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,

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 [15], 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

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

In the case addressed by Sidon, that is, for $h = 2$, results of Chowla, Erdős, Singer, and Turán [3,
5, 6, 16] 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 $n$,

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

Note that, for $h = 3$, the first inequality in (1) 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

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

Date: 2016/02/06.
The third author is the corresponding author (email:sjlee242@gmail.com).
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 sup-
ported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the
Ministry of Science, ICT & Future Planning (NRF-2013R1A1A050913) and 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 080070, 1301698, and 1102086. The fifth author was partially supported by
ERC Advanced Grant DMMCA and a Trinity College JRF.

© 2016. This manuscript version is made available under the Elsevier user license
http://www.elsevier.com/open-access/userlicense/1.0/
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 [13].

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 \(1 \leq s \leq n\), let

\[
Z_h(n, s) = |\{ S \subseteq [n] : |S| = s\text{ and } S\text{ is a }B_h\text{-set}\}|
\]

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 \(c'_hn^{1/h} \leq \log Z_h(n) \leq C_hn^{1/h} \log n\) for some positive constants \(c'_h\) and \(C_h\). One now knows that, in fact,

\[
\log Z_h(n) \leq C'_hn^{1/h}
\]

for some constant \(C'_h\). The case \(h = 2\) of [2] is proved in [11] (see also [14]), and the case of arbitrary \(h\) is dealt with in [4]. To establish (2), we considered the more refined question of estimating \(Z_h(n, s)\) for a range of \(s\). Roughly speaking, we obtained good bounds for \(Z_h(n, s)\) for \(s \geq n^{1/(h+1)}(\log n)^2\) and derived [2] summing over all relevant \(s\) (see [4] for details).

We believe that the problem of estimating \(Z_h(n, s)\) for the whole range of \(s\) is interesting in its own right. Moreover, it 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 \(h \geq 2\).

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

\[
Z_h(n, s) \geq \left(\frac{c'_hn}{s^h}\right)^s.
\]

(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 \(n\), we have

\[
Z_h(n, s) \geq (1 - \delta)^s \binom{n}{s}.
\]

The lower bound in (3) may be proved by considering a family of “blow-ups” of an \(s\)-element \(B_h\)-set contained in \([3hs^h]\); the existence of such a set was proved by Bose and Chowla [1]. On the other hand, the lower bound in (4) comes from the fact that, for \(s \leq \varepsilon n^{1/(2h-1)}\), a typical \(s\)-element subset of \([n]\) becomes a \(B_h\)-set after the deletion of a small fraction of its elements.

Now, the lower bound in (4) tells us that, for \(s \leq \varepsilon n^{1/(2h-1)}\), the trivial upper bound \(Z_h(n, s) \leq \binom{n}{s}\) is sharp up to a factor of \((1 + o(1))^s\). 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 (3). 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 \( h \geq 2 \) and a real number \( \delta > 0 \). For every \( s \geq n^{1/(2h-1)+\delta} \) and every large enough \( n \), we have
\[
Z_h(n, s) \leq \left( \frac{n}{s^{h-\delta}} \right)^s.
\]

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
\[
Z_3(n, s) \leq \left( \frac{n}{s^{3-\delta}} \right)^s.
\]

We believe that our methods for proving Theorem 1.4 can eventually be adapted to establish Conjecture 1.3 for every \( h \), but the case of general \( h \) 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 \( s^{o(s)} \). In particular, we see that the probability that a random \( s \)-element subset of \([n]\) is a \( B_3 \)-set is \( s^{-(2+o(1))s} \). In this second range, \( B_3 \)-sets are therefore scarcer. Finally, note that, by (1), 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 \( (n/s^{3-o(1)})^s \); 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 \((1-o(1))^s(n/s)^s = (\Theta(n/s))^s \).

Theorem 1.4 implies a result in probabilistic combinatorics, which confirms the case \( h = 3 \) of a conjecture put forward in [1]. 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 \( W \subset Z \) and \( x \in Z \) 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 \( \Theta \)-notation (with respect to \( n \to \infty \)); the implicit constants are always absolute constants. We omit floor \( \lfloor \cdot \rfloor \) and ceiling \( \lceil \cdot \rceil \) 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 chosen 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 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] [12], the cardinality of the largest $B_3$-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 $m = m(n) = n^{a+o(1)}$ for some constant $0 \leq a \leq 1$. 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

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

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 $0 \leq a \leq 1$ be a fixed constant. Suppose $m = m(n) = n^{a+o(1)}$. There exists a constant $b_3 = b_3(a)$ such that, asymptotically almost surely, we have

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

Furthermore,

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

The piecewise linear function $b_3$ in (5) 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 $p = m/n = n^{a-1+o(1)}$ 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 $F_3([n]_p) \leq |[n]_p|$, 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 $s \geq n^{1/5+\delta}$, this expectation is at most

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

Hence, if $n^{a+o(1)} = pm \ll s^{3-\delta}$, 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

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

Since $\delta > 0$ is arbitrary, the result follows. □
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 $C_T$). Given a set $T \subset [n]$, we define the collision graph $C_T$ 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

$$a + z_1 + z_2 = b + z_3 + z_4. \quad (6)$$

Observation 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 $C_T$.

Proof. Suppose on the contrary that $a, b \in S \setminus T$ with $a \neq b$ satisfies (6) 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. \[\square\]

In view of Observation 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 $C_T$ 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

$$R \geq e^{-\beta q} N. \quad (7)$$

Suppose

$$e(A) \geq \beta \left( \frac{|A|}{2} \right) \quad \text{for any } A \subset V(G) \text{ with } |A| \geq R. \quad (8)$$
Then, for all integers \(m \geq 0\), the number of independent sets in \(G\) of cardinality \(q + m\) is at most

\[
\binom{N}{q} \binom{R}{m}.
\]  

(9)

When applying Lemma 3.3, we shall often take \(R = \gamma N = \gamma|V(G)|\) for some number \(\gamma > 0\). Hypothesis (7) then becomes

\[
e^{\beta q \gamma} > 1
\]

(10)

and the bound (9) becomes

\[
\binom{|V(G)|}{q} \binom{\gamma|V(G)|}{m}.
\]

In order to prove our main result, Theorem 1.4, we shall enumerate all possible seed sets \(T\) and show that the corresponding graphs \(C_T\) are quite dense. In fact, they are sufficiently dense so 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 \(e_{C_T}(A)\) 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 \(T \subset [n]\). We denote by \(G^2 = G \times G\) the Cartesian product of the edge set of \(G\) with itself.

**Definition 3.4** (Representation count \(R_G\)). For an integer \(z \in [−n, n]\), let

\[
R_G(z) = |\{(z_1 z_2, z_3 z_4) \in G^2 : z = (z_1 + z_2) - (z_3 + z_4), \text{ all } z_i \text{ s distinct}\}|.
\]

(11)

Clearly \(R_G(−z) = R_G(z)\) for every \(z\). Also, \(T\) will always be a \(B_3\)-set, and hence we shall always have \(R_G(0) = 0\). Consequently, we shall only be interested in \(R_G(z)\) for \(z \in [n]\).

**Definition 3.5** (Collision multigraph \(\tilde{C}_G\)). Let \(\tilde{C}_G\) 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 \(C_T\) 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 \(C_T\) through the following proposition.

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

\[
e_{C_T}(A) \max_{z \in [−n, n]} R_G(z) \geq e_{\tilde{C}_G}(A).
\]

Proof. Note that \(\{a, b\} \in \tilde{C}_G[A]\) 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 \(\{a, b\} \in C_T[A]\) (see Definition 3.1). The proposition follows. \(\square\)

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 Observation 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

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

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

$$\max_{z \in [-n, n]} R_G(z) = \max_{z \in [n]} R_G(z) \leq \log Q_G. \quad (12)$$

In view of the definitions above and Observation 3.6 and Lemma 3.3, our goal is to enumerate $B_3$-sets $T \subset [n]$ and graphs $G \subset (\mathcal{F}_3)$ such that $\log Q_G$ is not too large, while at the same time $e_{e_G}(A)$ is large for every “large enough” set $A \subset [n]$. This discussion motivates our next definition.

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

$$\xi = \frac{1}{10^9 \log n} \quad \text{and} \quad \alpha_i = \xi^{2^i+1-1} \text{ for } i \geq 0,$$

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

\[ (a) \quad e(G_i) \geq (1 - \alpha_i)\left(\frac{|T|}{2}\right), \]

\[ (b) \quad Q_{G_i} \leq e \exp\left(\lambda n^{\xi} \sum_{j=1}^{[T]} \frac{1}{j}\right). \]

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

Our main objective is to use the Keitman–Winston lemma to obtain an upper bound on the number of $B_3$-sets. Since we are now aiming to prove (almost) tight upper bounds on the number of $B_3$-sets with $s$ elements for a wide range of $s$, unlike in our previous work, we may no longer use the Kleitman–Winston lemma “directly”; we are unable to obtain a sufficiently strong lower bounds on $e_{e_G}(A)$ for very small sets $T$. Instead, we shall define a collection of “seeds” $\{T^{(1)}, T^{(2)}, \ldots, T^{(m)}\}$ and companion sets $\{\tilde{T}^{(1)}, \tilde{T}^{(2)}, \ldots, \tilde{T}^{(m)}\}$ such that every $B_3$-set $S$ satisfies $T^{(j)} \subset S \subset T^{(j)} \cup \tilde{T}^{(j)}$ for some $j$. (For the definition of $\tilde{T}^{(j)}$, see Definition 3.6.) The actual bound comes from counting, for each pair $(T^{(j)}, \tilde{T}^{(j)})$, how many $B_3$-sets $S$ contain $T^{(j)}$ and are contained in $T^{(j)} \cup \tilde{T}^{(j)}$.

The sets $T^{(j)}$ will be the bounded sets in the very technical Definition 3.9. The set $\tilde{T}^{(j)}$ corresponding to $T^{(j)}$ will include all elements that cannot be added to $T^{(j)}$ to extend it to a larger bounded set (with the same parameters). In particular, for every $a \in \tilde{T}^{(j)}$, we know that $T^{(j)} \cup \{a\}$ is not a bounded set. A structural result about $T^{(j)}$ and $\tilde{T}^{(j)}$, Lemma 5.7, allows us to show that $\tilde{T}^{(j)}$ induces a rather dense subgraph in the collision graph $C_{T^{(j)}}$, see Corollary 5.11. Hence, by counting independent sets in $C_{T^{(j)}}$, we may bound the number of $B_3$-sets sandwiched between $T^{(j)}$ and $T^{(j)} \cup \tilde{T}^{(j)}$.
So far this explains what we intend to do with this boundedness property, but not why it is defined this way. This definition is shaped by our goals and also by the induction used in the proof. Ideally we would like to control \( \max_z R_{K(T)}(z) \), where \( K(T) \) is the complete graph on the vertex set \( T \). However, it turns out that this is not sufficiently flexible for our proof to work. Indeed, suppose that we would like to bound this maximum by some value \( L = o(|T|) \). Now consider a set of \( 4(L+1) \) elements such that there are \( L+1 \) distinct sums of type \( a_1 + a_2 - b_1 - b_2 \) yielding the same value \( z \), and notice that all extensions of this (pathological) set to some larger set \( T \) will fail the representation count bound. Considering subgraphs \( G \) of \( K(T) \) gives us enough flexibility to handle cases like the one described above by simply eliminating a few pairs from \( K(T) \).

The definition also allows us to vary how strong the bound on \( R_G \) (indirectly represented by \( Q_G \)) is. This strength is controlled by the parameter \( n^\varepsilon \), ranging between \( n^\varepsilon \) and \( n \), where \( n \) gives a trivial bound which holds for any \( B_3 \)-set \( T \) even with \( G = K(T) \). In tighter bounds on \( Q_G \), we allow the graph \( G \) to have smaller density as can be seen in the definition; see the role played by \( i \) in the bound on the density of \( G_i \). The role of this sequence of bounds will become more apparent after the proof of Corollary \( 5.11 \). We therefore postpone the explanation of the subtle details to Remark \( 5.12 \).

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

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

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

\[
\lambda = \lambda(s) = \frac{s^{5-25\varepsilon}}{n} \geq 1.
\]  

We classify the \( B_3 \)-sets of size \( s \) into two types, depending on whether or not the cardinality of their largest \((\lambda, \varepsilon)\)-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 \( s \in \left[ n^{1/(5-25\varepsilon)}, 3n^{1/3} \right] \) be a given integer. Let \( \lambda \) be as defined in \( (14) \). The number of \( B_3 \)-sets of cardinality \( s \) contained in \( [n] \) that contain a \((\lambda, \varepsilon)\)-bounded set larger than \( s^{1-6\varepsilon} \) is at most

\[
\left( \frac{n}{s^{3-12\varepsilon-o(1)}} \right)^s.
\]

**Proposition 3.11.** Let \( \varepsilon > 0 \), let \( n \) be a sufficiently large integer, and let \( s \in \left[ n^{1/(5-25\varepsilon)}, 3n^{1/3} \right] \) be a given integer. Let \( \lambda \) be as defined in \( (14) \). The number of \( B_3 \)-sets of cardinality \( s \) contained in \( [n] \) that do not contain any \((\lambda, \varepsilon)\)-bounded set larger than \( s^{1-6\varepsilon} \) is at most

\[
\left( \frac{n}{s^{3-8\varepsilon-o(1)}} \right)^s.
\]

9
Before we proceed with the formal proofs, let us briefly discuss our general approach. Every $B_3$-set with $s$ elements that contains a $(\lambda, \varepsilon)$-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 $P_{100\lambda, \varepsilon, 0}$ (see Lemma 4.2). Using Lemma 3.3, we shall be able to bound the number of possible extensions of any such set $T$ to a $B_3$-set with $s$ elements. This is because the graph $C_T$ 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 $(\lambda, \varepsilon)$-bounded subset of size at least $s^{1/7}$ (see Lemma 4.1). In particular, every such $B_3$-set contains a maximal $(\lambda, \varepsilon)$-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 $(\lambda, \varepsilon)$-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 $S \setminus T \subset \tilde{T}$ for some set $\tilde{T} \subset [n]$ such that the graph $C_T[\tilde{T}]$ 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 5.14).

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 4.2 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 $C_T$.

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 $(\lambda, \varepsilon)$-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 $(\lambda, \varepsilon)$-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 a 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

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

then

$$z_1 + z_2 + z'_1 + z'_2 = z_3 + z'_3 + z_4 + z_4.$$

$$z_1 + z_2 + z'_1 + z'_2 = z_3 + z'_3 + z_4 + z_4.$$
Since $T$ is a $B_4$-set, we must have, as multisets, $\{z_1, z_2, z_3', z_4'\} = \{z_1', z_2', z_3, z_4\}$. 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,

$$Q_G \leq en.$$ 

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 $(\lambda, \varepsilon)$-bounded set.

The second lemma allows one to pass to subsets of a convenient cardinality when dealing with $(\lambda, \varepsilon)$-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 $P_{\lambda, \varepsilon, i}$. For every $m$ satisfying $n^{1/100} \leq m \leq |T|$, there exists $T' \subset T$ with $|T'| = m$ such that $T'$ satisfies $P_{100\lambda, \varepsilon, i}$. 

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

$$e_{G_i}(T') \geq e(G_i) \left( \frac{|T| - 2}{m - 2} \right) \left( \frac{|T|}{m} \right)^{-1} = e(G_i) \left( \frac{m}{2} \right) \left( \frac{|T|}{2} \right)^{-1}. $$

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

$$e(G'_i) \geq (1 - \alpha_i) \left( \frac{|T|}{2} \right) \left( \frac{m}{2} \right) \left( \frac{|T|}{2} \right)^{-1} = (1 - \alpha_i) \left( \frac{m}{2} \right). $$

(15)

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 (i.e., $\sum_{j=1}^{n} 1/j \approx \log n$), we obtain, for every large enough $n$,

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

(16)

It follows from (15) and (16) that $T'$ satisfies $P_{100\lambda, \varepsilon, i}$.

4.2. **A technical lemma on the local density of $C_T$.** We now state a key technical lemma that will help us give lower bounds for the local density of the collision graph $C_T$. 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 $e_{C_T}(A)$ 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 $\xi$ and the $\alpha_i$ are defined in (13).

**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

$$e_{C_T}(A) = \Omega\left(\frac{\delta^2|A|D^2|T|^2}{\lambda n\varepsilon(\log n)^3}\right).$$

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 $(\lambda, \varepsilon)$-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 $C_T$ 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

$$e_{C_T}(A) = \Omega\left(\frac{|A|^2|T|^4}{\lambda n(\log n)^3}\right).$$

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 $(\lambda, \varepsilon)$-bounded subset with more than $s^{1-6\varepsilon}$ elements. Suppose that $T \subset S$ is a $(\lambda, \varepsilon)$-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 $C_T$ satisfies (8) with

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

and

$$\beta = \Omega\left(\frac{|T|^4}{\lambda n(\log n)^3}\right) \quad \Omega\left(\frac{(s^{1-6\varepsilon})^4}{s^{2-26\varepsilon}(\log n)^3}\right) = \Omega\left(\frac{1}{s^{1-\varepsilon}(\log n)^3}\right).$$

Take

$$q = (\log n)/\beta = O\left((\log n)^4s^{1-\varepsilon}\right) = o(s).$$
Note that (10) is satisfied. Hence, Observation 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
\[
\left(\frac{n}{s^{1-6\varepsilon}}\right)\left(\left(\frac{n}{s^{1-6\varepsilon}}\right) - 1\right) \leq n^s \left(\frac{C\varepsilon}{s^2-12\varepsilon(s-q-s^{1-6\varepsilon})}\right)^{s-q-s^{1-6\varepsilon}} = \left(\frac{n}{s^3-12\varepsilon-o(1)}\right)^s.
\]
By the discussion above, this completes the proof of Proposition 3.10. \qed

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 $A \subseteq [n]$ with $|A| \geq Cn/|T|^2$ be arbitrary. Recall that, by definition, there exists a graph $G_0 \subseteq \binom{T}{2}$ satisfying $\binom{a}{2}$ and $\binom{b}{2}$ of Definition 3.9 with $i = 0$ and $\lambda$ replaced by $100\lambda$. Our goal is to establish a lower bound on $e_{G_0}(A)$ and then apply Observation 3.6 to obtain the lemma. Let $H = H_T(A, [2n])$ (recall Definition 4.3). Let $c = 10^{-4}$ and let
\[
W = \{ w \in [2n] : \deg_H(w) \geq c |T|^2 |A|/n \}
\] (the vertices in $W$ have very large degree). Notice that, since $|H| \leq |A| |T|$, we have $|W| \leq n/(c|T|)$.

Also, set
\[
A' = \{ a \in A : |N_H(a) \cap W| < 0.1 |T| \}.
\]

Claim 5.2. If $|A'| \leq |A|/2$, then the conclusion of the lemma holds. More precisely,
\[
e_{C,T}(A \setminus A') = \Omega\left(\frac{\delta^2 |A|^2 |T|^4}{\lambda n (\log n)^3}\right).
\]

Proof. We apply Lemma 4.5 to the graph $H' = H_T(A \setminus A', W) \subseteq H$ with $i = 0$, $B = W$, $100\lambda$ in place of $\lambda$, 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 $\delta$ 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 $100\lambda$ in place of $\lambda$.

(iii) From the definition of $A'$ in (18), 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
\[
D = \frac{|H'|}{|W|} \geq \frac{|A|}{2} \cdot \frac{0.1 |T|}{|W|} \geq \frac{c |A| |T|^2}{20n}.
\]

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

From Lemma 4.5 with $D = \Omega(|A| |T|^2/n)$ as in (19) and $|A| \geq C n/|T|^2$, we conclude that
\[
e_{C,T}(A \setminus A') = \Omega\left(\frac{\delta^2 |A|^2 |T|^6}{\lambda n^2 (\log n)^2}\right) = \Omega\left(\frac{\delta^2 |A|^2 |T|^4}{\lambda n (\log n)^3}\right),
\]
as required. \qed

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


**Definition 5.3** (Auxiliary graph $\text{Aux}$). Let $\text{Aux}$ be a bipartite graph whose color classes are $A'$ and a disjoint copy of $[3n]$ and whose edge set is defined as follows. For $x \in A'$, the neighbors of $x$ in $\text{Aux}$ are all elements 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, $z_1z_2 \in G_0[T \setminus (W - x)]$.

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

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

$$x, x + z_1 + z_2 = x' + z_3 + z_4, x'$$

with $(z_1z_2, z_3z_4) \in G_0^2$ and $z_i$s distinct;

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

$$x, x + z_1 + z = x' + z_2 + z, x'$$

with $(z_1z, z_2z) \in G_0^2$.  

(20)

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 (11)). Therefore,

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

(21)

We have an inequality instead of equality in (21) above because, owing to the definition of $\text{Aux}$, the first edge of the path must come from $G_0[T \setminus (W - x)]$ and the second edge of the path from $G_0[T \setminus (W - x')]$, and hence not all pairs counted by $R_{G_0}(x - x')$ yield appropriate 2-paths in $\text{Aux}$.)

In order to estimate $e_{G_0}(A') = \sum_{x, x' \in A'} R_{G_0}(x - x')$ 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 $\{x, x'\} \subset A'$ ($x \neq x'$).

Let $\mathcal{D}$ denote the set of all degenerate paths in $\text{Aux}$, 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 $\phi: \mathcal{D} \to \mathcal{P}$, estimating $|\mathcal{P}|$ and bounding $|\phi^{-1}(P)|$ for all $P \in \mathcal{P}$.

**Claim 5.4.** We have

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

**Proof.** First we define a map $\phi: \mathcal{D} \to \mathcal{P}$ as follows. For a degenerate path in $\text{Aux}$ between $x$, $x' \in A'$ as in (20), 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 $\text{Aux}$, we also know that $x + z_1 = x' + z_2 \notin W$. Let $\phi$ 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 $|\phi^{-1}(P)| \leq |T|$ for any $P \in \mathcal{P}$. Hence $|\mathcal{D}| \leq |T| |\mathcal{P}|$.

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
Since \( G \) by using Jensen’s inequality:

\[
|D| \leq |T| |P| \leq |T| |A'| |T| \cdot \frac{c |T|^2 |A|}{n} \leq \frac{c |A|^2 |T|^4}{n}.
\]

\[\square\]

**Claim 5.5.** The number of 2-paths between vertices of \( A' \) in \( \text{Aux} \), namely the sum \( \sum_{x,x' \in A'} p(x,x') \), is at least \( \frac{|A|^2 |T|^4}{7 \times 64n} \).

**Proof.** The total number of 2-paths between pairs of vertices of \( A' \) in \( \text{Aux} \) can be bounded below by using Jensen’s inequality:

\[
\sum_{y \in [3n]} \left( \frac{\deg_{\text{Aux}}(y)}{2} \right) \geq 3n \left( \frac{|\text{Aux}|}{3n} \right)^2 \geq \frac{|\text{Aux}|^2}{7n}.
\]

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 \( \text{Aux} \) 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

\[
|T \setminus (W - x)| = |T| - |T \cap (W - x)| = |T| - |(T + x) \cap W| = |T| - |N_H(x) \cap W| > 0.9 |T|.
\]

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

\[
\deg_{\text{Aux}}(x) = e_{G_0}(T \setminus (W - x)) \geq \left( \frac{|T \setminus (W - x)|}{2} \right) - |K(T) \setminus G_0| \geq \left( \frac{0.9 |T|}{2} \right) - \alpha_0 \left( \frac{|T|}{2} \right) > \frac{|T|^2}{4}.
\]

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

Therefore

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

Since \( |A'| \geq |A|/2 \), the claim follows. \[\square\]

It follows from Claims 5.4 and 5.5 together with (21) and \( c = 1/10000 \), that

\[
e_{\tilde{G}_0}(A) \geq e_{\tilde{G}_0}(A') \geq \frac{|A|^2 |T|}{500n}.
\]
Since $Q_{G_0}$ satisfies $[b]$ with $i = 0$, and $100\lambda \geq 1$, we have

$$Q_{G_0} \leq \epsilon n \exp\left(100\lambda(1 + \log |T|)\right) \leq \exp(200\lambda \log n).$$

It follows by [12] and Observation [3.6] that

$$e_C(A) \geq \frac{e_{C_{G_0}}(A)}{\log Q_{G_0}} \geq \frac{e_{C_{G_0}}(A)}{200\lambda \log n} \geq \frac{|A|^2 |T|^4}{10^6 \lambda n \log n}.$$

Hence Lemma 5.1 is proved. \hfill \Box

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 whose all $(\lambda, \varepsilon)$-bounded subsets have fewer than $s^{1-6\varepsilon}$ elements. We shall do this by bounding, for any given $(\lambda, \varepsilon)$-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 $(\lambda, \varepsilon)$-bounded subset of $S$.

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

Definition 5.6. Given a $(\lambda, \varepsilon)$-bounded set $T$, let

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

Also, for $i \in \{0, 1, \ldots, [1/\varepsilon]\}$, let

$$\tilde{T}_i = \{x \in \tilde{T} : i \text{ is the smallest index such that } T \cup \{x\} \text{ does not satisfy } P_{\lambda, \varepsilon, i}\}.$$

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

$$\tilde{T} = \bigcup_{i=0}^{\lfloor 1/\varepsilon \rfloor} \tilde{T}_i.$$

The next lemma gives us important information on the sets $\tilde{T}_i$. The sets $B_i$, whose existence is asserted in this lemma, will be crucial for us to prove that $C_T[\tilde{T}_i]$ 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 $i \in \{0, 1, \ldots, [1/\varepsilon]\}$ and suppose that a set $T$ satisfies $P_{\lambda, \varepsilon, i}$. There exists a set $B_i = B_i(T) \subset [2n]$ with

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

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

$$|(T+x) \cap B_i| \geq \alpha_i |T|.$$
On the other hand, \( z \in [n] \), set
\[
 f_i(w, z) = \begin{cases} 
 1 & \text{if } z = \pm(w - a - b) \text{ for some } \{a, b\} \in G_i, \\
 0 & \text{otherwise}.
\end{cases}
\]

Also, for each \( w \in [2n] \), let
\[
 U_{i,w} = \sum_{z \in [n]} (\exp R_{G_i}(z)) e^2 f_i(w, z),
\]
which is motivated by Claim 5.9. In what follows, we will show that the set \( B_i \) defined by
\[
 B_i = \left\{ w \in [2n] : U_{i,w} > \frac{\lambda n^{i,z} Q_{G_i}}{(e + 1)|T|(|T| + 1)} \right\}
\]
satisfies the conclusions of the lemma.

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

Proof. We start by proving the following inequality, which will be used shortly:
\[
 \sum_{w \in [2n]} f_i(w, z) \leq 2 e(G_i).
\]
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
\[
 \sum_{w \in [2n]} U_{i,w} = \sum_{w \in [2n]} \sum_{z \in [n]} (\exp R_{G_i}(z)) e^2 f_i(w, z) = e^2 \sum_{z \in [n]} \exp R_{G_i}(z) \sum_{w \in [2n]} f_i(w, z) \\
 \leq e^2 Q_{G_i} 2 e(G_i) < e^2 |T|(|T| - 1) Q_{G_i}.
\]
On the other hand,
\[
 \sum_{w \in [2n]} U_{i,w} \geq \sum_{w \in B_i} U_{i,w} \geq |B_i| \frac{\lambda n^{i,z} Q_{G_i}}{(e + 1)|T|(|T| + 1)},
\]
which implies (22), concluding the proof of the claim.

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 \( x \in \tilde{T}_i \).

For \( y \in T \), denote by \( G_i \cup \{xy\} \) the graph with vertex set \( V(G_i) \cup \{x\} \) and edge set \( E(G_i) \cup \{xy\} \), and we consider
\[
 Q_{G_i \cup \{xy\}} - Q_{G_i} = \sum_{z \in [n]} \exp(R_{G_i}(z)) \left[ \exp(\sum_{z \in [n]} \exp(R_{G_i\cup\{xy\}}(z) - R_{G_i}(z)) - 1 \right).
\]

The following claim relates \( Q_{G_i \cup \{xy\}} - Q_{G_i} \) and \( U_{i,x+y} \).

Claim 5.9. We have
\[
 Q_{G_i \cup \{xy\}} - Q_{G_i} \leq U_{i,x+y}.
\]

Proof. We shall prove the claim by showing that every term in the sum (25) is bounded from above by its corresponding term in the sum (23) defining \( U_{i,x+y} \). 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 \( (x, z_1 z_2) \), \( z_1 z_2 \in G_i \), satisfying \( z = (x+y)-(z_1 z_2) \), or due to a pair \( (z_1 z_2, xy) \) such that \( z = (z_1 z_2)-(x+y) \),
where, in both cases, we require \( \{x, y\} \cap \{z_1, z_2\} = \emptyset \). If \( f_i(x + y, z) = 0 \) 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 \((\dagger)\) in (25) is 0.

Since \( T \) is a \( B_3 \)-set, there can be at most one edge \( \{a, b\} \in G_i \) such that \( z = (x + y) - a - b \) 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

\[
R_{G_i}(z) \leq R_{G_i \cup \{xy\}}(z) \leq R_{G_i}(z) + 2.
\]

In particular, the term \((\dagger)\) in (25) is 0, \( e - 1 \) or \( e^2 - 1 \).

To summarize, regardless of whether \( f_i(x + y, z) = 0 \) or \( 1 \), we have \((\dagger) \leq e^2 f_i(x + y, z)\). Therefore,

\[
Q_{G_i \cup \{xy\}} - Q_{G_i} \leq \sum_{z \in [n]} (\exp R_{G_i}(z)) e^2 f_i(x + y, z) = U_{i,x+y}.
\]

Next we show that the increase of the moment function caused by adding multiple edges incident to \( x \) to the graph \( G_i \) is essentially the sum of the increases caused by the edges \( xy \) being added.

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

\[
Q_{G'_i} - Q_{G_i} \leq (e + 1) \sum_{y \in Y} (Q_{G_i \cup \{xy\}} - Q_{G_i}).
\]

**Proof.** Since \( G'_i \setminus G_i = \{xy : y \in Y\} \) contains only edges incident to \( x \), the difference \( R_{G'_i}(z) - R_{G_i}(z) \) equals the number of solutions to \( z = \pm(x + y - a - b) \) such that \( y \in Y \) 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

\[
z = x + y - a - b = x + y' - a' - b'
\]

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, \( R_{G'_i}(z) - R_{G_i}(z) \leq 2 \) and

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

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

\[
\Delta_z \leq (e + 1) \cdot \exp(R_{G_i}(z)) \cdot (e - 1) = (e + 1) \left[ \exp(R_{G_i}(z) + 1) - \exp R_{G_i}(z) \right]
\]

\[
\leq (e + 1) \left[ \exp R_{G_i \cup \{xy^*\}}(z) - \exp R_{G_i}(z) \right] \leq (e + 1) \sum_{y \in Y} \left[ \exp R_{G_i \cup \{xy\}}(z) - \exp R_{G_i}(z) \right].
\]

(27)

Note that if \( R_{G'_i}(z) = R_{G_i}(z) \), then both the left-hand and the right-hand side of (27) are zero. In other words,

\[
\Delta_z \leq (e + 1) \sum_{y \in Y} \left[ \exp R_{G_i \cup \{xy\}}(z) - \exp R_{G_i}(z) \right]
\]
holds for all \( z \in [n] \). Consequently,

\[
Q_{G_i'} - Q_{G_i} = \sum_{z \in [n]} \Delta_z \leq \sum_{z \in [n]} (e + 1) \sum_{y \in Y} \left[ \exp(R_{G_i \cup \{xy\}}(z)) - \exp(R_{G_i}(z)) \right] = (e + 1) \sum_{y \in Y} (Q_{G_i \cup \{xy\}} - Q_{G_i}).
\]

Setting

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

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

\[
Q_{G_i'} \leq Q_{G_i} + (e + 1) \sum_{y \in Y} (Q_{G_i \cup \{xy\}} - Q_{G_i}) \leq Q_{G_i} + (e + 1) \sum_{y \in Y} U_{i,x+y}
\]

\[
\leq Q_{G_i} + (e + 1) \sum_{y \in Y} \frac{\lambda n^i e Q_{G_i}}{(e + 1)|T|(|T| + 1)} \leq Q_{G_i} \left(1 + \frac{\lambda n^i e}{|T| + 1}\right)
\]

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

Since our assumption that \( x \in \tilde{T}_i \) 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

\[
e(G'_i) = e(G_i) + |Y| < (1 - \alpha_i) \left(\frac{|T| + 1}{2}\right)
\]

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

\[
|Y| < (1 - \alpha_i) \left[\left(\frac{|T| + 1}{2}\right) - \left(\frac{|T|}{2}\right)\right] = (1 - \alpha_i)|T|.
\]

From the definition of \( Y \) in (28) and the fact that \( T \subset [n] \), \( x \in [n] \), it follows that \( Y = T \setminus (B_i - x) \).

Hence

\[
|(T + x) \cap B_i| = |T \cap (B_i - x)| = |T| - |Y| \geq \alpha_i |T|.
\]

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

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

\[
|H_T(A, B)| \geq \alpha_i |T| |A|.
\]

Together with Lemma 4.5, this yields the following corollary.

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

\[
|A| \geq s^{-2+8\varepsilon} n,
\]

19
we have
\[ e_{C_T}(A) = \Omega\left(\frac{\alpha_i^3|A|^2}{|T| n^\epsilon (\log n)^3}\right). \]  

**Proof.** Fix \( i \in \{0, \ldots, \lfloor 1/\epsilon \rfloor - 1\} \) and let \( B = B_i(T) \) be the set from Lemma 5.7. In particular, 
\[ |B| = O(|T|^4/(\lambda n^{i\epsilon})). \]  
Also let 
\[ \delta = \alpha_i. \]  

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 (13), we have 
\[ \delta^2 = \xi^2 = \xi^{2(2i+1) - 2} = \frac{\alpha_{i+1}}{\xi} > \frac{\alpha_{i+1}}{100\xi}. \]  

Our assumptions that \( i < \lfloor 1/\epsilon \rfloor \) and that \( T \) is \((\lambda, \epsilon)\)-bounded imply that \( T \) satisfies \( P_{\lambda, \epsilon, i+1} \) (see Definition 3.9). Moreover, we also assume that \( |T| \geq n^{1/100} \). Lemma 5.7 implies that every \( a \in A \subset \overline{T_i} \) satisfies 
\[ \deg_{H}(a) = |(T + a) \cap B| \geq \delta |T|. \]  

Finally, recalling that \( \lambda = s^{5-26\epsilon}/n \geq 1 \) (see (14)) and that \( s \geq n^{1/(5-26\epsilon)} \), we have the average degree \( D \) of the vertices in \( B \) satisfies 
\[ D = \frac{|H|}{|B|} \geq \frac{\delta |A||T|}{|B|} \geq \frac{\delta s^{2+8\epsilon} n \lambda n^{i\epsilon}}{|B|} = \Omega\left(\frac{\delta s^{2+8\epsilon} \lambda n^{i\epsilon}}{|T|^3}\right) \geq \Omega\left(\frac{\delta s^{3-17\epsilon} n^{i\epsilon}}{|T|^3}\right) \geq 5000, \]  
where we used that \( |T| < s^{1-6\epsilon} \) and that \( n \) is large. By Lemma 4.5 (with \( i + 1 \) in place of \( i \)), we have 
\[ e_{C_T}(A) = \Omega\left(\frac{\delta^2 |A|^2 |T|^2}{\lambda n^{(i+1)\epsilon} (\log n)^3}\right) \geq \Omega\left(\frac{\delta^2 |A|^2 |T|^2}{\lambda n^{(i+1)\epsilon} (\log n)^3} \cdot \frac{D \delta |T| |A|}{|B|}\right) \]  

Since \( |B| = O(|T|^4/(\lambda n^{i\epsilon})) \), the term \( |T|^4/|B| \lambda n^{i\epsilon} \) on the right hand side of (33) can be replaced by 1. Hence, from (32) and (33) it follows that (30) holds and the corollary is proved. \( \square \)

**Remark 5.12.** Having a sequence of graphs \( G \) and upper bounds on \( Q_G \) allows us to invoke Lemma 4.5 in the proof of Corollary 5.11. Recall that in Lemma 4.5 (and its proof in Section 6), we are essentially counting some special paths of length four in a bipartite graph. The advantage of considering these special paths is that we can bound their multiplicity (i.e., given two vertices, the number of special paths that connect them). But as the definition of special paths involves the graph \( G \), we now pay the price for the flexibility of allowing \( G \) to be a proper subgraph of \( K(T) \) in Definition 3.9 - it makes our path counting argument more complex. We also see that we need to work with a sequence of graphs \( G \), since the density of the graph \( G \) used in the definition of \( Q_G \) has to be so close to one that even the sparse graph \( H_T(A,B) \) in Lemma 4.5 admits many special paths. However, the lower bound \( \delta \) on the minimum degree of \( H_T(A,B) \) is derived from Lemma 5.7 and hence it depends on the density of the graph \( G \). If we were working with a single graph \( G \), it would put us in a position where the special paths could be rare or even non-existing, thus making the upper bound on \( Q_G \) useless. Our initial attempt at this argument did not involve the parameter \( i \).
in $P_{\lambda,\epsilon,i}$ and, of course, our failure to continue the proof motivated the more complex Definition 3.9 with this additional parameter.

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

**Claim 5.13.** The set $\overline{T}_{[1/\epsilon]}$ is empty.

**Proof.** Recall that $\overline{T}_{[1/\epsilon]}$ is the set of all $x$ such that $T \cup \{x\}$ is a $B_3$-set and there is no graph $G_{[1/\epsilon]} \subset (T \cup \{x\})$ satisfying both conditions $(a)$ and $(b)$ of Definition 3.9 with $i = [1/\epsilon]$ 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 $G_{[1/\epsilon]}$ be the complete graph $K(T \cup \{x\})$ on $T \cup \{x\}$, then conditions $(a)$ and $(b)$ hold. Condition $(a)$ follows immediately for $G_{[1/\epsilon]} = K(T \cup \{x\})$. As for $(b)$ observe first that, as a moment’s thought shows, we have $R_{G_{[1/\epsilon]}}(z) \leq |T| + 1$ for all $z$ since $T \cup \{x\}$ is a $B_3$-set. Hence

$$Q_{G_{[1/\epsilon]}} \leq n e^{|T|+1} \leq n e \exp\left(\frac{\lambda n}{|1/\epsilon|} \frac{|T|+1}{|T|} \sum_{j=1}^{\lceil 1/\epsilon \rceil} \frac{1}{j} \right),$$

which establishes $(b)$ and proves the claim.

The following claim is a simple consequence of Corollary 5.11 and Claim 5.13.

**Claim 5.14.** Suppose that $T$ is a $(\lambda, \epsilon)$-bounded set with cardinality at least $n^{1/100}$ and less than $s^{1-6\epsilon}$. For any set $A \subset \overline{T}$ with

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

we have

$$e_{C_T}(A) = \Omega\left(\frac{\alpha_{[1/\epsilon]-1}^3|A|^2}{|T| n^\epsilon (\log n)^3}\right).$$

**Proof.** Let $A \subset \overline{T}$ be as in the statement of the claim. By Claim 5.13 and the pigeonhole principle, there must be some $i \in \{0, 1, \ldots, \lfloor 1/\epsilon \rfloor - 1\}$ such that $|A \cap \overline{T}_i| \geq |A|/\lceil 1/\epsilon \rceil \geq s^{-2+8\epsilon} n$. Applying Corollary 5.11 to $A \cap \overline{T}_i$ in place of $A$ yields that

$$e_{C_T}(A) = \Omega\left(\frac{\alpha_i^3|A \cap \overline{T}_i|^2}{|T| n^\epsilon (\log n)^3}\right).$$

Since $\alpha_i \geq \alpha_{[1/\epsilon]-1}$ and $|A \cap \overline{T}_i| = \Theta(|A|)$, the claim follows.

We are finally ready to prove Proposition 3.11.

**Proof of Proposition 3.11.** In view of Claim 5.14, the graph $C_T[\overline{T}]$ satisfies (S) for

$$\beta = \Omega\left(\frac{\alpha_{[1/\epsilon]-1}}{|T| n^\epsilon (\log n)^3}\right) \quad \text{and} \quad R = \gamma n = \lceil 1/\epsilon \rceil s^{-2+8\epsilon} n \gg 1.$$
Set \(q = (\log n)/\beta\) and note that (10) 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

\[
q = O\left(|T| n^2 \left(\frac{\log n}{\alpha^3[1/\varepsilon]-1}\right) \leq O\left(s^{1-6\varepsilon} n^{5\varepsilon} (\log n)^{2[1/\varepsilon]+1}\right) = o(s).
\]

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 \((\lambda, \varepsilon)\)-bounded subset is at most

\[
\left(\frac{n^2}{q}\right) \left[\frac{1}{r}\right] (s^{-2+8\varepsilon}n) n^{-q-|T|} \leq n^{s-|T|} \left(\frac{1}{s^{-2+8\varepsilon}(s-q-|T|)}\right)^{s-q-|T|} \leq n^{s-|T|} \left(\frac{1}{s^{-3-8\varepsilon-o(1)}}\right)^{s}.
\]

In view of Lemma 4.1 a maximum \((\lambda, \varepsilon)\)-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

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

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

\[
\sum_{t=s^{1/7}}^{s^{1-6\varepsilon}} \left(\frac{n^2}{t}\right) n^{s-t} \left(\frac{1}{s^{-3-8\varepsilon-o(1)}}\right)^{s} \leq \left(\frac{n}{s^{-3-8\varepsilon-o(1)}}\right)^{s}.
\]

This completes the proof of Proposition 3.11 \(\square\)

6. PROOF OF LEMMA 4.5

Fix an integer \(i \geq 0\) and real numbers \(\varepsilon > 0, \lambda \geq 1, D \geq 5000\) and \(\delta \in (0, 1]\) satisfying \(\delta^2 \geq \alpha_i/100\varepsilon\). Suppose that a set \(T \subset [n]\) with at least \(n^{1/100}\) elements satisfies \(P_{\lambda, \varepsilon, i}\). 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 \(P_{\lambda, \varepsilon, i}\) 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) \(\{a - b, b' - a'\}\) and \(\{b - a', b' - a''\}\) are edges of \(G_i\), and

(c) the differences \(b - a, b' - a', b - a', b' - a''\) are all distinct.

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

\[
(z_1, z_2, z_3, z_4) \in G_i^2, \quad a'' - a = (z_1 + z_2) - (z_3 + z_4) \quad \text{and the } z_i\text{s are all distinct. (34)}
\]

Any solution to (34) 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 (11)) that the number of solutions to (34) is exactly \(4R_{G_i}(a'' - a)\).

(For completeness, we remark that not all solutions need to define paths in \(H\) since \(A\) and \(B\) are
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 $\{a, b\}$.

just subsets of $[n]$ and $[2n]$.) We conclude that the total number $N$ of special paths in $H$ satisfies

$$ N = O(|\tilde{G}_G|) $$

(see Definition 3.5). Recalling (12), given that $G_i$ satisfies (b) of Definition 3.9 we have

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

In view of Observation 3.6, inequalities (35) and (36) 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

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

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

1. $\text{deg}_{G_i}(b - a) \geq (1 - 4\alpha_i/\delta)|T|$ for every $(a, b) \in H'$;
2. $|H'| \geq |H|/8 \log n$;
3. $\text{deg}_{H'}(a) \geq \delta|T|/16 \log n$ for every $a \in A'$;
4. $d \leq \text{deg}_{H'}(b) \leq 12d$ 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

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

\[
\{b-a, b'-a'\}, \{b'-a'', b-a\} \in G_i,
\]

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

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

\[
\frac{4\alpha_i}{\delta} |T| \leq 400 \xi \delta |T| \leq \frac{13}{150} \deg_{H'}(a').
\]

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

\[
X = \{b' \in N_{H'}(a') \setminus \{b\} : \{b-a, b'-a'\} \in G_i\}
\]

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 for an illustration.

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

\[
N_P \geq \sum_{b' \in X} (\deg_{H'}(b') - 4) \geq \left(1 - \frac{4}{d}\right) \sum_{b' \in X} \deg_{H'}(b') \geq 0.98 \sum_{b' \in X} \deg_{H'}(b'), \quad (37)
\]

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

\[
\sum_{b' \in N_{H'}(a')} \deg_{H'}(b') = \sum_{b' \in X} \deg_{H'}(b') + \sum_{b' \in X^c} \deg_{H'}(b').
\]

Since the degrees in \( B' \) are all in \([d, 12d]\), we get

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

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

\[
\left(1 + \frac{12}{99}\right) \sum_{b' \in X} \deg_{H'}(b') \overset{(37)}{\leq} \left(1 + \frac{12}{99}\right) \frac{100}{98} N_P < \frac{4}{3} N_P.
\]

Let \( N_A \) 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 \( 3N_A/4 \) 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

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

whence the claim follows. \hfill \Box

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

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

Note that

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

On the other hand, it follows from the assumption on $G_i$ (see Definition 3.5(a)) that

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

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

$$|L| \leq \frac{\delta}{2}|T| \leq \frac{|H|}{2|B|}, \quad (38)$$

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 (38) and the assumption of the lemma that

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

Since the graph $H'$ that we construct in what follows is a subgraph of $H^*$, it will satisfy (6).

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

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

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

$$\frac{e(H^*) - e(H)/4}{\log |T|} \geq \frac{e(H)}{4 \log n} \quad (39)$$

edges of $H^*$ incident to $B_j$.  

25
Set \( d = (D/16)e^{j-1} \), and \( \hat{H} = H^*[A \cup B_j] \). Since we assume that \( H \) satisfies (I) it follows from (39) that
\[
e(\hat{H}) \geq \frac{\delta |T| |A|}{4 \log n}.
\]
In particular, the average degree of vertices of \( A \) in \( \hat{H} \) 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 \( \hat{H} \), together with all the incident edges. The number of edges deleted by this procedure is at most
\[
|A| \frac{\delta}{16 \log n} |T| + |B| d \leq \frac{1}{2} e(\hat{H}).
\]
Hence, at least \( e(\hat{H})/2 \geq e(H)/8 \log n \) 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) \( \square \)

7. Concluding remarks

Throughout the paper, we assumed that \( \delta \) is an arbitrary, but fixed, positive real number. Our argument allows one to take some function \( \delta = \delta(n) \) with \( \delta \to 0 \) as \( n \to \infty \). Nevertheless, 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 \( h \geq 2 \) be an integer. Suppose \( 0 \leq a \leq 1 \) is a fixed constant and \( m = m(n) = n^{a+o(1)} \). Then, asymptotically almost surely, we have
\[
F_h([n]_m) = n^{b+o(1)},
\]
where \( b = b(a) \) is given by
\[
b(a) = \begin{cases} 
  a & \text{for } 0 \leq a \leq 1/(2h - 1), \\
  1/(2h - 1) & \text{for } 1/(2h - 1) \leq a \leq h/(2h - 1), \\
  a/h & \text{for } h/(2h - 1) \leq a \leq 1.
\end{cases}
\]

The fact that \( b(a) \) is at least as large as stated in (40) 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 (40). The case \( h = 2 \) of Conjecture 7.1 is proved in [10, 11] and we established the case \( h = 3 \) in this paper.

**References**


26


Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA (D. Dellamonica Jr., Y. Kohayakawa and V. Rödl) 
E-mail address: domingos.junior@gmail.com, rodl@mathcs.emory.edu

Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, Brazil (Y. Kohayakawa) 
E-mail address: yoshi@ime.usp.br

Department of Mathematics, Duksung Women’s University, Seoul, South Korea 01369 (S. J. Lee) 
E-mail address: sanglee242@duksung.ac.kr, 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