HYPERGRAPH CONTAINERS

DAVID SAXTON AND ANDREW THOMASON

Abstract. We develop a notion of containment for independent sets in hypergraphs. For every r-uniform hypergraph G, we find a relatively small collection $\mathcal{C}$ of vertex subsets, such that every independent set of G is contained within a member of $\mathcal{C}$, and no member of $\mathcal{C}$ is large; the collection, which is in various respects optimal, reveals an underlying structure to the independent sets. The containers offer a straightforward and unified approach to many combinatorial questions concerned (usually implicitly) with independence.

With regard to colouring, it follows that many (including all simple) r-uniform hypergraphs of average degree d have list chromatic number at least $(1/(r-1)^2 + o(1)) \log_r d$. For $r = 2$ this improves Alon's bound and is tight. For $r \geq 3$, previous bounds were weak but the present inequality is close to optimal.

In the context of extremal graph theory, it follows that, for each ℓ-uniform hypergraph H of order k, there is a collection $\mathcal{C}$ of ℓ-uniform hypergraphs of order n each with $o(n^k)$ copies of H, such that every H-free ℓ-uniform hypergraph of order n is a subgraph of a hypergraph in $\mathcal{C}$, and $\log |\mathcal{C}| \leq cn^{\ell - 1/m(H)} \log n$ where $m(H)$ is a standard parameter (there is a similar statement for induced subgraphs). This yields simple proofs of many hitherto difficult results: these include the number of H-free hypergraphs, sparsity theorems of Conlon-Gowers and Schacht, and the full KŁR conjecture.

Likewise, for systems of linear equations the containers supply, for example, bounds on the number of solution-free sets (including Sidon sets, for which we give both lower and upper bounds) and the existence of solutions in sparse random subsets.

Balogh, Morris and Samotij have independently obtained related results.

1. Introduction

The notion of an independent set plays a fundamental role in the study of hypergraphs. An r-uniform hypergraph, or *r-graph*, G is a pair $(V(G), E(G))$ comprising two sets, the vertices $V(G)$ and edges $E(G)$ of G, where each edge $e \in E(G)$ is a set of r elements of $V(G)$. Hence a 2-graph is an ordinary graph. A set $I \subset V(G)$ is *independent* if there is no edge $e \in E(G)$ with $e \subset I$.

Whilst there are many theorems in the literature that can be phrased in terms of estimating the number of independent sets in certain hypergraphs (we shall mention some of these shortly), the question *per se* of how many independent sets there can be in a graph has attracted attention only relatively lately. The maximum number of independent sets in a graph of given average degree can be determined easily via the Kruskal-Katona theorem [33, 27], but for regular graphs the maximum is harder to find: following a good estimate by Alon [1], the exact value for bipartite graphs was determined by Kahn [26] via an elegant entropy argument, and his result was extended to all graphs by Zhao [53]. There are at most $(2^{d+1} - 1)^{n/2d} = 2^{n/2+O(n/d)}$ independent sets in a d-regular graph of order n (that is, having n vertices), and this number is attained by $n/2d$ disjoint copies of $K_{d,d}$.

2000 *Mathematics Subject Classification.* 05C65.

The first author was supported by a grant from the EPSRC.

It would be convenient for many purposes if there were at most $2^{o(n)}$ independent sets in an r-graph G of order n and average degree d, but examples like those just cited show this hope to be a forlorn one. Nevertheless, for the applications we have in mind, it is enough to find a good collection $\mathcal{C}$ of *containers* for independent sets: this is a family of subsets of $V(G)$ such that, for each independent set I, there is a set $C \in \mathcal{C}$ with $I \subset C$, and $|\mathcal{C}| = 2^{o(n)}$. Of course, we could just take $\mathcal{C} = \{V(G)\}$, but this collection would not be helpful: for $\mathcal{C}$ to be of use, a further condition is needed that each container $C \in \mathcal{C}$ is not large, in a sense made precise later.

Another immediate candidate for $\mathcal{C}$ is the collection of *maximal* independent sets, but this too can be large; for example, if d is even, adding a 1-factor into the vertex classes of each $K_{d,d}$ of the graph $(n/2d)K_{d,d}$ produces a $(d+1)$-regular graph with at least $2^{n/4}$ maximal independent sets. (The maximum number of maximal independent sets in any graph of order n was determined by Moon and Moser [36].)

The main purpose of this paper is to show that an r-graph G of average degree d and order n does have a small collection $\mathcal{C}$ of containers. Typically, but not always, $|\mathcal{C}| = 2^{n/d^{1/(r-1+o(1))}}$. Results of this kind were known previously in special cases. Sapozhenko [43, 44, 46, 45] treated regular 2-graphs. Containers for r-graphs were introduced and used in [47] in the case of simple regular r-graphs (a hypergraph is *simple* or *linear* if every pair of vertices lies in at most one edge). However, the most interesting applications require containers for non-regular r-graphs. Finding such containers presents significant difficulties and the method here is unrelated to that of [47].

We describe our main results about containers in §2. The fundamental result is Theorem 2.5 stated in §2.3. It is worth mentioning that the statement applies to all r-graphs G but it gives useful information only if d is large (though independently of n). Following §2, the containers are constructed in §3 and their properties analysed in §4, where Theorem 2.5 is proved. Some corollaries, more amenable to application, are developed in §5 and §6, and in §7 we give an example to show that the main result is, in a sense, best possible. Finally, in §8–12 we give details of some applications.

Before getting down to details, though, we outline those applications to be discussed. Each of them involves, implicitly, dealing with the independent sets in some hypergraph G, and in each case it suffices to find an appropriate set of containers. Once some simple parameter of G has been calculated, the existence of the desired containers follows straight away from the main theorem or one of its consequences in §5–§6, and the applications are finished off by routine arguments (plus, in §10, a non-trivial removal lemma). The results obtained are generally best possible.

Balogh, Morris and Samotij [6] have independently developed a container theorem akin to Corollary 2.7 together with applications, including a proof of Theorem 12.2 for balanced graphs.

1.1. **A little notation.** We use standard notation. In particular, for $m, n \in \mathbb{N}$ we let $[n] = \{1, \ldots, n\}$ and $[m, n] = \{m, \ldots, n\}$. For collections of subsets we write, for example, $[m,n]^{(s)} = \{\sigma \subset [m,n] : |\sigma| = s\}$, $[m,n]^{(>s)} = \{\sigma \subset [m,n] : |\sigma| > s\}$, and so on. As usual, $\mathcal{P}[n]$ denotes the collection of all subsets of $[n]$. If G is a hypergraph we write $e(G) = |E(G)|$ for the number of edges of G and $v(G) = |V(G)|$ for the number of vertices of G. If $S \subset V(G)$ then $G[S]$ denotes the subhypergraph of G induced by S, that is, $G[S] = (S, E(G) \cap \mathcal{P}S)$.

1.2. **List colourings.** A 2-graph G is said to be *k-choosable* if, whenever for each vertex $v \in V(G)$ we assign a list L_v of k colours to v, then it is possible to choose a colour for v from the list L_v, so that no two adjacent vertices receive the same colour. The *list chromatic number* $\chi_l(G)$ (also called the *choice number*) is the smallest k such that G is k-choosable. If all the lists are the same then a list colouring is just an ordinary k-colouring and so $\chi_l(G)$ is at least $\chi(G)$, the ordinary chromatic number of G. This natural definition was first studied by Vizing [52] and by Erdős, Rubin and Taylor [19]. One of the main discoveries of [19] is that $\chi_l(G)$ can be much larger than $\chi(G)$, because $\chi_l(K_{d,d}) = (1 + o(1)) \log_2 d$, whereas $\chi(K_{d,d}) = 2$.

In fact, unlike $\chi(G)$, $\chi_l(G)$ must grow with the minimum degree of the graph G. Alon [2, 3] showed that $\chi_l(G) \geq (1/2 + o(1)) \log_2 d$ holds for any graph G of minimum degree d.

There is a straightforward reason, as pointed out by Alon and Kostochka [4], why the same is not true for r-graphs if $r \geq 3$. Let F be some graph on n vertices, say $F = (n/2)K_2$, and let G be some r-graph each of whose edges contains an edge of F. Then $\chi_l(G) = \chi_l(F)$, so in this example $\chi_l(G) = 2$, whereas the average degree of G can be large. However, if we restrict to simple r-graphs the situation is different. Following work of Haxell and Pei [22] on Steiner systems, Haxell and Verstraëte [23] proved that $\chi_l(G) \geq (\log d/5 \log \log d)^{1/2}$ for simple d-regular 3-graphs G. Alon and Kostochka [4] showed $\chi_l(G) \geq (\log d)^{1/(r-1)}$ for simple d-regular r-graphs G, and in [47] this was improved to $\chi_l(G) = \Omega(\log d)$.

We extend this to all simple r-graphs, at the same time giving a better constant.

Theorem 1.1. *Let $r \in \mathbb{N}$ be fixed. Let G be a simple r-graph with average degree d. Then, as $d \to \infty$,*

$$\chi_l(G) \;\geq\; (1 + o(1)) \, \frac{1}{(r-1)^2} \log_r d$$

holds. Moreover, if G is regular then

$$\chi_l(G) \;\geq\; (1 + o(1)) \, \frac{1}{r-1} \log_r d \,.$$

Note that, for $r = 2$, this improves Alon's bound [3] by a factor of 2 and is best possible. We think that the bound given for regular r-graphs might hold for general r-graphs and moreover that it too might be best possible (see §8).

Theorem 1.1 is a weaker version of Corollary 8.2, which gives a bound for some non-simple r-graphs also. We do not give a general bound for all r-graphs because it would be rather complicated to state: however, in any particular instance, a bound can readily be derived from the results in §8. This would cover, for example the theorem of Alon and Kostochka [5] that if at least half the $(r-1)$-tuples of vertices of G lie in at least m edges then $\chi_l(G) \geq c_r \log m$.

1.3. **H-free graphs.** An ℓ-graph on vertex set $[N]$ is said to be *H-free* if it contains no subgraph isomorphic to the ℓ-graph H.

As far as H-free graphs are concerned, our main result is this: for any given ℓ-graph H, though there are many H-free ℓ-graphs, each of these is contained in one of a very few ℓ-graphs that are almost H-free. More exactly, there is a small collection $\mathcal{C}$ of ℓ-graphs, each H-free ℓ-graph being a subgraph of an ℓ-graph in $\mathcal{C}$, and no ℓ-graph in $\mathcal{C}$ having more than $o(N^{v(H)})$ copies of H. The main content of the theorem is that the number of containers is very small. For graphs at least, Szemerédi's regularity lemma gives a set of containers with $\log|\mathcal{C}| = o(N^2)$, but the size of $\mathcal{C}$ in the theorem is much smaller. It is expressed in terms of a parameter $m(H)$ that appears often in the literature.

Definition 1.2. For an ℓ-graph H with $e(H) \geq 2$, let

$$m(H) = \max_{H' \subset H,\, e(H')>1} \frac{e(H')-1}{v(H')-\ell}.$$

Sometimes, H is called (strictly) balanced if the maximum is attained (uniquely) when $H' = H$. However, this restriction is not needed in any of our arguments and it is ignored.

We then obtain the following, where $\pi(H) = \lim_{N\to\infty} \mathrm{ex}(N,H)\binom{N}{\ell}^{-1}$.

Theorem 1.3. *Let H be an ℓ-graph with $e(H) \geq 2$ and let $\epsilon > 0$. For some $c > 0$ and for N sufficiently large, there exists a collection $\mathcal{C}$ of ℓ-graphs on vertex set $[N]$ such that*

(a) *for every H-free ℓ-graph I on vertex set $[N]$ there exists $C \in \mathcal{C}$ with $I \subset C$,*
(b) *moreover, for every pair I, C in (a), there exists $T = (T_1, \ldots, T_s)$ where $T_i \subset I$, $s \leq c$ and $\sum_i |T_i| \leq cN^{\ell-1/m(H)}$, such that $C = C(T)$,*
(c) *for every ℓ-graph $C \in \mathcal{C}$, the number of copies of H in C is at most $\epsilon N^{v(H)}$, and $e(C) \leq (\pi(H)+\epsilon)\binom{N}{\ell}$,*
(d) $\log|\mathcal{C}| \leq cN^{\ell-1/m(H)}\log N$.

The condition (b) just says that C is determined by T, which is comprised of small subsets of $V(G)$ (actually subsets of I — this sometimes matters, because it means (b) is stronger than (d)).

The existence of $\mathcal{C}$ follows straightforwardly from the results in §2, as shown in §9, by considering the $e(H)$-graph $G = G(N,H)$, whose $n = \binom{N}{\ell}$ vertices are the ℓ-sets in $[N]$, and whose edges are subsets of $V(G)$ spanning a copy of H in $[N]$. The subsets of $V(G)$ are then ℓ-graphs with vertex set $[N]$, independent sets in G corresponding to H-free ℓ-graphs.

One corollary of Theorem 1.3 is the next one. In the case $\ell = 2$, it was proved for complete H by Erdős, Kleitman and Rothschild [17] and for general H by Erdős, Frankl and Rödl [16]. Nagle, Rödl and Schacht [37] proved it for general ℓ using hypergraph regularity methods. The proof here follows quite easily from Theorem 1.3 using the supersaturation theorem of Erdős and Simonovits [47]; details are in §9.

Corollary 1.4. *Let H be an ℓ-graph. The number of H-free ℓ-graphs on vertex set $[N]$ is $2^{(\pi(H)+o(1))\binom{N}{\ell}}$.*

For ℓ-graphs H which satisfy $\mathrm{ex}(N,H) = o(N^\ell)$ (when $\ell = 2$ this means H is bipartite), we have $\pi(H) = 0$, and Corollary 1.4 is unhelpful. Nevertheless our results can still be useful, provided appropriate information about $G(N,H)$ is available. The simplest case is $\ell = 2$ and $H = K_{2,2} = C_4$, where it is well known that $\mathrm{ex}(N, C_4) = (1/2 + o(1))N^{3/2}$ (Erdős, Rényi and Sós [18]), implying the trivial upper bound $2^{O(N^{3/2}\log N)}$ for the number of C_4-free graphs. A direct application of our results readily gives the bound $2^{O(N^{3/2})}$, but we don't give details because Kleitman and Winston [28] obtained a finer bound, namely $2^{(1.082+o(1))N^{3/2}}$. More generally, the number of $K_{s,t}$-free graphs has been well estimated by Balogh and Samotij [7].

Alongside the many results about H-free graphs, there is a corresponding corpus about *induced* H-free graphs, that is, graphs with no induced subgraph isomorphic to H. The number of induced H-free graphs was closely estimated by Prömel and Steger [38], and there have been many subsequent refinements.

If I is an induced H-free ℓ-graph, we need to ask what kind of object C must be in order that the inclusion $I \subset C$ is helpful; if C itself is just an ℓ-graph and $I \subset C$ means I is a subgraph of C, as in Theorem 1.3, then the induced subgraphs of I differ from those of C, which is no use. We borrow the notion of *2-coloured multigraph* from [34, 51]. A 2-coloured ℓ-multigraph C on vertex set $[N]$ is a pair of edge sets $C_R, C_B \subset [N]^{(\ell)}$, which we call the red and the blue edge sets. Let I be an ℓ-graph on $[N]$. Then we write $I \subset C$ if $E(I) \subset C_R$ and $[N]^{(\ell)} \setminus E(I) \subset C_B$. Thus edges in $C_R \cap C_B$ do not affect whether $I \subset C$.

Theorem 1.5. *Let H be an ℓ-graph and let $\epsilon > 0$. For some $c > 0$ and for N sufficiently large, there exists a collection $\mathcal{C}$ of 2-coloured ℓ-multigraphs on vertex set $[N]$ such that*

(a) *for every ℓ-graph I on vertex set $[N]$ with no induced copy of H there exists $C \in \mathcal{C}$ with $I \subset C$,*
(b) *for every $C \in \mathcal{C}$, the number of copies of H in C is at most $\epsilon N^{v(H)}$,*
(c) $\log |\mathcal{C}| \leq cN^{\ell-(v(H)-\ell)/\left(\binom{v(H)}{\ell}-1\right)} \log N$.

This theorem can be used to recover basic results, akin to Corollary 1.4, about the number of induced H-free ℓ-graphs. In fact we can state a probabilistic version just as readily. Let $G^{(\ell)}(N,p)$ be a random ℓ-graph obtained by choosing edges independently from the complete ℓ-graph $K_N^{(\ell)}$ with probability p. The next statement involves a function $h_p(H)$, whose definition is natural enough but which is deferred until §9.

Theorem 1.6. *Let $0 < p < 1$ and let H be an ℓ-graph. Then*

$$\mathbb{P}(G^{(\ell)}(N,p) \text{ is induced-}H\text{-free}) = 2^{-(h_p(H)+o(1))\binom{N}{\ell}}.$$

For graphs, that is, $\ell = 2$, this theorem was proved for $p = 1/2$ by Prömel and Steger [38, Theorem 1.3] and for general p by Bollobás and Thomason [8, Theorem 1.1] (as illuminated by Marchant and Thomason [35]). For general ℓ it was proved for $p = 1/2$ by Dotson and Nagle [13], using hypergraph regularity techniques.

It will be clear that similar arguments to those described in this section can be used to obtain container results about other structures, such as tournaments.

1.4. **Linear equations.** Let F be either a finite field or the set of integers $[N]$. We consider linear systems of equations $Ax = b$, where A is a $k \times r$ matrix with entries in F, $x \in F^r$ and $b \in F^k$. We call such a triple (F, A, b) a $k \times r$ *linear system.*

Definition 1.7. For a $k \times r$ linear system (F, A, b), a subset $I \subset F$ is *solution-free* if there is no $x \in I^r$ with $Ax = b$, and $\mathrm{ex}(F, A, b)$ is the maximum size of a solution-free subset.

The notion of a solution-free subset is analogous to that of an H-free hypergraph in the previous section. Once again, our contribution to this topic is a container theorem for solution-free sets. The statement (which extends to equations over abelian groups) is given in Theorem 10.2; it requires a few technical definitions so we omit it from the introduction.

Nevertheless we mention a consequence for counting solution-free subsets. For an equation $Ax = b$, how many solution-free subsets of F are there? A well-known instance of this question is to find the number of subsets $S \subset [N]$ containing no solution to $x + y = z$; the asymptotic answer, conjectured by Cameron and Erdős [11], was given by Green [24] and by Sapozhenko [46].

For a general system, every subset of a solution-free set is itself solution-free, so there are at least $2^{\mathrm{ex(F,A,b)}}$ solution-free sets. For a single equation (the case $k = 1$), it was shown by Green [25] that there are at most $2^{\mathrm{ex(F,A,b)+o(|F|)}}$ solution-free subsets. The same bound does not always hold for $k \geq 2$. If some variables are closely tied to other variables — say the equations imply that $x = y$ — then there can be significantly more than $2^{\mathrm{ex(F,A,b)}}$ solution-free sets. However, a natural condition on A rules out closely tied variables, and in this case Green's bound holds good.

Definition 1.8. We say that A has *full rank* if given any $b \in F^k$ there exists $x \in F^r$ with $Ax = b$. We then say that A is *abundant* if it has full rank and every $k \times (r-2)$ submatrix obtained by removing a pair of columns from A still has full rank.

Theorem 1.9. *Let (F, A, b) be a $k \times r$ linear system with A abundant. Then the number of solution-free subsets of F is $2^{\mathrm{ex(F,A,b)+o(|F|)}}$. Here $o(1) \to 0$ as $|F| \to \infty$, with A fixed in the case $F = [N]$.*

For example, take $A = (1, 1, -1)$ and $b = (0)$. Theorem 1.9 says that the number of sum-free subsets of $[N]$ is $2^{N/2+o(N)}$, giving a new proof of the weak form of the Cameron and Erdős conjecture, proved independently by Alon [1], by Calkin [10] and by Erdős and Granville (unpublished).

Similar results hold when F is an abelian group. For the proof of Theorem 1.9 we need a result for equations analogous to the supersaturation theorem for graphs: this is the removal lemma of Král', Serra and Vena [31, 32]. More information is given in §10.

1.5. **Sidon sets.** For linear systems where $\mathrm{ex}(F, A, b) = o(|F|)$, Theorem 1.9 is uninformative. One of the most prominent examples is that of Sidon sets. A set $A \subset [n]$ is *Sidon* if every sum of two elements is distinct, i.e., there are no solutions to $w + x = y + z$ with $\{w, x\} \neq \{y, z\}$. It is easy to see that a Sidon set has size at most $\lceil \sqrt{2n} \rceil$, since each of the $|S|(|S| - 1)/2$ values $x - y$, where $x, y \in S$ and $y < x$, are distinct and lie in $\{1, \ldots, n-1\}$. Erdős and Turán [21] improved this upper bound to $|S| \leq (1 + o(1))\sqrt{n}$, and there are examples achieving this bound.

It is natural to ask, as Cameron and Erdős did [11], how many Sidon sets there are, and the answer clearly lies between $2^{(1+o(1))\sqrt{n}}$ and $2^{O(\sqrt{n}\log n)}$. Neither of these bounds, it turns out, is tight.

Theorem 1.10. *There are between $2^{(1.16+o(1))\sqrt{n}}$ and $2^{(55+o(1))\sqrt{n}}$ Sidon subsets of $[n]$.*

The lower bound gives a negative answer to the open question of whether there are only $2^{(1+o(1))\sqrt{n}}$ Sidon sets. The upper bound, also proved by Kohayakawa, Lee, Rödl and Samotij [29] (in fact with a better constant), follows directly by plugging in the appropriate numbers into a container-counting theorem. For details see §11.

1.6. **Sparsity.** In recent times, there has been interest in the extent to which theorems holding for dense structures hold also for sparse random substructures. Our results can be applied in this context, and we give some illustrative examples involving the notions of H-free graphs and solution-free subsets already discussed.

The application of our results always fits a simple paradigm. Typically we want some statement to hold for a random substructure, with high probability; by considering an appropriate collection of containers, the fact that there is a small number of containers means that the work is reduced, via the union bound, to establishing a (generally much simpler) statement for a single container.

For example, consider a random ℓ-graph $G^{(\ell)}(N,p)$, as defined in §1.3. Evidently there are H-free subgraphs of $G^{(\ell)}(N,p)$ with $p\,\mathrm{ex}(N,H)$ edges, but are there significantly larger H-free subgraphs? Kohayakawa, Łuczak and Rödl [30] conjectured that if $p > cN^{-1/m(H)}$ then H-free subgraphs of $G^{(\ell)}(N,p)$ almost surely have at most $(1+o(1))p\,\mathrm{ex}(N,H)$ edges. This conjecture was recently proved by Conlon and Gowers [12] (for strictly balanced H) and by Schacht [48], using different methods. Our methods give an alternative proof. For each container $C \in \mathcal{C}$ given by Theorem 1.3, it is easily seen that with high probability $G^{(\ell)}(N,p)$ contains not much more than $pe(C) \le (\pi(H)+o(1))p\binom{N}{\ell}$ edges of C, and by the union bound this holds for all $C \in \mathcal{C}$, and hence also for all H-free ℓ-graphs.

Theorem 1.11. *Let H be an ℓ-graph and let $0 < \gamma < 1$. For some $c > 0$, for N sufficiently large and for $p \ge cN^{-1/m(H)}$, the following event holds with probability greater than $1 - \exp\{-\gamma^3 p\binom{N}{\ell}/512\}$:*

$$\textit{every } H\textit{-free subgraph of } G^{(\ell)}(N,p) \textit{ has at most } (\pi(H)+\gamma)p\binom{N}{\ell} \textit{ edges.}$$

Other related conjectures, including what has become known as the KŁR conjecture, were made in [30], not all of which have previously been proved in full, but they all follow from Theorem 1.3 in a similar way. Details are in §12.

The same arguments can be applied to solution sets of linear equations. Here is a typical consequence.

Theorem 1.12 (Conlon and Gowers [12], Schacht [48])**.** *Let $\ell \ge 3$ and $\epsilon > 0$. There exists a constant $c > 0$ such that for $p \ge cN^{-1/(\ell-1)}$, if $X \subset [N]$ is a random subset chosen with probability p, then with probability tending to 1 as $N \to \infty$, any subset of X of size $\epsilon|X|$ contains an arithmetic progression of length ℓ.*

Further examples and details can be found in §12.

2. Containers

A couple of simple notions are needed for the statement of the main theorem, and we define these now. They are the co-degree function and degree measure.

2.1. The co-degree function $\delta(G,\tau)$.

The present results about containers were originally motivated by the study of the list chromatic number of simple hypergraphs, described in §8. The main difficulties in the construction of containers are already present in the simple case. However the method can be adapted efficiently to any hypergraph. The size and number of the containers depends on the way the edges overlap, but the dependence can be encapsulated by a single parameter, which is usually quite straightforward to compute. This parameter appears in most of the theorems.

First, we define the degree of a subset of vertices, in the natural way.

Definition 2.1. The *degree* of set of vertices $\sigma \subset V(G)$ is the number of edges containing σ; that is,

$$d(\sigma) \;=\; |\{\, e \in E(G) \,:\, \sigma \subset e \,\}|\,.$$

If $|\sigma| = 1$, that is $\sigma = \{v\}$ where $v \in V(G)$, we generally write $d(v)$ instead of $d(\{v\})$.

We can now define the co-degree function $\delta(G,\tau)$. This is a function of the parameter τ, a parameter used in the construction of containers.

Definition 2.2. Let G be an r-graph of order n and average degree d. Let $\tau > 0$. Given $v \in V(G)$ and $2 \le j \le r$, let

$$d^{(j)}(v) \;=\; \max\{\, d(\sigma) \,:\, v \in \sigma \subset V(G),\, |\sigma| = j \,\}\,.$$

If $d > 0$ we define δ_j by the equation

$$\delta_j\, \tau^{j-1} n d \;=\; \sum_v d^{(j)}(v)\,.$$

Then the *co-degree function* $\delta(G,\tau)$ is defined by

$$\delta(G,\tau) \;=\; 2^{\binom{r}{2}-1} \sum_{j=2}^{r} 2^{-\binom{j-1}{2}} \delta_j\,.$$

If $d = 0$ we define $\delta(G,\tau) = 0$.

There is nothing significant about the powers of 2 in the definition; they are just constants needed for Lemma 4.2.

Remark 2.3. The parameter τ appears in the main theorem, Theorem 2.5, and the smaller that τ can be made the stronger the result becomes. The constraint on τ comes from a lower bound on $\delta(G,\tau)$. It can be seen that if τ decreases then the values of the δ_j increase, and hence so does $\delta(G,\tau)$; indeed $\delta(G,\tau) \to 0$ as $\tau \to \infty$. A typical application will have τ as small as possible subject to $\delta(G,\tau)$ being less than some constant, say $1/r!$.

Here are some observations intended to indicate the optimal size of τ. Observe first that, unless G has isolated vertices, $d^{(j)}(v) \ge 1$ holds for all v, and so $\delta_j \ge \tau^{1-j}/d$. The largest of these bounds is $\delta_r \ge \tau^{1-r}/d$ (τ is invariably less than one) and so, for fixed r and large d, it will always be that for $\delta(G,\tau)$ to be small we must choose τ at least as large as $d^{-1/(r-1)}$.

In a simple hypergraph, $d(\sigma) \le 1$ holds whenever $|\sigma| \ge 2$, and so $\delta_j \le \tau^{1-j}/d$. In this case the largest of the δ_j's is δ_r, and we can make $\delta(G,\tau)$ small by choosing τ just a little larger than $d^{-1/(r-1)}$. In fact, for any hypergraph whose edges are uniformly distributed, δ_r is once again the δ_j which dominates, as a simple calculation (which we omit) shows, so here again $\delta(G,\tau)$ is small if τ is larger than $d^{-1/(r-1)}$.

Sometimes, though, the dominant δ_j is not δ_r. One example of this is in the case of Sidon sets: when $|S| < n^{2/3}$ it is the value of δ_2 which is the most important. Another example is the hypergraph describing H-free ℓ-graphs: here the most important δ_j is determined by which subgraph $H' \subset H$ achieves the maximum of $(e(H')-1)/(v(H')-\ell)$, and this is how $m(H)$ enters in. But in each of our examples the values are easily checked.

In summary, we must always choose $\tau \ge d^{-1/(r-1)}$, and for simple or uniformly distributed hypergraphs the value need not be much larger. But there are applications which are far from uniformly distributed, where τ needs to be larger and where the behaviour of $\delta(G,\tau)$ will prove crucial.

2.2. **Degree measure.** We mentioned in the introduction that the containers must not be too large. For a substantial number of applications it suffices that $|C| \le (1-c)|G|$ for some constant c. This is achievable for regular hypergraphs but it clearly is unattainable in general; for example, if $G = K_{d,n-d}$ (which, for large n, has average degree close to $2d$)

then some container must have size at least $n - d$. Other applications require that the number of edges inside a container, that is, $e(G[C])$, is small. This is attainable in general but there are applications where what matters is $|C|$ rather than $e(G[C])$.

In fact we measure the size of containers by what we call *degree measure*. It turns out that if the degree measure is bounded then it is possible to recover all the properties of containers that are needed.

Definition 2.4. Let G be an r-graph of order n and average degree d. Let $S \subset V(G)$. The *degree measure* $\mu(S)$ of S is defined by

$$\mu(S) \,=\, \frac{1}{nd} \sum_{u \in S} d(u)\,.$$

Thus μ is a probability measure on $V(G)$. Note that if G is regular then $\mu(S) = |S|/n$, which is the *uniform* measure of S.

2.3. **The main theorem.** The essential idea for demonstrating that a collection of containers is small is this: each container is specified by just a small set of vertices, meaning that, given some small set $T \subset V(G)$, there is a construction which produces another larger set $C = C(T)$, and the construction is such that for any independent set I there is some $T \subset I$ which produces a C with $I \subset C$. If T is small, the number of choices for T is not large, and so the number of containers is not large. Actually, we shall generate C from not just one small set but an r-tuple $T = (T_{r-1}, \dots, T_0) \in \mathcal{P}^r I$ of small sets; the principle is the same.

We already introduced the parameter τ. Essentially, τ will be the value of $\mu(T)$, which is why we want τ to be as small as possible. In fact, the theorem guarantees $\mu(T_i) \le 2\tau/\zeta$ where ζ is some small constant at our disposal. Often we shall take $\zeta = 1/12r!$ but sometimes it is useful to choose a smaller value.

We use one more piece of shorthand. Let $T = (T_{r-1}, \dots, T_1, T_0) \in \mathcal{P}^r[n]$ and let $w \in [n]$. Then we define $T \cap [w] \,=\, (T_{r-1} \cap [w], \dots, T_1 \cap [w], T_0 \cap [w])$.

Theorem 2.5. *Let G be an r-graph with vertex set $[n]$, where $d(v)$ decreases with v. Let $\tau, \zeta > 0$ satisfy $\delta(G,\tau) \le \zeta$. Then there is a function $C : \mathcal{P}^r[n] \to \mathcal{P}[n]$, such that, for every independent set $I \subset [n]$ there exists $T = (T_{r-1}, \dots, T_0) \in \mathcal{P}^r I$ with*

(a) $I \subset C(T)$,
(b) $\mu(T_0), \dots, \mu(T_{r-1}) \le 2\tau/\zeta$,
(c) $|T_0|, \dots, |T_{r-1}| \le 2\tau n/\zeta^2$, *and*
(d) $\mu(C(T)) \le 1 - 1/r! + 4\zeta + 2r\tau/\zeta$.

Moreover C has the online property, meaning that $C(T) \cap [w] = C(T \cap [w]) \cap [w]$ for all $T \in \mathcal{P}^r[n]$ and $w \in [n]$.

In fact, the above is true for all sets $I \subset [n]$ for which either $G[I]$ is $\lfloor \tau^{r-1}\zeta e(G)/n \rfloor$-degenerate or $e(G[I]) \le 2r\tau^r e(G)/\zeta$.

Remark 2.6. It is worth making a few observations at this point.

- Roughly speaking, the theorem says that for each I there exists $T \subset I$ with $\mu(T) \le \tau$, $I \subset C(T)$ and $\mu(C) \le 1 - 1/r!$, provided τ is large enough to make $\delta(G,\tau)$ small.
- The online property is needed only for certain applications, principally Theorem 2.8. More is said about this in §4.4.

- Ordering the vertices by degrees is generally unnecessary: it is used only to obtain (c) and the online property simultaneously, and to accommodate b-degenerate graphs.
- The container construction method makes essentially no use of the independence of the sets I, so we include an extension to two kinds of sparse subset, where either $G[I]$ is b-degenerate or $e(G[I]) \le bn$, and b is small. As usual, we say $G[I]$ is *b-degenerate* if for every subset $J \subset I$ there is a vertex in $G[J]$ of degree at most b. This is equivalent to saying that every subgraph $G[J]$ is sparse; indeed if $G[I]$ is b-degenerate then $G[J]$ has at most $b|J|$ edges for all $J \subset I$, and conversely if $G[J]$ has at most $b|J|$ edges for all $J \subset I$ then $G[I]$ is rb-degenerate.
- The theorem is best possible in two senses. First, the examples in §7 show that our method cannot give $\mu(C) < 1 - 1/r!$. Secondly, the bound on the measure of the generating sets T_i, implicit in the constraint $\delta(G, \tau) \le \zeta$ which determines how small τ can be, can not be improved significantly. One way to see this is that an improvement here would give an improvement in some of the applications, say Theorem 1.11, but these are known to be best possible.

As mentioned in the introduction, Theorem 2.5 has a variety of consequences and weaker forms which are easier to apply directly. These are discussed in §5–§6, but we state a couple of them here for illustration.

2.4. **Tight containers.** In §5 we give a number of consequences of Theorem 2.5 concerning the number of edges $e(G[C])$ inside the container. The following is the weakest of these; it provides, in a handy format, a collection of containers each with few internal edges, the size of the collection being bounded by a simple function of τ.

Corollary 2.7. *Let G be an r-graph on vertex set $[n]$. Let $0 < \epsilon < 1/2$. Suppose that τ satisfies $\delta(G, \tau) \le \epsilon/12r!$ and $\tau \le 1/144r!^2 r$. Then there exists a constant $c = c(r)$ and a collection $\mathcal{C} \subset \mathcal{P}[n]$ such that*

(a) *for every independent set I there exists $T = (T_1, \ldots, T_s) \in \mathcal{P}^s I$ with $I \subset C(T) \in \mathcal{C}$, $|T_i| \le c\tau n$ and $s \le c \log(1/\epsilon)$,*
(b) *$e(G[C]) \le \epsilon e(G)$ for all $C \in \mathcal{C}$,*
(c) *$\log |\mathcal{C}| \le c \log(1/\epsilon) n\tau \log(1/\tau)$.*

Moreover, (a) holds for all sets $I \subset [n]$ for which either $G[I]$ is $\lfloor \epsilon\tau^{r-1} e(G)/12r!n \rfloor$-degenerate or $e(G[I]) \le 24\epsilon r! r\tau^r e(G)$.

2.5. **Uniformly bounded containers.** Next we give a consequence for applications when the size $|C|$ of the container is of interest. As mentioned earlier, examples such as $K_{d,n-d}$ show that it is not possible always to guarantee that $|C|$ is bounded. However, it is possible to bound $|C \cap [v]|$ for some initial segment $[v]$ of the vertex set $[n]$, and this can be done so that the number of different sets $|C \cap [v]|$ is small (a function of v rather than of n). More exactly, each container C "nominates" $v = g(C)$, so that $|C \cap [v]|$ is small and not many containers nominate any given v. This consequence of the main result is suitable for dealing with list colourings, for example. The statement is actually in terms of tuples $(C_1, \ldots, C_t)$ of containers rather than individual containers, since this is ultimately more efficient. The theorem is explained more in §6.

Theorem 2.8. *Let G be an r-graph on vertex set $[n]$, for which the degree sequence is decreasing. Let $0 < \zeta \le 1/12r!$. Suppose that $\delta(G,\tau) \le \zeta$, that $\tau \le \zeta^2/r$, and that $k \in [n]$ satisfies $\mu([k]) \le \zeta/2r!$. Let $t \in \mathbb{N}$.*

Then there exists a collection $\mathcal{C} \subset \mathcal{P}[n]$ and a map $g : \mathcal{C}^t \to [k,n]$, with the following properties:

(a) *for all independent sets I there is some $C \in \mathcal{C}$ with $I \subset C$,*

(b) *for all $v \in [n]$*

$$\log |\{ (C_1 \cap [v], \ldots, C_t \cap [v]) \, : \, g(C_1, \ldots, C_t) = v \}| \, \le \, \zeta^{-2} vtr\tau \log(1/\tau),$$

(c) *and for all $(C_1, \ldots, C_t) \in \mathcal{C}^t$*

$$\frac{1}{t} \sum_{i=1}^{t} |C_i \cap [v]| \, \le \, (1 - \frac{1}{r!} + 8\zeta)v \,,$$

where $v = g(C_1, \ldots, C_t)$.

Moreover, (a) holds for all sets $I \subset [n]$ for which either $G[I]$ is $\lfloor \tau^{r-1}\zeta e(G)/n \rfloor$-degenerate or $e(G[I]) \le 2r\tau^r e(G)/\zeta$.

3. The Algorithm

In this section we describe the method of building containers and establish the basic facts about them.

Sapozhenko [43, 44, 45] gives a way to build containers in the case of ordinary (2-uniform) graphs that are close to being regular. However even for graphs of this kind there is more than one approach which works. It is not obvious how to extend these graph methods to hypergraphs, each method offering a few plausible possibilities, most of which fail but some of which succeed. Indeed we eventually found more than one way to construct containers for regular 3-graphs, and one of these, which extended to regular r-graphs, was described in [47]: it is a random-based method. But we have found it difficult to find a method which fulfils the goals of working for general (not just regular) r-graphs and which also has the necessary online property. The method of [47] would be enough for some applications, such as list colouring, if it were true that each r-graph of average degree d contained a subgraph with degrees in the range $c_1 d^\alpha$ to $c_2 d^\alpha$, for some constants c_1, c_2 and α. But it is possible to construct examples of r-graphs where this is far from the truth (these examples bear some similarity to the 2-graphs of Pyber, Rödl and Szemerédi [39]).

The method given below fulfils all our requirements. The reason for the remarks just made is that, whilst we do attempt to motivate the construction, we cannot do so fully because it is hard to explain why one construction should work when another similar one does not. The devil is sometimes in the detail, and a prolonged discussion would be unjustified.

There are two aspects to the construction: the production of a suitable small set T from a set I, which we call *pruning*, and the production of C from T, which we call *building*. (In fact, as mentioned already, we produce not just one small set but an r-tuple $T = (T_{r-1}, \ldots, T_0) \in \mathcal{P}^r I$ of small sets.) These two aspects are very closely intertwined, and it is convenient to describe both in terms of a single algorithm which has two slightly differing modes, a *build mode* and a *prune mode*.

3.1. **Constructing containers.** Given an r-graph G with vertex set $[n]$, and a sequence of subsets $T_{r-1}, T_{r-2}, \dots, T_1$ of $[n]$, we produce a sequence of *s-multigraphs* P_s for $s = r-1, \dots, 1$. This means that P_s is s-uniform but multiple edges are allowed; in other words $E(P_s)$ is a *multiset*.

Each edge $\{u_{s-1}, u_{s-2}, \dots, u_0\} \in E(P_s)$ with $u_{s-1} < u_{s-2} < \cdots < u_0$ will come from an edge $\{v_{r-1}, v_{r-2}, \cdots, v_s, u_{s-1}, u_{s-2}, \dots, u_0\} \in E(G)$, where $v_{r-1} < \cdots < v_s < u_{s-1}$ and $v_j \in T_j$, $r-1 \geq j \geq s$. Equivalently, each edge of P_s is an edge of P_{s+1} whose first vertex, which is in T_s, has been removed. The reason P_s is defined as a multigraph, even if G itself does not have multiple edges, is so that distinct edges of G give rise to distinct edges of P_s.

The multigraph P_1 is 1-uniform: its edges are single vertices. If I is an independent set and the sets T_j are chosen within I, as they will be, then evidently the members of $E(P_1)$ cannot be vertices in I, and so the container C can be chosen from vertices not in $E(P_1)$. Our first aim, then, is to ensure that $E(P_1)$ is as large as possible, and to this end we attempt to make $E(P_s)$ large for each s. However this aim has to be balanced against keeping the sets T_s small.

Hence we shall choose a parameter τ, so that, roughly speaking, T_s will comprise a proportion τ of the vertices (in degree measure), and we might hope the size of $E(P_s)$ to be roughly τ times the size of $E(P_{s+1})$. This means the average degree of P_s will typically be around $\tau^{r-s}d$. The parameter τ is the same as that discussed in §2.1 and the constraint $\tau \geq d^{-1/(r-1)}$ described there is precisely what is needed to ensure that $E(P_1)$ contains something worthwhile.

However not every edge of G with its first $r-s$ vertices in $T_{r-1}, \dots, T_s$ will be admitted as an edge of P_s, but only a selection of these. We do not allow edges into P_s if they increase the degree of some vertex, or the degree of some subset $\sigma \subset [n]$, beyond some agreed threshold. We define the degree of σ in the multigraph P_s to be

$$d_s(\sigma) = |\{e \in E(P_s) : \sigma \subset e\}|,$$

where we note that here we are counting edges with multiplicity in the multiset $E(P_s)$. (Naturally we may write $d_s(v)$ instead of $d_s(\{v\})$ if $v \in [n]$.) There are several reasons for wanting to bound the degrees in P_s. One reason is the hope of keeping the vertex degrees near to τ^{r-s} times the degrees in G, so that degree measure in P_s relates to measure in G; in particular, small sets of vertices cannot account for most of the edges of P_s unless those sets have large measure in G. A second reason for controlling degrees of subsets is that only by doing so can we restrain the degrees of vertices at later stages: this comes out in the proof of Lemma 4.2.

So we proceed in the following way. We begin with $P_r = G$, and then apply the straightforward algorithm below to construct P_s from P_{s+1} using T_s, with s taking the values $r-1, r-2, \dots, 1$ in turn. During the application of the algorithm, the degrees $d_s(\sigma)$ in P_s will grow, as edges are added. We denote by Γ_s the collection of vertices and subsets whose degrees have reached their bound, and we do not permit the addition to P_s of any edge which contains a current member of Γ_s. The set Γ_s will grow too during the construction.

As mentioned before, the algorithm can be run in two modes: prune mode and build mode. The aim of successive runs of prune mode is to produce sets $T_{r-1}, \dots, T_1$, and the aim of build mode is to produce sets $C_{r-1}, \dots, C_1$ which will form the basis of a container. Thus in a single run of prune mode, an independent set I is input and some set T_s is output, whereas in build mode a set T_s is input and a set C_s is output. Both forms of the algorithm include the construction of P_s from P_{s+1}.

The essential difference between the modes is this. The algorithm inspects the vertices in $[n]$ one by one. In prune mode, where I is part of the input, if the vertex is in I a decision is made, according to some rule, whether to place the vertex into T_s. In build mode, where T_s is input but I is not, the vertex is inspected to see whether it passes the rule; if it passes, but the vertex is not in T_s, we know it could not have been in I, so it is removed from C_s (with the aim of making a container via C_s). In both modes, if the vertex is in T_s, appropriate edges are added to P_s.

Two real numbers are included in the input to the algorithm. The parameter τ is the more important and has already been discussed. The parameter ζ is a small constant, often in practice chosen to be $1/12r!$ but sometimes chosen smaller. It is used in the rule to decide membership of T_s.

The independence of the set I is not actually used by the algorithm, and it is useful to define the algorithm for general subsets $I \subset [n]$.

Algorithm

input an r-graph G on vertex set $[n]$
an $(s+1)$-multigraph P_{s+1} on vertex set $[n]$
parameters $\tau, \zeta > 0$
in prune mode a subset $I \subset [n]$
in build mode a subset $T_s \subset [n]$

output an s-multigraph P_s on vertex set $[n]$
in prune mode a subset $T_s \subset [n]$
in build mode a subset $C_s \subset [n]$

put $E(P_s) = \emptyset$ and $\Gamma_s = \emptyset$
in prune mode put $T_s = \emptyset$
in build mode put $C_s = [n]$

for $v = 1, 2, \ldots, n$ do:
let $F = \{ f \in [v+1, n]^{(s)} : \{v\} \cup f \in E(P_{s+1}),$ and $\forall \sigma \in \Gamma_s \ \sigma \not\subset f \}$
[*here F is a multiset with multiplicities inherited from $E(P_{s+1})$*]
in prune mode if $|F| \geq \zeta \tau^{r-s-1} d(v)$ and $v \in I$, add v to T_s
in build mode if $|F| \geq \zeta \tau^{r-s-1} d(v)$, remove v from C_s
if $v \in T_s$ then
add F to $E(P_s)$
for each $u \in [v+1, n]$, if $d_s(u) > \tau^{r-s} d(u)$, add $\{u\}$ to Γ_s
for each $\sigma \in [v+1, n]^{(>1)}$, if $d_s(\sigma) > 2^s \tau d_{s+1}(\sigma)$, add σ to Γ_s

The algorithm adds to P_s all s-edges which, with $v \in T_s$ as first vertex, form an edge of P_{s+1} and which do not contain (at that moment) any subset in Γ_s. The degree threshold for a vertex entering Γ_s is in terms of its degree $d(u)$ in the original graph G, whereas for a larger subset σ it is in terms of its degree in P_{s+1}; this difference is for technical reasons arising in the proof of Theorem 2.5.

Clearly, if in prune mode we construct T_s from I, then, when in build mode with the same set T_s as input, the condition $|F| \geq \zeta\tau^{r-s-1}d(v)$ can happen only if either $v \notin I$ or $v \in T_s$. Evidently, therefore, $I \subset C_s \cup T_s$, so $C_s \cup T_s$ is one option for a container for I. Another option for a container, as mentioned earlier, is $[n]$ minus the vertices in $E(P_1)$. In particular, if I is independent then $I \subset [n] - \Gamma_1$. (Strictly speaking, the set Γ_1 is a set of singletons of vertices rather than a set of vertices, but we identify Γ_1 with $\{u \in [n] : \{u\} \in \Gamma_1\}$. Furthermore Γ_1 is not output by the algorithm, but it is easily recoverable from P_1. Indeed, $E(P_1)$ is just a multiset of vertices, and $d_1(u)$ is the multiplicity of u in $E(P_1)$. By construction, $\Gamma_1 = \{u \in [n] : d_1(u) > \tau^{r-1}d(u)\}$.) So each of $C_s \cup T_s$, $1 \leq s \leq r-1$, and $[n] - \Gamma_1$ is a container for I; our aim is to ensure that at least one of these is a good container, meaning that is not close to $[n]$.

Here then is a way of viewing the operation of the algorithm. If Γ_1 is large then $[n] - \Gamma_1$ is a good container for I. If Γ_1 is not large then, since the degrees in P_1 are bounded, the average degree of P_1 must be small. But $P_r = G$, whose average degree is not small, so there must be some s for which P_{s+1} has large average degree (of order $\tau^{r-s-1}d$) but P_s has small average degree (much smaller than $\tau^{r-s}d$). Since the degrees are bounded, there must have been plenty of vertices of P_{s+1} which could have contributed edges to $E(P_s)$ but did not do so. Why did they not do so? Only because they are not in I and so not available for T_s. These are exactly the vertices which are removed from C_s: hence for this value of s, $C_s \cup T_s$ will be a good container for I.

3.2. **Properties of the construction.** We are thus lead to two important definitions.

Definition 3.1. Let G be an r-graph on vertex set $[n]$ and let $I \subset [n]$. Let $\tau, \zeta > 0$. Let $T_{r-1}, \ldots, T_1$ be the sets constructed by repeated applications of the algorithm in prune mode. Let $T_0 = I \cap \Gamma_1$. Then we define

$$T(G, I, \tau, \zeta) \,=\, (T_{r-1}, \ldots, T_1, T_0) \,\in\, \mathcal{P}^r I\,.$$

The r-tuple T is the fruit of running the algorithm in prune mode, from which the container for I will be built. As noted earlier, if I is an independent set then $T_0 = I \cap \Gamma_1 = \emptyset$. Hence the introduction of T_0 is unnecessary if we wish to find containers only for independent sets, but by introducing T_0 we can produce containers for non-independent sets too. It will turn out that T_0 is small, as desired, provided I is sparse; see §4.2.

Now comes the main definition — that of containers.

Definition 3.2. Let G be an r-graph on vertex set $[n]$ and let $T = (T_{r-1}, \ldots, T_1, T_0) \in \mathcal{P}^r[n]$. Let $\tau, \zeta > 0$. Let $C_{r-1}, \ldots, C_1$ be constructed by repeated applications of the algorithm in build mode, using $T_{r-1}, \ldots, T_1$. Let $C_0 = [n] - \Gamma_1$. The *container* $C(G, T, \tau, \zeta)$ is then

$$C(G, T, \tau, \zeta) \,=\, (C_{r-1} \cap C_{r-2} \cap \cdots \cap C_1 \cap C_0) \cup T_{r-1} \cup T_{r-2} \cdots \cup T_1 \cup T_0\,.$$

Lemma 3.3. *If $T = T(G, I, \tau, \zeta)$ then $I \subset C(G, T, \tau, \zeta)$.*

Proof. We noted earlier that $I \subset C_s \cup T_s$ for $s > 0$. Moreover, $I \subset C_0 \cup T_0$ by definition, since $C_0 = [n] - \Gamma_1$ and $T_0 = I \cap \Gamma_1$. Hence $I \subset C(G, T, \tau, \zeta)$. □

Before computing the size of the containers $C(G, T, \tau, \zeta)$ and the number of them, we make note of their online property, namely that $C(G, T, \tau, \zeta) \cap [w]$ is determined just by $T \cap [w]$.

Lemma 3.4. *Let G be an r-graph on vertex set $[n]$ and let $T \in \mathcal{P}^r[n]$. Then, for each $w \in [n]$, $C(G,T,\tau,\zeta) \cap [w] = C(G, T \cap [w], \tau, \zeta) \cap [w]$ holds.*

Proof. A little reflection on the algorithm makes the lemma clear. Suppose that $T_s \cap [w]$ is given, together with the multiset of edges $e \in E(P_{s+1})$ such that $e \cap [w] \neq \emptyset$. Then the algorithm, run with $v = 1, \dots, w$ rather than $v = 1, \dots, n$, will correctly produce the edges $e \in E(P_s)$ such that $e \cap [w] \neq \emptyset$ and will find all $\sigma \in \Gamma_s$ such that $\sigma \cap [w] \neq \emptyset$ (together with some other members of $E(P_s)$ and Γ_s that we shall not need). Moreover, when running in build mode, this restricted version of the algorithm will correctly determine $C_s \cap [w]$. Therefore, by running the restricted algorithm for $s = r-1, r-2, \dots, 1$, we can determine $C_{r-1} \cap [w], \dots, C_1 \cap [w]$. Finally, since we have determined $\Gamma_1 \cap [w]$ we can find $C_0 \cap [w]$, and because we are given $T_0 \cap [w]$ this means we know $C \cap [w]$. $\square$

4. Container calculations

In this section we estimate the measure of the tuples $T(G, I, \tau, \zeta)$ and of the containers $C(G, T, \tau, \zeta)$, thereby proving Theorem 2.5.

4.1. Degrees and co-degrees.

Before making these estimates we need information on how large the degrees can be in P_s. The intention behind the set Γ_s is to prevent degrees being much larger than the target degrees, namely $\tau^{r-s} d(u)$ for the vertex u; after the degree of u attains this level, no further edges containing u are added to P_s. However, when a vertex u enters Γ_s, it does so because some multiset F has been added to $E(P_s)$. Since F can include many edges that contain u, the degree $d_s(u)$ can increase significantly in one step, from an initial value at most the target value $\tau^{r-s} d(u)$ to something much larger. The extent of this problem depends ultimately on the way the edges of G overlap each other.

The reason Γ_s is defined the way it is in the algorithm, is to keep control of the degree problem without increasing τ more than is necessary. This can be expressed succinctly in terms of the co-degree function $\delta(G, \tau)$ introduced in §2.1.

First we need a small calculation.

Lemma 4.1. *For $2 \le s \le r$ and $2 \le j \le s$, let $a_s^{(j)}$ be given by the equations $a_r^{(j)} = \delta_j$ and $a_s^{(j)} = 2^s a_{s+1}^{(j)} + a_{s+1}^{(j+1)}$ for $s < r$, where δ_j was defined in Definition 2.2. Then $a_s^{(2)} \le 4^{2-s} \delta(G, \tau)$ holds for $s \ge 2$.*

Proof. Since $a_s^{(2)} \ge 2^s a_{s+1}^{(2)} \ge 4 a_{s+1}^{(2)}$, it is enough to prove that $a_2^{(2)} \le \delta(G, \tau)$. Now by dint of the definition it is clear that $a_s^{(j)}$ is a linear combination of the numbers $\delta_{j+\ell}$, $\ell \ge 0$. We claim that the coefficient of $\delta_{j+\ell}$ in $a_s^{(j)}$ is at most $2^{\binom{r}{2} - \binom{s+\ell}{2} + \ell}$. This is certainly true if $s = r$, since the only positive coefficient is that of δ_j (i.e. $\ell = 0$). For $s < r$ we may prove the claim on the assumption that it is true for $s+1$. If $\ell = 0$ then the coefficient of $\delta_{j+\ell}$ in $a_{s+1}^{(j+1)}$ is zero, and the claim follows because $2^{\binom{r}{2} - \binom{s+\ell}{2} + \ell} = 2^s 2^{\binom{r}{2} - \binom{s+1+\ell}{2} + \ell}$. If $\ell \ge 1$ we have

$$2^s \, 2^{\binom{r}{2} - \binom{s+1+\ell}{2} + \ell} + 2^{\binom{r}{2} - \binom{s+\ell}{2} + \ell - 1} = 2^{\binom{r}{2} - \binom{s+\ell}{2} + \ell} \left[2^{-\ell} + 2^{-1} \right] \le 2^{\binom{r}{2} - \binom{s+\ell}{2} + \ell}$$

and the claim follows in this case too. Hence the claim always holds, and so

$$a_2^{(2)} \;\le\; 2^{\binom{r}{2}} \sum_{\ell=0}^{r-2} 2^{-\binom{\ell+2}{2}+\ell} \delta_{2+\ell} \;=\; 2^{\binom{r}{2}-1} \sum_{j=2}^{r} 2^{-\binom{j-1}{2}} \delta_j \;=\; \delta(G,\tau)\,,$$

by definition of $\delta(G,\tau)$. □

Here is the main lemma about degrees in P_s, explaining the role of the co-degree function $\delta(G,\tau)$.

Lemma 4.2. *Let G be an r-graph on vertex set $[n]$ with average degree d. Let $P_r = G$ and let $P_{s-1},\dots,P_1$ be the multigraphs constructed by successive applications of the algorithm. Then*

$$\sum_{u\in U} d_s(u) \;\le\; \tau^{r-s}\, nd\,(\,\mu(U) + 4^{1-s}\delta(G,\tau)\,)$$

holds for all subsets $U \subset [n]$ and for $1 \le s \le r$.

Proof. By analogy with Definition 2.2 we define

$$d_s^{(j)}(u) \;=\; \max\{\, d_s(\sigma) \,:\, u\in\sigma\in[n]^{(j)}\,\}\,,$$

for $j \ge 2$, where here it is the final values of these quantities that are used — that is, we measure these quantities in the output multigraph P_s.

When $s = r$ the lemma is true by definition of $\mu(U)$, so from now on we assume $s \le r-1$. Suppose that $\sigma \in [n]^{(j)}$. If $\sigma \notin \Gamma_s$ then $d_s(\sigma) \le 2^s\tau d_{s+1}(\sigma)$. If $\sigma \in \Gamma_s$ then σ was added to Γ_s after some vertex $v \in T_s$ was inspected and F was added to $E(P_s)$. Before this took place, $d_s(\sigma) \le 2^s\tau d_{s+1}(\sigma)$ held; since the number of edges of F containing σ was at most $d_{s+1}(\sigma\cup\{v\})$, the final value of $d_s(\sigma)$ satisfies $d_s(\sigma) \le 2^s\tau d_{s+1}(\sigma) + d_{s+1}(\sigma\cup\{v\})$. This inequality holds for all $\sigma \in [n]^{(j)}$.

Let $u \in [n]$; then $d_s^{(j)}(u) = d_s(\sigma)$ for some $\sigma \in [n]^{(j)}$, so

$$d_s^{(j)}(u) \le 2^s\tau d_{s+1}(\sigma) + d_{s+1}(\sigma\cup\{v\}) \le 2^s\tau d_{s+1}^{(j)}(u) + d_{s+1}^{(j+1)}(u)\,. \tag{1}$$

We claim that

$$\sum_{u\in[n]} d_s^{(j)}(u) \le a_s^{(j)}\tau^{r-s+j-1} nd\,,$$

where $a_s^{(j)}$ was defined in Lemma 4.1. Indeed, for $s = r$ the claim (with equality) is just the definition of δ_j, and for $s \le r-1$ it follows immediately by induction (on $r-s$) from inequality (1) and the definition of $a_s^{(j)}$. Hence, for $s \ge 1$, we have by Lemma 4.1

$$\sum_{u\in[n]} d_{s+1}^{(2)}(u) \le 4^{1-s}\tau^{r-s} nd\,\delta(G,\tau)\,. \tag{2}$$

Now let $u \in U$. If $u \notin \Gamma_s$ then $d_s(u) \le \tau^{r-s}d(u)$. If $u \in \Gamma_s$ then u was added to Γ_s after some vertex $v \in T_s$ was inspected and F was added to $E(P_s)$. Since F has at most $d_{s+1}(\{u,v\})$ edges containing u, the degree of u is at most $\tau^{r-s}d(u) + d_{s+1}(\{u,v\})$. Now $d_{s+1}(\{u,v\}) \le d_{s+1}^{(2)}(u)$ so, using (2), we have

$$\sum_{u\in U} d_s(u) \;\le\; \sum_{u\in U} \tau^{r-s}d(u) + d_{s+1}^{(2)}(u) \;\le\; \tau^{r-s}nd\mu(U) + 4^{1-s}\tau^{r-s}nd\,\delta(G,\tau)\,,$$

which establishes the lemma. □

4.2. **The measure of the sets T_s.** We now estimate the measures of the sets T_s.

Lemma 4.3. *Let $I \subset [n]$ and $T = T(G, I, \tau, \zeta) = (T_{r-1}, \ldots, T_1, T_0)$. Then $\mu(T_s) \leq (\tau/\zeta)(1 + \delta(G, \tau))$ for $1 \leq s \leq r - 1$.*

Proof. The set T_s is output when the algorithm is run in prune mode. During the run of the algorithm, each vertex v which enters T_s contributes a set F of at least $\zeta\tau^{r-s-1}d(v)$ edges to $E(P_s)$. Therefore, writing d for the average degree of G, Lemma 4.2 yields

$$\begin{aligned}\zeta\tau^{r-s-1}nd\mu(T_s) = \sum_{v \in T_s} \zeta\tau^{r-s-1}d(v) \leq e(P_s) &\leq \sum_{u \in [n]} d_s(u) \\ &\leq \tau^{r-s}nd(1 + 4^{1-s}\delta(G, \tau))\end{aligned}$$

and this proves the lemma. □

The set T_0 needs a different argument. As noted before, $T_0 = \emptyset$ if I is independent.

It turns out that we can allow I to be b-degenerate, for fixed b, at essentially no cost compared with I being independent, though to handle this case we shall need to make sure the degree sequence of G is decreasing. On the other hand, requiring only that $e(G[I]) \leq bn$ will incur a small cost, which is that τ must increase typically from around $d^{-1/(r-1)}$ to around $d^{-1/r}$ in order to keep T_0 small (d being the average degree of G).

Lemma 4.4. *Let G be an r-graph on vertex set $[n]$ with average degree d. Let $I \subset [n]$ and $T = T(G, I, \tau, \zeta) = (T_{r-1}, \ldots, T_1, T_0)$. If $e(G[I]) \leq bn$ then $\mu(T_0) \leq \tau^{1-r}d^{-1}b$. If $G[I]$ is b-degenerate and $d(v) \geq \tau^{1-r}rb$ for $v \in [m]$, then $\mu(T_0 \cap [m]) \leq (\tau/\zeta)(1 + \delta(G, \tau))$.*

Proof. Recall that $T_0 = I \cap \Gamma_1$, so $d_1(v) > \tau^{r-1}d(v)$ for each $v \in T_0$. Let $J = T_{r-1} \cup \cdots \cup T_1 \cup T_0 \subset I$. Recall that distinct 1-edges $\{v\} \in \Gamma_1$ correspond to distinct r-edges $\{v_{r-1}, \ldots, v_1, v\} \subset I$ with $v_{r-1} < \cdots < v_1 < v$ and $v_s \in T_s$, so these edges lie in $G[J]$. It follows that $\tau^{r-1}nd\mu(T_0) \leq \sum_{v \in T_0} d_1(v) \leq e(G[J])$.

If $e(G[I]) \leq bn$ then we simply have $\tau^{r-1}nd\mu(T_0) \leq e(G[J]) \leq e(G[I]) \leq bn$, so $\mu(T_0) \leq \tau^{1-r}d^{-1}b$.

If $G[I]$ is b-degenerate then we consider J^* instead, where S^* denotes $S \cap [m]$ for any $S \subset [n]$. If $v \in T_0^*$ then each of the above-mentioned edges $\{v_{r-1}, \ldots, v_1, v\}$ lies within J^*, since $v_{r-1} < \cdots < v_1 < v$. Therefore $\tau^{r-1}nd\mu(T_0^*) \leq e(G[J^*]) \leq b|J^*|$, by b-degeneracy. Then, by definition of $[m]$,

$$\tau^{r-1}nd\mu(T_0^*) \leq b|J^*| \;\leq\; \frac{b}{\tau^{1-r}rb} \sum_{u \in J^*} d(u) \;=\; \frac{\tau^{r-1}}{r}nd\mu(J^*)\,.$$

Therefore $r\mu(T_0^*) \leq \mu(J^*) \leq \mu(T_{r-1}^*) + \cdots + \mu(T_0^*)$, and so $(r-1)\mu(T_0^*) \leq \mu(T_{r-1}^*) + \cdots + \mu(T_1^*) \leq \mu(T_{r-1}) + \cdots + \mu(T_1)$, from which the result follows via Lemma 4.3. □

4.3. **The measure of the container $C(G, T, \tau, \zeta)$.** We now prove the crucial fact that the measure of the container $C(G, T, \tau, \zeta)$ is bounded above by some constant less than one. This can be established with a fairly simple argument, but just a little more care yields a bound close to $1 - 1/r!$, and this is best possible, as shown in §7.

Lemma 4.5. *Let $T = (T_{r-1}, \ldots, T_0) \in \mathcal{P}^r[n]$. Then*

$$\mu(C(G, T, \tau, \zeta)) \;\leq\; 1 - \frac{1}{r!} + \frac{11}{4}\zeta + \frac{1}{4}\delta(G, \tau) + \sum_{s=0}^{r-1} \mu(T_s)\,.$$

Proof. Recall from Definition 3.2 that $C_0 = [n] - \Gamma_1$ and that $C_{r-1}, \ldots, C_1$ are constructed by the algorithm in build mode. Let $C = C_{r-1} \cap \cdots \cap C_0$. We define

$$\begin{aligned}
D_1 &= [n] \setminus C \\
D_2 &= \{v \in [n] : \{v\} \in \Gamma_2, v \notin D_1\} \\
D_3 &= \{v \in [n] : \{v\} \in \Gamma_3, v \notin (D_1 \cup D_2)\} \\
&\ \vdots \\
D_{r-1} &= \{v \in [n] : \{v\} \in \Gamma_{r-1}, v \notin (D_1 \cup \cdots \cup D_{r-2})\} \\
D_r &= [n] \setminus (D_1 \cup \cdots \cup D_{r-1})\,.
\end{aligned}$$

Evidently, $D_1, \ldots, D_r$ form a partition of $[n]$. Since $C(G, T, \tau, \zeta) = C \cup T_{r-1} \cdots \cup T_0$, it is enough to prove that $\mu(D_1) \geq 1/r! - 11\zeta/4 - \delta(G, \tau)/4$.

It is convenient to define $D_{<s} = D_1 \cup \cdots \cup D_{s-1}$, $D_{\leq s} = D_s \cup D_{<s}$ and $D_{>s} = [n] \setminus D_{\leq s}$. For $s \geq 2$ we also need certain subsets of the edges of P_s:

$$\begin{aligned}
X_s &= \{\, f \in E(P_s) \,:\, |f \cap D_{<s}| \geq 1,\ \ |f \cap D_{>s}| \geq 2 \,\} \\
Y_s &= \{\, f \in E(P_s) \,:\, f \subset D_{\geq s} \,\} \\
Z_s &= \{\, f \in Y_s \,:\, \sigma \subset f \text{ for some } \sigma \in \Gamma_{s-1}, |\sigma| \geq 2 \,\}\,.
\end{aligned}$$

We further define the numbers x_s, y_s, z_s by $|X_s| = x_s \tau^{r-s} nd$, $|Y_s| = y_s \tau^{r-s} nd$ and $|Z_s| = z_s \tau^{r-s} nd$, where d is the average degree of G. Observe that $X_2 = Z_2 = \emptyset$, that is, $x_2 = z_2 = 0$.

Note that $X_s \cap Y_s = \emptyset$ and each member of $E(P_s) \setminus Y_s$ meets $D_{<s}$. So

$$\begin{aligned}
\sum_{v \in D_s} d_s(v) &= \sum_{f \in E(P_s)} |f \cap D_s| \\
&= \sum_{f \in E(P_s) \setminus (X_s \cup Y_s)} |f \cap D_s| \,+\, \sum_{f \in X_s} |f \cap D_s| \,+\, \sum_{f \in Y_s} |f \cap D_s| \\
&\leq (s-1)|E(P_s) \setminus (X_s \cup Y_s)| + (s-3)|X_s| + s|Y_s| \\
&= (s-1)|E(P_s) \setminus Y_s| - 2|X_s| + s|Y_s| \\
&\leq (s-1) \sum_{v \in D_{<s}} d_s(v) - 2|X_s| + s|Y_s| \\
&\leq (s-1)\tau^{r-s}\, nd\,(\,\mu(D_{<s}) + 4^{1-s}\delta(G, \tau)) - 2|X_s| + s|Y_s|\,,
\end{aligned}$$

where the last line employs Lemma 4.2.

Now $d_s(v) \geq \tau^{r-s} d(v)$ for all $v \in D_s$: for $s < r$ this is because $\{v\} \in \Gamma_s$, and for $s = r$ it holds trivially. Hence $\tau^{r-s} nd\mu(D_s) \leq \sum_{v \in D_s} d_s(v)$, so we obtain $\mu(D_s) \leq (s-1)(\mu(D_{<s}) + 4^{1-s}\delta(G, \tau)) - 2x_s + sy_s$. Adding $\mu(D_{<s}) = \mu(D_{\leq s-1})$ to each side gives

$$\mu(D_{\leq s}) \leq s\mu(D_{\leq s-1}) - 2x_s + sy_s + (s-1)4^{1-s}\delta(G, \tau)$$

for each $s \geq 2$. Multiplying this inequality by $1/s!$ and summing over $s = 2, \ldots, r$, noting that $\mu(D_{\leq r}) = 1$, $D_{\leq 1} = D_1$ and $x_2 = 0$, we obtain

$$\frac{1}{r!} \,\leq\, \mu(D_1) - 2\sum_{s \geq 3} \frac{x_s}{s!} + \sum_{s \geq 2} \frac{y_s}{(s-1)!} + \frac{1}{4}\delta(G, \tau)\,, \tag{3}$$

where we used $\sum_{s \geq 2} 4^{1-s}(s-1)/s! < 1/4$.

Suppose $s \geq 2$ and $f \in Y_s \setminus Z_s$. If $\sigma \in \Gamma_{s-1}$ and $\sigma \subset f$ then $|\sigma| = 1$, say $\sigma = \{u\}$. But $\{u\} \in \Gamma_{s-1}$ implies $u \in D_{<s}$ by definition of D_{s-1}, which contradicts $f \in Y_s$. Thus f contains no member of Γ_{s-1}. Let v be the first vertex of f. Then $v \notin D_1$ since $s \geq 2$, so $v \in C \subset C_{s-1}$. By the construction of C_{s-1}, v is the first vertex of fewer than $\zeta\tau^{r-s}d(v)$ edges of P_s that contain no member of Γ_{s-1}, so it is the first vertex of fewer than $\zeta\tau^{r-s}d(v)$ edges in $Y_s \setminus Z_s$. Therefore $|Y_s| - |Z_s| \leq \sum_{v \in D_{\geq s}} \zeta\tau^{r-s}d(v) = \zeta\tau^{r-s}nd\mu(D_{\geq s})$. Hence $y_s - z_s \leq \zeta\mu(D_{\geq s}) \leq \zeta$. In particular $y_2 \leq \zeta$, because $z_2 = 0$.

Now let $s \geq 3$ and put $S = \{\sigma \in \Gamma_{s-1} : |\sigma| \geq 2, \sigma \subset D_{\geq s}\}$. By definition of Z_s, each member of Z_s contains a member of S. Therefore

$$\begin{aligned} z_s\tau^{r-s}nd \;=\; |Z_s| &\leq \sum_{\sigma \in S} d_s(\sigma) \\ &\leq \sum_{\sigma \in S} \frac{1}{\tau 2^{s-1}} d_{s-1}(\sigma) \quad \text{by definition of } \Gamma_{s-1} \\ &\leq \frac{1}{\tau} \left|\{f \in E(P_{s-1}) : \sigma \subset f \text{ for some } \sigma \in S\}\right| \\ &\leq \frac{1}{\tau} |X_{s-1} \cup Y_{s-1}| = (x_{s-1} + y_{s-1})\tau^{r-s}nd\,. \end{aligned}$$

Hence $z_s \leq x_{s-1} + y_{s-1}$ for $s \geq 3$. Since $y_s \leq z_s + \zeta$ this means $y_s \leq x_{s-1} + y_{s-1} + \zeta$; by repeating and applying both $x_2 = 0$ and $y_2 \leq \zeta$, this yields $y_s \leq x_{s-1} + x_{s-2} + \cdots + x_3 + (s-1)\zeta$ for $s \geq 3$. The inequality holds for $s = 2$ also. Substituting this inequality into inequality (3) we obtain

$$\frac{1}{r!} \;\leq\; \mu(D_1) + \sum_{s \geq 3} x_s \left(-\frac{2}{s!} + \sum_{j=s}^{r-1} \frac{1}{j!} \right) + \zeta \sum_{s \geq 2} \frac{1}{(s-2)!} + \frac{1}{4}\delta(G,\tau)\,.$$

The coeffient of x_s is negative, and so $1/r! \leq \mu(D_1) + 11\zeta/4 + \delta(G,\tau)/4$, which is what we needed to prove. □

4.4. **Proof of Theorem 2.5.** The simple and obvious choices for T and C in Theorem 2.5 are $T = T(G, I, \tau, \zeta)$ and $C(T) = C(G, T, \tau, \zeta)$. Indeed this choice works perfectly well for parts (a), (b) and (d) in the (most important) case when I is independent, and also when $e(G[I])$ is bounded. However for the case when $G[I]$ is b-degenerate a different choice is required. We need containers determined by small sets of vertices none of which has low degree.

The online property offers, as a by-product, a convenient way to achieve this aim, which furthermore has the side benefit of yielding (c). We label the vertex set so that the degrees are decreasing, and hence the vertices of small degree are specified by some terminal segment of $[n]$. We then run the algorithm just on the initial segment $[m]$ of vertices of high degree, and add the trailing segment $[m+1, n]$ to the container. These modified containers, determined by subsets of $[m]$, still have the online property.

Lemma 4.6. *Let G be an r-graph on vertex set $[n]$ and let $m \in [n]$. Let $I \subset [n]$, let $S = T(G, I, \tau, \zeta)$ and let $T = S \cap [m]$. Put*

$$C \;=\; C(G, T, \tau, \zeta) \cup [m+1, n]\,.$$

Then C is a container, determined by T, with the online property: that is, $I \subset C$, and $C \cap [w]$ is determined by $T \cap [w]$ for all $w \in [n]$.

Proof. The lemma is a simple consequence of Lemma 3.4. □

Proof of Theorem 2.5. Notice that the theorem is trivial if $\zeta \geq 1/4r!$, since in that case the function $C(T) = [n]$ works, with $T = (\emptyset, \ldots, \emptyset)$ representing all I. Recall too from §2.1 that $\delta(G, \tau) \to 0$ as $\tau \to \infty$. Hence the condition $\delta(G, \tau) \leq \zeta$ is satisfiable by making τ large enough, although if $\tau \geq \zeta/2r$ the theorem is again trivial.

Let d be the average degree of G. We define $m \in [n]$ by $[m] = \{v \in [n] : d(v) \geq \zeta d\}$. Let $S = T(G, I, \tau, \zeta) \in \mathcal{P}^r[n]$ and write $S = (S_{r-1}, \ldots, S_0)$. We then put $T = S \cap [m]$, so $T = (T_{r-1}, \ldots, T_0)$ where $T_s = S_s \cap [m]$, $0 \leq s \leq r-1$. Finally, we define $C(T) = C(G, T, \tau, \zeta) \cup [m+1, n]$. We shall show that the theorem holds with this choice of T and C.

First, note that (a) and the online property both hold by virtue of Lemma 4.6. To show that (b) holds, we apply Lemmas 4.3 and 4.4; notice that the sets denoted by T_s in those lemmas are here denoted by S_s. For $s \geq 1$ we obtain $\mu(T_s) \leq \mu(S_s) \leq 2\tau/\zeta$ by Lemma 4.3 because, as remarked, may assume $\delta(G, \tau) \leq \zeta < 1/4r!$. As for T_0, if I is independent then $T_0 = \emptyset$. If $G[I]$ is b-degenerate then Lemma 4.4 shows $\mu(T_0) = \mu(S_0 \cap [m]) \leq 2\tau/\zeta$, because $d(v) \geq \zeta d \geq \tau^{1-r} rb$ for $v \in [m]$. If $e(G[I]) \leq bn$ then Lemma 4.4 shows $\mu(T_0) \leq \mu(S_0) \leq \tau^{1-r} d^{-1} b \leq 2\tau/\zeta$. Hence (b) holds in every case.

Since $T_s \subset [m]$ for each s, we have $|T_s|\zeta d \leq \sum_{v \in T_s} d(v) = nd\mu(T_s)$, so (c) is a consequence of (b).

Now $\mu(C) \leq \mu(C(G, T, \tau, \zeta)) + \mu([m+1, n])$. Lemma 4.5 applied to $C(G, T, \tau, \zeta)$ shows $\mu(C(G, T, \tau, \zeta)) \leq 1 - 1/r! + 3\zeta + 2r\tau/\zeta$, because (b) holds. Finally, note that $\mu([m+1, n]) < \zeta$ by definition of $[m]$, so (d) holds and we are done. □

5. Tight containers

The method of Theorem 2.5 will not, in general, give containers of measure less than $1 - 1/r!$ (we shall give examples to show this in §7). On a more positive note, it is possible via iteration to obtain smaller, almost optimal, containers, at the cost of sacrificing the online property. Some applications, most notably list colouring, depend crucially on the online property, but for many applications it is unnecessary, and in such cases it is generally well worthwhile using smaller, iterated, containers.

Let G be an r-graph of order n and average degree d, and let I be an independent set in G. Suppose we have a container C for I with $\mu(C) \leq 1 - c$, such as that supplied by Theorem 2.5 where c is close to $1/r!$. It follows that the induced subgraph $G[C]$ cannot be dense; to be precise,

$$e(G[C]) \leq (1/r) \sum_{v \in C} d(v) = (1/r) nd\mu(C) = \mu(C) e(G) \leq (1-c) e(G)\,. \tag{4}$$

Theorem 2.5 can now be re-applied, this time to the hypergraph $G[C]$, to get a container $C' \subset C$, with $e(G[C']) \leq (1-c) e(G[C]) \leq (1-c)^2 e(G)$. Repeating this operation often enough will give a container with $o(e(G))$ edges.

The reason that the online property of Theorem 2.5 is lost following iteration is that, in order to compute, say, $C' \cap [v]$, it is necessary to know the whole of $G[C]$, for which it is necessary to know the whole of C, not just $C \cap [v]$; hence $C' \cap [v]$ cannot be computed online.

Another drawback of iteration is the increase in the number of sets T_i needed to specify the eventual container, leading to a greater number of possible containers. Usually in practice this turns out to be unimportant. Here is a simple lemma to help count the number of containers being generated.

Lemma 5.1. *There are at most* $\exp\{s\theta n(1+\log(1/\theta))\}$ *s-tuples of subsets* $T_1, \ldots, T_s \subset [n]$ *with* $|T_1| + \ldots + |T_s| \le s\theta n$*, where* $\theta \le 1$.

Proof. Let there be N_j such s-tuples with $|T_1| + \ldots + |T_s| = j$. We wish to bound $N = N_0 + N_1 + \ldots + N_{\lfloor s\theta n\rfloor}$. Plainly, the ordinary generating function $N_0 + N_1 x + N_2 x^2 + \ldots$ for the sequence (N_j) is equal to $((1 + x)^n)^s$. Therefore, since $\theta \le 1$, we have $\theta^{s\theta n} N \le (1 + \theta)^{ns} \le e^{s\theta n}$. □

The next theorem is a version of Theorem 2.5 more suited to iteration.

Theorem 5.2. *Let G be an r-graph on vertex set $[n]$. Suppose that $\delta(G, \tau) \le 1/12r!$ and that $\tau \le 1/144r!^2 r$. Then there exists a collection $\mathcal{C} \subset \mathcal{P}[n]$ such that*

(a) *for every independent set I there exists $T = (T_{r-1}, \ldots, T_0) \in \mathcal{P}^r(I)$ with $I \subset C(T) \in \mathcal{C}$ and $|T_i| \le 288r!^2\tau n$,*
(b) $\log|\mathcal{C}| \le 288rr!^2 n\tau\log(1/\tau)$*, and*
(c) $e(G[C]) \le (1 - 1/2r!)e(G)$ *for all $C \in \mathcal{C}$.*

Moreover, (a) holds for all sets $I \subset [n]$ for which either $G[I]$ is $\lfloor\tau^{r-1}e(G)/12r!n\rfloor$-degenerate or $e(G[I]) \le 24r!r\tau^r e(G)$.

Proof. We apply Theorem 2.5 to G with $\zeta = 1/12r!$ (we may assume the degree order of G is decreasing). For each set I we have $T = (T_{r-1}, \ldots, T_0)$ and a container $C(T)$ satisfying properties (a)–(d) of that theorem. Take $\mathcal{C}$ to be the collection of all such C. Since $\tau \le \zeta^2/r$, we have $2r\tau/\zeta \le 2\zeta$, so $\mu(C) \le 1 - 1/r! + 6\zeta = 1 - 1/2r!$. It follows from inequality (4) that $e(G[C]) \le (1 - 1/2r!)e(G)$.

Hence (a) and (c) of the present theorem are satisfied and it remains to check (b). Theorem 2.5 tells us that each container C is specified by sets $T_0, \ldots, T_{r-1}$ each of size at most θn, where $\theta = 2\tau/\zeta^2 = 288r!^2\tau \le 2/r \le 1$. By Lemma 5.1 we have

$$\log|\mathcal{C}| \le r\theta n(1 + \log(1/\theta)) \le r\theta n\log(1/\tau) = 288rr!^2 n\tau\log(1/\tau)\,,$$

establishing (b) and completing the proof. □

Theorem 5.2 is easy to apply iteratively. Though a similar theorem can be derived using the proof method of Theorem 3.1 of [47] in place of Theorem 2.5, which actually gives a better bound in (c), with a polynomial in r in place of $r!$, the bound in (b) is worse, with $\tau^{1/2}$ in place of τ. We remark, though, that it is not possible to iterate Theorem 3.1 of [47] directly, because it applies only to regular r-graphs, or, more precisely, to r-graphs with no large sparse subset. Even if G itself is regular, $G[C]$ might be far from regular.

The next theorem is the result of applying Theorem 5.2 repeatedly to end up with a collection of sparse containers. Its appearance is rather technical but it is in a form which can be readily applied.

Theorem 5.3. *Let G be an r-graph on vertex set $[n]$. Let $e_0 \le e(G)$. Suppose that, for each $U \subset [n]$ with $e(G[U]) \ge e_0$, the function $\tau(U)$ satisfies both $\delta(G[U], \tau) \le 1/12r!$ and $\tau(U) \le 1/144r!^2 r$. For $e_0 \le m \le e(G)$ define*

$$f(m) = \max\{ -|U|\tau(U)\log\tau(U) : U \subset [n],\, e(G[U]) \ge m\}$$
$$\tau^* = \max\{ \tau(U) : U \subset [n],\, e(G[U]) \ge e_0\}$$

Let $k = \log(e_0/e(G))/\log(1 - 1/2r!)$. Then there exists a collection $\mathcal{C} \subset \mathcal{P}[n]$ such that

(a) *for every independent set I there exists $T = (T_1, \dots, T_s) \in \mathcal{P}^s(I)$ with $I \subset C(T) \in \mathcal{C}$, $|T_i| \le 288r!^2\tau^* n$ and $s \le (k+1)r$,*
(b) *$e(G[C]) \le e_0$ for all $C \in \mathcal{C}$,*
(c) *$\log|\mathcal{C}| \le 288rr!^2 \sum_{0\le i<k} f(e_0/(1-1/2r!)^i)$.*

Moreover, (a) holds for all $I \subset [n]$ for which either $G[I]$ is $\lfloor \tau(U)^{r-1}e(G[U])/12rr!|U|\rfloor$-degenerate or $e(G[I]) \le 24r!r\tau(U)^r e(G[U])$, for all $U \subset [n]$ with $e(G[U]) \ge e_0$.

Proof. We will show that for all t with $e_0 \le t < e(G)/(1 - 1/2r!)$, there exists a collection $\mathcal{C}_t \subset \mathcal{P}[n]$ satisfying conditions (a)–(c), where the constant e_0 has been replaced by t in (a)–(c), and k is replaced by $k(t) = \log(t/e(G))/\log(1 - 1/2r!)$.

When $t \ge e(G)$, we may take $\mathcal{C}_t = \{[n]\}$. Otherwise, suppose $t < e(G)$. It is enough to show that $\mathcal{C}_t$ exists provided $\mathcal{D} = \mathcal{C}_{t/(1-1/2r!)}$ exists. Each $D \in \mathcal{D}$ is specified by a tuple $T' = (T_1, \dots, T_{s'})$ with $s' \le (k(t/(1-1/2r!))+1)r = k(t)r$. If $e(G[D]) \le t$, let $\mathcal{C}_t(D) = \{D\}$. Otherwise, apply Theorem 5.2 with $\tau = \tau(D) \le \tau^*$ to the r-graph $G[D]$, and let $\mathcal{C}_t(D)$ be the collection of containers given by the theorem. Then put $\mathcal{C}_t = \bigcup_{D\in\mathcal{D}} \mathcal{C}_t(D)$.

If $C \in \mathcal{C}(D)$ then C is specified completely by T', together with the r-tuple appearing in condition (a) of Theorem 5.2 if the theorem was applied. Hence C is specified completely by a tuple of size at most $(k(t)+1)r$, so satisfying condition (a). If $D \in \mathcal{D}$ then either $e(G[D]) \le t$ in which case $|\mathcal{C}_t(D)| = 1$, or $e(G[D]) > t$ in which case

$$\log|\mathcal{C}_t(D)| \le 288rr!^2|D|\tau(D)\log(1/\tau(D)) \le 288rr!^2 f(t).$$

Hence

$$\log|\mathcal{C}_t| \le \log|\mathcal{D}| + 288rr!^2 f(t) \le 288rr!^2 \sum_{0\le i<k(t)} f(t/(1-1/2r!)^i).$$

Finally for $C \in \mathcal{C}_t(D)$, note that $e(G[C]) \le t$, since if $e(G[D]) > t$ then by condition (c) of Theorem 5.2 $e(G[C]) \le (1 - 1/2r!)e(G[D]) \le t$. □

For certain applications the technical detail of Theorem 5.3 is not needed; what is required is a simple statement that a few iterations will produce a container with a negligible proportion of the original edges. Such a statement was presented earlier as Corollary 2.7.

Proof of Corollary 2.7. Let $e_0 = \epsilon e(G)$. Observe that for $U \subset [n]$, if $e(U) \ge \epsilon e(G)$ then $\delta(G[U], \tau) \le \delta(G, \tau)/\epsilon \le 1/12r!$. Therefore we may apply Theorem 5.3 to the graph G with $e_0 = \epsilon e(G)$ and $\tau(U) = \tau$ for all U. Then $\tau^* = \tau$ and $f(m) = n\tau\log(1/\tau)$. Hence we obtain a collection $\mathcal{C}$ satisfying conditions (a) and (b) of the corollary, and

$$\log|\mathcal{C}| \le 288r!^2 r\left(1 + \frac{\log\epsilon}{\log(1 - 1/2r!)}\right) n\tau\log(1/\tau),$$

giving condition (c). □

6. Uniformly bounded containers

Theorem 2.5 provides containers of bounded degree measure, and §5 provides containers with few edges inside. For some purposes, though, what is required is containers of bounded size, that is, of bounded uniform measure.

For regular hypergraphs the results of §5 can be used directly. If G is d-regular and $C \subset [n]$ is such that $e(G[C]) = o(nd)$ then $|C| \le (1 - 1/r + o(1))n$. However for non-regular hypergraphs we get no such information. Indeed, it is perfectly possible to have $\mu(C)$ and $e(G[C])$ both small whilst $|C|$ is close to n. Hence it is not possible to guarantee a container of bounded uniform measure.

Instead, what we can show is that there is some initial interval $[v] \subset [n]$ such that $|C \cap [v]|$ is bounded. The basic lemma which translates information about μ-measure into information about uniform measure is the following one. In the lemma, S is a multiset, so $\mu(S)$, $|S \cap [v]|$ and so on have their natural interpretations counting with multiplicities.

Lemma 6.1. *Let $\mu : [n] \to \mathbb{R}$ be a measure with $\mu(1) \ge \mu(2) \ge \cdots \ge \mu(n)$, and let $S \subset [n]$ be a multiset. Then*

$$\alpha\, \mu\left(\{\, v \in [n] \,:\, |S \cap [v]| \ge \alpha v \,\}\right) \,\le\, \mu(S)$$

holds for all $\alpha \ge 0$.

Proof. Let $W = \{v : |S \cap [v]| \ge \alpha v\}$. We must show $\alpha\mu(W) \le \mu(S)$. Let $W = \{w_1, \dots, w_k\}$ where $k = |W|$ and $w_1 < w_2 < \ldots < w_k$. Define the numbers $s_1, \dots, s_k$ by $s_1 = |S \cap [w_1]|$ and $s_i = |S \cap [w_{i-1}+1, w_i]|$ for $i \ge 2$. Then we have $\mu(S \cap [w_{i-1}+1, w_i]) \ge s_i \mu(w_i)$, because $\mu(1) \ge \mu(2) \ge \cdots \ge \mu(n)$. Therefore

$$\begin{aligned}
\mu(S) &\ge \mu(S \cap [w_1]) + \mu(S \cap [w_1+1, w_2]) + \cdots + \mu(S \cap [w_{k-1}+1, w_k]) \\
&\ge s_1\mu(w_1) + s_2\mu(w_2) + \cdots + s_k\mu(w_k) \\
&= \sum_{i=1}^{k} \alpha\mu(w_i) + (s_1 + \cdots + s_i - \alpha i)(\mu(w_i) - \mu(w_{i+1}))\,, \\
&= \alpha\mu(W) + \sum_{i=1}^{k} (s_1 + \cdots + s_i - \alpha i)(\mu(w_i) - \mu(w_{i+1}))\,, \qquad (5)
\end{aligned}$$

where $\mu(w_{k+1})$ is defined to be zero. Now $|S \cap [w_i]| = s_1 + \cdots + s_i$ holds for $1 \le i \le k$, and so $s_1 + \cdots + s_i \ge \alpha w_i$, because $w_i \in W$. In particular, $s_1 + \cdots + s_i \ge \alpha i$, since $w_i \ge i$. Moreover μ is a decreasing function, so each summand in (5) is non-negative, and the lemma follows. □

In fact we shall need not just that $|C \cap [v]|$ is bounded for a single container C but that the average $\sum_{i=1}^{t} |C_i \cap [v]|$ is bounded for a collection $C_1, \dots, C_t$. The next lemma prepares the way.

Lemma 6.2. *Let μ be a probability measure on $[n]$ with $\mu(1) \ge \mu(2) \ge \cdots \ge \mu(n) \ge 0$. Let $T_1, \dots, T_s, C_1, \dots, C_t$ be subsets of $[n]$, with $\mu(T_i) \le \lambda$ for $1 \le i \le s$ and $\mu(C_j) \le 1 - c - \eta$ for $1 \le j \le t$, where $c, \eta > 0$. Suppose moreover that $k \in [n]$ and $\mu([k]) \le \eta c$. Then there exists $v \in [k, n]$ with*

$$\frac{1}{s}\sum_{i=1}^{s} |T_i \cap [v]| < \frac{\lambda}{\eta} v \qquad \text{and} \qquad \frac{1}{t}\sum_{i=1}^{t} |C_i \cap [v]| < (1-c)v\,.$$

Proof. Let $U = \{v : \sum_{i=1}^{s} |T_i \cap [v]| \geq s\lambda v/\eta\}$. Writing S for the multiset which is the disjoint union of $T_1, \ldots, T_s$, so that $\mu(S) \leq s\lambda$ and $|S \cap [v]| = \sum_{i=1}^{s} |T_i \cap [v]|$, we can apply Lemma 6.1 with $\alpha = s\lambda/\eta$ to obtain $\mu(U) \leq \mu(S)/\alpha \leq \eta$.

In like manner, let $W = \{v : \sum_{i=1}^{t} |C_i \cap [v]| \geq t(1-c)v\}$. Writing now S for the multiset which is the disjoint union of $C_1, \ldots, C_t$, so that $\mu(S) \leq t(1-c-\eta)$ and $|S \cap [v]| = \sum_{i=1}^{t} |C_i \cap [v]|$, we apply Lemma 6.1 with $\alpha = t(1-c)$ to obtain $\mu(W) \leq t(1-c-\eta)/\alpha = 1-\eta/(1-c)$.

It follows that $\mu(U \cup W \cup [k]) < \eta + 1 - \eta/(1-c) + \eta c < 1$, so there exists $v \in [n]$ not contained in $U \cup W \cup [k]$. This v satisfies the conditions of the corollary. □

We can now prove the main result about containers and uniform measure, which was stated earlier as Theorem 2.8.

Proof of Theorem 2.8. Apply Theorem 2.5 to G to obtain a collection $\mathcal{C}$ of containers $C(T)$ for $T = (T_{r-1}, \ldots, T_0) \in \mathcal{P}^r[n]$. Since $\tau \leq \zeta^2/r$ we have $2r\tau/\zeta \leq 2\zeta$ so $\mu(C(T)) \leq 1 - 1/r! + 6\zeta$.

Let $(C_1, \ldots, C_t) \in \mathcal{C}^t$, where $t \in \mathbb{N}$. Each C_i is specified by an r-tuple of sets T_j, so the whole collection $(C_1, \ldots, C_t)$ is specified by rt sets which, after re-labelling, we call $T_1, \ldots, T_{rt}$, with $\mu(T_i) \leq 2\tau/\zeta$ for $1 \leq i \leq rt$. Let $c = 1/r! - 8\zeta$ and $\eta = 2\zeta$. Then $\mu([k]) \leq \zeta/2r! < \eta c$ so the conditions of Lemma 6.2 are satisfied with $s = rt$ and $\lambda = 2\tau/\zeta$, and we may choose $v \in [k, n]$ with

$$\frac{1}{s}\sum_{i=1}^{s} |T_i \cap [v]| < \frac{\tau}{\zeta^2} v \qquad \text{and} \qquad \frac{1}{t}\sum_{i=1}^{t} |C_i \cap [v]| < (1 - \frac{1}{r!} + 8\zeta)v\,.$$

Define $g(C_1, \ldots, C_t) = v$. Then (a) and (c) of the theorem are satisfied.

To obtain (b), we need that the containers have the online property: in other words, the t-tuple $(C_1 \cap [v], \ldots, C_t \cap [v])$ is determined by $T_1 \cap [v], \ldots, T_s \cap [v]$. This online property is guaranteed by Theorem 2.5. Hence the size of the set $Z = \{(C_1 \cap [v], \ldots, C_t \cap [v]) : g(C_1, \ldots, C_t) = v\}$ is bounded by the number of tuples $(T_1 \cap [v], \ldots, T_s \cap [v])$. Now $\sum_{i=1}^{s} |T_i \cap [v]| < s\theta v$, where $\theta = \tau/\zeta^2 < 1$. So by Lemma 5.1

$$\log |Z| \;\leq\; s\theta v(1 + \log(1/\theta)) \;\leq\; s\theta v \log(1/\tau) \;=\; \zeta^{-2} vtr\tau \log(1/\tau)\,,$$

which completes the proof. □

7. An example of large containers

Theorem 2.5 provides a small collection of containers for independent sets in an r-graph, each container having degree measure at most $1 - 1/r!$ (plus a term that is usually small). It is conceivable that an algorithm different to the one in §3 might yield smaller online containers. But in this section we describe examples to indicate that $1 - 1/r!$ is the limit of the present method.

Indeed, for each $r \geq 2$, there are examples of simple r-uniform hypergraphs G and independent sets I for which $\mu(C(G, T, \tau, \zeta))$ can be as large as $1 - 1/r! + o(1)$ when $T = T(G, I, \tau, \zeta)$. Because a precise detailed definition of the examples would be very long and opaque, we give instead a sketch showing how the examples work, together with a few words about how the details can be filled in.

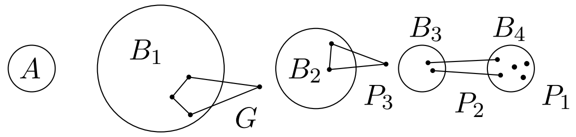

FIGURE 1. Approximate setup sufficient to produce a container of measure 23/24 in a 4-graph G.

7.1. **Sketch for $r = 4$.** Suppose G is a simple d-regular 4-graph with the following properties (see Figure 1). Let the vertex set be $A \cup B_1 \cup B_2 \cup B_3 \cup B_4$, where $|B_1| = 18k$, $|B_2| = 4k$, $|B_3| = k$, $|B_4| = k$, $|A| = o(k)$, so $\mu(B_4) = 1/4! + o(1)$ and so on, where $o(1) \to 0$ as $d \to \infty$. Since G is regular, the degree ordering used by Theorem 2.5 is not determined by the 4-graph itself, so we may suppose we have an ordering with $A < B_1 < B_2 < B_3 < B_4$. Let the induced subgraph of G on $B_1 \cup \cdots \cup B_4$ be approximately d-regular (so that only a few edges meeting $B_1 \cup \cdots \cup B_4$ also meet A), and suppose every edge has 3 vertices in B_1 and 1 vertex in $B_2 \cup B_3 \cup B_4$ (the sizes of the sets B_i have been chosen to ensure that this is possible). Suppose we run the algorithm with $\tau = d^{-1/3}$ (which is roughly the typical value of τ for a d-regular simple 4-graph).

First the algorithm is run with $s = 3$. After the vertices in A have been inspected but before any vertex in B_1 has been inspected, the partially constructed subgraph of P_3 will contain 3-edges that depend on I and on the edges of G that contain a vertex of A. But suppose for now that we can choose I, G so that at this point, P_3 contains a $d^{2/3}$-regular graph on $B_2 \cup B_3 \cup B_4$, in which every edge has two vertices in B_2 and one vertex in $B_3 \cup B_4$ (see Figure 1). Then since Γ_3 contains vertices with degree at least $\tau d = d^{2/3}$, all vertices in $B_2 \cup B_3 \cup B_4$ will be in Γ_3. Therefore, as the algorithm for P_3 proceeds to inspect the vertices after A, all vertices in B_1 will be kept in C_3 (since, for each vertex $v \in B_1$, there are very few edges whose first vertex is v that do not meet Γ_3). Furthermore, all the vertices in $B_2 \cup B_3$ will be kept in C_3 since they are not the first vertex of any 4-edge in G.

Similarly, if we can arrange things so that when the algorithm is run with $s = 2$, $B_3 \cup B_4$ contains a $d^{1/3}$-regular bipartite graph, and when the algorithm is run with $s = 1$, B_4 contains a 1-regular 1-graph, then the sets $B_1 \cup B_2 \cup B_3$ will be kept in C_2 and C_1. Also Γ_1 contains no singletons from $B_1 \cup B_2 \cup B_3$, so $B_1 \cup B_2 \cup B_3 \subset C$, and so $\mu(C) \geq 23/24 + o(1)$ as required.

This broad setup generalizes to all r by taking sets $B_1, \ldots, B_r$ of size $|B_i| = ((r-i+1)! - (r-i)!)k$, such that the induced graph of G on $B_1 \cup \cdots \cup B_r$ is approximately d-regular with every edge having $r-1$ vertices in B_1 and 1 vertex in $B_2 \cup \cdots \cup B_r$. Then provided that the P_s are similarly generalized, we will have $B_1 \cup \cdots \cup B_{r-1} \subset C$ and so $\mu(C) \geq 1 - 1/r! + o(1)$.

7.2. **Priming the P_s.** In the example above, we did not specify how to guarantee that the graphs P_s are as stated. This is achieved by adding edges to G containing vertices from A and specifying $I \subset A$ appropriately; we indicate how this may be achieved for $r = 4$.

Suppose $|A| = 2kd^{-1/3}$. Choose an equipartition $D_1 \cup D_2$ of B_2, and consider the 4 vertex classes $A, D_1, D_2, B_3 \cup B_4$. The last three classes each have size $2k$. Add edges to G forming a simple 4-partite 4-graph between these classes, such that in this set of edges the vertices in A have degree d and the vertices in $B_2 \cup B_3 \cup B_4$ have degree $d^{2/3}$. Now let $I = A$. Consider what happens when the algorithm is run on this graph. Each vertex

$v \in A$ will be added to T_3 and the corresponding edges will be added to P_3, producing the required $d^{2/3}$-regular graph on $B_2 \cup B_3 \cup B_4$.

Suppose now that $A = A_1 \cup A_2 \cup A_3$, $|A_1| = kd^{-2/3}$, $|A_2| = kd^{-1/3}$. Consider the 4 vertex classes A_1, A_2, B_3, B_4. Add edges forming a simple 4-partite 4-graph between these classes, such that in this set of edges the vertices in A_1 have degree d, the vertices in A_2 have degree $d^{2/3}$, and the vertices in $B_3 \cup B_4$ have degree $d^{1/3}$. Add also edges between A_2 and A_3 so that every vertex in $A_2 \cup A_3$ has degree d in G. Let $I = A_1 \cup A_2$, and suppose the degree ordering on A is $A_1 < A_2 < A_3$. When the algorithm is run, the vertices in A_1 will be added to T_3. Let $a_1 \in A_1$ be the last vertex in the degree order in A_1. After vertex a_1 has been inspected by the algorithm, the graph P_3 will be a 3-partite graph between classes A_2, B_3, B_4, where the vertices in A_2 have degree $d^{2/3}$ and the vertices in $B_3 \cup B_4$ have degree $d^{1/3}$. Every vertex $v \in A_2$ will be added to T_2, and the graph P_2 will then be the required $d^{1/3}$-regular graph on $B_3 \cup B_4$.

The graph P_1 can be achieved similarly. This argument overlooks the point that the graphs P_1, P_2, P_3 cannot be simultaneously produced as stated; for example, it would require the degree in P_3 of a vertex in $b \in B_2$ to be $d^{2/3} + d^{1/3} + 1$, whereas the maximum degree, by construction of P_3, is $d^{2/3} + 1$. Nonetheless, we can adjust some of the edge sets slightly so that the example works. These constructions generalize to all r.

8. List colourings

In [47], a lower bound for the list colouring number of a regular hypergraph was proved. Theorem 2.1 of that paper, based on a simple probabilistic argument, gave a bound of approximately $(\log k)/\log(1/c)$ provided there is a collection $\mathcal{C}$ of containers for the independent sets, with $|C| \le (1-c)n$ for each $C \in \mathcal{C}$ and with $|\mathcal{C}| \le e^{n/k}$. This proof fails to work for a general hypergraph because it is not possible to find containers of bounded size, but Theorem 2.8 provides conditions under which we can recover the proof.

Let $[t]$ be a set of colours. We say that a collection of lists $\{L_u \subset [t] : u \in [n]\}$ is *$\mathcal{C}$-compatible* if there is a colouring function $f : [n] \to [t]$ and a tuple $(C_1, \dots, C_t) \in \mathcal{C}^t$, such that, for each $u \in [n]$, $f(u) \in L_u$ and $u \in C_{f(u)}$.

Theorem 8.1. *Let $0 < \epsilon, c < 1$. Then there exists $k_0 = k_0(\epsilon, c)$, such that the following property holds for all $k > k_0$.*

Let $\ell = \lfloor (1-\epsilon)\log k/\log(1/c) \rfloor$ and let $t = \lfloor 2\ell^2/c \rfloor$. Let $n > k$ and let $\mathcal{C} \subset \mathcal{P}[n]$. Suppose that there is a map $g : \mathcal{C}^t \to [k, n]$, such that

$$(a) \qquad \frac{1}{t}\sum_{i=1}^{t} |C_i \cap [v]| \le (1-c)v$$

holds for every $(C_1, \dots, C_t) \in \mathcal{C}^t$, where $v = g(C_1, \dots, C_t)$. Suppose moreover that

$$(b) \qquad |\{ (C_1 \cap [v], \dots, C_t \cap [v]) : g(C_1, \dots, C_t) = v \}| \le e^{vt/k}$$

holds for all $v \in [n]$. Then there is a collection of lists $\{L_u : u \in [n]\}$, each of size $|L_u| = \ell$, which is not $\mathcal{C}$-compatible.

Proof. For each $u \in [n]$, let $L_u \in [t]^{(\ell)}$ be a subset of $[t]$ of size ℓ chosen uniformly and independently at random, and let $\mathcal{L} = \{L_u : u \in [n]\}$ be the collection of lists.

Given a tuple $(C_1, \dots, C_t) \in \mathcal{C}^t$, we define, for each $u \in [n]$, the set of colours

$$B_u = B_u(C_1, \dots, C_t) = \{i \in [t] : u \in C_i\}.$$

We say that $\mathcal{L}$ *fits* $(C_1,\ldots,C_t)$ if $L_u \cap B_u \neq \emptyset$ for each $u \in [n]$. This is the same as saying there is a function $f : [n] \to [t]$ with $f(u) \in L_u \cap B_u$, or in other words, $f(u) \in L_u$ and $u \in C_{f(u)}$. Hence we shall prove the theorem by showing that, with positive probability, $\mathcal{L}$ fits no tuple $(C_1,\ldots,C_t)$, since then $\mathcal{L}$ is not $\mathcal{C}$-compatible.

In fact, we claim something stronger: with positive probability, $\mathcal{L}$ *rejects* every tuple $(C_1,\ldots,C_t)$, meaning that there is some $u \in [v]$ with $L_u \cap B_u = \emptyset$, where $v = g(C_1,\ldots,C_t)$. To establish the claim, fix for the time being some tuple $(C_1,\ldots,C_t)$ and let $v = g(C_1,\ldots,C_t)$. Let $u \in [v]$ and let p_u be the probability that $L_u \cap B_u = \emptyset$, or equivalently $L_u \subset [t] \setminus B_u$. Then

$$p_u = \Pr(L_u \cap B_u = \emptyset) = \binom{z_u}{\ell}\binom{t}{\ell}^{-1} \quad \text{where } z_u = \max\{\ell - 1, t - |B_u|\}\,.$$

We note here that $\ell \geq 1$ if k_0 is large enough and thus $ct > \ell$. Write z for the average of the values z_u for $u \in [v]$; then by condition (a) of the lemma we have

$$vz = \sum_{u\in[v]} z_u \geq \sum_u t - |B_u| = vt - \sum_u |B_u| = vt - \sum_{i=1}^{t} |C_i \cap [v]| \geq vct\,.$$

So, since the function $\binom{z_u}{\ell}$ is convex for $z_u \geq \ell - 1$, we have

$$\begin{aligned}\sum_{u\in[v]} p_u = \sum_{u\in[v]} \binom{z_u}{\ell}\binom{t}{\ell}^{-1} &\geq v\binom{z}{\ell}\binom{t}{\ell}^{-1} \\ &\geq v\binom{ct}{\ell}\binom{t}{\ell}^{-1} \geq v(c - (\ell-1)/t)^{\ell}\,.\end{aligned}$$

Since $\ell \geq 1$ we have $(\ell-1)/t \leq (\ell-1)/(2\ell^2/c - 1) \leq c/2\ell$, and so $(c - (\ell-1)/t)^\ell \geq c^\ell(1 - 1/2\ell)^\ell \geq c^\ell/2$. Hence the probability that $\mathcal{L}$ fails to reject $(C_1,\ldots,C_t)$ is

$$\begin{aligned}\Pr(B_u \cap L_u \neq \emptyset \text{ for all } u \in [v]) &= \prod_{u\in[v]} (1 - p_u) \\ &\leq \exp\{-\sum_{u\in[v]} p_u\} \leq \exp\{-vc^\ell/2\}\,.\end{aligned}$$

Notice that the probability of $(C_1,\ldots,C_t)$ not being rejected depends only on the tuple $(C_1\cap[v],\ldots,C_t\cap[v])$ which, by condition (b) of the lemma, takes at most $\exp\{vt/k\}$ values as $(C_1,\ldots,C_t)$ ranges over $\mathcal{C}^t$. Hence if we fix v and write P_v for the probability that there is some tuple $(C_1,\ldots,C_t)$ with $v = g(C_1,\ldots,C_t)$ which is not rejected, then

$$\begin{aligned}P_v &\leq \exp\{vt/k - vc^\ell/2\} \\ &\leq \exp\left\{\frac{v}{2k}\left[\frac{4}{c}\left(\frac{(1-\epsilon)\log k}{\log 1/c}\right)^2 - k^\epsilon\right]\right\} \\ &\leq \exp\left\{-\frac{v}{2k}k^{\epsilon/2}\right\} \quad \text{if } k_0 \text{ is large enough} \\ &\leq \exp\left\{-\frac{1}{2}v^{\epsilon/2}\right\} \quad \text{since } k \leq v \\ &\leq v^{-2} \quad \text{if } k_0 \text{ is large enough.}\end{aligned}$$

Finally, if we consider all tuples $(C_1, \ldots, C_t) \in \mathcal{C}^t$, the probability that one of them is not rejected is at most

$$\sum_{v \in [k,n]} P_v \le \sum_{v \ge k} v^{-2} < 1$$

if k_0 is large. This establishes our claim and so proves the theorem. □

We can now prove a lower bound on the list chromatic number of a hypergraph. It is awkward to state the most general conditions under which a non-trivial bound can be obtained, so instead we state a couple of typical results. The first applies to many r-graphs of average degree d, and certainly to simple r-graphs; indeed, Theorem 1.1 stated earlier is the special case when G is simple.

Corollary 8.2. *Let $r \in \mathbb{N}$ be fixed. Let G be an r-graph with average degree d. Suppose that $d^{(j)}(v) \le d^{(r-j)/(r-1)+o(1)}$ for every $v \in V(G)$ and for $2 \le j \le r$ (recall Definition 2.2), where $o(1) \to 0$ as $d \to \infty$. Then*

$$\chi_l(G) \ge (1 + o(1)) \frac{1}{(r-1)^2} \log_r d\,.$$

Moreover, if G is regular then

$$\chi_l(G) \ge (1 + o(1)) \frac{1}{r-1} \log_r d\,.$$

Proof. We shall apply Theorem 2.8 to G. Choose $\zeta = \zeta(d)$ so that, as $d \to \infty$, then $\zeta = o(1)$, $\zeta = d^{o(1)}$ and $d^{(j)}(v) \le d^{(r-j)/(r-1)}\zeta^{-1}$ for all $v \in V(G)$ and $2 \le j \le r$. Let $\tau = d^{-1/(r-1)}\zeta^{-3}$. Then $\zeta \le 1/12r!$ because $\zeta = o(1)$. Also, recalling Definition 2.2, $\delta_j = \sum_v d^{(j)}(v)/\tau^{j-1}nd \le \zeta^2 = o(\zeta)$, so $\delta(G, \tau) \le \zeta$. Moreover $\tau \le \zeta^2/r$ because $\zeta = d^{o(1)}$.

Let $k = \lfloor \zeta^3/\tau \log(1/\tau) \rfloor$. Then $\log k = (1/(r-1) + o(1)) \log d$. Let $v \in V(G)$. For each $u \in V(G)$, at most $d^{(2)}(v) \le d^{(r-2)/(r-1)}\zeta^{-1}$ edges contain both u and v, and so $d(v) \le nd^{(r-2)/(r-1)}\zeta^{-1}$. It follows that $\mu([k]) \le (1/nd)knd^{(r-2)/(r-1)}\zeta^{-1} \le \zeta^5 \le \zeta/2r!$. So the conditions of Theorem 2.8 are satisfied. It follows that there exists a collection $\mathcal{C}$ of containers for the independent sets of G, satisfying properties (b) and (c) of Theorem 2.8, and since $\zeta^{-2}r\tau \log(1/\tau) < 1/k$ it follows that conditions (a) and (b) of Theorem 8.1 are satisfied, with $c = 1/r! - 8\zeta \ge (1 + o(1))r^{-(r-1)}$.

Consequently there are lists of size $(1 + o(1)) \log k/ \log(1/c)$ that are not $\mathcal{C}$-compatible, which is to say lists of size at least $(1/(r-1)+o(1)) \log d/ \log(1/c) \ge (1/(r-1)^2+o(1)) \log_r d$. Since $\mathcal{C}$ is a set of containers for the independent sets of G, the first claim of the corollary follows.

The proof for regular graphs is similar, except that we take $c = 1/r + o(1)$. To achieve this we make use of Corollary 2.7 instead of Theorem 2.8. With τ, ζ and k defined as before, we can take $\epsilon = \zeta$ in Corollary 2.7 because $\delta(G, \tau) = o(\zeta)$. We obtain a collection $\mathcal{C}$ of containers such that $e(G[C]) \le \zeta e(G) = o(e(G))$ for all $C \in \mathcal{C}$. Because G is regular this means that $|C| \le (1 - c)n$, where $c = 1/r + o(1)$ and $n = |G|$. We can now apply Theorem 8.1 by defining $g(C_1, \ldots, C_t) = n$ for all $(C_1, \ldots, C_t)$; note that condition (b) of the theorem is satisfied because, by Corollary 2.7, $\log |\mathcal{C}| \le c(r) \log(1/\epsilon)n\tau \log(1/\tau) < n/k$. The remainder of the proof is the same. □

The bound given for r-graphs of average degree d is weaker than that for regular r-graphs because we only had containers of measure $1 - 1/r!$ available, rather than $1 - 1/r$. Probably this is an artifact of our algorithm, and $\chi_l(G) \geq (1/(r-1)+o(1))\log_r d$ holds for r-graphs of average degree d.

The bound for regular graphs is tight. Indeed, let $K(r,m)$ be the complete r-partite r-graph with m vertices in each class. Suppose that lists of size ℓ are given to the vertices. Randomly choose, for each colour in the palette, a vertex class on which that colour is forbidden to be used; then the expected number of vertices with no available colour is $rmr^{-\ell}$ which is less than one if $\ell > 1+\log_r m$, and so $\chi_l \leq 2+\log_r m$ (see Haxell and Verstraëte [23]). This graph is d-regular where $d = m^{r-1}$ so $\chi_l \leq 2+(1/(r-1))\log_r d$. Note that $d^{(j)}(v) = m^{r-j} = d^{(r-j)/(r-1)}$.

It is not hard to construct an m-regular simple subgraph G of $K(r,m)$, and so (putting $d = m$) we have simple d-regular r-graphs with $\chi_l \leq 2+\log_r d$. Quite possibly $\chi_l \leq 2+(1/(r-1))\log_r d$ in this case too, because a subgraph of G with $d^{1-1/(r-1)}$ vertices in each class is likely to be very sparse, and a random colouring might be repairable if $rdr^{-\ell} < d^{1-1/(r-1)}$, or $\ell > 1+(1/(r-1))\log_r d$. But this argument is far from rigorous.

As an illustration of the use of containers for non-independent sets we finish with the next result.

Corollary 8.3. *Let G be a graph with average degree d. Then, for each $u \in V(G)$ there is a list L_u of $(1+o(1))\log_2 d$ colours, such that it is not possible to choose a colour $c(u) \in L_u$ with the vertices of each colour spanning a planar graph.*

Proof. We follow the proof of Corollary 8.2 with $r = 2$, except we use a set $\mathcal{C}$ of containers for those subsets I for which $G[I]$ is planar. Since a planar graph is 5-degenerate, we can apply Theorem 2.8 and continue with the proof exactly as before, provided $5 \leq \tau d\zeta/r$. But $\tau = d^{-1}\zeta^{-3}$ so this condition holds comfortably. $\square$

9. H-free graphs

Theorem 1.3 is obtained by a routine application of Corollary 2.7 to the following hypergraph, whose independent sets correspond to H-free ℓ-graphs on vertex set $[N]$.

Definition 9.1. Let H be an ℓ-graph. Let $r = e(H)$. The r-graph $G(N,H)$ has vertex set $[N]^{(\ell)}$, where $B = \{v_1,...,v_r\} \in V(G)^{(r)}$ is an edge whenever B, considered as an ℓ-graph with vertices in $[N]$, is isomorphic to H.

Lemma 9.2. *Let H be an ℓ-graph with $r = e(H) \geq 2$ and let $\gamma \leq 1$. For N sufficiently large, $\delta(G(N,H),\gamma^{-1}N^{-1/m(H)}) \leq r2^{r^2}v(H)!^2\gamma$.*

Proof. Let $G = G(N,H)$. Consider $\sigma \subset [N]^{(\ell)}$ (so σ is both a set of vertices of G and an ℓ-graph on vertex set $[N]$). The degree of σ in G is the number of ways of extending σ to an ℓ-graph isomorphic to H. If σ as an ℓ-graph is not isomorphic to any subgraph of H, then clearly $d(\sigma) = 0$. Otherwise, let $v(\sigma)$ be the number of vertices in σ considered as an ℓ-graph, so there exists $V \subset [N]$, $|V| = v(\sigma)$ with $\sigma \subset V^{(\ell)}$. Edges of G containing σ correspond to copies of H in $[N]^{(\ell)}$ containing σ, each such copy given by a choice of $v(H)-v(\sigma)$ vertices in $[N]-V$ and a permutation of the vertices of H. Hence $d(\sigma) = c_\sigma\binom{N-v(\sigma)}{v(H)-v(\sigma)}$ for some integer c_σ in the range $1 \leq c_\sigma \leq v(H)!$. Thus for N sufficiently large,

$$1/v(H)! \leq d(\sigma)N^{-v(H)+v(\sigma)} \leq v(H)!.$$

For $v \in V(G)$ and $1 \le j \le e(H)$, the quantity $d^{(j)}(v)$ is the maximum of $d(\sigma)$ over all $\sigma \subset [N]^{(\ell)}$ with $v \in \sigma$ and $|\sigma| = j$. Thus

$$1/v(H)! \le d^{(j)}(v) N^{-v(H)+f(j)} \le v(H)!, \quad \text{where } f(j) = \min_{H' \subset H,\, e(H')=j} v(H').$$

Let $\tau = \gamma^{-1} N^{-1/m(H)}$. Since $f(1) = \ell$ and $\gamma \le 1$, for $2 \le j \le e(H)$

$$\delta_j = \frac{\sum_v d^{(j)}(v)}{\tau^{j-1} \sum_v d(v)} \le v(H)!^2 \tau^{1-j} N^{\ell - f(j)} \le v(H)!^2 N^{\ell - f(j) + (j-1)/m(H)} \gamma.$$

By definition of $f(j)$ and $m(H)$, $\ell - f(j) + (j-1)/m(H) \le 0$. Hence $\delta_j \le v(H)!^2 \gamma$ and so, with $r = e(H)$,

$$\delta(G, \tau) = 2^{\binom{r}{2}-1} \sum_{j=2}^{r} 2^{-\binom{j-1}{2}} \delta_j \le r 2^{r^2} v(H)!^2 \gamma$$

as claimed. □

We use a well-known supersaturation theorem to bound the number of edges in containers.

Proposition 9.3 (Erdős and Simonovits [20])**.** *Let H be an ℓ-graph and let $\epsilon > 0$. There exists N_0 and $\eta > 0$ such that if C is an ℓ-graph on $N \ge N_0$ vertices containing at most $\eta N^{v(H)}$ copies of H then $e(C) \le (\pi(H) + \epsilon)\binom{N}{\ell}$.*

Proof of Theorem 1.3. Let η be given by Proposition 9.3, let $\beta = \min\{\epsilon, \eta\}$, let $G = G(N, H)$, let $r = e(H)$ and let $\tau = 12 r! r 2^{r^2} v(H)!^2 N^{-1/m(H)}/\beta$. By Lemma 9.2, $\delta(G, \tau) \le \beta/12r!$. The vertex set of G is $[N]^{(\ell)}$, $n = \binom{N}{\ell}$; and the edge set of G is the set of ℓ-graphs isomorphic to H, of which there are at most $v(H)!\binom{N}{v(H)}$. We claim that the collection $\mathcal{C}$ given by Corollary 2.7 applied with β for N sufficiently large satisfies the conditions of the theorem. Write c' for the constant $c = c(r)$ in Corollary 2.7.

Condition (a): $\mathcal{C}$ covers the independent sets of G, which are precisely the H-free ℓ-graphs on vertex set $[N]$.

Condition (b): for each I, the corresponding $C \in \mathcal{C}$ with $I \subset C$ is specified by a tuple $T = (T_1, \ldots, T_s) \in \mathcal{P}^s(I)$ with s bounded by a constant depending on r and ϵ. Furthermore $|T_i| \le c' \tau n \le c N^{\ell - 1/m(H)}$ for c sufficiently large as a function of H, ϵ.

Condition (c): the number of copies of H in C is at most $\beta e(G) \le \beta N^{v(H)}$. Note also that since $\beta \le \eta$, Proposition 9.3 implies that $e(C) \le (\pi(H) + \epsilon)\binom{N}{\ell}$.

Condition (d): this holds as $\log|\mathcal{C}| \le c' \log(1/\beta) n \tau \log(1/\tau) \le c N^{\ell - 1/m(H)} \log N$ for c and N sufficiently large. □

Proof of Corollary 1.4. We may assume that $e(H) \ge 2$, since otherwise the result is trivial. Let $\epsilon > 0$ and let $\mathcal{C} \subset \mathcal{P}([N]^{(\ell)})$ be the collection given by Theorem 1.3 (we may assume that N is sufficiently large). Note that $|\mathcal{C}| = 2^{o(N^\ell)}$ and $e(C) \le (\pi(H) + \epsilon)\binom{N}{\ell}$ for each $C \in \mathcal{C}$. Since every H-free ℓ-graph on N vertices is a subgraph of some graph $C \in \mathcal{C}$, this means there are at most $\sum_{C \in \mathcal{C}} 2^{e(C)} \le 2^{(\pi(H) + \epsilon + o(1))\binom{N}{\ell}}$ H-free graphs on N vertices. Since $\epsilon > 0$ was arbitrary this completes the proof. □

Proof of Theorem 1.5. The proof is very similar to Theorem 1.3, and we only sketch the details.

Let $r = \binom{v(H)}{\ell}$. Let G be the r-graph whose vertex set is two copies of $[N]^{(\ell)}$, denoted by V_R and V_B (vertices in V_R correspond to ℓ-edges and vertices in V_B correspond to non-ℓ-edges), and whose edges correspond to copies of H; thus $f \in (V_R \cup V_B)^{(r)}$ is an edge of G whenever $f \cap V_R$ and $f \cap V_B$ are the edges and non-edges, respectively, of an ℓ-graph isomorphic to H with vertices in $[N]$. Note that every induced-H-free ℓ-graph $I \subset [N]^{(\ell)}$ corresponds to an independent set of G, namely the set of vertices in V_R corresponding to the edges of I together with the set of vertices in V_B corresponding to non-edges of I.

G has very similar properties to $G' = G(N, K^{(\ell)}_{v(H)})$ given in Definition 9.1, where $K^{(\ell)}_{v(H)}$ is the complete ℓ-graph on $v(H)$ vertices. In particular, for fixed τ, the δ_j for G differ only by a constant factor by those for G'. Let $m = m(K^{(\ell)}_{v(H)}) = \left(\binom{v(H)}{\ell} - 1\right)/(v(H) - \ell)$, so that $\delta(G, N^{-1/m}/\epsilon) = O(\epsilon)$.

As in the proof of Theorem 1.3, we may apply Corollary 2.7 with $\tau = O(N^{-1/m}/\epsilon)$ to obtain the required collection $\mathcal{C} \subset \mathcal{P}(V_R \cup V_B)$, where each $C \in \mathcal{C}$ is identified with a 2-coloured ℓ-multigraph in the obvious way. □

We now prove Theorem 1.6. First we must define the function $h_p(H)$. For a 2-coloured ℓ-multigraph J with red and blue edge sets J_R and J_B, define

$$H_p(J) = -|J_R - J_B| \log_2 p - |J_B - J_R| \log_2(1 - p).$$

Observe that the probability that a random ℓ-graph on the same vertex set is a subgraph of J is $2^{-H_p(J)}$. Let

$$\mathrm{hex}(H, N) = \min\{\, H_p(J) : J_R \cup J_B = [N]^{(\ell)}, H \not\subset J \,\}\,.$$

Then we put $h_p(H) = \lim_{N\to\infty} \mathrm{hex}(H, N)\binom{N}{\ell}^{-1}$.

As always, we need a supersaturation lemma for the proof of Theorem 1.6. The proof of the following is essentially the same as that of Proposition 9.3, and we do not give details.

Lemma 9.4. *Let H be an ℓ-graph and let $0 < \epsilon, p < 1$. There exists N_0 and $\eta > 0$ such that if C is a 2-coloured ℓ-multigraph on $N \ge N_0$ vertices containing at most $\eta N^{v(H)}$ copies of H then $H_p(C) \ge (h_p(H) - \epsilon)\binom{N}{\ell}$.*

Proof of Theorem 1.6. The lower bound on the stated probability follows from the definition of $h_p(H)$. Let $\epsilon > 0$ and let η be given by Lemma 9.4. Let $\mathcal{C}$ be the collection of 2-coloured ℓ-multigraphs given by Theorem 1.5 satisfying $|\mathcal{C}| = 2^{o(1)\binom{N}{\ell}}$ and for every $C \in \mathcal{C}$, the number of copies of H in C is at most $\eta N^{v(H)}$. By Lemma 9.4, for each $C \in \mathcal{C}$ we have $H_p(C) \ge (h_p(H) - \epsilon)\binom{N}{\ell}$. Since every induced-$H$-free graph on vertex set $[N]$ is contained in some $C \in \mathcal{C}$,

$$\mathbb{P}(G^{(\ell)}(N,p) \text{ is induced-}H\text{-free}) \le \sum_{C\in\mathcal{C}} 2^{-H_p(C)} \le 2^{-(h_p(H)-\epsilon+o(1))\binom{N}{\ell}}.$$

Since $\epsilon > 0$ was arbitrary this completes the proof of Theorem 1.6. □

10. Linear equations

In this section, F denotes a finite field, the set of integers $[N]$, or an abelian group. We consider linear systems of the form $Ax = b$, where A is a $k \times r$ matrix. As in the introduction, when F is a finite field or $[N]$, then A has entries in F. Our methods apply to abelian groups as well: in this case, let A have integer entries, where integer-group multiplication ax, $a \in \mathbb{Z}$, $x \in F$, is a copies of x, $x + \cdots + x$; or $-a$ copies of $-x$ if a is negative. The definitions of *full rank* and *abundant* given in the introduction extend to abelian groups.

Often one wishes to discount solutions to an equation $Ax = b$ where the vector x contains repeated values. For example, in forbidding a 3-term arithmetic progression, we take $A = (1, 1, -2)$ and $b = (0)$ and discount solutions of the form $x + x - 2x = 0$. To accommodate this setup, we let $Z \subset F^r$ be a set of discounted solutions. We also call (F, A, b, Z) a $k \times r$ *linear system*, where a solution to the system is a vector $x \in F^r - Z$ such that $Ax = b$. A subset $I \subset F$ is *solution-free* if there is no $x \in I^r - Z$ with $Ax = b$.

In this setup, we could also define $\mathrm{ex}(F, A, b, Z)$ to be the maximum size of a solution-free subset when $Z \neq \emptyset$. However, this turns out to be unnecessary, since typically $|Z| = o(|F|^{r-k})$ so $\mathrm{ex}(F, A, b, Z) = \mathrm{ex}(F, A, b) + o(|F|)$ by Proposition 10.7.

Definition 10.1. Let (F, A, b, Z) be a $k \times r$ linear system with A abundant. When F is a finite field or $[N]$, define (as in Rödl and Rucińksi [40])

$$m_F(A) = \max_{J \subset [r],\, |J| \geq 2} \frac{|J| - 1}{|J| - 1 + \mathrm{rank}(A_J) - k},$$

where the matrix A_J is the $k \times (r - |J|)$ submatrix of A obtained by deleting columns indexed by J. Otherwise, when F is an abelian group, let t be the maximum value of j for which A_J has full rank whenever $|J| = j$, and let

$$m_F(A) = \frac{k + t - 1}{t - 1}.$$

It can readily be checked that if A is abundant then the denominators appearing in the definition of $m_F(A)$ are strictly positive. The separate definition of $m_F(A)$ when F is an abelian group is necessary since the rank of an integer matrix over an abelian group is not well-defined; in general, when the pair (F, A) could either be considered a finite field or an abelian group with A integer valued, the value of the second definition is at least as big as the value of the first definition. This is since $\mathrm{rank}(A_J) = k$ when $|J| \leq t$, and is otherwise at least $\max\{0, k + t - |J|\}$.

This is our main theorem for linear equations. The *determinantal* of a $k \times r$ integer matrix is the greatest common divisor of the determinants of its $k \times k$ submatrices.

Theorem 10.2. *Let $\epsilon > 0$ and let (F, A, b, Z) be a $k \times r$ linear system with A abundant. There are constants c, β, depending on A, ϵ in the case $F = [N]$, and depending only on k, r, ϵ otherwise, such that if $|F| \geq c$ and $|Z| \leq \beta |F|^{r-k}$ then there exists $\mathcal{C} \subset \mathcal{P}(F)$ satisfying*

(a) *for every solution-free subset $I \subset F$ there exists $C \in \mathcal{C}$ with $I \subset C$,*
(b) *moreover, for each pair I, C in (a), there exists $T = (T_1, \ldots, T_s)$ where $T_i \subset I$, $s \leq c$ and $\sum_i |T_i| \leq c|F|^{1-1/m_F(A)}$, such that $C = C(T)$,*
(c) *for every $C \in \mathcal{C}$, the number of solutions to $Ax = b$ with $x \in C^r - Z$ is at most $\epsilon |F|^{r-k}$,*
(d) *if F is a finite field or $[N]$, or F is an abelian group and the determinantal of A is coprime to $|F|$, then $|C| \leq ex(F, A, b) + \epsilon |F|$,*
(e) $\log |\mathcal{C}| \leq c|F|^{1-1/m_F(A)} \log |F|$.

Theorem 10.3. *Let (F, A, b, Z) be a $k \times r$ linear system with $|Z| = o(|F|^{r-k})$ and A abundant, and additionally with the determinantal of A coprime to $|F|$ in the case that F*

is an abelian group. Then the number of solution-free subsets of F *is* $2^{\mathrm{ex(F,A,b)+o(|F|)}}$. *Here* $o(1) \to 0$ *as* $|F| \to \infty$, *with* A *fixed in the case* $F = [N]$.

Note that Theorem 10.3 generalizes Theorem 1.9 by including abelian groups and allowing $Z \neq \emptyset$.

If A is not abundant, then the conclusion of Theorem 10.3 need not hold. For example, let $A = (1, 1)$, $b = (0)$, and consider the cyclic group C_n for n odd. Observe that the pairs (x, y) such that $x + y = 0$ and $x \neq y$ partition $C_n \setminus \{0\}$. Therefore $\mathrm{ex}(C_n, A, b) = (n + 1)/2$. However, one can construct a solution-free set by including either x or y or neither for each pair (x, y), so there are at least $3^{(n-1)/2}$ solution-free sets. There are similar examples with larger values of k and $r > k + 2$.

Additionally, when $F = [N]$, the condition that A is fixed as $|F| = N \to \infty$ is necessary. For example, for the equation $w + x + (10N)y - (10N)z = N$, the maximum size of a solution-free subset of $[N]$ is $N/2$ (since for every pair $w, x \in [N]$ with $w + x = N$, a solution-free set can include at most one of x or y), but there are at least $3^{(N-1)/2}$ solution free sets, since for every $w, x \in [N]$ with $w + x = N$ and $w \neq x$, we can include either x or y or neither to form a solution-free set.

Theorem 10.2 follows from an application of Corollary 2.7 to the following hypergraph, whose independent sets correspond to solution-free subsets of F.

Definition 10.4. Let (F, A, b, Z) be a $k \times r$ linear system. The r-partite r-graph $G = G(F, A, b, Z)$ has vertex set $V(G) = X_1 \cup \cdots \cup X_r$, where each X_i is a disjoint copy of F, and edge set $E(G) = \{x = (x_1, \ldots, x_r) \in X_1 \times \cdots \times X_r - Z : Ax = b\}$.

Fact 10.5. Let F be a finite field or abelian group, let A be a $k \times \ell$ matrix and let $b \in F^k$. If A has full rank then there are $|F|^{\ell-k}$ solutions to $Ax = b$. More generally if F is a finite field, there are at most $|F|^{\ell-\mathrm{rank(A)}}$ solutions to $Ax = b$.

Proof. If A has full rank, then for every $b_1, b_2 \in F^k$ there exists $x \in F^\ell$ with $Ax = b_2 - b_1$. Thus if x_1 is a solution to $Ax_1 = b_1$ then $A(x_1 + x) = b_2$, so by symmetry every $b \in F^k$ has $|F|^\ell/|F|^k$ solutions to $Ax = b$. The case when F is a finite field is standard. □

Lemma 10.6. *Let* (F, A, b, Z) *be a* $k \times r$ *linear system where* F *is a finite field or abelian group,* A *is an abundant matrix and* $|Z| \leq |F|^{r-k}/2$. *Let* $G = G(F, A, b, Z)$, $\gamma \leq 1$ *and* $\tau = |F|^{-1/m_F(A)}/\gamma$. *Then* $\delta(G, \tau) \leq r2^{r^2}\gamma$.

Proof. The number of edges in G is the number of solutions to $Ax = b$ not in Z. The matrix A has full rank, so by Fact 10.5 the number of edges of G is $|F|^{r-k} - |Z| \geq |F|^{r-k}/2$.

For $v \in V(G)$ and $j \geq 2$, $d^{(j)}(v)$ is the maximum over all sets $\sigma = \{y_1, \ldots, y_j\}$ with $v \in \sigma$ of the number of edges of G containing σ. If σ contains two vertices in the same part X_i then there are no edges containing σ. Otherwise suppose $y_\ell \in X_{i_\ell}$ for $\ell = 1, \ldots, j$, where $1 \leq i_1 < \cdots < i_j \leq r$. Let $J = \{i_1, \ldots, i_j\}$. The number of edges containing σ is at most the number of solutions to $Ax = b$ with $x_{i_\ell} = y_\ell$ for $\ell = 1, \ldots, j$, which for some $b^* \in F^k$ is the number of solutions to $A_J x^* = b^*$, $x^* \in F^{r-j}$. We now split the proof into two cases depending on whether F is a finite field or an abelian group.

When F is an abelian group: If $j \leq t$ (recall Definition 10.1), then A_J has full rank by assumption, and so Fact 10.5 implies the number of solutions is at most $|F|^{r-j-k}$. When $2 \leq j \leq t$, we have $d^{(j)}(v) \leq |F|^{r-j-k}$, so

$$\delta_j = \frac{\sum_v d^{(j)}(v)}{\tau^{j-1} \sum_v d(v)} \leq \frac{2|F|^{r-j-k}}{\tau^{j-1}|F|^{r-k-1}} = 2|F|^{1-j}\tau^{1-j}.$$

When $t+1 \le j \le t+k$, the bound $d^{(j)}(v) \le d^{(t)}(v)$ implies that

$$\delta_j \le \frac{2|F|^{r-t-k}}{\tau^{j-1}|F|^{r-k-1}} = 2|F|^{1-t}\tau^{1-j} \le 2|F|^{1-t}\tau^{1-t-k},$$

When $t+k+1 \le j \le r$, the bound $d^{(j)}(v) \le |F|^{r-j}$ implies that

$$\delta_j \le \frac{2|F|^{r-j}}{\tau^{j-1}|F|^{r-k-1}} = 2|F|^{k-j+1}\tau^{1-j}.$$

Since $\tau = \gamma^{-1}|F|^{-1/m_F(A)} = \gamma^{-1}|F|^{-(t-1)/(k+t-1)}$ and $\gamma \le 1$, this implies that $\delta_j \le 2\gamma$ for all j.

When F is a finite field: By Fact 10.5 the number of solutions to $A_J x^* = b^*$ is at most $|F|^{r-j-\mathrm{rank(A_J)}}$. Hence

$$d^{(j)}(v) \le \max_{J\subset[r],\,|J|=j} |F|^{r-j-\mathrm{rank(A_J)}}.$$

Using $\tau = \gamma^{-1}|F|^{-1/m_F(A)}$ and $\gamma \le 1$, this implies that

$$\delta_j = \frac{\sum_v d^{(j)}(v)}{\tau^{j-1}\sum_v d(v)} \le 2 \max_{J\subset[r],\,|J|=j} |F|^{1-j+k-\mathrm{rank(A_J)}+(j-1)/m_F(A)}\gamma.$$

The exponent is at most 0 be definition of $m_F(A)$ so $\delta_j \le 2\gamma$.

In both cases, $\delta(G,\tau) = 2^{\binom{r}{2}-1}\sum_{j=2}^{r} 2^{-\binom{j-1}{2}}\delta_j \le r2^{r^2}\gamma$. □

The proof of part (d) of Theorem 10.2 requires the following removal lemma of Král', Serra and Venna. Note that if A has determinantal coprime to $|F|$ then in particular A has full rank.

Proposition 10.7 (Král', Serra and Vena [31, 32]). *Let (F,A,b) be a $k\times r$ linear system where F is a finite field or abelian group and A has full rank. Suppose further that A has determinantal coprime to $|F|$ in the case that F is an abelian group. Let $\alpha > 0$. Then there exists $\eta = \eta(\alpha,k,r) > 0$ such that for all $Y_1,\dots,Y_r \subset F$ and $b \in F^k$, if there are at most $\eta|F|^{r-k}$ solutions to $Ax=b$ with $x_i \in Y_i$, then there are sets $Y_1' \subset Y_1,\dots,Y_r' \subset Y_r$ with $|Y_i'| \le \alpha|F|$ such that there is no solution to $Ax=b$ with $x_i \in Y_i \setminus Y_i'$.*

Proof of Theorem 10.2. We may assume that F is a finite field or abelian group. Indeed, $[N]$ can be embedded into the finite field $\mathbb{Z}_p$ for a sufficiently large prime p. Taking p in the range $4k!|A|^k N \le p \le 8k!|A|^k N$, where $|A|$ is the sum of the absolute values of A, guarantees that A is still abundant in $\mathbb{Z}_p$ and that a solution to $Ax=b$ (mod p) is also a solution to $Ax=b$ (provided, say, $|b_i| \le p/2$; but we may assume this since otherwise there are no solutions to $Ax=b$ in $[N]$). Then the result of this theorem for $(\mathbb{Z}_p,A,b,Z)$ implies the result for $([N],A,b,Z)$, since p/N is bounded by a constant depending only on A.

Let η be given by Proposition 10.7 applied with $\alpha = \epsilon/r$. Let $\beta = \min\{\epsilon,\eta/2,1/12r!\}$. Note that $\beta = \beta(\epsilon,k,r)$.

Let $G = G(F,A,b,Z)$ and let $\tau = 12r!r2^{r^2}|F|^{-1/m_F(A)}/\beta$ so that by Lemma 10.6, $\delta(G,\tau) \le \beta/12r!$. Note also that $\tau < 1/144r!^2 r$ for $|F|$ sufficiently large, so Corollary 2.7 gives a collection of sets $\mathcal{D}$ covering the independent sets of G. For $D \in \mathcal{D}$, let $\pi_i(D) = D \cap X_i \subset F$ be the part of D in the ith copy of F and let

$$\mathcal{C} = \{C(D) : D \in \mathcal{D}\} \subset \mathcal{P}(F) \qquad \text{where } C(D) = \pi_1(D)\cap\cdots\cap\pi_r(D).$$

We claim that $\mathcal{C}$ satisfies the conditions of the theorem.

Condition (a): consider a solution-free set $I \subset F$. The subset J of $V(G)$ formed by taking a copy of I in each X_i is an independent set in G. In particular, it is contained in some $D \in \mathcal{D}$, hence $I \subset C(D)$.

Condition (b): for I, C in (a), there exists a tuple $T = (T_1, \ldots, T_s)$ such that $T_i \subset J$, $D = D(T)$, $|T_i| \le 288r!^2\tau n \le c|F|^{1-1/m_F(A)}$ and $s \le c$ for c sufficiently large as a function of r and β. Thus C is specified by the rs-tuple $(T_1 \cap X_1, \ldots, T_1 \cap X_r, \ldots, T_s \cap X_r) \subset I^{rs}$. This verifies condition (b) (with rs in place of s).

Condition (c): consider $C \in \mathcal{C}$. Each solution to $Ax = b$ with $x \in C^r - Z$ corresponds to an edge of $G[D]$, of which there are at most $\beta e(G) = \beta|F|^{r-k}$.

Condition (d): there are at most $\beta|F|^{r-k} + |Z| \le \eta|F|^{r-k}$ such $x \in C^r$ with $Ax = b$, so by Proposition 10.7 there exists a set $Y = \cup_i Y_i \subset C$, $|Y| \le \epsilon|F|$ such that there are no $x \in (C - Y)^r$ with $Ax = b$, implying that $|C| \le \mathrm{ex}(F, A, b) + \epsilon|F|$.

Condition (e): observe that

$$\log|\mathcal{C}| \le \log|\mathcal{D}| \le c(r)\log(1/\beta)|F|\tau\log(1/\tau) \le c|F|^{1-1/m_F(A)}\log|F|,$$

for c sufficiently large as a function of r and β. □

Proof of Theorem 10.3. The proof is the same as the proof of Corollary 1.4, where the collection given by Theorem 10.2 is used instead of the collection given by Theorem 1.3. □

11. Sidon sets

In this section we prove Theorem 1.10. To prove the upper bound we construct, in the natural way, the hypergraph representing the solutions to $w + x = y + z$ in a subset $S \subset [n]$. We then apply Theorem 5.3 in an entirely mechanical way; all that is needed is to set appropriate values and to check the conditions. We remark that the codegree function δ for this hypergraph exhibits a change in behaviour when $|S| < n^{2/3}$, as the dominant contribution then comes from δ_2 rather than δ_4 (see equation (6)); Kohayakawa, Lee, Rödl and Samotij [29] noticed an interesting behavioural change at the same point, for a closely related problem.

Proof of Theorem 1.10. Lower bound construction. Suppose $n = 4p(p-1)$ for some prime p. Ruzsa [41] shows that there is a set $S \subset [p(p-1)]$ of size $p-1$ such that every sum of two elements of S is distinct modulo $p(p-1)$. Thus for any $U_1, U_2, U_3, U_4 \subset S$ satisfying $U_i \cap U_j = \emptyset$ for $i \ne j$, the set

$$U_1 \cup (U_2 + p(p-1)) \cup (U_3 + 2p(p-1)) \cup (U_4 + 3p(p-1))$$

is a Sidon subset of $[4p(p-1)]$, where $V + x := \{v + x : v \in V\}$. This gives $5^{p-1} = \sqrt{5}^{(1+o(1))\sqrt{n}} > 2^{(1.16+o(1))\sqrt{n}}$ Sidon subsets of $[n] = [4p(p-1)]$. The general case follows by embedding $[4p(p-1)]$ into $[n]$, where p is the largest prime such that $4p(p-1) < n$, and using the fact that the ratio of successive primes tends to 1. (We note that any construction for large modular Sidon sets could have been used here; this includes the classical constructions of Singer [50] and of Bose [9].)

Upper bound. Let G be the 4-graph on vertex set $[n]$, where $\{w, x, y, z\} \in [n]^{(4)}$ is an edge whenever $w + x = y + z$. Sidon sets correspond to independent sets in G (although the converse is not always true, since solutions to $w + x = y + z$ where $w = x$ or $y = z$ do not correspond to edges of G). We shall apply Theorem 5.3 to the graph G; our task is to bound $f(m)$. To this end, let $\beta = 3 \times 10^{14}$, let $u_0 = \beta\sqrt{n}$, and consider $U \subset [n]$ where $u = |U| \geq u_0$.

For $i \in [n-1]$, let $t_i = |\{\{x,y\} \in U^{(2)} : x < y, y - x = i\}|$. Note that $\sum_i t_i = \binom{u}{2}$. Each pair of sets $\{w,z\} \neq \{y,x\}$ with $w - z = y - x$ corresponds to an edge with $w + x = y + z$, and each such edge corresponds to the two pairs $\{w,z\} \neq \{y,x\}$ and $\{w,y\} \neq \{x,z\}$. Hence the number of edges in $G[U]$ satisfies

$$m = e(G[U]) = \frac{1}{2}\sum_{i=1}^{n-1}\binom{t_i}{2} \geq \frac{n-1}{2}\binom{\frac{1}{n-1}\sum_i t_i}{2} \geq u^4/20n,$$

where the last inequality holds for $u \geq u_0$. Let $e_0 = \beta^4 n/20$, as used for Theorem 5.3. Thus $m \geq e_0$. Let $k = 12r! = 288$. For the application of Theorem 5.3, put

$$\tau = \tau(U) = \max\{24ku^2/m, (4ku/m)^{1/3}\}.$$

We must check that the conditions of Theorem 5.3 hold.

Recall the definition of $d^{(j)}(w)$. In $G[U]$, observe that $d^{(2)}(w) \leq u/2 + u = 3u/2$, since for $x \in U$, the number of solutions of the form $w + x = y + z$ is at most $u/2$ and the number of solutions of the form $w + y = x + z$ is at most u; similarly $d^{(3)}(w) \leq 3$ and $d^{(4)}(w) \leq 1$. Hence

$$\delta_2 \leq \frac{3u^2}{8\tau m} \qquad \delta_3 \leq \frac{3u}{4\tau^2 m} \qquad \delta_4 \leq \frac{u}{4\tau^3 m},$$

and (since $\tau < 1/12$)

$$\delta = 32\delta_2 + 16\delta_3 + 4\delta_4 \leq \frac{12u^2}{\tau m} + \frac{2u}{\tau^3 m}. \tag{6}$$

Then both terms on the right hand side of (6) are less than $1/2k$, so $\delta \leq 1/12r!$ is satisfied. To apply Theorem 5.3, we also require $\tau \leq 1/144r!^2 r = 1/331776$.

If $\tau \leq 24ku^2/m$, then the constraint $\tau \leq 1/144r!^2 r$ is automatically satisfied (since $m \geq u^4/20n$ and $u \geq \beta\sqrt{n}$), and

$$\begin{aligned} u\tau\log(1/\tau) &\leq (24ku^3/m)\log(m/(24ku^2)) \\ &\leq 20^{3/4}48k\sqrt{n}\left(\frac{n}{m}\right)^{1/4}\log\frac{(m/n)^{1/4}}{(24k)^{1/2}(20)^{1/4}} \\ &=: f_1(m), \end{aligned}$$

where the first inequality holds since $\tau\log(1/\tau)$ is an increasing function of τ when $\tau < 1/e$, and the second inequality holds since $u^3\log(m/(24ku^2))$ is an increasing function of u when $u \leq e^{-1/3}\sqrt{m/24k}$, and $u \leq (20nm)^{1/4}$ which is less than $e^{-1/3}\sqrt{m/24k}$ because $m \geq e_0$.

Alternatively, if $\tau \leq (4ku/m)^{1/3}$ then the constraint $\tau \leq 1/144r!^2 r$ is automatically satisfied, and

$$\begin{aligned} u\tau\log(1/\tau) &\leq (4ku^4/27m)^{1/3}\log(m/4ku) \\ &\leq 6k^{1/3}n^{1/3}\log n \\ &=: f_2(m) \qquad \text{when } m \leq e(G), \end{aligned}$$

where the second inequality holds since $u^{4/3}\log(m/4ku)$ is an increasing function of u when $u \leq e^{-3/4}m/4k$ and $u \leq (20nm)^{1/4}$, together with the bound $m \leq n^4$. Let $f_2(m) = 0$ for $m > e(G)$.

Let $\alpha = 1 - 1/2r!$ and $m_i = e_0/\alpha^i = \beta^4 n/20\alpha^i$. The conditions of Theorem 5.3 hold, so let $\mathcal{C}$ be the collection of containers given by Theorem 5.3 for the graph G, where each $C \in \mathcal{C}$ satisfies $e(C) \le e_0$ (and hence $|C| \le u_0$), and $\log|\mathcal{C}| \le 288rr!^2 \sum_{i\ge 0} f(m_i)$. Since f_1 and f_2 are non-increasing functions of m, $f(m) \le \max\{f_1(m), f_2(m)\}$ for $m \ge e_0$.

Note that $\sum_{i\ge 0}\gamma^i = 1/(1-\gamma)$ and $\sum_{i\ge 0} i\gamma^i = \gamma/(1-\gamma)^2$, so

$$\begin{aligned}
288rr!^2 \sum_{i\ge 0} f_1(m_i) &= 288rr!^2 20^{3/4} 48k\sqrt{n} \sum_{i\ge 0} \frac{(20\alpha^i)^{1/4}}{\beta} \log \frac{\beta}{\alpha^{i/4}\sqrt{480k}} \\
&= 288rr!^2 \frac{960k\sqrt{n}}{\beta} \left(\frac{\alpha^{1/4}\log(1/\alpha)}{4(1-\alpha^{1/4})^2} + \frac{\log(\beta/\sqrt{480k})}{1-\alpha^{1/4}} \right) \\
&< \frac{7\sqrt{n}}{2}.
\end{aligned}$$

Observe that $m_i \ge n^4 > e(G)$ when $i \ge 3\log n/\log(1/\alpha)$ (and hence $f_2(m_i) = 0$), so

$$\sum_{i\ge 0} f_2(m_i) = o(\sqrt{n}).$$

Each Sidon set in $[n]$ is a subset of size at most $(1+o(1))\sqrt{n}$ of some $C \in \mathcal{C}$, $|C| \le u_0$. By Lemma 5.1 the number of such subsets is at most $\exp\{\theta u_0(1+\log(1/\theta))\}$, where θu_0 is the maximum size of a Sidon set, so $\theta = 1/\beta + o(1)$. Letting $\mathcal{S}$ be the collection of Sidon subsets of $[n]$,

$$\begin{aligned}
\frac{\log|\mathcal{S}|}{\sqrt{n}} &\le \frac{\theta u_0(1+\log(1/\theta))}{\sqrt{n}} + \frac{288rr!^2}{\sqrt{n}} \sum_{i\ge 0} f_1(m_i) + \frac{288rr!^2}{\sqrt{n}} \sum_{i\ge 0} f_2(m_i) \\
&< 1 + \log\beta + 7/2 + o(1) \;<\; 55\log 2 + o(1),
\end{aligned}$$

which completes the verification. □

12. Sparsity

In this section we prove Theorem 1.11 and related theorems. Note that the condition $p \ge cN^{-1/m(H)}$ in Theorem 1.11 is tight up to the value of c. Indeed, if $p = o(N^{-1/m(H)})$, it is readily checked that for some subgraph $H' \subset H$ with $m(H') = m(H)$, the expected number of copies of H' is much less than the number of edges, and removing very few edges will result in an H-free subgraph.

As a further illustration of the paradigm described in §1.6, we prove two other conjectures of Kohayakawa, Łuczak and Rödl [30]. The first of these has already been proved, by Conlon and Gowers [12] for strictly balanced graphs and by Samotij [42], following Schacht [48], for all graphs. It states that, for non-bipartite H, not only does every H-free subgraph I of a random graph have at most $(1+o(1))p\pi(H)\binom{N}{2}$ edges, but in the case that I has close to $p\pi(H)\binom{N}{2}$ edges, it can be made $(\chi(H)-1)$-partite by removing a small number of edges. The dense ($p = 1$) version of this theorem is the stability theorem of Erdős and Simonovits [14, 15, 49].

Theorem 12.1. *Let H be a 2-graph with $\pi(H) > 0$ and let $0 < \gamma < 1$. There exist constants $\epsilon, c > 0$ such that for N sufficiently large and for $p \ge cN^{-1/m(H)}$, the following is true. Let E_0 be the event that there exists an H-free subgraph $I \subset G(N,p)$ with $e(I) \ge$*

$(1-\frac{1}{\chi(H)-1}-\epsilon)p\binom{N}{2}$ which cannot be made $(\chi(H)-1)$-partite by removing at most $\gamma p\binom{N}{2}$ edges. Then $\mathbb{P}(E_0) \le \exp\{-\epsilon^2 p\binom{N}{2}\}$.

The other conjecture from [30], sometimes known as the KŁR conjecture, has a more technical statement. Let G be a graph. For $U, W \subset V(G)$, write $E_G(U,W) \subset E(G)$ for the set of edges of G with one vertex in U and one vertex in W. Let $e_G(U,W) = |E_G(U,W)|$ and write $d_G(U,W) = e_G(U,W)/(|U||W|)$ for the *edge density*. For $0 < \eta, p \le 1$, say that the pair (U,W) is *(η,p)-regular* if for every $U' \subset U$ with $|U'| \ge \eta|U|$ and $W' \subset W$ with $|W'| \ge \eta|W|$, the edge density satisfies

$$|d_G(U',W') - d_G(U,W)| \le \eta p.$$

This extends the notion of regularity to sparse graphs of density p.

Let H be a graph on vertex set $[h]$. In what follows, $V_1 \cup \cdots \cup V_h$ is a partition of $[N] = [hn]$, where each part has size $|V_i| = n$. If G is a graph on vertex set $[N]$, say that G is (H,η,p)-regular if for every pair (V_i,V_j) with $\{i,j\} \in E(H)$, the bipartite subgraph of G between V_i and V_j is (η,p)-regular.

Let $G = G(n,M,H)$ denote a graph chosen uniformly at random from all h-partite graphs with parts $V_1,\ldots,V_h$, having $e_G(V_i,V_j) = M$ if $\{i,j\} \in E(H)$ and $e_G(V_i,V_j) = 0$ otherwise. We say that $G(n,M,H)$ is H-free if there is no set of vertices $v_1,\ldots,v_h$ with $v_i \in V_i$ such that $\{v_i,v_j\}$ is an edge of G whenever $\{i,j\} \in E(H)$.

A proof of the next theorem in the case that H is balanced was given by Balogh, Morris and Samotij [6]. The theorem verifies the KŁR conjecture.

Theorem 12.2. *Let H be a graph and let $\alpha > 0$. Then there exist $\eta, c > 0$, such that for n sufficiently large, if $M \ge cn^{2-1/m(H)}$, then*

$$\mathbb{P}(G(n,M,H) \text{ is both } H\text{-free and } (H,\eta,M/n^2)\text{-regular}) \le \alpha^M.$$

Thus Theorem 12.2 says that although there might be sparse (H,η,p)-regular graphs with no copy of H (unlike in the dense case), there are extremely few of them.

A similar result to Theorem 1.11 applies to other structures, such as solutions to linear equations as in §10. So we can say, for example, that if J is a solution-free subset of F of maximum size, and if a subset $X \subset F$ is chosen with probability p, then almost certainly the maximum size of a solution-free subset of X is $p(|J|/|F| + o(1))|F|$, provided p is not too small.

Theorem 12.3. *Let $0 < \gamma < 1$ and let (F,A,b,Z) be a $k\times r$ linear system with A abundant, and additionally with the determinantal of A coprime to F in the case that F is an abelian group. Then there exist constants $\epsilon, c > 0$, depending on γ, A when $F = [N]$, and depending only on γ, k, r otherwise, such that for $|F| \ge c$, $|Z| \le \epsilon|F|^{r-k}$ and $p \ge c|F|^{-1/m_F(A)}$, if $X \subset F$ is a random subset with each element included independently with probability p, then the following event holds with probability greater than $1 - \exp\{-\gamma^3 p|F|/512\}$:*

every solution-free subset has at most $p(ex(F,A,b) + \gamma|F|)$ elements.

When F is a finite field, the condition $p \ge c|F|^{-1/m_F(A)}$ in Theorem 12.3 is known to be tight up to the value of the constant c appearing, at least under some slightly more restrictive assumptions on (F,A,b,Z). See Rödl and Ruciński [40].

Theorems 1.11, 12.1 and 12.3 follow by applying the following lemma to the collections given by Theorems 1.3 and 10.2.

Lemma 12.4. *Given $0 < \nu < 1$ and $s \geq 1$, there is a constant $\phi = \phi(\nu, s)$ such that the following holds. Let M be a set, $|M| = n$, and let $\mathcal{I} \subset \mathcal{P}(M)$. Let $t \geq 1$, let $\phi t/n \leq p \leq 1$ and let $\nu n/2 \leq d \leq n$. Suppose for each $I \in \mathcal{I}$ there exists both $T_I = (T_1, \ldots, T_{s'}) \in \mathcal{P}^{s'}(I)$ and $D = D(T_I) \subset M$, where $s' \leq s$, $\sum_i |T_i| \leq t$ and $|D(T_I)| \leq d$. Let $X \subset M$ be a random subset where each element is chosen independently with probability p. Then*

$$\mathbb{P}\left(|D(T) \cap X| > (1+\nu)pd \text{ for some } I \subset X,\ I \in \mathcal{I}\right) \leq \exp\{-\nu^2 pd/32\}. \tag{7}$$

Proof. Consider $I \in \mathcal{I}$ and $T = T_I = (T_1, \ldots, T_{s'})$. Let $J(T) = T_1 \cup \cdots \cup T_{s'}$. Let E_T be the event that

$$J(T) \subset X \quad \text{and} \quad |D(T) \cap X| \geq (1+\nu)pd.$$

The event E_T is contained in $F_T \cap G_T$, where F_T is the event that $J(T) \subset X$ and G_T is the event that $|(D(T) - J(T)) \cap X| \geq (1+\nu)pd - |J(T)|$. Since F_T and G_T are independent, $\mathbb{P}(E_T) \leq \mathbb{P}(F_T)\mathbb{P}(G_T)$. Now $|J(T)| \leq t \leq pn/\phi \leq 2pd/\phi\nu \leq \nu pd/2$ if ϕ is large. So by standard estimates (e.g., Chernoff's bound),

$$\mathbb{P}(G_T) \leq \mathbb{P}(\mathrm{Bin}(d,p) \geq (1+\nu/2)pd) \leq \exp\{-\nu^2 pd/16\},$$

where $\mathrm{Bin}(n,p)$ is the binomial random variable. Note that $\mathbb{P}(F_T) = p^{|J(T)|}$. Let $x = pn/t \geq \phi$, so $t \leq 2pd/x\nu$. If ϕ is large we may assume $p(n-t) > t$, so

$$\sum_T \mathbb{P}(F_T) \leq \sum_{i=0}^{t} \binom{n}{i} 2^{si} p^i \leq (t+1)\left(\frac{ne2^s p}{t}\right)^t \leq (xe^2 2^s)^t \leq (xe^2 2^s)^{\frac{2pd}{x\nu}} \leq \exp\{\nu^2 pd/32\}$$

holds if ϕ, and therefore x, is large. If there exists $I \subset X$, $I \in \mathcal{I}$ with $|D(T_I) \cap X| \geq (1+\delta)pd$, then the event E_{T_I} holds. Hence the probability in (7) is bounded by

$$\sum_T \mathbb{P}(F_T)\mathbb{P}(G_T) \leq \exp\{\nu^2 pd/32\}\exp\{-\nu^2 pd/16\} \leq \exp\{-\nu^2 pd/32\}$$

as claimed. □

Proof of Theorem 1.11. Let $\mathcal{I}$ be the set of H-free ℓ-graphs on vertex set $[N]$. Let $\epsilon = \gamma/4$ and $M = [N]^{(\ell)}$. For $I \in \mathcal{I}$, let $T = T_I$, $C = C(T)$ and $c' = c(H, \epsilon)$ be given by Theorem 1.3. Our aim is to apply Lemma 12.4 with $D(T) = C(T)$ and

$$\nu = \gamma/2, \quad d = (\pi(H) + \epsilon)\binom{N}{\ell}, \quad s = c', \quad t = c' N^{\ell - 1/m(H)}.$$

The conditions of Lemma 12.4 then hold with $n = \binom{N}{\ell}$, noting that $d \geq \nu n/2$ and that $p \geq cN^{-1/m(H)} \geq \phi t/n$ if c is large enough. Finally, note that in (7), each H-free ℓ-graph $I \in \mathcal{I}$ is contained in $C(T_I)$ and $(1+\nu)pd \leq (\pi(H) + \gamma)p\binom{N}{\ell}$, so the probability in the statement of the theorem is bounded by

$$\exp\{-\nu^2 pd/32\} \leq \exp\left\{-\gamma^3 p\binom{N}{\ell}/512\right\},$$

completing the proof. □

Proof of Theorem 12.1. Notice that $\pi(H) = 1 - 1/(\chi(H) - 1)$ and $\chi(H) \geq 3$. It is a standard exercise, using either the stability arguments of Erdős and Simonovits or using Szemerédi's regularity lemma, that there exists $\epsilon > 0$ such that if C is a 2-graph on vertex set $[N]$ for N sufficiently large with $e(C) \geq (1 - \frac{1}{\chi(H)-1} - 11\epsilon)\binom{N}{2}$ and such that C contains at most $\epsilon\binom{N}{v(H)}$ copies of H, then there exists a subgraph $F \subset C$ of size $e(F) \leq (\gamma/2)\binom{N}{2}$

such that $C - F$ is $(\chi(H) - 1)$-partite. We may and shall assume that $\epsilon \le 1/66$ and $65\epsilon^2 \le \gamma^3$.

Let $\mathcal{I}$ be the set of H-free graphs on vertex set $[N]$. For $I \in \mathcal{I}$ let $T = T_I$, $C = C(T)$ and $c' = c(H, \epsilon)$ be given by Theorem 1.3 with ϵ as above. Let

$$\mathcal{I}_1 = \left\{ I \in \mathcal{I} : e(C(T_I)) \ge \left(1 - \frac{1}{\chi(H) - 1} - 11\epsilon\right)\binom{N}{2} \right\},$$
$$\mathcal{I}_2 = \mathcal{I} - \mathcal{I}_1.$$

For $I \in \mathcal{I}_1$ let $F = F(T_I) \subset C(T_I)$ be as above, so that $C(T_I) - F(T_I)$ is $(\chi(H)-1)$-partite.

Let $X = G(N, p)$. Let E_1 be the event that there exists $I \subset X$, $I \in \mathcal{I}_1$ such that $|F(T_I) \cap X| \ge \gamma p\binom{N}{2}$. Let E_2 be the event that there exists $I \subset X$, $I \in \mathcal{I}_2$ such that $|C(T_I) \cap X| \ge (1 - \frac{1}{\chi(H)-1} - \epsilon)p\binom{N}{2}$. Observe that $E_0 \subset E_1 \cup E_2$.

The probability of E_2 is bounded by applying Lemma 12.4 to the collection $\mathcal{I}_2$, with $M = [N]^{(2)}$, $n = \binom{N}{2}$, $D(T_I) = C(T_I)$, $\nu = 10\epsilon$, $d = (1 - \frac{1}{\chi(H)-1} - 11\epsilon)\binom{N}{2} \ge \frac{1}{3}\binom{N}{2}$, $s = c'$ and $t = c'N^{2-1/m(H)}$; provided $p \ge cN^{-1/m(H)}$ and c, N are sufficiently large,

$$\mathbb{P}(E_2) \le \exp\left\{-\nu^2 pd/32\right\} \le \exp\left\{-25\epsilon^2 p\binom{N}{2}/24\right\}.$$

The probability of E_1 is bounded by applying Lemma 12.4 to the collection $\mathcal{I}_1$, with $D(T_I) = F(T_I)$, $\nu = \gamma$, $d = (\gamma/2)\binom{N}{2}$, $s = c'$ and $t = c'N^{2-1/m(H)}$; provided $p \ge cN^{-1/m(H)}$ and c, N are sufficiently large,

$$\mathbb{P}(E_1) \le \exp\left\{-\nu^2 pd/32\right\} = \exp\left\{-p\gamma^3\binom{N}{2}/64\right\} \le \exp\left\{-65p\epsilon^2\binom{N}{2}/64\right\}.$$

Since $\mathbb{P}(E_0) \le \mathbb{P}(E_1) + \mathbb{P}(E_2)$ and pN^2 is large, this completes the proof of Theorem 12.1. □

Proof of Theorem 12.2. Let $N = hn$, let $p = M/n^2$ and let $X = G(n, M, H)$. For ease of notation we shall often identify graphs with their edge sets. Let $\nu = \nu(H, \alpha)$, $\eta = \eta(H, \nu)$ and $\epsilon = \epsilon(H, \eta)$ be sufficiently small constants to be chosen later. Let $\mathcal{C}$ be given by Theorem 1.3 applied with H and ϵ. For $T = (T_1, \dots, T_s)$, let $J(T) = T_1 \cup \dots \cup T_s$, and define the following probabilistic events:

$$E_T : J(T) \subset X \subset C(T) \text{ and } X \text{ is } (H, \eta, p)\text{-regular},$$
$$F_T : J(T) \subset X,$$
$$G_T : X \subset C(T) \text{ and } X - J(T) \text{ is } (H, 2\eta, p)\text{-regular}.$$

Theorem 1.3 tells us that if X is H-free then there exists $T = (T_1, \dots, T_s)$ with $J(T) \subset X \subset C(T)$, and if in addition X is (H, η, p)-regular then E_T holds. By the union bound, it is hence sufficient to show that $\sum_T \mathbb{P}(E_T) \le \alpha^M$.

Theorem 1.3 also states that $s \le c'$ and $|J(T)| \le c'N^{2-1/m(H)}$, where $c' = c'(\epsilon, H)$, so it is easily checked that if X is (H, η, p)-regular then $X - J(T)$ is $(H, 2\eta, p)$-regular provided that $p \ge (c'/\eta^3)N^{-1/m(H)}$. This holds provided $M \ge cn^{2-1/m(H)}$ for c sufficiently large, which we shall assume. Therefore $E_T \subset F_T \cap G_T$ and so $\mathbb{P}(E_T) \le \mathbb{P}(F_T)\mathbb{P}(G_T|F_T)$.

In order to bound $\mathbb{P}(G_T|F_T)$, let T be fixed, let $J = J(T)$ and let $C = C(T) \in \mathcal{C}$. Consider an η-regular Szemerédi partition of C refining the partition $V_1, \dots, V_h$, i.e., for r sufficiently large depending only on η, and for each $i \in [h]$ an equitable partition $V_i = V_{i,1} \cup \dots \cup V_{i,r}$ so that all but at most $\eta r^2 e(H)$ of the bipartite graphs between pairs $(V_{i,x}, V_{j,y})$ with $\{i, j\} \in E(H)$ are η-regular.

Provided η and ϵ are sufficiently small then, by standard arguments, for every choice of $x_1, \ldots, x_h \in [r]$ there exists some pair $\{i,j\} \in E(H)$ such that the pair (V_{i,x_i}, V_{j,x_j}) is either irregular or has density $d_C(V_{i,x_i}, V_{j,x_j}) < \nu$ (since otherwise there would be more than ϵN^h copies of H in C, contradicting the conditions of Theorem 1.3). In particular, by averaging, there exists $\{i_0, j_0\} \in E(H)$ such that at least $r^2/e(H)$ of the pairs $(V_{i_0,x}, V_{j_0,y})$ with $x, y \in [r]$ are either irregular or of density less than ν. Let

$$R = \{(V_{i_0,x}, V_{j_0,y}) : x, y \in [r] \text{ and } d_C(V_{i_0,x}, V_{j_0,y}) < \nu\},$$
$$S = \{\{u, w\} : u \in U,\ w \in W, \text{ for some } (U, W) \in R\}.$$

Since there are at most $\eta r^2 e(H)$ irregular pairs $(V_{i_0,x}, V_{j_0,y})$ with $x, y \in [r]$, this implies that $|R| \geq r^2/2e(H)$ for η sufficiently small. Allowing for the sets $V_{i,x}$ not being all exactly the same size, this implies that $|S| \geq n^2/4e(H)$. Note that, by the definition of R, $|S \cap C| \leq \nu|S| \leq \nu n^2$ holds.

Write $P(i,j)$ for the set of pairs of vertices $\{\{u,v\} : u \in V_i, v \in V_j\}$. If G_T holds, then $X - J$ is $(H, 2\eta, p)$-regular and so

$$|S \cap (X - J)| \geq (1 - 2\eta)p|S| \geq M/8e(H). \tag{8}$$

However $|S \cap C| \leq \nu n^2$ and if G_T holds then $X \subset C$, so the probability of (8) is small. Specifically, in generating the random graph $(X - J) \cap P(i_0, j_0)$ when conditioned on $J \subset X$, we are selecting a set of $M - |J \cap P(i_0, j_0)| \leq M$ edges uniformly from at least $n^2 - |J \cap P(i_0, j_0)| \geq n^2/2$ possible edges, and for (8) to hold, we must select at least $M/8e(H)$ edges from a set of νn^2 possibilities. This probability is at most

$$\mathbb{P}(G_T | F_T) \leq \binom{M}{M/8e(H)} \left(\frac{\nu n^2}{n^2/2}\right)^{M/8e(H)} \leq 2^M (2\nu)^{M/8e(H)} \leq (\alpha/2)^M,$$

provided $\nu = \nu(H, \alpha)$ is small enough.

Thus $\sum_T \mathbb{P}(E_T) \leq \sum_T \mathbb{P}(F_T)\mathbb{P}(G_T|F_T) \leq (\alpha/2)^M \sum_T \mathbb{P}(F_T)$, and to complete the proof it is enough to show $\sum_T \mathbb{P}(F_T) \leq 2^M$. Now

$$\sum_T \mathbb{P}(F_T) = \sum_T \mathbb{P}(J(T) \subset X) = \sum_T \mathbb{E}\mathbf{1}_{J(T) \subset X} = \mathbb{E}\,|\{T : J(T) \subset X\}|\,.$$

Since $T = (T_1, \ldots, T_s)$ where $s \leq c'$, Lemma 5.1 tells us that $|\{T : J(T) \subset X\}| \leq \exp\{c'\theta|X|(1 + \log(1/\theta))\}$ where $\theta|X|$ is the average size of the T_i. Now $|X| = Me(H) \geq ce(H)n^{2-1/m(H)}$ and $\theta|X| \leq c'n^{2-1/m(H)}$. We can make θ as small as we wish to by making c large, thus ensuring $|\{T : J(T) \subset X\}| \leq 2^M$, as desired. □

Proof of Theorem 12.3. Repeat the proof of Theorem 1.11 with $M = F$ and use Theorem 10.2 instead of Theorem 1.3 for the collection $\mathcal{C}$. □

Proof of Theorem 1.12. Let $([N], A, b, Z)$ be the $(\ell - 2) \times \ell$ linear system corresponding to forbidding an ℓ-term arithmetic progression in $[N]$. For example if $\ell = 3$ then $A = (1, 1, -2)$, $b = (0)$ and Z is the set of solutions of the form $x + x - 2x = 0$ that are discounted, so $|Z| = N$. It can readily be checked that $m_{[N]}(A) = \ell - 1$. Since $\mathrm{ex}(F, A, b) = 0$, the result immediately follows by applying Theorem 12.3. □

References

[1] N. Alon, Independent sets in regular graphs and sum-free subsets of finite groups, *Israel Journal of Mathematics* **73** (1991), 247–256.
[2] N. Alon, Restricted colorings of graphs, in *Surveys in Combinatorics 1993*, London Mathematical Society Lecture Notes Series, **187** (K. Walker, ed.), Cambridge University Press, Cambridge, (1993), 1–33.
[3] N. Alon, Degrees and choice numbers, *Random Structures and Algorithms* **16** (2000), 364–368.
[4] N. Alon and A. Kostochka, Hypergraph list coloring and Euclidean Ramsey theory. (submitted).
[5] N. Alon and A. Kostochka, Dense uniform hypergraphs have high list chromatic number, (preprint).
[6] J. Balogh, R. Morris and W. Samotij, Independent sets in hypergraphs, (preprint).
[7] J. Balogh and W. Samotij, The number of $K_{s,t}$-free graphs, *J. London Mathematical Soc.* **83** (2011), 368–388.
[8] B. Bollobás and A. Thomason, The structure of hereditary properties and colourings of random graphs, *Combinatorica* **20** (2000) 173–202.
[9] R.C. Bose, An affine analogue of Singer's theorem, *J. Indian Math. Soc.* **6** (1942), 1–15.
[10] N.J. Calkin, On the number of sum-free sets, *Bulletin of the London Mathematical Society* **22** (1990), pp141.
[11] P.J. Cameron and P. Erdős, On the number of sets of integers with various properties, Number Theory (Banff, AB, 1988), de Gruyter, Berlin (1990), 61–79.
[12] D. Conlon and W.T. Gowers, Combinatorial theorems in sparse random sets, submitted.
[13] R. Dotson and B. Nagle, Hereditary properties of hypergraphs, *J. Combinatorial Theory (Ser. B)* **99** (2009), 460–473.
[14] P. Erdős, Some recent results on extremal problems in graph theory (Results), in: P. Rosenstiehl (Ed.), Theory of Graphs (International Symposium, Rome, 1966), Gordon and Breach, NewYork, Dunod, Paris, 1967, pp. 118–123.
[15] P. Erdős, On some new inequalities concerning extremal properties or graphs, in: P. Erdős, G. Katona (Eds.), Theory of Graphs, Academic Press, New York, 1968, pp. 77–81.
[16] 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.
[17] P. Erdős, D.J. Kleitman and B.L. Rothschild, Asymptotic enumeration of K_n-free graphs, *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973)* **2** (1976), 19–27.
[18] P. Erdős, A. Rényi and V.T. Sós, On a problem of graph theory, *Studia Sci. Math. Hungar.* **1** (1966), 51–57.
[19] P. Erdős, A.L. Rubin and H. Taylor, Choosability in graphs, *Proc West Coast Conf. on Combinatorics, Graph Theory and Computing*, Congressus Numerantium XXVI (1979), 125–157.
[20] P. Erdős and M. Simonovits, Supersaturated graphs and hypergraphs, *Combinatorica* **3** (1983), 181–192.
[21] P. Erdős and P. Turán, On a problem of Sidon in additive number theory, and on some related problems, *Journal of the London Mathematical Society* **1** (1941), 212.
[22] P. Haxell and M. Pei, On list coloring Steiner triple systems, *J. Combinatorial Designs* **17** (2009), 314-322.
[23] P. Haxell and J. Verstraëte, List coloring hypergraphs, *Electronic J. Combinatorics* **17** (2010) R129, 12pp.
[24] B. Green, The Cameron-Erdős conjecture, *Bull. London Math. Soc.* **36** (2004), 769-778.
[25] B. Green, A Szemerédi-type regularity lemma in abelian groups, with applications, *Geometric and Functional Analysis* **15** (2005), 340–376.
[26] J. Kahn, An entropy approach to the hard-core model on bipartite graphs, *Combin. Probab. Comput.* **10** (2001), 219–237.
[27] Gy.A. Katona, A theorem of finite sets, in "Theory of graphs (Proc. Colloq., Tihany, 1966)" Academic Press, New York, (1968), 187–207.
[28] D.J. Kleitman and K.J. Winston, On the number of graphs without 4-cycles, *Disc. Mathematics* **41** (1982), 167–172.

[29] Y. Kohayakawa, S.J. Lee, V. Rödl and W. Samotij, The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers, (preprint).

[30] Y. Kohayakawa, T. Łuczak and V. Rödl, On K_4-free subgraphs of random graphs, *Combinatorica* **17** (1997), 173–213.

[31] D. Král', O. Serra and L. Vena, A removal lemma for systems of linear equations over finite fields, *Israel Journal of Mathematics (to appear)* (2010).

[32] D. Král', O. Serra and L. Vena, On the removal lemma for linear systems over abelian groups, *Arxiv preprint arXiv:1106.4243* (2011).

[33] J.B. Kruskal, The number of simplices in a complex, in "Mathematical optimization techniques", Univ. of California Press, Berkeley, Calif. (1963), 251–278.

[34] E. Marchant and A. Thomason, Extremal graphs and multigraphs with two weighted colours, in "Fete of Combinatorics and Computer Science" *Bolyai Soc. Math. Stud.*, **20** (2010), 239–286.

[35] E. Marchant and A. Thomason, The structure of hereditary properties and 2-coloured multigraphs, *Combinatorica* **31** (2011), 85–93.

[36] J.W. Moon and L. Moser, On cliques in graphs, *Israel J. Math.* **3** (1965) 23–28.

[37] B. Nagle, V. Rödl and M. Schacht, Extremal hypergraph problems and the regularity method, in "Topics in discrete mathematics", *Algorithms Combin.* **26** (2006), 247–278.

[38] H.-J. Prömel and A. Steger, Excluding induced subgraphs III: a general asymptotic, *Rand. Struct. Alg.* **3** (1992), 19–31.

[39] L. Pyber, V. Rödl and E. Szemerédi, Dense graphs without 3-regular subgraphs, *J. Combinatorial Theory (Ser. B)* **63** (1995), 41–54.

[40] V. Rödl and A. Rucińksi, Rado partition theorem for random subsets of integer, *Proc. of the London Math. Soc.* **74** (1997), 481–502.

[41] I. Ruzsa, Solving a linear equation in a set of integers I, *Acta Arith* **65** (1993), 259–282.

[42] W. Samotij, Stability results for random discrete structures, (preprint).

[43] A.A. Sapozhenko, On the number of connected subsets with given cardinality of the boundary in bipartite graphs. (in Russian) *Metody Diskret. Analiz.* **45** 45 (1987), 42–70, 96.

[44] A.A. Sapozhenko, On the number of independent sets in extenders, *Discrete Math. Appl.* **11** (2001), 155–161.

[45] A.A. Sapozhenko, Independent sets in quasi-regular graphs, *European J. Combin.* **27** (2006), 1206–1210.

[46] A.A. Sapozhenko, The Cameron-Erdős conjecture, *Discrete Math.* **308** (2008), 4361–4369.

[47] D. Saxton and A. Thomason, List colourings of regular hypergraphs, *Combinatorics, Probability and Computing* (2011).

[48] M. Schacht, Extremal results for random discrete structures, submitted.

[49] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, in: P. Erdős, G. Katona (Eds.), Theory of Graphs (Proceedings of the Colloquium, Tihany, 1966), Academic Press, New York, 1968, pp. 279–319.

[50] J. Singer, A theorem in finite projective geometry and some applications to number theory, *Trans. Amer. Math. Soc.* **43** (1938), 377–385.

[51] A. Thomason, Graphs, colours, weights and hereditary properties, in 'Surveys in Combinatorics, 2011 (R. Chapman ed.), *LMS Lecture Note Series* **392** (2011), 333–364.

[52] V.G. Vizing, Coloring the vertices of a graph in prescribed colors, *Diskret. Analiz* No. 29, Metody Diskret. Anal. v Teorii Kodov i Shem 101 (1976), 3–10, 101 (in Russian).

[53] Y. Zhao, The number of independent sets in a regular graph, *Combin. Probab. Comput.* **19** (2010), 315–320.

Department of Pure Mathematics and Mathematical Statistics, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK

E-mail address: `d.saxton@dpmms.cam.ac.uk`

E-mail address: `a.g.thomason@dpmms.cam.ac.uk`