THE NUMBER OF B_3-SETS OF A GIVEN CARDINALITY

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

Abstract. A set S of integers is a B_3*-set* if all the sums of the form $a_1 + a_2 + a_3$, with a_1, a_2 and $a_3 \in S$ and $a_1 \leq a_2 \leq a_3$, are distinct. We obtain asymptotic bounds for the number of B_3-sets of a given cardinality contained in the interval $[n] = \{1, \ldots, n\}$. We use these results to estimate the maximum size of a B_3-set contained in a typical (random) subset of $[n]$ of a given cardinality. These results confirm conjectures recently put forward by the authors [*On the number of* B_h*-sets*, submitted].

1. Introduction

Let $h \geq 2$ be an integer. A set S of integers is a B_h*-set* if for any z there is at most one sequence $a_1 \leq \cdots \leq a_h$ satisfying $z = a_1 + \cdots + a_h$ if we require that $a_i \in S$ for every $i = 1, \ldots, h$. The study of B_h-sets goes back to Sidon [14], who asked how large B_2-sets, or *Sidon sets*, can be if one imposes that they should be subsets of $[n] = \{1, \ldots, n\}$. Let

$$F_h(n) = \max\{|S| \colon S \subset [n] \text{ is a } B_h\text{-set}\}. \tag{1}$$

In the case addressed by Sidon, that is, for $h = 2$, results of Chowla, Erdős, Singer, and Turán [3, 5, 6, 15] from the 1940s tell us that $F_2(n) = (1 + o(1))\sqrt{n}$. The case of general h is less well understood. Bose and Chowla [1] showed that $F_h(n) \geq (1 + o(1))n^{1/h}$ for $h \geq 3$, while an easy argument gives that, for every $h \geq 3$ and large n,

$$F_h(n) \leq (h \cdot h! \cdot n)^{1/h} \leq h^2 n^{1/h}. \tag{2}$$

Note that, for $h = 3$, the first inequality in (2) gives that $F_3(n) \leq 3n^{1/3}$ for all large enough n. For general h, successively better bounds of the form $F_h(n) \leq c_h n^{1/h}$ have been obtained. The latest bounds are due to Green [7], who proved that

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

where $o(1) \to 0$ as $h \to \infty$. For a wealth of material on Sidon sets and on B_h-sets, the reader is referred to the classical monograph of Halberstam and Roth [8] and to a survey by O'Bryant [12].

Date: 2014/02/05, 10:29pm.

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 the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MSIP) (No. 2013042157). The fourth author was supported by the NSF grants DMS 0800070, 1301698, and 1102086. The fifth author was partially supported by ERC Advanced Grant DMMCA and a Trinity College JRF.

A related problem to bounding $F_h(n)$ is the problem of *estimating how many B_h-sets $[n]$ contains.* In fact, this problem was raised by Cameron and Erdős [2] in 1990 for $h = 2$. Let us introduce the following definition.

Definition 1.1 ($Z_h(n)$, $Z_h(n, s)$)**.** *For non-negative integers $1 \le s \le n$, let*

$$Z_h(n,s) = \Big|\big\{S \subset [n] \,:\, |S| = s \text{ and } S \text{ is a } B_h\text{-set}\big\}\Big|. \tag{4}$$

Furthermore, let $Z_h(n) = \sum_s Z_h(n, s)$.

In view of the fact that $F_h(n) = \Theta(n^{1/h})$, one sees that $c'_h n^{1/h} \le \log Z_h(n) \le C_h n^{1/h} \log n$ for some positive constants c'_h and C_h. One now knows that, in fact,

$$\log Z_h(n) \le C'_h n^{1/h} \tag{5}$$

for some constant C'_h. The case $h = 2$ of (5) is proved in [11] (see also [13]), and the arbitrary h case is dealt with in [4]. As it turns out, to establish (5), we considered the more refined question of estimating $Z_h(n, s)$. Roughly speaking, we obtained good bounds for $Z_h(n, s)$ for $s \ge n^{1/(h+1)}(\log n)^2$ and derived (5) summing over all relevant s (see [4] for details).

The problem of estimating $Z_h(n, s)$ for the whole range of s is interesting in its own right, and has an application to a certain problem in probabilistic combinatorics (we shall come back to this application in Sections 2 and 7). To develop a feel for the problem of estimating $Z_h(n, s)$, let us state lower bounds for this quantity, proved in [4].

Proposition 1.2 (Lower bounds for $Z_h(n, s)$)**.** *The following bounds hold for every $h \ge 2$.*

(*i*) *There is a constant $c'_h > 0$ such that, for all n and s, we have*

$$Z_h(n,s) \ge \left\lfloor \left(\frac{c'_h n}{s^h}\right)^s \right\rfloor. \tag{6}$$

(*ii*) *For any $\delta > 0$, there is an $\varepsilon > 0$ such that, for any $s \le \varepsilon n^{1/(2h-1)}$ and any large enough n, we have*

$$Z_h(n,s) \ge (1-\delta)^s \binom{n}{s}. \tag{7}$$

The lower bound in (6) may be proved coupling Bose and Chowla's construction [1] and a simple product construction. On the other hand, the lower bound in (7) comes from the fact that, for $s \le \varepsilon n^{1/(2h-1)}$, a typical s-element subset of $[n]$ becomes a B_h-set after the deletion of a small fraction of its elements.

Now, the lower bound in (7) tells us that, for $s \le \varepsilon n^{1/(2h-1)}$, the trivial upper bound $Z_h(n, s) \le \binom{n}{s}$ is sharp up to a factor of the form $(1 + o(1))^s$. The problem is, then, to obtain good upper bounds for $Z_h(n, s)$ for s of order $n^{1/(2h-1)}$ or larger, perhaps coming close to matching (6). We believe that this is possible, and put forward such a conjecture in [4], which we reproduce below for convenience.

Conjecture 1.3. *Fix an integer $h \ge 2$ and a real number $\delta > 0$. For every $s \ge n^{1/(2h-1)+\delta}$ and every large enough n, we have*

$$Z_h(n,s) \le \left(\frac{n}{s^{h-\delta}}\right)^s. \tag{8}$$

Conjecture 1.3 is proved for $h = 2$ in [10, 11]. The main result of this paper establishes Conjecture 1.3 for $h = 3$.

Theorem 1.4 (Main result)**.** *For every $\delta > 0$, there exists an integer n_0 such that if $n \geq n_0$ and $n^{1/5+\delta} \leq s \leq 3n^{1/3}$, then*

$$Z_3(n, s) \leq \left(\frac{n}{s^{3-\delta}}\right)^s. \tag{9}$$

We believe that our methods for proving Theorem 1.4 can eventually be adapted to establish Conjecture 1.3 for every h, but the general h case brings considerable new difficulties, and will be addressed elsewhere.

Let us compare the bounds we have for $Z_3(n, s)$ as s varies. For $s \ll n^{1/5}$, Proposition 1.2(ii) tells us that $Z_h(n, s)$ is, up to a multiplicative factor of $(1 - o(1))^s$, equal to the total number $\binom{n}{s}$ of s-element subsets of $[n]$. In this range, one might therefore say that B_h-sets are 'relatively abundant'. On the other hand, for any given $\delta > 0$, for $n^{1/5+\delta} \leq s \ll n^{1/3}$, Theorem 1.4 and Proposition 1.2(i) applied for $h = 3$ determine $Z_3(n, s)$ up to a multiplicative factor of the form $s^{o(s)}$, and we see that the probability that a random s-element subset of $[n]$ is a B_3-set is roughly of the form $s^{-(2+o(1))s}$. In this second range, B_3-sets are therefore scarcer. Finally, note that, by (2), if $s > 3n^{1/3}$ and n is large, then $Z_3(n, s) = 0$.

The discussion above tells us that there is a sudden change of behaviour around $s_0 = n^{1/5}$. Indeed, roughly speaking, for s considerably larger than this 'critical' value s_0, we have that $Z_3(n, s)$ is of the form $\left(n/s^{3-o(1)}\right)^s$; this is in contrast to the fact that, as we have already seen, for s of smaller order than s_0, we have that $Z_3(n, s)$ is of the form $(1 - o(1))^s \binom{n}{s} = (\Theta(n/s))^s$.

Theorem 1.4 implies a result in probabilistic combinatorics, which confirms the case $h = 3$ of a conjecture put forward in [4]. We shall discuss this corollary of Theorem 1.4 in Section 2.

Notation and organization of the paper. Throughout this paper we identify a graph with the set of its edges. In particular, if G is a graph, then $e \in G$ means that e is an edge of G; moreover, we write both $|G|$ and $e(G)$ for the number of edges in G. If $e = \{x, y\}$ is an edge in a graph, we sometimes write xy for e. As usual, edges are unordered pairs of vertices; however, if H is a bipartite graph with vertex classes A and B, it will be convenient to think of the edge set of H as a subset of $A \times B$ in the natural way. For a set $A \subset V(G)$ we denote by $e(A) = e_G(A)$ the number of edges in the subgraph induced by A, which is denoted by $G[A]$. If T is a set, we denote by $K(T)$ the complete graph with vertex set T. For a set $W \subset \mathbb{Z}$ and $x \in \mathbb{Z}$ the set $W + x$ is defined as the set of all numbers $w + x$ with $w \in W$.

We write $a \ll b$ as shorthand for the statement $a/b \to 0$ as $n \to \infty$. We use the standard O, o and Θ-notation (with respect to $n \to \infty$); the implicit constants are always absolute constants. We omit floor $\lfloor\ \rfloor$ and ceiling $\lceil\ \rceil$ symbols when they are not essential. We sometimes write a/bc for $a/(bc)$. We are mostly interested in large n; in our statements and inequalities we often tacitly assume that n is larger than a suitably large constant.

This paper is organized as follows. In Section 2 we state and prove the result in probabilistic combinatorics very briefly alluded to above. The remainder of the paper is devoted to the proof of Theorem 1.4, the structure of which is presented in Figure 1. The general approach used in the proof

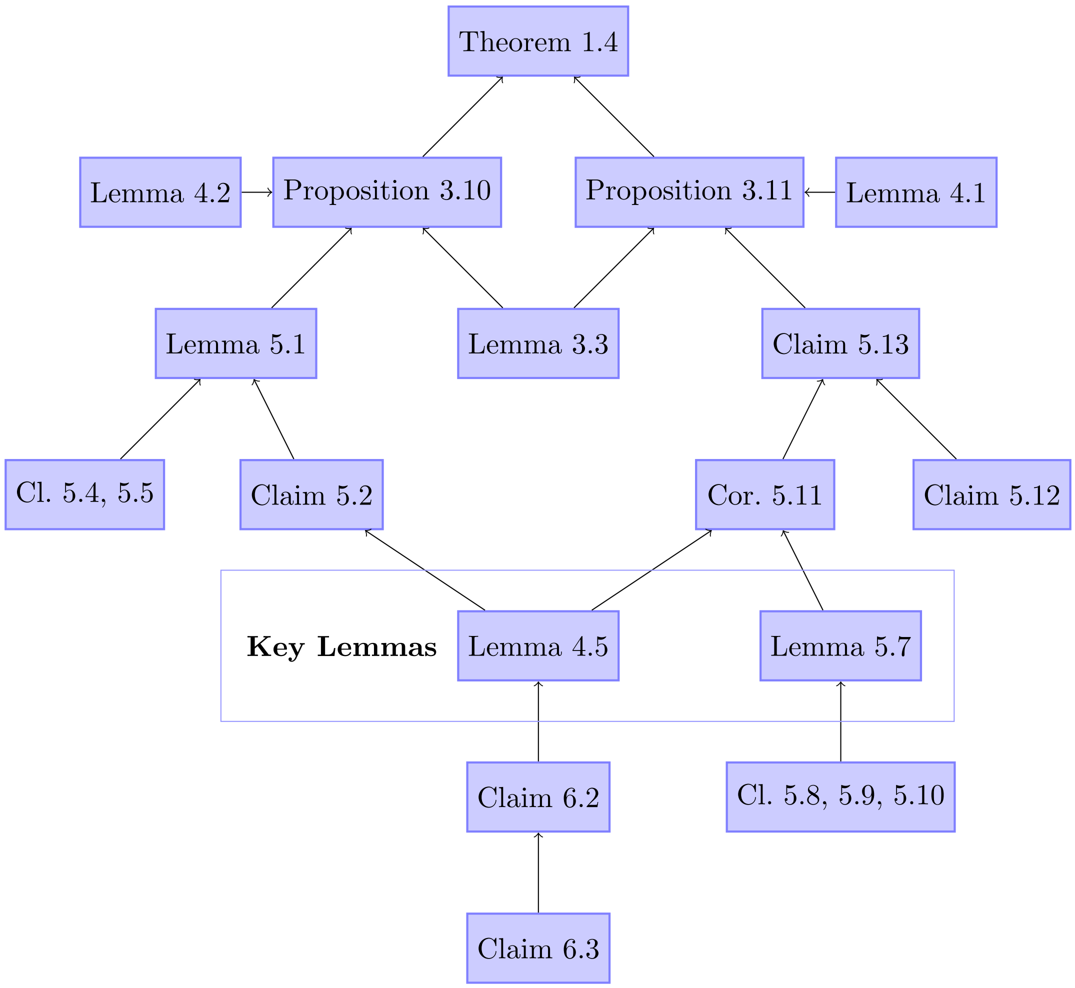

FIGURE 1. A diagram illustrating the flow of the proof of our main result

is described in Section 3. Section 4 gives some auxiliary lemmas, one of which, Lemma 4.5, plays a central technical rôle. The proof of this lemma is given in Section 6. The two main propositions that together imply Theorem 1.4 (Propositions 3.10 and 3.11; see Section 3) are proved in Section 5.

2. LARGEST B_3-SETS CONTAINED IN RANDOM SETS OF INTEGERS

In [10, 11], the cardinality of the largest B_2-sets, i.e., *Sidon sets*, contained in random sets of integers was investigated. Given an integer function $0 \leq m = m(n) \leq n$, let us denote by $[n]_m$ an m-element subset of $[n]$ chosen uniformly at random from all such sets. Given a set R, let $F_h(R)$ be the cardinality of the largest B_h-sets contained in R. We are interested in the random variable $F_h([n]_m)$.

For simplicity, let us suppose $m = m(n) = (1+o(1))n^a$ for some constant $0 < a < 1$. It is proved in [10, 11] that, asymptotically almost surely, that is, with probability tending to 1 as $n \to \infty$, one has $F_2([n]_m) = n^{b_2+o(1)}$, where

$$b_2 = b_2(a) = \begin{cases} a & \text{if } 0 \le a \le 1/3, \\ 1/3 & \text{if } 1/3 \le a \le 2/3, \\ a/2 & \text{if } 2/3 \le a \le 1. \end{cases} \tag{10}$$

Therefore, $F_2([n]_m)$ undergoes a sudden change of behaviour at $a = 1/3$ and at $2/3$. Furthermore, somewhat unexpectedly, $F_2([n]_m)$ does not change considerably as we vary a from $1/3$ to $2/3$. It is natural to ask whether a similar result holds for arbitrary h; indeed, in [4], we put forward a conjecture that states that this is the case. Theorem 1.4 implies that this conjecture holds for $h = 3$. Our result is as follows.

Theorem 2.1 (B_3-sets contained in random sets of integers)**.** *Let $0 \le a \le 1$ be a fixed constant. Suppose $m = m(n) = (1 + o(1))n^a$. There exists a constant $b_3 = b_3(a)$ such that, asymptotically almost surely, we have*

$$F_3([n]_m) = n^{b_3+o(1)}. \tag{11}$$

Furthermore,

$$b_3 = b_3(a) = \begin{cases} a & \text{if } 0 \le a \le 1/5, \\ 1/5 & \text{if } 1/5 \le a \le 3/5, \\ a/3 & \text{if } 3/5 \le a \le 1. \end{cases} \tag{12}$$

The piecewise linear function b_3 in (12) is given in Figure 2.

Proof of Theorem 2.1. We shall be somewhat sketchy in the more routine parts of the argument. We first observe that one may switch to the so called binomial model $[n]_p$. To be more precise, let $p = m/n = (1 + o(1))n^{a-1}$ and put each $x \in [n]$ in $[n]_p$ with probability p, independently of all other elements in $[n]$. A standard argument tells us that it suffices to prove that $F_3([n]_p) = n^{b_3+o(1)}$ with probability $1 - o(1/\sqrt{m})$.

The required lower bound for $F_3([n]_p)$ is established in [4]. Since $F_3([n]_p) \le |[n]_p|$, standard arguments prove Theorem 2.1 in the range $a \in [0, 1/5]$. We may use Theorem 1.4 to bound the random variable $F_3([n]_p)$ from above, in probability, as follows. The expected number of B_3-sets of size s in $[n]_p$ is $p^s Z_3(n, s)$. For any given $\delta > 0$, Theorem 1.4 implies that, for $s \ge n^{1/5+\delta}$, this expectation is at most

$$\left(p\,\frac{n}{s^{3-\delta}}\right)^s. \tag{13}$$

Hence, if $(1 + o(1))n^a = pn \ll s^{3-\delta}$, then this expectation is $o(1/\sqrt{m})$. In particular, for every $a > 1/5$, with suitably large probability, the largest B_3-sets contained in $[n]_p$ have cardinality at most

$$\max\{n^{1/5+\delta}, n^{a/3+\delta}\} = n^{b_3(a)+\delta}.$$

Since $\delta > 0$ is arbitrary, the result follows. $\square$

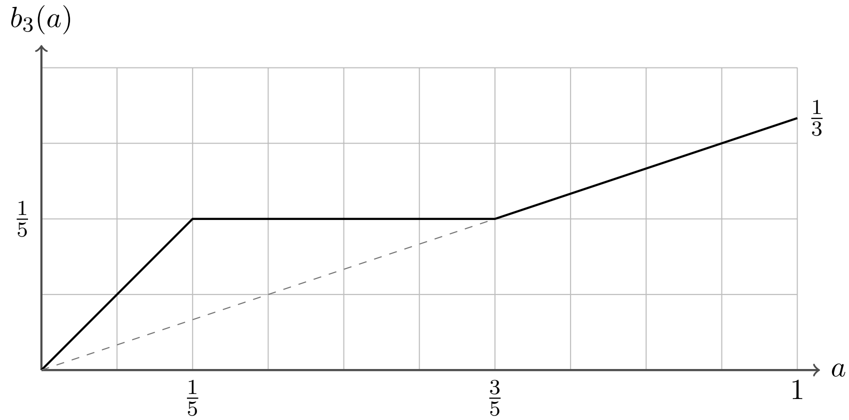

FIGURE 2. The graph of the piecewise linear function b_3 from Theorem 2.1

3. THE PROOF OF THEOREM 1.4

Theorem 1.4 follows in a straightforward manner from two propositions, Propositions 3.10 and 3.11, stated at the end of this section. We need some preparations to be able to state those two propositions. The following definition introduces a central object in the proof.

Definition 3.1 (Collision graph $\mathcal{C}_T$)**.** *Given a set $T \subset [n]$, we define the collision graph $\mathcal{C}_T$ on the vertex set $[n]$ by letting $\{a, b\}$ with a, $b \in [n]$ and $a \neq b$ be an edge whenever there exist z_1, z_2, z_3, $z_4 \in T$ such that*

$$a + z_1 + z_2 = b + z_3 + z_4. \tag{14}$$

Proposition 3.2. *Suppose that $S \subset [n]$ is a B_3-set. Then for every $T \subset S$, the set $S \setminus T$ is an independent set in $\mathcal{C}_T$.*

Proof. Suppose on the contrary that a, $b \in S \setminus T$ with $a \neq b$ satisfies (14) with z_1, z_2, z_3, $z_4 \in T$. From the fact that S is a B_3-set we deduce that the multisets $\{a, z_1, z_2\}$ and $\{b, z_3, z_4\}$ coincide. Since $a \in S \setminus T$ and z_3 and $z_4 \in T$, we obtain that $a = b$, which is a contradiction. □

In view of Proposition 3.2, our general strategy for estimating the number of B_3-sets of a given size s will be as follows: we first enumerate *seed* B_3-sets T with $|T| \ll s$ and then we bound the number of independent sets in $\mathcal{C}_T$ for each such T. The following lemma, which is implicit in the work of Kleitman and Winston [9] (see also [11, Lemma 3.1]), will be used to bound the number of independent sets.

Lemma 3.3. *Let G be a graph on N vertices, let q be an integer and let $0 \leq \beta \leq 1$ and R be real numbers with*

$$R \geq e^{-\beta q} N. \tag{15}$$

Suppose

$$e(A) \geq \beta \binom{|A|}{2} \text{ for any } A \subset V(G) \text{ with } |A| \geq R. \tag{16}$$

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

$$\binom{N}{q}\binom{R}{m}. \tag{17}$$

When applying Lemma 3.3, we shall often take $R = \gamma N = \gamma|V(G)|$ for some number $\gamma > 0$. Hypothesis (15) then becomes

$$e^{\beta q}\gamma > 1 \tag{18}$$

and the bound (17) becomes

$$\binom{|V(G)|}{q}\binom{\gamma|V(G)|}{m}. \tag{19}$$

In order to prove our main result, Theorem 1.4, we shall enumerate all possible seed sets T and show that the corresponding graphs $\mathcal{C}_T$ are quite dense. In fact, they are dense enough that we can apply Lemma 3.3 to establish that the number of extensions of T to a significantly larger B_3-set S is rather small. More precisely, for every B_3-set T we are interested in proving lower bounds for $e_{\mathcal{C}_T}(A)$ for arbitrary but somewhat large $A \subset [n]$. It turns out that we shall need to consider two separate cases, depending on the structure of T. We now need some definitions.

Let G be a graph on the vertex set $T \subset [n]$ and let $z \in [n]$ be arbitrary. Denote by $G^2 = G \times G$ the Cartesian product of the edge set of G with itself.

Definition 3.4 (Representation count R_G)**.** *Let*

$$R_G(z) = \left|\left\{(z_1z_2, z_3z_4) \in G^2 \ : \ z = (z_1+z_2) - (z_3+z_4), \ \textit{all } z_i\textit{s distinct}\right\}\right|. \tag{20}$$

Clearly $R_G(-z) = R_G(z)$ for every z. In what follows, T will always be a B_3-set, and hence we shall always have $R_G(0) = 0$. Finally, we mention that we shall only be interested in $R_G(z)$ for $z \in \{-n+1, \dots, -1\} \cup \{1, \dots, n-1\} \subset [-n, n]$.

Definition 3.5 (Collision multigraph $\widetilde{\mathcal{C}}_G$)**.** *Let $\widetilde{\mathcal{C}}_G$ be the multigraph with vertex set $[n]$ in which the multiplicity of each $\{a,b\} \in \binom{[n]}{2}$ is exactly $R_G(b-a)$.*

The reason we introduce this multigraph version of $\mathcal{C}_T$ is that it will be easier to estimate from below the number of multi-edges that are induced by subsets $A \subset [n]$. We can then establish bounds for $\mathcal{C}_T$ through the following proposition.

Proposition 3.6. *For every non-empty graph $G \subset \binom{T}{2}$ and $A \subset [n]$ we have*

$$e_{\mathcal{C}_T}(A) \max_{z \in [-n,n]} R_G(z) \geq e_{\widetilde{\mathcal{C}}_G}(A). \tag{21}$$

Proof. Note that $\{a,b\} \in \widetilde{\mathcal{C}}_G[A]$ implies that $R_G(b-a) \geq 1$ which means that $b - a = (z_1+z_2) - (z_3+z_4)$ for some $z_i \in T$, and thus $\{a,b\} \in \mathcal{C}_T[A]$ (see Definition 3.1). The proposition follows. □

A substantial part of this paper is dedicated to proving the existence of suitable graphs G for which we can bound $\max_{z \in [-n,n]} R_G(z)$ and then apply Proposition 3.6.

Remark 3.7. Note that, in the definition of $R_G(z)$, the elements z_1, z_2, z_3 and z_4 are required to be all distinct. This restriction allows us to avoid the "degenerate" case in which $z = z_1 - z_3$

for $z_1, z_3 \in T$. In this case, for every $x \in N_G(z_1) \cap N_G(z_3)$ we have $z = (z_1 + x) - (z_3 + x)$ with $(z_1x, z_3x) \in G^2$, which means that, without the restriction, the value of $R_G(z)$ could be as large as $|N_G(z_1) \cap N_G(z_3)|$, which, in turn, could be as large as $|T| - 2$.

We now define a quantity Q_G that will help us bound $\max_{z\in[-n,n]} R_G(z)$.

Definition 3.8 (Moment generating function of R_G)**.** *Let*

$$Q_G = \sum_{z\in[n]} \exp R_G(z). \tag{22}$$

As already observed, R_G vanishes at 0 for any B_3-set T and is an even function. Thus

$$\max_{z\in[-n,n]} R_G(z) = \max_{z\in[n]} R_G(z) \leq \log Q_G. \tag{23}$$

In view of the definitions above and Proposition 3.6 and Lemma 3.3, our goal is to enumerate B_3-sets $T \subset [n]$ and graphs $G \subset \binom{T}{2}$ such that $\log Q_G$ is not too large, while at the same time $e_{\widetilde{\mathcal{C}}_G}(A)$ is large for every "large enough" set $A \subset [n]$. This discussion motivates our next definition.

Definition 3.9 (Bounded set)**.** *Let*

$$\xi = \frac{1}{10^6 \log n} \quad \text{and} \quad \alpha_i = \xi^{2^{i+1}-1} \text{ for } i \geq 0, \tag{24}$$

and let $\varepsilon > 0$ be a fixed constant. Given $\lambda \geq 1$ and a non-negative integer i, a set T is said to satisfy $\mathcal{P}_{\lambda,\varepsilon,i}$ if it is a B_3-set and there exists a graph G_i on the vertex set T such that

(a) $e(G_i) \geq \big(1 - \alpha_i\big)\binom{|T|}{2}$,

(b) $Q_{G_i} \leq en \exp\left(\lambda n^{i\varepsilon} \sum_{j=1}^{|T|} \frac{1}{j}\right)$.

A set T is called (λ, ε)-bounded if it satisfies $\mathcal{P}_{\lambda,\varepsilon,i}$ for $i = 0, 1, \ldots, \lceil 1/\varepsilon \rceil$.

We now summarize our strategy for proving Theorem 1.4. Fix a positive constant $\delta > 0$. We may and shall suppose that $\delta \leq 1$. Let $\varepsilon = \delta/13$ and note that

$$\frac{1}{5 - 25\varepsilon} \leq \frac{1}{5} + \delta \qquad \text{and} \qquad 3 - 13\varepsilon = 3 - \delta \leq 3 - 12\varepsilon. \tag{25}$$

Let s be an integer satisfying the assumptions of the theorem. In particular, $s \geq n^{1/(5-25\varepsilon)}$ by our choice of ε. Our goal is to estimate the number of B_3-sets of cardinality s. For the remainder of the paper, we let

$$\lambda = \lambda(s) = \frac{s^{5-25\varepsilon}}{n} \geq 1. \tag{26}$$

We classify the B_3-sets of size s into two types, *depending on whether or not the cardinality of their largest (λ, ε)-bounded subsets is greater than $s^{1-6\varepsilon}$*. We shall prove the following two propositions, estimating the number of B_3-sets of cardinality s of each type separately. These two propositions together easily imply Theorem 1.4.

Proposition 3.10. *Let $\varepsilon > 0$, let n be a sufficiently large integer, and let $s \in [n^{1/(5-25\varepsilon)}, 3n^{1/3}]$ be a given integer. Let λ be as defined in* (26)*. The number of B_3-sets of cardinality s contained* *in* $[n]$ *that* ***contain a*** (λ,ε)***-bounded set larger than*** $s^{1-6\varepsilon}$ *is at most*

$$\left(\frac{n}{s^{3-12\varepsilon-o(1)}}\right)^s. \tag{27}$$

Proposition 3.11. *Let $\varepsilon > 0$, let n be a sufficiently large integer, and let $s \in [n^{1/(5-25\varepsilon)}, 3n^{1/3}]$ be a given integer. Let λ be as defined in* (26)*. The number of B_3-sets of cardinality s contained in $[n]$ that* ***do not contain any*** (λ,ε)***-bounded set larger than*** $s^{1-6\varepsilon}$ *is at most*

$$\left(\frac{n}{s^{3-8\varepsilon-o(1)}}\right)^s. \tag{28}$$

Before we proceed with the formal proofs, let us briefly discuss our general approach. Every B_3-set with s elements that contains a (λ,ε)-bounded set with at least $s^{1-6\varepsilon}$ elements will be shown to contain a set T with $|T| = s^{1-6\varepsilon}$ which satisfies $\mathcal{P}_{100\lambda,\varepsilon,0}$ (see Lemma 4.2). Using Lemma 3.3, we shall be able bound the number of possible extensions of any such set T to a B_3-set with s elements. This is because the graph $\mathcal{C}_T$ will be shown to satisfy an appropriate local density condition (see Lemma 5.1). Showing this is the main difficulty in this part of the argument. The details are given in Section 5.1.

The proof of Proposition 3.11 is somewhat more complicated. First we show that any B_3-set of cardinality s must contain a (λ,ε)-bounded subset of size at least $s^{1/7}$ (see Lemma 4.1). In particular, every such B_3-set contains a *maximal* (λ,ε)-bounded subset with at least $s^{1/7}$ elements. Our strategy will therefore be to estimate, for each B_3-set T with $|T| < s^{1-6\varepsilon}$, the number of B_3-sets S such that $T \subset S$ and T is a *maximal* (λ,ε)-bounded subset of S. The maximality of T will be shown to imply that the set of elements that can appear in $S \setminus T$ admits a certain structure (see Definition 5.6 and Lemma 5.7). More concretely, we shall show that $S \setminus T \subset \widetilde{T}$ for some set $\widetilde{T} \subset [n]$ such that the graph $\mathcal{C}_T[\widetilde{T}]$ satisfies certain local density conditions that allow us to use Lemma 3.3 to bound the total number of such possible extensions S of T appropriately (the precise local density condition is given in Corollary 5.13)

The remainder of the paper is devoted to proving Propositions 3.10 and 3.11.

4. Auxiliary lemmas

We now give three auxiliary lemmas. The two lemmas in Section 4.1 are quite simple, while the lemma given in Section 4.2, Lemma 4.5, is somewhat more technical. However, Lemma 4.2 will be one of the key lemmas that will allow us to prove local density results for certain induced subgraphs of the collision graph $\mathcal{C}_T$.

4.1. Bounded sets.

Our first lemma states that for any $\lambda \geq 1$ and any $\varepsilon > 0$, every B_3-set S contains a (λ,ε)-bounded subset whose size is at least a small power of $|S|$.

Lemma 4.1. *For any $\lambda \geq 1$, $\varepsilon > 0$, and B_3-set $S \subset [n]$ there exists a (λ,ε)-bounded set $T \subset S$ of cardinality $|T| \geq |S|^{1/7}$.*

Proof. Observe that S contains a B_4-set with $\lceil |S|^{1/7} \rceil$ elements. Indeed, one may construct such set greedily by starting from an empty set and sequentially adding to it elements of S. As long as the constructed set T has fewer than $|S|^{1/7}$ elements, one can always add to T an arbitrary element from the (non-empty) set $S \setminus (4T - 3T)$, which assures that T remains a B_4-set.

Hence, we may choose a B_4-set $T \subset S$ with $|T| = \lceil |S|^{1/7} \rceil$. Let G be the complete graph on the vertex set T. The fact that T is a B_4-set implies that $R_G(z) \in \{0, 1\}$ for every z. Indeed, if some $(z_1 z_2, z_3 z_4), (z_1' z_2', z_3' z_4') \in G^2$, with $|\{z_1, z_2, z_3, z_4\}| = 4$ and $|\{z_1', z_2', z_3', z_4'\}| = 4$, satisfy

$$(z_1 + z_2) - (z_3 + z_4) = z = (z_1' + z_2') - (z_3' + z_4'), \tag{29}$$

then

$$z_1 + z_2 + z_3' + z_4' = z_1' + z_2' + z_3 + z_4. \tag{30}$$

Since T is a B_4-set, we must have, as multisets,

$$\{z_1, z_2, z_3', z_4'\} = \{z_1', z_2', z_3, z_4\}. \tag{31}$$

This forces $\{z_1, z_2\} = \{z_1', z_2'\}$ and $\{z_3, z_4\} = \{z_3', z_4'\}$. Consequently $R_G(z) \leq 1$.

In particular, Q_G is a sum of n terms which are either $e^0 = 1$ or $e^1 = e$. Therefore,

$$Q_G \leq en. \tag{32}$$

Clearly, $G_i = G = K(T)$ satisfies both (a) and (b) of Definition 3.9 for any $i \geq 0$, $\lambda \geq 1$, and $\varepsilon > 0$. Hence, T is a (λ, ε)-bounded set. □

The second lemma allows one to pass to subsets of a convenient cardinality when dealing with (λ, ε)-bounded sets. Moreover, we shall see that we may carry out this procedure without significantly affecting the "boundedness" parameters.

Lemma 4.2. *Let $\lambda \geq 1$ and an integer $i \geq 0$ be given and suppose that $T \subset [n]$ satisfies $\mathcal{P}_{\lambda,\varepsilon,i}$. For every m satisfying $n^{1/100} \leq m \leq |T|$, there exists $T' \subset T$ with $|T'| = m$ such that T' satisfies $\mathcal{P}_{100\lambda,\varepsilon,i}$.*

Proof. Let G_i be a graph whose existence is asserted in the definition of $\mathcal{P}_{\lambda,\varepsilon,i}$. A simple averaging argument shows that there exists a $T' \subset T$ with $|T'| = m$ such that

$$e_{G_i}(T') \geq e(G_i) \binom{|T| - 2}{m - 2} \binom{|T|}{m}^{-1} = e(G_i) \binom{m}{2} \binom{|T|}{2}^{-1}. \tag{33}$$

Taking $G_i' = G_i[T']$ and recalling that G_i satisfies (a) of Definition 3.9 yields

$$\begin{aligned} e(G_i') &\geq (1 - \alpha_i) \binom{|T|}{2} \binom{m}{2} \binom{|T|}{2}^{-1} \\ &= (1 - \alpha_i) \binom{m}{2}. \end{aligned} \tag{34}$$

Using the facts that $n^{1/100} < m \leq |T| \leq 3n^{1/3}$, that G_i satisfies (b) of Definition 3.9, and well-known estimates for the harmonic numbers, we obtain, for every large enough n,

$$Q_{G_i'} \leq Q_{G_i} \leq en \exp\left(\lambda n^{i\varepsilon} \sum_{j=1}^{|T|} \frac{1}{j} \right) \leq en \exp\left(100 \lambda n^{i\varepsilon} \sum_{j=1}^{|T'|} \frac{1}{j} \right). \tag{35}$$

It follows from (34) and (35) that T' satisfies $\mathcal{P}_{100\lambda,\varepsilon,i}$. □

4.2. **A technical lemma on the local density of $\mathcal{C}_T$.** We now state a key technical lemma that will help us give lower bounds for the local density of the collision graph $\mathcal{C}_T$. We need the following definition.

Definition 4.3 (Bipartite graph $H_T(A,B)$)**.** *Given a set $T \subset [n]$, we define the graph H_T on the vertex set $([n] \times \{1\}) \cup ([2n] \times \{2\})$ by letting $\{(a,1),(b,2)\}$ be an edge whenever $b-a \in T$. For $A \subset [n]$ and $B \subset [2n]$, we denote by $H_T(A,B)$ the subgraph of H_T induced by $(A\times\{1\})\cup(B\times\{2\})$.*

Remark 4.4. In the definition above, we wish to have the disjoint union of $[n]$ and $[2n]$ as the vertex set of H_T. The standard way of producing such a disjoint union involves the use of the Cartesian product, as above. In what follows, we shall be less formal and we shall refer to vertices as $a \in A$, $b \in B$, etc, instead of $(a,1) \in A \times \{1\}$, $(b,2) \in B \times \{2\}$, etc.

The following technical lemma allows us to obtain a lower bound on $e_{\mathcal{C}_T}(A)$ in terms of the edge-density of the graph $H_T(A,B)$ for every sufficiently large B_3-set T that satisfies $\mathcal{P}_{\lambda,\varepsilon,i}$. When applying this lemma, we have to come up with a suitable set B (see the proofs of Claim 5.2 and Corollary 5.11). Recall that ξ and the α_i are defined in 24.

Lemma 4.5. *The following holds for every integer $i \geq 0$, and every $\varepsilon > 0$, $\lambda \geq 1$, $D \geq 5000$ and $\delta \in (0,1]$ satisfying $\delta^2 \geq \alpha_i/100\xi$. Suppose that a set $T \subset [n]$ with at least $n^{1/100}$ elements satisfies $\mathcal{P}_{\lambda,\varepsilon,i}$. Moreover, suppose that $A \subset [n]$ and $B \subset [2n]$ are such that the graph $H = H_T(A,B)$ satisfies*

(I) *every vertex of A has degree at least $\delta\,|T|$;*

(II) *the average degree of the vertices in B is D.*

Then

$$e_{\mathcal{C}_T}(A) = \Omega\left(\frac{\delta^2|A|\,D^2|T|^2}{\lambda n^{i\varepsilon}(\log n)^3}\right). \tag{36}$$

The proof of Lemma 4.5 will be given in Section 6.

5. Proofs of Propositions 3.10 and 3.11

5.1. **Sets containing a large bounded subset.** Let us now prove Proposition 3.10, which deals with the case in which the "seed" set contains a (λ,ε)-bounded set of cardinality greater than $n^{1-6\varepsilon}$. Our main tools will be Lemma 3.3 and the following estimate on the number of edges induced by small sets of vertices in the collision graph $\mathcal{C}_T$ when T is a bounded set.

Lemma 5.1. *There exists an absolute constant $C > 0$ such that the following holds. If T satisfies $\mathcal{P}_{100\lambda,\varepsilon,0}$ and $|T| \geq n^{1/100}$, then for any $A \subset [n]$ with $|A| \geq (C/|T|^2)n$, we have*

$$e_{\mathcal{C}_T}(A) = \Omega\left(\frac{|A|^2|T|^4}{\lambda n(\log n)^3}\right). \tag{37}$$

We give the proof of Proposition 3.10 before proving Lemma 5.1.

Proof of Proposition 3.10. We wish to estimate the number of B_3-sets S of size s that contain a (λ,ε)-bounded subset with more than $s^{1-6\varepsilon}$ elements. Suppose that $T \subset S$ is a (λ,ε)-bounded set with $|T| \geq s^{1-6\varepsilon}$. By Definition 3.9, T must satisfy $\mathcal{P}_{\lambda,\varepsilon,i}$ for $i = 0$. By Lemma 4.2, we may assume without loss of generality that the cardinality of T is exactly $s^{1-6\varepsilon}$ and T satisfies $\mathcal{P}_{100\lambda,\varepsilon,0}$. Lemma 5.1 then implies that $\mathcal{C}_T$ satisfies (16) with

$$R = \gamma n = \frac{Cn}{|T|^2} = \frac{Cn}{s^{2-12\varepsilon}} \tag{38}$$

and

$$\beta = \Omega\left(\frac{|T|^4}{\lambda n(\log n)^3}\right) \overset{(26)}{=} \Omega\left(\frac{(s^{1-6\varepsilon})^4}{s^{5-25\varepsilon}(\log n)^3}\right) = \Omega\left(\frac{1}{s^{1-\varepsilon}(\log n)^3}\right). \tag{39}$$

Take

$$q = (\log n)/\beta = O\big((\log n)^4 s^{1-\varepsilon}\big) = o(s). \tag{40}$$

Note that (18) is satisfied. Hence, Proposition 3.2 and Lemma 3.3 yield that the number of B_3-sets of cardinality s that contain a set T satisfying $\mathcal{P}_{100\lambda,\varepsilon,0}$ with cardinality $s^{1-6\varepsilon}$ is at most

$$\begin{aligned}\binom{n}{s^{1-6\varepsilon}}\binom{n}{q}\binom{Cn/s^{2-12\varepsilon}}{s-q-s^{1-6\varepsilon}} &\leq n^s\left(\frac{Ce}{s^{2-12\varepsilon}(s-q-s^{1-6\varepsilon})}\right)^{s-q-s^{1-6\varepsilon}} \\ &= n^s\left\{\left(\frac{Ce}{s^{2-12\varepsilon}(1-o(1))s}\right)^{(s-q-s^{1-6\varepsilon})/s}\right\}^s = \left\{n\left(\frac{Ce}{(1-o(1))s^{3-12\varepsilon}}\right)^{1-o(1)}\right\}^s \\ &= \left(\frac{n}{s^{3-12\varepsilon-o(1)}}\right)^s.\end{aligned} \tag{41}$$

By the discussion above, this completes the proof of Proposition 3.10. □

It now remains to prove Lemma 5.1.

Proof of Lemma 5.1. Let $C = 10^9$. We shall show that this choice of C will do. Suppose that T satisfies $\mathcal{P}_{100\lambda,\varepsilon,0}$ and let $A \subset [n]$ with $|A| \geq Cn/|T|^2$ be arbitrary. Recall that, by definition, there exists a graph $G_0 \subset \binom{T}{2}$ satisfying (a) and (b) of Definition 3.9 with $i = 0$ and λ replaced by 100λ. Our goal is to establish a lower bound on $e_{\widetilde{\mathcal{C}}_{G_0}}(A)$ and then apply Proposition 3.6 to obtain the lemma. Let $H = H_T(A, [2n])$ (recall Definition 4.3). Let $c = 10^{-4}$ and let

$$W = \big\{w \in [2n] \,:\, \deg_H(w) \geq c\,|T|^2|A|/n\big\} \tag{42}$$

(the vertices in W have very large degree). Notice that, since $|H| \leq |A|\,|T|$, we have $|W| \leq n/(c|T|)$. Also, set

$$A' = \{a \in A \,:\, |N_H(a) \cap W| < 0.1\,|T|\}. \tag{43}$$

Claim 5.2. *If $|A'| \leq |A|/2$, then the conclusion of the lemma holds. More precisely,*

$$e_{\mathcal{C}_T}(A \setminus A') = \Omega\left(\frac{\delta^2|A|^2|T|^4}{\lambda n(\log n)^3}\right). \tag{44}$$

Proof. We apply Lemma 4.5 to the graph $H' = H_T(A \setminus A', W) \subset H$ with $i = 0$, $B = W$, 100λ in place of λ, and $A \setminus A'$ in place of A. We proceed in steps.

(i) Set $\delta = 0.1$ and notice that $\delta^2 = 0.01 = \alpha_0/100\xi$, and thus δ satisfies the condition of the lemma.

(*ii*) We assumed that T satisfies $\mathcal{P}_{100\lambda,\varepsilon,0}$ and that $|T| \geq n^{1/100}$, and therefore T satisfies the conditions of the lemma, with 100λ in place of λ.

(*iii*) From the definition of A' in (43), it follows that $\deg_H(a) = |N_H(a) \cap W| \geq \delta\,|T|$ for all $a \in A \setminus A'$.

(*iv*) Finally, the average degree of the vertices in W is

$$D = \frac{|H'|}{|W|} \geq \frac{|A|}{2} \cdot \frac{0.1\,|T|}{|W|} \geq \frac{c\,|A|\,|T|^2}{20n}. \tag{45}$$

As $C = 10^9$, $c = 10^{-4}$ and $|A| \geq Cn/|T|^2$, we have $D \geq cC/20 = 5000$.

From Lemma 4.5, with $D = \Omega(|A|\,|T|^2/n)$ as in (45) and $|A| \geq Cn/|T|^2$, we conclude that

$$e_{\mathcal{C}_T}(A \setminus A') = \Omega\bigg(\delta^2 \frac{|A|^3|T|^6}{\lambda n^2(\log n)^3}\bigg) = \Omega\bigg(\frac{\delta^2|A|^2|T|^4}{\lambda n(\log n)^3}\bigg), \tag{46}$$

as required. □

In view of Claim 5.2, let us assume that $|A'| \geq |A|/2$.

Definition 5.3 (Auxiliary graph Aux)**.** *Let* Aux *be a bipartite graph with classes consisting of* A' *and a disjoint copy of* $[3n]$ *as follows. For* $x \in A'$*, the neighbors of* x *in* Aux *are all elements of the form* $y = x + z_1 + z_2$ *for some* $z_1z_2 \in G_0$ *such that* $x + z_1,\ x + z_2 \notin W$*. In other words,* $z_1z_2 \in G_0[T \setminus (W - x)]$*.*

Suppose that a pair of distinct $x,\ x' \in A'$ is connected by a path of length two in Aux. We classify this path as follows:

non-degenerate path: if the path is of the form

$$x,\ x + z_1 + z_2 = x' + z_3 + z_4,\ x' \quad \text{with } (z_1z_2, z_3z_4) \in G_0^2 \text{ and } z_i\text{s distinct;}$$

degenerate path: if the path is of the form

$$x,\ x + z_1 + z = x' + z_2 + z,\ z' \quad \text{with } (z_1z, z_2z) \in G_0^2. \tag{47}$$

Note that the two cases above are exhaustive since elements in an edge of G_0 are necessarily distinct (i.e., G_0 has no loops). Denote by $d(x, x')$ the number of degenerate paths between x and x' and by $p(x, x')$ the total number of 2-paths connecting them.

Note that a non-degenerate path between x, x' corresponds to an ordered pair of edges of G_0 counted by $R_{G_0}(x' - x) = R_{G_0}(x - x')$ (see (20)). Therefore,

$$R_{G_0}(x - x') \geq p(x, x') - d(x, x'). \tag{48}$$

(We have an inequality instead of equality in (48) above because, owing to the definition of Aux, the first edge of the pair must come from $G_0[T \setminus (W - x)]$ and the second edge of the pair from $G_0[T \setminus (W - x')]$, and hence not all pairs counted by $R_{G_0}(x - x')$ yields appropriate 2-paths in Aux.) In order to estimate $e_{\widetilde{\mathcal{C}}_{G_0}}(A') = \sum_{x,x' \in A'} R_{G_0}(x - x')$ we bound the number of degenerate paths and estimate the total number of 2-paths. Here and in what follows, we write $\sum_{x,x' \in A'}$ for the sum over all unordered pairs $\{x, x'\} \subset A$ $(x \neq x')$.

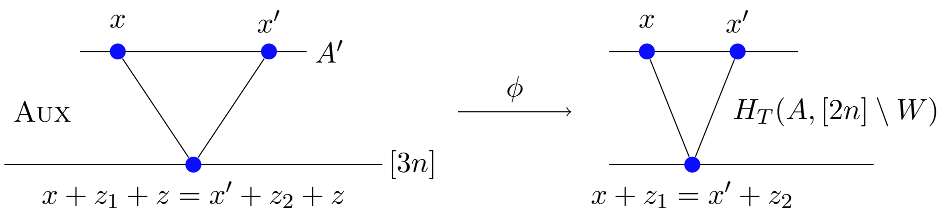

FIGURE 3. Owing to the definition of Aux, we know that $x + z_1 = x' + z_2$ does not belong to W. Consequently, $(x, x + z_1, x')$ is a two-path in $H_T(A, [2n] \setminus W)$, and hence a member of $\mathcal{P}$.

Let $\mathcal{D}$ denote the set of all degenerate paths in Aux, so that $|\mathcal{D}| = \sum_{x,x' \in A'} d(x, x')$. Also, let $\mathcal{P}$ be the set of all paths of length two in $H_T(A', [2n] \setminus W)$ with both endpoints in A'. We will provide an upper bound for $|\mathcal{D}|$ by defining a map $\phi\colon \mathcal{D} \to \mathcal{P}$, estimating $|\mathcal{P}|$ and bounding $|\phi^{-1}(P)|$ for all $P \in \mathcal{P}$.

Claim 5.4. *We have*

$$|\mathcal{D}| \le \frac{c\,|A|^2|T|^4}{n}.$$

Proof. First we define a map $\phi\colon \mathcal{D} \to \mathcal{P}$ as follows. For a degenerate path in Aux between $x, x' \in A'$ as in (47), we infer that x, x' are connected by a path of length two in H (i.e., x, $x + z_1 = x' + z_2$, x'). Given the definition of Aux, we also know that $x + z_1 = x' + z_2 \notin W$. Let ϕ map the degenerate path $x, x + z_1 + z = x' + z_2 + z, x'$ to the path $x, x + z_1 = x' + z_2, x'$, which indeed belongs to $\mathcal{P}$—see Figure 3. Since there are at most $|T|$ choices for $z \in T$ such that both $\{z_1, z\}$ and $\{z_2, z\}$ are edges in G_0, we conclude that $|\phi^{-1}(P)| \le |T|$ for any $P \in \mathcal{P}$. Hence $|\mathcal{D}| \le |T|\,|\mathcal{P}|$.

The cardinality of $\mathcal{P}$ can be bounded from above by $|A'|\,|T| \cdot c\,|T|^2|A|/n$. Indeed, there are at most $|A'|\,|T|$ choices for the first edge of the path and since the path's middle vertex, which is determined by the first edge, is not in W, it follows from (42) that the number of choices for the second edge is at most $c\,|T|^2|A|/n$. Consequently,

$$|\mathcal{D}| \le |T|\,|\mathcal{P}| \le |T| \cdot |A'|\,|T| \cdot \frac{c\,|T|^2|A|}{n} \le \frac{c\,|A|^2|T|^4}{n}. \qquad \square$$

Claim 5.5. *The number of 2-paths between vertices of A' in* Aux*, namely the sum $\sum_{x,x' \in A'} p(x, x')$, is at least*

$$\frac{|A|^2|T|^4}{7 \times 64n}. \tag{49}$$

Proof. The total number of 2-paths between pairs of vertices of A' in Aux can be bounded below by using Jensen's inequality:

$$\sum_{y \in [3n]} \binom{\deg_{\text{Aux}}(y)}{2} \ge 3n \binom{|\text{Aux}|/3n}{2} \ge \frac{|\text{Aux}|^2}{7n}. \tag{50}$$

The claim will follow after we obtain a lower bound for $|\text{Aux}|$.

Note that by construction (see Definition 5.3) and the fact that T is a B_3-set, the degree of any $x \in A'$ in Aux is precisely $e_{G_0}(T \setminus (W - x))$. Since $|N_H(x) \cap W| < 0.1\,|T|$ and $N_H(x) =$

$T + x \subset [2n]$, we have

$$\begin{aligned} |T \setminus (W - x)| &= |T| - |T \cap (W - x)| = |T| - |(T + x) \cap W| \\ &= |T| - |N_H(x) \cap W| > 0.9\,|T|. \end{aligned} \tag{51}$$

Since G_0 satisfies (a) with $i = 0$, we must have

$$\deg_{\mathrm{Aux}}(x) = e_{G_0}\big(T \setminus (W - x)\big) \geq \binom{|T \setminus (W - x)|}{2} - |K(T) \setminus G_0| \geq \binom{0.9\,|T|}{2} - \alpha_0 \binom{|T|}{2} > \frac{|T|^2}{4}. \tag{52}$$

Since $x \in A'$ was arbitrary, we have $|\mathrm{Aux}| \geq |A'||T|^2/4$.

Therefore

$$\sum_{x,x' \in A'} p(x, x') \geq \frac{|\mathrm{Aux}|^2}{7n} \geq \frac{(|A'|\,|T|^2/4)^2}{7n}. \tag{53}$$

Since $|A'| \geq |A|/2$, the claim follows. □

It follows from Claims 5.4 and 5.5, together with (48) and $c = 1/10000$, that

$$e_{\widetilde{\mathcal{C}}_{G_0}}(A) \geq e_{\widetilde{\mathcal{C}}_{G_0}}(A') \overset{(48)}{\geq} \sum_{x,x' \in A'} \big(p(x, x') - d(x, x')\big) \overset{\text{Cl. 5.4 \& 5.5}}{\geq} \frac{|A|^2\,|T|^4}{500n}. \tag{54}$$

Since Q_{G_0} satisfies (b) with $i = 0$, and $100\lambda \geq 1$, we have

$$Q_{G_0} \leq en \exp\big(100\lambda(1 + \log|T|)\big) \leq \exp(200\lambda \log n). \tag{55}$$

It follows by (23) and Proposition 3.6 that

$$e_{\mathcal{C}}(A) \geq \frac{e_{\widetilde{\mathcal{C}}_{G_0}}(A)}{\log Q_{G_0}} \geq \frac{e_{\widetilde{\mathcal{C}}_{G_0}}(A)}{200\lambda \log n} \geq \frac{|A|^2\,|T|^4}{10^5 \lambda n \log n}. \tag{56}$$

Hence Lemma 5.1 is proved. □

5.2. **Sets not containing a large bounded subset.** We now turn to the proof of Proposition 3.11, that is, we enumerate B_3-sets such that all of its (λ, ε)-bounded subsets have fewer than $s^{1-6\varepsilon}$ elements. We shall do this by bounding, for any given (λ, ε)-bounded set T, the number of ways one can extend T to a B_3-set S in such a way that T remains a maximal (λ, ε)-bounded subset of S.

In what follows, we show that extensions preserving a (λ, ε)-bounded set T as maximal must admit certain structural properties that severely restrict the number of possible extensions.

Definition 5.6. *Given a (λ, ε)-bounded set T, let*

$$\widetilde{T} = \big\{x \in [n] \,:\, T \cup \{x\} \text{ is a } B_3\text{-set but not a } (\lambda, \varepsilon)\text{-bounded set}\big\}. \tag{57}$$

Also, for $i \in \{0, 1, \ldots, \lceil 1/\varepsilon \rceil\}$, let

$$\widetilde{T}_i = \big\{x \in \widetilde{T} \,:\, i \text{ is the smallest index such that } T \cup \{x\} \text{ does not satisfy } \mathcal{P}_{\lambda,\varepsilon,i}\big\}. \tag{58}$$

Note that, by definition, if a B_3-set S contains T and T is a maximal (λ, ε)-bounded subset of S, then $S \setminus T \subset \widetilde{T}$. Note that, clearly, the sets $\widetilde{T}_i$ partition $\widetilde{T}$ and

$$\widetilde{T} = \bigcup_{i=0}^{\lceil 1/\varepsilon \rceil} \widetilde{T}_i. \tag{59}$$

The next lemma gives us important information on the sets $\widetilde{T}_i$. The sets B_i, whose existence is asserted in this lemma, will be crucial for us to prove that $\mathcal{C}_T[\widetilde{T}_i]$ satisfies a local density condition, as specified in Corollary 5.11. The B_i will be used in an application of Lemma 4.5 in the proof of Corollary 5.11.

Lemma 5.7. *Let $i \in \{0, 1, \ldots, \lceil 1/\varepsilon \rceil\}$ and suppose that a set T satisfies $\mathcal{P}_{\lambda,\varepsilon,i}$. There exists a set $B_i = B_i(T) \subset [2n]$ with*

$$|B_i| < \frac{e^2(e+1)|T|^4}{\lambda n^{i \cdot \varepsilon}} \tag{60}$$

such that, for every $x \in \widetilde{T}_i$,

$$|(T + x) \cap B_i| \geq \alpha_i \, |T|. \tag{61}$$

Proof. Since T satisfies $\mathcal{P}_{\lambda,\varepsilon,i}$, there exists a graph G_i on the vertex set T that satisfies (a) and (b) of Definition 3.9. Let us fix such a graph G_i for the remainder of the proof of Lemma 5.7. For technical reasons, it will be convenient to introduce the following definition: for each $w \in [2n]$ and $z \in [n]$, set

$$f_i(w, z) = \begin{cases} 1 & \text{if } z = \pm(w - a - b) \text{ for some } \{a, b\} \in G_i, \\ 0 & \text{otherwise.} \end{cases} \tag{62}$$

Also, for each $w \in [2n]$, let

$$U_{i,w} = \sum_{z \in [n]} (\exp R_{G_i}(z)) \, e^2 f_i(w, z). \tag{63}$$

In what follows, we will show that the set B_i defined by

$$B_i = \left\{ w \in [2n] \;:\; U_{i,w} > \frac{\lambda n^{i\varepsilon} Q_{G_i}}{(e+1)|T|(|T|+1)} \right\} \tag{64}$$

satisfies the conclusions of the lemma.

Claim 5.8. *We have $|B_i| < e^2(e+1)|T|^4/(\lambda n^{i \cdot \varepsilon})$.*

Proof. We start by proving the following inequality, which will be used shortly:

$$\text{for any } z \in [n], \text{ we have } \sum_{w \in [2n]} f_i(w, z) \leq 2\, e(G_i). \tag{65}$$

This inequality holds since each edge $\{a, b\}$ of G_i may only contribute to the sum on the left hand side with the two entries $f_i(a + b + z, z)$ and $f_i(a + b - z, z)$. Now observe that

$$\begin{aligned} \sum_{w \in [2n]} U_{i,w} &= \sum_{w \in [2n]} \sum_{z \in [n]} (\exp R_{G_i}(z)) \, e^2 f_i(w, z) \\ &= e^2 \sum_{z \in [n]} \exp R_{G_i}(z) \sum_{w \in [2n]} f_i(w, z) \leq e^2 Q_{G_i} 2\, e(G_i) < e^2 \, |T|(|T| - 1) Q_{G_i}. \end{aligned} \tag{66}$$

On the other hand,

$$\sum_{w\in[2n]} U_{i,w} \geq \sum_{w\in B_i} U_{i,w} \geq |B_i| \frac{\lambda n^{i\cdot\varepsilon} Q_{G_i}}{(e+1)|T|(|T|+1)}, \tag{67}$$

which implies (60), concluding the proof of the claim. □

It remains to prove that for every $x \in \widetilde{T}_i$, we have $|(T+x)\cap B_i| \geq \alpha_i\,|T|$. Fix an arbitrary $x \in \widetilde{T}_i$. For $y \in T$, denote by $G_i \cup \{xy\}$ the graph with vertex set $V(G_i)\cup\{x\}$ and edge set $E(G_i)\cup\{xy\}$. Let

$$D_{i,xy} = Q_{G_i\cup\{xy\}} - Q_{G_i}. \tag{68}$$

Expanding, we obtain

$$D_{i,xy} = \sum_{z\in[n]} \exp\bigl(R_{G_i}(z)\bigr) \underbrace{\Bigl\{\exp\Bigl(R_{G_i\cup\{xy\}}(z) - R_{G_i}(z)\Bigr) - 1\Bigr\}}_{(\ddagger)}. \tag{69}$$

The following claim relates $D_{i,xy}$ and $U_{i,x+y}$.

Claim 5.9. *We have*

$$D_{i,xy} \leq U_{i,x+y}. \tag{70}$$

Proof. Let $w = x + y$. We shall prove the claim by showing that every term in the sum (69) is bounded above by its corresponding term in the sum (63) defining $U_{i,w}$. Let $z \in [n]$ be arbitrary.

Note that any difference between $R_{G_i\cup\{xy\}}(z)$ and $R_{G_i}(z)$ must be either due to a pair (xy, z_1z_2), $z_1z_2 \in G_i$, satisfying

$$z = (x+y) - (z_1+z_2) = w - (z_1+z_2), \tag{71}$$

or due to a pair (z_1z_2, xy) such that $z = (z_1+z_2) - w$, where, in both cases, we require $\{x,y\}\cap\{z_1,z_2\} = \emptyset$. If $f_i(w,z) = 0$ then there are no such pairs and we must have $R_{G_i\cup\{xy\}}(z) = R_{G_i}(z)$. In this case, the term $(\ddagger)$ in (69) is 0.

Since T is a B_3-set, there can be at most one edge $\{a,b\} \in G_i$ such that $z = w - a - b$ and at most one edge $\{a',b'\} \in G_i$ for which $-z = (x+y) - a' - b'$. Therefore, we always have $R_{G_i\cup\{xy\}}(z) \leq R_{G_i}(z) + 2$. Consequently, in this case

$$R_{G_i}(z) \leq R_{G_i\cup\{xy\}}(z) \leq R_{G_i}(z) + 2. \tag{72}$$

In particular, the term $(\ddagger)$ in (69) is 0, $e-1$ or e^2-1.

To summarize, regardless of whether $f_i(w,z)$ is 0 or 1, we have

$$(\ddagger) \leq e^2 f_i(w,z). \tag{73}$$

Therefore,

$$D_{i,xy} \leq \sum_{z\in[n]} \left(\exp R_{G_i}(z)\right) e^2 f_i(w,z) = U_{i,w} = U_{i,x+y}.$$ □

Next we show that the effect in the moment function caused by adding multiple edges incident to x to the graph G_i is essentially the sum of the effects of each edge xy being added.

Claim 5.10. *For any $Y \subset T$, letting $G_i' = G_i \cup \{xy \,:\, y \in Y\}$, we have*

$$Q_{G_i'} - Q_{G_i} \le (e+1) \sum_{y \in Y} D_{i,xy}. \tag{74}$$

Proof. Since $G_i' \setminus G_i = \{xy \colon y \in Y\}$ contains only edges incident to x, the difference $R_{G_i'}(z) - R_{G_i}(z)$ equals the number of solutions to $z = \pm(x + y - a - b)$ for some $y \in Y$ and $\{a, b\} \in G_i$ with $\{x, y\} \cap \{a, b\} = \emptyset$. Let us bound the number of such solutions. To this end, suppose that

$$z = x + y - a - b = x + y' - a' - b' \tag{75}$$

for y, $y' \in Y$ and $\{a, b\}$, $\{a', b'\} \in G_i$ with $\{x, y\} \cap \{a, b\} = \emptyset$ and $\{x, y'\} \cap \{a', b'\} = \emptyset$. Then $y + a' + b' = y' + a + b$ and, since these elements come from the B_3-set T, we conclude that $\{y, a', b'\} = \{y', a, b\}$. It follows that $y = y'$ and $\{a, b\} = \{a', b'\}$. Hence, there is at most one solution to $z = x + y - a - b$ and at most one solution to $-z = x + y - a - b$. Consequently, $R_{G_i'}(z) - R_{G_i}(z) \le 2$ and

$$\Delta_z := \exp R_{G_i'}(z) - \exp R_{G_i}(z) \le (\exp R_{G_i}(z))\,(e^2 - 1). \tag{76}$$

Moreover, $R_{G_i'}(z) > R_{G_i}(z)$ only if $R_{G_i \cup \{xy^*\}}(z) > R_{G_i}(z)$ for some $y^* \in Y$, and, therefore, in that case, we have

$$\begin{aligned}
\Delta_z &\le (e+1) \cdot \exp\big(R_{G_i}(z)\big) \cdot (e-1) \\
&= (e+1)\big\{\exp\big(R_{G_i}(z) + 1\big) - \exp R_{G_i}(z)\big\} \\
&\le (e+1)\big\{\exp R_{G_i \cup \{xy^*\}}(z) - \exp R_{G_i}(z)\big\} \\
&\le (e+1) \cdot \sum_{y \in Y} \big\{\exp R_{G_i \cup \{xy\}}(z) - \exp R_{G_i}(z)\big\}.
\end{aligned} \tag{77}$$

Note that if $R_{G_i'}(z) = R_{G_i}(z)$, then both the left-hand and the right-hand side of (77) are zero. In other words,

$$\Delta_z \le (e+1) \sum_{y \in Y} \Big(\exp\big(R_{G_i \cup \{xy\}}(z)\big) - \exp\big(R_{G_i}(z)\big)\Big) \tag{78}$$

holds for all $z \in [n]$. Consequently,

$$\begin{aligned}
Q_{G_i'} - Q_{G_i} &= \sum_{z \in [n]} \Delta_z \\
&\le \sum_{z \in [n]} (e+1) \sum_{y \in Y} \Big(\exp\big(R_{G_i \cup \{xy\}}(z)\big) - \exp\big(R_{G_i}(z)\big)\Big) \\
&= (e+1) \sum_{y \in Y} D_{i,xy}.
\end{aligned}$$

□

Setting

$$Y = \{y \in T \,:\, x + y \in [2n] \setminus B_i\} \tag{79}$$

in Claim 5.10 yields that $G_i' = G_i \cup \{xy \,:\, y \in Y\}$ satisfies

$$\begin{aligned}
Q_{G_i'} &\leq Q_{G_i} + (e+1)\sum_{y\in Y} D_{i,xy} \\
&\overset{(70)}{\leq} Q_{G_i} + (e+1)\sum_{y\in Y} U_{i,x+y} \\
&\overset{(64)+(79)}{\leq} Q_{G_i} + (e+1)\sum_{y\in Y} \frac{\lambda n^{i\varepsilon} Q_{G_i}}{(e+1)|T|(|T|+1)} \\
&\leq Q_{G_i}\left(1 + \frac{\lambda n^{i\varepsilon}}{|T|+1}\right) \\
&\overset{\text{Def. 3.9}(b)}{\leq} en \exp\left(\lambda n^{i\varepsilon} \sum_{j=1}^{|T|} \frac{1}{j}\right) \exp\left(\frac{\lambda n^{i\varepsilon}}{|T|+1}\right) \\
&\leq en \exp\left(\lambda n^{i\varepsilon} \sum_{j=1}^{|T|+1} \frac{1}{j}\right),
\end{aligned} \tag{80}$$

which means that G_i' satisfies (b) of Definition 3.9 with $T \cup \{x\}$ in place of T.

Since our assumption that $x \in \widetilde{T}_i$ implies that $T \cup \{x\}$ does not satisfy $\mathcal{P}_{\lambda,\varepsilon,i}$, the graph G_i' must fail (a) of Definition 3.9. Thus

$$e(G_i') = e(G_i) + |Y| < \big(1 - \alpha_i\big)\binom{|T|+1}{2} \tag{81}$$

and, as G_i satisfies (a), we conclude that

$$|Y| < \big(1-\alpha_i\big)\left\{\binom{|T|+1}{2} - \binom{|T|}{2}\right\} = \big(1-\alpha_i\big)|T|. \tag{82}$$

From the definition of Y in (79) and the fact that $T \subset [n]$, $x \in [n]$, it follows that

$$Y = T \setminus (B_i - x). \tag{83}$$

Hence

$$|(T+x) \cap B_i| = |T \cap (B_i - x)| = |T| - |Y| \geq \alpha_i\,|T|. \tag{84}$$

Since $x \in \widetilde{T}_i$ was arbitrary, the proof of Lemma 5.7 is complete. □

Recall Definition 4.3 from Section 4.2. Lemma 5.7 implies that for every $i \in \{0, 1, \ldots, \lceil 1/\varepsilon \rceil\}$, there exists a $B = B_i$ with $|B| = O(|T|^4/(\lambda n^{i\varepsilon}))$ such that for every $A \subset \widetilde{T}_i$,

$$|H_T(A, B)| \geq \alpha_i\,|T|\,|A|. \tag{85}$$

Together with Lemma 4.5, this yields the following corollary.

Corollary 5.11. *Suppose that T is a (λ, ε)-bounded set with cardinality at least $n^{1/100}$ and less than $s^{1-6\varepsilon}$. For every $i \in \{0, \ldots, \lceil 1/\varepsilon \rceil - 1\}$ and any set $A \subset \widetilde{T}_i$ with*

$$|A| \geq s^{-2+8\varepsilon} n, \tag{86}$$

we have

$$e_{\mathcal{C}_T}(A) = \Omega\biggl(\frac{\alpha_i^3|A|^2}{|T|\,n^{\varepsilon}(\log n)^3}\biggr). \tag{87}$$

Proof. Fix $i \in \{0,\ldots,\lceil 1/\varepsilon\rceil - 1\}$ and let $B = B_i(T)$ be the set from Lemma 5.7. In particular, $|B| = O\bigl(|T|^4/(\lambda n^{i\varepsilon})\bigr)$. Also let

$$\delta = \alpha_i. \tag{88}$$

We now show that the the graph $H = H_T(A,B)$ satisfies all the conditions of Lemma 4.5 with $i+1$ in place of i. We proceed in steps. From (24), we have

$$\delta^2 = \alpha_i^2 = \xi^{2(2^{i+1}-1)} = \xi^{2^{i+2}-2} = \frac{\alpha_{i+1}}{\xi} > \frac{\alpha_{i+1}}{100\xi}. \tag{89}$$

Our assumptions that $i < \lceil 1/\varepsilon\rceil$ and that T is (λ,ε)-bounded imply that T satisfies $\mathcal{P}_{\lambda,\varepsilon,i+1}$ (see Definition 3.9). Moreover, we also assume that $|T| \ge n^{1/100}$. Lemma 5.7 implies that every $a \in A \subset \widetilde{T}_i$ satisfies

$$\deg_H(a) = |(T+a)\cap B| \ge \delta\,|T|. \tag{90}$$

Finally, recalling that $\lambda = s^{5-25\varepsilon}/n \ge 1$ (see (26)) and that $s \ge n^{1/(5-25\varepsilon)}$, we have that the average degree D of the vertices in B satisfies

$$D = \frac{|H|}{|B|} \ge \frac{\delta|A||T|}{|B|} \ge \delta\frac{s^{-2+8\varepsilon}n\,|T|}{|B|} = \Omega\biggl(\delta\frac{s^{-2+8\varepsilon}n\lambda n^{i\varepsilon}}{|T|^3}\biggr) = \Omega\biggl(\delta\frac{s^{3-17\varepsilon}n^{i\varepsilon}}{|T|^3}\biggr) \ge \delta s^{\varepsilon} \ge 5000, \tag{91}$$

where we used that $|T| < s^{1-6\varepsilon}$ and that n is large. By Lemma 4.5 (with $i+1$ in place of i), we have

$$\begin{aligned} e_{\mathcal{C}_T}(A) &= \Omega\biggl(\delta^2\frac{|A|\,|T|^2}{\lambda n^{(i+1)\varepsilon}(\log n)^3}\Bigl(\frac{|H|}{|B|}\Bigr)^2\biggr) \\ &\overset{(91)}{=} \Omega\biggl(\delta^2\frac{|A|\,|T|^2}{\lambda n^{(i+1)\varepsilon}(\log n)^3}\cdot D\frac{\delta\,|T|\,|A|}{|B|}\biggr) \\ &= \Omega\biggl(\delta^3\frac{D\,|A|^2|T|^3}{|B|\lambda n^{(i+1)\varepsilon}(\log n)^3}\biggr) \\ &\overset{(88)}{=} \Omega\biggl(\alpha_i^3\frac{D}{|T|\,n^{\varepsilon}(\log n)^3}\frac{|T|^4}{|B|\lambda n^{i\varepsilon}}|A|^2\biggr). \end{aligned} \tag{92}$$

Since $|B| = O\bigl(|T|^4/(\lambda n^{i\varepsilon})\bigr)$, the term $|T|^4/|B|\lambda n^{i\varepsilon}$ on the right hand side of (92) can be replaced by 1. Hence, from (91) and (92) it follows that (87) holds and the corollary is proved. $\square$

Let $s \in [n^{1/(5-25\varepsilon)}, 3n^{1/3}]$ be fixed and $t < s^{1-6\varepsilon}$ be an integer. In order to prove Proposition 3.11, we will estimate how many B_3-sets have a maximal (λ,ε)-bounded set T with cardinality t. As we observed above, if T is a maximal (λ,ε)-bounded subset of S, then $S\setminus T \subset \widetilde{T}$ (recall Definition 5.6). Therefore, it suffices to prove an upper bound for the number of B_3-sets S satisfying $S\setminus T \subset \widetilde{T}$. For that we shall apply Lemma 3.3 to the graph $\mathcal{C}_T[\widetilde{T}]$. Therefore we have to show that $\mathcal{C}_T[\widetilde{T}]$ satisfies the conditions of the lemma. We need the following claim.

Claim 5.12. *The set $\widetilde{T}_{\lceil 1/\varepsilon\rceil}$ is empty.*

Proof. Recall that $\widetilde{T}_{\lceil 1/\varepsilon \rceil}$ is the set of all x such that $T \cup \{x\}$ is a B_3-set and there is no graph $G_{\lceil 1/\varepsilon \rceil} \subset \binom{T\cup\{x\}}{2}$ satisfying both conditions (a) and (b) of Definition 3.9 with $i = \lceil 1/\varepsilon \rceil$ and $T \cup \{x\}$ in place of T. The claim will follow after we show that for any x such that $T \cup \{x\}$ is a B_3-set, if we let $G_{\lceil 1/\varepsilon \rceil}$ be the complete graph $K(T \cup \{x\})$ on $T \cup \{x\}$, then conditions (a) and (b) hold. Condition (a) follows immediately for $G_{\lceil 1/\varepsilon \rceil} = K(T \cup \{x\})$. As for (b), observe first that, as a little thought shows, we have $R_{G_{\lceil 1/\varepsilon \rceil}}(z) \leq |T| + 1$ for all z since $T \cup \{x\}$ is a B_3-set. Hence

$$Q_{G_{\lceil 1/\varepsilon \rceil}} \leq n e^{|T|+1} \leq en \exp\bigg(\lambda n^{\lceil 1/\varepsilon \rceil \varepsilon} \sum_{j=1}^{|T|+1} \frac{1}{j}\bigg), \tag{93}$$

which establishes (b) and proves the claim. □

The following claim is a simple consequence of Corollary 5.11 and Claim 5.12.

Claim 5.13. *Suppose that T is a (λ, ε)-bounded set with cardinality at least $n^{1/100}$ and less than $s^{1-6\varepsilon}$. For any set $A \subset \widetilde{T}$ with*

$$|A| \geq \lceil 1/\varepsilon \rceil s^{-2+8\varepsilon} n, \tag{94}$$

we have

$$e_{\mathcal{C}_T}(A) = \Omega\bigg(\frac{\alpha^3_{\lceil 1/\varepsilon \rceil - 1} |A|^2}{|T|\, n^{\varepsilon} (\log n)^3}\bigg). \tag{95}$$

Proof. Let $A \subset \widetilde{T}$ be as in the statement of the claim. By Claim 5.12 and the pigeonhole principle, there must be some $i \in \{0, 1, \ldots, \lceil 1/\varepsilon \rceil - 1\}$ such that $|A \cap \widetilde{T}_i| \geq |A| / \lceil 1/\varepsilon \rceil \geq s^{-2+8\varepsilon} n$. Applying Corollary 5.11 to $A \cap \widetilde{T}_i$ in place of A yields that

$$e_{\mathcal{C}_T}(A) = \Omega\bigg(\frac{\alpha_i^3 |A \cap \widetilde{T}_i|^2}{|T|\, n^{\varepsilon} (\log n)^3}\bigg). \tag{96}$$

Since $\alpha_i \geq \alpha_{\lceil 1/\varepsilon \rceil - 1}$ and $|A \cap \widetilde{T}_i| = \Theta(|A|)$, the claim follows. □

We are finally ready to prove Proposition 3.11.

Proof of Proposition 3.11. In view of Claim 5.13, the graph $\mathcal{C}_T[\widetilde{T}]$ satisfies (16) for

$$\beta = \Omega\bigg(\frac{\alpha^3_{\lceil 1/\varepsilon \rceil - 1}}{|T|\, n^{\varepsilon} (\log n)^3}\bigg) \quad \text{and} \quad R = \gamma n = \lceil 1/\varepsilon \rceil s^{-2+8\varepsilon} n \gg 1. \tag{97}$$

Set $q = (\log n)/\beta$ and note that (18) is satisfied as well. Moreover, since the assumptions of Proposition 3.11 require that $s > n^{1/5}$ and $|T| \leq s^{1-6\varepsilon}$, we have

$$q = O\bigg(|T|\, n^{\varepsilon} \frac{(\log n)^4}{\alpha^3_{\lceil 1/\varepsilon \rceil - 1}}\bigg) \overset{(24)}{=} O\Big(s^{1-6\varepsilon} s^{5\varepsilon} (\log n)^{32^{\lceil 1/\varepsilon \rceil}+1}\Big) = o(s). \tag{98}$$

From Lemma 3.3 we conclude that the number of extensions of T into a B_3-set of size s such that T is a maximal (λ, ε)-bounded subset is at most

$$\binom{n}{q}\binom{\lceil 1/\varepsilon\rceil s^{-2+8\varepsilon}n}{s-q-|T|} \le n^{s-|T|}\left(\frac{\lceil 1/\varepsilon\rceil e}{s^{2-8\varepsilon}(s-q-|T|)}\right)^{s-q-|T|} \le n^{s-|T|}\left(\frac{1}{s^{3-8\varepsilon-o(1)}}\right)^{s}. \tag{99}$$

In view of Lemma 4.1, a maximum (λ, ε)-bounded subset of a B_3-set of cardinality s always contains at least $s^{1/7}$ elements, hence we can assume, without loss of generality, that

$$n^{1/100} \ll s^{1/7} \le t = |T| < s^{1-6\varepsilon}.$$

In particular, considering all possible choices of seed set T, the number of B_3-sets that do not contain any (λ, ε)-bounded subset of size larger than $s^{1-6\varepsilon}$ is at most

$$\sum_{t=s^{1/7}}^{s^{1-6\varepsilon}} \binom{n}{t} n^{s-t}\left(\frac{1}{s^{3-8\varepsilon-o(1)}}\right)^{s} \le \left(\frac{n}{s^{3-8\varepsilon-o(1)}}\right)^{s}.$$

This completes the proof of Proposition 3.11. □

6. Proof of Lemma 4.5

Fix an integer $i \ge 0$ and real numbers $\varepsilon > 0$, $\lambda \ge 1$, $D \ge 5000$ and $\delta \in (0,1]$ satisfying $\delta^2 \ge \alpha_i/100\xi$. Suppose that a set $T \subset [n]$ with at least $n^{1/100}$ elements satisfies $\mathcal{P}_{\lambda,\varepsilon,i}$. Moreover, suppose that $A \subset [n]$ and $B \subset [2n]$ are such that the graph $H = H_T(A, B)$ satisfies the two conditions from the statement of the lemma. The fact that T satisfies $\mathcal{P}_{\lambda,\varepsilon,i}$ means that, in particular, we may choose a graph G_i on the vertex set T that satisfies (a) and (b) of Definition 3.9.

Definition 6.1 (Special paths)**.** *A 4-path (a, b, a', b', a'') in H is said to be a G_i-special path, or simply a special path, if*

(a) a, a', $a'' \in A$ and b, $b' \in B$,
(b) $\{b - a, b' - a'\}$ and $\{b - a', b' - a''\}$ are edges of G_i, and
(c) the differences $b - a$, $b' - a'$, $b - a'$, and $b' - a''$ are all distinct.

Note that a 4-path (a, b, a', b', a'') between a and $a'' \in A$ is special if, letting $z_1 = b-a$, $z_2 = b'-a'$, $z_3 = b - a'$, and $z_4 = b' - a''$, we have

$$(z_1z_2, z_3z_4) \in G_i^2, \;\; a'' - a = (z_1 + z_2) - (z_3 + z_4) \;\text{ and the } z_i\text{s are all distinct.} \tag{100}$$

We claim that for any a, $a'' \in A$, the number of special paths from a to a'' is at most $4R_{G_i}(a'' - a)$. Indeed, if an ordered 4-tuple (z_1, z_2, z_3, z_4) is a solution to (100), then the sequence of elements

$$a,\; b := a + z_1,\; a' := a + z_1 - z_3,\; b' := a + z_1 - z_3 + z_2,\; a'' = a + z_1 - z_3 + z_2 - z_4 \tag{101}$$

forms a special path in H provided that $a' \in A$ and b, $b' \in B$. Any solution to (100) remains a solution after swapping z_1 with z_2 or z_3 with z_4. Therefore, it follows from the definition of R_{G_i} (see (20)) that the number of solutions to (100) is exactly $4R_{G_i}(a'' - a)$. (For completeness,

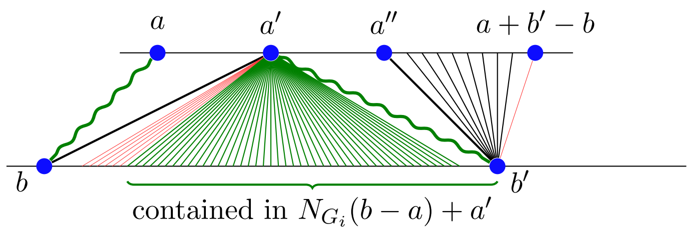

FIGURE 4. Counting semi-special paths extending $P = (a, b, a')$ in the graph H' from Claim 6.3. The first two edges are determined by P and the third edge $\{a', b'\}$ must be such that $\{b - a, b' - a'\} \in G_i$. In view of the properties of H', most of the H'-neighbors of a' produce extensions of P to a semi-special path. Note that the fourth edge may be any edge incident to b' except for $\{a', b'\}$ and possibly $\{b', a + b' - b\}$, $\{b', a' + b' - b\}$ and $\{b', a\}$. For instance, if $a + b' - b$ is a neighbor of b', then it cannot be used to produce a semi-special path since the difference $b' - (a + b' - b) = b - a$ would repeat the difference of the first edge $\{a, b\}$.

we remark that not all solutions need to define paths in H since A and B are just subsets of $[n]$ and $[2n]$.) We conclude that the total number N of special paths in H satisfies

$$N = O\big(|\widetilde{\mathcal{C}}_{G_i}|\big) \tag{102}$$

(see Definition 3.5). Recalling (23), given that G_i satisfies (b) of Definition 3.9, we have

$$4R_{G_i}(a'' - a) \leq 4 \log Q_{G_i} \leq 4\lambda n^{i\varepsilon} \sum_{j=1}^{|T|} \frac{1}{j} \leq 4\lambda n^{i\varepsilon} \log n. \tag{103}$$

In view of Proposition 3.6, inequalities (102) and (103) tell us that Lemma 4.5 will be proved if we establish the following claim.

Claim 6.2. *The total number N of special paths satisfies*

$$N = \Omega\bigg(\frac{\delta^2 |A|\, D^2 |T|^2}{(\log n)^2}\bigg). \tag{104}$$

In order to prove Claim 6.2, we will first construct a subgraph $H' \subset H$ satisfying certain properties that will enable us to estimate the number of special paths N in H.

Claim 6.3. *There exists $d \geq D/16$ and $H' \subset H$ with vertex classes $A' \subset A$ and $B' \subset B$ such that*

- (i) $\deg_{G_i}(b - a) \geq \big(1 - 4\alpha_i/\delta\big)|T|$ *for every* $(a, b) \in H'$;
- (ii) $|H'| \geq |H|/8 \log n$;
- (iii) $\deg_{H'}(a) \geq \delta|T|/16 \log n$ *for every* $a \in A'$;
- (iv) $d \leq \deg_{H'}(b) \leq 12d$ *for every* $b \in B'$.

We postpone the proof of Claim 6.3 and now establish Claim 6.2.

Proof of Claim 6.2. Let $P = (a, b, a')$ be an arbitrary path of length two in the graph H' obtained from Claim 6.3, with $a, a' \in A'$ and $b \in B'$. Consider all possible extensions of this path to a path of length four, say (a, b, a', b', a'') with the condition that the differences

$$b-a,\ b-a',\ b'-a',\ b'-a'' \tag{105}$$

are all distinct and, moreover $\{b-a, b'-a'\} \in G_i$. Call such (oriented) paths *semi-special*. Note that if both $P^{\rightarrow} = (a, b, a', b', a'')$ and $P^{\leftarrow} = (a'', b', a', b, a)$ are semi-special, then we must have both

$$\{b-a, b'-a'\},\ \{b'-a'', b-a'\} \in G_i, \tag{106}$$

and the differences $b-a$, $b-a'$, $b'-a'$, $b'-a''$ are all distinct. This means that the paths $P^{\rightarrow}$ and $P^{\leftarrow}$ are in fact special (recall Definition 6.1). We shall later use this simple fact.

Since H' satisfies (iii), we have $\deg_{H'}(a') \geq \delta|T|/16\log n$. Moreover, by condition (i), we have $\deg_{G_i}(b-a) \geq \big(1-4\alpha_i/\delta\big)|T|$. As we require that $\delta^2 \geq \alpha_i/100\xi$, it follows that the number of non-neighbors of $b-a$ in G_i is at most

$$\frac{4\alpha_i}{\delta}|T| \leq 400\xi\delta\,|T| \overset{(24)}{\leq} \frac{\deg_{H'}(a')}{150}. \tag{107}$$

Consequently, at least 99.3% of the neighbors b' of a' in H' are such that $\{b-a, b'-a'\} \in G_i$. Let

$$X = \big\{b' \in N_{H'}(a') \setminus \{b\}\ :\ \{b-a, b'-a'\} \in G_i\big\} \tag{108}$$

and $X^c = N_{H'}(a') \setminus X$. Note that $|X| \geq 0.993\,|N_{H'}(a')| - 1 \geq 0.99|N_{H'}(a')|$. For each $b' \in X$, we have at least $\deg_{H'}(b') - 4 \geq d-4$ possible choices for $a'' \in N_{H'}(b')$ that produce a semi-special path, namely, the only requirement is that $b'-a''$ must be different from the other three differences and a'' cannot coincide with a (in fact, one sees that this last condition is automatically satisfied, if one recalls that T is a B_3-set). See Figure 4 for an illustration.

From the discussion above, the number N_P of semi-special paths that start with P satisfies

$$N_P \geq \sum_{b'\in X}(\deg_{H'}(b') - 4) \geq \left(1-\frac{4}{d}\right)\sum_{b'\in X}\deg_{H'}(b') \geq 0.98\sum_{b'\in X}\deg_{H'}(b'), \tag{109}$$

where in the last inequality we used the fact that $d \geq D/16 > 200$.

On the other hand, the total number of 4-paths starting with P is at most

$$\sum_{b'\in N_{H'}(a')}\deg_{H'}(b') = \sum_{b'\in X}\deg_{H'}(b') + \sum_{b'\in X^c}\deg_{H'}(b'). \tag{110}$$

Since the degrees in B' are all in $[d, 12d]$, we get

$$\sum_{b'\in X^c}\deg_{H'}(b') \leq 12d\,|X^c| \leq \frac{12}{99}d\,|X| \leq \frac{12}{99}\sum_{b'\in X}\deg_{H'}(b'). \tag{111}$$

Hence, the total number of 4-paths starting with $P = (a, b, a')$ is bounded from above by

$$\left(1+\frac{12}{99}\right)\sum_{b'\in X}\deg_{H'}(b') \overset{(109)}{\leq} \left(1+\frac{12}{99}\right)\frac{100}{98}N_P < \frac{4}{3}N_P. \tag{112}$$

Let N_4 be the *total* number of paths in H' of length 4 starting and ending in A'. We proved above that the number N_P of semi-special paths that start with P corresponds to more than $3/4$ of the total number of 4-paths that starting with P. Since our argument holds for every P, we conclude that there are more than $3N_4/4$ semi-special paths in H'. Considering the involution that takes 4-paths $P = (a, b, a', b', a'')$ to their reverse $P^{\leftarrow} = (a'', b', a', b, a)$, we see that more than $1/2$ of the 4-paths in H' starting and ending in A' are semi-special in *both* directions, and thus are special. That is, there are more than $N_4/2$ special paths in H'. Finally, we estimate N_4 by first picking an edge $ab \in H'$, then picking a neighbor $a' \neq a$ of b, and so on. This yields

$$N_4 \geq \frac{|H|}{8\log n}(d-1)\left(\frac{\delta|T|}{16\log n} - 1\right)(d-2), \tag{113}$$

whence the claim follows. □

It only remains to prove Claim 6.3.

Proof of Claim 6.3. This proof will be divided into three simple steps. First we define a set L of low degree vertices in G_i and show that this set is quite small. In the second step, we partition the class B according to the degrees of the vertices in $H_{T\setminus L}(A, B)$ and select one part B_j that is incident to a good fraction of the edges while at the same time B_j has vertices with roughly the same degree. Finally, we delete the vertices of low degree in $H_{T\setminus L}(A, B_j)$ to obtain the desired graph.

We assume that n, and therefore $|T|$, are sufficiently large for the calculations that follow to hold. Let

$$L = \big\{x \in T \,:\, \deg_{G_i}(x) < \big(1 - 4\alpha_i/\delta\big)|T|\big\}. \tag{114}$$

Note that

$$2\,e(G_i) = \sum_{x\in T} \deg_{G_i}(x) \leq |L|\big(1 - 4\alpha_i/\delta\big)|T| + (|T| - |L|)|T| = |T|^2 - \frac{4\alpha_i}{\delta}|T|\,|L|. \tag{115}$$

On the other hand, it follows from the assumption on G_i (see Definition 3.9(a)) that

$$2\,e(G_i) \geq \big(1 - \alpha_i\big)|T|(|T| - 1) \geq |T|^2 - 2\alpha_i\,|T|^2. \tag{116}$$

A straightforward comparison of the two inequalities above yields the following (non-tight) bound

$$|L| \leq \frac{\delta}{2}|T| \leq \frac{|H|}{2|A|}, \tag{117}$$

and hence $|L||A| \leq |H|/2$. Let $H^* = H_{T\setminus L}(A, B) \subset H$ be the subgraph of H consisting of all the edges $ab \in H$ ($a \in A$, $b \in B$) such that $b - a \in T \setminus L$. It follows from (117) and the assumption of the lemma that

$$e(H^*) = e(H) - |L|\,|A| \geq e(H)/2. \tag{118}$$

Since the graph H' that we construct in what follows is a subgraph of H^*, it will satisfy (i).

Let $I_0 = [0, D/4)$, and $I_j = [(D/4)e^{j-1}, (D/4)e^j)$ for $j \geq 1$. For $j \geq 0$, let

$$B_j = \{b \in B \,:\, \deg_{H^*}(b) \in I_j\}. \tag{119}$$

Note that $B_j = \emptyset$ for $j \geq \log|T|$ since the maximum degree is at most $|T|$. Moreover, the number of edges incident to B_0 is at most $|B|\,D/4 = e(H)/4$. In particular, by the pigeonhole principle, there exists $1 \le j \le \log|T|$ such that there are at least

$$\frac{e(H^*) - e(H)/4}{\log|T|} \ge \frac{e(H)}{4\log n} \tag{120}$$

edges of H^* incident to B_j.

Set $d = (D/16)e^{j-1}$, and $\widehat{H} = H^*[A \cup B_j]$. Since we assume that H satisfies (I), it follows from (120) that

$$e(\widehat{H}) \ge \frac{\delta\,|T|\,|A|}{4\log n}. \tag{121}$$

In particular, the average degree of vertices of A in $\widehat{H}$ is at least $\delta|T|/4\log n$ and the degrees of vertices in B_j are all in $[4d, 12d]$. While there exists a vertex from A with degree smaller than $\frac{\delta}{16\log n}|T|$, or a vertex from B with degree less than d, remove the vertex from the graph $\widehat{H}$, together with all the incident edges. The number of edges deleted by this procedure is at most

$$|A|\,\frac{\delta}{16\log n}|T| + |B|\,d \le \frac{1}{2}e(\widehat{H}). \tag{122}$$

Hence, at least $e(\widehat{H})/2 \ge e(H)/8\log n$ edges remain after the deletion procedure above. Let H' be the graph obtained after the procedure and observe that it satisfies (*ii*), (*iii*), and (*iv*). □

7. Concluding remarks

In this whole paper, we considered δ to be an arbitrary, but fixed positive real number. Our argument allows one to take some function $\delta = \delta(n)$ with $\delta \to 0$ as $n \to \infty$. Here, we opted for simplicity and did not attempt to optimize the argument to obtain the smallest possible $\delta = \delta(n)$.

We close by restating our conjectured answer (see [4]) to the problem addressed in Section 2. We believe that Theorem 2.1, concerning the cardinality of the largest B_3-sets contained in the random sets $[n]_m$, is a particular case of a more general result.

Conjecture 7.1. *Let $h \ge 2$ be an integer. Suppose $0 \le a \le 1$ is a fixed constant and $m = m(n) = (1 + o(1))n^a$. Then, asymptotically almost surely, we have $F_h([n]_m) = n^{b+o(1)}$, where $b = b(a)$ is given by*

$$b(a) = \begin{cases} a & \text{for } 0 \le a \le 1/(2h-1), \\ 1/(2h-1) & \text{for } 1/(2h-1) \le a \le h/(2h-1), \\ a/h & \text{for } h/(2h-1) \le a \le 1. \end{cases} \tag{123}$$

The fact that $b(a)$ is at least as large as stated in (123) is proved in [4]. On the other hand, a routine argument shows that, if true, Conjecture 1.3 implies the upper bound for $b(a)$ conjectured in (123). The case $h = 2$ of Conjecture 7.1 is proved in [10, 11] and we established the case $h = 3$ in this paper.

References

1. R. C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helv. **37** (1962/1963), 141–147.

2. 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.
3. 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.
4. D. Dellamonica, Jr., Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij, *On the number of B_h-sets*, submitted.
5. P. Erdős, *On a problem of Sidon in additive number theory and on some related problems. Addendum*, J. London Math. Soc. **19** (1944), 208.
6. 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.
7. B. Green, *The number of squares and $B_h[g]$ sets*, Acta Arith. **100** (2001), no. 4, 365–390.
8. H. Halberstam and K. F. Roth, *Sequences*, second ed., Springer-Verlag, New York, 1983.
9. D. J. Kleitman and K. J. Winston, *On the number of graphs without* 4*-cycles*, Discrete Math. **41** (1982), no. 2, 167–172.
10. Y. Kohayakawa, S. Lee, and V. Rödl, *The maximum size of a Sidon set contained in a sparse random set of integers*, Proceedings of the 22nd Annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2011) (Philadelphia, PA, USA), Society for Industrial and Applied Mathematics, 2011, pp. 159–171.
11. Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij, *The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers*, Random Structures Algorithms, to appear.
12. K. O'Bryant, *A complete annotated bibliography of work related to Sidon sequences*, Electron. J. Combin. (2004), Dynamic surveys 11, 39 pp. (electronic).
13. D. Saxton and A. Thomason, *Hypergraph containers*, arXiv:1204.6595 (2012).
14. S. Sidon, *Ein Satz über trigonometrische Polynome und seine Anwendung in der Theorie der Fourier-Reihen*, Math. Ann. **106** (1932), no. 1, 536–539.
15. J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Trans. Amer. Math. Soc. **43** (1938), no. 3, 377–385.

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 Mathematical Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea (S. J. Lee)

E-mail address: sjlee242@gmail.com

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel, and Trinity College, Cambridge CB2 1TQ, UK (W. Samotij)

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