THE NUMBER OF B_h-SETS OF A GIVEN CARDINALITY

DOMINGOS DELLAMONICA JR., YOSHIHARU KOHAYAKAWA, SANG JUNE LEE, VOJTĚCH RÖDL, AND WOJCIECH SAMOTIJ

Abstract. For any integer $h \geqslant 2$, a set A of integers is called a *B_h-set* if all sums $a_1 + \ldots + a_h$, with $a_1, \ldots, a_h \in A$ and $a_1 \leqslant \ldots \leqslant a_h$, are distinct. We obtain essentially sharp asymptotic bounds for the number of B_h-sets of a given cardinality that are contained in the interval $\{1, \ldots, n\}$. As a consequence of these bounds, we determine, for any integer $m \leqslant n$, the cardinality of the largest B_h-set contained in a typical m-element subset of $\{1, \ldots, n\}$.

1. Introduction

Let $h \geqslant 2$ be an integer. We call a set A of integers a *B_h-set* if all sums of the form $a_1 + \ldots + a_h$, where $a_1, \ldots, a_h \in A$ satisfy $a_1 \leqslant \ldots \leqslant a_h$, are distinct. The study of B_h-sets goes back to the work of Sidon [36], who, motivated by the study of certain trigonometric series, considered infinite sequences $k_1 < k_2 < \ldots$ for which the number of representations of each integer M as $k_i + k_j$, with $i \leqslant j$, is uniformly bounded. In particular, Sidon asked (see also [14]) to determine the maximum number of elements in such a sequence that are not larger than a given integer n, when the upper bound on the number of representations as above is one. For each $h \geqslant 2$ and n, let $[n] := \{1, \ldots, n\}$, and define

$$F_h(n) = \max\{|A| \colon A \subset [n] \text{ is a } B_h\text{-set}\}.$$

In other words, Sidon was interested in the asymptotic behavior of the function F_2. (This is why B_2-sets are now usually referred to as *Sidon sets*.) The results of Chowla, Erdős, Singer, and Turán [6, 14, 13, 37] yield that $F_2(n) = (1 + o(1))\sqrt{n}$, which answers the question of Sidon. The asymptotic behavior of the function F_h in the case $h > 2$ is less well understood, even though the problem of estimating it has received considerable amount of attention. Bose and Chowla [3] showed that $F_h(n) \geqslant (1 + o(1))n^{1/h}$ for each $h \geqslant 3$. On the other hand, an easy counting argument gives that for all h and n,

$$F_h(n) \leqslant (h \cdot h! \cdot n)^{1/h} \leqslant hn^{1/h}.$$

Date: 2017/10/23, 7:17pm.
2010 *Mathematics Subject Classification.* 11B75 (primary), and 05A16, 05D40 (secondary) .
The third author was the corresponding author.
The second author was partially supported by FAPESP (2013/03447-6, 2013/07699-0), CNPq (310974/2013-5, 459335/2014-6), NSF (DMS 1102086) and NUMEC/USP (Project MaCLinC). The third author was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2013R1A1A1059913) and the Ministry of Education (NRF-2016R1D1A1B03933404). The fourth author was supported by the NSF grants DMS 0800070, 1301698, and 1102086. The fifth author was partially supported by a Trinity College JRF and a grant from the Israel Science Foundation.

Successively better bounds of the form $F_h(n) \leqslant c_h n^{1/h}$ for sufficiently large n were given in [5, 7, 12, 19, 26, 27, 29, 35]. Currently, the best bounds are due to Green [15], who proved that

$$c_3 < 1.519, \quad c_4 < 1.627 \quad \text{and} \quad c_h \leqslant \frac{1}{2e}\left(h + \left(\frac{3}{2} + o(1)\right)\log h\right),$$

where $o(1)$ is some function tending to 0 as $h \to \infty$. For a wealth of material on B_h-sets, the reader is referred to the classical monograph of Halberstam and Roth [16] and to the more recent survey of O'Bryant [31].

In this work, we shall be interested in the problem of *enumerating* B_h-sets. Let $\mathcal{Z}_n^h$ be the family of all B_h-sets contained in $[n]$. In 1990, Cameron and Erdős [4] proposed the problem of estimating $|\mathcal{Z}_n^2|$, that is, the number of Sidon sets contained in $[n]$. Recalling the definition of $F_h(n)$ and observing that the property of being a B_h-set is preserved under taking subsets, one easily obtains

$$2^{F_h(n)} \leqslant |\mathcal{Z}_n^h| \leqslant \sum_{t=0}^{F_h(n)} \binom{n}{t}. \tag{1}$$

Since $(1 + o(1))n^{1/h} \leqslant F_h(n) \leqslant hn^{1/h}$, one deduces from (1) that

$$(1 + o(1))n^{1/h} \leqslant \log_2 |\mathcal{Z}_n^h| \leqslant c_h n^{1/h} \log n, \tag{2}$$

where c_h is some positive constant.

The logarithmic gap between the lower and the upper bounds in (2) was first closed in the case of Sidon sets [23], that is, when $h = 2$, and subsequently [10] for arbitrary h.

Theorem 1 ([23, 10])**.** *For every $h \geqslant 2$, there exists a constant C_h such that $|\mathcal{Z}_n^h| \leqslant 2^{C_h n^{1/h}}$ holds for all n.*

Another proof of Theorem 1 in the case $h = 2$ was later given by Saxton and Thomason [33] (see [28] for a similar result for $[n]^d$, $d \geqslant 2$). Let us also mention that Saxton and Thomason [33] proved that, perhaps somewhat surprisingly, we have $\log_2 |\mathcal{Z}_n^2| \geqslant (1.16 + o(1))F_2(n)$.

In fact, both [23] and [10] considered a somewhat refined version of the original question posed by Cameron and Erdős. This refinement was motivated by the problem of estimating the maximum size of a B_h-set contained in a *random* set of integers, which was the main focus of these two papers; for details, we refer the reader to §1.1. For a nonnegative integer t, let $\mathcal{Z}_n^h(t)$ be the family of all B_h-sets contained in $[n]$ that have precisely t elements. The main results of [10, 23] are estimates on the cardinality of $\mathcal{Z}_n^h(t)$ for a wide range of t.

In order to establish a lower bound for $|\mathcal{Z}_n^h(t)|$, in [10] we exhibited two large subfamilies of $\mathcal{Z}_n^h(t)$. One of them is constructed using a standard deletion argument. The resulting family is very large, but the construction works only if $t \leqslant \varepsilon_h n^{1/(2h-1)}$ for some constant $\varepsilon_h > 0$. The second one is built using a certain blow-up operation. The resulting family is much smaller, but the construction is valid for all $t \leqslant F_h(n)$. The lower bounds on $|\mathcal{Z}_n^h(t)|$ that are implied by the existence of these two families can be summarized as follows.

Proposition 2 ([10])**.** *The following holds for every $h \geqslant 2$.*

(*i*) *For every $\delta > 0$, there exists a constant $\varepsilon = \varepsilon(h, \delta) > 0$ such that for each $t \leqslant \varepsilon n^{1/(2h-1)}$,*

$$|\mathcal{Z}_n^h(t)| \geqslant (1-\delta)^t \binom{n}{t}.$$

(*ii*) *There is a constant $c_h > 0$ such that for every $t \leqslant F_h(n)$,*

$$|\mathcal{Z}_n^h(t)| \geqslant \left(\frac{c_h n}{t^h}\right)^t.$$

In fact, the deletion argument yielding (*i*) in Proposition 2 can be pushed a little further using the following idea of Kohayakawa, Kreuter, and Steger [22]. Instead of exhibiting a t-element B_h-set in a random subset of $[n]$ with of $t + o(t)$ elements, one may find it in a much larger random subset of $[n]$ with the help of an extension of the powerful result of Ajtai, Komlós, Pintz, Spencer, and Szemerédi [1] on uncrowded hypergraphs due to Duke, Lefmann, and Rödl [11]. Similar ideas were used by the authors in the context of Sidon sets [23]. We postpone the proof of Proposition 3 to Appendix A.

Proposition 3. *For every $h \geqslant 2$ and $\varepsilon > 0$, there are positive constants c_h and C_h such that for all sufficiently large n and t satisfying $\varepsilon n^{1/(2h-1)} \leqslant t \leqslant c_h(n \log n)^{1/(2h-1)}$,*

$$|\mathcal{Z}_n^h(t)| \geqslant \exp\left(-\frac{C_h t^{2h-1}}{n}\right)^t \cdot \binom{n}{t}.$$

The assertions of Propositions 2 and 3 may be summarized as follows. If $t \ll n^{1/(2h-1)}$, then B_h-sets constitute a sizeable $(1-o(1))^t$-proportion of all t-element subsets of $[n]$ and if $t \gg (n \log n)^{1/(2h-1)}$, then we only know that this proportion is at least (merely) $(c_h'/t^{h-1})^t$ for some constant $c_h' > 0$. It turns out that the ratio of $|\mathcal{Z}_n^h(t)|$ to $\binom{n}{t}$ undergoes a dramatic change when t is around $(n \log n)^{1/(2h-1)}$. A fairly straightforward corollary of the so-called container theorems proved independently by Balogh, Morris, and Samotij [2] and by Saxton and Thomason [33] (applied to the $2h$-uniform hypergraph of solutions to the equation $a_1 + \ldots + a_h = b_1 + \ldots + b_h$ which are contained in $[n]$) is that when $t \gg n^{1/(2h-1)}$, then $|\mathcal{Z}_n^h(t)| \leqslant \big(o(1)\big)^t \binom{n}{t}$. The main result of [10] is that a much stronger estimate, $|\mathcal{Z}_n^h(t)| \leqslant (c_h n/t^h)^t$ for some constant $c_h > 0$, holds under the stronger assumption that $t \geqslant n^{1/(h+1)}(\log n)^2$, which matches the lower bound given by Proposition 2. We conjectured in [10] that this best-possible estimate continues to hold (up to a $t^{o(t)}$ multiplicative factor) under the much weaker (and almost optimal) assumption that $t \geqslant n^{1/(2h-1)+o(1)}$. In the current work, we prove this conjecture, determining $|\mathcal{Z}_n^h(t)|$ up to a multiplicative factor of $t^{o(t)}$ for almost all t.

Theorem 4 (Main result)**.** *For every $h \geqslant 2$ and $\varepsilon > 0$ and all sufficiently large integers n and $t \geqslant n^{1/(2h-1)+\varepsilon}$, we have*

$$|\mathcal{Z}_n^h(t)| \leqslant \left(\frac{n}{t^{h-\varepsilon}}\right)^t. \tag{3}$$

As a consequence of Theorem 4 and Proposition 2, we have the following corollary.

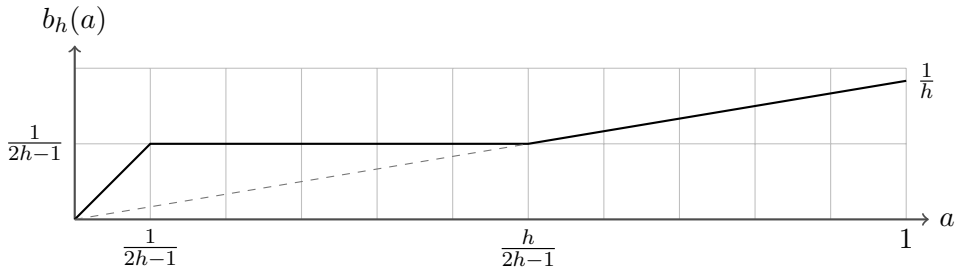

FIGURE 1. The graph corresponding to the piece-wise linear function $b_h(a)$ of Theorem 6.

Corollary 5. *Fix any $\varepsilon > 0$. The asymptotic behavior of $|\mathcal{Z}_n^h(t)|$ is given by*

$$\log|\mathcal{Z}_n^h(t)| = \begin{cases} \big(1+o(1)\big)t\log(n/t) & \text{if } t \ll n^{1/(2h-1)}, \\ t\log\big(n/t^{h-\varepsilon}\big) & \text{if } t \geqslant n^{1/(2h-1)+\varepsilon}. \end{cases}$$

Note that there is an n^{ε} gap in the threshold above as well as in the actual asymptotic estimate. It would be interesting to obtain precise asymptotics for the logarithm of $|\mathcal{Z}_n^h(t)|$ for all $t \geqslant n^{1/(2h-1)}$ (see Conjecture 27).

1.1. **Largest B_h-sets contained in random sets of integers.** In recent years, a major trend in probabilistic combinatorics has been to prove 'sparse random' analogues of classical results in extremal combinatorics and additive number theory. This trend was initiated around twenty years ago with the work of Haxell, Kohayakawa, Łuczak, and Rödl [17, 18, 24, 25] and recently culminated in the breakthrough work of Conlon and Gowers [8] and Schacht [34], which provides general tools for 'transferring' extremal and structural results from the dense to the sparse random environment. This trend provides strong motivation for our work on Sidon sets and B_h-sets, including this paper, owing to the fact that estimating $|\mathcal{Z}_n^h(t)|$ is very closely tied to the problem of determining the maximum size of a B_h-set contained in a sparse random set of integers.

Given a set R of integers, let $F_h(R)$ denote the maximum size of a B_h-set contained in R. Note that this definition generalizes the one made earlier, as $F_h(n) = F_h([n])$. Let $[n]_m$ be a uniformly chosen random m-element subset of $[n]$. We want to study the distribution of the random variable $F_h([n]_m)$ for all m. A standard deletion argument implies that with probability tending to 1 as $n \to \infty$, or *asymptotically almost surely* (a.a.s. for short), we have

$$F_h([n]_m) = (1+o(1))m \quad \text{if } m = m(n) \ll n^{1/(2h-1)}.$$

On the other hand, the transference theorems of Schacht [34] and Conlon and Gowers [8] imply that, a.a.s.,

$$F_h([n]_m) = o(m) \quad \text{if } m = m(n) \gg n^{1/(2h-1)}.$$

These two observations were the starting point in [10, 23], where much more precise information on $F_h([n]_m)$ was provided. As a consequence of Theorem 4, we may now describe the exact behavior (up to $n^{o(1)}$ factors) of $F_h([n]_m)$ for the whole range of m.

Theorem 6. *Let $h \geqslant 2$ be given and set, for any $a \in [0,1]$,*

$$b_h(a) = \begin{cases} a, & \text{if } 0 \leqslant a \leqslant 1/(2h-1), \\ 1/(2h-1), & \text{if } 1/(2h-1) \leqslant a \leqslant h/(2h-1), \\ a/h, & \text{if } h/(2h-1) \leqslant a \leqslant 1. \end{cases}$$

Then for every $m = m(n) = n^a$ we have, a.a.s.,

$$F_h([n]_m) = n^{b_h(a)+o(1)}.$$

Proof sketch. The upper bound on $F_h([n]_m)$ follows from a simple counting argument, namely, for any t, the probability that $F_h([n]_m) \geqslant t$ is at most

$$|\mathcal{Z}_n^h(t)| \cdot \binom{n-t}{m-t}\binom{n}{m}^{-1}.$$

Our main result, Theorem 4, shows that for any $t \geqslant n^{1/(2h-1)+\varepsilon}$ ($\varepsilon > 0$), the above expression becomes $o(1)$ when $m < t^{h-\delta}$ ($\delta = \delta(\varepsilon) > 0$). This translates to the claimed upper bound on $F_h([n]_m)$ when $m = n^a$ for some $h/(2h-1) \leqslant a \leqslant 1$. When $m \leqslant n^{1/(2h-1)}$, the claimed upper bound follows from the trivial bound $F_h([n]_m) \leqslant m$. Finally, when $m = n^a$ for some $1/(2h-1) \leqslant a \leqslant h/(2h-1)$, the claimed upper bound follows from the monotonicity of $b_h(a)$ with respect to a.

The lower bounds on $F_h([n]_m)$ asserted in the theorem were already proved in [10, Theorem 2.5] and therefore we omit their proofs here. □

2. Proof outline

We devote this section to an overview of the high level structure of the proof of our main result, Theorem 4.

Let us start by recalling the general strategy for proving upper bounds on $|\mathcal{Z}_n^h(t)|$ that was used in [10, 23]. The high-level idea there was to bound the number of sets of size $t-s$ one can add to a given B_h-set of size s (the 'seed' set) so that the resulting t-element set still has the B_h property (here, one has to consider a suitable size s for the seed set). Having achieved this, one may derive a bound on $|\mathcal{Z}_n^h(t)|$ by summing over the choices one has for the seed set.

It will be convenient to explain the high level view of the proof strategy for the case $h = 3$ and leave the several details and additional complications of the general case for later sections. Keep in mind that this section is very informal and often these intuitive descriptions only serve as rough approximations of rather long and technical definitions and proofs.

Collision graph and independent sets. Observe that if two distinct elements $x, y \in [n] \setminus S$ satisfy

$$x - y = a_1 + a_2 - b_1 - b_2, \quad \text{for some } \{a_1, a_2\}, \{b_1, b_2\} \in \binom{S}{2}, \tag{4}$$

then $S \cup \{x, y\}$ is clearly not a B_3-set and hence x and y cannot simultaneously belong to any $T \in \mathcal{Z}_n^3(t)$ with $T \supset S$. This motivates our next definition.

Definition 7 (Collision graph $\mathrm{CG}_S^{(3)}$)**.** Let S be a B_3-set. Denote by $\mathrm{CG}_S^{(3)}$ the graph on the vertex set $[n]$ whose edges are all pairs of distinct elements $x, y \in [n]$ that satisfy (4).

The above observation is equivalent to noting that $T \setminus S$ must be an independent set in the collision graph $\mathrm{CG}_S^{(3)}$. Therefore, the number of extensions of S to a B_3 set of cardinality t is not larger than the number of independent sets in $\mathrm{CG}_S^{(3)}$ with cardinality $t - |S|$. The number of such independent sets can be bounded with the use of the following lemma, implicit in the work of Kleitman and Winston [20] (see also the survey [32]), which provides an upper bound on the number of independent sets in graphs that have many edges in each sufficiently large vertex subset. A proof of this lemma is given in [10].

Lemma 8. *Let G be a graph on N vertices, let q be an integer, and let $0 \leqslant \beta \leqslant 1$ and R be real numbers satisfying*

$$R \geqslant e^{-\beta q} N. \tag{5}$$

Suppose that

$$e_G(A) \geqslant \beta \binom{|A|}{2} \text{ for every } A \subset V(G) \text{ with } |A| \geqslant R. \tag{6}$$

Then, for all integers $m \geqslant 0$, the number of independent sets of cardinality $q + m$ in G is at most

$$\binom{N}{q}\binom{R}{m}. \tag{7}$$

Lemma 8 effectively reduces the problem of counting extensions of S into larger B_3-sets to the problem of verifying that $\mathrm{CG}_S^{(3)}$ satisfies condition (6) for appropriately chosen q, m, R, and β. To get the most out of Lemma 8 we take $q \ll m$ and $R = n/t^{2-3\varepsilon}$.

It is not very difficult to show that for every suitably large set $A \subset [n] \setminus S$ there are many *quadruples* $(x, y, \{a_1, a_2\}, \{b_1, b_2\})$ with x and $y \in A$ that satisfy the equality in (4). This, however, does not immediately imply that $e_{\mathrm{CG}_S^{(3)}}(A)$ is large because a single edge of $\mathrm{CG}_S^{(3)}$ may correspond to many different quadruples. This is where the notion of boundedness (roughly described below) comes into play.

***Bounded* sets (Def. 22)**. A key concept used by our method is that of a *bounded* set. Roughly speaking, we call a B_3-set S *bounded* if, for every $0 \neq w \in \mathbb{Z}$, the number of representations of w of the form $w = a_1 + a_2 - b_1 - b_2$, with $a_1, a_2, b_1, b_2 \in S$, can be "controlled". More specifically, there exists a very dense graph G with vertex set S such that it is possible to find strong upper bounds on the number of representations as above which also satisfy $\{a_1, a_2\}, \{b_1, b_2\} \in E(G)$. (The actual definition is somewhat more involved, but for the purpose of this outline, we shall keep things informal.)

Our so-far informal definition implies that for a bounded set S, and any fixed pair $(x, y) \in S^2$, with $x \neq y$, the number of quadruples $(x, y, \{a_1, a_2\}, \{b_1, b_2\})$ that satisfy (4) with $\{a_1, a_2\}, \{b_1, b_2\} \in E(G)$ is also bounded. The main goal then, is to establish lower bounds on the number of such quadruples. Before that, we must show that every B_3-sets extends some bounded set S.

Containers method [2, 33] **(Thm. 9)**. Let us fix $\varepsilon > 0$ and $t \geqslant n^{1/5+\varepsilon}$ (as in the statement of Theorem 4 with $h = 3$). Consider the family $\mathcal{F} = \mathcal{F}(t) = \mathcal{F}_{\texttt{small}}(t) \cup \mathcal{F}_{\texttt{large}}(t)$ of pairs of sets $(S, \widetilde{S})$ defined as follows:

(1) $\mathcal{F}_{\texttt{large}}(t)$ consists of all pairs $(S, \widetilde{S})$ where $S \subset [n]$ is a bounded B_3-set with precisely $t^{1-\varepsilon}$ elements, and

$$\widetilde{S} = \big\{x \in [n] \setminus S \colon S \cup \{x\} \text{ is a } B_3\text{-set}\big\}.$$

(2) $\mathcal{F}_{\texttt{small}}(t)$ consists of all pairs $(S, \widetilde{S})$ where $S \subset [n]$ is a bounded B_3-set with fewer than $t^{1-\varepsilon}$ elements, and

$$\widetilde{S} = \big\{x \in [n] \setminus S \colon S \cup \{x\} \text{ is a } B_3\text{-set which is } \textit{not} \text{ bounded}\big\}.$$

We first show that for every $T \in \mathcal{Z}_n^3(t)$ there exists some rather large set $S^* \subset T$ which is bounded. If $T \in \mathcal{Z}_n^3(t)$ contains a bounded set S^* with $|S^*| \geqslant t^{1-\varepsilon}$, then there exists $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{large}}(t)$ such that $S \subset S^* \subset T \subset S \cup \widetilde{S}$. On the other hand, if all bounded sets contained in T have cardinality smaller than $t^{1-\varepsilon}$, then there must exist some $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{small}}(t)$ for which $S \subset T \subset S \cup \widetilde{S}$ (indeed, any maximal bounded subset $S \subset T$ can be taken). This argument shows that it is possible to obtain upper bounds for $|\mathcal{Z}_n^3(t)|$ by estimating, for each pair $(S, \widetilde{S}) \in \mathcal{F}$, how many $T \in \mathcal{Z}_n^3(t)$ satisfy $S \subset T \subset S \cup \widetilde{S}$.

Estimating the density of $\mathrm{CG}_S^{(3)}$ by counting paths in an auxiliary graph (Thm. 21). We will use an auxiliary bipartite graph H with color classes A and $[3n]$, in which $(x, w) \in A \times [3n]$ is an edge of H if and only if there exists an edge $\{a_1, a_2\} \in G$ such that $w = x + a_1 + a_2$.

Simpler case. For $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{large}}(t)$, one can prove a lower bound on the number of quadruples as follows. Since G is dense, the number of edges of H is

$$|A| \cdot e(G) \geqslant R \cdot \frac{1}{2}\binom{|S|}{2} = \frac{n}{2t^{2-3\varepsilon}}\binom{t^{1-\varepsilon}}{2} \gg n.$$

In particular, the average degree in the class $[3n]$ is larger than, say, 100. For every $w \in [3n]$ with degree $d \geqslant 2$ we can form $\binom{d}{2}$ quadruples of the form $(x, y, \{a_1, a_2\}, \{b_1, b_2\})$ that satisfy

$$x + a_1 + a_2 = w = y + b_1 + b_2$$

with $\{a_1, a_2\}, \{b_1, b_2\} \in G$ (and thus (4) is satisfied). A simple application of the Cauchy–Schwarz inequality can be used to give a lower bound on the number of quadruples and such a bound is sufficient (in the sense of Lemma 8) to settle the case when $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{large}}(t)$.

Difficult case (Lemma 20). The case when $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{small}}(t)$ is substantially more difficult. Since S itself is small, even if G was the complete graph on S, it is possible that $R \cdot e(G) = o(n)$ and the auxiliary graph H defined above is therefore too sparse to be useful. Our strategy for this case involves studying the additive structure of $\widetilde{S}$. Indeed, we are able show that there is some rather small set $W = W(S) \subset [3n]$ such that for every $x \in \widetilde{S}$ there are many $w \in W$ such that $w = x + a_1 + a_2$ for some $\{a_1, a_2\} \in G$. This means that either

(a) $|\widetilde{S}|$ is small and, given that we need to count sets T with $S \subset T \subset S \cup \widetilde{S}$, we may bound the number of extensions trivially by $\binom{|\widetilde{S}|}{t-|S|}$ or

(b) $|\widetilde{S}|$ is large, which means that the induced subgraph $H[\widetilde{S} \cup W]$ is such that the average degree of W is large, and thus the argument used in the simpler case above applies. Indeed when $|\widetilde{S}|$ is large we invoke Theorem 9 below (see (9)).

For the general case $h \geqslant 3$, the boundedness condition becomes substantially more technical (in particular, we now have to deal with hypergraphs). The proofs are also more difficult, in particular due to the fact that for the case $h = 3$, one can leverage the definition of a B_3-set to control the alternating sums of length four. On the other hand, for larger values of h, the generalization of (4) requires us to control alternating sums of length $2(h-1)$. The gap between h and $2(h-1)$ becomes non-trivial after $h > 3$ and demands a much more careful treatment. Moreover, the estimates on quadruples as outlined above are in reality quite involved, and we devote the entirety of Section 6 to establishing them.

2.1. **Notation.** For an integer x and a set $A \subset \mathbb{Z}$, let us use the following notation:

$$x \boxplus A = x + \sum_{a \in A} a \qquad \text{and} \qquad x \boxminus A = x - \sum_{a \in A} a.$$

Moreover, for a hypergraph $\mathcal{H}$ with $V(\mathcal{H}) \subset \mathbb{Z}$, and integer x, let

$$x \boxplus \mathcal{H} = \big\{x \boxplus e \colon e \in E(\mathcal{H})\big\}. \tag{8}$$

We shall often let $\mathcal{H}$ in the above definition be the complete k-uniform hypergraph with vertex set S, writing

$$x \boxplus \binom{S}{k}.$$

We sometimes write $|G|$ or $|\mathcal{H}|$ to denote the number of edges in a graph G or in a hypergraph $\mathcal{H}$. We also abuse the notation and denote by $e \in G$ ($e \in \mathcal{H}$) the fact that e is an edge of the graph (hypergraph). For the sake of clarity of our presentation, we write 'k-graph' instead of 'k-uniform hypergraph'. More notation will be introduced and used locally when needed.

2.2. **Proof summary.** Recall that given $S \subset [n]$, we have defined the collision graph $\mathrm{CG}_S^{(3)}$ as the graph with vertex set $[n]$ whose edges are all pairs of distinct elements $x, y \in [n]$ that satisfy (4). More generally, for all $h \geq 2$, the edges of $\mathrm{CG}_S^{(h)}$ are formed by all pairs of distinct $x, y \in S$ that satisfy

$$x \boxplus A = y \boxplus B \quad \text{for some } A, B \in \binom{S}{h-1}.$$

As we have already observed above, if T is a B_h-set, then for every $S \subset T$, the set $T \setminus S$ is independent in the graph $\mathrm{CG}_S^{(h)}$. We shall show that Theorem 4 follows from Lemma 8 and the following statement, whose proof constitutes most of the remainder of this paper.

For convenience, from now on we will omit the superscript in $\mathrm{CG}_S^{(h)}$.

Theorem 9. *For every $h \geqslant 2$ and $\delta \in (0, 1/2)$ the following is true for all sufficiently large n. Suppose that $n^{1/(2h-1)+\delta} \leqslant t \leqslant 2hn^{1/(2h-1)+\delta}$.*

There exists a family $\mathcal{F}$ of pairs of sets $(S, \widetilde{S})$ with $S, \widetilde{S} \subset [n]$ and $|S| \leqslant t^{1-\delta}$ that has the following property. For every $T \in \mathcal{Z}_n^h(t)$, there is $(S, \widetilde{S}) \in \mathcal{F}$ such that $S \subset T \subset S \cup \widetilde{S}$ and the graph CG_S satisfies

$$e_{\mathrm{CG}_S}(A) \geqslant \frac{1}{t^{1-\delta/2}} \binom{|A|}{2} \quad \text{for every } A \subset \widetilde{S} \text{ with } |A| \geqslant \frac{n}{t^{h-1-8h^2\delta}}. \tag{9}$$

Moreover, for each set S, there is at most one $\widetilde{S}$ such that $(S, \widetilde{S}) \in \mathcal{F}$, and thus $|\mathcal{F}| < n^{t^{1-\delta}}$.

We postpone the (fairly straightforward) derivation of Theorem 4 to §5.2 and focus on Theorem 9 instead. First, we need several additional definitions which generalize the concepts already introduced in [9].

In analogy with the simplified outline given earlier in this section for $h = 3$, we would like to generalize the concept of boundedness in terms of a dense $(h-1)$-graph $\mathcal{G} \subset \binom{S}{h-1}$ whose edges satisfy a certain condition. It turns out that hypergraphs of all uniformities $\{2, 3, \ldots, h-1\}$ will be needed in the course of the proof. Therefore, the definitions that follow include such a sequence of hypergraphs.

Definition 10 (Representation count)**.** For a k-graph $\mathcal{G}$ and an ℓ-graph $\mathcal{H}$ with $V(\mathcal{G}), V(\mathcal{H}) \subset [n]$ and an integer z, we let $R_{\mathcal{G},\mathcal{H}}(z)$ be the number of pairs $(e, f) \in \mathcal{G} \times \mathcal{H}$ that satisfy

$$z = e \boxminus f \qquad \text{and} \qquad e \cap f = \emptyset.$$

Moreover, let

$$\|R_{\mathcal{G},\mathcal{H}}\| = \max_z R_{\mathcal{G},\mathcal{H}}(z).$$

For brevity, we shall often write $R_{\mathcal{G}}$ for $R_{\mathcal{G},\mathcal{G}}$.

Remark 11. If S is a B_h-set, then for any k and ℓ with $k + \ell \leqslant h$ we must have $\big\|R_{\binom{S}{k},\binom{S}{\ell}}\big\| = 1$.

Indeed, if there were some w for which $R_{\binom{S}{k},\binom{S}{\ell}}(w) \geqslant 2$, then one could obtain distinct $a_1, \ldots, a_k$, $b_1, \ldots, b_\ell \in S$, and distinct $c_1, \ldots, c_k$, $d_1, \ldots, d_\ell \in S$ such that

$$a_1 + \cdots + a_k - (b_1 + \cdots + b_\ell) = w = c_1 + \cdots + c_k - (d_1 + \cdots + d_\ell).$$

Since S is obviously a $B_{k+\ell}$-set as well, the last equality implies

$$\{a_1, \ldots, a_k, d_1, \ldots, d_\ell\} = \{b_1, \ldots, b_\ell, c_1, \ldots, c_k\},$$

which forces $\{a_1, \ldots, a_k\} = \{c_1, \ldots, c_k\}$ and $\{b_1, \ldots, b_\ell\} = \{d_1, \ldots, d_\ell\}$, thus showing that the representations are in fact identical (up to a permutation of the labels).

Definition 12 (Collision multigraph)**.** Given an $(h-1)$-graph $\mathcal{G}$ with $V(\mathcal{G}) \subset [n]$, let $\widetilde{\mathrm{CG}}_{\mathcal{G}}$ be the multigraph on the vertex set $[n]$, where the multiplicity of each pair $x, y \in [n]$ equals $R_{\mathcal{G}}(x - y)$.

Observe that for every $S \subset [n]$ and every $(h-1)$-graph $\mathcal{G}$ with $V(\mathcal{G}) = S$, the set of pairs with non-zero multiplicity in $\widetilde{\mathrm{CG}}_{\mathcal{G}}$ is a subgraph of CG_S (and in fact, it is equal to CG_S when $\mathcal{G}$ is the complete $(h-1)$-graph on S). Moreover, for every $A \subset [n]$,

$$e_{\widetilde{\mathrm{CG}}_{\mathcal{G}}}(A) \leqslant \|R_{\mathcal{G}}\| \cdot e_{\mathrm{CG}_S}(A), \tag{10}$$

where $e_{\widetilde{\mathrm{CG}}_{\mathcal{G}}}(A)$ counts pairs of vertices of A with their multiplicities in $\widetilde{\mathrm{CG}}_{\mathcal{G}}$. In view of (10), a natural approach to proving a strong lower bound on $e_{\mathrm{CG}_S}(A)$ is to construct an $(h-1)$-graph $\mathcal{G}$ with $V(\mathcal{G}) = S$ for which the ratio $e_{\widetilde{\mathrm{CG}}_{\mathcal{G}}}(A)/\|R_{\mathcal{G}}\|$ is large.

Since there seems to be no easy way of controlling $\|R_{\mathcal{G}}\|$ 'directly', similarly as in [9], we shall instead maintain an upper bound on the *moment generating function* of $R_{\mathcal{G}}$, defined as follows.

Definition 13 (Moment generating function of $R_{\mathcal{G},\mathcal{H}}$)**.** Given a k-graph $\mathcal{G}$ and an ℓ-graph $\mathcal{H}$ with $V(\mathcal{G}), V(\mathcal{H}) \subset [n]$ and a positive real λ, we let

$$Q_{\mathcal{G},\mathcal{H}}(\lambda) = \sum_{z=-\ell n}^{kn} \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big). \tag{11}$$

Note that the range of the above sum includes all z for which $R_{\mathcal{G},\mathcal{H}}(z) \neq 0$. Note also that $R_{\mathcal{G},\mathcal{H}}(z) = R_{\mathcal{H},\mathcal{G}}(-z)$ and thus $Q_{\mathcal{G},\mathcal{H}} = Q_{\mathcal{H},\mathcal{G}}$.

One reason why we are interested in the moment generating function is the following trivial relationship between $R_{\mathcal{G},\mathcal{H}}$ and $Q_{\mathcal{G},\mathcal{H}}(\lambda)$.

Remark 14. For every $\lambda > 0$,

$$\|R_{\mathcal{G},\mathcal{H}}\| = \max_z R_{\mathcal{G},\mathcal{H}}(z) \leqslant \frac{1}{\lambda} \log Q_{\mathcal{G},\mathcal{H}}(\lambda). \tag{12}$$

To explore the above relationship between $\|R_{\mathcal{G},\mathcal{H}}\|$ and $Q_{\mathcal{G},\mathcal{H}}(\lambda)$ we will develop some auxiliary results in Sections 3.

Even though we are mainly interested in $\|R_{\mathcal{G}}\|$ when $\mathcal{G}$ is an $(h-1)$-graph, it will be necessary to involve hypergraphs of different uniformities since the smaller hypergraphs are needed in several parts of the proof, as can be evidenced by the following observation.

Observation 15. Let $\mathcal{G} \subset \binom{S}{h-1}$, $x \in [n] \setminus S$, and $\widehat{\mathcal{G}} \subset \binom{S\cup\{x\}}{h-1}$ be such that $\mathcal{G} = \widehat{\mathcal{G}}[S]$. Denote by $\mathcal{N}_x$ the neighborhood of x in $\widehat{\mathcal{G}}$, that is $\big\{f \setminus \{x\} \colon f \in \widehat{\mathcal{G}}, x \in f\big\}$. Then, for any integer z,

$$R_{\widehat{\mathcal{G}}}(z) = R_{\mathcal{G}}(z) + R_{\mathcal{G},\mathcal{N}_x}(z+x) + R_{\mathcal{N}_x,\mathcal{G}}(z-x). \tag{13}$$

Indeed, from Definition 10 of $R_{\widehat{\mathcal{G}}}(z)$ we have,

$$\begin{aligned} R_{\widehat{\mathcal{G}}}(z) = R_{\mathcal{G}}(z) + {} & \\ & \big|\big\{(e,f) \in \mathcal{G} \times (\widehat{\mathcal{G}} \setminus \mathcal{G}) \colon e \boxminus f = z, e \cap f = \emptyset\big\}\big| + \\ & \big|\big\{(e,f) \in (\widehat{\mathcal{G}} \setminus \mathcal{G}) \times \mathcal{G} \colon e \boxminus f = z, e \cap f = \emptyset\big\}\big| + \\ & \big|\big\{(e,f) \in (\widehat{\mathcal{G}} \setminus \mathcal{G})^2 \colon e \boxminus f = z, e \cap f = \emptyset\big\}\big|. \end{aligned}$$

The last term is zero since $e, f \in \widehat{\mathcal{G}} \setminus \mathcal{G}$ must be such that $x \in e \cap f$. We also have

$$\begin{aligned} R_{\mathcal{G},\mathcal{N}_x}(z+x) &= \big|\big\{(e,f') \in \mathcal{G} \times \mathcal{N}_x \colon e \boxminus f' = z+x, e \cap f' = \emptyset\big\}\big| \\ &= \big|\big\{(e,f) \in \mathcal{G} \times (\widehat{\mathcal{G}} \setminus \mathcal{G}) \colon e \boxminus f = z, e \cap f = \emptyset\big\}\big|. \end{aligned}$$

Similarly we obtain the term $R_{\mathcal{N}_x,\mathcal{G}}(z-x)$ in (13).

The above observation indicates that in order to estimate $\|R_{\mathcal{G}}\|$ for an $(h-1)$-graph it may be necessary to also estimate $\|R_{\mathcal{G},\mathcal{H}}\|$ with $\mathcal{H}$ of smaller uniformity. Inductively, bounds for the representation count among hypergraphs of all possible uniformities in $\{2, \ldots, h-1\}$ are needed.

This motivates the following definition. First, given a positive integer m, let H_m denote the m^{th} harmonic number, that is,

$$\mathsf{H}_m = \sum_{j=1}^{m} \frac{1}{j}$$

and recall that $0 \leqslant \mathsf{H}_m - \log m \leqslant 1$ for every m.

Definition 16. Let h and $n \geqslant 2$ be integers, let $\alpha \in [0,1)$, and let $\lambda > 0$. We shall say that a set $S \in \mathcal{Z}_n^h$ satisfies property $\mathcal{P}_h(\lambda, \alpha)$ if there exist hypergraphs $\mathcal{G}^{(k)} \subset \binom{S}{k}$ for each $k \in [h-1]$ such that

(*a*) $|\mathcal{G}^{(k)}| \geqslant (1 - 2^k\alpha)\binom{|S|}{k}$ and

(*b*) for all k and $\ell \in [h-1]$,

$$Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \leqslant (2hn+1) \cdot \exp(\mathsf{H}_{|S|}), \tag{14}$$

where

$$\xi_j = (2\log n)^{-j} \quad \text{for all integers } j. \tag{15}$$

Remark 17. Note that if $\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(h-1)}$ satisfy condition (*b*) of the above definition, then

$$\|R_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}\| \overset{(12)}{\leqslant} \frac{\log Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell})}{\lambda \cdot \xi_{k+\ell}} \overset{(14)}{\leqslant} \frac{\mathsf{H}_{|S|} + \log(2hn+1)}{\lambda \cdot \xi_{k+\ell}} \leqslant \frac{2\log n}{\lambda \cdot \xi_{k+\ell}} \overset{(15)}{=} \frac{1}{\lambda \cdot \xi_{k+\ell+1}}.$$

In words, λ is a parameter that can be adjusted to directly control the bounds on the maximum representation counts.

Finally, given a set $S \subset [n]$ satisfying $\mathcal{P}_h(\lambda, \alpha)$, we let

$$\widetilde{S}_{\lambda,\alpha} = \big\{x \in [n] \setminus S \colon S \cup \{x\} \text{ is a } B_h\text{-set that does not satisfy } \mathcal{P}_h(\lambda,\alpha)\big\}. \tag{16}$$

2.3. **Organization of the proof.** In Section 3 we prove two technical lemmas that explain how adding a single vertex (together with edges containing it) to a hypergraph affects the moment generating function $Q_{\mathcal{G},\mathcal{H}}$. We then show in Section 4 that each pair of sets $(S, \widetilde{S}_{\lambda,\alpha})$ defined in (16) possesses a certain additive structure. In Section 5 we state a technical result, Theorem 21, which asserts that for every sufficiently dense hypergraph $\mathcal{H} \subset \binom{S}{h-1}$, the multigraph $\widetilde{\mathrm{CG}}_{\mathcal{H}}$ has many edges in each large subset of $\widetilde{S}_{\lambda,\alpha}$. We then use this technical theorem to prove Theorem 9. A fairly straightforward derivation of our main result, Theorem 4, from Theorem 9 is presented in §5.2. The fairly long and technical proof of Theorem 21 is postponed to Section 6. The flow of the proof of Theorem 4 is given in Figure 2.

3. Extension lemmas

In this section we prove two technical lemmas that we shall later use to bound the moment generating functions $Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\cdot)$. The first lemma shows that if we extend two hypergraphs $\mathcal{G}$ and $\mathcal{H}$ to form $\widehat{\mathcal{G}}$ and $\widehat{\mathcal{H}}$ by adding to them a single vertex, then we may bound the increase of the moment function, $Q_{\widehat{\mathcal{G}},\widehat{\mathcal{H}}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda)$, in terms of the increases of the moment function caused by extending $\mathcal{G}$ and $\mathcal{H}$ separately, that is, $Q_{\widehat{\mathcal{G}},\mathcal{H}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda)$ and $Q_{\mathcal{G},\widehat{\mathcal{H}}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda)$, provided that the neighborhoods of the new vertex in $\mathcal{G}$ and $\mathcal{H}$ are two 'well-behaved' hypergraphs $\mathcal{N}$ and $\mathcal{M}$, respectively.

Lemma 18. *Let k and $\ell \geqslant 2$ be integers and let $\lambda > 0$. Suppose that*

- *$\mathcal{G}$ is a k-graph and $\mathcal{N}$ is a $(k-1)$-graph with $V(\mathcal{G}) = V(\mathcal{N}) \subset [n]$,*

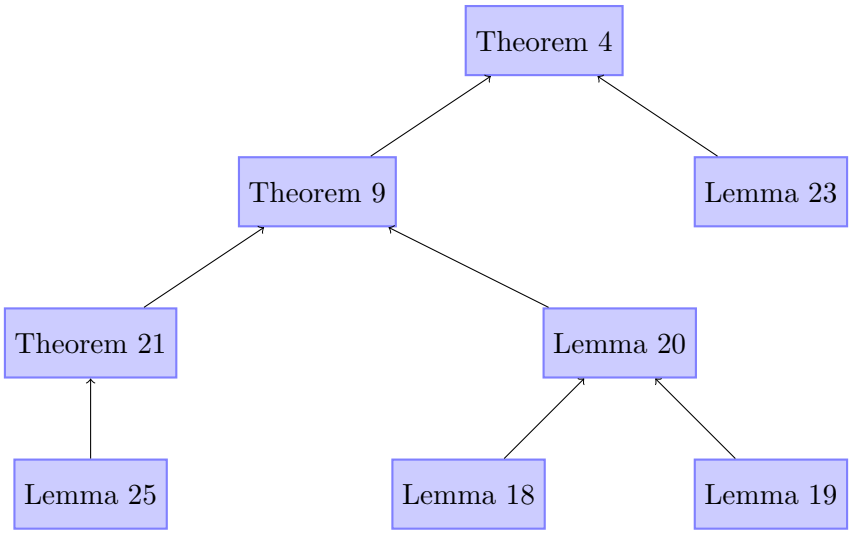

FIGURE 2. A diagram illustrating the flow of the proof of our main result.

- $\mathcal{H}$ *is an* ℓ*-graph and* $\mathcal{M}$ *is an* $(\ell-1)$*-graph with* $V(\mathcal{H}) = V(\mathcal{M}) \subset [n]$*,*
- $\|R_{\mathcal{N},\mathcal{H}}\|, \|R_{\mathcal{G},\mathcal{M}}\| \leqslant 1/\lambda$*,*
- x *is an arbitrary element of* $[n]$ *not in* $V(\mathcal{G}) \cup V(\mathcal{H})$*.*

Then the hypergraphs $\widehat{\mathcal{G}}$ *and* $\widehat{\mathcal{H}}$ *defined by*

$$\widehat{\mathcal{G}} = \mathcal{G} \cup \big\{\{x\} \cup e \colon e \in \mathcal{N}\big\} \quad \text{and} \quad \widehat{\mathcal{H}} = \mathcal{H} \cup \big\{\{x\} \cup f \colon f \in \mathcal{M}\big\}$$

satisfy

$$Q_{\widehat{\mathcal{G}},\widehat{\mathcal{H}}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda) \leqslant 2\left(\big(Q_{\widehat{\mathcal{G}},\mathcal{H}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda)\big) + \big(Q_{\mathcal{G},\widehat{\mathcal{H}}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda)\big)\right). \tag{17}$$

Proof. With the same argument used in Observation 15 one can establish that

$$R_{\widehat{\mathcal{G}},\widehat{\mathcal{H}}}(z) = R_{\mathcal{G},\mathcal{H}}(z) + R_{\mathcal{N},\mathcal{H}}(z-x) + R_{\mathcal{G},\mathcal{M}}(z+x) \tag{18}$$

for every integer z. Now, let

$$\begin{aligned} N(z) &= R_{\mathcal{N},\mathcal{H}}(z-x) = R_{\widehat{\mathcal{G}},\mathcal{H}}(z) - R_{\mathcal{G},\mathcal{H}}(z), \\ M(z) &= R_{\mathcal{G},\mathcal{M}}(z+x) = R_{\mathcal{G},\widehat{\mathcal{H}}}(z) - R_{\mathcal{G},\mathcal{H}}(z) \end{aligned} \tag{19}$$

and observe that

$$Q_{\widehat{\mathcal{G}},\mathcal{H}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda) = \sum_{z=-\ell n}^{kn} \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \Big(\underbrace{\exp\big(\lambda \cdot N(z)\big) - 1}_{a}\Big), \tag{20}$$

$$Q_{\mathcal{G},\widehat{\mathcal{H}}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda) = \sum_{z=-\ell n}^{kn} \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \Big(\underbrace{\exp\big(\lambda \cdot M(z)\big) - 1}_{b}\Big). \tag{21}$$

Since (18) holds, we have

$$\exp\big(\lambda \cdot R_{\widehat{\mathcal{G}},\widehat{\mathcal{H}}}(z)\big) = \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \exp\big(\lambda \cdot N(z)\big) \cdot \exp\big(\lambda \cdot M(z)\big), \tag{22}$$

and hence,

$$Q_{\widehat{\mathcal{G}},\widehat{\mathcal{H}}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda) = \sum_{z=-\ell n}^{kn} \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \Big(\underbrace{\exp\big(\lambda \cdot N(z)\big) \cdot \exp\big(\lambda \cdot M(z)\big) - 1}_{ab+a+b} \Big).$$

Therefore, in order to establish (17), it is enough to show that for every z,

$$\underbrace{\exp\big(\lambda \cdot N(z)\big) \cdot \exp\big(\lambda \cdot M(z)\big) - 1}_{ab+a+b} \leqslant 2 \cdot \Big(\underbrace{\exp\big(\lambda \cdot N(z)\big) - 1}_{a} + \underbrace{\exp\big(\lambda \cdot M(z)\big) - 1}_{b} \Big).$$

To prove the above inequality, first let $a = \exp\big(\lambda \cdot N(z)\big) - 1$ and $b = \exp\big(\lambda \cdot M(z)\big) - 1$ and notice that the inequality becomes $ab + a + b \leqslant 2(a + b)$, or simply $ab \leqslant a + b$.

Our assumption that $\|R_{\mathcal{N},\mathcal{H}}\|, \|R_{\mathcal{G},\mathcal{M}}\| \leqslant 1/\lambda$, together with (19) imply that

$$0 \leqslant \lambda N(z) = \lambda R_{\mathcal{N},\mathcal{H}}(z - x) \leqslant \lambda \, \|R_{\mathcal{N},\mathcal{H}}\| \leqslant 1,$$

and similarly, $0 \leqslant \lambda M(z) \leqslant 1$. This means that $a, b \in [0, e - 1] \subset [0, 2]$. In particular, $a + b \leqslant 4$. Consquently, by the AM–GM inequality: $ab \leqslant \big(\frac{a+b}{2}\big)^2 = (a + b)\frac{a+b}{4} \leqslant a + b$. □

Our second lemma shows how one can extend a hypergraph by adding one vertex together with edges containing it in a way that causes only a minor increase in the moment function.

Lemma 19. *Let $k \geqslant 2$ and $\ell \geqslant 1$ be integers and let $\lambda > 0$. Suppose that*

- *$\mathcal{G}$ is a k-graph and $\mathcal{N}$ is a $(k-1)$-graph with $V(\mathcal{G}) = V(\mathcal{N}) \subset [n]$,*
- *$\mathcal{H}$ is an ℓ-graph and $V(\mathcal{H}) \subset [n]$ is a B_ℓ-set*[1]*,*
- *$\|R_{\mathcal{N},\mathcal{H}}\| \leqslant 1/\lambda$.*

Then for every integer $M \geqslant 1$, there exists a set $\Gamma \subset [kn]$ with $|\Gamma| \leqslant M$ such that for any $x \in [n] \setminus V(\mathcal{G})$, the k-graph $\widehat{\mathcal{G}}$ on $V(\mathcal{G}) \cup \{x\}$ defined by

$$\widehat{\mathcal{G}} = \mathcal{G} \cup \big\{ \{x\} \cup e \colon e \in \mathcal{N} \text{ and } x \boxplus e \notin \Gamma \big\} \tag{23}$$

satisfies

$$Q_{\widehat{\mathcal{G}},\mathcal{H}}(\lambda) \leqslant Q_{\mathcal{G},\mathcal{H}}(\lambda) \cdot \left(1 + \frac{2\lambda|\mathcal{H}||\mathcal{N}|}{M}\right).$$

Proof. For integers w and z, define

$$I(w, z) = \mathbf{1}\big[z = w \boxminus f \text{ for some } f \in \mathcal{H}\big].$$

and

$$u_w = \sum_{z=-\ell n}^{kn} \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) I(w, z). \tag{24}$$

Set

$$\Gamma = \left\{ w \in [kn] \colon u_w \geqslant \frac{|\mathcal{H}| \cdot Q_{\mathcal{G},\mathcal{H}}(\lambda)}{M} \right\}. \tag{25}$$

Claim 1. $|\Gamma| \leqslant M$.

[1] If $\ell = 1$, then this condition is vacuous as every set of numbers is a B_1-set.

Proof. Observe first that, for every z,

$$\sum_{w\in[kn]} I(w,z) \leqslant |\mathcal{H}|.$$

Indeed, each value w such that $I(w,z)=1$ has some associated $f_w \in \mathcal{H}$ satisfying $w \boxminus f_w = z$, and since we clearly cannot have $f_w = f_{w'}$ for distinct w, w', the inequality follows. Therefore,

$$\sum_{w\in[kn]} u_w = \sum_{z=-\ell n}^{kn} \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \sum_{w\in[kn]} I(w,z) \leqslant |\mathcal{H}| \cdot Q_{\mathcal{G},\mathcal{H}}(\lambda).$$

On the other hand, as $\Gamma \subset [kn]$,

$$\sum_{w\in[kn]} u_w \geqslant |\Gamma| \cdot \frac{|\mathcal{H}| \cdot Q_{\mathcal{G},\mathcal{H}}(\lambda)}{M}.$$

Combining the two previous inequalities completes the proof of the claim. □

Let

$$\mathcal{N}_x = \{e \in \mathcal{N} \colon x \oplus e \notin \Gamma\} \tag{26}$$

and consider the k-graph $\widehat{\mathcal{G}}$ defined in (23), namely

$$\widehat{\mathcal{G}} = \mathcal{G} \cup \big\{\{x\} \cup e \colon e \in \mathcal{N}_x\big\}.$$

Observe that for any integer z,

$$R_{\widehat{\mathcal{G}},\mathcal{H}}(z) \leqslant R_{\mathcal{G},\mathcal{H}}(z) + R_{\mathcal{N}_x,\mathcal{H}}(z-x).$$

It follows that

$$\begin{aligned} \exp\big(\lambda \cdot R_{\widehat{\mathcal{G}},\mathcal{H}}(z)\big) &\leqslant \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \exp\big(\lambda \cdot R_{\mathcal{N}_x,\mathcal{H}}(z-x)\big) \\ &\leqslant \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \big(1 + 2\lambda \cdot R_{\mathcal{N}_x,\mathcal{H}}(z-x)\big), \end{aligned} \tag{27}$$

where the last inequality follows from the fact that $e^x \leqslant 1+2x$ for all $x \in [0,1]$ and our assumption that

$$\lambda \cdot R_{\mathcal{N}_x,\mathcal{H}}(z-x) \leqslant \lambda \cdot \|R_{\mathcal{N}_x,\mathcal{H}}\| \leqslant \lambda \cdot \|R_{\mathcal{N},\mathcal{H}}\| \leqslant 1.$$

Moreover,

$$R_{\mathcal{N}_x,\mathcal{H}}(z-x) \leqslant \sum_{e\in\mathcal{N}_x} \sum_{f\in\mathcal{H}} \mathbf{1}\big[z = (x \boxplus e) \boxminus f\big] = \sum_{e\in\mathcal{N}_x} I(x \boxplus e, z),$$

where the last equality follows because $V(\mathcal{H})$ is a B_ℓ-set and hence for given x, e, and z, there is at most one $f \in \mathcal{H}$ such that $z = (x \boxplus e) \boxminus f$. Consequently, from (27), we have

$$\exp\big(\lambda \cdot R_{\widehat{\mathcal{G}},\mathcal{H}}(z)\big) \leqslant \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \left(1 + 2\lambda \sum_{e\in\mathcal{N}_x} I(x \boxplus e, z)\right).$$

Summing the above inequality over all $z \in [-\ell n, kn]$, and recalling (24) yields

$$Q_{\widehat{\mathcal{G}},\mathcal{H}}(\lambda) \leqslant Q_{\mathcal{G},\mathcal{H}}(\lambda) + 2\lambda \sum_{e\in\mathcal{N}_x} u_{x\boxplus e}.$$

From the definitions of Γ and $\mathcal{N}_x$ (see (25) and (26)) we finally conclude that

$$Q_{\widehat{\mathcal{G}},\mathcal{H}}(\lambda) \leqslant Q_{\mathcal{G},\mathcal{H}}(\lambda) \cdot \left(1 + \frac{2\lambda|\mathcal{H}||\mathcal{N}_x|}{M}\right) \leqslant Q_{\mathcal{G},\mathcal{H}}(\lambda) \cdot \left(1 + \frac{2\lambda|\mathcal{H}||\mathcal{N}|}{M}\right). \qquad \square$$

4. The additive structure of $(S, \widetilde{S}_{\lambda,\alpha})$

In this section we show that if S satisfies property $\mathcal{P}_h(\lambda,\alpha)$ of Definition 16, then the pair $(S, \widetilde{S}_{\lambda,\alpha})$ possesses some stringent additive structure. In particular, one can partition $\widetilde{S}_{\lambda,\alpha}$ into $\bigcup_{k=2}^{h-1} \widetilde{S}_{\lambda,\alpha,k}$ in such a way that the number of elements of the form $x \boxplus e$, with $x \in \widetilde{S}_{\lambda,\alpha,k}$ and $e \in \binom{S}{k-1}$, belonging to a fairly small set $\Gamma_k = \Gamma_k(S)$ is disproportionally large. In later sections, we shall exploit this structure to derive a strong lower bound on $e_{\widetilde{\mathrm{CG}}_{\mathcal{H}}}(A)$ for all sufficiently large $A \subset \widetilde{S}_{\lambda,\alpha}$ and every sufficiently dense $\mathcal{H} \subset \binom{S}{h-1}$.

Our argument in the next lemma can be summarized as follows. Since S has property $\mathcal{P}_h(\lambda,\alpha)$, there are some $\mathcal{G}^{(1)}, \dots, \mathcal{G}^{(h-1)}$ with vertex set S satisfying (a) and (b) of Definition 16. Given an $x \in \widetilde{S}_{\lambda,\alpha}$, using Lemmas 18 and 19 from Section 3, we shall extend each $\mathcal{G}^{(k)}$ to a $\widehat{\mathcal{G}}^{(k)} \subset \binom{S\cup\{x\}}{k}$ so that condition (b) of Definition 16 is satisfied. By the definition of $\widetilde{S}_{\lambda,\alpha}$ (see (16)), some $\widehat{\mathcal{G}}^{(k)}$ must fail condition (a). In particular, the degree of x in the construction must be small. By the definition of the extension (see (23)) we derive the conclusion of the lemma.

Lemma 20. *Let $\lambda \in (0,1]$, let $\alpha \in [0,1]$, and suppose that a set $S \in \mathcal{Z}_n^h$ satisfies property $\mathcal{P}_h(\lambda,\alpha)$. Then there exist sets $\Gamma_2, \dots, \Gamma_{h-1} \subset [hn]$ with the following properties:*

- *(i) $|\Gamma_k| \leqslant (|S|+1)^{k+h-1} \cdot \lambda$, for every $k \in \{2,\dots,h-1\}$.*
- *(ii) For every $x \in \widetilde{S}_{\lambda,\alpha}$ there is some $k \in \{2,\dots,h-1\}$ such that*

$$\left|x \boxplus \binom{S}{k-1} \cap \Gamma_k\right| \geqslant 2^{k-1}\alpha\binom{|S|}{k-1}.$$

Proof. Let λ, α, and S be as in the statement of the lemma and fix arbitrary hypergraphs $\mathcal{G}^{(1)}, \dots, \mathcal{G}^{(h-1)}$ that satisfy conditions (a) and (b) of Definition 16. In particular, it follows from Remark 17 that for every $k \in \{2,\dots,h-1\}$ and $\ell \in [h-1]$, we have $\|R_{\mathcal{G}^{(k-1)},\mathcal{G}^{(\ell)}}\| \leqslant 1/(\lambda \cdot \xi_{k+\ell})$, and hence we may apply Lemma 19 with

$$\mathcal{G} = \mathcal{G}^{(k)}, \quad \mathcal{N} = \mathcal{G}^{(k-1)}, \quad \mathcal{H} = \mathcal{G}^{(\ell)}, \quad \lambda = \lambda \cdot \xi_{k+\ell}, \quad M = 8(|S|+1)^{k+\ell} \cdot \lambda \cdot \xi_{k+\ell}$$

to obtain a set $\Gamma_{k,\ell} \subset [kn]$ with

$$|\Gamma_{k,\ell}| \leqslant 8(|S|+1)^{k+\ell} \cdot \lambda \cdot \xi_{k+\ell} \tag{28}$$

such that for any $x \in [n] \setminus S$ the k-graph $\widehat{\mathcal{G}}_\ell^{(k)}(x)$ defined by

$$\widehat{\mathcal{G}}_\ell^{(k)}(x) = \mathcal{G}^{(k)} \cup \big\{\{x\} \cup e \colon e \in \mathcal{G}^{(k-1)} \text{ and } x \boxplus e \notin \Gamma_{k,\ell}\big\} \tag{29}$$

satisfies

$$\begin{aligned} Q_{\widehat{\mathcal{G}}_\ell^{(k)}(x),\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) &\leqslant Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \cdot \left(1 + \frac{|\mathcal{G}^{(\ell)}| \cdot |\mathcal{G}^{(k-1)}|}{4(|S|+1)^{k+\ell}}\right) \\ &\leqslant Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \cdot \left(1 + \frac{1}{4(|S|+1)}\right). \end{aligned} \tag{30}$$

For each $k \in \{2, \dots, h-1\}$, we let

$$\Gamma_k = \bigcup_{\ell=1}^{h-1} \Gamma_{k,\ell} \tag{31}$$

and observe that (28) implies that condition (i) from the statement of this lemma is satisfied; see the definition of ξ_j in (15).

Now, fix some $x \in \widetilde{S}_{\lambda,\alpha}$, let $\widehat{\mathcal{G}}^{(1)} = \mathcal{G}^{(1)} \cup \big\{\{x\}\big\}$, and define for each $k \in \{2, \dots, h-1\}$,

$$\widehat{\mathcal{G}}^{(k)} = \bigcap_{\ell=1}^{h-1} \widehat{\mathcal{G}}_\ell^{(k)}(x) \overset{(29)}{=} \mathcal{G}^{(k)} \cup \big\{\{x\} \cup e \colon e \in \mathcal{G}^{(k-1)} \text{ and } x \boxplus e \notin \Gamma_k\big\}. \tag{32}$$

Since $S \cup \{x\}$ is a B_h-set, then $\|R_{\widehat{\mathcal{G}}^{(k)},\widehat{\mathcal{G}}^{(\ell)}}\| \leqslant 1$ for every $k, \ell \in [h-1]$ satisfying $k + \ell \leqslant h$ and therefore

$$Q_{\widehat{\mathcal{G}}^{(k)},\widehat{\mathcal{G}}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \leqslant (2hn+1) \cdot \exp(\lambda \cdot \xi_{k+\ell}) \overset{(15)}{\leqslant} (2hn+1) \cdot e \leqslant (2hn+1) \cdot \exp(\mathsf{H}_{|S|+1}),$$

as we have assumed that $\lambda \leqslant 1$. It follows from (30), and the fact that $\widehat{\mathcal{G}}^{(k)} \subset \widehat{\mathcal{G}}_\ell^{(k)}(x)$, that for every $k, \ell \in \{2, \dots, h-1\}$,

$$Q_{\widehat{\mathcal{G}}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \leqslant Q_{\widehat{\mathcal{G}}_\ell^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \leqslant Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \cdot \left(1 + \frac{1}{4(|S|+1)}\right). \tag{33}$$

Since $Q_{\mathcal{G},\mathcal{H}}(\cdot) = Q_{\mathcal{H},\mathcal{G}}(\cdot)$, the same bound above applies to $Q_{\mathcal{G}^{(k)},\widehat{\mathcal{G}}^{(\ell)}}(\lambda \cdot \xi_{k+\ell})$. Consequently, Lemma 18 implies that for all $k, \ell \in \{2, \dots, h-1\}$,

$$\begin{aligned} Q_{\widehat{\mathcal{G}}^{(k)},\widehat{\mathcal{G}}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) &\leqslant Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \cdot \left(1 + \frac{1}{|S|+1}\right) \\ &\quad \text{—by condition } (b) \text{ of Def. 16—} \\ &\leqslant (2hn+1) \cdot \exp(\mathsf{H}_{|S|}) \cdot \left(1 + \frac{1}{|S|+1}\right) \\ &\leqslant (2hn+1) \cdot \exp(\mathsf{H}_{|S|+1}). \end{aligned} \tag{34}$$

In other words, the hypergraphs $\widehat{\mathcal{G}}^{(1)}, \dots, \widehat{\mathcal{G}}^{(h-1)}$ satisfy condition (b) of Definition 16 with S replaced by $S \cup \{x\}$. Since $x \in \widetilde{S}_{\lambda,\alpha}$, the set $S \cup \{x\}$ does not satisfy property $\mathcal{P}_h(\lambda, \alpha)$ and hence condition (a) of Definition 16 has to be violated, that is, there must be some $k \in [h-1]$ for which $|\widehat{\mathcal{G}}^{(k)}| < (1 - 2^k\alpha)\binom{|S|+1}{k}$. Together with the fact that $|\mathcal{G}^{(k)}| \geqslant (1 - 2^k\alpha)\binom{|S|}{k}$, we have

$$|\widehat{\mathcal{G}}^{(k)}| - |\mathcal{G}^{(k)}| < (1 - 2^k\alpha)\binom{|S|+1}{k} - (1 - 2^k\alpha)\binom{|S|}{k} = (1 - 2^k\alpha)\binom{|S|}{k-1}. \tag{35}$$

This is clearly not true if $k = 1$, as $|\widehat{\mathcal{G}}^{(1)}| = |\mathcal{G}^{(1)}| + 1$, hence let us consider $k \in \{2, \dots, h-1\}$. By the definition of $\widehat{\mathcal{G}}^{(k)}$ in (32),

$$|\widehat{\mathcal{G}}^{(k)}| - |\mathcal{G}^{(k)}| = |\mathcal{G}^{(k-1)}| - |\{e \in \mathcal{G}^{(k-1)} \colon x \boxplus e \in \Gamma_k\}| = |\mathcal{G}^{(k-1)}| - \big|\big(x \boxplus \mathcal{G}^{(k-1)}\big) \cap \Gamma_k\big|,$$

where in the last equality we used the fact that S is a B_{k-1}-set and therefore no two distinct $e, e' \in \mathcal{G}^{(k-1)}$ may satisfy $x \boxplus e = x \boxplus e'$. Since $\mathcal{G}^{(k-1)}$ satisfies condition (a) of Definition 16, we

have

$$|\widehat{\mathcal{G}}^{(k)}| - |\mathcal{G}^{(k)}| \geqslant (1 - 2^{k-1}\alpha)\binom{|S|}{k-1} - \big|\big(x \boxplus \mathcal{G}^{(k-1)}\big) \cap \Gamma_k\big|. \tag{36}$$

Combining (35) and (36) yields

$$\left|x \boxplus \binom{S}{k-1} \cap \Gamma_k\right| \geqslant \big|\big(x \boxplus \mathcal{G}^{(k-1)}\big) \cap \Gamma_k\big| \geqslant 2^{k-1}\alpha\binom{|S|}{k-1},$$

which yields condition (ii) of this lemma, as x was arbitrary. □

5. Proof of the main result

In this section we derive our main result, Theorem 4, from the main result of the previous two sections, Lemma 20, and the following technical statement, Theorem 21 below, which provides lower bounds on $e_{\widetilde{\mathrm{CG}}_{\mathcal{H}}}(A)$ for various sets A and $(h-1)$-graphs $\mathcal{H}$. As an attentive reader will surely notice, the assumptions of Theorem 21 are suited for invoking the theorem with $A \subset \widetilde{S}_{\lambda,\alpha,\ell}$ and $\Gamma = \Gamma_\ell$ from Lemma 20. We postpone the proof of Theorem 21 to Section 6.

Theorem 21. *Let $h \geqslant 2$, $\ell \in [h-1]$, $\beta \in (0,1]$, n be a sufficiently large integer, and $d \geqslant (128h\log_2 n)^{\ell+2}$. Let $S \in \mathcal{Z}_n^h$, with $\beta\,|S| > n^{1/(100h^2)}$, and $\mathcal{H} \subset \binom{S}{h-1}$, with*

$$|\mathcal{H}| \geqslant \left(1 - \frac{\beta^h}{(\log_2 n)^{7h^2}}\right)\binom{|S|}{h-1}. \tag{37}$$

Suppose that the sets $A \subset [n]$ and $\Gamma \subset [hn]$ satisfy

$$\sum_{a\in A}\left|a \boxplus \binom{S}{\ell} \cap \Gamma\right| \geqslant \max\left\{\beta\binom{|S|}{\ell}\cdot|A|,\quad d\cdot|\Gamma|\right\}. \tag{38}$$

Then

$$e_{\widetilde{\mathrm{CG}}_{\mathcal{H}}}(A) \geqslant \frac{\beta^h\cdot d}{(\log_2 n)^{7h^2}}|A|\binom{|S|}{h-1}. \tag{39}$$

We are now ready to prove Theorem 9, which implies an upper bound on $|\mathcal{Z}_n^h(t)|$ for t in a narrow range around $n^{1/(2h-1)+o(1)}$. It is then easy to show that this bound extends to all t satisfying $n^{1/(2h-1)+o(1)} \leqslant t \leqslant F_h(n)$; see §5.2.

Before we embark on the proof of Theorem 9, let us formally define the notion of bounded B_h-sets.

Definition 22 (Bounded B_h-sets)**.** Let $h \geqslant 2$ and $\rho > 0$ be given. A set $S \subset [n]$ satisfies property $\mathcal{P}_h(\rho)$ if it satisfies, for all $i \in \{0,1,\ldots,\lceil 1/\rho\rceil\}$, property $\mathcal{P}_h(\lambda_i,\alpha_i)$, where

$$\lambda_i = n^{-i\rho},\quad \alpha_0 = \frac{1}{2^h(\log_2 n)^{7h^2}},\quad \text{and}\quad \alpha_{j+1} = \frac{(\alpha_j/2)^h}{(\log_2 n)^{7h^2}}\quad \text{for } j = 0,1,\ldots,\lceil 1/\rho\rceil - 1. \tag{40}$$

The reason for having such a range of parameters (λ_i,α_i) in the definition of boundedness will become aparent in §5.1.2.

5.1. **Proof of Theorem 9.** Let h, δ, n, and t be given as in the statement of Theorem 9. We shall construct a family $\mathcal{F} = \mathcal{F}(t) = \mathcal{F}_{\texttt{small}}(t) \cup \mathcal{F}_{\texttt{large}}(t)$ of pairs of sets $(S,\widetilde{S})$ with the property

that every $T \in \mathcal{Z}_n^h(t)$ satisfies $S \subset T \subset S \cup \widetilde{S}$ for some $(S, \widetilde{S}) \in \mathcal{F}$ and, more importantly, such that every pair $(S, \widetilde{S})$ satisfies (9). To this end, let

$$\rho = \frac{\delta}{4}\left(\frac{1}{2h-1} + \delta\right). \tag{41}$$

Note that $\alpha_i = (\log n)^{-\Theta(1)}$ for every $i \in \{0, 1, \dots, \lceil 1/\rho \rceil\}$ since δ, h, and ρ are absolute constants.

Define $\mathcal{F}_{\texttt{large}}(t)$ to be the set of all pairs $(S, \widetilde{S})$ such that the following hold:

- (I) $S \in \mathcal{Z}_n^h$.
- (II) $|S| = t^{1-\delta}$.
- (III) There exists $\mathcal{G} \subset \binom{S}{h-1}$ satisfying (cf. (a) and (b) of Definition 16):
  - $|\mathcal{G}| \geqslant (1 - 2^{h-1}\alpha_0)\binom{|S|}{h-1}$.
  - $Q_{\mathcal{G},\mathcal{G}}(\lambda_0 \xi_{2h-2}/2) \leqslant (2hn+1)\exp\big(\mathsf{H}_{|S|}\big)$.
- (IV) The set $\widetilde{S}$ is defined as

$$\widetilde{S} = \big\{x \in [n] \setminus S \colon S \cup \{x\} \text{ is a } B_h\text{-set}\big\}. \tag{42}$$

Define $\mathcal{F}_{\texttt{small}}(t)$ to be the set of all pairs $(S, \widetilde{S})$ such that the following hold:

- (i) S satisfies $\mathcal{P}_h(\rho)$.
- (ii) $n^{1/(8h^2)} \leqslant |S| < t^{1-\delta}$.
- (iii) The set $\widetilde{S}$ is defined as

$$\widetilde{S} = \big\{x \in [n] \setminus S \colon S \cup \{x\} \text{ is a } B_h\text{-set which does } \textit{not} \text{ satisfy } \mathcal{P}_h(\rho)\big\}. \tag{43}$$

Claim 2. *For every $T \in \mathcal{Z}_n^h(t)$, there exists $(S, \widetilde{S}) \in \mathcal{F}$ such that $S \subset T \subset S \cup \widetilde{S}$.*

Proof. Given a $T \in \mathcal{Z}_n^h(t)$, let $\mathcal{S}$ be the family of all subsets of T that satisfy $\mathcal{P}_h(\rho)$ and have at least $n^{1/(8h^2)}$ elements. We first show that $\mathcal{S} \neq \emptyset$. For that, observe that one can form a B_{2h-2}-set $X \subset T$ by greedily picking elements from T one-by-one until no more elements can be selected. The elements that cannot be added to X are of the form

$$x_1 + \cdots + x_{2h-2} - (y_1 + \cdots + y_{2h-3})$$

with $x_i \in X$ for all $i \in [2h-2]$ and $y_j \in X$ for all $j \in [2h-3]$. Hence, if X was obtained by the greedy procedure, we must have $|X|^{4h-5} \geqslant |T|$. In particular, $|X| \geqslant t^{1/(4h-5)} \geqslant n^{1/(8h^2)}$. For every $k = 1, \dots, h-1$, let $\mathcal{G}^{(k)} = \binom{X}{k}$. Since X is a B_{2h-2}-set, it follows by Remark 11 that $\|R_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}\| = 1$ when $k + \ell \leqslant 2h-2$. Therefore, for each $\lambda \in (0,1]$, and $k, \ell \in [h-1]$,

$$Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda) \leqslant 2hne^{\lambda} < (2hn+1) \cdot \exp(\mathsf{H}_{|X|}).$$

It follows that X satisfies $\mathcal{P}_h(\lambda, \alpha)$ for any $\lambda \leqslant 1$ and any $\alpha \geqslant 0$. In particular, it satisfies $\mathcal{P}_h(\rho)$, which shows that $X \in \mathcal{S}$.

Pick some largest $S \in \mathcal{S}$. If $|S| < t^{1-\delta}$, then let $\widetilde{S}$ be the set defined in (43) so that $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{small}}(t)$. Since S is the largest subset of T which satisfies $\mathcal{P}_h(\rho)$ we clearly have $T \setminus S \subset \widetilde{S}$, which is the conclusion of the claim. Hence we may assume, for the remainder of the proof of the claim, that $|S| \geqslant t^{1-\delta}$.

We shall now construct a subset $S' \subset S$ and $\mathcal{G} \subset \binom{S'}{h-1}$ that satisfy (I)–(III) with S replaced by S'. By assumption, S satisfies $\mathcal{P}_h(\rho)$ and hence, in particular, $S \in \mathcal{P}_h(\lambda_0, \alpha_0)$. Therefore, there exists a $\mathcal{G}_0^{(h-1)} \subset \binom{S}{h-1}$ satisfying the conditions of Definition 16 with $\lambda = \lambda_0$ and $\alpha = \alpha_0$. Consider an arbitrary subset $S' \in \binom{S}{t^{1-\delta}}$ and let $\mathcal{G} = \mathcal{G}_0^{(h-1)}[S']$. Since $|S'| \geqslant |S|^{1-\delta}$, then $\mathsf{H}_{|S|} \leqslant 2\mathsf{H}_{|S'|}$ and consequently,

$$Q_{\mathcal{G},\mathcal{G}}(\lambda_0\xi_{2h-2}) \leqslant Q_{\mathcal{G}_0^{(h-1)},\mathcal{G}_0^{(h-1)}}(\lambda_0\xi_{2h-2}) \leqslant (2hn+1)\exp(2\mathsf{H}_{|S'|}).$$

Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
Q_{\mathcal{G},\mathcal{G}}(\lambda_0\xi_{2h-2}/2) &= \sum_{z=-(h-1)n}^{(h-1)n} \exp\bigl(\lambda_0\xi_{2h-2}\cdot R_{\mathcal{G}}(z)\bigr)^{1/2} \\
&\leqslant \Biggl(\bigl((2h-2)n+1\bigr)\sum_{z=-(h-1)n}^{(h-1)n} \exp\bigl(\lambda_0\xi_{2h-2}\cdot R_{\mathcal{G}}(z)\bigr)\Biggr)^{1/2} \\
&= \Bigl(\bigl((2h-2)n+1\bigr)Q_{\mathcal{G},\mathcal{G}}(\lambda_0\xi_{2h-2})\Bigr)^{1/2} \\
&\leqslant (2hn+1)\exp\bigl(\mathsf{H}_{|S'|}\bigr).
\end{aligned}$$

As the choice of S' above was arbitrary, we may select S' which maximizes $|\mathcal{G}|$. By averaging over all sets of cardinality $|S'| = t^{1-\delta}$, we have

$$|\mathcal{G}| \geqslant \bigl|\mathcal{G}_0^{(h-1)}\bigr|\binom{t-(h-1)}{|S'|-(h-1)}\binom{t}{|S'|}^{-1} \geqslant (1-2^{h-1}\alpha_0)\binom{|S'|}{h-1}.$$

Consequently, S' and its corresponding $\widetilde{S}'$, defined as in (42), form a pair $(S', \widetilde{S}') \in \mathcal{F}_{\texttt{large}}(t)$. Since $T \supset S$ is a B_h-set, it follows that $T \setminus S' \subset \widetilde{S}'$. This completes the proof of the claim. □

So far we have constructed a family $\mathcal{F}$ that satisfies the first assertion of the theorem. It remains to show that the second assertion also holds, that is, that for all $(S, \widetilde{S}) \in \mathcal{F}$, the graph CG_S satisfies (9). In order to prove it, we shall consider two cases, depending on whether $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{large}}(t)$ or $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{small}}(t)$.

5.1.1. *Case when* $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{large}}(t)$. Our definition of $\mathcal{F}_{\texttt{large}}(t)$ (see (I)–(IV)) guarantees the existence of an $(h-1)$-graph $\mathcal{G} \subset \binom{S}{h-1}$ that satisfies (III). Let $A \subset \widetilde{S}$ be an arbitrary set with $|A| \geqslant \frac{n}{t^{h-1-8h^2\delta}}$. We shall apply[2] Theorem 21 with

$$\ell = h-1, \quad \beta = 1, \quad \mathcal{H} = \mathcal{G}, \quad \Gamma = [hn], \quad d = |A|\binom{|S|}{h-1}/(hn). \tag{44}$$

Indeed, the conditions of the theorem are satisfied because of the following:

- For every $a \in A$, we have $a \boxplus \binom{S}{h-1} \subset [hn] = \Gamma$.
- $|A|\binom{|S|}{h-1} \geqslant \frac{n}{t^{h-1-8h^2\delta}}\left(\frac{t^{1-\delta}}{h-1}\right)^{h-1} \geqslant nt^{\delta}$ and thus $d \geqslant t^{\delta} \gg (128h\log_2 n)^{h+1}$.
- We have $|S| = t^{1-\delta} > n^{1/(4h)}$ and thus $\beta|S| > n^{1/(100h^2)}$.

[2]It is possible to use a direct argument for this case, as described in the outline of Section 2, however, we instead use Theorem 21 to reduce the length of the proof.

- Since $|\mathcal{G}| \geqslant (1-2^{h-1}\alpha_0)\binom{|S|}{h-1}$, it follows from (40) that $\mathcal{H}=\mathcal{G}$ satisfies (37), as $2^{h-1}\alpha_0 = \frac{1}{2(\log_2 n)^{7h^2}} < \frac{\beta^h}{(\log_2 n)^{7h^2}}$.

Hence,

$$e_{\widetilde{\mathrm{CG}}_{\mathcal{G}}}(A) \geqslant (\log n)^{-O(1)} \cdot d|A|\binom{|S|}{h-1} \overset{(44)}{\geqslant} \frac{1}{hn\cdot(\log n)^{O(1)}}|A|^2\binom{|S|}{h-1}^2.$$

On the other hand, from (III) and Remark 14 we conclude that

$$\|R_{\mathcal{G}}\| = \frac{2}{\lambda_0\xi_{2h-2}}\log Q_{\mathcal{G},\mathcal{G}}(\lambda_0\xi_{2h-2}/2) \leqslant \frac{2\log\Big((2hn+1)\exp\big(\mathsf{H}_{|S|}\big)\Big)}{\lambda_0\xi_{2h-2}} \overset{(15)}{=} (\log n)^{O(1)}.$$

Therefore,

$$e_{\mathrm{CG}_S}(A) \geqslant \frac{e_{\widetilde{\mathrm{CG}}_{\mathcal{G}}}(A)}{\|R_{\mathcal{G}}\|} \geqslant \frac{|S|^{2h-2}}{n\cdot(\log n)^{O(1)}}|A|^2. \tag{45}$$

We claim that (45) implies

$$e_{\mathrm{CG}_S}(A) \geqslant \frac{|A|^2}{|S|} \geqslant t^{\delta-1}\binom{|A|}{2},$$

which gives the conclusion of the theorem. For this it is enough to show that for all sufficiently large n,

$$|S|^{2h-1} > n^{1+\delta}.$$

As we have $|S| = t^{1-\delta}$, $\delta < 1/2$, and $t \geqslant n^{1/(2h-1)+\delta}$, the above inequality follows by taking the logarithm of both sides and observing that

$$(1-\delta)(2h-1)\left(\frac{1}{2h-1}+\delta\right)\log n > (1-\delta)(1+3\delta)\log n > (1+\delta)\log n.$$

This completes the proof of Theorem 9 in the case $(S,\widetilde{S}) \in \mathcal{F}_{\texttt{large}}(t)$.

5.1.2. *Case when* $(S,\widetilde{S}) \in \mathcal{F}_{\texttt{small}}(t)$. Recalling the definition of $\mathcal{F}_{\texttt{small}}(t)$, we can naturally partition $\widetilde{S}$ as

$$\widetilde{S} = \bigcup_{i=1}^{\lceil 1/\rho\rceil} \widetilde{S}_i,$$

where $\widetilde{S}_i$ is the set of all $x \in \widetilde{S}$ such that i is the smallest index for which $S\cup\{x\}$ does not satisfy $\mathcal{P}_h(\lambda_i,\alpha_i)$. Note that $\widetilde{S}_i \subset \widetilde{S}_{\lambda_i,\alpha_i}$, where $\widetilde{S}_{\lambda_i,\alpha_i}$ is the set introduced by Definition 16.

Claim 3. $\widetilde{S}_{\lceil 1/\rho\rceil} = \emptyset$.

Proof. This is true because $\lambda_{\lceil 1/\rho\rceil}$ is so small that the bound on Q is trivially true even for complete hypergaphs. More formally, assume for the sake of a contradiction that $x \in \widetilde{S}_{\lceil 1/\rho\rceil}$. For any $k \in [h-1]$, let $\mathcal{K}^{(k)} = \binom{S\cup\{x\}}{k}$ and observe that since $\lambda_{\lceil 1/\rho\rceil} \leqslant n^{-1}$, then for all $k,\ell \in [h-1]$,

$$\lambda_{\lceil 1/\rho\rceil}\cdot\|R_{\mathcal{K}^{(k)},\mathcal{K}^{(\ell)}}\| \leqslant n^{-1}\cdot\big(|S|+1\big)^{k+\ell-1} \leqslant n^{-1}\cdot t^{(1-\delta)(2h-2)} < 1.$$

Consequently,

$$Q_{\mathcal{K}^{(k)},\mathcal{K}^{(\ell)}}(\lambda_{\lceil 1/\rho\rceil}\xi_{k+\ell}) = \sum_{w=-kn}^{\ell n}\exp\big(\lambda_{\lceil 1/\rho\rceil}\xi_{k+\ell}R_{\mathcal{K}^{(k)},\mathcal{K}^{(\ell)}}(w)\big) < \sum_{w=-kn}^{\ell n} e \leqslant (2hn+1)e.$$

It follows that the family of hypergraphs $\mathcal{K}^{(k)}$, $k \in [h-1]$, satisfies the conditions of Definition 16 with $\lambda = \lambda_{\lceil 1/\rho \rceil}$ and $\alpha = \alpha_{\lceil 1/\rho \rceil}$. Therefore $S \cup \{x\} \in \mathcal{P}_h(\lambda_{\lceil 1/\rho \rceil}, \alpha_{\lceil 1/\rho \rceil})$ and thus $x \notin \widetilde{S}_{\lceil 1/\rho \rceil}$, which is a contradiction. □

Now for each $i \in \{0, 1, \dots, \lceil 1/\rho \rceil - 1\}$ we apply Lemma 20 with $\lambda = \lambda_i$ and $\alpha = \alpha_i$ to obtain sets $\Gamma_{i,2}, \dots, \Gamma_{i,h-1} \subset [hn]$ satisfying the following:

- $|\Gamma_{i,k}| \leqslant (|S|+1)^{h+k-1} \cdot \lambda_i$ for every $k \in \{2, \dots, h-1\}$.
- For every $x \in \widetilde{S}_i$ there is a $k \in \{2, \dots, h-1\}$ such that

$$\left| x \boxplus \binom{S}{k-1} \cap \Gamma_{i,k} \right| \geqslant 2^{k-1} \alpha_i \binom{|S|}{k-1}.$$

We then further partition

$$\widetilde{S}_i = \bigcup_{k=2}^{h-1} \widetilde{S}_{i,k},$$

where $x \in \widetilde{S}_{i,k}$ if k is the smallest index for which the second condition above holds.

Choose an arbitrary set $A \subset \widetilde{S}$ with $|A| \geqslant \frac{n}{t^{h-1-8h^2\delta}}$. Let $i \in \{0, 1, \dots, \lceil 1/\rho \rceil - 1\}$ and $k \in \{2, \dots, h-1\}$ be such that $A_{i,k} = A \cap \widetilde{S}_{i,k}$ satisfies

$$|A_{i,k}| \geqslant \frac{|A|}{(h-2)\lceil 1/\rho \rceil} = \Omega\left(\frac{n}{t^{h-1-8h^2\delta}} \right). \tag{46}$$

Finally, let $\mathcal{H}$ denote the $(h-1)$-graph $\mathcal{G}^{(h-1)}$ whose existence is guaranteed by the fact that S satisfies $\mathcal{P}_h(\lambda_{i+1}, \alpha_{i+1})$. (Note that in view of Claim 3, we must have $i \leqslant \lceil 1/\rho \rceil - 1$, so this is indeed well-defined.) Here is the point in the proof where we actually need the variety of (λ, α) pairs for which the set S must satisfy $\mathcal{P}_h(\lambda, \alpha)$. Indeed, we use $\mathcal{P}_h(\lambda_i, \alpha_i)$ to obtain the sets $\Gamma_{i,k}$ $(k = 2, \dots, h-1)$, and $\mathcal{P}_h(\lambda_{i+1}, \alpha_{i+1})$ to obtain $\mathcal{H}$. This is simply due to the fact that the density of $\mathcal{H}$ must be so large, that the $(h-1)$-graph one can obtain through $\mathcal{P}_h(\lambda_i, \alpha_i)$ would not be sufficient to satisfy the conditions of Theorem 21. In effect, we are allowing a slightly worse upper bound on $\|R_{\mathcal{H}}\|$ for the benefit of a denser $\mathcal{H}$.

Recall from Definition 16 that

- $|\mathcal{H}| \geqslant (1 - 2^{h-1}\alpha_{i+1}) \binom{|S|}{h-1}$ and
- $Q_{\mathcal{H},\mathcal{H}}\big(\lambda_{i+1}\xi_{2h-2}\big) \leqslant (2hn+1)\exp(\mathsf{H}_{|S|})$, which by Remark 14 means that

$$\|R_{\mathcal{H}}\| \leqslant \frac{\log\Big((2hn+1)\exp\big(\mathsf{H}_{|S|}\big)\Big)}{\lambda_{i+1}\xi_{2h-2}} \overset{(15)}{=} \frac{(\log n)^{O(1)}}{\lambda_{i+1}}. \tag{47}$$

We shall now apply Theorem 21 with

$$\ell = k-1, \quad A = A_{i,k}, \quad \beta = \alpha_i, \quad \mathcal{H} \text{ as above,}$$
$$\Gamma = \Gamma_{i,k}, \quad d = 2^{k-1}\alpha_i |A_{i,k}| \binom{|S|}{k-1} / |\Gamma_{i,k}|.$$

First, let us verify that the conditions of the theorem are satisfied for our choice of parameters:

- For every $a \in A_{i,k} \subset \widetilde{S}_{i,k}$, we have $\big|a \boxplus \binom{S}{k-1} \cap \Gamma_{i,k}\big| \geqslant 2^{k-1}\alpha_i \binom{|S|}{k-1} \geqslant \beta \binom{|S|}{k-1}$.

- We also have
$$\sum_{a\in A_{i,k}}\left|a\boxplus\binom{S}{k-1}\cap\Gamma_{i,k}\right|\geqslant 2^{k-1}\alpha_i\,|A_{i,k}|\binom{|S|}{k-1}=d\,|\Gamma_{i,k}|.$$
- Since $|\Gamma_{i,k}|\leqslant(|S|+1)^{h+k-1}\cdot\lambda_i$, we see that d satisfies
$$d=2^{k-1}\alpha_i\,|A_{i,k}|\binom{|S|}{k-1}/|\Gamma_{i,k}|\overset{(46)}{\geqslant}(\log n)^{-O(1)}\cdot\frac{n|S|^{k-1}}{t^{h-1-8h^2\delta}(|S|+1)^{h+k-1}\lambda_i}$$
Since $|S|<t\leqslant 2hn^{1/(2h-1)+\delta}$ and $\lambda_i\leqslant 1$, we have
$$d\geqslant(\log n)^{-O(1)}\cdot\frac{n}{t^{2h-1-8h^2\delta}}\geqslant(\log n)^{-O(1)}\cdot n^{4h^2\delta/(2h-1)}\gg(128h\log_2 n)^{k+1}.$$
- The set S, by the definition of $\mathcal{F}_{\texttt{small}}(t)$, has cardinality at least $n^{1/(8h^2)}$ and therefore $\beta|S|\gg n^{1/(100h^2)}$.
- By our choice of β and the fact that $|\mathcal{H}|\geqslant(1-2^{h-1}\alpha_{i+1})\binom{|S|}{h-1}$, it follows from (40) that $\mathcal{H}$ satisfies (37). Indeed,
$$2^{h-1}\alpha_{i+1}=\frac{2^h(\alpha_i/2)^h}{2(\log_2 n)^{7h^2}}=\frac{\beta^h}{2(\log_2 n)^{7h^2}}.\tag{48}$$

Hence, by Theorem 21,
$$e_{\widetilde{\mathrm{CG}}_{\mathcal{H}}}(A_{i,k})\geqslant(\log n)^{-O(1)}\cdot d\,|A_{i,k}|\binom{|S|}{h-1}.$$
Recalling (47), we conclude that
$$\begin{aligned}e_{\mathrm{CG}_S}(A)\geqslant e_{\mathrm{CG}_S}(A_{i,k})&\geqslant\frac{e_{\widetilde{\mathrm{CG}}_{\mathcal{H}}}(A_{i,k})}{\|R_{\mathcal{H}}\|}\geqslant(\log n)^{-O(1)}\cdot\lambda_{i+1}\cdot d\binom{|S|}{h-1}|A_{i,k}|\\&=(\log n)^{-O(1)}\cdot\frac{\lambda_{i+1}|S|^{h+k-2}}{|\Gamma_{i,k}|}|A_{i,k}|^2\\&\geqslant(\log n)^{-O(1)}\cdot\frac{\lambda_{i+1}}{\lambda_i\,|S|}|A_{i,k}|^2\\&=(\log n)^{-O(1)}\cdot\frac{n^{-\rho}}{|S|}\binom{|A|}{2},\end{aligned}\tag{49}$$
where we used the definition of λ_i,λ_{i+1} in (40) and the fact that $|A_{i,k}|=\Omega\big(|A|\big)$. From the fact that $|S|<t^{1-\delta}$, and $n^{-\rho}\geqslant t^{-\delta/4}$ (see (41) and recalling that $t\geqslant n^{1/(2h-1)+\delta}$), we obtain
$$e_{\mathrm{CG}_S}(A)\geqslant\frac{1}{t^{1-3\delta/4}(\log n)^{O(1)}}\binom{|A|}{2}\gg\frac{1}{t^{1-\delta/2}}\binom{|A|}{2}.$$
This concludes the proof of Theorem 9.

5.2. **Deriving Theorem 4 from Theorem 9.** Our proof of Theorem 4 has two independent parts. First, we derive the claimed bound on $|\mathcal{Z}_n^h(t)|$ only for t in a narrow interval around $n^{1/(2h-1)+\varepsilon}$. Second, we extend this bound to all larger t using the following statement, Lemma 23 below, which was already implicitly proved in [10, Section 5]. For completeness, we include the proof of Lemma 23 in Appendix A.

Lemma 23. *Let $h \geqslant 2$ and suppose that $N \geqslant 2hn$. Then, for every t,*

$$|\mathcal{Z}_N^h(t)| \geqslant |\mathcal{Z}_n^h(t)| \cdot \left(\frac{N}{2hn}\right)^t.$$

Proof of Theorem 4. Suppose that $h \geqslant 2$, fix some $\varepsilon > 0$ and let δ be a sufficiently small positive constant. For any sufficiently large n, define

$$\ell_h(n) = n^{1/(2h-1)+\delta}.$$

We will first show that

$$|\mathcal{Z}_n^h(t)| \leqslant \left(\frac{n}{t^{h-\varepsilon/2}}\right)^t \quad \text{for all } \ell_h(n) \leqslant t \leqslant 2h\ell_h(n). \tag{50}$$

To this end, invoke Theorem 9 to obtain a family $\mathcal{F}$ with the property that every $T \in \mathcal{Z}_n^h(t)$ satisfies $S \subset T \subset S \cup \widetilde{S}$ for some $(S, \widetilde{S}) \in \mathcal{F}$ and such that every $(S, \widetilde{S}) \in \mathcal{F}$ has $|S| \leqslant t^{1-\delta}$ and satisfies (9). We may bound $|\mathcal{Z}_n^h(t)|$ from above by the sum, over all $(S, \widetilde{S}) \in \mathcal{F}$, of the number $F_h(S, \widetilde{S}, t)$ of B_h-sets of cardinality t that contain S and are contained in $S \cup \widetilde{S}$.

Fix an arbitrary $(S, \widetilde{S}) \in \mathcal{F}$. If $|\widetilde{S}| < \frac{n}{t^{h-1-8h^2\delta}}$, then condition (9) is vacuous, but on the other hand,

$$F_h(S, \widetilde{S}, t) \leqslant \binom{\frac{n}{t^{h-1-8h^2\delta}}}{t-|S|} \quad \text{when } |\widetilde{S}| < \frac{n}{t^{h-1-8h^2\delta}}. \tag{51}$$

Otherwise, when $|\widetilde{S}| \geqslant \frac{n}{t^{h-1-8h^2\delta}}$, as $F_h(S, \widetilde{S}, t)$ is at most the number of $(t-|S|)$-element independent sets in $\mathrm{CG}_S[\widetilde{S}]$, we invoke Lemma 8 with

$$\begin{gathered} G = \mathrm{CG}_S[\widetilde{S}], \quad N = |\widetilde{S}|, \quad R = \frac{n}{t^{h-1-8h^2\delta}}, \\ \beta = \frac{1}{t^{1-\delta/2}}, \quad q = \lceil \beta^{-1} \log n \rceil, \quad \text{and} \quad m = t - q - |S|. \end{gathered} \tag{52}$$

Note that the conditions of the lemma are satisfied by our choice of parameters as

$$R = \frac{n}{t^{h-1-8h^2\delta}} \gg 1 \geqslant e^{-\beta q} n \geqslant e^{-\beta q} N$$

and condition (9) implies that G satisfies (6). It follows from Lemma 8 that

$$F_h(S, \widetilde{S}, t) \leqslant \binom{|\widetilde{S}|}{q} \binom{R}{t-q-|S|} \quad \text{when } |\widetilde{S}| \geqslant \frac{n}{t^{h-1-8h^2\delta}}. \tag{53}$$

As $q \ll t/(\log n)$ and $|\widetilde{S}| \leqslant n$, in view of both (51) and (53), we have

$$F_h(S, \widetilde{S}, t) \leqslant e^{o(t)} \binom{\frac{n}{t^{h-1-8h^2\delta}}}{t} \leqslant \left(\frac{n}{t^{h-8h^2\delta}}\right)^{t+o(t)}.$$

Since $|S| \leqslant t^{1-\delta}$ for each $(S, \widetilde{S}) \in \mathcal{F}$,

$$|\mathcal{Z}_n^h(t)| = \sum_{(S,\widetilde{S}) \in \mathcal{F}} F_h(S, \widetilde{S}, t) \leqslant n^{t^{1-\delta}} \left(\frac{n}{t^{h-8h^2\delta}}\right)^{t+o(t)} = \left(\frac{n}{t^{h-8h^2\delta}}\right)^{t+o(t)}.$$

Consequently, (50) holds provided that $\delta = \delta(\varepsilon)$ is sufficiently small.

We now extend the upper bound given in (50) to all t up to $F_h(n)$. Suppose that $2h\ell_h(n) < t \leqslant F_h(n)$ and let N be the largest integer such that $t \geqslant \ell_h(N)$. Note that $t < \ell_h(N+1) < \ell_h(N)+1 < 2h\ell_h(N)$ and $\ell_h(2hn) < 2h\ell_h(n) < t$, thus $N > 2hn$. From (50), we conclude that

$$|\mathcal{Z}_N^h(t)| \leqslant \left(\frac{N}{t^{h-\varepsilon/2}}\right)^t.$$

Lemma 23 then implies that

$$\left(\frac{N}{t^{h-\varepsilon/2}}\right)^t \geqslant |\mathcal{Z}_N^h(t)| \geqslant |\mathcal{Z}_n^h(t)| \cdot \left(\frac{N}{2hn}\right)^t.$$

Consequently,

$$|\mathcal{Z}_n^h(t)| \leqslant \left(\frac{n}{t^{h-\varepsilon}}\right)^t$$

for all t with $\ell_h(n) \leqslant t \leqslant F_h(n)$, provided that n is sufficiently large. □

6. Proof of Theorem 21

Let S, A, Γ, and $\mathcal{H}$ be as in the statement of Theorem 21. Recall that our goal is to construct many quadruples $(a_1, a_2, e_1, e_2) \in A^2 \times \mathcal{H}^2$ with

$$a_1 \boxplus e_1 = a_2 \boxplus e_2 \qquad \text{and} \qquad e_1 \cap e_2 = \emptyset. \tag{54}$$

We shall reduce this task to the task of counting certain paths in a pair of bipartite graphs sharing one color class. Our argument will have two parts. In the first part, termed the *pre-processing stage*, we construct the aforementioned pair of bipartite graphs from the sets S, A, and Γ. Significant effort is put into making these two graphs highly degree-regular. In the second part, we count certain paths in these graphs, which we term *special* and *semi-special*, that correspond to quadruples (a_1, a_2, e_1, e_2) that satisfy (54). Our counting arguments rely heavily on the degree-regularity inherited from the pre-processing stage.

6.1. **A warm-up.**

This section is fairly technical and thus it might be helpful to start with a simpler setup. Unfortunately, this simplified setup is unrealistic for our application. However, it will allow us to demonstrate some subtleties of the proof. Suppose in this simplification that $\mathcal{H}$ is the complete $(h-1)$-graph on S and that, moreover, we allow $e_1 \cap e_2 \neq \emptyset$ in the quadruples (a_1, a_2, e_1, e_2) that we are counting.

Consider a bipartite graph J with color classes A and Γ, where $(a, w) \in A \times \Gamma$ is an edge of J whenever $w = a + s_1 + \cdots + s_\ell$ with all $s_i \in S$. Observe that if (a_1, w) and (a_2, w) both belong to J, then we have a quadruple $(a_1, a_2, e_1^{(\ell)}, e_2^{(\ell)}) \in A^2 \times \binom{S}{\ell}^2$ with $a_1 \boxplus e_1^{(\ell)} = w = a_2 \boxplus e_2^{(\ell)}$.

In particular, when $\ell = h-1$, we may simply count paths of length two in J which start and end in A in order to estimate the number of quadruples of the desired type. Establishing a lower bound is fairly easy since we know, from the theorem's hypotheses, that the average degree in Γ is somewhat large, and therefore we can apply the Cauchy–Schwarz inequality to estimate the number of 2-paths in J with central vertex in Γ. Indeed, the left-hand side of (38) equals $e(J)$ and

therefore, the number of 2-paths is at least

$$\sum_{w\in\Gamma}\binom{\deg_J(w)}{2} = \Theta(1)\sum_{w\in\Gamma}\deg_J(w)^2 \geqslant \Theta(1)\frac{e(J)^2}{|\Gamma|} \overset{(38)}{\geqslant} \Theta(1)\beta d\,|A|\binom{|S|}{h-1}.$$

Note that the right-hand side of the above inequality is somewhat larger than the lower bound stated in the theorem (which is natural, considering we simplified the setup).

We now need to handle the case when $\ell < h-1$. To that end, we shall show how to extend the auxiliary graph J and how to map certain types of paths in this extended graph into quadruples in $A^2 \times \binom{S}{h-1}^2$. An important observation, which we prove in detail later, is that since a typical $a \in A$ has about $\beta\binom{|S|}{\ell}$ neighbors in J, there must be some subset $Z \subset \{a \boxplus e\colon a \in A, e \in \binom{S}{\ell-1}\}$ such that for each $z \in Z$ there are, say, at least $\frac{\beta}{1000}|S|$ elements $s \in S$ with $z+s \in \Gamma$. Let us extend J by including Z as a third color class (making it a tripartite graph) and adding the edges $(z,w) \in Z \times \Gamma$ such that $w = z+s$ for some $s \in S$. In this extended graph J, consider a path P starting with an edge (a_1,b_1) from A to Γ, then a zig-zag $b_1,z_1,b_2,z_2,\dots,b_{h-\ell-1},z_{h-\ell-1},b_{h-\ell}$ between Γ and Z, then finally an edge $(b_{h-\ell},a_2)$ from Γ to A (see Figure 4).

Observe that the edges in the path P correspond to a set of $2h-2$ elements from S. Indeed, the first edge (a_1,b_1) is 'generated' by ℓ elements of S, while each of the $2h-2\ell-2$ edges in the zig-zag between Γ and Z are generated by single elements; finally, $(b_{h-\ell},a_2)$ is generated by ℓ elements. Regrouping these elements yields two $(h-1)$-tuples e_1,e_2 such that $a_1 \boxplus e_1 = a_2 \boxplus e_2$.

To summarize, we can define an auxiliary graph and count certain types of paths in it with the purpose of extracting estimates on the number of quadruples $(a_1,a_2,e_1,e_2) \in A^2 \times \binom{S}{h-1}^2$. In the full proof, we have to deal with the following complications:

(*1*) We must have $e_1 \cap e_2 = \emptyset$.

(*2*) Both $(h-1)$-tuples e_1,e_2 must belong to $\mathcal{H}$, which is very dense but not necessarily complete.

We address both of these issues in §6.2 by pre-processing the auxiliary graph.

To deal with (*1*), we count paths by starting at an arbitrary edge from A to Γ and appending edges to the path one by one, provided that each such edge is *valid*, that is, the generator(s) of the edge being added are distinct from all the previous generators. This is fairly straightforward for all the edges in the zig-zag between Γ and Z, since each edge has a single element from S as their generator. Only at the very last edge of the path, from Γ to A, is there a potential problem, namely, such edges are generated by ℓ elements from S, and it is possible that too many such ℓ-sets contain some generator that already appeared in the path. This issue is handled by introducing the concept of *congestion*, and bounding it in the processed auxiliary graph.

We establish a lower bound on the number of paths as above by using the minimum degree of the vertices in each class to determine how many valid edges there are at each step. On the other hand, we use the maximum degree of each class to estimate from above how many paths have, say, $e_1 \notin \mathcal{H}$. To ensure that there are many paths that satisfy (*2*) it will therefore be important to have minimum and maximum degrees as close as possible. Hence, one of the objectives of the pre-processing Lemma 25 is to find a relatively dense subgraph with balanced degrees.

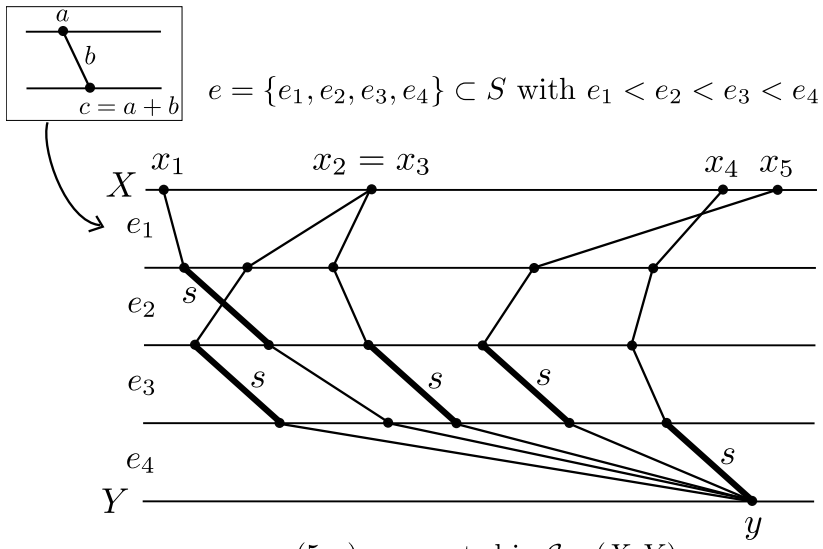

FIGURE 3. A congested vertex.

6.2. **Pre-processing stage.** In this subsection we introduce some key definitions and state the pre-processing lemma. In the next subsection, we use this lemma to establish Theorem 21. The proof of the pre-processing lemma is quite technical and is postponed until §6.4.

Roughly speaking, in the pre-processing stage we obtain a pair of bipartite graphs sharing a class with the property that *both* graphs are highly degree-regular in the shared class. This regularity is useful when we need to count, with great accuracy, certain special paths in §6.3.

Define, for any $S, X, Y \subset \mathbb{Z}$ and $k \geqslant 1$, the bipartite graph

$$\mathcal{C}_{S,k}(X,Y) = \left\{(a,b) \in X \times Y : b = a \boxplus e \text{ for some } e \in \binom{S}{k}\right\}. \tag{55}$$

Definition 24 (Congestion)**.** For k, S, X, and Y as above, $d \geqslant 1$, and $s \in S$, we say that a vertex $y \in Y$ is (d,s)-*congested in* $\mathcal{C}_{S,k}(X,Y)$ if there are at least d tuples $e \in \binom{S}{k}$ such that $s \in e$ and $y \boxminus e \in X$. We simply say that $y \in Y$ is d-*congested in* $\mathcal{C}_{S,k}(X,Y)$ if it is (d,s)-congested in $\mathcal{C}_{S,k}(X,Y)$ for some $s \in S$ (see Figure 3).

Lemma 25. *Let $h \geqslant 2$, $\ell \in [h-1]$, $\beta \in (0,1)$, n be a sufficiently large integer, and $d \geqslant (128h\log_2 n)^{\ell+2}$. Suppose that $S \in \mathcal{Z}_n^h$, with $|S| \geqslant 2h$, $X \subset [n]$, and $\Gamma_0 \subset [hn]$ are such that $\mathcal{C}_0 = \mathcal{C}_{S,\ell}(X,\Gamma_0)$ satisfies*

$$|\mathcal{C}_0| \geqslant \max\left\{\beta\binom{|S|}{\ell} \cdot |X|, \quad d \cdot |\Gamma_0|\right\}. \tag{56}$$

Then for some $1 \leqslant k \leqslant \ell$, condition (1) below holds.

(1) There exist sets $\bar{\Gamma}, Z \subset \mathbb{Z}$ and numbers δ_1 and δ_2 such that the graphs $\mathcal{C}_1 = \mathcal{C}_{S,k}(X,\bar{\Gamma})$ and $\mathcal{C}_2 = \mathcal{C}_{S,1}(Z,\bar{\Gamma})$ satisfy the following conditions:

(1-a) No vertex of $\bar{\Gamma}$ is $\left\lceil \dfrac{\delta_1}{16h\log_2 n} \right\rceil$-congested in $\mathcal{C}_1$.

(*1-b*) *For all* $b \in \bar{\Gamma}$,

$$\frac{d}{(4\log_2 n)(128h\log_2 n)^{\ell-k}} \leqslant \delta_1 \leqslant \deg_{\mathcal{C}_1}(b) \leqslant 2\delta_1.$$

(*1-c*) *For all* $b \in \bar{\Gamma}$,

$$4h \leqslant \delta_2 \leqslant \deg_{\mathcal{C}_2}(b) \leqslant 8\delta_2.$$

(*1-d*) *For all* $z \in Z$, *we have*

$$\deg_{\mathcal{C}_2}(z) \geqslant \frac{\beta\,|S|}{(\log_2 n)^5(256h^2\log_2 n)^{\ell-k}}.$$

(*1-e*) $|\mathcal{C}_1| \geqslant \dfrac{|\mathcal{C}_0|}{12(\log_2 n)^3 \cdot (128h\,|S|\log n)^{\ell-k}}$.

6.3. **Completing the proof of Theorem 21.** Recall that we are tasked with counting the number $\mathcal{Q}$ of quadruples $(a_1, a_2, e_1, e_2) \in A^2 \times \mathcal{H}^2$ that satisfy (54). We will recast this goal in terms of counting the number of certain paths in an auxiliary graph.

The following notation will be convenient in the arguments that follow. For a fixed B_k-set S,

$$\forall (x,y) \in \mathcal{C}_{S,k}(X,Y), \quad e_{y-x} = e_{S,k,y-x} \in \binom{S}{k} \text{ is the unique } k\text{-set satisfying } y = x \boxplus e, \tag{57}$$

where $\mathcal{C}_{S,k}(X,Y)$ is the graph defined in (55). Since S and k will be understood from context, we will use the short version e_{y-x}.

Let h, ℓ, β, n, d, S, A, $\mathcal{H}$, and Γ satisfy all the requirements in the statement of Theorem 21. We shall invoke Lemma 25 with h, ℓ, β, n, d, S, $X = A$, and $\Gamma_0 = \Gamma$. Note that the assumptions of Theorem 21 match those of Lemma 25, namely, we have $d \geqslant (128h\log_2 n)^{\ell+2}$, $\beta > 0$, $X \subset [n]$, $S \in \mathcal{Z}_n^h$, $|S| \gg 2h$, $\Gamma_0 \subset [hn]$, and $\mathcal{C}_0 = \mathcal{C}_{S,\ell}(X, \Gamma_0)$ satisfies

$$|\mathcal{C}_0| = \sum_{x \in X} \left| x \boxplus \binom{S}{\ell} \cap \Gamma \right| \geqslant \max\left\{ \beta \binom{|S|}{\ell} \cdot |X|, \quad d \cdot |\Gamma| \right\},$$

where the inequality follows by the exact same requirement imposed by Theorem 21 on $A = X$. Lemma 25 then implies that there exist $k \in [\ell]$, sets $\bar{\Gamma}, Z \subset \mathbb{Z}$, and numbers δ_1 and δ_2 such that all conditions in (*1*) hold.

6.3.1. *The case* $k = h-1$. It will be easier and instructive to deal first with the case $k = \ell = h-1$. Here we will only need the graph $\mathcal{C}_1 = \mathcal{C}_{S,\ell}(A, \bar{\Gamma})$.

Consider the map ϕ that takes each 2-path (a_1, b, a_2) in $\mathcal{C}_1$, with $a_1, a_2 \in A$, to the quadruple $(a_1, a_2, e_{b-a_1}, e_{b-a_2})$. First note that ϕ is one-to-one. Indeed, for any 2-path (a_1, c, a_2), we have $e_{b-a_1} = e_{c-a_1}$ if and only if $b = a_1 \boxplus e_{b-a_1} = a_1 \boxplus e_{c-a_1} = c$. Therefore, one possible way to obtain a lower bound on the number of quadruples satisfying (54) is to establish how many $b \in \bar{\Gamma}$ are such that $b = a_1 \boxplus e_1 = a_2 \boxplus e_2$ with $e_1, e_2 \in \mathcal{H}$ and $e_1 \cap e_2 = \emptyset$. This task is divided in two steps:

- We first estimate from below the number of 2-paths $(a_1, a_1 \boxplus e_1 = a_2 \boxplus e_2, a_2)$ with $e_1 \cap e_2 = \emptyset$. We will refer to such paths as *semi-special*.
- We then bound from above the number of those 2-paths counted before for which either $e_1 \notin \mathcal{H}$ or $e_2 \notin \mathcal{H}$.

Let us now perform the first step above. We start building the 2-path by taking an arbitrary edge $(a_1, b) \in \mathcal{C}_1$. Then we need to choose $a_2 \in N_{\mathcal{C}_1}(b)$ such that $e_{b-a_2} \cap e_{b-a_1} = \emptyset$ (recall the notation (57)). Consider the set

$$E_b := \big\{e_{b-a_2} \colon a_2 \in N_{\mathcal{C}_1}(b)\big\}$$

Condition (*1-a*) states that no vertex of $\bar{\Gamma}$ is $\lceil \frac{\delta_1}{16h\log_2 n} \rceil$-congested in $\mathcal{C}_1 = \mathcal{C}_{S,h-1}(A, \bar{\Gamma})$. Recalling Definition 24, this implies that for every $s \in S$, the number of $(h-1)$-tuples in E_b containing s is at most $\frac{\delta_1}{16h\log_2 n}$. Since there are $h-1$ values $s \in e_{b-a_1} \subset S$, it follows that there are at least $|E_b| - (h-1)\frac{\delta_1}{16h\log_2 n}$ elements $e_{b-a_2} \in E_b$ such that $e_{b-a_2} \cap e_{b-a_1} = \emptyset$. Since $|E_b| = \deg_{\mathcal{C}_1}(b)$ and, by condition (*1-b*), we have $\deg_{\mathcal{C}_1}(b) \geqslant \delta_1$, it follows that there are more than $\delta_1/2$ choices for a_2. In total, we have found more than

$$|\mathcal{C}_1|\frac{\delta_1}{2} \overset{(1\text{-}e)}{\geqslant} \frac{\delta_1 |\mathcal{C}_0|}{12(\log_2 n)^3} \geqslant \frac{\delta_1 |A|}{12(\log_2 n)^3}\beta\binom{|S|}{h-1} \tag{58}$$

semi-special paths.

Now we must exclude all the 2-paths (a_1, b, a_2) counted above such that either $e_{b-a_1} \notin \mathcal{H}$ or $e_{b-a_2} \notin \mathcal{H}$. For a fixed $e \in \binom{S}{h-1} \setminus \mathcal{H}$, the number of paths with $e_{b-a_1} = e$ is at most $|A| \cdot (2\delta_1)$. Indeed, if we first choose $a_1 \in A$, then $b = a \boxplus e$ is determined and by condition (*1-b*), there are at most $\deg_{\mathcal{C}_1}(b) \leqslant 2\delta_1$ choices for $a_2 \in N_{\mathcal{C}_1}(b)$. The case when $e_{b-a_2} = b$ is symmetric, and therefore there are at most

$$4\delta_1 |A| \cdot \left|\binom{S}{h-1} \setminus \mathcal{H}\right| \overset{(37)}{\leqslant} 4\delta_1 |A| \cdot \frac{\beta^h}{(\log_2 n)^{7h^2}}\binom{|S|}{h-1}$$

2-paths which fail to satisfy $e_{b-a_1}, e_{b-a_2} \in \mathcal{H}$. Therefore,

$$\mathcal{Q} \overset{(58)}{\geqslant} \frac{\delta_1 |A|}{12(\log_2 n)^3}\beta\binom{|S|}{h-1} - 4\delta_1 |A| \cdot \frac{\beta^h}{(\log_2 n)^{7h^2}}\binom{|S|}{h-1} \geqslant \frac{\beta\delta_1}{24(\log_2 n)^3}|A|\binom{|S|}{h-1},$$

which yields the conclusion of the theorem since $\delta_1 \geqslant \frac{d}{4\log_2 n}$ (see (*1-b*)).

6.3.2. *General case.* Since the simpler case when $k = \ell = h-1$ was handled in our warm-up (§6.3.1), we assume from now on that

$$k < h-1 \quad \text{and} \quad k \leqslant \ell \leqslant h-1.$$

Let us define an auxiliary tripartite graph G with parts $A, \bar{\Gamma}, Z$ defined as follows. We place a copy of $\mathcal{C}_1$ between A and $\bar{\Gamma}$ and a copy of $\mathcal{C}_2$ between Z and $\bar{\Gamma}$. Formally, we have

$$V(G) = (A \times \{1\}) \cup (\bar{\Gamma} \times \{2\}) \cup (Z \times \{3\}),$$
$$E(G) = \big\{\big((a,1),(b,2)\big) \colon (a,b) \in \mathcal{C}_1\big\} \cup \big\{\big((y,2),(z,3)\big) \colon (z,y) \in \mathcal{C}_2\big\},$$

but we will drop this cumbersome definition and simply assume that $V(G) = A \cup \bar{\Gamma} \cup Z$ where the elements of these three sets come from three disjoint copies of $\mathbb{Z}$.

Definition 26 (Special path; see Figure 4)**.** A *special path* in G is a path of the form

$$P = (a_1, b_1, z_1, b_2, z_2, \ldots, b_{h-k}, a_2)$$

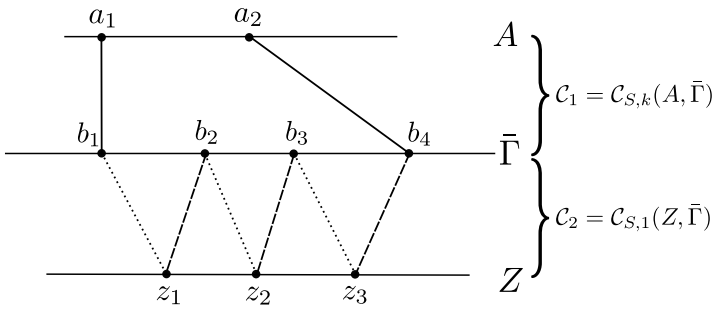

FIGURE 4. A special path in G.

such that, letting

$$e_1(P) = e_{b_1 - a_1} \cup \{b_{i+1} - z_i \colon i \in [h-k-1]\},$$
$$e_2(P) = e_{b_{h-k} - a_2} \cup \{b_i - z_i \colon i \in [h-k-1]\},$$

the following hold:

(*SP-1*) $a_1, a_2 \in A$,
(*SP-2*) $b_i \in \bar{\Gamma}$, for $i = 1, \dots h-k$,
(*SP-3*) $z_i \in Z$, for $i = 1, \dots, h-k-1$,
(*SP-4*) $|e_1(P) \cup e_2(P)| = 2(h-1)$,
(*SP-5*) $e_1(P) \in \mathcal{H}$, $e_2(P) \in \mathcal{H}$.

Note that if $a_1, a_2 \in A$ are connected by a special path P, then

$$\begin{aligned} a_1 \boxplus e_1(P) &= \underbrace{a_1 \boxplus e_{b_1 - a_1}}_{b_1} + (b_2 - z_1) + (b_3 - z_2) + \cdots + (b_{h-k} - z_{h-k-1}) \\ a_2 \boxplus e_2(P) &= \underbrace{a_2 \boxplus e_{b_{h-k} - a_2}}_{b_{h-k}} + (b_1 - z_1) + (b_2 - z_2) + \cdots + (b_{h-k-1} - z_{h-k-1}). \end{aligned} \tag{59}$$

Hence,

$$a_1 \boxplus e_1(P) = \sum_{i=1}^{h-k} b_i - \sum_{i=1}^{h-k-1} z_i = a_2 \boxplus e_2(P).$$

Together with the condition $|e_1(P) \cup e_2(P)| = 2(h-1)$ of (*SP-4*), which implies that $e_1(P) \cap e_2(P) = \emptyset$, and condition (*SP-5*), we see that $(a_1, a_2, e_1(P), e_2(P)) \in A^2 \times \mathcal{H}^2$ is a quadruple that satisfies (54). On the other hand, a quadruple $(a_1, a_2, e_1, e_2) \in A^2 \times \mathcal{H}^2$ corresponds to at most

$$\big((h-1)!\big)^2 < h^{2h}$$

special paths between a_1 and a_2 with $e_1(P) = e_1$ and $e_2(P) = e_2$. Indeed, after fixing orderings $e_1 = (e_{1,1}, \dots, e_{1,h-1})$ and $e_2 = (e_{2,1}, \dots, e_{2,h-1})$, the path $P = (a_1, b_1, z_1, b_2, z_2, \dots, b_{h-k}, a_2)$ defined

below is special (provided that it appears in the graph G):

$$(a_1, b_1 = a_1 + e_{1,1} + \cdots + e_{1,k}),$$
$$(b_1, z_1 = b_1 - e_{2,1}),$$
$$(z_1, b_2 = z_1 + e_{1,k+1}),$$
$$\vdots$$
$$(b_{h-k-1}, z_{h-k-1} = b_{h-k-1} - e_{2,h-k-1}),$$
$$(z_{h-k-1}, b_{h-k} = z_{h-k-1} + e_{1,h-1}),$$
$$(b_{h-k}, a_2 = b_{h-k} + e_{2,h-k} + e_{2,h-k+1} + \cdots + e_{2,h-1}).$$

Consequently, letting N be the number of special paths, we have

$$\mathcal{Q} \geqslant \frac{N}{h^{2h}}. \tag{60}$$

Our goal is now to provide a lower bound for N. To that end, we will proceed similarly to the warm-up case above, in two steps:

- We first estimate from below the number N^* of paths that satisfy (*SP-1*)–(*SP-4*) but not necessarily (*SP-5*). We shall call such paths *semi-special*.
- We then bound from above the number of semi-special paths P such that either $e_1(P) \notin \mathcal{H}$ or $e_2(P) \notin \mathcal{H}$.

The first edge of a semi-special path could be any $(a_1, b_1) \in \mathcal{C}_1$, hence there are $|\mathcal{C}_1|$ choices. Our choice of z_1 must be such that $z_1 \in N_{\mathcal{C}_2}(b_1)$ and $b_1 - z_1 \notin e_{b_1-a_1}$. According to condition (*1-c*), we have $\deg_{\mathcal{C}_2}(b_1) \geqslant \delta_2 \geqslant 4h$, and hence there are more than $\delta_2/2$ choices for z_1. Similarly, we must have $b_2 \in N_{\mathcal{C}_2}(z_1)$ and $b_2 - z_1 \notin e_{b_1-a_1} \cup \{b_1 - z_1\}$. According to condition (*1-d*), and the assumption that $\beta\,|S| \geqslant n^{1/(100h^2)}$, we have

$$\deg_{\mathcal{C}_2}(z_1) \geqslant \frac{\beta\,|S|}{(\log_2 n)^5 (256h^2 \log_2 n)^{\ell-k}} \geqslant \frac{n^{1/(100h^2)}}{(\log n)^{\Theta(1)}} \gg 4h$$

and hence there are more than $\deg_{\mathcal{C}_2}(z_1)/2$ choices for b_2. Continuing in this fashion, we construct a path arriving at $b_{h-k} \in \bar{\Gamma}$ which needs to be extended to some $a_2 \in N_{\mathcal{C}_1}(b_{h-k})$, under the restriction

$$e_{b_{h-k}-a_2} \cap \underbrace{\big(e_{b_1-a_1} \cup \{b_1 - z_1, b_2 - z_1, b_2 - z_2, b_3 - z_2, \ldots, b_{h-k} - z_{h-k-1}\}\big)}_{e'} = \emptyset.$$

Consider the set

$$E_{b_{h-k}} := \big\{e_{b_{h-k}-a_2} \colon a_2 \in N_{\mathcal{C}_1}(b_{h-k})\big\}.$$

Condition (*1-a*) states that no vertex of $\bar{\Gamma}$ is $\lceil \frac{\delta_1}{16h \log_2 n} \rceil$-congested in $\mathcal{C}_1 = \mathcal{C}_{S,k}(A, \bar{\Gamma})$. Recalling Definition 24, this implies that for all $s \in S$, the number of k-tuples in $E_{b_{h-k}}$ containing s is at most $\frac{\delta_1}{16h \log_2 n}$. Since there are fewer than $2h$ elements $s \in e' \subset S$, it follows that there are at least $|E_{b_{h-k}}| - \frac{\delta_1}{8 \log_2 n}$ tuples $e_{b_{h-k}-a_2} \in E_{b_{h-k}}$ such that $e_{b_{h-k}-a_2} \cap e' = \emptyset$. Since $|E_{b_{h-k}}| = \deg_{\mathcal{C}_1}(b_{h-k})$ and, by condition (*1-b*), we have $\deg_{\mathcal{C}_1}(b_{h-k}) \geqslant \delta_1$, it follows that more than $\delta_1/2$ elements $a_2 \in N_{\mathcal{C}_1}(b_{h-k})$ may be selected for the final vertex of the path. The above argument shows that:

- The number of choices for (a_1, b_1) is $|\mathcal{C}_1|$.
- Each element $z_1, \dots, z_{h-k-1}$ can be chosen from among at least $\delta_2/2$ alternatives.
- Each element b_i, $i \in \{2, 3, \dots, h-k\}$ can be chosen from among at least $\deg_{\mathcal{C}_2}(z_{i-1})/2 \geqslant \frac{\beta\,|S|}{2\cdot(\log_2 n)^5(256h^2\log_2 n)^{\ell-k}}$ alternatives.
- There are at least $\delta_1/2$ choices for the final vertex a_2.

Consequently,

$$N^* \geqslant |\mathcal{C}_1|\left(\frac{\delta_2}{2}\cdot\frac{\beta\,|S|}{2\cdot(\log_2 n)^5(256h^2\log_2 n)^{\ell-k}}\right)^{h-k-1}\cdot\frac{\delta_1}{2}.$$

From (*1-e*) we obtain

$$|\mathcal{C}_1| \geqslant \frac{|\mathcal{C}_0|}{12(\log_2 n)^3(128h\,|S|\log_2 n)^{\ell-k}} \overset{(56)}{\geqslant} \frac{\Omega(1)\cdot\beta\,|A|\binom{|S|}{\ell}}{|S|^{\ell-k}(\log_2 n)^{3+\ell-k}} = \frac{\Omega(1)\cdot\beta\,|A|\,|S|^k}{(\log_2 n)^{3+\ell-k}}.$$

Therefore, it follows that

$$\begin{aligned} N^* &\geqslant \Omega(1)\frac{\beta^{h-k}\,|A|\,|S|^{h-1}\delta_1\delta_2^{h-k-1}}{(\log_2 n)^{3+\ell-k+(5+\ell-k)(h-k-1)}} \\ &> \frac{\beta^h\delta_1\delta_2^{h-k-1}}{(\log_2 n)^{6h^2}}|A|\binom{|S|}{h-1}, \end{aligned} \tag{61}$$

where in the last inequality we used the fact that

$$3+\ell-k+(5+\ell-k)(h-k-1) < (5+\ell-k)(h-k) < (5+h)h < 6h^2,$$

and since this inequality is strict, the constant factors of the first inequality in (61) are easily absorbed by $(\log_2 n)^{6h^2-(3+\ell-k+(5+\ell-k)(h-k-1))}$.

Next we will bound the number of semi-special paths P such that $e_1(P) \notin \mathcal{H}$. Here we will rely on the almost regularity of the auxiliary graph to show that such semi-special paths are unusual. Fix an arbitrary $e \in \binom{S}{h-1} \setminus \mathcal{H}$ and one of the $(h-1)!$ orderings of its elements, say $(e_1, \dots, e_{h-1})$. Pick an element $a_1 \in A$ to be the first vertex of the path and notice that $b_1 = a_1 + e_1 + \cdots + e_k$ is determined. According to condition (*1-c*), there are at most $4\delta_2$ choices for $z_1 \in N_{\mathcal{C}_2}(b_1)$. Once z_1 is chosen, the value of b_2 must satisfy $b_2 = z_1 + e_{k+1}$, and so, continuing this construction process, we eventually arrive at b_{h-1}. From b_{h-1}, condition (*1-b*) shows that we have at most $2\delta_1$ candidates for $a_2 \in A$. To summarize, the number of semi-special paths P such that $e_1(P) \notin \mathcal{H}$ is at most

$$\left|\binom{S}{h-1}\setminus\mathcal{H}\right|(h-1)!|A|(4\delta_2)^{h-k-1}2\delta_1 \overset{(37)}{\leqslant} O(1)\cdot\frac{\beta^h\binom{|S|}{h-1}}{(\log_2 n)^{7h^2}}|A|\,\delta_1\delta_2^{h-k-1} \overset{(61)}{=} o(N^*). \tag{62}$$

Since the same is true for the number of semi-special paths P such that $e_2(P) \notin \mathcal{H}$, we conclude that $N \geqslant \frac{N^*}{2}$ and thus

$$\mathcal{Q} \overset{(60)}{\geqslant} \frac{N^*}{2h^{2h}} \geqslant \frac{1}{2h^{2h}}\frac{\beta^h\delta_1\delta_2^{h-k-1}}{(\log_2 n)^{6h^2}}|A|\binom{|S|}{h-1} \geqslant \frac{\beta^h d}{(\log_2 n)^{7h^2}}|A|\binom{|S|}{h-1},$$

where in the last inequality we used condition (*1-b*), that is, the fact that

$$\delta_1 \geqslant \frac{d}{(4\log_2 n)(128h\log_2 n)^{\ell-k}} \geqslant d(128h\log_2 n)^{-h}.$$

This completes the proof of Theorem 21. □

6.4. **Proof of Lemma 25.** The following proof, although extensive, is far from deep. It is essentially a recipe for obtaining (almost) degree-regular subgraphs and getting rid of some degeneracy which is present here in the form of congestion (see Def. 24). The length of the argument is simply due to the fact that once a subgraph is taken to force degrees in one class to behave well, it affects the degrees in the other class, so that we need to take several 'regularization' steps in the appropriate order to obtain the desired graph.

Proof of Lemma 25. We start this proof by letting $k \in [\ell]$ be the smallest integer such that the following holds:

- There exist $\Gamma \subset \mathbb{Z}$ and

$$\alpha \geqslant \beta \cdot (256h^2 \log_2 n)^{k-\ell}, \quad D \geqslant d \cdot (128h \log_2 n)^{k-\ell}, \tag{63}$$

  such that $\mathcal{C} = \mathcal{C}_{S,k}(X, \Gamma)$ satisfies

$$|\mathcal{C}| \geqslant \max\left\{\alpha\binom{|S|}{k} \cdot |X|, \quad D \cdot |\Gamma|, \quad \frac{|\mathcal{C}_0|}{(128h\,|S| \log n)^{\ell-k}}\right\}. \tag{64}$$

Note that such a minimum value of k must exist, since for $k = \ell$, all these conditions are satisfied by the assumptions of the lemma on $\Gamma = \Gamma_0$, $\alpha = \beta$, and $D = d$. We then fix k, Γ, α, D, and $\mathcal{C}$ as above and define

$$\Gamma^{\mathtt{cong}} = \left\{b \in \Gamma \colon b \text{ is } \left\lceil \frac{\deg_{\mathcal{C}}(b)}{32h \log_2 n} \right\rceil\text{-congested in } \mathcal{C}\right\}. \tag{65}$$

Claim 4. *We have*

$$|\mathcal{C}_{S,k}(X, \Gamma^{\mathtt{cong}})| < \frac{|\mathcal{C}|}{2}. \tag{66}$$

Proof. First notice that if $k = 1$, then no vertex can be d-congested in $\mathcal{C}$ for any $d > 1$. Hence, the only vertices in $\Gamma^{\mathtt{cong}}$ are those $b \in \Gamma$ with $\deg_{\mathcal{C}}(b) \leqslant 32h \log_2 n$. Since $D \geqslant d \cdot (128h \log_2 n)^{k-\ell} \geqslant (128h \log_2 n)^{k+2}$, we have

$$|\mathcal{C}_{S,k}(X, \Gamma^{\mathtt{cong}})| = \sum_{b \in \Gamma^{\mathtt{cong}}} \deg_{\mathcal{C}}(b) \leqslant 32h \log_2 n \cdot |\Gamma^{\mathtt{cong}}| \leqslant 32h \log_2 n \cdot |\Gamma| \leqslant \frac{D}{4} \cdot |\Gamma| \overset{(64)}{\leqslant} \frac{|\mathcal{C}|}{4},$$

which establishes the claim for $k = 1$. Hence let us assume that $k \geqslant 2$ and, for the sake of a contradiction, that (66) fails. We will show that this assumption contradicts the minimality of k.

For every $b \in \Gamma^{\mathtt{cong}}$, let $s_b \in S$ be a canonical choice of an element such that b is $\big(\lceil \frac{\deg_{\mathcal{C}}(b)}{32h \log_2 n} \rceil, s_b\big)$-congested in $\mathcal{C}$. Let

$$\begin{aligned} \Gamma^{\mathtt{cong}}_+ &= \big\{b \in \Gamma^{\mathtt{cong}} \colon \deg_{\mathcal{C}}(b) \geqslant D/4\big\}, \\ \Gamma' &= \big\{b - s_b \colon b \in \Gamma^{\mathtt{cong}}_+\big\}, \\ \mathcal{C}' &= \mathcal{C}_{S,k-1}(X, \Gamma'). \end{aligned} \tag{67}$$

Note that by construction, for any $y = b - s_b \in \Gamma'$, there must be at least $d = \lceil \deg_{\mathcal{C}}(b)/(32h \log_2 n) \rceil$ distinct tuples $e_1, \ldots, e_d \in \binom{S}{k}$ such that $s_b \in e_i$ and $b \boxminus e_i \in X$ for all $i = 1, \ldots, d$. Hence, setting $f_i = e_i \setminus \{s_b\}$ for each i, we obtain a collection of d distinct $(k-1)$-tuples such that $y \boxminus f_i \in X$

for all i. Since S is a B_{k-1}-set, this implies that $\deg_{\mathcal{C}'}(y) \geqslant d$. In general, we then have

$$\forall y \in \Gamma', \quad \deg_{\mathcal{C}'}(y) \geqslant \max_{\substack{b \in \Gamma_+^{\mathtt{cong}} \\ y = b - s_b}} \left\{ \left\lceil \frac{\deg_{\mathcal{C}}(b)}{32h \log_2 n} \right\rceil \right\} \overset{(67)}{\geqslant} \frac{D}{128h \log_2 n}. \tag{68}$$

For each $y \in \Gamma'$, there are at most $|S|$ representations of the form $y = b - s_b$ with $b \in \Gamma_+^{\mathtt{cong}}$. Therefore the maximum in the above inequality may be replaced by the average over all $b \in \Gamma_+^{\mathtt{cong}}$ such that $y = b - s_b$, yielding

$$\begin{aligned} |\mathcal{C}'| &= \sum_{y \in \Gamma'} \deg_{\mathcal{C}'}(y) \geqslant \sum_{y \in \Gamma'} \frac{1}{|S|} \sum_{\substack{b \in \Gamma_+^{\mathtt{cong}} \\ y = b - s_b}} \frac{\deg_{\mathcal{C}}(b)}{32h \log_2 n} \\ &= \frac{1}{32h\,|S| \log_2 n} \sum_{b \in \Gamma_+^{\mathtt{cong}}} \deg_{\mathcal{C}}(b) = \frac{|\mathcal{C}_{S,k}(X, \Gamma_+^{\mathtt{cong}})|}{32h\,|S| \log_2 n}. \end{aligned}$$

Since we assumed the converse of (66), it follows from (64) and the definitions of $\Gamma^{\mathtt{cong}}$ and $\Gamma_+^{\mathtt{cong}}$ that

$$|\mathcal{C}_{S,k}(X, \Gamma_+^{\mathtt{cong}})| = |\mathcal{C}_{S,k}(X, \Gamma^{\mathtt{cong}})| - |\mathcal{C}_{S,k}(X, \Gamma^{\mathtt{cong}} \setminus \Gamma_+^{\mathtt{cong}})| \overset{(67)}{\geqslant} \frac{|\mathcal{C}|}{2} - \frac{D}{4}|\Gamma| \overset{(64)}{\geqslant} \frac{|\mathcal{C}|}{4}.$$

We conclude that

$$|\mathcal{C}'| \geqslant \frac{|\mathcal{C}|}{128h\,|S| \log_2 n} \geqslant \alpha \binom{|S|}{k} \cdot |X| / (128h\,|S| \log_2 n) = \alpha' \binom{|S|}{k-1} \cdot |X|,$$

where

$$\alpha' := \frac{\alpha \binom{|S|}{k}}{128h\,|S| \binom{|S|}{k-1} \log_2 n} \geqslant \frac{\alpha(|S| - k + 1)}{128hk\,|S| \log_2 n} \geqslant \frac{\alpha}{256h^2 \log_2 n}.$$

Together with (68), we obtain

$$|\mathcal{C}'| \geqslant \max\left\{ \alpha' \binom{|S|}{k-1} \cdot |X|, \quad \frac{D}{128h \log_2 n} |\Gamma'|, \quad \frac{|\mathcal{C}|}{128h\,|S| \log_2 n} \right\},$$

which contradicts the minimality of k (see (64)). □

Define for all $j \geqslant 0$,

$$\Gamma_j = \left\{ b \in \Gamma \setminus \Gamma^{\mathtt{cong}} \colon \deg_{\mathcal{C}}(b) \in [2^j, 2^{j+1} - 1] \right\}. \tag{69}$$

Since the maximum degree in $\mathcal{C}$ is bounded by $|S|^k$ and S is a B_h-set, implying that $|S|^k \leqslant |S|^{h-1} \ll n$, we have $B_j = \emptyset$ for $j \geqslant \log_2 n$. Pick $0 \leqslant j^* \leqslant \log_2 n$ such that $|\mathcal{C}_{S,k}(X, \Gamma_{j^*})|$ is maximum and let $\Gamma^* = \Gamma_{j^*}$. We then have

$$\Gamma^* \subset \Gamma \setminus \Gamma^{\mathtt{cong}}, \qquad \forall b \in \Gamma^*,\ \deg_{\mathcal{C}}(b) \in [2^{j^*}, 2^{j^*+1} - 1], \quad \text{and} \quad |\mathcal{C}_{S,k}(X, \Gamma^*)| \overset{(66)}{\geqslant} \frac{|\mathcal{C}|}{2 \log_2 n}. \tag{70}$$

Thus,

$$\frac{D\,|\Gamma|}{2 \log_2 n} \overset{(64)}{\leqslant} \frac{|\mathcal{C}|}{2 \log_2 n} \leqslant |\mathcal{C}_{S,k}(X, \Gamma^*)| \leqslant (2^{j^*+1} - 1) \cdot |\Gamma^*| \leqslant 2^{j^*+1} \cdot |\Gamma|,$$

which implies that $2^{j^*} \geqslant D/(4\log_2 n)$. Let $\delta_1 = 2^{j^*}$ and observe that

$$\delta_1 \geqslant \frac{D}{4\log_2 n} \quad \text{and} \quad \forall b \in \Gamma^*,\ \deg_{\mathcal{C}}(b) \in [\delta_1, 2\delta_1 - 1]. \tag{71}$$

Claim 5. *For any $\bar{\Gamma} \subset \Gamma^*$, we have*

$$|\mathcal{C}_{S,k}(X, \bar{\Gamma})| \geqslant \frac{|\bar{\Gamma}|}{4\,|\Gamma^*|\log_2 n}|\mathcal{C}|.$$

Proof. We have

$$2\delta_1\,|\Gamma^*| \overset{(71)}{\geqslant} |C_{S,k}(X, \Gamma^*)| \overset{(70)}{\geqslant} \frac{|\mathcal{C}|}{2\log_2 n}$$

and

$$|\mathcal{C}_{S,k}(X, \bar{\Gamma})| \overset{(71)}{\geqslant} \delta_1\,|\bar{\Gamma}|.$$

The lower bound on δ_1 obtained from the first inequality, when substituted into the second inequality immediately yields the claim. □

Claim 6. *Any subset $\bar{\Gamma} \subset \Gamma^*$ satisfies conditions (1-a) and (1-b).*

Proof. In view of (71), condition (*1-b*) follows immediately for any subset of Γ^*. We now check that condition (*1-a*) is also satisfied. Recall (65) and (70). By definition, every $b \in \bar{\Gamma} \subset \Gamma \setminus \Gamma^{\mathtt{cong}}$ is *not* $\lceil \frac{\deg_{\mathcal{C}}(b)}{32h\log_2 n} \rceil$-congested in $\mathcal{C}$. Since $\deg_{\mathcal{C}_1}(b) = \deg_{\mathcal{C}}(b) \leqslant 2\delta_1$, it follows that b is also not $\lceil \frac{\delta_1}{16h\log_2 n} \rceil$-congested in $\mathcal{C}_1 = \mathcal{C}_{S,k}(X, \bar{\Gamma})$. □

In view of Claim 6, it suffices to construct subsets $Z \subset \mathbb{Z}$ and $\bar{\Gamma} \subset \Gamma^*$ that will satisfy properties (*1-c*)–(*1-e*). The next claim will bring us closer to that goal.

Claim 7. *There are sets $\bar{\Gamma} \subset \Gamma^*$ and $Z \subset \mathbb{Z}$ with $|\bar{\Gamma}| \geqslant \dfrac{|\Gamma^*|}{3(\log_2 n)^2}$ and an integer δ_2 such that conditions (1-c) and (1-d) hold.*

Proof. Consider the auxiliary $(k+1)$-partite graph with parts

$$X_0 = X, \qquad X_1 = X_0 + S, \qquad \dots \qquad X_{k-1} = X_{k-2} + S, \qquad X_k = \Gamma^*$$

and edges joining $a \in X_i$ and $b \in X_{i+1}$ whenever $b - a \in S$, for $i \in \{0, \dots, k-1\}$; see Figure 5. Let us call a path of length $m \in [k]$ in this graph *proper* if it is of the form $(x_0, x_1, \dots, x_m)$ with $x_i \in X_i$ for all $i \in \{0, 1, \dots, m\}$ and, moreover, the differences $x_i - x_{i-1}$ are all distinct for $i \in [m]$.

Notice that for each vertex $b \in X_k$, there are exactly $k!\deg_{\mathcal{C}}(b)$ proper paths of length k ending at b. Indeed, since S is a B_k-set, for each $a \in N_{\mathcal{C}}(b) \subset X = X_0$ there exists a unique $e \in \binom{S}{k}$ such that $b = a \boxplus e$. Any ordering $(e_1, \dots, e_k)$ of e corresponds to the proper path $(a, a + e_1, a + e_1 + e_2, \dots, b)$. Conversely, if $(a, x_1, \dots, x_{k-1}, b)$ is a proper path, then $a \in N_{\mathcal{C}}(b)$ and the set of consecutive differences in the path gives an ordering of some $e \in \binom{S}{k}$ such that $a \boxplus e = b$. We will use this fact shortly.

For each $u \in X_{k-1}$, let P_u denote the number of proper paths (of length $k-1$) ending at u. For each $j \geqslant 0$, let

$$X_{k-1,j} = \big\{u \in X_{k-1} \colon P_u \in [2^j, 2^{j+1} - 1]\big\}. \tag{72}$$

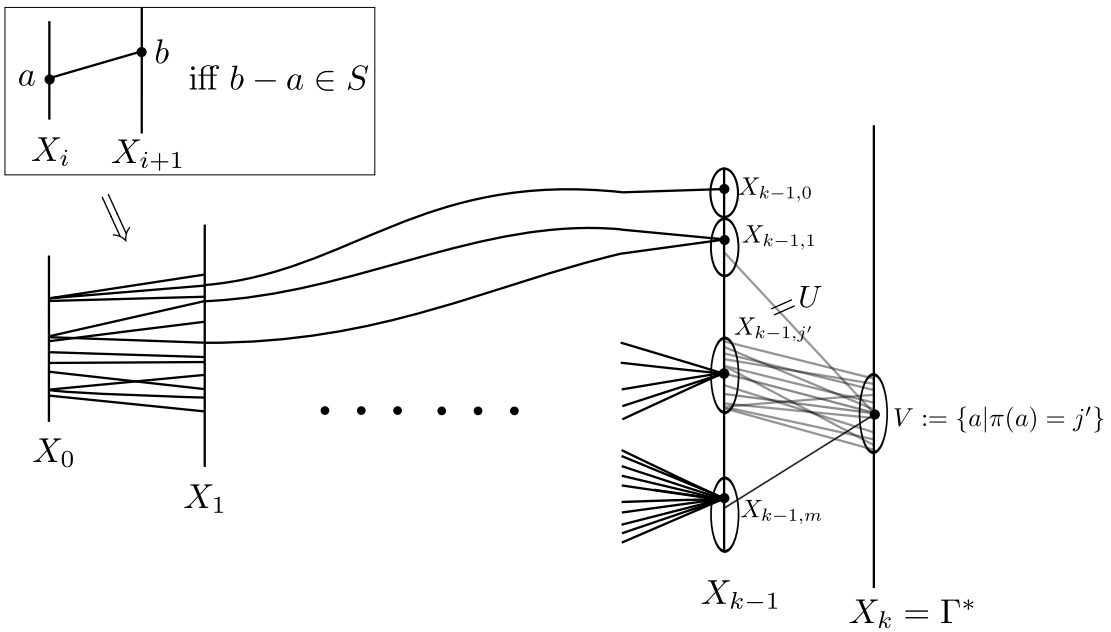

FIGURE 5. The auxiliary $(k+1)$-partite graph with parts $X_0, X_1, \ldots, X_k$.

Since S is a B_k-set, we have $P_u \leqslant |S|^{k-1} < n$ for all $u \in X_{k-1}$. Hence, $X_{k-1,j} = \emptyset$ whenever $j \geqslant \log_2 n$ and therefore

$$X_{k-1} = \bigcup_{j=0}^{\log_2 n} X_{k-1,j}.$$

For every $b \in X_k$, let $\pi(b) \in [\log_2 n]$ be the index such that among all the proper paths (of length k) ending at b, the largest number visits the set $X_{k-1,\pi(b)}$. In particular, more than $k! \deg_{\mathcal{C}}(b)/\log_2 n$ proper paths ending at b have a final edge of the form (u, b) for some $u \in X_{k-1,\pi(b)}$.

Let j' be such that $\pi^{-1}(j')$ has maximum size. For brevity, let $U = X_{k-1,j'}$ and $V = \pi^{-1}(j')$. By construction, we have that, for every $v \in V$, the number of proper paths passing through U and ending at v is at least

$$\frac{k! \deg_{\mathcal{C}}(v)}{\log_2 n}. \tag{73}$$

We will prove that

$$\deg_{\mathcal{C}_{S,1}(U,V)}(v) \geqslant 16h \quad \text{for every } v \in V. \tag{74}$$

Suppose for the sake of a contradiction, that for some $v \in V$, the above inequality fails. By (73), there must be some $u \in U$ such that at least

$$\frac{k! \deg_{\mathcal{C}}(v)}{16h \log_2 n}$$

proper paths end in (u, v). However, as we will show, this implies that v is $(d, v-u)$-congested in $\mathcal{C}$ with $d \geqslant \frac{\deg_{\mathcal{C}}(v)}{16h \log_2 n}$, which contradicts the fact that $V \subset X_k = \Gamma^*$ is disjoint from $\Gamma^{\mathtt{cong}}$. To show that v is congested, note that for each proper path $(x_0, \ldots, x_{k-2}, u, v)$, the k-set $e = \{x_1 - x_0, \ldots, x_{k-2} - x_{k-3}, u - x_{k-2}, v - u\} \in \binom{S}{k}$ satisfies $v - u \in e$ and $v \boxminus e = x_0 \in X_0 = X$.

Since the same k-set can be obtained by at most $(k-1)!$ proper paths ending in (u,v), there must be at least $\frac{k!\deg_{\mathcal{C}}(v)}{(k-1)!16h\log_2 n}$ such k-sets, which proves that v is $(d, v-u)$-congested as claimed. The obtained contradiction proves that (74) holds.

Since we chose $V = \pi^{-1}(j')$ of maximum size, we have

$$|V| \geqslant \frac{X_k}{\log_2 n} = \frac{|\Gamma^*|}{\log_2 n}.$$

Define, for every $m \geqslant 0$

$$V_m = \left\{v \in V\colon \deg_{\mathcal{C}_{S,1}(U,V)}(v) \in [16h\cdot 2^m, 16h\cdot 2^{m+1}-1]\right\} \tag{75}$$

and similarly as before, notice that $V_m = \emptyset$ for $m \geqslant \log_2 n \geqslant \log_2 |S|$. Observe also that (74) implies that

$$V = \bigcup_{m=0}^{\log_2 n} V_m.$$

Now, pick an m' with $0 \leqslant m' \leqslant \log_2 n$ such that

$$|V_{m'}| \geqslant \frac{|V|}{\log_2 n} \geqslant \frac{|\Gamma^*|}{(\log_2 n)^2}. \tag{76}$$

Using Claim 5, (73), and (76), we obtain that the total number N of proper paths (of length k) whose final edge is a pair in $U \times V_{m'}$ satisfies

$$N \geqslant \sum_{v\in V_{m'}} \frac{k!\deg_{\mathcal{C}}(v)}{\log_2 n} \geqslant \frac{k!}{\log_2 n}\cdot |\mathcal{C}_{S,k}(X, V_{m'})| \geqslant \frac{k!}{\log_2 n}\cdot \frac{|V_{m'}|}{4|\Gamma^*|\log_2 n}|\mathcal{C}| \overset{(76)}{\geqslant} \frac{k!\,|\mathcal{C}|}{4(\log n)^4}. \tag{77}$$

Since there are fewer than

$$|X|\,|S|^{k-1}$$

proper $(k-1)$-paths, by (72) and our choice of $U = X_{k-1,j'}$, we must have

$$2^{j'}|U| \leqslant \sum_{u\in U} P_u \leqslant |X|\,|S|^{k-1}.$$

Since

$$N \leqslant \sum_{u\in U} P_u \deg_{\mathcal{C}_{S,1}(U,V_{m'})}(u) \leqslant 2^{j'+1}\sum_{u\in U}\deg_{\mathcal{C}_{S,1}(U,V_{m'})}(u),$$

it follows that

$$\begin{aligned}\frac{1}{|U|}\sum_{u\in U}\deg_{\mathcal{C}_{S,1}(U,V_{m'})}(u) \;&\geqslant\; \frac{N}{2^{j'+1}|U|} \geqslant \frac{N}{2\,|X|\,|S|^{k-1}} \overset{(77)}{\geqslant} \frac{k!\,|\mathcal{C}|}{8(\log_2 n)^4|X|\,|S|^{k-1}} \\ &\overset{(64)}{\geqslant} \frac{k!\,\alpha\binom{|S|}{k}\cdot|X|}{8(\log_2 n)^4|X|\,|S|^{k-1}} \gg \frac{\alpha\,|S|}{(\log_2 n)^5}.\end{aligned} \tag{78}$$

We are now ready to construct the sets $Z \subset U$, $\bar{\Gamma} \subset V_{m'}$. Set

$$\delta_2 = 4h\cdot 2^{m'}. \tag{79}$$

We begin by setting $Z = U$, $\bar{\Gamma} = V_{m'}$ and then successively remove vertices:

- $z \in Z$ such that $\deg_{\mathcal{C}_{S,1}(Z,\bar{\Gamma})}(z) < \frac{\alpha\,|S|}{(\log_2 n)^5}$ and

- $b \in \bar{\Gamma}$ such that $\deg_{\mathcal{C}_{S,1}(Z,\bar{\Gamma})}(b) < \delta_2$.

From (63) we have $\alpha \geqslant \beta \cdot (256h^2 \log_2 n)^{k-\ell}$ and from (79), we have $\delta_2 \geqslant 4h$. Moreover, by the definition of $V_{m'}$ in (75), we have $\deg_{\mathcal{C}_{S,1}(Z,\bar{\Gamma})}(b) \leqslant 8\delta_2$ for all $b \in \bar{\Gamma} \subset V_{m'}$. Consequently, *if* the sets obtained from the above iterative process are not empty, they must satisfy (*1-c*) and (*1-d*). We shall show an explicit lower bound on $|\bar{\Gamma}|$.

Note that the total number of edges lost because a vertex $z \in Z$ was deleted is bounded by

$$|U| \cdot \frac{\alpha\,|S|}{(\log_2 n)^5} \overset{(78)}{\ll} |\mathcal{C}_{S,1}(U, V_{m'})|$$

The number of edges lost due to a vertex $b \in \bar{\Gamma}$ being deleted is bounded by

$$|V_{m'}|\,\delta_2 \overset{(75),(79)}{<} \sum_{v \in V_{m'}} \frac{\deg_{\mathcal{C}_{S,1}(U,V_{m'})}(v)}{4} \leqslant \frac{|\mathcal{C}_{S,1}(U, V_{m'})|}{4}.$$

We conclude that fewer than $|\mathcal{C}_{S,1}(U, V_{m'})|/3$ edges were lost in total. Therefore

$$|\bar{\Gamma}| \cdot 8\delta_2 \geqslant \sum_{b \in \bar{\Gamma}} \deg_{\mathcal{C}_{S,1}(Z,\bar{\Gamma})}(b) = |\mathcal{C}_{S,1}(Z, \bar{\Gamma})| \geqslant \frac{2\,|\mathcal{C}_{S,1}(U, V_{m'})|}{3} \overset{(75),(79)}{\geqslant} \frac{2\,|V_{m'}| \cdot 4\delta_2}{3}.$$

It follows that

$$|\bar{\Gamma}| \geqslant \frac{|V_{m'}|}{3} \overset{(76)}{\geqslant} \frac{|\Gamma^*|}{3(\log_2 n)^2}.$$

This completes the proof of the claim. □

Let $\bar{\Gamma}$ be the set whose existence is asserted by Claim 7. By Claim 6, it satisfies (*1-a*)–(*1-d*). By Claim 5,

$$|\mathcal{C}_1| = |\mathcal{C}_{S,k}(X, \bar{\Gamma})| \geqslant \frac{|\mathcal{C}|}{12(\log_2 n)^3} \overset{(64)}{\geqslant} \frac{|\mathcal{C}_0|}{12(\log_2 n)^3 \cdot (128h\,|S| \log n)^{\ell-k}},$$

which establishes (*1-e*) and completes the proof of the pre-processing lemma. □

7. Concluding Remarks

In this paper, we have established essentially tight bounds for the number of B_h-sets contained in the set $\{1, \ldots, n\}$ of almost every given cardinality t. There remains, however, a small 'threshold gap', that is, an interval of values of t for which the precise asymptotics of $|\mathcal{Z}_n^h(t)|$ is not determined here. This interval is of the form $[\varepsilon(n \log n)^{1/(2h-1)}, n^{1/(2h-1)+\varepsilon}]$, where $\varepsilon = \varepsilon(n)$ is some function of n that slowly converges to 0 as $n \to \infty$. Outside this interval, the value of $|\mathcal{Z}_n^h(t)|$ is determined within an $n^{\varepsilon t}$ multiplicative factor (and more precisely for t sufficiently far from the endpoints of that interval).

There are therefore two directions in which our result could be refined. The first of them would be to improve the upper bound on $|\mathcal{Z}_n^h(t)|$ from $(n/t^{h+o(1)})^t$ to $(f(n)n/t^h)^t$ for some small explicit function $f(n)$; our methods give $f(n) = n^{c/\log\log n}$ for some small positive constant c. The second direction would be to narrow the threshold gap.

It is conceivable that our methods could be used to obtain somewhat stronger upper bounds, however it would most likely require a great deal of effort. In order to improve our estimates, one needs to 'balance' the values of αs and λs better. The sequence of λs must be longer so that the

ratio of consecutive values allows to obtain better bounds in (49). At the same time, the sequence of αs has to decrease quickly enough so that condition (37) in Theorem 21 is satisfied when we apply it in the proof of Theorem 9; see (48).

As for narrowing the gap, one could adapt the proof given in §5.1 by requiring A to be larger, therefore allowing t to be smaller. The cost one would pay for that is a weaker upper bound on $|\mathcal{Z}_n^h(t)|$, which would be a result of applying Lemma 8 with larger values of R. The obtained upper bounds would still be similar to those proved by Theorem 4. In view of the lower bound of Proposition 2 (i) and Proposition 3, it is clearly not possible to reduce t below $(n \log n)^{1/(2h-1)}$. A careful analysis of our proof shows that (45) is where a lower bound on t of that form is required. More precisely, for our application of Lemma 8 to work, we need $q = o(t)$, $|S| = o(t)$, and $\beta q = \Omega(\log n)$. For that reason, t must be at least $n^{1/(2h-1)}(\log n)^{\Theta(1)}$ for (45) to yield anything useful. In fact, it seems that any proof based on Lemma 8 would require a threshold gap with factor at least $\log n$. Still, we believe that the following is true:

Conjecture 27. *For every $h \geqslant 2$, there exists a constant C_h such that*

$$|\mathcal{Z}_n^h(t)| \leqslant \left(\frac{C_h n}{t^h}\right)^t$$

for every n and t satisfying $t \geqslant C_h(n \log n)^{1/(2h-1)}$.

Another interesting question is whether Theorem 9 is needed at all. It is conceivable that the set $\widetilde{S}$ is always small (in the sense that (9) is vacuous). Although proving such a statement would not in itself improve the bounds we obtained in this paper, it would help to understand better the structure of B_h-sets.

7.1. **Related work.** Some recent results in extremal combinatorics have used the so-called *containers method* based on the main results of [2, 33]. This method was recently applied by Morris and Saxton [30] to show that for every integer $h \geqslant 2$, the number of C_{2h}-free graphs with vertex set $[n]$ is at most $2^{O(n^{1+1/h})}$, which extends the results of [20, 21]. In fact, they proved that for every $m \gg n^{1+1/(2h-1)}(\log n)^2$, the number $f_{n,m}(C_{2h})$ of C_{2h}-free graphs with vertex set $[n]$ that have exactly m edges satisfies

$$f_{n,m}(C_{2h}) \leqslant \left(\frac{Cn^{h+1}}{m^h} \cdot \left(\log \frac{n^{h+1}}{m^h}\right)^{h-1}\right)^m. \tag{80}$$

The problems of counting C_{2h}-free graphs and B_h-sets seem to be related. Given a t-element B_h-set $T \subset [n]$, one may define an auxiliary bipartite graph G_T on $[hn] \times \{1, 2\}$ by placing an edge between $(x, 1)$ and $(y, 2)$ whenever $y - x \pmod n$ is an element of T. This graph G_T has htn edges and is 'essentially' C_{2h}-free[3]. In particular, the bound (80) may be viewed as an analogue of our Theorem 4. However, we are not aware of any rigorous connection between these two results.

One might still ask whether the argument of [30] could be adapted to our setting. As in most applications of the containers method, the heart of [30] is proving a sufficiently strong supersaturation

[3]The graph G_T contains $\Theta(nt^h)$ copies of C_{2h} which correspond to 'trivial' equalities of the form $a_1 + \ldots + a_h = a_{\pi(1)} + \ldots + a_{\pi(h)}$, where π is some permutation of $[h]$.

result for copies of C_{2h} in n-vertex graphs with more that $Dn^{1+1/h}$ edges; see [30, Theorem 1.5]. It is conceivable that one could obtain some supersaturation theorem for solutions to the equation $a_1 + \ldots + a_h = b_1 + \ldots + b_h$ in subsets of $[n]$ with more than $Dn^{1/h}$ elements using the methods of [30]. However, a supersaturation statement that would be necessary for our application does not seem to follow from [30, Theorem 1.5] as it is unclear how to 'translate' condition (b) there to our setting. We did not pursue this direction further, mainly because our research leading to the current work was carried out largely in parallel to [30].

The obvious advantage of the approach of Morris and Saxton is that their upper bound on $f_{n,m}(C_{2h})$ is larger than the (theoretical) lower bound[4] of $\left(cn^{h+1}/m^h\right)^m$ only by a factor of $\left(\log(n^{h+1}/m^h)\right)^{(h-1)m}$. On the other hand, our approach has the advantage of being entirely self-contained, since it relies merely on Lemma 8, which is a simple result on graphs as opposed to the much more involved hypergraph version of that result proved in [2, 33].

Acknowledgement. The authors are indebted to the anonymous referee, whose careful reading of the paper and valuable comments and suggestions helped us greatly. In particular, we would like to thank the referee for pointing out a mistake in the statement of Conjecture 27 in the previous version of the paper.

References

1. M. Ajtai, J. Komlós, J. Pintz, J. Spencer, and E. Szemerédi, *Extremal uncrowded hypergraphs*, J. Combin. Theory Ser. A **32** (1982), 321–335.
2. J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc. **28** (2015), 669–709.
3. R. C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helv. **37** (1962/1963), 141–147.
4. 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, pp. 61–79.
5. S. Chen, *On the size of finite Sidon sequences*, Proc. Amer. Math. Soc. **121** (1994), 353–356.
6. S. Chowla, *Solution of a problem of Erdős and Turán in additive-number theory*, Proc. Nat. Acad. Sci. India. Sect. A. **14** (1944), 1–2.
7. J. Cilleruelo, *New upper bounds for finite B_h sequences*, Adv. Math. **159** (2001), 1–17.
8. D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, Ann. of Math. (2) **184** (2016), 367–454.
9. D. Dellamonica Jr., Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij, *The number of B_3-sets of a given cardinality*, J. Combin. Theory Ser. A **142** (2016), 44–76.
10. ———, *On the number of B_h-sets*, Combin. Probab. Comput. **25** (2016), 108–129.
11. R. A. Duke, H. Lefmann, and V. Rödl, *On uncrowded hypergraphs*, Random Structures Algorithms **6** (1995), 209–212.
12. A. G. D'yachkov and V. V. Rykov, *B_s-sequences*, Mat. Zametki **36** (1984), 593–601.
13. P. Erdős, *On a problem of Sidon in additive number theory and on some related problems. Addendum*, J. London Math. Soc. **19** (1944), 208.
14. P. Erdős and P. Turán, *On a problem of Sidon in additive number theory, and on some related problems*, J. London Math. Soc. **16** (1941), 212–215.
15. B. Green, *The number of squares and $B_h[g]$ sets*, Acta Arith. **100** (2001), 365–390.

[4]One could derive such a bound in a similar fashion to [30, Proposition 1.4] and our Lemma 23 from the existence of n-vertex graphs with $\Omega(n^{1+1/h})$ edges which admit no non-backtracking closed walks of length $2h$.

16. H. Halberstam and K. F. Roth, *Sequences*, second ed., Springer-Verlag, New York, 1983.
17. P. E. Haxell, Y. Kohayakawa, and T. Łuczak, *Turán's extremal problem in random graphs: forbidding even cycles*, J. Combin. Theory Ser. B **64** (1995), 273–287.
18. ———, *Turán's extremal problem in random graphs: forbidding odd cycles*, Combinatorica **16** (1996), 107–122.
19. X. D. Jia, *On finite Sidon sequences*, J. Number Theory **44** (1993), 84–92.
20. D. J. Kleitman and K. J. Winston, *On the number of graphs without* 4*-cycles*, Discrete Math. **41** (1982), 167–172.
21. D. J. Kleitman and D. B. Wilson, *On the number of graphs which lack small cycles*, manuscript, 15 pp, 1996.
22. Y. Kohayakawa, B. Kreuter, and A. Steger, *An extremal problem for random graphs and the number of graphs with large even-girth*, Combinatorica **18** (1998), 101–120.
23. 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*, Random Structures Algorithms **46** (2015), 1–25.
24. Y. Kohayakawa, T. Łuczak, and V. Rödl, *Arithmetic progressions of length three in subsets of a random set*, Acta Arith. **75** (1996), 133–163.
25. ———, *On* K^4*-free subgraphs of random graphs*, Combinatorica **17** (1997), 173–213.
26. M. N. Kolountzakis, *The density of* $B_h[g]$ *sequences and the minimum of dense cosine sums*, J. Number Theory **56** (1996), 4–11.
27. F. Krückeberg, B_2*-Folgen und verwandte Zahlenfolgen*, J. Reine Angew. Math. **206** (1961), 53–60.
28. S. J. Lee, *On sidon sets in a random set of vectors*, to appear in J. Korean Math. Soc.
29. B. Lindström, *A remark on* B_4*-sequences*, J. Combinatorial Theory **7** (1969), 276–277.
30. R. Morris and D. Saxton, *The number of* $C_{2\ell}$*-free graphs*, Adv. Math. **298** (2016), 534–580.
31. K. O'Bryant, *A complete annotated bibliography of work related to Sidon sequences*, Electron. J. Combin. (2004), Dynamic surveys 11, 39 pp. (electronic).
32. W. Samotij, *Counting independent sets in graphs*, European J. Combin. **48** (2015), 5–18.
33. D. Saxton and A. Thomason, *Hypergraph containers*, Invent. Math. **201** (2015), 925–992.
34. M. Schacht, *Extremal results for random discrete structures*, Ann. of Math. (2) **184** (2016), 333–365.
35. I. E. Shparlinskiĭ, *On* B_s*-sequences*, Combinatorial analysis, No. 7 (Russian), Moskov. Gos. Univ., Moscow, 1986, pp. 42–45, 163.
36. S. Sidon, *Ein Satz über trigonometrische Polynome und seine Anwendung in der Theorie der Fourier-Reihen*, Math. Ann. **106** (1932), 536–539.
37. J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Trans. Amer. Math. Soc. **43** (1938), 377–385.

Appendix A. Omitted proofs

Our proof of Proposition 3 will use the following result of Duke, Lefmann, and Rödl [11], which strengthens the powerful result of Ajtai, Komlós, Pintz, Spencer, and Szemerédi [1]. Recall that a hypergraph is called simple if each pair of its edges intersects in at most one vertex.

Theorem 28. *For every $r \geqslant 3$, there exists a positive constant δ such that the following holds. Every simple r-uniform hypergraph $\mathcal{H}$ with N vertices and average degree at most d^{r-1} contains an independent set of cardinality $\alpha(\mathcal{H})$, where*

$$\alpha(\mathcal{H}) \geqslant \delta \cdot \frac{(\log d)^{1/(r-1)}}{d} \cdot N.$$

Proof of Proposition 3. Suppose that $h \geqslant 2$, fix a positive ε, and let δ be the constant whose existence is guaranteed by Theorem 28 invoked with $r = 2h$. Let

$$C_h = \max\left\{\frac{\log(e/\varepsilon)}{\varepsilon^{2h-1}}, \left(\frac{8}{\delta}\right)^{2h-1}\right\} \qquad \text{and} \qquad c_h = (8h^2 C_h)^{-1/(2h-1)}.$$

Suppose that

$$\varepsilon n^{1/(2h-1)} \leqslant t \leqslant c_h (n \log n)^{1/(2h-1)}$$

and let

$$m = n^{1/(2h-1)} \exp\left(C_h t^{2h-1}/n\right).$$

The claimed lower bound on $|\mathcal{Z}_n^h(t)|$ will follow once we show that a uniformly chosen random m-element subset $R \subset [n]$ contains a t-element B_h-set with probability at least $1/2$. Indeed, we will then have

$$\begin{aligned}|\mathcal{Z}_n^h(t)| &\geqslant \frac{\frac{1}{2}\cdot\binom{n}{m}}{\binom{n-t}{m-t}} = \frac{1}{2}\cdot\binom{m}{t}^{-1}\binom{n}{t} \geqslant \left(\frac{em}{t}\right)^{-t}\binom{n}{t} \\ &= \left(\frac{en^{1/(2h-1)}}{t}\cdot\exp\left(\frac{C_h t^{2h-1}}{n}\right)\right)^{-t}\binom{n}{t} \geqslant \left(\exp\left(-\frac{2C_h t^{2h-1}}{n}\right)\right)^{t}\binom{n}{t};\end{aligned}$$

the last inequality follows as $C_h \geqslant \varepsilon^{-(2h-1)}\log(e/\varepsilon)$. It therefore suffices to prove that the probability that R contains a t-element B_h-set is at least $1/2$, as discussed above.

To this end, observe first that

$$m \leqslant n^{1/(2h-1)} \exp\left(C_h \cdot c_h^{2h-1} \log n\right) = n^{1/(2h-1)+1/(8h^2)} \ll n^{2/(4h-3)}.$$

In particular, letting X be the number of solutions to the equation $a_1 + \ldots + a_h = b_1 + \ldots + b_h$ with $|\{a_1, \ldots, a_h, b_1, \ldots, b_h\}| < 2h$ that are contained in R, we have

$$\mathbb{E}[X] \leqslant K \cdot \sum_{k=3}^{2h-1} n^{k-1}\left(\frac{m}{n}\right)^k \leqslant 2K \cdot \frac{m^{2h-1}}{n} \ll m,$$

where K is a constant depending only on h. Moreover, letting Y be the number of solutions $(a_1, \ldots, a_h, b_1, \ldots, b_h)$ to the above equation with $2h$ distinct coordinates, all belonging to the

random set R, we have

$$\mathbb{E}[Y] \leqslant n^{2h-1} \cdot \left(\frac{m}{n}\right)^{2h} = \frac{m^{2h}}{n}.$$

Finally, letting Z be the number of pairs $(a_1, \ldots, a_h, b_1, \ldots, b_h)$ and $(a'_1, \ldots, a'_h, b'_1, \ldots, b'_h)$ of solutions with $2h$ distinct coordinates that satisfy $|\{a_1, \ldots, a_h, b_1, \ldots, b_h\} \cap \{a'_1, \ldots, a'_h, b'_1, \ldots, b'_h\}| \geqslant 2$ and are both contained in the random set R, we have

$$\mathbb{E}[Z] \leqslant K \cdot \sum_{k=2}^{2h-1} n^{4h-2-k} \left(\frac{m}{n}\right)^{4h-k} \leqslant 2K \cdot \frac{m^{4h-2}}{n^2} \ll m.$$

It follows from Markov's inequality that

$$\mathbb{P}(X > 3\mathbb{E}[Y] \text{ or } X > 12\mathbb{E}[X] \text{ or } Z > 12\mathbb{E}[Z]) < 1/2. \tag{81}$$

Let $\mathcal{G}$ be the $2h$-uniform hypergraph with vertex set $[n]$ whose edges are all $\{a_1, \ldots, a_h, b_1, \ldots, b_h\}$ with $2h$ distinct elements such that $a_1 + \ldots + a_h = b_1 + \ldots + b_h$. It follows from (81) that, with probability at least $1/2$, there is $S \subset [n]$ with $|S| = m/2$ such that the subhypergraph $\mathcal{H} = \mathcal{G}[S]$ induced on S is simple, has at most $3m^{2h}/n$ edges, and such that, moreover, there is no solution to $a_1 + \cdots + a_h = b_1 + \cdots + b_h$ with $\{a_1, \ldots, a_h, b_1, \ldots, b_h\}$ a subset of S with less than $2h$ elements. In particular, as the average degree d of $\mathcal{H}$ satisfies

$$d^{2h-1} = \frac{2h \cdot e(\mathcal{H})}{v(\mathcal{H})} \leqslant \frac{12hm^{2h-1}}{n} \leqslant \left(\frac{4m}{n^{1/(2h-1)}}\right)^{2h-1},$$

whence $d \leq 4m/n^{1/(2h-1)}$. Note that $4m/n^{1/(2h-1)} = \exp(C_h t^{2h-1}/n)$. Therefore, Theorem 28 implies that

$$\alpha(\mathcal{H}) \geqslant \delta \frac{n^{1/(2h-1)}}{4m} \left(\log \frac{4m}{n^{1/(2h-1)}}\right)^{1/(2h-1)} \frac{m}{2} \geqslant \frac{\delta}{8} C_h^{1/(2h-1)} t \geqslant t.$$

Since every independent set of $\mathcal{H}$ is a B_h-set (as R' does not contain any solutions with repeated coordinates), this completes the proof. □

Proof of Lemma 23. Fix some $h \geqslant 2$ and suppose that n and N are integers satisfying $N \geqslant 2hn$. We shall show that there exists a subset $U \subset [N]$ and a projection $\pi \colon U \to [n]$ such that the following holds:

(*a*) If $A \subset [n]$ is a B_h-set, then any set $B \subset \pi^{-1}(A)$ with $|B \cap \pi^{-1}(x)| = 1$ for every $x \in A$ is also a B_h-set.

(*b*) For every $x \in [n]$, we have $|\pi^{-1}(x)| \geqslant N/(2hn)$.

Observe that the existence of such U and π immediately implies the assertion of the lemma. Indeed, for every $A \in \mathcal{Z}_n^h(t)$, we may construct at least $(N/(2hn))^t$ different $B \in \mathcal{Z}_N^h(t)$ by choosing for each $x \in A$ one of at least $N/(2hn)$ elements of $\pi^{-1}(x)$ to be included in B. Moreover, each B constructed in this way satisfies $\pi(B) = A$.

In order to define the projection π and its domain $U \subset [N]$, we first partition $[N]$ into intervals

$$I_j = \left(\frac{j}{n}N, \frac{j+1}{n}N\right] \cap \mathbb{Z}, \quad j = 0, \ldots, n-1.$$

Furthermore, we subdivide each of the intervals above into h subintervals of (almost) equal lengths, namely,

$$I_{j,k} = \left(\left(\frac{j}{n} + \frac{k}{hn} \right) N, \left(\frac{j}{n} + \frac{k+1}{hn} \right) N \right] \cap \mathbb{Z}, \quad j = 0, \ldots, n-1 \text{ and } k = 0, \ldots, h-1.$$

We then define the domain of π by

$$U = \bigcup_{j=0}^{n-1} I_{j,0}.$$

The projection π is then defined by letting $\pi(x) = j + 1$, where j is the unique index such that $x \in I_{j,0}$. Condition (b) is clearly satisfied as for every j,

$$|I_{j,0}| \geqslant \left\lfloor \frac{N}{hn} \right\rfloor \geqslant \frac{N}{2hn},$$

where the last inequality follows from our assumption that $N \geqslant 2hn$.

It remains to prove that condition (a) is also satisfied. Let $A \subset [n]$ be a B_h-set and let $B \subset \pi^{-1}(A)$ be a set satisfying $|B \cap \pi^{-1}(A)| = 1$. This ensures that $\pi|_B$ is a bijection between B and A. Let $(b_1, \ldots, b_h) \in B^h$ be an arbitrary h-tuple with $b_1 \leqslant \ldots \leqslant b_h$ and let ℓ be the unique index such that $b_1 + \ldots + b_h \in I_\ell$. We claim that $\ell + h = \pi(b_1) + \ldots + \pi(b_h)$. Indeed, for each $i \in [h]$, let $j_i = \pi(b_i) - 1$, so that $b_i \in I_{j_i,0}$, and observe that

$$b_1 + \ldots + b_j \in \left(\frac{j_1 + \ldots + j_h}{n} N, \frac{j_1 + \ldots + j_h + 1}{n} N \right] \cap \mathbb{Z} = I_{j_1 + \ldots + j_h}.$$

Since A is a B_h set and π is one-to-one, it follows that no other h-tuple $(b'_1, \ldots, b'_h) \in B^h$ with $b'_1 \leqslant \ldots \leqslant b'_h$ can satisfy $\pi(b'_1) + \ldots + \pi(b'_h) = \ell + h$. In particular, no other h-tuple $(b'_1, \ldots, b'_h)$ with $b'_1 \leqslant \ldots \leqslant b'_h$ satisfies $b'_1 + \ldots + b'_h \in I_\ell$ and hence B must be a B_h-set (recall that ℓ is the unique index such that $b_1 + \ldots + b_h \in I_\ell$). □

Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA (D. Dellamonica Jr. and V. Rödl)

E-mail address: domingos.junior@gmail.com, rodl@mathcs.emory.edu

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, Brazil (Y. Kohayakawa)

E-mail address: yoshi@ime.usp.br

Department of Mathematics, Duksung Women's University, Seoul 01369, South Korea (S. J. Lee)

E-mail address: sanglee242@duksung.ac.kr, sjlee242@gmail.com

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel (W. Samotij)

E-mail address: samotij@post.tau.ac.il