

1 **THE NUMBER OF B_3 -SETS OF A GIVEN CARDINALITY**

2 DOMINGOS DELLAMONICA JR., YOSHIHARU KOHAYAKAWA, SANG JUNE LEE, VOJTĚCH RÖDL,
3 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 $[n] = \{1, \dots, n\}$. 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].

4 1. INTRODUCTION

5 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
6 sequence $a_1 \leq \dots \leq a_h$ satisfying $z = a_1 + \dots + a_h$ if we require that $a_i \in S$ for every $i = 1, \dots, h$.
7 The study of B_h -sets goes back to Sidon [14], who asked how large B_2 -sets, or *Sidon sets*, can be
8 if one imposes that they should be subsets of $[n] = \{1, \dots, n\}$. Let

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

9 In the case addressed by Sidon, that is, for $h = 2$, results of Chowla, Erdős, Singer, and Turán [3,
10 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
11 understood. Bose and Chowla [1] showed that $F_h(n) \geq (1 + o(1))n^{1/h}$ for $h \geq 3$, while an easy
12 argument gives that, for every $h \geq 3$ and large n ,

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

13 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
14 general h , successively better bounds of the form $F_h(n) \leq c_h n^{1/h}$ have been obtained. The latest
15 bounds are due to Green [7], who proved that

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

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

Date: 2014/02/07, 2:59am.

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 (2013R1A1A1059913) and 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

37 The lower bound in (6) may be proved coupling Bose and Chowla's construction [1] and a simple
 38 product construction. On the other hand, the lower bound in (7) comes from the fact that, for
 39 $s \leq \varepsilon n^{1/(2h-1)}$, a typical s -element subset of $[n]$ becomes a B_h -set after the deletion of a small
 40 fraction of its elements.

41 Now, the lower bound in (7) tells us that, for $s \leq \varepsilon n^{1/(2h-1)}$, the trivial upper bound $Z_h(n, s) \leq$
 42 $\binom{n}{s}$ is sharp up to a factor of the form $(1 + o(1))^s$. The problem is, then, to obtain good upper
 43 bounds for $Z_h(n, s)$ for s of order $n^{1/(2h-1)}$ or larger, perhaps coming close to matching (6). We
 44 believe that this is possible, and put forward such a conjecture in [4], which we reproduce below
 45 for convenience.

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

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

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

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

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

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

55 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
 56 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 -
 57 element subsets of $[n]$. In this range, one might therefore say that B_h -sets are ‘relatively abundant’.
 58 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)
 59 applied for $h = 3$ determine $Z_3(n, s)$ up to a multiplicative factor of the form $s^{o(s)}$, and we see that
 60 the probability that a random s -element subset of $[n]$ is a B_3 -set is roughly of the form $s^{-(2+o(1))s}$.
 61 In this second range, B_3 -sets are therefore scarcer. Finally, note that, by (2), if $s > 3n^{1/3}$ and n is
 62 large, then $Z_3(n, s) = 0$.

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

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

69 **Notation and organization of the paper.** Throughout this paper we identify a graph with the
 70 set of its edges. In particular, if G is a graph, then $e \in G$ means that e is an edge of G ; moreover,
 71 we write both $|G|$ and $e(G)$ for the number of edges in G . If $e = \{x, y\}$ is an edge in a graph,
 72 we sometimes write xy for e . As usual, edges are unordered pairs of vertices; however, if H is a
 73 bipartite graph with vertex classes A and B , it will be convenient to think of the edge set of H as
 74 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
 75 of edges in the subgraph induced by A , which is denoted by $G[A]$. If T is a set, we denote by $K(T)$
 76 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
 77 set of all numbers $w + x$ with $w \in W$.

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

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

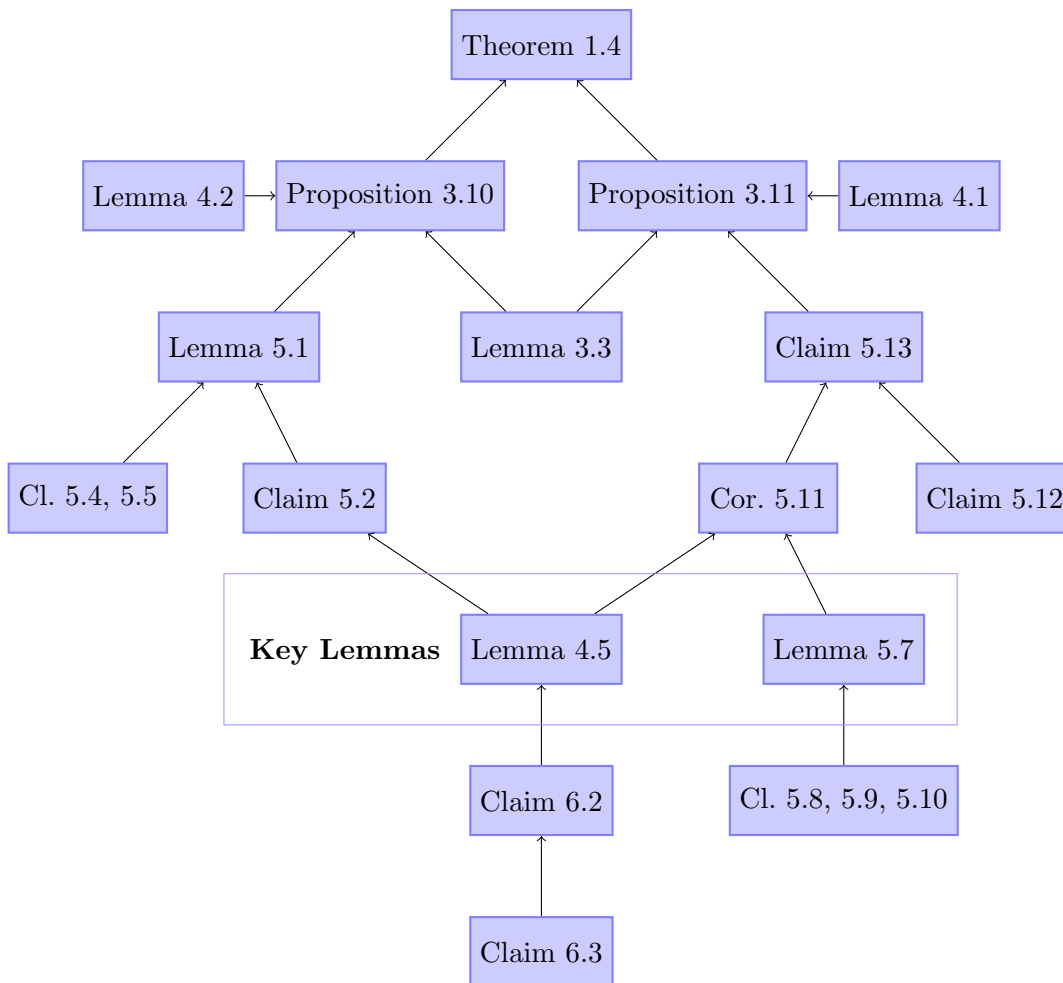


FIGURE 1. A diagram illustrating the flow of the proof of our main result

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

89

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

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

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

96 has $F_2([n]_m) = n^{b_2+o(1)}$, where

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

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

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

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

105 Furthermore,

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

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

107 *Proof of Theorem 2.1.* We shall be somewhat sketchy in the more routine parts of the argument.
 108 We first observe that one may switch to the so called binomial model $[n]_p$. To be more precise,
 109 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
 110 other elements in $[n]$. A standard argument tells us that it suffices to prove that $F_3([n]_p) = n^{b_3+o(1)}$
 111 with probability $1 - o(1/\sqrt{m})$.

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

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

117 Hence, if $(1 + o(1))n^a = pn \ll s^{3-\delta}$, then this expectation is $o(1/\sqrt{m})$. In particular, for every $a >$
 118 $1/5$, with suitably large probability, the largest B_3 -sets contained in $[n]_p$ have cardinality at most

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

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

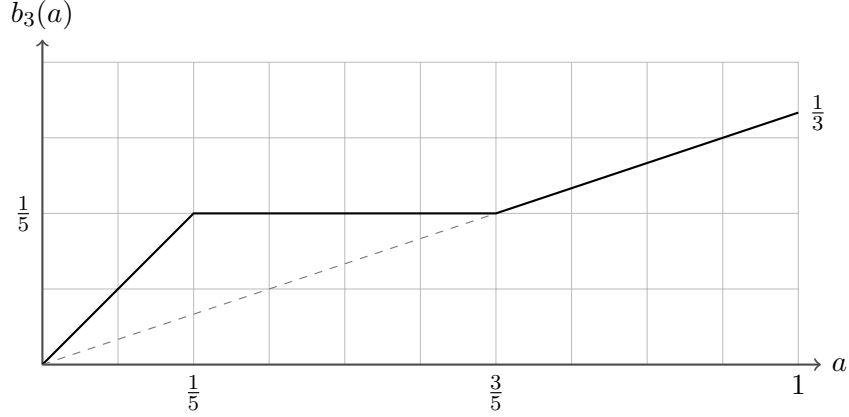


FIGURE 2. The graph of the piecewise linear function b_3 from Theorem 2.1

3. THE PROOF OF THEOREM 1.4

120

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

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

$$a + z_1 + z_2 = b + z_3 + z_4. \quad (14)$$

127 **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*
 128 *independent set in \mathcal{C}_T .*

129 *Proof.* Suppose on the contrary that $a, b \in S \setminus T$ with $a \neq b$ satisfies (14) with $z_1, z_2, z_3, z_4 \in T$.
 130 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.
 131 Since $a \in S \setminus T$ and z_3 and $z_4 \in T$, we obtain that $a = b$, which is a contradiction. \square

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

137 **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*
 138 *numbers with*

$$R \geq e^{-\beta q} N. \quad (15)$$

139 *Suppose*

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

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

$$\binom{N}{q} \binom{R}{m}. \quad (17)$$

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

$$e^{\beta q} \gamma > 1 \quad (18)$$

143 and the bound (17) becomes

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

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

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

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

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

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

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

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

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

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

162 *Proof.* Note that $\{a, b\} \in \tilde{\mathcal{C}}_G[A]$ implies that $R_G(b - a) \geq 1$ which means that $b - a = (z_1 + z_2) -$
 163 $(z_3 + z_4)$ for some $z_i \in T$, and thus $\{a, b\} \in \mathcal{C}_T[A]$ (see Definition 3.1). The proposition follows. \square

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

166 **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
 167 to be all distinct. This restriction allows us to avoid the “degenerate” case in which $z = z_1 - z_3$

168 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
 169 $(z_1x, z_3x) \in G^2$, which means that, without the restriction, the value of $R_G(z)$ could be as large
 170 as $|N_G(z_1) \cap N_G(z_3)|$, which, in turn, could be as large as $|T| - 2$.

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

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

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

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

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

174 In view of the definitions above and Proposition 3.6 and Lemma 3.3, our goal is to enumerate B_3 -
 175 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 $e_{\tilde{C}_G}(A)$
 176 is large for every “large enough” set $A \subset [n]$. This discussion motivates our next definition.

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

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

178 and let $\varepsilon > 0$ be a fixed constant. Given $\lambda \geq 1$ and a non-negative integer i , a set T is said to
 179 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

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

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

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

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

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

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

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

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

192 **Proposition 3.10.** *Let $\varepsilon > 0$, let n be a sufficiently large integer, and let $s \in [n^{1/(5-25\varepsilon)}, 3n^{1/3}]$
 193 be a given integer. Let λ be as defined in (26). The number of B_3 -sets of cardinality s contained*

194 in $[n]$ that **contain a (λ, ε) -bounded set larger than $s^{1-6\varepsilon}$ is at most**

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

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

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

198 Before we proceed with the formal proofs, let us briefly discuss our general approach. Every B_3 -
 199 set with s elements that contains a (λ, ε) -bounded set with at least $s^{1-6\varepsilon}$ elements will be shown
 200 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
 201 shall be able bound the number of possible extensions of any such set T to a B_3 -set with s elements.
 202 This is because the graph \mathcal{C}_T will be shown to satisfy an appropriate local density condition (see
 203 Lemma 5.1). Showing this is the main difficulty in this part of the argument. The details are given
 204 in Section 5.1.

205 The proof of Proposition 3.11 is somewhat more complicated. First we show that any B_3 -set
 206 of cardinality s must contain a (λ, ε) -bounded subset of size at least $s^{1/7}$ (see Lemma 4.1). In
 207 particular, every such B_3 -set contains a *maximal* (λ, ε) -bounded subset with at least $s^{1/7}$ elements.
 208 Our strategy will therefore be to estimate, for each B_3 -set T with $|T| < s^{1-6\varepsilon}$, the number of B_3 -
 209 sets S such that $T \subset S$ and T is a *maximal* (λ, ε) -bounded subset of S . The maximality of T will
 210 be shown to imply that the set of elements that can appear in $S \setminus T$ admits a certain structure
 211 (see Definition 5.6 and Lemma 5.7). More concretely, we shall show that $S \setminus T \subset \tilde{T}$ for some set
 212 $\tilde{T} \subset [n]$ such that the graph $\mathcal{C}_T[\tilde{T}]$ satisfies certain local density conditions that allow us to use
 213 Lemma 3.3 to bound the total number of such possible extensions S of T appropriately (the precise
 214 local density condition is given in Corollary 5.13)

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

216 4. AUXILIARY LEMMAS

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

221 **4.1. Bounded sets.** Our first lemma states that for any $\lambda \geq 1$ and any $\varepsilon > 0$, every B_3 -set S
 222 contains a (λ, ε) -bounded subset whose size is at least a small power of $|S|$.

223 **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*
 224 *cardinality $|T| \geq |S|^{1/7}$.*

225 *Proof.* Observe that S contains a B_4 -set with $\lceil |S|^{1/7} \rceil$ elements. Indeed, one may construct such
 226 set greedily by starting from an empty set and sequentially adding to it elements of S . As long as

227 the constructed set T has fewer than $|S|^{1/7}$ elements, one can always add to T an arbitrary element
 228 from the (non-empty) set $S \setminus (4T - 3T)$, which assures that T remains a B_4 -set.

229 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
 230 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
 231 $(z_1 z_2, z_3 z_4), (z'_1 z'_2, z'_3 z'_4) \in G^2$, with $|\{z_1, z_2, z_3, z_4\}| = 4$ and $|\{z'_1, z'_2, z'_3, z'_4\}| = 4$, satisfy

$$(z_1 + z_2) - (z_3 + z_4) = z = (z'_1 + z'_2) - (z'_3 + z'_4), \quad (29)$$

232 then

$$z_1 + z_2 + z'_3 + z'_4 = z'_1 + z'_2 + z_3 + z_4. \quad (30)$$

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

$$\{z_1, z_2, z'_3, z'_4\} = \{z'_1, z'_2, z_3, z_4\}. \quad (31)$$

234 This forces $\{z_1, z_2\} = \{z'_1, z'_2\}$ and $\{z_3, z_4\} = \{z'_3, z'_4\}$. Consequently $R_G(z) \leq 1$.

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

$$Q_G \leq en. \quad (32)$$

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

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

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

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

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

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

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

247 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
 248 estimates for the harmonic numbers, we obtain, for every large enough n ,

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

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

250 **4.2. A technical lemma on the local density of \mathcal{C}_T .** We now state a key technical lemma that
 251 will help us give lower bounds for the local density of the collision graph \mathcal{C}_T . We need the following
 252 definition.

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

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

260 The following technical lemma allows us to obtain a lower bound on $e_{\mathcal{C}_T}(A)$ in terms of the
 261 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
 262 applying this lemma, we have to come up with a suitable set B (see the proofs of Claim 5.2 and
 263 Corollary 5.11). Recall that ξ and the α_i are defined in 24.

264 **Lemma 4.5.** *The following holds for every integer $i \geq 0$, and every $\varepsilon > 0$, $\lambda \geq 1$, $D \geq 5000$*
 265 *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*
 266 *satisfies $\mathcal{P}_{\lambda, \varepsilon, i}$. Moreover, suppose that $A \subset [n]$ and $B \subset [2n]$ are such that the graph $H = H_T(A, B)$*
 267 *satisfies*

- 268 (I) *every vertex of A has degree at least $\delta|T|$;*
- 269 (II) *the average degree of the vertices in B is D .*

270 *Then*

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

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

272 5. PROOFS OF PROPOSITIONS 3.10 AND 3.11

273 **5.1. Sets containing a large bounded subset.** Let us now prove Proposition 3.10, which deals
 274 with the case in which the ‘‘seed’’ set contains a (λ, ε) -bounded set of cardinality greater than $n^{1-6\varepsilon}$.
 275 Our main tools will be Lemma 3.3 and the following estimate on the number of edges induced by
 276 small sets of vertices in the collision graph \mathcal{C}_T when T is a bounded set.

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

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

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

280 *Proof of Proposition 3.10.* We wish to estimate the number of B_3 -sets S of size s that contain a
 281 (λ, ε) -bounded subset with more than $s^{1-6\varepsilon}$ elements. Suppose that $T \subset S$ is a (λ, ε) -bounded

282 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
 283 assume without loss of generality that the cardinality of T is exactly $s^{1-6\varepsilon}$ and T satisfies $\mathcal{P}_{100\lambda,\varepsilon,0}$.
 284 Lemma 5.1 then implies that \mathcal{C}_T satisfies (16) with

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

285 and

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

286 Take

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

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

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

289 By the discussion above, this completes the proof of Proposition 3.10. \square

290 It now remains to prove Lemma 5.1.

291 *Proof of Lemma 5.1.* Let $C = 10^9$. We shall show that this choice of C will do. Suppose that T
 292 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
 293 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 .
 294 Our goal is to establish a lower bound on $e_{\tilde{\mathcal{C}}_{G_0}}(A)$ and then apply Proposition 3.6 to obtain the
 295 lemma. Let $H = H_T(A, [2n])$ (recall Definition 4.3). Let $c = 10^{-4}$ and let

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

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

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

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

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

299 *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
 300 place of λ , and $A \setminus A'$ in place of A . We proceed in steps.

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

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

305 (iii) From the definition of A' in (43), it follows that $\deg_H(a) = |N_H(a) \cap W| \geq \delta|T|$ for all
 306 $a \in A \setminus A'$.

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

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

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

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

$$e_{C_T}(A \setminus A') = \Omega\left(\delta^2 \frac{|A|^3|T|^6}{\lambda n^2(\log n)^3}\right) = \Omega\left(\frac{\delta^2|A|^2|T|^4}{\lambda n(\log n)^3}\right), \quad (46)$$

310 as required. □

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

312 **Definition 5.3** (Auxiliary graph AUX). *Let AUX be a bipartite graph with classes consisting of A'
 313 and a disjoint copy of $[3n]$ as follows. For $x \in A'$, the neighbors of x in AUX are all elements
 314 of the form $y = x + z_1 + z_2$ for some $z_1 z_2 \in G_0$ such that $x + z_1, x + z_2 \notin W$. In other words,
 315 $z_1 z_2 \in G_0[T \setminus (W - x)]$.*

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

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

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

319 **degenerate path:** if the path is of the form

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

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

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

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

325 (We have an inequality instead of equality in (48) above because, owing to the definition of AUX,
 326 the first edge of the pair must come from $G_0[T \setminus (W - x)]$ and the second edge of the pair from
 327 $G_0[T \setminus (W - x')]$, and hence not all pairs counted by $R_{G_0}(x - x')$ yields appropriate 2-paths in AUX.)
 328 In order to estimate $e_{\tilde{C}_{G_0}}(A') = \sum_{x, x' \in A'} R_{G_0}(x - x')$ we bound the number of degenerate paths
 329 and estimate the total number of 2-paths. Here and in what follows, we write $\sum_{x, x' \in A'}$ for the sum
 330 over all unordered pairs $\{x, x'\} \subset A$ ($x \neq x'$).

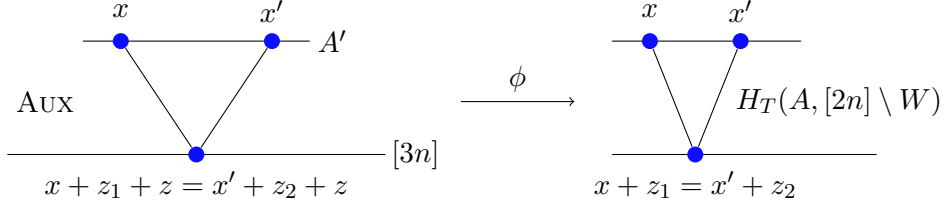


FIGURE 3. Owing to the definition of AUX, we know that $x + z_1 = x' + z_2$ does not belong to W . Consequently, $(x, x + z_1, x')$ is a two-path in $H_T(A, [2n] \setminus W)$, and hence a member of \mathcal{P} .

331 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}
 332 be the set of all paths of length two in $H_T(A', [2n] \setminus W)$ with both endpoints in A' . We will provide
 333 an upper bound for $|\mathcal{D}|$ by defining a map $\phi: \mathcal{D} \rightarrow \mathcal{P}$, estimating $|\mathcal{P}|$ and bounding $|\phi^{-1}(P)|$ for
 334 all $P \in \mathcal{P}$.

335 **Claim 5.4.** *We have*

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

336 *Proof.* First we define a map $\phi: \mathcal{D} \rightarrow \mathcal{P}$ as follows. For a degenerate path in AUX between $x, x' \in A'$
 337 as in (47), we infer that x, x' are connected by a path of length two in H (i.e., $x, x + z_1 = x' + z_2,$
 338 x'). Given the definition of AUX, we also know that $x + z_1 = x' + z_2 \notin W$. Let ϕ map the degenerate
 339 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
 340 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
 341 in G_0 , we conclude that $|\phi^{-1}(P)| \leq |T|$ for any $P \in \mathcal{P}$. Hence $|\mathcal{D}| \leq |T| |\mathcal{P}|$.

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

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

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

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

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

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

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

351 Note that by construction (see Definition 5.3) and the fact that T is a B_3 -set, the degree of
 352 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 $N_H(x) =$

353 $T + x \subset [2n]$, we have

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

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

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

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

356 Therefore

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

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

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

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

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

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

360 It follows by (23) and Proposition 3.6 that

$$e_C(A) \geq \frac{e_{\tilde{C}_{G_0}}(A)}{\log Q_{G_0}} \geq \frac{e_{\tilde{C}_{G_0}}(A)}{200\lambda \log n} \geq \frac{|A|^2 |T|^4}{10^5 \lambda n \log n}. \quad (56)$$

361 Hence Lemma 5.1 is proved. \square

362 **5.2. Sets not containing a large bounded subset.** We now turn to the proof of Proposi-
 363 tion 3.11, that is, we enumerate B_3 -sets such that all of its (λ, ε) -bounded subsets have fewer
 364 than $s^{1-6\varepsilon}$ elements. We shall do this by bounding, for any given (λ, ε) -bounded set T , the number
 365 of ways one can extend T to a B_3 -set S in such a way that T remains a maximal (λ, ε) -bounded
 366 subset of S .

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

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

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

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

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

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

$$\tilde{T} = \bigcup_{i=0}^{\lceil 1/\varepsilon \rceil} \tilde{T}_i. \quad (59)$$

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

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

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

379 *such that, for every $x \in \tilde{T}_i$,*

$$|(T+x) \cap B_i| \geq \alpha_i |T|. \quad (61)$$

380 *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)
 381 of Definition 3.9. Let us fix such a graph G_i for the remainder of the proof of Lemma 5.7. For
 382 technical reasons, it will be convenient to introduce the following definition: for each $w \in [2n]$ and
 383 $z \in [n]$, set

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

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

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

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

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

386 satisfies the conclusions of the lemma.

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

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

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

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

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

391 On the other hand,

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

392 which implies (60), concluding the proof of the claim. \square

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

$$D_{i,xy} = Q_{G_i \cup \{xy\}} - Q_{G_i}. \quad (68)$$

396 Expanding, we obtain

$$D_{i,xy} = \sum_{z \in [n]} \exp(R_{G_i}(z)) \underbrace{\left\{ \exp\left(R_{G_i \cup \{xy\}}(z) - R_{G_i}(z)\right) - 1 \right\}}_{(\ddagger)}. \quad (69)$$

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

398 **Claim 5.9.** *We have*

$$D_{i,xy} \leq U_{i,x+y}. \quad (70)$$

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

401 Note that any difference between $R_{G_i \cup \{xy\}}(z)$ and $R_{G_i}(z)$ must be either due to a pair $(xy, z_1 z_2)$,
 402 $z_1 z_2 \in G_i$, satisfying

$$z = (x + y) - (z_1 + z_2) = w - (z_1 + z_2), \quad (71)$$

403 or due to a pair $(z_1 z_2, xy)$ such that $z = (z_1 + z_2) - w$, where, in both cases, we require $\{x, y\} \cap$
 404 $\{z_1, z_2\} = \emptyset$. If $f_i(w, z) = 0$ then there are no such pairs and we must have $R_{G_i \cup \{xy\}}(z) = R_{G_i}(z)$.
 405 In this case, the term (\ddagger) in (69) is 0.

406 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
 407 at most one edge $\{a', b'\} \in G_i$ for which $-z = (x + y) - a' - b'$. Therefore, we always have
 408 $R_{G_i \cup \{xy\}}(z) \leq R_{G_i}(z) + 2$. Consequently, in this case

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

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

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

$$(\ddagger) \leq e^2 f_i(w, z). \quad (73)$$

411 Therefore,

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

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

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

$$Q_{G'_i} - Q_{G_i} \leq (e+1) \sum_{y \in Y} D_{i,xy}. \quad (74)$$

415 *Proof.* Since $G'_i \setminus G_i = \{xy : y \in Y\}$ contains only edges incident to x , the difference $R_{G'_i}(z) - R_{G_i}(z)$
 416 equals the number of solutions to $z = \pm(x + y - a - b)$ for some $y \in Y$ and $\{a, b\} \in G_i$ with
 417 $\{x, y\} \cap \{a, b\} = \emptyset$. Let us bound the number of such solutions. To this end, suppose that

$$z = x + y - a - b = x + y' - a' - b' \quad (75)$$

418 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
 419 $y + a' + b' = y' + a + b$ and, since these elements come from the B_3 -set T , we conclude that $\{y, a', b'\} =$
 420 $\{y', a, b\}$. It follows that $y = y'$ and $\{a, b\} = \{a', b'\}$. Hence, there is at most one solution
 421 to $z = x + y - a - b$ and at most one solution to $-z = x + y - a - b$. Consequently, $R_{G'_i}(z) - R_{G_i}(z) \leq 2$
 422 and

$$\Delta_z := \exp R_{G'_i}(z) - \exp R_{G_i}(z) \leq (\exp R_{G_i}(z)) (e^2 - 1). \quad (76)$$

423 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
 424 case, we have

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

425 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
 426 other words,

$$\Delta_z \leq (e+1) \sum_{y \in Y} (\exp(R_{G_i \cup \{xy\}}(z)) - \exp(R_{G_i}(z))) \quad (78)$$

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

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

428 Setting

$$Y = \{y \in T : x + y \in [2n] \setminus B_i\} \quad (79)$$

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

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

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

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

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

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

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

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

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

435 Hence

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

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

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

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

439 Together with Lemma 4.5, this yields the following corollary.

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

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

442 we have

$$e_{\mathcal{C}_T}(A) = \Omega\left(\frac{\alpha_i^3 |A|^2}{|T| n^\varepsilon (\log n)^3}\right). \quad (87)$$

443 *Proof.* Fix $i \in \{0, \dots, \lceil 1/\varepsilon \rceil - 1\}$ and let $B = B_i(T)$ be the set from Lemma 5.7. In particular,
444 $|B| = O(|T|^4/(\lambda n^{i\varepsilon}))$. Also let

$$\delta = \alpha_i. \quad (88)$$

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

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

447 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
448 Definition 3.9). Moreover, we also assume that $|T| \geq n^{1/100}$. Lemma 5.7 implies that every $a \in$
449 $A \subset \tilde{T}_i$ satisfies

$$\deg_H(a) = |(T+a) \cap B| \geq \delta |T|. \quad (90)$$

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

$$D = \frac{|H|}{|B|} \geq \frac{\delta |A| |T|}{|B|} \geq \delta \frac{s^{-2+8\varepsilon} n |T|}{|B|} = \Omega\left(\delta \frac{s^{-2+8\varepsilon} n \lambda n^{i\varepsilon}}{|T|^3}\right) = \Omega\left(\delta \frac{s^{3-17\varepsilon} n^{i\varepsilon}}{|T|^3}\right) \geq \delta s^\varepsilon \geq 5000, \quad (91)$$

452 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
453 have

$$\begin{aligned} e_{\mathcal{C}_T}(A) &= \Omega\left(\delta^2 \frac{|A| |T|^2}{\lambda n^{(i+1)\varepsilon} (\log n)^3} \left(\frac{|H|}{|B|}\right)^2\right) \\ &\stackrel{(91)}{=} \Omega\left(\delta^2 \frac{|A| |T|^2}{\lambda n^{(i+1)\varepsilon} (\log n)^3} \cdot D \frac{\delta |T| |A|}{|B|}\right) \\ &= \Omega\left(\delta^3 \frac{D |A|^2 |T|^3}{|B| \lambda n^{(i+1)\varepsilon} (\log n)^3}\right) \\ &\stackrel{(88)}{=} \Omega\left(\alpha_i^3 \frac{D}{|T| n^\varepsilon (\log n)^3} \frac{|T|^4}{|B| \lambda n^{i\varepsilon}} |A|^2\right). \end{aligned} \quad (92)$$

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

456 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,
457 we will estimate how many B_3 -sets have a maximal (λ, ε) -bounded set T with cardinality t . As we
458 observed above, if T is a maximal (λ, ε) -bounded subset of S , then $S \setminus T \subset \tilde{T}$ (recall Definition 5.6).
459 Therefore, it suffices to prove an upper bound for the number of B_3 -sets S satisfying $S \setminus T \subset \tilde{T}$. For
460 that we shall apply Lemma 3.3 to the graph $\mathcal{C}_T[\tilde{T}]$. Therefore we have to show that $\mathcal{C}_T[\tilde{T}]$ satisfies
461 the conditions of the lemma. We need the following claim.

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

463 *Proof.* Recall that $\tilde{T}_{\lceil 1/\varepsilon \rceil}$ is the set of all x such that $T \cup \{x\}$ is a B_3 -set and there is no graph $G_{\lceil 1/\varepsilon \rceil} \subset$
464 $(T \cup \{x\})$ satisfying both conditions (a) and (b) of Definition 3.9 with $i = \lceil 1/\varepsilon \rceil$ and $T \cup \{x\}$ in place
465 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 $G_{\lceil 1/\varepsilon \rceil}$
466 be the complete graph $K(T \cup \{x\})$ on $T \cup \{x\}$, then conditions (a) and (b) hold. Condition (a)
467 follows immediately for $G_{\lceil 1/\varepsilon \rceil} = K(T \cup \{x\})$. As for (b), observe first that, as a little thought
468 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

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

469 which establishes (b) and proves the claim. \square

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

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

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

473 *we have*

$$e_{\mathcal{C}_T}(A) = \Omega\left(\frac{\alpha_{\lceil 1/\varepsilon \rceil - 1}^3 |A|^2}{|T| n^\varepsilon (\log n)^3}\right). \quad (95)$$

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

$$e_{\mathcal{C}_T}(A) = \Omega\left(\frac{\alpha_i^3 |A \cap \tilde{T}_i|^2}{|T| n^\varepsilon (\log n)^3}\right). \quad (96)$$

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

478 We are finally ready to prove Proposition 3.11.

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

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

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

$$q = O\left(|T| n^\varepsilon \frac{(\log n)^4}{\alpha_{\lceil 1/\varepsilon \rceil - 1}^3}\right) \stackrel{(24)}{=} O\left(s^{1-6\varepsilon} s^{5\varepsilon} (\log n)^{32\lceil 1/\varepsilon \rceil + 1}\right) = o(s). \quad (98)$$

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

$$\begin{aligned} \binom{n}{q} \binom{\lceil 1/\varepsilon \rceil s^{-2+8\varepsilon} n}{s-q-|T|} &\leq n^{s-|T|} \left(\frac{\lceil 1/\varepsilon \rceil e}{s^{2-8\varepsilon}(s-q-|T|)} \right)^{s-q-|T|} \\ &\leq n^{s-|T|} \left(\frac{1}{s^{3-8\varepsilon-o(1)}} \right)^s. \end{aligned} \quad (99)$$

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

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

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

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

488 This completes the proof of Proposition 3.11. \square

489 6. PROOF OF LEMMA 4.5

490 Fix an integer $i \geq 0$ and real numbers $\varepsilon > 0$, $\lambda \geq 1$, $D \geq 5000$ and $\delta \in (0, 1]$ satisfying $\delta^2 \geq$
491 $\alpha_i/100\xi$. Suppose that a set $T \subset [n]$ with at least $n^{1/100}$ elements satisfies $\mathcal{P}_{\lambda, \varepsilon, i}$. Moreover, suppose
492 that $A \subset [n]$ and $B \subset [2n]$ are such that the graph $H = H_T(A, B)$ satisfies the two conditions
493 from the statement of the lemma. The fact that T satisfies $\mathcal{P}_{\lambda, \varepsilon, i}$ means that, in particular, we may
494 choose a graph G_i on the vertex set T that satisfies (a) and (b) of Definition 3.9.

495 **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
496 simply a special path, if*

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

500 Note that a 4-path (a, b, a', b', a'') between a and $a'' \in A$ is special if, letting $z_1 = b-a$, $z_2 = b'-a'$,
501 $z_3 = b-a'$, and $z_4 = b'-a''$, we have

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

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

$$a, b := a + z_1, a' := a + z_1 - z_3, b' := a + z_1 - z_3 + z_2, a'' = a + z_1 - z_3 + z_2 - z_4 \quad (101)$$

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

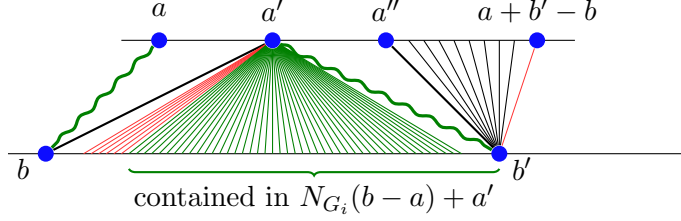


FIGURE 4. Counting semi-special paths extending $P = (a, b, a')$ in the graph H' from Claim 6.3. The first two edges are determined by P and the third edge $\{a', b'\}$ must be such that $\{b - a, b' - a'\} \in G_i$. In view of the properties of H' , most of the H' -neighbors of a' produce extensions 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 $\{a, b\}$.

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

$$N = O(|\tilde{\mathcal{C}}_{G_i}|) \quad (102)$$

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

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

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

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

$$N = \Omega\left(\frac{\delta^2 |A| D^2 |T|^2}{(\log n)^2}\right). \quad (104)$$

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

515 **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*

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

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

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

523 of length four, say (a, b, a', b', a'') with the condition that the differences

$$b - a, b - a', b' - a', b' - a'' \quad (105)$$

524 are all distinct and, moreover $\{b - a, b' - a'\} \in G_i$. Call such (oriented) paths *semi-special*. Note
 525 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
 526 both

$$\{b - a, b' - a'\}, \{b' - a'', b - a'\} \in G_i, \quad (106)$$

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

529 Since H' satisfies (iii), we have $\deg_{H'}(a') \geq \delta|T|/16 \log n$. Moreover, by condition (i), we have
 530 $\deg_{G_i}(b - a) \geq (1 - 4\alpha_i/\delta)|T|$. As we require that $\delta^2 \geq \alpha_i/100\xi$, it follows that the number of
 531 non-neighbors of $b - a$ in G_i is at most

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

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

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

533 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
 534 have at least $\deg_{H'}(b') - 4 \geq d - 4$ possible choices for $a'' \in N_{H'}(b')$ that produce a semi-special
 535 path, namely, the only requirement is that $b' - a''$ must be different from the other three differences
 536 and a'' cannot coincide with a (in fact, one sees that this last condition is automatically satisfied,
 537 if one recalls that T is a B_3 -set). See Figure 4 for an illustration.

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

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

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

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

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

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

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

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

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

543 Let N_4 be the *total* number of paths in H' of length 4 starting and ending in A' . We proved
 544 above that the number N_P of semi-special paths that start with P corresponds to more than 3/4
 545 of the total number of 4-paths that starting with P . Since our argument holds for every P , we

546 conclude that there are more than $3N_4/4$ semi-special paths in H' . Considering the involution that
 547 takes 4-paths $P = (a, b, a', b', a'')$ to their reverse $P^\leftarrow = (a'', b', a', b, a)$, we see that more than $1/2$
 548 of the 4-paths in H' starting and ending in A' are semi-special in *both* directions, and thus are
 549 special. That is, there are more than $N_4/2$ special paths in H' . Finally, we estimate N_4 by first
 550 picking an edge $ab \in H'$, then picking a neighbor $a' \neq a$ of b , and so on. This yields

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

551 whence the claim follows. \square

552 It only remains to prove Claim 6.3.

553 *Proof of Claim 6.3.* This proof will be divided into three simple steps. First we define a set L of
 554 low degree vertices in G_i and show that this set is quite small. In the second step, we partition
 555 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
 556 incident to a good fraction of the edges while at the same time B_j has vertices with roughly the
 557 same degree. Finally, we delete the vertices of low degree in $H_{T \setminus L}(A, B_j)$ to obtain the desired
 558 graph.

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

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

561 Note that

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

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

$$2e(G_i) \geq (1 - \alpha_i)|T|(|T| - 1) \geq |T|^2 - 2\alpha_i|T|^2. \quad (116)$$

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

$$|L| \leq \frac{\delta}{2}|T| \leq \frac{|H|}{2|A|}, \quad (117)$$

564 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
 565 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
 566 the lemma that

$$e(H^*) = e(H) - |L||A| \geq e(H)/2. \quad (118)$$

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

568 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

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

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

571 there exists $1 \leq j \leq \log |T|$ such that there are at least

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

572 edges of H^* incident to B_j .

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

$$e(\widehat{H}) \geq \frac{\delta |T| |A|}{4 \log n}. \quad (121)$$

575 In particular, the average degree of vertices of A in \widehat{H} is at least $\delta |T| / 4 \log n$ and the degrees
576 of vertices in B_j are all in $[4d, 12d]$. While there exists a vertex from A with degree smaller
577 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} ,
578 together with all the incident edges. The number of edges deleted by this procedure is at most

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

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

581

7. CONCLUDING REMARKS

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

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

588 **Conjecture 7.1.** *Let $h \geq 2$ be an integer. Suppose $0 \leq a \leq 1$ is a fixed constant and $m = m(n) =$
589 $(1 + o(1))n^a$. Then, asymptotically almost surely, we have $F_h([n]_m) = n^{b+o(1)}$, where $b = b(a)$ is
590 given by*

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

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

595

REFERENCES

596 [1] R. C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helv. **37** (1962/1963),
597 141–147.

- 598 [2] P. J. Cameron and P. Erdős, *On the number of sets of integers with various properties*, Number theory (Banff,
599 AB, 1988), de Gruyter, Berlin, 1990, pp. 61–79.
- 600 [3] S. Chowla, *Solution of a problem of Erdős and Turán in additive-number theory*, Proc. Nat. Acad. Sci. India.
601 Sect. A. **14** (1944), 1–2.
- 602 [4] D. Dellamonica, Jr., Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij, *On the number of B_h -sets*, submitted.
- 603 [5] P. Erdős, *On a problem of Sidon in additive number theory and on some related problems. Addendum*, J. London
604 Math. Soc. **19** (1944), 208.
- 605 [6] P. Erdős and P. Turán, *On a problem of Sidon in additive number theory, and on some related problems*, J.
606 London Math. Soc. **16** (1941), 212–215.
- 607 [7] B. Green, *The number of squares and $B_h[g]$ sets*, Acta Arith. **100** (2001), no. 4, 365–390.
- 608 [8] H. Halberstam and K. F. Roth, *Sequences*, second ed., Springer-Verlag, New York, 1983.
- 609 [9] D. J. Kleitman and K. J. Winston, *On the number of graphs without 4-cycles*, Discrete Math. **41** (1982), no. 2,
610 167–172.
- 611 [10] Y. Kohayakawa, S. Lee, and V. Rödl, *The maximum size of a Sidon set contained in a sparse random set
612 of integers*, Proceedings of the 22nd Annual ACM–SIAM Symposium on Discrete Algorithms (SODA 2011)
613 (Philadelphia, PA, USA), Society for Industrial and Applied Mathematics, 2011, pp. 159–171.
- 614 [11] Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij, *The number of Sidon sets and the maximum size of Sidon
615 sets contained in a sparse random set of integers*, Random Structures Algorithms, to appear.
- 616 [12] K. O’Byrant, *A complete annotated bibliography of work related to Sidon sequences*, Electron. J. Combin. (2004),
617 Dynamic surveys 11, 39 pp. (electronic).
- 618 [13] D. Saxton and A. Thomason, *Hypergraph containers*, arXiv:1204.6595 (2012).
- 619 [14] S. Sidon, *Ein Satz über trigonometrische Polynome und seine Anwendung in der Theorie der Fourier-Reihen*,
620 Math. Ann. **106** (1932), no. 1, 536–539.
- 621 [15] J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Trans. Amer. Math.
622 Soc. **43** (1938), no. 3, 377–385.

623 DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EMORY UNIVERSITY, ATLANTA, GA 30322, USA
624 (D. Dellamonica Jr., Y. Kohayakawa and V. Rödl)
625 *E-mail address:* `domingos.junior@gmail.com`, `rodl@mathcs.emory.edu`

626 INSTITUTO DE MATEMÁTICA E ESTATÍSTICA, UNIVERSIDADE DE SÃO PAULO, RUA DO MATÃO 1010, 05508–
627 090 SÃO PAULO, BRAZIL (Y. Kohayakawa)
628 *E-mail address:* `yoshi@ime.usp.br`

629 DEPARTMENT OF MATHEMATICAL SCIENCES, KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY
630 (KAIST), DAEJEON, SOUTH KOREA (S. J. Lee)
631 *E-mail address:* `sjlee242@gmail.com`

632 SCHOOL OF MATHEMATICAL SCIENCES, TEL AVIV UNIVERSITY, TEL AVIV 69978, ISRAEL, AND TRINITY COLLEGE,
633 CAMBRIDGE CB2 1TQ, UK (W. Samotij)
634 *E-mail address:* `samotij@post.tau.ac.il`