}\right) \leqslant \exp\left(\frac{N^2}{100R^2}\right),$$

provided that n is sufficiently large, such that, for every $C \in \mathcal{C}$, either $|C| \leqslant \alpha \cdot v(\mathcal{H})$ or there is a subset $W \subseteq C$ with $|W| \geqslant (1-\beta)|C| \geqslant |C| - \beta(k+1)\binom{N}{2}$ and $e(\mathcal{H}[W]) < E$.

It remains to show that $|X_1(C) \cup X_2(C)| \geqslant \left(\frac{N}{4R}\right)^2$ for every $C \in \mathcal{C}$. Suppose that this were not true and set $X(C) = X_1(C) \cup X_2(C)$. If $|C| \leqslant \alpha v(\mathcal{H}) = \frac{k+1}{3k}\binom{N}{2} \leqslant \frac{1}{2}\binom{N}{2}$, then

$$|X(C)| \geqslant \binom{N}{2} - |C| \geqslant \frac{1}{2}\binom{N}{2} \geqslant \left(\frac{N}{4R}\right)^2,$$

so we may assume that $|C| > \alpha v(\mathcal{H})$. In particular, there must be a $W \subseteq C$ with $|C \setminus W| \leqslant \beta(k+1)\binom{N}{2}$ such that $e(\mathcal{H}[W]) < E$. Observe that

$$|X(W)| \leqslant |X(C)| + |C \setminus W| \leqslant \left(\frac{N}{4R}\right)^2 + \beta(k+1)\binom{N}{2} \leqslant \frac{N^2}{8R^2} \qquad (44)$$

and let $c\colon E(K_N) \to [\![k+1]\!]$ be an arbitrary colouring such that $(e, c_e) \in W$ for every $e \notin X(W)$ and $c_e = k + 1$ otherwise. By Lemma 8.1, either $|X(W)| = |c^{-1}(k+1)| \geqslant (1/2) \cdot (N/R)^2$ or the colouring c has at least $(1/2) \cdot (N/R)^n$ monochromatic copies of K_n in colours $1, \ldots, k$. However, the former inequality contradicts (44) and thus the latter must hold. Let K be an arbitrary copy of K_n in K_N that c colours with some $i \in [\![k]\!]$. Since $E(K) \cap X_1(W) = \emptyset$, then $E(K) \times \{0, i\} \subseteq W$ and, as a result, any injection $\varphi\colon V(H) \to V(K)$ corresponds to an edge of $\mathcal{H}[W]$. This implies that $e(\mathcal{H}[W]) \geqslant (1/2) \cdot (N/R)^n \geqslant E$, contradicting our assumption. $\qquad\square$

References

1. N. Alon, *A non-linear lower bound for planar epsilon-nets*, Discrete Comput. Geom. **47** (2012), 235–244.

2. N. Alon and J. H. Spencer, *The probabilistic method*, fourth ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016.

3. B. Aronov, E. Ezra, and M. Sharir, *Small-size ϵ-nets for axis-parallel rectangles and boxes*, SIAM J. Comput. **39** (2010), 3248–3282.

4. J. Balogh, B. Bollobás, and M. Simonovits, *The typical structure of graphs without given excluded subgraphs*, Random Structures Algorithms **34** (2009), 305–318.

5. J. Balogh, N. Bushaw, M. Collares, H. Liu, R. Morris, and M. Sharifzadeh, *The typical structure of graphs with no large cliques*, Combinatorica **37** (2017), 617–632.

6. J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc. **28** (2015), 669–709.

7. ———, *The method of hypergraph containers*, Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. 3, 2018, pp. 3045–3078.

8. J. Balogh, R. Morris, W. Samotij, and L. Warnke, *The typical structure of sparse K_{r+1}-free graphs*, Trans. Amer. Math. Soc. **368** (2016), 6439–6485.

9. J. Balogh and J. Solymosi, *On the number of points in general position in the plane*, Discrete Anal. (2018), Paper No. 16, 20.

10. A. Bernshteyn, M. Delcourt, H. Towsner, and A. Tserunyan, *A short nonalgorithmic proof of the containers theorem for hypergraphs*, Proc. Amer. Math. Soc. **147** (2019), 1739–1749.

11. B. Bollobás and V. Nikiforov, *The number of graphs with large forbidden subgraphs*, European J. Combin. **31** (2010), 1964–1968.

12. D. Conlon, D. Dellamonica, S. La Fleur, V. Rödl, and M. Schacht, *A note on induced Ramsey numbers*, A journey through discrete mathematics, Springer, Cham, 2017, pp. 357–366.

13. D. Conlon, J. Fox, and B. Sudakov, *On two problems in graph Ramsey theory*, Combinatorica **32** (2012), 513–535.

14. D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, Ann. of Math. (2) **184** (2016), 367–454.

15. W. Deuber, *Generalizations of Ramsey's theorem*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. I, 1975, pp. 323–332. Colloq. Math. Soc. János Bolyai, Vol. 10.

16. P. Erdős, *On some problems in graph theory, combinatorial analysis and combinatorial number theory*, Graph theory and combinatorics (Cambridge, 1983), Academic Press, London, 1984, pp. 1–17.

17. P. Erdős, P. Frankl, and V. Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs Combin. **2** (1986), 113–121.

18. P. Erdős, A. Hajnal, and L. Pósa, *Strong embeddings of graphs into colored graphs*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. I, 1975, pp. 585–595. Colloq. Math. Soc. János Bolyai, Vol. 10.

19. P. Erdős and G. Szekeres, *A combinatorial problem in geometry*, Compositio Math. **2** (1935), 463–470.

20. P. Erdős, D. J. Kleitman, and B. L. Rothschild, *Asymptotic enumeration of K_n-free graphs*, Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II, Accad. Naz. Lincei, Rome, 1976, pp. 19–27. Atti dei Convegni Lincei, No. 17.

21. J. Folkman, *Graphs with monochromatic complete subgraphs in every edge coloring*, SIAM J. Appl. Math. **18** (1970), 19–24.

22. J. Fox and B. Sudakov, *Density theorems for bipartite graphs and related Ramsey-type results*, Combinatorica **29** (2009), 153–196.

23. E. Friedgut, V. Rödl, and M. Schacht, *Ramsey properties of random discrete structures*, Random Structures Algorithms **37** (2010), 407–436.

24. H. Furstenberg and Y. Katznelson, *A density version of the Hales-Jewett theorem*, J. Anal. Math. **57** (1991), 64–119.

25. R. E. Greenwood and A. M. Gleason, *Combinatorial relations and chromatic graphs*, Canadian J. Math. **7** (1955), 1–7.

26. D. Haussler and E. Welzl, *ϵ-nets and simplex range queries*, Discrete Comput. Geom. **2** (1987), 127–151.

27. S. Janson, *Poisson approximation for large deviations*, Random Structures Algorithms **1** (1990), 221–229.

28. P. Kolaitis, H. J. Prömel, and B. Rothschild, *K_{l+1}-free graphs: asymptotic structure and a $0-1$ law*, Trans. Amer. Math. Soc. **303** (1987), 637–671.

29. J. Komlós, J. Pach, and G. Woeginger, *Almost tight bounds for ϵ-nets*, Discrete Comput. Geom. **7** (1992), 163–173.

30. H. Lefmann, *A note on Ramsey numbers*, Studia Sci. Math. Hungar. **22** (1987), 445–446.

31. J. Matoušek, R. Seidel, and E. Welzl, *How to net a lot with little: small ϵ-nets for disks and halfspaces*, Proceedings of the Sixth Annual Symposium on Computational Geometry, 1990, pp. 16–22.

32. R. Morris and D. Saxton, *The number of $C_{2\ell}$-free graphs*, Adv. Math. **298** (2016), 534–580.

33. F. Mousset, R. Nenadov, and A. Steger, *On the number of graphs without large cliques*, SIAM J. Discrete Math. **28** (2014), 1980–1986.

34. R. Nenadov and A. Steger, *A short proof of the random Ramsey theorem*, Combin. Probab. Comput. **25** (2016), 130–144.

35. J. Nešetřil and V. Rödl, *The Ramsey property for graphs with forbidden complete subgraphs*, J. Combinatorial Theory Ser. B **20** (1976), 243–249.

36. J. Pach and G. Tardos, *Tight lower bounds for the size of epsilon-nets*, J. Amer. Math. Soc. **26** (2013), 645–658.

37. H. J. Prömel and A. Steger, *Random l-colorable graphs*, Random Structures Algorithms **6** (1995), 21–37.

38. F. P. Ramsey, *On a problem of formal logic*, Proc. London Math. Soc. (2) **30** (1929), 264–286.

39. V. Rödl, *The dimension of a graph and generalized Ramsey theorems*, Master's thesis, Charles University, 1973.

40. V. Rödl, A. Ruciński, and M. Schacht, *An exponential-type upper bound for Folkman numbers*, Combinatorica **37** (2017), 767–784.

41. ———, *Ramsey properties of random graphs and Folkman numbers*, Discuss. Math. Graph Theory **37** (2017), 755–776.

42. D. Saxton and A. Thomason, *Hypergraph containers*, Invent. Math. **201** (2015), 925–992.

43. ———, *Simple containers for simple hypergraphs*, Combin. Probab. Comput. **25** (2016), 448–459.

44. M. Schacht, *Extremal results for random discrete structures*, Ann. of Math. (2) **184** (2016), 333–365.

Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Moscow Institute of Physics and Technology, Russian Federation
Email address: `jobal@illinois.edu`

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel
Email address: `samotij@tauex.tau.ac.il`