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 [32], who, motivated by the study 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 [12]) 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. Define, for each $h \geqslant 2$ and n, letting $[n] := \{1, \ldots, n\}$,

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

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 [5, 12, 11, 33] 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 [2] 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: 2015/07/02, 22:54.

2010 *Mathematics Subject Classification.* 11B75 (primary), and 05A16, 05D40 (secondary) .

The second author was partially supported by FAPESP (2013/03447-6, 2013/07699-0), CNPq (308509/2007-2 and 477203/2012-4), NSF (DMS 1102086) and NUMEC/USP (Project MaCLinC/USP). 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). 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 [4, 6, 10, 17, 23, 24, 26, 31]. The currently best known bounds are due to Green [13], 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 [14] and to the more recent survey of O'Bryant [28].

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 [3] 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 [22], that is, when $h = 2$, and subsequently [9] for arbitrary h.

Theorem 1 ([22, 9])**.** *For every $h \geqslant 2$, there exists a constant C_h such that $|\mathcal{Z}_n^h| \leqslant 2^{C_h n^{1/h}}$ for all n.*

Let us mention here that another proof of Theorem 1 in the case $h = 2$ was later given by Saxton and Thomason [29], who also showed that, perhaps somewhat surprisingly, $\log_2 |\mathcal{Z}_n^2| \geqslant (1.16 + o(1))F_2(n)$. Also, for Sidon sets in $[n]^d$ for an integer $d \geq 2$, a similar result was given in [25].

In fact, both [22] and [9] 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 [9, 22] were 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 [9] we constructed 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 ([9])**.** *The following holds for every $h \geqslant 2$:*

(*i*) *For every $\delta > 0$, there exists an $\varepsilon > 0$ 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 other words, if $t \ll n^{1/(2h-1)}$, then B_h-sets constitute a sizable $(1-o(1))^t$-proportion of all t-element subsets of $[n]$ and if $t \gg n^{1/(2h-1)}$, then we only know that this proportion is at least (mere) $(c_h'/t^{h-1})^t$ 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 at $t \sim n^{1/(2h-1)}$. A fairly straightforward corollary of the so-called container theorems proved independently by Balogh, Morris, and Samotij [1] and by Saxton and Thomason [29] (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 [9] 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 [9] 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 3 (Main result)**.** *For every $h \geqslant 2$, $\varepsilon > 0$, and all sufficiently large integers n and $t \geqslant n^{1/(2h-1)+\varepsilon}$,*

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

1.1. **Largest B_h-sets contained in random sets of integers.** In the 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 [15, 16, 20, 21] and recently culminated in the breakthrough work of Conlon and Gowers [7] and Schacht [30], 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 and B_h-sets, including this paper, due to the fact that estimating $|\mathcal{Z}_n^h(t)|$ is very closely tied with 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

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

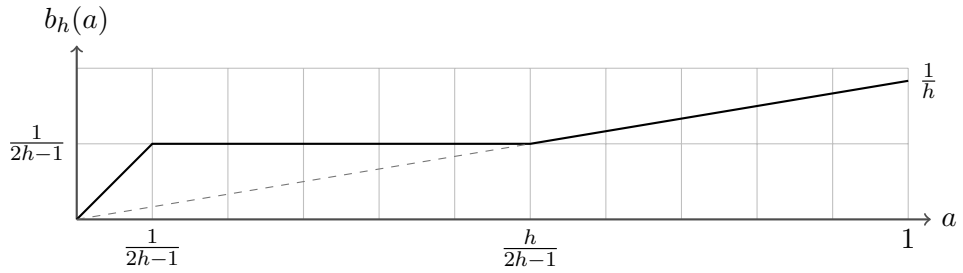

FIGURE 1. The graph corresponding to the piece-wise linear function $b_h(a)$ of Theorem 4.

On the other hand, the transference theorems of Schacht [30] and Conlon and Gowers [7] 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 [9, 22], where much more precise information on $F_h([n]_m)$ was provided. As a consequence of Theorem 3, 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 4. *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 $m = m(n) = n^{a+o(1)}$ for some $a \in [0,1]$ 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

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

Our main result, Theorem 3, shows that for any $t \geqslant n^{1/(2h-1)+o(1)}$, the above expression becomes $o(1)$ when $m < t^{h-o(1)}$. This translates to the upper bound on $F_h([n]_m)$ when $m = n^{a+o(1)}$ for some $h/(2h-1) \leqslant a \leqslant 1$. When $m \leqslant n^{1/(2h-1)+o(1)}$, the claimed upper bound follows from the trivial bound $F_h([n]_m) \leqslant m$. Finally, when $m = n^{a+o(1)}$ for some $1/(2h-1) \leqslant a \leqslant h/(2h-1)$, the claimed upper bound follows from the monotonicity of $b_h(\cdot)$.

The lower bounds on $F_h([n]_m)$ asserted in the theorem were already proved in [9] and therefore we omit their proofs here. □

2. Proof outline

We devote this section to a detailed outline of the proof of our main result, Theorem 3.

2.1. **Background.** Let us start by recalling the general strategy for proving upper bounds on $|\mathcal{Z}_n^h(t)|$ that was used in [9, 22]. The high-level idea there was to bound the number of sets of a given size that one can add to a given B_h-set so that the resulting larger set still has the B_h property. (Having achieved this, one may easily derive a bound on $|\mathcal{Z}_n^h(t)|$ using induction on t.) More precisely, that a B_h-set $S \subset [n]$ of cardinality s is given and we would like to extend it to a larger B_h-set $T \subset [n]$ of cardinality t. The core of [9, 22] is showing that if s is somewhat large, then the number of such extensions is very small. To see why this might be true, observe first that if two distinct elements $x, y \in [n] \setminus S$ satisfy

$$\begin{gathered} x + a_1 + \ldots + a_{h-1} = y + b_1 + \ldots + b_{h-1} \\ \text{for some } \{a_1, \ldots, a_{h-1}\}, \{b_1, \ldots, b_{h-1}\} \in \binom{S}{h-1}, \end{gathered} \tag{4}$$

then $S \cup \{x, y\}$ is clearly not a B_h-set and hence x and y cannot simultaneously belong to T. This motivates our next definition.

Definition 5 (Collision graph CG_S)**.** Let S be a B_h-set. Denote by CG_S 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 CG_S. Therefore, the number of extensions of S to a B_h set of cardinality t is not larger than the number of independent sets in CG_S which have $t - s$ elements. The number of such independent sets can be bounded with the use of the following lemma, implicit in the work of Kleitman and Winston [19], 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 [9].

Lemma 6. *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 6 effectively reduces the problem of counting extensions of S into larger B_h-sets to the problem of verifying that CG_S satisfies condition (6) for appropriately chosen R and β. It is not very difficult to show that for every sufficiently large set $A \subset [n] \setminus S$, there are many *quadruples* $(x, y, \{a_1, \ldots, a_{h-1}\}, \{b_1, \ldots, b_{h-1}\})$ with $x, y \in A$ that satisfy the equality in (4). This, however, does not immediately imply that $e_{\mathrm{CG}_S}(A)$ is large because a single edge of CG_S may correspond to many different quadruples. In [9], we proved an upper bound on the maximum number of such quadruples that a given pair $\{x, y\}$ can be contained in. Unfortunately, this bound becomes

too weak when $|S| \ll n^{1/(h+1)}$. Consequently, we obtained an upper bound on $|\mathcal{Z}_n^h(t)|$ only for $t \leqslant h^2 n^{1/(h+1)} (\log n)^{1+1/(h+1)}$.

2.2. **Special case $h = 3$.** In [8], we enhanced the above strategy and proved Theorem 3 in the case $h = 3$. In the present work, we build on this refined strategy and make further adjustments and generalizations, many of which are highly technical. Therefore we believe that it is instructive to overview the argument used in [8] in that special case first. We shall omit many details in order to simplify the exposition.

Consider all B_3-sets S such that no two distinct $x, y \in [n]$ form 'too many' quadruples satisfying (4), or equivalently, no number $z \neq 0$ admits too many representations as $a_1 + a_2 - b_1 - b_2$, where $a_1, a_2, b_1, b_2 \in S$. (The actual requirement is somewhat more technical.) For convenience, let us call these sets *bounded*.

Those t-element B_h-sets which contain a bounded subset of cardinality $t^{1-\varepsilon}$ may be counted as before, as long as $t \geqslant n^{1/5+o(1)}$. The heart of [8] is estimating the number of $T \in \mathcal{Z}_n^3(t)$ whose largest bounded subset has fewer than $t^{1-\varepsilon}$ elements. The key fact used here is that if S is a maximal bounded subset of T, then $T \setminus S$ is necessarily contained in a set $\widetilde{S}$ that possesses some highly non-random arithmetic properties.

More precisely, given any $\varepsilon > 0$ and $t \geqslant n^{1/5+\varepsilon}$, we consider a family $\mathcal{F} = \mathcal{F}(t) = \mathcal{F}_{\texttt{small}}(t) \cup \mathcal{F}_{\texttt{large}}(t)$ of pairs of sets $(S, \widetilde{S})$ defined as follows:

(1) $\mathcal{F}_{\texttt{large}}(t)$ contains all bounded B_3-sets S with precisely $t^{1-\varepsilon}$ elements, each paired with

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

(2) $\mathcal{F}_{\texttt{small}}(t)$ contains all bounded B_3-sets S with fewer than $t^{1-\varepsilon}$ elements, each paired with

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

Clearly, the family $\mathcal{F}$ has the property that for every $T \in \mathcal{Z}_n^3(t)$, there exists a pair $(S, \widetilde{S}) \in \mathcal{F}$ such that $S \subset T \subset S \cup \widetilde{S}$. This allows us to bound $|\mathcal{Z}_n^3(t)|$ by estimating, for each pair $(S, \widetilde{S}) \in \mathcal{F}$, how many $T \in \mathcal{Z}_n^3(t)$ satisfy $S \subset T \subset S \cup \widetilde{S}$. Since $T \setminus S$ is an independent set in $\mathrm{CG}_S[\widetilde{S}]$, we may derive an upper bound using Lemma 6. In order to obtain a sufficiently strong bound, we have to prove that $\mathrm{CG}_S[\widetilde{S}]$ satisfies the assumptions of Lemma 6 for suitable parameters β and R. This is straightforward when $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{large}}$. Showing this for $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{small}}$ is highly non-trivial and requires a considerable amount of effort.

2.3. **Notation.** Given two sets $A, B \subset \mathbb{Z}$, we let

$$A \oplus B = \sum_{a \in A} a + \sum_{b \in B} b \qquad \text{and} \qquad A \ominus B = \sum_{a \in A} a - \sum_{b \in B} b.$$

For an integer x, we abbreviate $\{x\} \oplus A$ and $\{x\} \ominus A$ with $x \oplus A$ and $x \ominus A$, so that

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

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

$$x \oplus \mathcal{H} = \big\{x \oplus e \colon e \in \mathcal{H}\big\}. \tag{8}$$

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

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

For the sake of clarity of our presentation, we shall often write $\mathrm{polylog}(n)$ to denote any function that is bounded above by $(\log n)^C$ for some absolute constant C. Morevoer, we shall from now on write 'k-graph' instead of 'k-uniform hypergraph'.

More notation will be introduced and used locally when needed.

2.4. **Proof summary.** Recall that given an $S \subset [n]$, let have defined the collision graph CG_S to the graph on the vertex set $[n]$ whose edges are all pairs of distinct elements $x, y \in [n]$ that satisfy (4), that is

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

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 CG_S. We shall show that Theorem 3 follows from Lemma 6 and the following statement, whose proof will take most of this paper.

Theorem 7. *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

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

We postpone the (fairly straightforward) derivation of Theorem 3 to §5.2 and focus on Theorem 7 instead. First, we need several additional definitions which generalize concepts already introduced in [8].

Definition 8 (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

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

Moreover, let

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

For the sake of brevity, we shall often write $R_{\mathcal{G}}$ to denote $R_{\mathcal{G},\mathcal{G}}$.

Definition 9 (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 $x, y \in [n]$ is $R_{\mathcal{G}}(x - y)$.

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. 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.

Similarly as in the case $h = 3$ described in §2.2, the family $\mathcal{F}$ from the statement of Theorem 7 will be partitioned into subsets $\mathcal{F}_{\texttt{large}}$ and $\mathcal{F}_{\texttt{small}}$ in the following manner. Very roughly speaking, we will say that a set S is *bounded* if S is a B_h-set and there exists a companion hypergraph $\mathcal{G} \subset \binom{S}{h-1}$ with the following properties: $\mathcal{G}$ is sufficiently dense and $\|R_{\mathcal{G}}\|$ is somewhat small. With this informal description, we describe the pairs in $\mathcal{F}$.

(1) $\mathcal{F}_{\texttt{large}}$ consists of all pairs formed by a bounded set S of 'large' size ($n^{1/(2h-1)+\delta}$ elements) and

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

It will not be very difficult to show that $\mathrm{CG}_S[\widetilde{S}]$ satisfies the assumptions of Lemma 6.

(2) $\mathcal{F}_{\texttt{small}}$ consists of all pairs formed by a bounded set S of 'small' size (fewer than $n^{1/(2h-1)+\delta}$ elements) and

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

In other words, $\widetilde{S}$ contains all the elements $x \in [n] \setminus S$ such that $S \cup \{x\}$ is a B_h-set but for which one cannot find a companion hypergraph $\overline{\mathcal{G}} \subset \binom{S \cup \{x\}}{h-1}$ in such a way that the density of $\overline{\mathcal{G}}$ is large and $\|R_{\overline{\mathcal{G}}}\|$ is appropriately bounded. The precise definition of bounded sets will force an atypical additive structure on the pair $(S, \widetilde{S})$, which eventually allow us to prove that $\mathrm{CG}_S[\widetilde{S}]$ satisfies the assumptions of Lemma 6.

It is easy to see that for every $T \in \mathcal{Z}_n^h(t)$, there is an $(S, \widetilde{S}) \in \mathcal{F}$ such that $S \subset T \subset S \cup \widetilde{S}$. Indeed, given such a set T, we let S' be the largest bounded subset of T. If $|S'| \geqslant n^{1/(2h-1)+\delta}$, then we let S be an arbitrary subset of S' with precisely $n^{1/(2h-1)+\delta}$ elements and obtain a pair $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{large}}$. Otherwise, we let $S = S'$ and obtain a pair $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{small}}$. The real content of Theorem 7 is the fact that (9) holds for every $(S, \widetilde{S}) \in \mathcal{F}$. Since the arguments involving $\mathcal{F}_{\texttt{large}}$ are somewhat standard, we shall focus the remainder of this section on the precise notion of boundedness (which determines the definitions of $\mathcal{F}_{\texttt{large}}$, $\mathcal{F}_{\texttt{small}}$, and $\widetilde{S}$) and the structure of the pairs $(S, \widetilde{S}) \in \mathcal{F}_{\texttt{small}}$.

First of all, since there seems to be no easy way of controlling $\|R_{\mathcal{G}}\|$ 'directly', similarly as in [8], we shall instead maintain an upper bound on the *moment generating function* of $R_{\mathcal{G}}$, defined as follows.

Definition 10 (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 11. 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 this end, we shall construct the set S and the hypergraph $\mathcal{G}$ with $V(\mathcal{G}) = S$ step-by-step, adding to $\mathcal{G}$ one vertex at a time. There is an apparent difficulty to be overcome in this approach. On the one hand, in order to guarantee that the multigraph $\widetilde{\mathrm{CG}}_{\mathcal{G}}$ has many edges, we should make sure that $\mathcal{G}$ is relatively dense. On the other hand, the more edges we add to $\mathcal{G}$, the more difficult it is to guarantee that $\|R_{\mathcal{G}}\|$ stays small.

Suppose that a B_h-set S is bounded, that is, there is an $(h-1)$-graph $\mathcal{G}$ with vertex set S that is relatively dense and such that $Q_{\mathcal{G},\mathcal{G}}(\lambda)$ is somewhat small and, consequently, $\|R_{\mathcal{G}}\|$ is also small. Suppose moreover that we are trying to add a new vertex $x \notin S$ to the $(h-1)$-graph $\mathcal{G}$ in order to form a new $(h-1)$-graph $\overline{\mathcal{G}}$ with vertex set $S \cup \{x\}$. Clearly, it suffices to decide which $(h-2)$-tuples of elements of S will form an edge together with the new vertex x. Denote by $\mathcal{N}_x$ the hypergraph formed by all such $(h-2)$-tuples. One easily sees that

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

In view of this, it seems useful to control $\|R_{\mathcal{G},\mathcal{N}_x}\| = \|R_{\mathcal{N}_x,\mathcal{G}}\|$. One can achieve this by requiring that $\mathcal{N}_x \subset \mathcal{G}^{(h-2)}$ for some $(h-2)$-graph $\mathcal{G}^{(h-2)}$ such that $Q_{\mathcal{G},\mathcal{G}^{(h-2)}}(\lambda)$ is somewhat small and, consequently, $\|R_{\mathcal{G},\mathcal{N}_x}\| \leqslant \|R_{\mathcal{G},\mathcal{G}^{(h-2)}}\|$ is also small. Continuing in this fashion, one realizes that it is useful to construct an entire family of hypergraphs, one k-graph $\mathcal{G}^{(k)} \subset \binom{S}{k}$ for each $k \in [h-1]$, such that each $\mathcal{G}^{(k)}$ is relatively dense and at the same time $Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda)$ is somewhat small for all pairs $k, \ell \in [h-1]$. 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 12. Let $h, n \geqslant 2$ 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 $k \in [h-1]$, such that

(*a*) $|\mathcal{G}^{(k)}| \geqslant (1 - 2^k\alpha)\binom{|S|}{k}$ for each $k \in [h-1]$;

(*b*) $Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \leqslant (2hn+1) \cdot \exp(\mathsf{H}_{|S|})$ for all $k, \ell \in [h-1]$, where

$$\xi_j = (2\log n)^{-j} \quad \text{for all integers } j. \tag{14}$$

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

$$\widetilde{S}_{\lambda,\alpha} = \big\{x \in [n] \setminus S \colon S \cup \{x\} \notin \mathcal{P}_h(\lambda, \alpha)\big\}. \tag{15}$$

Remark 13. Note that if $\mathcal{G}^{(1)},\dots,\mathcal{G}^{(h-1)}$ satisfy condition (b) of the above definition, then

$$\|R_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}\| \leqslant \frac{\log Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda\cdot\xi_{k+\ell})}{\lambda\cdot\xi_{k+\ell}} \leqslant \frac{2\log n}{\lambda\cdot\xi_{k+\ell}} = \frac{1}{\lambda\cdot\xi_{k+\ell+1}}.$$

Let us tentatively say that a set $S\in\mathcal{Z}_n^h$ is bounded if it satisfies property $\mathcal{P}_h(\lambda,\alpha)$ for some given parameters $\alpha\leqslant 2^{-h}$ and $\lambda\geqslant n^{-\delta}$. (In reality, the definition is somewhat more complicated, but this approximation will suffice for now.) Assume this definition of boundedness and suppose that $(S,\widetilde{S})\in\mathcal{F}_{\texttt{small}}$. In particular, $S\in\mathcal{P}_h(\lambda,\alpha)$ and $\widetilde{S}=\widetilde{S}_{\lambda,\alpha}$. Roughly speaking, it means that there are realtively dense hypergraphs $\mathcal{G}^{(1)},\dots,\mathcal{G}^{(h-1)}$ such that $Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda)$ is somewhat small for all pairs $k,\ell\in[h-1]$ but for each $x\in\widetilde{S}$, it is not possible to choose, for every $k\in\{2,\dots,h-1\}$, sufficiently many edges of $\mathcal{G}^{(k-1)}$ to form the neighborhood of x in a k-graph $\overline{\mathcal{G}}^{(k)}$ with vertex set $S\cup\{x\}$ in such a way that $Q_{\overline{\mathcal{G}}^{(k)},\mathcal{G}^{(\ell)}}(\lambda)$ is not much larger than $Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda)$ for all k and ℓ.

In the first key step of the proof, we shall show that there exist sets $\Gamma_2,\dots,\Gamma_{h-1}\subset[n]$ which are fairly small but for each $x\in\widetilde{S}$, there is some $k\in\{2,\dots,h-1\}$ such that the set $x\oplus\binom{S}{k-1}$, defined as in (8), has an unusually large intersection with Γ_k. More precisely, $|\Gamma_k|\leqslant\lambda\cdot(|S|+1)^{k+h-1}$ and

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

This constitutes what we called earlier 'an atypical additive structure' on the pair $(S,\widetilde{S})$. In the second key step of the proof, we shall exploit the existence of the sets $\Gamma_2,\dots,\Gamma_{h-1}$ to derive a strong lower bound on $\widetilde{\mathrm{CG}}_{\mathcal{H}}(A)$ for all sufficiently large $A\subset\widetilde{S}$ and some $\mathcal{H}\subset\binom{S}{h-1}$. The caveat here is that we cannot take $\mathcal{H}=\mathcal{G}^{(h-1)}$ as our argument requires that $e(\mathcal{H})\geqslant(1-\alpha')\binom{|S|}{k}$ for some $\alpha'\ll\alpha$. But this only means that $e_{\mathrm{CG}_S}(A)\geqslant e_{\widetilde{\mathrm{CG}}_{\mathcal{H}}}(A)/\|R_{\mathcal{H}}\|$ and we have no strong a priori bound on $\|R_{\mathcal{H}}\|$. This is why in the real definition of bounded sets (Definition 18), we shall require that S admits a whole family of collections $\mathcal{G}^{(1)},\dots,\mathcal{G}^{(h-1)}$, one for each pair (λ,α) coming from a sequence $(\lambda_i,\alpha_i)_i$, where $\alpha_{i+1}\ll\alpha_i$ and λ_{i+1}/λ_i is not very small for each i. This way, we may let the hypergraph $\mathcal{H}$ above be the $(h-1)$-graph $\mathcal{G}^{(h-1)}$ constructed for the next pair (λ,α) in our sequence. This $\mathcal{H}$ is sufficiently dense, since $\alpha_{i+1}\ll\alpha_i$. But now $\|R_{\mathcal{H}}\|$ is not much larger than the 'original' bound on $\|R_{\mathcal{G}^{(h-1)}}\|$, since λ_{i+1} is not much smaller than λ_i.

2.5. **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 hypergraph affects the moment function $Q_{\mathcal{G},\mathcal{H}}(\cdot)$. We then show in Section 4 that each pair of sets $(S,\widetilde{S}_{\lambda,\alpha})$ defined in (15) possesses a certain additive structure. In Section 5 we state a technical result, Theorem 17, which asserts that for every sufficiently dense $\mathcal{H}\subset\binom{S}{h-1}$, 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 7. A fairly straightforward derivation of our main results, Theorem 3, from Theorem 7 is presented in §5.2. The fairly long and technical proof of Theorem 17 is postponed to Section 6. In Section 7, we conlude the paper with a short discussion.

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 when we extend two hypergraphs $\mathcal{G}$ and $\mathcal{H}$ to form $\overline{\mathcal{G}}$ and $\overline{\mathcal{H}}$ by adding to them a single vertex, then we may bound the increase of the moment function, $Q_{\overline{\mathcal{G}},\overline{\mathcal{H}}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda)$, in terms of the increases of the moment function caused by extending $\mathcal{G}$ and $\mathcal{H}$ separately, that is, $Q_{\overline{\mathcal{G}},\mathcal{H}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda)$ and $Q_{\mathcal{G},\overline{\mathcal{H}}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda)$, 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 14. *Let $k, \ell \geqslant 2$ be integers and let $\lambda > 0$. Suppose that*

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

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

satisfy

$$Q_{\overline{\mathcal{G}},\overline{\mathcal{H}}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda) \leqslant 2\left(\big(Q_{\overline{\mathcal{G}},\mathcal{H}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda)\big) + \big(Q_{\mathcal{G},\overline{\mathcal{H}}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda)\big)\right). \tag{16}$$

Proof. As $R_{\overline{\mathcal{G}},\overline{\mathcal{H}}}(z)$ counts only pairs $(e,f) \in \overline{\mathcal{G}} \times \overline{\mathcal{H}}$ that satisfy $e \cap f = \emptyset$, we have

$$R_{\overline{\mathcal{G}},\overline{\mathcal{H}}}(z) = R_{\mathcal{G},\mathcal{H}}(z) + R_{\mathcal{N},\mathcal{H}}(z-x) + R_{\mathcal{G},\mathcal{M}}(z+x) \tag{17}$$

for every integer z. Now, let

$$\begin{aligned} N(z) &= R_{\mathcal{N},\mathcal{H}}(z-x) = R_{\overline{\mathcal{G}},\mathcal{H}}(z) - R_{\mathcal{G},\mathcal{H}}(z), \\ M(z) &= R_{\mathcal{G},\mathcal{M}}(z+x) = R_{\mathcal{G},\overline{\mathcal{H}}}(z) - R_{\mathcal{G},\mathcal{H}}(z) \end{aligned} \tag{18}$$

and observe that

$$Q_{\overline{\mathcal{G}},\mathcal{H}}(\lambda) - Q_{\mathcal{G},\mathcal{H}}(\lambda) = \sum_{z=-\ell n}^{kn} \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \Big(\underbrace{\exp\big(\lambda \cdot N(z)\big) - 1}_{a}\Big), \tag{19}$$

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

Since (17) holds, we have

$$\exp\big(\lambda \cdot R_{\overline{\mathcal{G}},\overline{\mathcal{H}}}(z)\big) = \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \exp\big(\lambda \cdot N(z)\big) \cdot \exp\big(\lambda \cdot M(z)\big), \tag{21}$$

and hence,

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

Therefore, in order to establish (16), 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 (18) 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, $a+b\leqslant 4$. Consequently, $ab\leqslant\big(\frac{a+b}{2}\big)^2\leqslant a+b$ by the AM–GM inequality. □

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 15. *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 $\overline{\mathcal{G}}$ on $V(\mathcal{G})\cup\{x\}$ defined by

$$\overline{\mathcal{G}}=\mathcal{G}\cup\big\{\{x\}\cup e\colon e\in\mathcal{N}\text{ and }x\oplus e\notin\Gamma\big\}\tag{22}$$

satisfies

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

Proof. For integers w and z, define

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

and

$$u_w=\sum_{z=-\ell n}^{kn}\exp\big(\lambda\cdot R_{\mathcal{G},\mathcal{H}}(z)\big)I(w,z),\tag{23}$$

see Figure 2. Set

$$\Gamma=\left\{w\in[kn]\colon u_w\geqslant\frac{|\mathcal{H}|\cdot Q_{\mathcal{G},\mathcal{H}}(\lambda)}{M}\right\}.\tag{24}$$

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

Proof. Observe first that for every z,

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

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

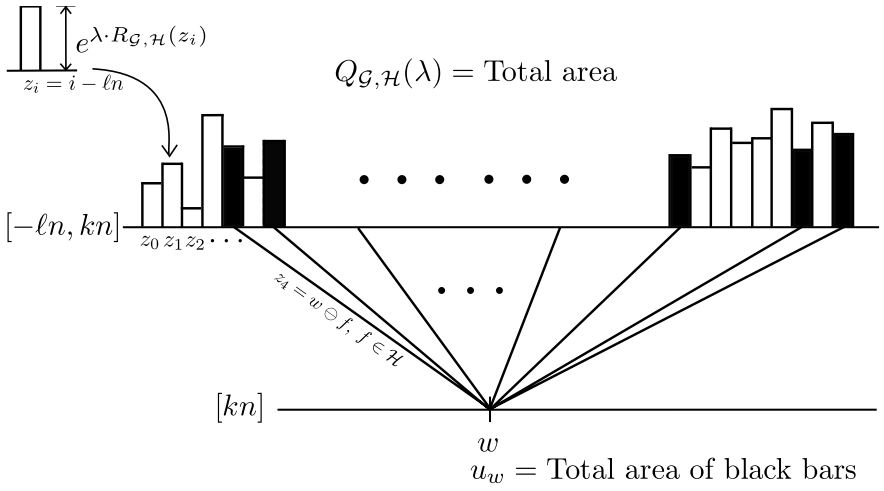

FIGURE 2. The definition of u_w.

Indeed, each value w such that $I(w,z) = 1$ has some associated $f_w \in \mathcal{H}$ satisfying $w \ominus f_w = z$, 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

$$\mathcal{N}_x = \{e \in \mathcal{N} \colon x \oplus e \notin \Gamma\} \tag{25}$$

and consider the k-graph $\overline{\mathcal{G}}$ defined in (22), namely

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

Observe that for any integer z,

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

It follows that

$$\begin{aligned} \exp\big(\lambda \cdot R_{\overline{\mathcal{G}},\mathcal{H}}(z)\big) &\leqslant \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \exp\big(\lambda \cdot R_{\mathcal{N}_x,\mathcal{H}}(z-x)\big) \\ &\leqslant \exp\big(\lambda \cdot R_{\mathcal{G},\mathcal{H}}(z)\big) \cdot \big(1 + 2\lambda \cdot R_{\mathcal{N}_x,\mathcal{H}}(z-x)\big)\,, \end{aligned} \tag{26}$$

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,

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

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 $z=(x\oplus e)\ominus f$. Consequently, from (26), we have

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

Summing the above inequality over all $z\in[-\ell n, kn]$, and recalling (23) yields

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

From the definitions of Γ and $\mathcal{N}_x$, see (24), (25), we finally conclude that

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

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

In this section we show that if S satisfies $\mathcal{P}_h(\lambda,\alpha)$, 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 a large fraction of all numbers of the form $x\oplus e$ with $x\in\widetilde{S}_{\lambda,\alpha,k}$ and $e\in\binom{S}{k-1}$ belong to a fairly small set Γ_k. 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 $\mathcal{H}\subset\binom{S}{h-1}$.

Lemma 16. *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 a $k\in\{2,\dots,h-1\}$ such that*

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

Proof. Our argument here 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 12.Given an $x\in\widetilde{S}_{\lambda,\alpha}$, using Lemmas 14 and 15 from Section 3, we shall extend each $\mathcal{G}^{(k)}$ to a $\overline{\mathcal{G}}^{(k)}\subset\binom{S\cup\{x\}}{k}$ so that condition (b) of Definition 12 is satisfied. By the definition of $\widetilde{S}_{\lambda,\alpha}$, some $\overline{\mathcal{G}}^{(k)}$ must fail condition (a). We shall derive the conclusion of the lemma from this fact.

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 12. In particular, it follows from Remark 13 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 15 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

$$|\Gamma_{k,\ell}| \leqslant 8(|S|+1)^{k+\ell} \cdot \lambda \cdot \xi_{k+\ell} \tag{27}$$

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

$$\overline{\mathcal{G}}_\ell^{(k)}(x) = \mathcal{G}^{(k)} \cup \big\{\{x\} \cup e \colon e \in \mathcal{G}^{(k-1)} \text{ and } x \oplus e \notin \Gamma_{k,\ell}\big\} \tag{28}$$

satisfies

$$\begin{aligned} Q_{\overline{\mathcal{G}}_\ell^{(k)}(x),\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) &\leqslant Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \cdot \left(1 + \frac{|\mathcal{G}^{(\ell)}| \cdot |\mathcal{G}^{(k-1)}|}{4(|S|+1)^{k+\ell}}\right) \\ &\leqslant Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \cdot \left(1 + \frac{1}{4(|S|+1)}\right). \end{aligned} \tag{29}$$

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

$$\Gamma_k = \bigcup_{\ell=1}^{h-1} \Gamma_{k,\ell} \tag{30}$$

and observe that (27) implies that condition (i) from the statement of this lemma is satisfied, see the definition of ξ_j in (14).

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

$$\overline{\mathcal{G}}^{(k)} = \bigcap_{\ell=1}^{h-1} \overline{\mathcal{G}}_\ell^{(k)}(x) \overset{(28)}{=} \mathcal{G}^{(k)} \cup \big\{\{x\} \cup e \colon e \in \mathcal{G}^{(k-1)} \text{ and } x \oplus e \notin \Gamma_k\big\}. \tag{31}$$

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

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

as we have assumed that $\lambda \leqslant 1$. It follows from (29), and the fact that $\overline{\mathcal{G}}^{(k)} \subset \overline{\mathcal{G}}_\ell^{(k)}(x)$, that for every $k, \ell \in \{2, \ldots, h-1\}$,

$$Q_{\overline{\mathcal{G}}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \leqslant Q_{\overline{\mathcal{G}}_\ell^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \leqslant Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \cdot \left(1 + \frac{1}{4(|S|+1)}\right). \tag{32}$$

Since $Q_{\mathcal{G},\mathcal{H}}(\cdot) = Q_{\mathcal{H},\mathcal{G}}(\cdot)$, the same bound above applies to $Q_{\mathcal{G}^{(k)},\overline{\mathcal{G}}^{(\ell)}}(\lambda \cdot \xi_{k+\ell})$. Consequently, Lemma 14 implies that for all $k, \ell \in \{2, \ldots, h-1\}$,

$$\begin{aligned} Q_{\overline{\mathcal{G}}^{(k)},\overline{\mathcal{G}}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) &\leqslant Q_{\mathcal{G}^{(k)},\mathcal{G}^{(\ell)}}(\lambda \cdot \xi_{k+\ell}) \cdot \left(1 + \frac{1}{|S|+1}\right) \\ &\quad \text{—by condition } (b) \text{ of Def. 12—} \\ &\leqslant (2hn+1) \cdot \exp(\mathsf{H}_{|S|}) \cdot \left(1 + \frac{1}{|S|+1}\right) \\ &\leqslant (2hn+1) \cdot \exp(\mathsf{H}_{|S|+1}). \end{aligned} \tag{33}$$

In other words, the hypergraphs $\overline{\mathcal{G}}^{(1)}, \dots, \overline{\mathcal{G}}^{(h-1)}$ satisfy condition (b) of Definition 12 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 12 has to be violated, that is, there must be some $k \in [h-1]$ for which $|\overline{\mathcal{G}}^{(k)}| < (1-2^k\alpha)\binom{|S|+1}{k}$. Together with the fact that $|\mathcal{G}^{(k)}| \geqslant (1-2^k\alpha)\binom{|S|}{k}$, we have

$$|\overline{\mathcal{G}}^{(k)}| - |\mathcal{G}^{(k)}| < (1-2^k\alpha)\binom{|S|+1}{k} - (1-2^k\alpha)\binom{|S|}{k} = (1-2^k\alpha)\binom{|S|}{k-1}. \tag{34}$$

This is clearly not true if $k = 1$, as $|\overline{\mathcal{G}}^{(1)}| = |\mathcal{G}^{(1)}| + 1$, hence let us consider $k \in \{2, \dots, h-1\}$. By the definition of $\overline{\mathcal{G}}^{(k)}$ in (31),

$$|\overline{\mathcal{G}}^{(k)}| - |\mathcal{G}^{(k)}| = |\mathcal{G}^{(k-1)}| - |\{e \in \mathcal{G}^{(k-1)} \colon x \oplus e \in \Gamma_k\}| = |\mathcal{G}^{(k-1)}| - \left|\left(x \oplus \mathcal{G}^{(k-1)}\right) \cap \Gamma_k\right|,$$

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 $x \oplus e = x \oplus e'$. Since $\mathcal{G}^{(k-1)}$ satisfies condition (a) of Definition 12, we have

$$|\overline{\mathcal{G}}^{(k)}| - |\mathcal{G}^{(k)}| \geqslant (1-2^{k-1}\alpha)\binom{|S|}{k-1} - \left|\left(x \oplus \mathcal{G}^{(k-1)}\right) \cap \Gamma_k\right|. \tag{35}$$

Combining (34) and (35) yields

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

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 3, from the main result of the previous two sections, Lemma 16, and the following technical statement, Theorem 17 below, that 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 17 are suited for invoking the theorem with $A \subset \widetilde{S}_{\lambda,\alpha,\ell}$ and $\Gamma = \Gamma_\ell$ from Lemma 16. We postpone the proof of Theorem 17 to Section 6.

Theorem 17. *Let $h \geqslant 2$, $\ell \in [h-1]$, $\beta \in (0,1)$, n be a sufficiently large integer, and $d \geqslant (128h\log_2 n)^{\ell+2}$. Fix some $S \in \mathcal{Z}_n^h$, $A \subset [n]$, and $\mathcal{H} \subset \binom{S}{h-1}$, with $\beta\,|S| > n^{1/(100h^2)}$, and*

$$|\mathcal{H}| \geqslant \left(1 - \frac{\beta^h}{(\log_2 n)^{7h^2}}\right)\binom{|S|}{h-1}. \tag{36}$$

Suppose that there exists a set $\Gamma \subset [hn]$ such that

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

Then,

$$e_{\widetilde{\mathrm{CG}}_{\mathcal{H}}}(A) \geqslant \frac{\beta^h \cdot d}{(\log_2 n)^{7h^2}}|A|\binom{|S|}{h-1}. \tag{37}$$

We are now ready to prove Theorem 7, which implies an upper bound on $|\mathcal{Z}_n^h(t)|$ for t in a narrow range interval 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 7, let us formally define the notion of bounded B_h-sets. As we tried to explain at the end of §2.4, it is insufficient to consider property $\mathcal{P}_h(\lambda, \alpha)$ for merely one pair (λ, α). This is because in order to successfully apply Theorem 17 with $A \subset \widetilde{S}_{\lambda,\alpha}$ and $\Gamma = \Gamma_\ell$ from Lemma 16, we need an $\mathcal{H} \subset \binom{S}{h-1}$ satisfying (36) with $\beta \approx \alpha$, whereas the only suitable candidate for $\mathcal{H}$, the $(h-1)$-graph $\mathcal{G}^{(h-1)}$ from the definition of $\mathcal{P}_h(\lambda, \alpha)$, is not sufficiently dense. That is why for us a bounded set will be one that satisfies $\mathcal{P}_h(\lambda, \alpha)$ not for one but for an entire sequence of pairs (λ_i, α_i) with α_i decreasing sufficiently fast so that (36) will be satisfied with $\mathcal{H}$ being the 'next' $\mathcal{G}^{(h-1)}$ and, at the same time, λ_i decreasing sufficiently slowly so that $\|R_{\mathcal{G}^{(h-1)}}\|$ does not increase too much while we move to the 'next' $\mathcal{G}^{(h-1)}$.

Definition 18. 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

$$\lambda_i = n^{-i\rho}, \quad \alpha_0 = \frac{1}{2^h (\log_2 n)^{7h^2}}, \quad \text{and} \quad \alpha_{i+1} = \frac{(\alpha_i/2)^h}{(\log_2 n)^{7h^2}} \quad \text{for } i = 0, 1, \ldots, \lceil 1/\rho \rceil - 1. \tag{38}$$

5.1. **Proof of Theorem 7.** Let h, δ, n, and t be given as in the statement of Theorem 7. 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

$$\rho = \frac{\delta}{4}\left(\frac{1}{2h-1} + \delta\right). \tag{39}$$

Note that $\alpha_i = \mathrm{polylog}(n)^{-1}$ for every $i \in \{0, 1, \ldots, \lceil 1/\rho \rceil\}$ 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:

- (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 12):
  - $|\mathcal{G}| \geqslant (1 - 2^{h-1}\alpha_0)\binom{|S|}{h-1}$.
  - $Q_{\mathcal{G},\mathcal{G}}(\lambda_0 \xi_{2h-2}/2) \leqslant (2hn+1)\exp\big(\mathsf{H}_{|S|}\big)$.
- (IV) The set $\widetilde{S}$ is defined as

$$\widetilde{S} = \big\{x \in [n] \setminus S \colon S \cup \{x\} \text{ is a } B_h\text{-set}\big\}. \tag{40}$$

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

$$\begin{gathered} S \text{ satisfies } \mathcal{P}_h(\rho),\ n^{1/(8h^2)} \leqslant |S| < t^{1-\delta}, \text{ and} \\ \widetilde{S} = \big\{x \in [n] \setminus S \colon S \cup \{x\} \text{ is a } B_h\text{-set which does } \textit{not} \text{ satisfy } \mathcal{P}_h(\rho)\big\}. \end{gathered} \tag{41}$$

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, $|X| \geqslant t^{1/(4h-5)} \geqslant n^{1/(8h^2)}$. Let $\mathcal{K} = \binom{X}{h-1}$. Since X is a B_{2h-2}-set, it follows that $\|R_{\mathcal{K}}\| = 1$. Therefore, for each $\lambda \in (0,1]$,

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

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 (41) 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 $T \setminus S \subset \widetilde{S}$. Hence we may assume 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 12 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

$$\begin{aligned}
Q_{\mathcal{G},\mathcal{G}}(\lambda_0\xi_{2h-2}/2) &= \sum_{z=-(h-1)n}^{(h-1)n} \exp\big(\lambda_0\xi_{2h-2} \cdot R_{\mathcal{G}}(z)\big)^{1/2} \\
&\leqslant \left(\big((2h-2)n+1\big) \sum_{z=-(h-1)n}^{(h-1)n} \exp\big(\lambda_0\xi_{2h-2} \cdot R_{\mathcal{G}}(z)\big)\right)^{1/2} \\
&= \Big(\big((2h-2)n+1\big) Q_{\mathcal{G},\mathcal{G}}(\lambda_0\xi_{2h-2})\Big)^{1/2} \\
&\leqslant (2hn+1)\exp\big(\mathsf{H}_{|S'|}\big).
\end{aligned}$$

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

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

Consequently, S' and its corresponding $\widetilde{S}'$, defined as in (40), 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. $\square$

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 $(S,\widetilde{S}) \in \mathcal{F}_{\mathtt{large}}(t)$ or $(S,\widetilde{S}) \in \mathcal{F}_{\mathtt{small}}(t)$.

5.1.1. *Case when* $(S,\widetilde{S}) \in \mathcal{F}_{\mathtt{large}}(t)$. Our definition of $\mathcal{F}_{\mathtt{large}}(t)$, 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 Theorem 17 with

$$\ell = h-1, \quad \beta = 1, \quad \mathcal{H} = \mathcal{G}, \quad \Gamma = [hn], \quad d = |A|\binom{|S|}{h-1}/(hn). \tag{42}$$

Indeed, the conditions of the theorem are satisfied as:

- For every $a \in A$, we have $a \oplus \binom{S}{h-1} \subset [hn] = \Gamma$.
- $|A|\binom{|S|}{h-1} \geqslant \frac{n}{t^{h-1-8h^2\delta}}\left(\frac{t^{1-\delta}}{h-1}\right)^{h-1} \geqslant nt^{\delta}$ and thus $d \geqslant t^{\delta} \gg (128h\log_2 n)^{h+1}$.
- We have $|S| = t^{1-\delta} > n^{1/(4h)}$ and thus $\beta|S| > n^{1/(100h^2)}$.
- Since $|\mathcal{G}| \geqslant (1-2^{h-1}\alpha_0)\binom{|S|}{h-1}$, it follows from (38) that $\mathcal{H} = \mathcal{G}$ satisfies (36), as $2^{h-1}\alpha_0 = \frac{1}{2(\log_2 n)^{7h^2}} < \frac{\beta^h}{(\log_2 n)^{7h^2}}$.

Hence,

$$e_{\widetilde{\mathrm{CG}}_{\mathcal{G}}}(A) \geqslant \mathrm{polylog}(n)^{-1} \cdot d|A|\binom{|S|}{h-1} \overset{(42)}{\geqslant} \frac{1}{hn \cdot \mathrm{polylog}(n)}|A|^2\binom{|S|}{h-1}^2.$$

On the other hand, from (III) and Remark 11 we conclude that

$$\|R_{\mathcal{G}}\| = \frac{2}{\lambda_0\xi_{2h-2}}\log Q_{\mathcal{G},\mathcal{G}}(\lambda_0\xi_{2h-2}/2) \leqslant \frac{2\log\Big((2hn+1)\exp\big(\mathsf{H}_{|S|}\big)\Big)}{\lambda_0\xi_{2h-2}} \overset{(14)}{=} \mathrm{polylog}(n).$$

Therefore,

$$e_{\mathrm{CG}_S}(A) \geqslant \frac{e_{\widetilde{\mathrm{CG}}_{\mathcal{G}}}(A)}{\|R_{\mathcal{G}}\|} \geqslant \frac{|S|^{2h-2}}{n \cdot \mathrm{polylog}(n)}|A|^2. \tag{43}$$

We claim that (43) 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 7 in the case $(S,\widetilde{S}) \in \mathcal{F}_{\mathtt{large}}(t)$.

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 $\widetilde{S}$ into

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

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 12.

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

Proof. 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 $k, \ell \in [h-1]$,

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

Consequently,

$$Q_{\mathcal{K}^{(k)},\mathcal{K}^{(\ell)}}(\lambda_{\lceil 1/\rho \rceil} \xi_{k+\ell}) < (2hn+1)e.$$

It follows that the family of hypergraphs $\mathcal{K}^{(k)}$, $k \in [h-1]$, satisfies the conditions of Definition 12 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 16 with $\lambda = \lambda_i$ and $\alpha = \alpha_i$ to obtain sets $\Gamma_{i,2}, \dots, \Gamma_{i,h-1} \subset [hn]$ satisfying:

- $|\Gamma_{i,k}| \leqslant (|S|+1)^{h+k-1} \cdot \lambda_i$ 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

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

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 $A \subset \widetilde{S}$ 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

$$|A_{i,k}| \geqslant \frac{|A|}{(h-2)\lceil 1/\rho \rceil} = \Omega\bigg(\frac{n}{t^{h-1-8h^2\delta}}\bigg). \tag{44}$$

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.) Recall from Definition 12 that:

- $|\mathcal{H}| \geqslant (1 - 2^{h-1}\alpha_{i+1}) \binom{|S|}{h-1}$;

- $Q_{\mathcal{H},\mathcal{H}}\big(\lambda_{i+1}\xi_{2h-2}\big) \leqslant (2hn+1)\exp(\mathsf{H}_{|S|})$, which by Remark 11 means that

$$\|R_{\mathcal{H}}\| \leqslant \frac{\log\Big((2hn+1)\exp\big(\mathsf{H}_{|S|}\big)\Big)}{\lambda_{i+1}\xi_{2h-2}} \overset{(14)}{=} \frac{\mathrm{polylog}(n)}{\lambda_{i+1}}. \tag{45}$$

We shall now apply Theorem 17 with

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

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 $\big|a \oplus \binom{S}{k-1} \cap \Gamma_{i,k}\big| \geqslant 2^{k-1}\alpha_i\binom{|S|}{k-1} \geqslant \beta\binom{|S|}{k-1}$.
- We also have

$$\sum_{a\in A_{i,k}}\left|a \oplus \binom{S}{k-1} \cap \Gamma_{i,k}\right| \geqslant 2^{k-1}\alpha_i\,|A_{i,k}|\binom{|S|}{k-1} = d\,|\Gamma_{i,k}|.$$

- Since $|\Gamma_{i,k}| \leqslant |S|^{h+k-1}\cdot\lambda_i$, we see that d satisfies

$$d = 2^{k-1}\alpha_i\,|A_{i,k}|\binom{|S|}{k-1}/|\Gamma_{i,k}| \overset{(44)}{\geqslant} \mathrm{polylog}(n)^{-1}\cdot\frac{n|S|^{k-1}}{t^{h-1-8h^2\delta}\,|S|^{h+k-1}\lambda_i}$$

  Since $|S| < t \leqslant 2hn^{1/(2h-1)+\delta}$ and $\lambda_i \leqslant 1$, we have

$$d \geqslant \mathrm{polylog}(n)^{-1}\cdot\frac{n}{t^{2h-1-8h^2\delta}} \geqslant \mathrm{polylog}(n)^{-1}\cdot n^{4h^2\delta/(2h-1)} \gg (128h\log_2 n)^{k+1}.$$

- The set S, by the definition of $\mathcal{F}_{\mathtt{small}}(t)$, 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 (38) that $\mathcal{H}$ satisfies (36). Indeed,

$$2^{h-1}\alpha_{i+1} = \frac{2^h(\alpha_i/2)^h}{2(\log_2 n)^{7h^2}} = \frac{\beta^h}{2(\log_2 n)^{7h^2}}. \tag{46}$$

Hence by Theorem 17,

$$e_{\widetilde{\mathrm{CG}}_{\mathcal{H}}}(A_{i,k}) \geqslant \mathrm{polylog}(n)^{-1}\cdot d\,|A_{i,k}|\binom{|S|}{h-1}.$$

Recalling (45), we conclude that

$$\begin{aligned} e_{\mathrm{CG}_S}(A) \geqslant e_{\mathrm{CG}_S}(A_{i,k}) &\geqslant \frac{e_{\widetilde{\mathrm{CG}}_{\mathcal{H}}}(A_{i,k})}{\|R_{\mathcal{H}}\|} \geqslant \mathrm{polylog}(n)^{-1}\cdot\lambda_{i+1}\cdot d\binom{|S|}{h-1}|A_{i,k}| \\ &= \mathrm{polylog}(n)^{-1}\cdot\frac{\lambda_{i+1}|S|^{h+k-2}}{|\Gamma_{i,k}|}|A_{i,k}|^2 \\ &= \mathrm{polylog}(n)^{-1}\cdot\frac{\lambda_{i+1}}{\lambda_i\,|S|}|A_{i,k}|^2 \\ &= \mathrm{polylog}(n)^{-1}\cdot\frac{n^{-\rho}}{|S|}\binom{|A|}{2}, \end{aligned} \tag{47}$$

where we used the definition of λ_i, λ_{i+1} in (38) 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 (39) and recall that $t \geqslant n^{1/(2h-1)+\delta}$), we obtain

$$e_{\mathrm{CG}_S}(A) \geqslant \frac{1}{t^{1-3\delta/4}\,\mathrm{polylog}(n)}\binom{|A|}{2} \gg \frac{1}{t^{1-\delta/2}}\binom{|A|}{2}.$$

This concludes the proof of Theorem 7.

5.2. **Deriving Theorem 3 from Theorem 7.** Our proof of Theorem 3 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 19 below, which was already implicitly proved in [9, Section 5]. For completeness, we include the proof of Lemma 19 in the Appendix A.

Lemma 19. *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 3. 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

$$|\mathcal{Z}_n^h(t)| \leqslant \left(\frac{n}{t^{h-\varepsilon/2}}\right)^t \quad \text{for all } \ell_h(n) \leqslant t \leqslant 2h\ell_h(n). \tag{48}$$

To this end, invoke Theorem 7 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,

$$F_h(S, \widetilde{S}, t) \leqslant \binom{\frac{n}{t^{h-1-8h^2\delta}}}{t-|S|} \quad \text{when } |\widetilde{S}| < \frac{n}{t^{h-1-8h^2\delta}}. \tag{49}$$

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 6 with

$$\begin{gathered} G = \mathrm{CG}_S[\widetilde{S}], \quad N = |\widetilde{S}|, \quad R = \frac{n}{t^{h-1-8h^2\delta}}, \\ \beta = \frac{1}{t^{1-\delta/2}}, \quad q = \lceil \beta^{-1} \log n \rceil, \quad \text{and} \quad m = t - q - |S|. \end{gathered} \tag{50}$$

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 6 that

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

As $q \ll t/(\log n)$ and $|\widetilde{S}| \leqslant n$, in view of both (49) and (51), we have

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

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

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

Consequently, (48) holds provided that $\delta = \delta(\varepsilon)$ is sufficiently small.

We no extend the upper bound given in (48) 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 (48), we conclude that

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

Lemma 19 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 17

Let S, A, Γ, and $\mathcal{H}$ be as in the statement of Theorem 17. 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

$$a_1 \oplus e_1 = a_2 \oplus e_2 \qquad \text{and} \qquad e_1 \cap e_2 = \emptyset. \tag{52}$$

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 (52). Our counting arguments rely heavily on the degree-regularity inherited from the pre-processing stage. We start with some definitions needed in both parts of the proof.

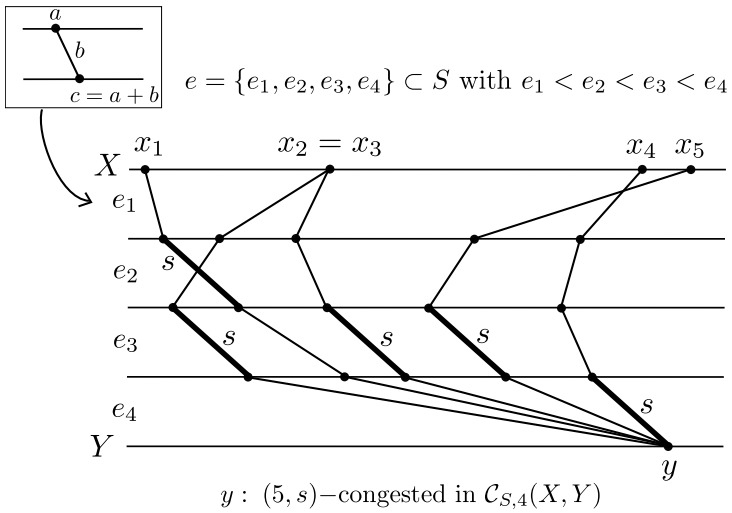

FIGURE 3. A congested vertex.

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

$$\mathcal{C}_{S,k}(X,Y) = \left\{(a,b) \in X \times Y : b = a \oplus e \text{ for some } e \in \binom{S}{k}\right\}. \tag{53}$$

Definition 20 (Congestion, see Figure 3)**.** For k, S, X, 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 $y \ominus e \in X$. 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 $s \in S$.

6.1. **Pre-processing stage.** In this subsection we state and prove the pre-processing lemma. In the next subsection, we use this lemma to establish Theorem 17. The presentation is quite long and technical. Since understanding the details of this proof is not necessary to follow the subsequent arguments, any reader who initially skips these next pages will perhaps be more motivated to return to them after seeing how this lemma integrates into our proof in §6.2.

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 §6.2.

Lemma 21. *Let $h \geqslant 2$, $\ell \in [h-1]$, $\beta \in (0,1)$, n be a sufficiently large integer, and $d \geqslant (128h\log_2 n)^{\ell+2}$. 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*

$$|\mathcal{C}_0| \geqslant \max\left\{\beta\binom{|S|}{\ell}\cdot|X|,\quad d\cdot|\Gamma_0|\right\}. \tag{54}$$

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

Proof of Lemma 21. 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

$$\alpha \geqslant \beta \cdot (256h^2\log_2 n)^{k-\ell}, \quad D \geqslant d \cdot (128h\log_2 n)^{k-\ell}, \tag{55}$$

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

$$|\mathcal{C}| \geqslant \max\left\{\alpha\binom{|S|}{k} \cdot |X|, \quad D \cdot |\Gamma|, \quad \frac{|\mathcal{C}_0|}{(128h\,|S|\log n)^{\ell-k}}\right\}. \tag{56}$$

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

$$\Gamma^{\mathtt{cong}} = \left\{b \in \Gamma\colon b \text{ is } \left\lceil \frac{\deg_{\mathcal{C}}(b)}{32h\log_2 n}\right\rceil\text{-congested in } \mathcal{C}\right\}. \tag{57}$$

Claim 4. *We have*

$$|\mathcal{C}_{S,k}(X,\Gamma^{\mathtt{cong}})| < \frac{|\mathcal{C}|}{2}. \tag{58}$$

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

$$|\mathcal{C}_{S,k}(X,\Gamma^{\mathtt{cong}})| = \sum_{b\in\Gamma^{\mathtt{cong}}} \deg_{\mathcal{C}}(b) \leqslant 32h\log_2 n \cdot |\Gamma^{\mathtt{cong}}| \leqslant 32h\log_2 n \cdot |\Gamma| \leqslant \frac{D}{4}\cdot|\Gamma| \overset{(56)}{\leqslant} \frac{|\mathcal{C}|}{4},$$

which establishes the claim for $k = 1$. Hence let us assume that $k \geqslant 2$ and, for the sake of a contradiction, that (58) 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

$$\begin{aligned}\Gamma^{\mathtt{cong}}_{+} &= \big\{b \in \Gamma^{\mathtt{cong}}\colon \deg_{\mathcal{C}}(b) \geqslant D/4\big\},\\ \Gamma' &= \big\{b - s_b\colon b \in \Gamma^{\mathtt{cong}}_{+}\big\},\\ \mathcal{C}' &= \mathcal{C}_{S,k-1}(X,\Gamma').\end{aligned} \tag{59}$$

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 $e_1, \dots, e_d \in \binom{S}{k}$ such that $s_b \in e_i$ and $b \ominus e_i \in X$ for all $i = 1, \dots, d$. Hence, setting $f_i = e_i \setminus \{s_b\}$ for each i, we obtain a collection of d distinct $(k-1)$-tuples such that $y \ominus f_i \in X$ 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

$$\forall y \in \Gamma', \quad \deg_{\mathcal{C}'}(y) \geqslant \max_{\substack{b \in \Gamma_+^{\mathtt{cong}} \\ y = b - s_b}} \left\{ \left\lceil \frac{\deg_{\mathcal{C}}(b)}{32h\log_2 n} \right\rceil \right\} \overset{(59)}{\geqslant} \frac{D}{128h\log_2 n}. \tag{60}$$

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 (58), it follows from (56) and the definitions of $\Gamma^{\mathtt{cong}}$ and $\Gamma_+^{\mathtt{cong}}$ that

$$|\mathcal{C}_{S,k}(X, \Gamma_+^{\mathtt{cong}})| = |\mathcal{C}_{S,k}(X, \Gamma^{\mathtt{cong}})| - |\mathcal{C}_{S,k}(X, \Gamma^{\mathtt{cong}} \setminus \Gamma_+^{\mathtt{cong}})| \overset{(59)}{\geqslant} \frac{|\mathcal{C}|}{2} - \frac{D}{4}|\Gamma| \overset{(56)}{\geqslant} \frac{|\mathcal{C}|}{4}.$$

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 (60), 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 (56)). □

Define for all $j \geqslant 0$,

$$\Gamma_j = \left\{ b \in \Gamma \setminus \Gamma^{\mathtt{cong}} \colon \deg_{\mathcal{C}}(b) \in [2^j, 2^{j+1} - 1] \right\}. \tag{61}$$

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

$$\Gamma^* \subset \Gamma \setminus \Gamma^{\mathtt{cong}}, \qquad \forall b \in \Gamma^*,\ \deg_{\mathcal{C}}(b) \in [2^{j^*}, 2^{j^*+1} - 1], \quad \text{and} \quad |\mathcal{C}_{S,k}(X, \Gamma^*)| \overset{(58)}{\geqslant} \frac{|\mathcal{C}|}{2\log_2 n}. \tag{62}$$

Thus,

$$\frac{D\,|\Gamma|}{2\log_2 n} \overset{(56)}{\leqslant} \frac{|\mathcal{C}|}{2\log_2 n} \leqslant |\mathcal{C}_{S,k}(X,\Gamma^*)| \leqslant (2^{j^*+1}-1)\cdot|\Gamma^*| \leqslant 2^{j^*+1}\cdot|\Gamma|,$$

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

$$\delta_1 \geqslant \frac{D}{4\log_2 n} \quad \text{and} \quad \forall b \in \Gamma^*,\ \deg_{\mathcal{C}}(b) \in [\delta_1, 2\delta_1 - 1]. \tag{63}$$

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

$$2\delta_1\,|\Gamma^*| \overset{(63)}{\geqslant} |C_{S,k}(X,\Gamma^*)| \overset{(62)}{\geqslant} \frac{|\mathcal{C}|}{2\log_2 n}$$

and

$$|\mathcal{C}_{S,k}(X,\bar{\Gamma})| \overset{(63)}{\geqslant} \delta_1\,|\bar{\Gamma}|.$$

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 (63), condition (*1-b*) follows immediately for any subset of Γ^*. We now check that condition (*1-a*) is also satisfied. Recall (57) and (62). 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 $|\bar{\Gamma}| \geqslant \frac{|\Gamma^*|}{3(\log_2 n)^2}$ 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 4. 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 $b = a \oplus e$. 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 $a \oplus e = b$. We will use this fact shortly.

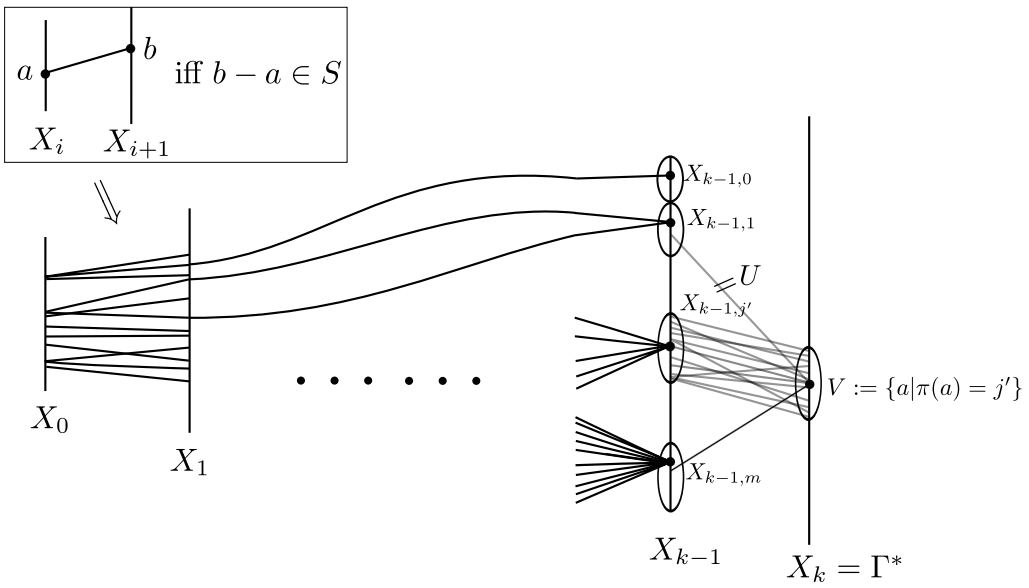

FIGURE 4. The auxiliary $(k+1)$-partite graph with parts $X_0, X_1, \ldots, X_k$.

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

$$X_{k-1,j} = \big\{u \in X_{k-1} \colon P_u \in [2^j, 2^{j+1} - 1]\big\}. \tag{64}$$

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

$$\forall v \in V, \quad \#\text{ of proper paths passing through } U \text{ and ending at } v \text{ is } \geqslant \frac{k! \deg_{\mathcal{C}}(v)}{\log_2 n}. \tag{65}$$

We will prove that

$$\deg_{\mathcal{C}_{S,1}(U,V)}(v) \geqslant 16h \quad \text{for every } v \in V. \tag{66}$$

Suppose for the sake of a contradiction, that for some $v \in V$, the above inequality fails. By (65), 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 $(x_0, \dots, x_{k-2}, u, v)$, the k-set $e = \{x_1 - x_0, \dots, x_{k-2} - x_{k-3}, u - x_{k-2}, v - u\} \in \binom{S}{k}$ satisfies $v - u \in e$ and $v \ominus e = x_0 \in X_0 = X$. 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 (66) 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 $m \geqslant 0$

$$V_m = \big\{v \in V\colon \deg_{\mathcal{C}_{S,1}(U,V)}(v) \in [16h \cdot 2^m, 16h \cdot 2^{m+1} - 1]\big\}, \tag{67}$$

and similarly as before, notice that $V_m = \emptyset$ for $m \geqslant \log_2 n \geqslant \log_2 |S|$. Observe also that (66) 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

$$|V_{m'}| \geqslant \frac{|V|}{\log_2 n} \geqslant \frac{|\Gamma^*|}{(\log_2 n)^2}. \tag{68}$$

Using Claim 5, (65), and (68), 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

$$N \geqslant \sum_{v \in V_{m'}} \frac{k!\deg_{\mathcal{C}}(v)}{\log_2 n} \geqslant \frac{k!}{\log_2 n} \cdot |\mathcal{C}_{S,k}(X, V_{m'})| \geqslant \frac{k!}{\log_2 n} \cdot \frac{|V_{m'}|}{4|\Gamma^*|\log_2 n}|\mathcal{C}| \overset{(68)}{\geqslant} \frac{k!\,|\mathcal{C}|}{4(\log n)^4}. \tag{69}$$

Since there are fewer than $|X|\,|S|^{k-1}$ proper $(k-1)$-paths, by (64) 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

$$\begin{aligned} \frac{1}{|U|} \sum_{u \in U} \deg_{\mathcal{C}_{S,1}(U,V_{m'})}(u) \;&\geqslant\; \frac{N}{2^{j'+1}|U|} \geqslant \frac{N}{2\,|X|\,|S|^{k-1}} \overset{(69)}{\geqslant} \frac{k!\,|\mathcal{C}|}{8(\log_2 n)^4|X|\,|S|^{k-1}} \\ &\overset{(56)}{\geqslant} \frac{k!\,\alpha\binom{|S|}{k} \cdot |X|}{8(\log_2 n)^4|X|\,|S|^{k-1}} \gg \frac{\alpha\,|S|}{(\log_2 n)^5}. \end{aligned} \tag{70}$$

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

$$\delta_2 = 4h \cdot 2^{m'}. \tag{71}$$

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 (55) we have $\alpha \geqslant \beta \cdot (256h^2 \log_2 n)^{k-\ell}$ and from (71), we have $\delta_2 \geqslant 4h$. Moreover, by the definition of $V_{m'}$ in (67), 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

$$|U| \cdot \frac{\alpha\,|S|}{(\log_2 n)^5} \overset{(70)}{\ll} |\mathcal{C}_{S,1}(U, V_{m'})|.$$

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

$$|V_{m'}|\,\delta_2 \overset{(67),(71)}{<} \sum_{v \in V_{m'}} \frac{\deg_{\mathcal{C}_{S,1}(U,V_{m'})}(v)}{4} \leqslant \frac{|\mathcal{C}_{S,1}(U, V_{m'})|}{4}.$$

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

$$|\bar{\Gamma}| \cdot 8\delta_2 \geqslant \sum_{b \in \bar{\Gamma}} \deg_{\mathcal{C}_{S,1}(Z,\bar{\Gamma})}(b) = |\mathcal{C}_{S,1}(Z, \bar{\Gamma})| \geqslant \frac{2\,|\mathcal{C}_{S,1}(U, V_{m'})|}{3} \overset{(67),(71)}{\geqslant} \frac{2\,|V_{m'}| \cdot 4\delta_2}{3}.$$

It follows that

$$|\bar{\Gamma}| \geqslant \frac{|V_{m'}|}{3} \overset{(68)}{\geqslant} \frac{|\Gamma^*|}{3(\log_2 n)^2}.$$

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,

$$|\mathcal{C}_1| = |\mathcal{C}_{S,k}(X, \bar{\Gamma})| \geqslant \frac{|\mathcal{C}|}{12(\log_2 n)^3} \overset{(56)}{\geqslant} \frac{|\mathcal{C}_0|}{12(\log_2 n)^3 \cdot (128h\,|S| \log n)^{\ell-k}},$$

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

6.2. **Completing the proof of Theorem 17.** 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 (52). 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,

$$\forall (x, y) \in \mathcal{C}_{S,k}(X, Y), \quad e_{y-x} = e_{S,k,y-x} \in \binom{S}{k} \text{ is the unique } k\text{-set satisfying } y = x \oplus e, \tag{72}$$

where $\mathcal{C}_{S,k}(X, Y)$ is the graph defined in (53). 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 17. We shall invoke Lemma 21 with h, ℓ, β, n, d, S, $X = A$, and $\Gamma_0 = \Gamma$. Note that the assumptions of Theorem 17 match those of Lemma 21, 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

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

where the inequality follows by the exact same requirement imposed by Theorem 17 on $A = X$. Lemma 21 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.

6.2.1. *Warm-up: 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 $b = a_1 \oplus e_{b-a_1} = a_1 \oplus e_{c-a_1} = c$. Therefore, one possible way to obtain a lower bound on the number of quadruples satisfying (52) is to establish how many $b \in \bar{\Gamma}$ are such that $b = a_1 \oplus e_1 = a_2 \oplus e_2$ 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 $(a_1, a_1 \oplus e_1 = a_2 \oplus e_2, a_2)$ 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 (72)). 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 20, 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

$$|\mathcal{C}_1|\frac{\delta_1}{2} \overset{(1\text{-}e)}{\geqslant} \frac{\delta_1 |\mathcal{C}_0|}{12(\log_2 n)^3} \geqslant \frac{\delta_1 |A|}{12(\log_2 n)^3}\beta\binom{|S|}{h-1} \tag{73}$$

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 $b = a \oplus e$ 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

$$4\delta_1 |A| \cdot \left|\binom{S}{h-1} \setminus \mathcal{H}\right| \overset{(36)}{\leqslant} 4\delta_1 |A| \cdot \frac{\beta^h}{(\log_2 n)^{7h^2}}\binom{|S|}{h-1}$$

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

$$\mathcal{Q} \overset{(73)}{\geqslant} \frac{\delta_1 |A|}{12(\log_2 n)^3}\beta\binom{|S|}{h-1} - 4\delta_1 |A| \cdot \frac{\beta^h}{(\log_2 n)^{7h^2}}\binom{|S|}{h-1} \geqslant \frac{\beta\delta_1}{24(\log_2 n)^3}|A|\binom{|S|}{h-1},$$

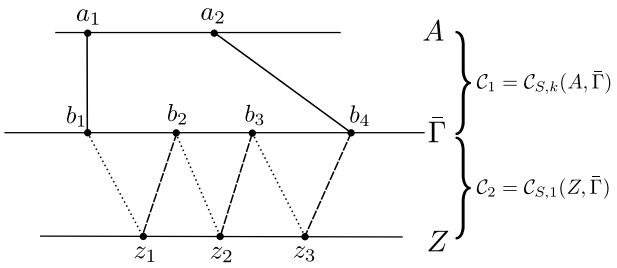

FIGURE 5. A special path in G.

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

6.2.2. *General case.* Since the simpler case when $k = \ell = h-1$ was handled in our warm-up (§6.2.1), 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 22 (Special path, see Figure 5)**.** A *special path* in G is a path of the form

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

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

$$\begin{aligned} a_1 \oplus e_1(P) &= \underbrace{a_1 \oplus e_{b_1 - a_1}}_{b_1} + (b_2 - z_1) + (b_3 - z_2) + \dots + (b_{h-k} - z_{h-k-1}) \\ a_2 \oplus e_2(P) &= \underbrace{a_2 \oplus e_{b_{h-k} - a_2}}_{b_{h-k}} + (b_1 - z_1) + (b_2 - z_2) + \dots + (b_{h-k-1} - z_{h-k-1}), \end{aligned} \tag{74}$$

and hence,

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

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 (52). 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 $e_1 = (e_{1,1}, \ldots, e_{1,h-1})$ and $e_2 = (e_{2,1}, \ldots, e_{2,h-1})$, the path $P = (a_1, b_1, z_1, b_2, z_2, \ldots, b_{h-k}, a_2)$ defined below is special (provided that it appears in the graph G):

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

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

$$\mathcal{Q} \geqslant \frac{N}{h^{2h}}. \tag{75}$$

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

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

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 20, this implies that for all $s \in S$, the number of k-tuples in $E_{b_{h-k}}$ containing s is at most $\frac{\delta_1}{16h\log_2 n} n$. 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 $z_1, \ldots, z_{h-k-1}$ can be chosen from among at least $\delta_2/2$ alternatives.
- Each element b_i, $i \in \{2, 3, \ldots, h-k\}$ 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

$$|\mathcal{C}_1| \geqslant \frac{|\mathcal{C}_0|}{12(\log_2 n)^3(128h\,|S|\log_2 n)^{\ell-k}} \overset{(54)}{\geqslant} \frac{\Omega(1)\cdot\beta\,|A|\binom{|S|}{\ell}}{|S|^{\ell-k}(\log_2 n)^{3+\ell-k}} = \frac{\Omega(1)\cdot\beta\,|A|\,|S|^k}{(\log_2 n)^{3+\ell-k}}.$$

Therefore, it follows that

$$\begin{aligned} N^* &\geqslant \Omega(1)\frac{\beta^{h-k}\,|A|\,|S|^{h-1}\delta_1\delta_2^{h-k-1}}{(\log_2 n)^{3+\ell-k+(5+\ell-k)(h-k-1)}} \\ &> \frac{\beta^h\delta_1\delta_2^{h-k-1}}{(\log_2 n)^{6h^2}}|A|\binom{|S|}{h-1}, \end{aligned} \tag{76}$$

where in the last inequality we used the fact that

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

Since this inequality is strict, the constant factors of the first inequality in (76) 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 $e_1(P) \notin \mathcal{H}$. Fix an arbitrary $e \in \binom{S}{h-1}\setminus\mathcal{H}$ and one of the $(h-1)!$ orderings of its elements, say $(e_1, \ldots, e_{h-1})$. 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

$$\left|\binom{S}{h-1} \setminus \mathcal{H}\right|(h-1)!|A|(4\delta_2)^{h-k-1}2\delta_1 \overset{(36)}{\leqslant} O(1)\cdot \frac{\beta^h\binom{|S|}{h-1}}{(\log_2 n)^{7h^2}}|A|\,\delta_1\delta_2^{h-k-1} \overset{(76)}{=} o(N^*). \tag{77}$$

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

$$\mathcal{Q} \overset{(75)}{\geqslant} \frac{N^*}{2h^{2h}} \geqslant \frac{1}{2h^{2h}}\frac{\beta^h\delta_1\delta_2^{h-k-1}}{(\log_2 n)^{6h^2}}|A|\binom{|S|}{h-1} \geqslant \frac{\beta^h d}{(\log_2 n)^{7h^2}}|A|\binom{|S|}{h-1},$$

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 17. □

7. Concluding Remarks

In this paper, we have established essentially tight bounds for the number of B_h-sets contained in the set $\{1,\dots,n\}$ of almost every given cardinality t. There is, 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 $[\varepsilon n^{1/(2h-1)}, n^{1/(2h-1)+\varepsilon}]$, 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 at most $n^{\varepsilon t}$ multiplicative factor.

There are therefore two directions in which our result could be refined. The first of them would be to improve the upper bound on $|\mathcal{Z}_n^h(t)|$ to $\left(\frac{f(n)n}{t^h}\right)^t$, where $f(n)$ is some explicit function; 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 (47). At the same time, the sequence of αs has to decrease quickly enough so that condition (36) in Theorem 17 is satisfied when we apply it in the proof of Theorem 7, see (46).

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 6 with larger values of R. The obtained upper bounds would still be similar to those proved by Theorem 3. In view of the lower bound of Proposition 2 (*i*), it is clearly not possible to reduce t below $n^{1/(2h-1)}$. A careful analysis of our proof shows that (43) is where a lower bound on t of that form is required. More precisely, for our application of Lemma 6 to work, we need $q = o(t)$, $|S| = o(t)$, and $\beta q = \Omega(\log n)$. For that reason, t must be at least $n^{1/(2h-1)}\,\mathrm{polylog}(n)$ for (43) to yield anything useful. In fact, it seems that any

proof based on Lemma 6 would require a threshold gap with factor at least $\log n$. Still, we believe that the following is true:

Conjecture 23. *For every $h \geqslant 2$, there exists a constant C_h such that*

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

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

7.1. **Related work.** Some recent results in extremal combinatorics have used the so-called *containers method* based on the main results of [1, 29]. This method was recently applied by Morris and Saxton [27] 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 [19, 18]. 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

$$f_{n,m}(C_{2h}) \leqslant \left(\frac{Cn^{h+1}}{m^h} \cdot \left(\log \frac{n^{h+1}}{m^h}\right)^{h-1}\right)^m. \tag{78}$$

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[2]. In particular, the bound (78) may be viewed as an analogue of our Theorem 3. However, we are not aware of any rigorous connection between these two results.

One might still ask whether the argument of [27] could be adapted to our setting. As in most applications of the containers method, the heart of [27] is proving a sufficiently strong supersaturation result for copies of C_{2h} in n-vertex graphs with more that $Dn^{1+1/h}$ edges, see [27, 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 [27]. However, a supersaturation statement that would be necessary for our application does not seem to follow from [27, 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 [27].

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[3] of $\left(\frac{cn^{h+1}}{m^h}\right)$ only by a factor of $\left(\log \frac{n^{h+1}}{m^h}\right)^{(h-1)m}$. On the other hand, our approach has the advantage of being entirely self-contained, since it relies merely on the simple Lemma 6 as opposed to this much more involved hypergraph version proved in [1, 29].

[2]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 $[h]$.

[3]One could derive such a bound in a similar fashion to [27, Proposition 1.4] and our Lemma 19 from the existence of n-vertex graphs with $\Omega(n^{1+1/h})$ edges which admit no non-backtracking closed walks of length $2h$.

References

1. J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, to appear in Journal of the American Mathematical Society.
2. R. C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helv. **37** (1962/1963), 141–147.
3. 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.
4. Sheng Chen, *On the size of finite Sidon sequences*, Proc. Amer. Math. Soc. **121** (1994), no. 2, 353–356.
5. 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.
6. Javier Cilleruelo, *New upper bounds for finite B_h sequences*, Adv. Math. **159** (2001), no. 1, 1–17.
7. D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, submitted.
8. D. Dellamonica Jr., Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij, *The number of B_3-sets of a given cardinality*, to appear in J. Combin. Theory Ser. A.
9. ———, *On the number of B_h-sets*, to appear in Combinatorics, Probability and Computing.
10. A. G. D'yachkov and V. V. Rykov, *B_s-sequences*, Mat. Zametki **36** (1984), no. 4, 593–601.
11. P. Erdős, *On a problem of Sidon in additive number theory and on some related problems. Addendum*, J. London Math. Soc. **19** (1944), 208.
12. 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.
13. Ben Green, *The number of squares and $B_h[g]$ sets*, Acta Arith. **100** (2001), no. 4, 365–390.
14. Heini Halberstam and Klaus Friedrich Roth, *Sequences*, second ed., Springer-Verlag, New York, 1983.
15. 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.
16. ———, *Turán's extremal problem in random graphs: forbidding odd cycles*, Combinatorica **16** (1996), 107–122.
17. Xing De Jia, *On finite Sidon sequences*, J. Number Theory **44** (1993), no. 1, 84–92.
18. Daniel J. Kleitman and David Bruce Wilson, *On the number of graphs which lack small cycles*, manuscript, 15 pp, 1996.
19. Daniel J. Kleitman and Kenneth J. Winston, *On the number of graphs without 4-cycles*, Discrete Math. **41** (1982), no. 2, 167–172.
20. Y. Kohayakawa, T. Łuczak, and V. Rödl, *Arithmetic progressions of length three in subsets of a random set*, Acta Arith. **75** (1996), 133–163.
21. ———, *On K^4-free subgraphs of random graphs*, Combinatorica **17** (1997), 173–213.
22. Yoshiharu Kohayakawa, Sang June Lee, Vojtěch Rödl, and Wojciech 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), no. 1, 1–25. MR 3291291
23. Mihail N. Kolountzakis, *The density of $B_h[g]$ sequences and the minimum of dense cosine sums*, J. Number Theory **56** (1996), no. 1, 4–11.
24. Fritz Krückeberg, *B_2-Folgen und verwandte Zahlenfolgen*, J. Reine Angew. Math. **206** (1961), 53–60.
25. S. J. Lee, *On sidon sets in a random set of vectors*, to appear in J. Korean Math. Soc.
26. Bernt Lindström, *A remark on B_4-sequences*, J. Combinatorial Theory **7** (1969), 276–277.
27. R. Morris and D. Saxton, *The number of $C_{2\ell}$-free graphs*, submitted.
28. Kevin O'Bryant, *A complete annotated bibliography of work related to Sidon sequences*, Electron. J. Combin. (2004), Dynamic surveys 11, 39 pp. (electronic).
29. D. Saxton and A. Thomason, *Hypergraph containers*, to appear in Inventiones mathematicae.
30. M. Schacht, *Extremal results for random discrete structures*, submitted.

31. I. E. Shparlinskiĭ, *On B_s-sequences*, Combinatorial analysis, No. 7 (Russian), Moskov. Gos. Univ., Moscow, 1986, pp. 42–45, 163.
32. S. Sidon, *Ein Satz über trigonometrische Polynome und seine Anwendung in der Theorie der Fourier-Reihen*, Math. Ann. **106** (1932), no. 1, 536–539.
33. James Singer, *A theorem in finite projective geometry and some applications to number theory*, Trans. Amer. Math. Soc. **43** (1938), no. 3, 377–385.

Appendix A. Omitted proofs

Proof of Lemma 19. 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., Y. Kohayakawa, 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 132-714, 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