

1 THE NUMBER OF SIDON SETS AND THE MAXIMUM SIZE OF SIDON  
2 SETS CONTAINED IN A SPARSE RANDOM SET OF INTEGERS

3 YOSHIHARU KOHAYAKAWA, SANG JUNE LEE, VOJTĚCH RÖDL, AND WOJCIECH SAMOTIJ

ABSTRACT. A set  $A$  of non-negative integers is called a *Sidon set* if all the sums  $a_1 + a_2$ , with  $a_1 \leq a_2$  and  $a_1, a_2 \in A$ , are distinct. A well-known problem on Sidon sets is the determination of the maximum possible size  $F(n)$  of a Sidon subset of  $[n] = \{0, 1, \dots, n - 1\}$ . Results of Chowla, Erdős, Singer and Turán from the 1940s give that  $F(n) = (1 + o(1))\sqrt{n}$ . We study Sidon subsets of sparse random sets of integers, replacing the ‘dense environment’  $[n]$  by a sparse, random subset  $R$  of  $[n]$ , and ask how large a subset  $S \subset R$  can be, if we require that  $S$  should be a Sidon set.

Let  $R = [n]_m$  be a random subset of  $[n]$  of cardinality  $m = m(n)$ , with all the  $\binom{n}{m}$  subsets of  $[n]$  equiprobable. We investigate the random variable  $F([n]_m) = \max |S|$ , where the maximum is taken over all Sidon subsets  $S \subset [n]_m$ , and obtain quite precise information on  $F([n]_m)$  for the whole range of  $m$ , as illustrated by the following abridged version of our results. Let  $0 \leq a \leq 1$  be a fixed constant and suppose  $m = m(n) = (1 + o(1))n^a$ . We show that there is a constant  $b = b(a)$  such that, almost surely, we have  $F([n]_m) = n^{b+o(1)}$ . As it turns out, the function  $b = b(a)$  is a continuous, piecewise linear function of  $a$  that is non-differentiable at two ‘critical’ points:  $a = 1/3$  and  $a = 2/3$ . Somewhat surprisingly, between those two points, the function  $b = b(a)$  is constant.

Our approach is based on estimating the number of Sidon sets of a given cardinality contained in  $[n]$ . Our estimates also directly address a problem raised by Cameron and Erdős [*On the number of sets of integers with various properties*, Number theory (Banff, AB, 1988), de Gruyter, Berlin, 1990, pp. 61–79].

4 1. INTRODUCTION

5 Recent years have witnessed vigorous research in the classical area of additive combinatorics. An  
6 attractive feature of these developments is that applications in theoretical computer science have  
7 motivated some of the striking research in the area (see, e.g., [32]). For a modern treatment of the  
8 subject, the reader is referred to [31].

9 Among the best known concepts in additive number theory is the notion of a *Sidon set*. A set  $A$   
10 of non-negative integers is called a *Sidon set* if all the sums  $a_1 + a_2$ , with  $a_1 \leq a_2$  and  $a_1, a_2 \in A$ ,  
11 are distinct. A well-known problem on Sidon sets is the determination of the maximum possible  
12 size  $F(n)$  of a Sidon subset of  $[n] = \{0, 1, \dots, n - 1\}$ . In 1941, Erdős and Turán [12] showed  
13 that  $F(n) \leq \sqrt{n} + O(n^{1/4})$ . In 1944, Chowla [8] and Erdős [11], independently of each other,  
14 observed that a result of Singer [29] implies that  $F(n) \geq \sqrt{n} - O(n^{5/16})$ . Consequently, it is known  
15 that  $F(n) = (1 + o(1))\sqrt{n}$ . For a wealth of related material, the reader is referred to the classical

---

*Date:* Fri 16<sup>th</sup> Dec, 2011, 8:42pm.

The first author was partially supported by CNPq (Proc. 308509/2007-2 and 484154/2010-9) and he is grateful to NUMEC/USP, Núcleo de Modelagem Estocástica e Complexidade of the University of São Paulo, and Project MaCLinC/USP, for supporting this research. The third author was supported by the NSF grant DMS 0800070. The fourth author was partially supported by ERC Advanced Grant DMMCA and a Trinity College JRF.

Parts of this work appeared in preliminary form in SODA 2011.

16 monograph of Halberstam and Roth [15] and to a recent survey by O’Bryant [24] and the references  
17 therein.

18 We investigate Sidon sets contained in *random sets of integers*, and obtain essentially tight bounds  
19 on their relative density. Our approach is based on finding upper bounds for the number of Sidon  
20 sets of a given cardinality contained in  $[n]$ . Besides being the key to our probabilistic results, our  
21 upper bounds also address a problem of Cameron and Erdős [7].

22 We discuss our bounds on the number of Sidon sets and our probabilistic results in the next two  
23 subsections.

24 **1.1. A problem of Cameron and Erdős.** Let  $\mathcal{Z}_n$  be the family of Sidon sets contained in  $[n]$ .  
25 Over two decades ago, Cameron and Erdős [7] proposed the problem of estimating  $|\mathcal{Z}_n|$ . Observe  
26 that one trivially has

$$2^{F(n)} \leq |\mathcal{Z}_n| \leq \sum_{1 \leq i \leq F(n)} \binom{n}{i} = n^{(1/2+o(1))\sqrt{n}}. \quad (1)$$

27 Cameron and Erdős [7] improved the lower bound in (1) by showing that  $\limsup_n |\mathcal{Z}_n| 2^{-F(n)} = \infty$   
28 and asked whether the upper bound could also be strengthened. Our result is as follows.

29 **Theorem 1.1.** *There is a constant  $c$  for which  $|\mathcal{Z}_n| \leq 2^{cF(n)}$ .*

30 Our proof method gives that the constant  $c$  in Theorem 1.1 may be taken to be arbitrarily close  
31 to  $\log_2(32e) = 6.442 \dots$  (for large enough  $n$ ). We do not make any attempts to optimize this  
32 constant as it seems that our approach cannot yield a sharp estimate for  $\log_2 |\mathcal{Z}_n|$ . It remains an  
33 interesting open question whether  $\log_2 |\mathcal{Z}_n| = (1 + o(1))F(n)$ .

34 **1.2. Probabilistic results.** We investigate Sidon subsets of sparse, *random* sets of integers, that  
35 is, we replace the ‘environment’  $[n]$  by a sparse, random subset  $R$  of  $[n]$ , and ask how large a  
36 subset  $S \subset R$  can be, if we require that  $S$  should be a Sidon set.

37 Investigating how classical extremal results in ‘dense’ environments transfer to ‘sparse’ settings has  
38 proved to be a deep line of research. A fascinating example along these lines occurs in the celebrated  
39 work of Tao and Green [14], where Szemerédi’s classical theorem on arithmetic progressions [30]  
40 is transferred to certain sparse, pseudorandom sets of integers and to the set of primes themselves  
41 (see [25, 26, 31] for more in this direction). Much closer examples to our setting are a version of  
42 Roth’s theorem on 3-term arithmetic progressions [27] for random subsets of integers [22], and the  
43 far reaching generalizations due to Conlon and Gowers [9] and Schacht [28]. For the sake of brevity,  
44 we shall not discuss this further and refer the reader to [9], [28], [16, Chapter 8] and [20, Section 4].

45 Let us now state a weak, but less technical version of our main probabilistic results. Let  $F(R) =$   
46  $\max |S|$ , where the maximum is taken over all Sidon subsets  $S \subset R$ . Let  $[n]_m$  be a random subset  
47 of  $[n]$  of cardinality  $m = m(n)$ , with all the  $\binom{n}{m}$  subsets of  $[n]$  equiprobable. We are interested in  
48 the random variable  $F([n]_m)$ .

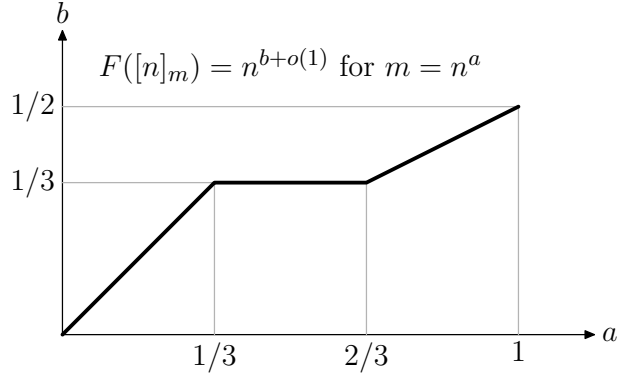


FIGURE 1. The graph of  $b = b(a)$

49 Standard methods give that, almost surely, that is, with probability tending to 1 as  $n \rightarrow \infty$ , we  
 50 have  $F([n]_m) = (1 - o(1))m$  if  $m = m(n) \ll n^{1/3}$ . On the other hand, the results of Schacht [28]  
 51 and Conlon and Gowers [9] imply that, if  $m = m(n) \gg n^{1/3}$ , then, almost surely, we have

$$F([n]_m) = o(m). \quad (2)$$

52 Thus  $F([n]_m)$  undergoes a sudden change of behaviour at  $m = n^{1/3+o(1)}$ . The following abridged  
 53 version of our results already gives us quite precise information on  $F([n]_m)$  for the whole range  
 54 of  $m$ .

55 **Theorem 1.2.** *Let  $0 \leq a \leq 1$  be a fixed constant. Suppose  $m = m(n) = (1 + o(1))n^a$ . There exists*  
 56 *a constant  $b = b(a)$  such that almost surely*

$$F([n]_m) = n^{b+o(1)}. \quad (3)$$

57 *Furthermore,*

$$b(a) = \begin{cases} a & \text{if } 0 \leq a \leq 1/3, \\ 1/3 & \text{if } 1/3 \leq a \leq 2/3, \\ a/2 & \text{if } 2/3 \leq a \leq 1. \end{cases} \quad (4)$$

58 Thus, the function  $b = b(a)$  is piecewise linear. The graph of  $b = b(a)$  is given in Figure 1. The  
 59 point  $(a, b) = (1, 1/2)$  in the graph is clear from the Erdős–Turán and Chowla results [8, 11, 12]  
 60 mentioned above. The behaviour of  $b = b(a)$  in the interval  $0 \leq a \leq 1/3$  is not hard to establish.  
 61 The fact that the point  $(1/3, 1/3)$  could be an interesting point in the graph is suggested by the  
 62 results of Schacht [28] and Conlon and Gowers [9]. It is somewhat surprising that, besides the  
 63 point  $a = 1/3$ , there is a second value at which  $b = b(a)$  is ‘critical’, namely,  $a = 2/3$ . Finally,  
 64 we find it rather interesting that  $b = b(a)$  should be *constant* between those two critical points.  
 65 We state our results in full in Section 2.1. Our results in fact determine  $F([n]_m)$  up to a *constant*  
 66 multiplicative factor for  $m \leq n^{2/3-\delta}$  for any fixed  $\delta > 0$  and for  $m \geq n^{2/3}(\log n)^{8/3}$ . For the missing  
 67 range of  $m$ , around  $n^{2/3}$ , our lower and upper bounds differ by a factor of  $O((\log n)/\log \log n)$ .

69 **2.1. Statement of the main results.** We prove a more detailed result than Theorem 1.1.  
 70 Let  $\mathcal{Z}_n(t)$  be the family of Sidon sets of cardinality  $t$  contained in  $[n]$ .

71 **Theorem 2.1.** *Let  $0 < \sigma < 1$  be a real number. For any large enough  $n$  and  $t \geq 2s_0$ , where  $s_0 =$   
 72  $(2(1 - \sigma)^{-1}n \log n)^{1/3}$ , we have*

$$|\mathcal{Z}_n(t)| \leq n^{3s_0} \left( \frac{32en}{\sigma t^2} \right)^t. \quad (5)$$

73 Theorem 1.1 follows from Theorem 2.1 by summing over all  $t$  (see Section 3.2). Our next result  
 74 covers values of  $t$  smaller than the ones covered in Theorem 2.1.

75 **Theorem 2.2.** *Let  $n$  and  $t$  be integers with*

$$3 \cdot 2^3 n^{1/3} \leq t \leq 4(n \log n)^{1/3}. \quad (6)$$

76 *Then*

$$|\mathcal{Z}_n(t)| \leq \left( \frac{6 \cdot 2^{3/2} n}{t} \exp \left( -\frac{t^3}{3 \cdot 2^7 n} \right) \right)^t. \quad (7)$$

77 Let us now turn to our probabilistic results. Instead of working with the *uniform model*  $[n]_m$  of  
 78 random subsets of  $[n]$ , it will be more convenient to work with the so called *binomial model*  $[n]_p$ ,  
 79 in which each element of  $[n]$  is put in  $[n]_p$  with probability  $p$ , independently of all other elements.  
 80 Routine methods allow us to translate our results on  $[n]_p$  below to the corresponding results on  $[n]_m$ ,  
 81 where  $p = m/n$  (see Section 2.2 for details).

82 We state our results on  $F([n]_p)$  split into theorems covering different ranges of  $p = p(n)$ . Our first  
 83 result corresponds to the range  $0 \leq a \leq 1/3$  in Theorem 1.2.

84 **Theorem 2.3.** *For  $n^{-1} \ll p = p(n) \ll n^{-2/3}$ , we almost surely have*

$$F([n]_p) = (1 + o(1))np. \quad (8)$$

85 *For  $n^{-1} \ll p \leq 2n^{-2/3}$ , we almost surely have*

$$\left( \frac{1}{3} + o(1) \right) np \leq F([n]_p) \leq (1 + o(1))np, \quad (9)$$

86 **Remark 2.4.** *One may in fact prove the following result: if  $p = \gamma n^{-2/3}$  for some constant  $\gamma$ , then*

$$\left( 1 - \frac{1}{12} \gamma^3 + o(1) \right) np \leq F([n]_p) \leq \left( 1 - \frac{1}{12} \gamma^3 + \frac{1}{12} \gamma^6 + o(1) \right). \quad (10)$$

87 Our next result covers the range  $1/3 \leq a < 2/3$  in Theorem 1.2.

88 **Theorem 2.5.** *For any  $\delta > 0$ , there is a positive constant  $c_2 = c_2(\delta)$  such that if  $2n^{-2/3} \leq p =$   
 89  $p(n) \leq n^{-1/3-\delta}$ , then we almost surely have*

$$c_1 (n \log(n^2 p^3))^{1/3} \leq F([n]_p) \leq c_2 (n \log(n^2 p^3))^{1/3}, \quad (11)$$

90 where  $c_1$  is a positive absolute constant.

91 We now turn to the point  $a = 2/3$  in Theorem 1.2.

92 **Theorem 2.6.** For any  $0 \leq \delta < 1/3$ , there is a positive constant  $c_3 = c_3(\delta)$  such that if  $1 \leq \alpha =$   
 93  $\alpha(n) \leq n^\delta$  and  $p = p(n) = \alpha^{-1}n^{-1/3}(\log n)^{2/3}$ , then we almost surely have

$$c_3(n \log n)^{1/3} \leq F([n]_p) \leq c_4(n \log n)^{1/3} \frac{\log n}{\log(\alpha + \log n)},$$

94 where  $c_4$  is an absolute constant.

95 We remark that Theorems 2.5 and 2.6 consider ranges that overlap (functions  $p = p(n)$  of the  
 96 form  $n^{-1/3-\delta'}$  for some  $0 < \delta' < 1/3$  are covered by both theorems). Finally, we consider the  
 97 range  $2/3 \leq a \leq 1$  in Theorem 1.2.

98 **Theorem 2.7.** There exist positive absolute constants  $c_5$  and  $c_6$  for which the following holds.  
 99 If  $1 \leq \alpha = \alpha(n) \leq (\log n)^2$  and  $p = p(n) = \alpha^{-1}n^{-1/3}(\log n)^{8/3}$ , then we almost surely have

$$c_5\sqrt{np} \leq F([n]_p) \leq c_6\sqrt{np} \cdot \frac{\sqrt{\alpha}}{1 + \log \alpha}.$$

100 Furthermore, if  $n^{-1/3}(\log n)^{8/3} \leq p = p(n) \leq 1$ , then, almost surely,

$$c_5\sqrt{np} \leq F([n]_p) \leq c_6\sqrt{np}.$$

101 **2.2. The uniform model and the binomial model.** We now discuss how to translate Theo-  
 102 rems 2.3, 2.5–2.7 on  $[n]_p$  in Section 2.1 to the corresponding results on  $[n]_m$ . Before we proceed,  
 103 let us make the following definition.

104 **Definition 2.8.** We shall say that an event in the probability space of the random sets  $[n]_p$  or  
 105 in the probability space of the random sets  $[n]_m$  holds with overwhelming probability, abbreviated  
 106 as *w.o.p.*, if the probability of failure of that event is  $O(n^{-C})$  for any constant  $C$ , that is, if the  
 107 probability of failure is ‘superpolynomially’ small.

108 For us, the following consequence of Pittel’s inequality (see, e.g., [6, p. 35] and [17, p. 17]) will  
 109 suffice for translating results on  $[n]_p$  to results on  $[n]_m$ .

110 **Lemma 2.9.** Let  $1 \leq m = m(n) < n$  and  $p = p(n)$  be such that  $p = m/n$ . Let  $P$  be an event in  
 111 the probability space of the random sets  $[n]_p$ . If  $[n]_p$  is in  $P$  w.o.p., then  $[n]_m$  is in  $P \cap \binom{[n]}{m}$  w.o.p.

112 *Proof.* Let  $Q$  be the complement of  $P$ . We shall show that, for any constant  $C > 0$ , there exists  
 113 a constant  $C' > 0$ , where  $C' \rightarrow \infty$  as  $C \rightarrow \infty$ , such that the following holds. If  $\mathbb{P}\left[[n]_p \text{ is in } Q\right] =$   
 114  $O(n^{-C})$ , then  $\mathbb{P}\left[[n]_m \text{ is in } Q \cap \binom{[n]}{m}\right] = O(n^{-C'})$ .

115 Pittel’s inequality (see [6, p. 35] and [17, p. 17]) states that

$$\mathbb{P}\left[[n]_m \text{ is in } Q \cap \binom{[n]}{m}\right] = O(\sqrt{m}) \cdot \mathbb{P}\left[[n]_p \text{ is in } Q\right]. \quad (12)$$

116 Since, by hypothesis,  $\mathbb{P}[[n]_p \text{ is in } Q] = O(n^{-C})$  holds for any constant  $C > 0$ , inequality (12)  
 117 implies that

$$\mathbb{P}[[n]_m \text{ is in } Q \cap \binom{[n]}{m}] = O(\sqrt{m} \cdot n^{-C}) = O(\sqrt{n} \cdot n^{-C}) = O(n^{-C+1/2}),$$

118 which completes the proof of Lemma 2.9. □

119 Every result in Theorems 2.5–2.7 will be proved with ‘w.o.p.’ rather than with ‘almost surely’.  
 120 By Lemma 2.9, we can translate each such result on  $[n]_p$  to the corresponding result on  $[n]_m$ ,  
 121 where  $p = m/n$ . For example, Theorem 2.5 implies the following uniform version: *For any  $\delta > 0$ ,*  
 122 *there is a positive constant  $c_2 = c_2(\delta)$  such that if  $2n^{1/3} \leq m = m(n) \leq n^{2/3-\delta}$ , then, with*  
 123 *overwhelming probability, we have*

$$c_1 \left( n \log \frac{m^3}{n} \right)^{1/3} \leq F([n]_m) \leq c_2 \left( n \log \frac{m^3}{n} \right)^{1/3},$$

124 where  $c_1$  is a positive absolute constant.

125 Finally, we remark that one may use the usual deletion method to prove that the result on  $[n]_m$   
 126 corresponding to Theorem 2.3 holds almost surely.

127 **2.3. Organization.** Our results on the number of Sidon sets are proved in Section 3. In Section 4,  
 128 we consider the upper bounds in Theorems 2.5–2.7. Section 5 contains some preparatory lemmas  
 129 for the proof of Theorems 2.3 and for the proofs of the lower bounds in Theorems 2.5–2.7. The  
 130 proof of Theorem 2.3 is given in Section 6. In Section 7, we give the proofs of the lower bounds in  
 131 Theorems 2.5–2.7.

132 For simplicity, we omit ‘floor’ and ‘ceiling’ symbols in our formulae, when they are not essential.  
 133 For the sake of clarity of the presentation, we often write  $a/bc$  instead of the less ambiguous  $a/(bc)$ .

### 134 3. THE NUMBER OF SIDON SETS

135 The proofs of Theorems 2.1 and 2.2 are based on a method introduced by Kleitman and Winston [19]  
 136 (see [2, 4, 5, 13, 21] for other applications of this method).

137 **3.1. Independent sets in locally dense graphs.** We start with the following lemma, which  
 138 gives an upper bound for the number of independent sets in graphs that are ‘locally dense’.

139 **Lemma 3.1.** *Let  $G$  be a graph on  $N$  vertices, let  $q$  be an integer and let  $0 \leq \beta \leq 1$  and  $R$  be real*  
 140 *numbers with*

$$R \geq e^{-\beta q} N. \tag{13}$$

141 *Suppose the number of edges  $e(U)$  induced in  $G$  by any set  $U \subset V(G)$  with  $|U| \geq R$  satisfies*

$$e(U) \geq \beta \binom{|U|}{2}. \tag{14}$$

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

$$\binom{N}{q} \binom{R}{r}. \quad (15)$$

143 *Proof.* Fix an integer  $r \geq 0$ . We describe a deterministic algorithm that associates to every inde-  
 144 pendent set  $I$  of size  $q + r$  in  $G$  a pair  $(S_0, A)$  of disjoint sets with  $S_0 \subset I \subset S_0 \cup A \subset V(G)$  and  
 145 with  $|S_0| = q$  and  $|A| \leq R$ . Furthermore, if, for some inputs  $I$  and  $I'$ , the algorithm outputs  $(S_0, A)$   
 146 and  $(S'_0, A')$  with  $S_0 = S'_0$ , then  $A = A'$ . A moment's thought now reveals that the number of  
 147 independent sets in  $G$  with  $q + r$  elements is at most as given in (15), as claimed. We now proceed  
 148 to describe the algorithm.

149 At all times, our algorithm maintains a partition of  $V(G)$  into sets  $S$ ,  $X$ , and  $A$  (short for *selected*,  
 150 *excluded*, and *available*). As the algorithm evolves,  $S$  increases,  $X$  increases and  $A$  decreases. The  
 151 vertices in  $A$  will be labelled  $v_1, \dots, v_{|A|}$ , where, for every  $i$ , the vertex  $v_i$  has maximum degree  
 152 in  $G[\{v_i, \dots, v_{|A|}\}]$  (the graph induced by  $\{v_i, \dots, v_{|A|}\}$  in  $G$ ); we break ties arbitrarily by giving  
 153 preference to vertices that come earlier in some arbitrary predefined ordering of  $V(G)$ .

154 We start the algorithm with  $A = V(G)$  and  $S = X = \emptyset$ . Crucially, at all times we maintain  $S \subset$   
 155  $I \subset S \cup A$ . The algorithm works as follows. While  $|S| < q$ , we repeat the following. Let  $a = |A|$   
 156 and suppose  $A = \{v_1, \dots, v_a\}$ , with the vertex labelling convention described above. Let  $i$  be the  
 157 smallest index such that  $v_i$  belongs to our independent set  $I$ , move  $v_1, \dots, v_{i-1}$  from  $A$  to  $X$  (they  
 158 are not in  $I$  by the choice of  $i$ ), and move  $v_i$  from  $A$  to  $S$  ( $v_i$  is in  $I$ ). Observe that  $A$  has already  
 159 lost  $i$  members in this iteration and  $S$  has gained one. Let  $U = \{v_i, \dots, v_a\}$ . If  $|U| \geq R$ , we further  
 160 move all neighbours of  $v_i$  in  $A$  to  $X$  (since  $I$  is an independent set and  $v_i \in I$ ). Otherwise, i.e.,  
 161 if  $|U| < R$ , consider the first  $q - |S|$  members  $v_{i_1}, \dots, v_{i_{q-|S|}}$  of  $I \cap A$  and move them from  $A$  to  $S$   
 162 (note that  $i < i_1 < \dots < i_{q-|S|} \leq a$  and we now have  $|S| = q$ ).

163 The procedure above defines an increasing sequence of sets  $S$ . Once we obtain a set  $S$  with  $|S| = q$ ,  
 164 we let  $S_0 = S$ , output  $(S_0, A)$  and stop the algorithm. Inspection shows that  $A$  depends only on  $S_0$   
 165 and not on  $I$ , i.e., if  $(S_0, A)$  and  $(S_0, A')$  are both outputs by the algorithm (for some inputs  $I$   
 166 and  $I'$ ), then  $A = A'$ . We now use our assumption on  $G$  to show that  $|A| \leq R$ .

167 We consider two cases: The first case is the case in which the body of the while loop of the algorithm  
 168 is executed with  $|U| < R$  at an iteration. The second case is the case in which we have  $|U| \geq R$   
 169 during the  $q$  iterations of the while loop. Observe that one of two cases must occur.

170 First, we consider the first case. At the iteration with  $|U| < R$ , the set  $A$  lost the first  $i$  vertices  
 171 (and possibly others) and hence at the end of this iteration we have  $|A| \leq a - i = |U| - 1 < R$ .  
 172 Moreover,  $|S|$  becomes of cardinality  $q$  and the algorithm stops.

173 Next, we consider the second case in which we have  $|U| \geq R$  during the  $q$  iterations of the while loop.  
 174 In each iteration, consider an execution of the body of the while loop of the algorithm when  $|U| \geq R$   
 175 and (only) the vertex  $v_i$  is moved to  $S$ . In this execution,  $A$  loses, in total,  $i + d(v_i, U)$  vertices,  
 176 where  $d(v_i, U)$  is the degree of  $v_i$  in the graph  $G[U]$ . Recall that we are considering the case  $|U| \geq R$

177 and that  $v_i$  has maximum degree in the graph  $G[U]$ . Applying (14), we see that  $d(v_i, U) \geq \beta(|U| - 1)$ .  
 178 Therefore, at the end of this iteration,  $A$  has cardinality

$$a - (i + d(v_i, U)) \leq a - (a - |U| + 1 + \beta(|U| - 1)) \leq |U| - \beta|U| \leq (1 - \beta)a.$$

179 In the second case, the cardinality of  $A$  decreases by a factor of  $1 - \beta$  in the  $q$  iterations of the  
 180 while loop and, at the end,  $A$  has at most  $N(1 - \beta)^q \leq Ne^{-\beta q} \leq R$  elements.  $\square$

181 **3.2. Proof of Theorem 2.1.** We derive Theorem 2.1 from the following lemma.

182 **Lemma 3.2.** *Let  $n$ ,  $s$  and  $q$  be integers and let  $0 < \sigma < 1$  be a real number such that*

$$\frac{s^2 q}{n} \geq \frac{2}{1 - \sigma} \log \frac{\sigma s}{2}. \quad (16)$$

183 *Then, for any integer  $r \geq 0$ , we have*

$$|\mathcal{Z}_n(s + q + r)| \leq |\mathcal{Z}_n(s)| \binom{n}{q} \binom{2n/\sigma s}{r}. \quad (17)$$

184 To obtain the bound for  $|\mathcal{Z}_n(t)|$  in Theorem 2.1, we apply Lemma 3.2 iteratively.

185 *Proof of Theorem 2.1.* Fix integers  $n$  and  $t$ , with  $t \geq 2s_0$ , where  $s_0$  is as given in the statement  
 186 of our theorem, that is,  $s_0 = (2(1 - \sigma)^{-1}n \log n)^{1/3}$ . We may clearly suppose that  $t \leq F(n) =$   
 187  $(1 + o(1))\sqrt{n}$ , as otherwise  $\mathcal{Z}_n(t) = \emptyset$ . Let  $K$  be the largest integer satisfying  $t2^{-K} \geq 2s_0$ . We  
 188 define three sequences  $(s_k)_{0 \leq k \leq K}$ ,  $(q_k)_{0 \leq k \leq K}$  and  $(r_k)_{0 \leq k \leq K}$  as follows. We let  $q_0 = s_0$  and  $r_0 =$   
 189  $t2^{-K} - s_0 - q_0$ . Moreover, we let  $s_1 = t2^{-K} \geq 2s_0$ ,  $q_1 = q_0/4$  and  $r_1 = t2^{-K+1} - s_1 - q_1$ .  
 190 For  $k = 2, \dots, K$ , we let  $s_k = 2s_{k-1} = t2^{-K+k-1}$ ,  $q_k = q_{k-1}/4 = q_0 4^{-k}$  and  $r_k = t2^{-K+k} - s_k - q_k$ .  
 191 We apply Lemma 3.2 with parameters  $s_k$ ,  $q_k$  and  $r_k$  for  $k = 0, \dots, K$ , to obtain from (17) that

$$|\mathcal{Z}_n(t2^{-K+k})| = |\mathcal{Z}_n(s_k + q_k + r_k)| \leq |\mathcal{Z}_n(s_k)| \binom{n}{q_k} \binom{2n/\sigma s_k}{r_k} \quad (18)$$

192 for all  $k$ . It suffices to check (16) to justify these applications of Lemma 3.2. Since  $s_k^2 q_k \geq s_0^2 q_0 =$   
 193  $2(1 - \sigma)^{-1}n \log n > 2(1 - \sigma)^{-1}n \log(\sigma s_k/2)$  for all  $0 \leq k \leq K$ , inequality (16) holds for  $n$ ,  $s_k$  and  $q_k$ .  
 194 Using that  $s_k = s_{k-1} + q_{k-1} + r_{k-1} = t2^{-K+k-1}$  for  $k \geq 1$  and that  $|\mathcal{Z}_n(s_0)| \leq \binom{n}{s_0}$ , we obtain  
 195 from (18) that

$$|\mathcal{Z}_n(t)| \leq \binom{n}{s_0} \prod_{0 \leq k \leq K} \binom{n}{q_k} \prod_{0 \leq k \leq K} \binom{2n/\sigma s_k}{r_k}. \quad (19)$$

196 Note that

$$\binom{n}{s_0} \leq \left(\frac{en}{s_0}\right)^{s_0} \leq n^{2s_0/3} \quad (20)$$

197 and that

$$\prod_{0 \leq k \leq K} \binom{n}{q_k} \leq n^{\sum_{0 \leq k \leq K} q_k} \leq n^{q_0 \sum_{0 \leq k \leq K} 1/4^k} \leq n^{4q_0/3} = n^{4s_0/3}. \quad (21)$$

198 We now proceed to estimate the last factor of the right-hand side of (19). First note that, by the  
 199 choice of  $K$ , we have  $(r_0 + s_0 + q_0)/2 = t2^{-K-1} < 2s_0$ , and hence  $r_0 < 2s_0$ . Therefore, we have



$$\binom{2n/\sigma s_0}{r_0} \leq \left(\frac{2en}{\sigma s_0 r_0}\right)^{r_0} \leq n^{r_0/3} \leq n^{2s_0/3} \leq n^{s_0} \quad (22)$$

200 for all large  $n$ . We now note that

$$\prod_{1 \leq k \leq K} \binom{2n/\sigma s_k}{r_k} = \prod_{1 \leq k \leq K} \binom{2n/\sigma s_{K-k+1}}{r_{K-k+1}} \leq \prod_{1 \leq k \leq K} \binom{2n/\sigma s_{K-k+1}}{r_{K-k+1} + q_{K-k+1}}. \quad (23)$$

201 To justify the inequality in (23) above, we check that

$$r_{K-k+1} + q_{K-k+1} \leq \frac{2n}{3\sigma s_{K-k+1}}. \quad (24)$$

202 Recalling that  $r_{K-k+1} + q_{K-k+1} = s_{K-k+1} = t2^{-k}$ , we see that (24) is equivalent to  $t2^{-k} \leq \sqrt{2n/3\sigma}$ .

203 However,

$$\frac{t}{2^k} \leq \frac{t}{2} \leq \frac{1}{2}F(n) = \left(\frac{1}{2} + o(1)\right)\sqrt{n} \leq \sqrt{\frac{2n}{3}} \leq \sqrt{\frac{2n}{3\sigma}} \quad (25)$$

204 for all large enough  $n$ . We continue (23) by noticing that

$$\begin{aligned} \prod_{1 \leq k \leq K} \binom{2n/\sigma s_{K-k+1}}{r_{K-k+1} + q_{K-k+1}} &= \prod_{1 \leq k \leq K} \binom{2n/\sigma t2^{-k}}{t2^{-k}} \leq \prod_{1 \leq k \leq K} \left(\frac{2^{2k+1}en}{\sigma t^2}\right)^{t2^{-k}} \\ &\leq \left(\frac{2en}{\sigma t^2}\right)^{t \sum_{k \geq 1} 2^{-k}} 2^{2t \sum_{k \geq 1} k2^{-k}} = \left(\frac{2en}{\sigma t^2}\right)^t 2^{4t} = \left(\frac{32en}{\sigma t^2}\right)^t. \end{aligned} \quad (26)$$

205 Inequality (5) now follows from (19), (20), (21), (22) and (26).  $\square$

206 It now remains to prove Lemma 3.2.

207 *Proof of Lemma 3.2.* Let  $S_0 \subset [n]$  be an arbitrary Sidon set with  $|S_0| = s$ . We show that the  
208 number of Sidon sets  $S \subset [n]$  with  $S_0 \subset S$  and  $|S| = s + q + r$  is at most  $\binom{n}{q} \binom{2n/\sigma s}{r}$ , whence our  
209 lemma will follow.

210 Let  $G$  be the graph on  $V = [n] \setminus S_0$  satisfying that  $\{a_1, a_2\}$  ( $a_1 \neq a_2$ ) is an edge in  $G$  if and only if  
211 there are  $b_1$  and  $b_2 \in S_0$  such that  $a_1 + b_1 = a_2 + b_2$ . Observe that if  $S \subset [n]$  is a Sidon set containing  
212  $S_0$ , then  $S \setminus S_0$  is an independent set in  $G$ . Let  $N = |V| = n - s$ ,  $\beta = (1 - \sigma)s^2/2n$  and  $R = 2n/\sigma s$ .  
213 We wish to apply Lemma 3.1 to  $G$  with  $\beta$  and  $R$  as just defined, to obtain an upper bound for the  
214 number of independent sets of cardinality  $q + r$ . Note that (13) follows from (16). Now let  $U \subset V$   
215 with  $|U| \geq R$  be given. We check (14) as follows.

216 Let  $J$  be the bipartite graph with (disjoint) vertex classes  $[2n]$  and  $U$ , with  $w \in [2n]$  adjacent  
217 to  $a \in U$  in  $J$  if and only if  $w = a + b$  for some  $b \in S_0$ . Note that  $a_1$  and  $a_2 \in U$  have a common  
218 neighbour  $w \in [2n]$  if and only if there are  $b_1$  and  $b_2 \in S_0$  with  $a_1 + b_1 = w = a_2 + b_2$ , that is, if  
219 and only if  $\{a_1, a_2\}$  is an edge of  $G$ .

220 Now note that  $J$  contains no 4-cycle: if  $a_1, a_2 \in U$  with  $a_1 \neq a_2$  are both adjacent to both  $w$  and  
221  $w' \in [2n]$  with  $w \neq w'$ , then  $a_1 + b_1 = w = a_2 + b_2$  for some  $b_1$  and  $b_2 \in S_0$  and  $a_1 + b'_1 = w' = a_2 + b'_2$   
222 for some  $b'_1$  and  $b'_2 \in S_0$ . But then  $b_1 - b'_1 = b_2 - b'_2$ , and hence  $b_1 + b'_2 = b'_1 + b_2$ . As  $b_1, b'_1, b_2$

223 and  $b'_2 \in S_0$  and  $S_0$  is a Sidon set, we have  $\{b_1, b'_2\} = \{b'_1, b_2\}$ . Since  $a_1 \neq a_2$ , we have  $b_1 \neq b_2$ ,  
 224 whence  $b_1 = b'_1$ , implying that  $w = a_1 + b_1 = a_1 + b'_1 = w'$ .

225 The remarks above give that  $e(U) = \sum_{w \in [2n]} \binom{d_J(w)}{2}$ , where  $d_J(w)$  denotes the degree of  $w$  in  $J$ .  
 226 Note that  $\sum_{w \in [2n]} d_J(w) = \sum_{a \in U} d_J(a) = |U||S_0| = |U|s$ . Using the convexity of the func-  
 227 tion  $f(x) = \binom{x}{2}$  and Jensen's inequality and recalling that  $|U| \geq R = 2n/\sigma s$ , i.e.,  $1 \leq \sigma \frac{|U|s}{2n}$ ,  
 228 we obtain

$$e(U) = \sum_{w \in [2n]} \binom{d_J(w)}{2} \geq 2n \binom{|U|s/2n}{2} = \frac{|U|s}{2} \left( \frac{|U|s}{2n} - 1 \right) \geq \frac{1}{4} (1 - \sigma) \frac{s^2}{n} |U|^2 \geq \beta \binom{|U|}{2},$$

229 as required in (14). Recall that a Sidon set  $S \subset [n]$  containing  $S_0$  is such that  $S \setminus S_0$  is an  
 230 independent set in  $G$ . Therefore, our required bound for the number of such  $S$  with  $|S| = s + q + r$   
 231 follows from the upper bound (15) for the number of independent sets of cardinality  $q + r$  in  $G$ .  $\square$

232 We conclude this section by deriving Theorem 1.1 from Theorem 2.1.

233 *Proof of Theorem 1.1.* Let  $\sigma = 32/33$  in Theorem 2.1. Then  $s_0 = (2(1 - \sigma)^{-1} n \log n)^{1/3} =$   
 234  $(66n \log n)^{1/3}$ . For large enough  $n$ , we have

$$|\mathcal{Z}_n| = \sum_{0 \leq t \leq F(n)} |\mathcal{Z}_n(t)| \leq \sum_{0 \leq t < 2s_0} \binom{n}{t} + \sum_{2s_0 \leq t \leq F(n)} n^{3s_0} \left( \frac{33en}{t^2} \right)^t. \quad (27)$$

235 Note that

$$\sum_{0 \leq t < 2s_0} \binom{n}{t} \leq 2s_0 \binom{n}{2s_0} \leq n^{2s_0}, \quad (28)$$

236 and that since  $f(t) = (33en/t^2)^t$  is increasing on the interval  $(0, \sqrt{33n/e})$ ,

$$\sum_{2s_0 \leq t \leq F(n)} n^{3s_0} \left( \frac{33en}{t^2} \right)^t \leq \sqrt{n} \cdot n^{3s_0} (33e)^{\sqrt{n}(1+o(1))} \leq (33e)^{\sqrt{n}(1+o(1))} \leq (33e)^{F(n)(1+o(1))}. \quad (29)$$

237 Combining (27) together with (28) and (29) implies that  $|\mathcal{Z}_n| \leq 2^{cF(n)}$  for a suitable constant  $c$ .  $\square$

238 **3.3. Proof of Theorem 2.2.** We derive Theorem 2.2 from the following more general but technical  
 239 estimate.

240 **Lemma 3.3.** *Let  $n$  and  $t$  be integers. Suppose  $s$  is an integer and  $\sigma$  is a real number such that,*  
 241 *letting  $\omega = t/s$ , we have*

$$\omega \geq 4, \quad 0 < \sigma < 1 \quad \text{and} \quad \frac{s^3}{n} \geq \frac{2}{1 - \sigma} \log \frac{\sigma s}{2}. \quad (30)$$

242 *Then*

$$|\mathcal{Z}_n(t)| \leq \left( \frac{12\omega n}{(t\sigma)^{1-2/\omega t}} \right)^t. \quad (31)$$

243 *Proof.* We invoke Lemma 3.2 with  $q = s$ . Note that, then, (30) implies (16). We now let  $r$  in  
 244 Lemma 3.2 be  $t - 2s$  and obtain that

$$|\mathcal{Z}_n(t)| \leq \binom{n}{s} \binom{n}{s} \binom{2n/\sigma s}{t-2s}. \quad (32)$$

245 The right-hand side of (32) is

$$\begin{aligned} \binom{n}{s}^2 \binom{2n/\sigma s}{t-2s} &\leq \left(\frac{en}{s}\right)^{2s} \left(\frac{2en}{\sigma s(t-2s)}\right)^{t-2s} = \left(\frac{en}{s}\right)^{2s} \left(\frac{en}{s}\right)^{t-2s} \left(\frac{2}{\sigma(t-2s)}\right)^{t-2s} \\ &= \left(\frac{e\omega n}{t}\right)^t \left(\frac{2}{\sigma t(1-2/\omega)}\right)^{t(1-2/\omega)} = \left(C \frac{n}{t^{2-2/\omega} \sigma^{1-2/\omega}}\right)^t, \end{aligned}$$

246 where  $C = 2^{1-2/\omega} e\omega / (1-2/\omega)^{1-2/\omega} = 2^{1-2/\omega} e\omega^{2-2/\omega} / (\omega-2)^{1-2/\omega}$ . As  $\omega \geq 4$ , we have  $\omega-2 \geq \omega/2$ ,  
 247 and hence  $C \leq e\omega 4^{1-2/\omega} < 12\omega$ , completing the proof of Lemma 3.3.  $\square$

248 *Proof of Theorem 2.2.* We shall apply Lemma 3.3. Let  $\omega = 4$  and  $s = t/\omega = t/4$ . Let  $\lambda =$   
 249  $\exp(t^3/(3 \cdot 2^6 n))$  and set  $\sigma = 2\lambda/s = 8\lambda/t \leq 1/3$ , where the last inequality follows from (6). It  
 250 follows that  $2/(1-\sigma) \leq 3$ , and hence

$$\frac{s^3}{n} = \frac{t^3}{4^3 n} = 3 \log \lambda \geq \frac{2}{1-\sigma} \log \lambda,$$

251 whence the third condition in (30) holds. We thus conclude that (31) holds. Let us now estimate  
 252 the right-hand side of (31).

253 Note that  $t\sigma = 4s\sigma = 8\lambda$ , and therefore  $(t\sigma)^{1-2/\omega} = (8\lambda)^{1/2}$  and

$$\frac{12\omega n}{(t\sigma)^{1-2/\omega} t} = \frac{12 \cdot 4n}{(8\lambda)^{1/2} t} = \frac{6 \cdot 8n}{8^{1/2} \lambda^{1/2} t} = \frac{6 \cdot 2^{3/2} n}{\lambda^{1/2} t} = \frac{6 \cdot 2^{3/2} n}{t} \exp\left(-\frac{t^3}{3 \cdot 2^7 n}\right). \quad (33)$$

254 Inequality (7) follows from (31) and (33), and Theorem 2.2 is proved.  $\square$

#### 255 4. THE UPPER BOUNDS IN THEOREMS 2.5–2.7

256 We shall apply Lemma 3.3 and Theorem 2.1 in order to prove the upper bounds in Theorem 2.5  
 257 and Theorems 2.6–2.7, respectively.

258 **4.1. Proof of the upper bound in Theorem 2.5.** Let  $\delta > 0$  be given. We show that there is a  
 259 constant  $c_2 = c_2(\delta)$  such that if  $2n^{-2/3} \leq p = p(n) \leq n^{-1/3-\delta}$ , then w.o.p. we have

$$F([n]_p) \leq c_2 (n \log n^2 p^3)^{1/3}.$$

260 To this end, we apply Lemma 3.3. We first define several auxiliary constants used to set  $t$ ,  $\omega$  and  $\sigma$   
 261 in Lemma 3.3. Choose  $\eta > 0$  small enough so that

$$(1-3\delta) \left(\frac{1}{3} + \eta\right) < \frac{1}{3}. \quad (34)$$

262 Choose  $\omega \geq 4$  so that

$$\left(\frac{1}{3} + \eta\right) \left(1 - \frac{2}{\omega}\right) > \frac{1}{3}. \quad (35)$$

263 Finally, choose  $c = c_2$  so that

$$\left(\frac{c}{\omega}\right)^3 > 3 \left(\frac{1}{3} + \eta\right) \quad \text{and} \quad c > \frac{24\omega}{2^{(1+3\eta)(1-2/\omega)}}. \quad (36)$$

264 Now set  $t = c(n \log n^2 p^3)^{1/3}$ ,  $s = t/\omega$ ,  $\sigma = 2(n^2 p^3)^{1/3+\eta}/s$  and  $\xi = 24\omega/c2^{(1+3\eta)(1-2/\omega)}$ . Note that

$$t \geq c(n \log 8)^{1/3} \geq cn^{1/3} \quad \text{and} \quad \xi < 1. \quad (37)$$

265 We first check that condition (30) holds for large enough  $n$ . We have  $\omega \geq 4$  by the choice of  $\omega$ .  
 266 Moreover, we have  $\sigma \rightarrow 0$  as  $n \rightarrow \infty$  because of (34). Finally, from (36) and the fact that  $\sigma \rightarrow 0$ ,  
 267 we have

$$\frac{s^3}{n} = \left(\frac{c}{\omega}\right)^3 \log n^2 p^3 \geq 3 \left(\frac{1}{3} + \eta\right) \log n^2 p^3 \geq \frac{2(1/3 + \eta)}{1 - \sigma} \log n^2 p^3 = \frac{2}{1 - \sigma} \log \frac{\sigma s}{2},$$

268 which completes the verification of (30). Hence Lemma 3.3 implies that

$$\mathbb{P}([n]_p \text{ contains a Sidon set of size } t) \leq |\mathcal{Z}_n(t)| p^t \leq \left(\frac{12\omega n p}{t(t\sigma)^{1-2/\omega}}\right)^t. \quad (38)$$

269 Making use of the first equation of (37) and the fact that  $t\sigma = \omega s\sigma = 2\omega(n^2 p^3)^{1/3+\eta}$ , we see that  
 270 the upper bound in (38) is at most

$$\begin{aligned} \left(\frac{12\omega n p}{cn^{1/3}(2\omega)^{1-2/\omega}(n^2 p^3)^{(1/3+\eta)(1-2/\omega)}}\right)^t &\leq \left(\frac{12\omega}{c(2\omega)^{1-2/\omega}} \cdot \frac{n^{2/3} p}{(n^2 p^3)^{(1/3+\eta)(1-2/\omega)}}\right)^t \\ &= \left(\frac{12\omega^{2/\omega}}{2^{1-2/\omega} c (n^2 p^3)^{(1/3+\eta)(1-2/\omega)-1/3}}\right)^t, \end{aligned} \quad (39)$$

271 which, by (35) and the assumption  $p \geq 2n^{-2/3}$ , is at most

$$\left(\frac{12\omega}{2^{1/2} c (2^3)^{(1/3+\eta)(1-2/\omega)-1/3}}\right)^t \leq \left(\frac{24\omega}{c2^{(1+3\eta)(1-2/\omega)}}\right)^t = \xi^t. \quad (40)$$

272 To complete the proof, it suffices to recall (37).

273 **4.2. Proof of the upper bound in Theorem 2.6.** Suppose  $1 \leq \alpha = \alpha(n) \leq n^{1/3}$ , and let  
 274  $p = p(n) = \alpha^{-1} n^{-1/3} (\log n)^{2/3}$ . We show that w.o.p.

$$F([n]_p) \leq c_4 (n \log n)^{1/3} \frac{\log n}{\log(\alpha + \log n)} \quad (41)$$

275 for some absolute constant  $c_4$ . To this end, we use Theorem 2.1. Let  $\sigma = 3/4$ ,  $s_0 = 2(n \log n)^{1/3}$   
 276 and  $t = \omega s_0$ , where

$$\omega = 11e \frac{\log n}{\log(\alpha + \log n)}, \quad (42)$$

277 and note that  $\omega \geq 2$  for sufficiently large  $n$ . Hence, by Theorem 2.1 and the union bound, the  
 278 probability that  $[n]_p$  contains a Sidon set with at least  $t$  elements can be bounded as follows:

$$\mathbb{P}(F([n]_p) \geq t) \leq |\mathcal{Z}_n(t)|p^t \leq n^{3s_0} \left(\frac{44enp}{t^2}\right)^t = n^{3s_0} \left(\frac{44enp}{\omega^2 s_0^2}\right)^{\omega s_0} \leq \left[\left(\frac{11e}{\alpha\omega^2}\right)^\omega n^3\right]^{s_0}, \quad (43)$$

279 where the last inequality follows from  $p = \alpha^{-1}n^{-1/3}(\log n)^{2/3}$  and  $s_0 = 2(n \log n)^{1/3}$ .

280 For the proof of (41), it suffices to show that the base of the exponential in the right-hand side  
 281 of (43) is bounded away from 1, that is, whether

$$\left(\frac{11e}{\alpha\omega^2}\right)^\omega n^3 < 1 - \varepsilon \quad (44)$$

282 for some absolute constant  $\varepsilon > 0$ . Since  $\omega \geq 11e$  for sufficiently large  $n$ , then we have

$$\left(\frac{\alpha\omega^2}{11e}\right)^\omega \geq (\alpha\omega)^\omega = \exp(\omega \log(\alpha\omega)). \quad (45)$$

283 We claim that

$$2 \log(\alpha\omega) \geq \log(\alpha + \log n). \quad (46)$$

284 Observe that since  $\omega \geq 2$ , then (46) is trivially satisfied if  $\alpha \geq \log n$ . On the other hand, if  
 285  $\alpha \leq \log n$ , then  $\omega \geq (\log n)/\log \log n$  and hence

$$2 \log(\alpha\omega) \geq 2 \log \omega \geq 2 \log \log n - 2 \log \log \log n \geq \log(2 \log n) \geq \log(\alpha + \log n).$$

286 It follows from (42), (45) and (46) that

$$\left(\frac{\alpha\omega^2}{11e}\right)^\omega \geq \exp(\omega \log(\alpha\omega)) \geq \exp(5e \log n) \geq 2n^3$$

287 and hence (44) holds, completing the proof of (41).

288 **4.3. Proof of the upper bounds in Theorem 2.7.** Suppose that  $\beta = \beta(n) \geq 1$  and let  $p =$   
 289  $p(n) = \beta n^{-1/3}(\log n)^{2/3}$ . Let  $\sigma = 3/4$ ,  $s_0 = 2(n \log n)^{1/3}$  and  $t = \omega s_0$  for some  $\omega \geq 2$ . Similarly as  
 290 in the proof of the upper bound in Theorem 2.6, see (43), using Theorem 2.1, we estimate

$$\mathbb{P}(F([n]_p) \geq t) \leq |\mathcal{Z}_n(t)|p^t \leq \left[\left(\frac{11e\beta}{\omega^2}\right)^\omega n^3\right]^{s_0}. \quad (47)$$

291 We split into two cases, depending on the order of magnitude of  $\beta$ .

292 (*Case I*) If  $\beta(n) \leq (\log n)^2$ , then we let  $\alpha = \beta^{-1}(\log n)^2$  and  $\omega = (11e \log n)/\log(e\alpha)$  so that  
 293  $t = \omega s_0 = 22e\sqrt{n\beta} \cdot \sqrt{\alpha}/\log(e\alpha)$ . Note that

$$\left(\frac{11e\beta}{\omega^2}\right)^\omega = \left(\frac{11e(\log n)^2}{\alpha\omega^2}\right)^\omega = \left(\frac{(\log(e\alpha))^2}{11e\alpha}\right)^{11e(\log(e\alpha))^{-1} \log n}. \quad (48)$$

294 Since the function  $f(x) = \left(\frac{x^2}{11e^x}\right)^{1/x} = \frac{1}{e} \left(\frac{x^2}{11}\right)^{1/x}$  is bounded by  $e^{\sqrt{4/11}/e-1} = 0.459\dots$  on  
 295 the interval  $[1, \infty)$ , it follows from (48) that (we let  $x = \log(e\alpha)$ )

$$\left(\frac{11e\beta}{\omega^2}\right)^\omega \leq \left(\frac{1}{2}\right)^{11e\log n} \leq n^{-4},$$

296 which, together with (47), proves that w.o.p. we have

$$F([n]_p) \leq t = c_6\sqrt{np} \cdot \frac{\sqrt{\alpha}}{1 + \log \alpha},$$

297 where  $c_6$  is an absolute constant.

298 (*Case II*) If  $\beta(n) \geq (\log n)^2$ , then we let  $\omega = 11e\sqrt{\beta}$  so that  $t = \omega s_0 = 22e\sqrt{np}$ . By (47), we have

$$\mathbb{P}(F([n]_p) \geq t) \leq \left[(11e)^{-11e\sqrt{\beta}} n^3\right]^{s_0} \leq \left[(11e)^{-\log n} n^3\right]^{s_0} \leq e^{-s_0},$$

299 which proves that w.o.p. we have

$$F([n]_p) \leq t = c_6\sqrt{np},$$

300 where  $c_6$  is an absolute constant.

## 301 5. NONTRIVIAL SOLUTIONS IN RANDOM SETS

302 **5.1. Estimating the number of nontrivial solutions.** A *solution* of the equation  $x_1 + x_2 =$   
 303  $y_1 + y_2$  is a quadruplet  $(a_1, a_2, b_1, b_2) \in [n]^4 = [n] \times [n] \times [n] \times [n]$  with  $a_1 + a_2 = b_1 + b_2$ . A  
 304 solution  $(a_1, a_2, b_1, b_2)$  of  $x_1 + x_2 = y_1 + y_2$  is called *trivial* if it is of the form  $(a_1, a_2, a_1, a_2)$  or  
 305  $(a_1, a_2, a_2, a_1)$ . Otherwise, it is called a *nontrivial* solution. Let us define a hypergraph and a  
 306 random variable that will be important for us.

307 **Definition 5.1.** *Let*

$$\mathcal{S} = \left\{ \{a_1, a_2, a_3, a_4\} : (a_1, a_2, a_3, a_4) \text{ is a nontrivial solution of } x_1 + x_2 = y_1 + y_2 \right\}. \quad (49)$$

308 *We think of  $\mathcal{S}$  as a hypergraph on the vertex set  $[n]$ . As usual, for  $R \subset [n]$ , we let  $\mathcal{S}[R]$  denote the*  
 309 *subhypergraph of  $\mathcal{S}$  induced on  $R$ . Let  $X$  be the random variable  $|\mathcal{S}[[n]_p]|$ , that is, the number of*  
 310 *hyperedges of  $\mathcal{S}$  induced by  $[n]_p$ .*

311 In Lemma 5.4 below, we give an estimate for  $X$  that will be used in the proof of Theorem 2.3 and  
 312 in the proofs of the lower bounds in Theorems 2.5–2.7.

313 To estimate  $X$ , we have to deal with the issue of ‘repeated entries’ in a hyperedge  $\{a_1, a_2, b_1, b_2\} \in \mathcal{S}$ .  
 314 Indeed, if  $\{a_1, a_2, a_3, a_4\} \in \mathcal{S}$ , with  $a_1 \leq a_2 \leq a_3 \leq a_4$ , we may have  $a_2 = a_3$ , but no other equality  
 315 can occur. Hence the hypergraph  $\mathcal{S}$  has hyperedges of size 4 and 3. Based on this, we make the  
 316 following definition.

317 **Definition 5.2.** *For  $i = 3$  and  $4$ , let  $\mathcal{S}_i$  be the subhypergraph of  $\mathcal{S}$  with all the hyperedges of size  $i$ .*  
 318 *Furthermore, let  $X_i := |\mathcal{S}_i[[n]_p]|$ .*

319 We clearly have

$$\mathcal{S} = \mathcal{S}_4 \cup \mathcal{S}_3 \quad \text{and} \quad \mathcal{S}_4 \cap \mathcal{S}_3 = \emptyset \quad (50)$$

320 and hence

$$X = X_4 + X_3. \quad (51)$$

321 In order to estimate  $X$ , we estimate  $X_4$  and  $X_3$  separately.

322 **Lemma 5.3.** *Fix  $\delta > 0$ . The following assertions hold w.o.p.*

323 (i) *If  $p \geq n^{-3/4+\delta}$ , then  $X_4 = n^3 p^4 (1/12 + o(1))$ .*

324 (ii) *If  $p \gg n^{-1}$ , then  $X_3 = O(\max\{n^2 p^3, n^{3\delta}\})$ .*

325 We remark that the constant implicit in the big- $O$  notation in (ii) above is an absolute constant.

326 The proof of Lemma 5.3 is based on a concentration result due to Kim and Vu [18]. We shall

327 introduce the Kim–Vu polynomial concentration result in Section 5.2 and prove Lemma 5.3 in

328 Section 5.3. Assuming Lemma 5.3, we are ready to estimate  $X$ .

329 **Lemma 5.4.** *Fix  $\delta > 0$  and suppose  $p \geq n^{-3/4+\delta}$ . Then, w.o.p.,  $X = n^3 p^4 (1/12 + o(1))$ .*

330 *Proof.* Let  $X = X([n]_p)$  be as defined in Definition 5.1 and recall (51). From the assumption

331 that  $p \geq n^{-3/4+\delta}$ , we see that the estimates for  $X_4$  and  $X_3$  given in Lemma 5.3(i) and (ii) do

332 hold w.o.p. Since the inequality  $np \gg 1$  yields  $n^2 p^3 \ll n^3 p^4$  and we also have  $n^{3\delta} \ll n^{4\delta} \leq n^3 p^4$ ,

333 because  $p \geq n^{-3/4+\delta}$ , we infer  $\max\{n^2 p^3, n^{3\delta}\} \ll n^3 p^4$ , and hence, w.o.p.,  $X_3 \ll X_4$ . It follows

334 from (51) and the estimate in Lemma 5.3(i) that  $X = n^3 p^4 (1/12 + o(1))$  holds w.o.p.  $\square$

335 It now remains to prove Lemma 5.3. We first introduce the main tool we shall use in the proof of  
336 that lemma, due to Kim and Vu [18].

337 **5.2. The Kim–Vu polynomial concentration result.** Let  $\mathcal{H} = (V, E)$  be a hypergraph on the

338 vertex set  $V = [n]$ . We assume each hyperedge  $e \in E(\mathcal{H})$  has a real weight  $w(e)$ . Let  $[n]_p$  be a

339 random subset of  $[n]$  obtained by choosing each element  $i \in [n]$  independently with probability  $p$

340 and let  $\mathcal{H}[[n]_p]$  be the subhypergraph of  $\mathcal{H}$  induced on  $[n]_p$ . Let  $Y$  be the sum of the weights of all

341 the hyperedges in  $\mathcal{H}[[n]_p]$ , i.e.,  $Y = \sum_{e \in \mathcal{H}[[n]_p]} w(e)$ . Kim and Vu obtained a concentration result

342 for the random variable  $Y$ . We now proceed to present their result [18] (see also Theorem 7.8.1 in

343 Alon and Spencer [3]).

344 We start by introducing basic definitions and notation (we follow [3]). Let  $k$  be the maximum

345 cardinality of the hyperedges in  $\mathcal{H}$ . For a set  $A \subset [n]$  ( $|A| \leq k$ ), let  $Y_A$  be the sum of the weights of

346 all the hyperedges in  $\mathcal{H}[[n]_p]$  containing  $A$ , i.e.,  $Y_A = \sum_{A \subset e \in \mathcal{H}[[n]_p]} w(e)$ . Let  $E_A = \mathbb{E}(Y_A \mid A \subset [n]_p)$

347 be the expectation of  $Y_A$  conditioned on the event that  $A$  should be contained in  $[n]_p$ . Let  $E_i$  be

348 the maximum value of  $E_A$  over all  $A \subset [n]$  with  $|A| = i$ . Note that  $E_0 = \mathbb{E}(Y)$ . Let  $\mu = \mathbb{E}(Y)$  and

349 set

$$E' = \max\{E_i : 1 \leq i \leq k\} \quad \text{and} \quad E = \max\{E', \mu\}. \quad (52)$$

350 **Theorem 5.5** (Kim–Vu polynomial concentration inequality). *With the above notation, we have,*  
 351 *for every  $\lambda > 1$ ,*

$$\mathbb{P}\left[|Y - \mu| > a_k(EE')^{1/2}\lambda^k\right] < 2e^2e^{-\lambda}n^{k-1},$$

352 *where  $a_k = 8^k(k!)^{1/2}$ .*

353 **5.3. Proof of Lemma 5.3.** We prove (i) and (ii) of Lemma 5.3 separately.

354 *Proof of Lemma 5.3(i).* We need to show that, for  $p \geq n^{-3/4+\delta}$ , where  $\delta > 0$  is fixed, we have  
 355  $X_4 = n^3p^4(1/12 + o(1))$  w.o.p. We first estimate the expectation  $\mu(X_4)$  of  $X_4$ .

356 Suppose  $\{i, j, k, l\} \in \mathcal{S}_4$  with  $0 \leq i < j < k < l \leq n-1$ . Note that  $i+l = j+k$ . Let us fix  
 357  $0 \leq i \leq n-1$ . If  $j \geq (n+i)/2$ , then  $l = j+k-i > 2j-i \geq n+i-i = n$ , which contradicts  $l \leq n-1$ .  
 358 Hence we have  $i < j < (n+i)/2$ . For fixed  $i$  and  $j$ , if  $k > n+i-j-1$ , then  $l = j+k-i > n-1$ ,  
 359 which contradicts  $l \leq n-1$ . Therefore we have  $j < k \leq n+i-j-1$ . Once  $i, j$  and  $k$  are chosen,  
 360 the value of  $l$  is determined by the condition  $i+l = j+k$ . Consequently,

$$|\mathcal{S}_4| \sim \sum_{i=0}^{n-1} \sum_{j=i}^{(n+i)/2} \sum_{k=j}^{n+i-j-1} 1 = \sum_{i=0}^{n-1} \sum_{j=i}^{(n+i)/2} (n+i-2j) \sim n^3 \int_0^1 \int_x^{(1+x)/2} (1+x-2y)dydx \sim \frac{1}{12}n^3.$$

361 Hence

$$\mu(X_4) = |\mathcal{S}_4|p^4 = \left(\frac{1}{12} + o(1)\right)n^3p^4. \quad (53)$$

362 Next we apply Theorem 5.5 to prove that  $X_4$  is concentrated around its expectation  $\mu(X_4)$ . To this  
 363 end, we compute the quantities  $E_i$  ( $1 \leq i \leq 4$ ) and  $E'$  and  $E$  defined in (52). We first estimate  $E_1$ .  
 364 For  $a \in [n]$ , consider the quantity  $E_{\{a\}}$ . The number of hyperedges in  $\mathcal{S}_4$  containing  $a$  is  $O(n^2)$   
 365 and the probability that one such hyperedge is in  $[n]_p$ , conditioned on  $a \in [n]_p$ , is  $p^3$ . We conclude  
 366 that, for any  $a \in [n]$ , we have  $E_{\{a\}} = O(n^2p^3)$ . Consequently,  $E_1 = \max\{E_A : |A| = 1\} = O(n^2p^3)$ .  
 367 A similar argument gives that  $E_i = \max\{E_A : |A| = i\} = O(n^{3-i}p^{4-i})$  for all  $1 \leq i < 4$ . Therefore,  
 368 since  $np \gg 1$ , we have  $E_i = O(n^2p^3)$  for all  $1 \leq i < 4$ . Also, clearly,  $E_4 = \max\{E_A : |A| = 4\} = 1$ .  
 369 Thus

$$E' = \max\{E_i : 1 \leq i \leq 4\} = O(\max\{n^2p^3, 1\}), \quad (54)$$

370 and  $E = \max\{E', \mu(X_4)\} = O(\max\{n^2p^3, 1, n^3p^4\})$ . Since  $p \geq n^{-3/4+\delta} > n^{-3/4}$ , we have

$$E = O(n^3p^4). \quad (55)$$

371 In view of (54) and (55), a simple computation implies the following:

372 (Case I) If  $n^{-3/4+\delta} \leq p \leq n^{-2/3}$ , then

$$E' = O(1) \quad \text{and} \quad E = O(n^3p^4). \quad (56)$$

373 (Case II) If  $p \geq n^{-2/3}$ , then

$$E' = O(n^2p^3) \quad \text{and} \quad E = O(n^3p^4). \quad (57)$$

374 We now estimate  $X_4$  for each case separately.



375 (Case I) Suppose  $n^{-3/4+\delta} \leq p \leq n^{-2/3}$ . In this case, (56) implies that

$$(EE')^{1/2} = O(n^3 p^4 \cdot 1)^{1/2} = O(n^3 p^4)^{1/2}. \quad (58)$$

376 Set  $\lambda = (n^3 p^4)^{1/12}$ . By the assumption  $p \geq n^{-3/4+\delta}$ , we have

$$\lambda = (n^3 p^4)^{1/12} \geq n^{\delta/3}. \quad (59)$$

377 Also  $n^3 p^4 \geq n^{4\delta} \gg 1$ , and hence combining (58) and  $\lambda = (n^3 p^4)^{1/12}$  implies that

$$(EE')^{1/2} \lambda^4 = O(n^3 p^4)^{1/2} (n^3 p^4)^{1/3} = O(n^3 p^4)^{5/6} = o(n^3 p^4). \quad (60)$$

378 Theorem 5.5 together with (59) then yields that

$$\mathbb{P}\left[|X_4 - \mu(X_4)| > a_4 (EE')^{1/2} \lambda^4\right] < 2e^2 e^{-\lambda n^3} \leq 2e^2 e^{-n^{\delta/3} n^3},$$

379 where  $a_4 = 8^4 (4!)^{1/2}$ . Given (60), we have that w.o.p.

$$X_4 = \mu(X_4) + o(n^3 p^4). \quad (61)$$

380 (Case II) Suppose  $p \geq n^{-2/3}$ . In this case, (57) yields that

$$(EE')^{1/2} = O(n^3 p^4 n^2 p^3)^{1/2} = O\left(\frac{n^3 p^4}{(np)^{1/2}}\right). \quad (62)$$

381 Set  $\lambda = (np)^{1/12}$ . By the assumption  $p \geq n^{-2/3}$ ,

$$\lambda \geq (n^{1/3})^{1/12} = n^{1/36}. \quad (63)$$

382 Since  $np \gg 1$ , combining (62) and  $\lambda = (np)^{1/12}$  implies that

$$(EE')^{1/2} \lambda^4 = O\left(\frac{n^3 p^4}{(np)^{1/2}}\right) (np)^{1/3} = O\left(\frac{n^3 p^4}{(np)^{1/6}}\right) = o(n^3 p^4). \quad (64)$$

383 Theorem 5.5 together with (63) then yields that

$$\mathbb{P}\left[|X_4 - \mu(X_4)| > a_4 (EE')^{1/2} \lambda^4\right] < 2e^2 e^{-\lambda n^3} \leq 2e^2 e^{-n^{1/36} n^3},$$

384 where  $a_4 = 8^4 (4!)^{1/2}$ . Given (64), we have that w.o.p.

$$X_4 = \mu(X_4) + o(n^3 p^4). \quad (65)$$

385 In view of (53), it follows from (61) and (65) that, for  $p \geq n^{-3/4+\delta}$ , we have  $X_4 = n^3 p^4 (1/12 + o(1))$   
 386 w.o.p. This completes the proof of (i) of Lemma 5.3.  $\square$

387 *Proof of Lemma 5.3(ii).* Fix  $\delta > 0$ . We show that, w.o.p.,  $X_3 = O(\max\{n^2 p^3, n^{3\delta}\})$  for  $p \gg n^{-1}$ .

388 First we estimate the expectation  $\mu(X_3)$  of  $X_3$ . Since  $|\mathcal{S}_3| = O(n^2)$ , we have

$$\mu(X_3) = O(n^2 p^3). \quad (66)$$

389 Next, we prove a concentration result for  $X_3$  applying Theorem 5.5. To this end, we estimate  
 390 the quantities  $E_i$  ( $1 \leq i \leq 3$ ). As in the proof of Lemma 5.3(i), one may check that  $E' =$   
 391  $\max_{1 \leq i \leq 3} E_i = O(\max\{np^2, p, 1\})$  and hence  $E = \max\{E', \mu(X_3)\} = O(\max\{np^2, p, 1, n^2 p^3\})$ . By

392 the assumption  $np \gg 1$ , we infer

$$E' = O(\max\{np^2, 1\}) \quad \text{and} \quad E = O(\max\{n^2p^3, 1\}). \quad (67)$$

393 Based on (67), we consider the cases  $p \geq n^{-2/3+\delta}$  and  $n^{-1} \ll p \leq n^{-2/3+\delta}$  separately.

394 We first suppose  $p \geq n^{-2/3+\delta}$ . From (67), we have  $E' = O(\max\{np^2, 1\})$  and  $E = O(n^2p^3)$ . A proof  
395 similar to the proofs of (61) and (65) shows that, for  $p \geq n^{-2/3+\delta}$ , w.o.p.,  $X_3 = \mu(X_3) + o(n^2p^3)$ .

396 This together with (66) implies that for  $p \geq n^{-2/3+\delta}$ , w.o.p.,

$$X_3 = O(n^2p^3). \quad (68)$$

397 We now suppose  $n^{-1} \ll p \leq n^{-2/3+\delta}$ . In this case, (67) yields that  $E' = O(1)$  and  $E = O(n^{3\delta})$  and  
398 hence, setting  $\lambda = n^{\delta/2}$ , we have

$$(EE')^{1/2}\lambda^3 = O(n^{(3/2)\delta})n^{(3/2)\delta} = O(n^{3\delta}). \quad (69)$$

399 Theorem 5.5 with  $\lambda = n^{\delta/2}$  yields

$$\mathbb{P}\left[|X_3 - \mu(X_3)| > a_3(EE')^{1/2}\lambda^3\right] < 2e^2e^{-\lambda n^2} \leq 2e^2e^{-n^{\delta/2}} n^2, \quad (70)$$

400 where  $a_3 = 8^3(3!)^{1/2}$ . Inequality (70) together with (69) implies that, for  $n^{-1} \ll p \leq n^{-2/3+\delta}$ ,  
401 w.o.p.,  $X_3 = \mu(X_3) + O(n^{3\delta})$ . Since, under the assumption  $p \leq n^{-2/3+\delta}$ , we have  $\mu(X_3) =$   
402  $O(n^2p^3) = O(n^{3\delta})$ , we infer that, for  $n^{-1} \ll p \leq n^{-2/3+\delta}$ , w.o.p.,

$$X_3 = O(n^{3\delta}). \quad (71)$$

403 Combining (68) and (71) completes the proof of (ii) of Lemma 5.3.  $\square$

404

## 6. PROOF OF THEOREM 2.3

405 **6.1. Theorem 2.3 for smaller  $p = p(n)$ .** We first consider the case in which  $n^{-1} \ll p \ll n^{-2/3}$ .

406 *Proof of (8) in Theorem 2.3.* Suppose  $n^{-1} \ll p \ll n^{-2/3}$ . We show that (8) holds almost surely,  
407 using the usual deletion method. Let  $\mathcal{S}$ ,  $\mathcal{S}[[n]_p]$  and  $X$  be as in Definition 5.1. If we delete one vertex  
408 from each hyperedge in  $\mathcal{S}[[n]_p]$ , the remaining vertex set is an independent set of  $\mathcal{S}[[n]_p]$ , and hence  
409 it is a Sidon set contained in  $[n]_p$ . Consequently,  $F([n]_p) \geq |[n]_p| - |\mathcal{S}[[n]_p]| = |[n]_p| - X$ . Since  
410 trivially  $F([n]_p) \leq |[n]_p|$ , we have  $|[n]_p| - X \leq F([n]_p) \leq |[n]_p|$ . Note that the Chernoff bound gives  
411 that, for  $p \gg n^{-1}$ , we almost surely have  $|[n]_p| = np + o(np)$ . Therefore, in order to show (8), it only  
412 remains to show that  $X = o(np)$  almost surely. Recall that  $X_i$  is the number of edges of cardinality  $i$   
413 in  $\mathcal{S}[[n]_p]$  ( $i \in \{3, 4\}$ ), and that  $X = X_3 + X_4$  (see Definition 5.2 and (51)). Equations (53) and (66),  
414 together with  $n^{-1} \ll p \ll n^{-2/3}$ , imply that  $\mathbb{E}(X) = \Theta(n^3p^4) + O(n^2p^3) = \Theta(n^3p^4) = o(np)$ . Hence  
415 Markov's inequality gives that we almost surely have  $X = o(np)$ , and our result follows.  $\square$

416 **6.2. Theorem 2.3 for larger  $p = p(n)$ .** We now consider the wider range  $n^{-1} \ll p \leq 2n^{-2/3}$ .

417 *Proof of (9) in Theorem 2.3.* We have already shown that, if  $n^{-1} \ll p \ll n^{-2/3}$ , then  $F([n]_p) =$   
418  $(1 + o(1))np$  holds almost surely. Therefore, it suffices to show that (9) holds if, e.g.,  $n^{-2/3}/\log n \leq$   
419  $p \leq 2n^{-2/3}$ . We proceed as in the proof of (8), given in Section 6.1 above. We have already observed  
420 that  $|[n]_p| = np(1 + o(1))$  almost surely as long as  $p \gg n^{-1}$ , and therefore  $F([n]_p) \leq np(1 + o(1))$   
421 almost surely in this range of  $p$ . It now suffices to recall that  $F([n]_p) \geq |[n]_p| - X$  and to prove that,  
422 almost surely, we have  $X \leq (2/3 + o(1))np$  if  $n^{-2/3}/\log n \leq p \leq 2n^{-2/3}$ . But with this assumption  
423 on  $p$ , Lemma 5.4 tells us that, w.o.p.,

$$X = \frac{1}{12}n^3p^4 + o(n^3p^4) = \frac{1}{12}n^3p^4 + o(np) \leq \left(\frac{2}{3} + o(1)\right)np, \quad (72)$$

424 as required. □

## 425 7. THE LOWER BOUNDS IN THEOREMS 2.5–2.7

426 Let us first state a simple monotonicity result (see, e.g., [17, Lemma 1.10]) that will be used a few  
427 times in this section.

428 **Fact 7.1.** *Let  $p = p(n)$  and  $q = q(n)$  be such that  $0 \leq p < q \leq 1$ , and let  $a = a(n) > 0$  and*  
429  *$b = b(n) > 0$  be functions of  $n$ .*

430 (i) *If  $F([n]_p) \geq a$  holds w.o.p., then  $F([n]_q) \geq a$  holds w.o.p.*

431 (ii) *If  $F([n]_q) \leq b$  holds w.o.p., then  $F([n]_p) \leq b$  holds w.o.p.*

432 Statements (i) and (ii) in Fact 7.1 are, in fact, equivalent. We state them both explicitly just for  
433 convenience.

434 **7.1. Proofs of the lower bounds in Theorems 2.5 and 2.6.** The lower bounds in Theorems 2.5  
435 and 2.6 rely on a result on independent sets in hypergraphs. Before stating the relevant result, we  
436 introduce some definitions. A hypergraph is called *simple* if any two of its hyperedges share at most  
437 one vertex. A hypergraph is  *$r$ -uniform* if all its hyperedges have cardinality  $r$ . We shall use the  
438 following extension of a celebrated result due to Ajtai, Komlós, Pintz, Spencer and Szemerédi [1],  
439 obtained by Duke, Lefmann and the third author [10].

440 **Lemma 7.2.** *Let  $\mathcal{H}$  be a simple  $r$ -uniform hypergraph,  $r \geq 3$ , with  $N$  vertices and average degree*  
441 *at most  $t^{r-1}$  for some  $t$ . Then  $\mathcal{H}$  has an independent set of size at least*

$$c \frac{(\log t)^{1/(r-1)}}{t} N, \quad (73)$$

442 where  $c = c(r)$  is a positive constant that depends only on  $r$ .

443 We now briefly discuss how to obtain a lower bound on  $F([n]_p)$  using Lemma 7.2. Let  $\mathcal{S}[[n]_p]$   
444 be the hypergraph in Definition 5.1. Since an independent set of  $\mathcal{S}[[n]_p]$  is a Sidon set contained  
445 in  $[n]_p$ , independent sets in  $\mathcal{S}[[n]_p]$  give lower bounds for  $F([n]_p)$ . To apply Lemma 7.2, we shall  
446 obtain a simple 4-uniform subhypergraph  $\mathcal{S}^*$  of  $\mathcal{S}[[n]_p]$  by deleting suitable vertices from  $\mathcal{S}[[n]_p]$ .

447 Lemma 7.2 will then tell us that  $\mathcal{S}^*$  has a suitably large independent set, and this will yield our  
 448 lower bound on  $F([n]_p)$ . In fact, we obtain the following result.

449 **Lemma 7.3.** *There is an absolute constant  $d > 0$  such that, for  $p \geq 2n^{-2/3}$ , w.o.p.  $F([n]_p) \geq$   
 450  $d(n \log(n^2 p^3))^{1/3}$  holds.*

451 Lemma 7.3 easily implies the lower bounds in Theorems 2.5 and 2.6. The proof of Lemma 7.3 will  
 452 be given in Section 7.3.

453 **7.2. Proof of the lower bound in Theorem 2.7.** For larger  $p = p(n)$ , it turns out that, instead  
 454 of using Lemma 7.2, it is better to make use of the fact that  $[n]$  contains a Sidon set of cardinality  
 455  $(1 + o(1))\sqrt{n}$  (see Section 1). An immediate use of this fact gives the lower bound  $(1 + o(1))p\sqrt{n}$ ,  
 456 but one can, in fact, do better. The following is a particular case of a very general theorem of  
 457 Komlós, Sulyok and Szemerédi [23].

458 **Lemma 7.4.** *There is an absolute constant  $d > 0$  such that, for every sufficiently large  $m$  and  
 459 every set of integers  $A$  with  $|A| = m$ , we have*

$$F(A) \geq d \cdot F([m]).$$

460 Since the Chernoff bound gives that, for  $p \gg 1/n$ , we almost surely have  $|[n]_p| = (1 + o(1))np$ ,  
 461 Lemma 7.4 together with  $F([m]) \geq (1 + o(1))\sqrt{m}$  gives the lower bound in Theorem 2.7. Clearly,  
 462 to have this result with ‘w.o.p.’, it suffices to assume  $p \gg (\log n)/n$ . There is an alternative, simple  
 463 proof of the following fact:

464 (\*) if  $(\log n)^2/n \ll p \leq 1/3$ , then, w.o.p.,

$$F([n]_p) \geq \left( \frac{1}{3\sqrt{2}} + o(1) \right) \sqrt{np}. \quad (74)$$

465 Fact 7.1 then implies that, for  $p \gg (\log n)^2/n$ , we have, w.o.p.,  $F([n]_p) \geq (1/3\sqrt{6} + o(1))\sqrt{np}$ .

466 *Proof of (\*).* Let  $(\log n)^2/n \ll p \leq 1/3$ . We shall show that (74) holds w.o.p. We define a partition  
 467 of  $[n] = \{0, \dots, n-1\}$  into equal length intervals, and consider a family of intervals in the partition  
 468 satisfying the property that, if we choose an arbitrary element from each interval, the set of chosen  
 469 elements forms a Sidon set. We shall choose the length of the intervals so that  $[n]_p$  will intersect  
 470 each interval in a constant number of elements on average. A simple analysis of this construction  
 471 yields that (74) holds w.o.p. The details are as follows.

472 Let  $\mathcal{I} = \{I_i: 0 \leq i < \lceil n/x \rceil\}$  be the partition of  $[n]$  into consecutive intervals with  $x = \lfloor 1/p \rfloor$   
 473 elements each. More precisely, let  $I_i = [xi, x(i+1) - 1] \cap [n]$  for all  $0 \leq i < \lceil n/x \rceil$ . In what follows,  
 474 we ignore  $I_{\lceil n/x \rceil - 1}$  if this interval has fewer than  $x$  elements. Let  $\mathcal{I}_{\text{even}} = \{I_0, I_2, I_4, \dots\} \subset \mathcal{I}$  be  
 475 the set of all intervals with even indices and let  $y = |\mathcal{I}_{\text{even}}|$ . Note that  $y \geq (1/2)\lceil n/x \rceil - 1 \geq$   
 476  $(1/2)\lfloor np \rfloor - 1 = (1/2 + o(1))np$ . By the Chowla–Erdős result [8, 11], there exists a Sidon subset  $S$

477 of  $[y]$  with

$$|S| = (1 + o(1))\sqrt{y} = \left(\frac{1}{\sqrt{2}} + o(1)\right)\sqrt{np}. \quad (75)$$

478 We “identify”  $[y]$  and  $\mathcal{I}_{\text{even}}$  by the bijection  $i \mapsto I_{2i}$ . Let  $\{a_i : i \in S\}$  be a set of integers with  $a_i \in I_{2i}$   
 479 for all  $i \in S$ . We claim that  $\{a_i : i \in S\}$  is a Sidon set. Suppose  $a_{i_1} + a_{i_2} = a_{j_1} + a_{j_2}$ , where  $i_1, i_2,$   
 480  $j_1$  and  $j_2 \in S$ . Observe that

$$a_{i_1} + a_{i_2} \in I_{2i_1+2i_2} \cup I_{2i_1+2i_2+1} \quad \text{and} \quad a_{j_1} + a_{j_2} \in I_{2j_1+2j_2} \cup I_{2j_1+2j_2+1}, \quad (76)$$

481 which, together with the assumption that  $a_{i_1} + a_{i_2} = a_{j_1} + a_{j_2}$ , implies that  $i_1 + i_2 = j_1 + j_2$ . Since  $S$   
 482 is a Sidon set, we have  $\{i_1, i_2\} = \{j_1, j_2\}$ , whence  $\{a_{i_1}, a_{i_2}\} = \{a_{j_1}, a_{j_2}\}$ . This shows that  $\{a_i : i \in S\}$   
 483 is indeed a Sidon set.

484 We now consider a random set  $[n]_p$ . An interval  $I_{2i}$  ( $i \in S$ ) is said to be *occupied* if  $I_{2i}$  contains  
 485 at least one element of  $[n]_p$ . Let  $\mathcal{I}_{\text{occ}}$  be the family of occupied intervals. By the above claim, we  
 486 have  $F([n]_p) \geq |\mathcal{I}_{\text{occ}}|$ . Let us estimate  $|\mathcal{I}_{\text{occ}}|$ . Note that each interval  $I_{2i}$  ( $i \in S$ ) is independently  
 487 occupied with probability

$$\tilde{p} = 1 - (1 - p)^x = 1 - (1 - p)^{\lfloor 1/p \rfloor} \geq 1 - e^{-p(1/p-1)} \geq 1 - e^{-1+p} \geq 1 - e^{-2/3} > 1/3, \quad (77)$$

488 where the third inequality follows from the assumption  $p \leq 1/3$ . Thus, under the assumption  
 489  $(\log n)^2/n \ll p \leq 1/3$ , the Chernoff bound, (75) and (77) give that, w.o.p.,

$$|\mathcal{I}_{\text{occ}}| = (1 + o(1))\mathbb{E}(|\mathcal{I}_{\text{occ}}|) = (1 + o(1))|S|\tilde{p} \geq \left(\frac{1}{\sqrt{2}} + o(1)\right)\sqrt{np} \cdot \frac{1}{3} = \left(\frac{1}{3\sqrt{2}} + o(1)\right)\sqrt{np}.$$

490 Recalling that  $F([n]_p) \geq |\mathcal{I}_{\text{occ}}|$ , statement (\*) follows.  $\square$

491 **7.3. Proof of Lemma 7.3.** In Lemma 7.5 below, we prove Lemma 7.3 for a narrower range of  $p$ .  
 492 We shall then invoke monotonicity (Fact 7.1) to obtain Lemma 7.3 in full.

493 **Lemma 7.5.** *There is an absolute constant  $d > 0$  such that, for  $2n^{-2/3} \leq p \ll n^{-2/3+1/15}$ , we*  
 494 *have  $F([n]_p) \geq d(n \log n^2 p^3)^{1/3}$  w.o.p.*

495 *Proof.* Let  $\mathcal{S}[[n]_p]$ ,  $\mathcal{S}_i[[n]_p]$ ,  $X$  and  $X_i$  be as in Definitions 5.1 and 5.2. Recall that the size of an  
 496 independent set of  $\mathcal{S}[[n]_p]$  gives a lower bound on  $F([n]_p)$ .

497 We wish to apply Lemma 7.2. However, since  $\mathcal{S}[[n]_p]$  may be neither simple nor uniform, we  
 498 consider a suitable *induced* subhypergraph  $\mathcal{S}^* \subset \mathcal{S}[[n]_p]$ , as discussed just after the statement of  
 499 Lemma 7.2. We have  $\mathcal{S}[[n]_p] = \mathcal{S}_3[[n]_p] \cup \mathcal{S}_4[[n]_p]$ . Let  $\tilde{\mathcal{S}}_4$  be the set of all hyperedges in  $\mathcal{S}_4[[n]_p]$   
 500 that share at least two vertices with some other hyperedge in  $\mathcal{S}_4[[n]_p]$ . If we delete one vertex from  
 501 each hyperedge of  $\mathcal{S}_3[[n]_p] \cup \tilde{\mathcal{S}}_4$ , the remaining induced subhypergraph  $\mathcal{S}^*$  of  $\mathcal{S}[[n]_p]$  is both simple  
 502 and 4-uniform. To apply Lemma 7.2 to  $\mathcal{S}^*$ , we now estimate  $|V(\mathcal{S}^*)|$  and the average degree of  $\mathcal{S}^*$ .

503 First we consider  $|V(\mathcal{S}^*)|$ . Note that  $|[n]_p| - X_3 - |\tilde{\mathcal{S}}_4| = |[n]_p| - |\mathcal{S}_3[[n]_p]| - |\tilde{\mathcal{S}}_4| \leq |V(\mathcal{S}^*)| \leq |[n]_p|$ .  
 504 We shall show the following two facts.

505 **Fact 7.6.** *Fix  $\delta > 0$  and suppose  $n^{-1+\delta} \ll p \ll n^{-1/2}$ . We have, w.o.p.,  $X_3 = o(np)$ .*

506 **Fact 7.7.** Fix  $\delta > 0$  and suppose  $n^{-1+\delta} \ll p \ll n^{-2/3+1/15}$ . We have, w.o.p.,  $|\tilde{\mathcal{S}}_4| = o(np)$ .

507 Since the Chernoff bound gives that  $|[n]_p| = np + o(np)$  w.o.p. for  $p \gg (\log n)/n$ , Facts 7.6 and 7.7  
508 imply that, w.o.p., we have

$$|V(\mathcal{S}^*)| = np(1 + o(1)). \quad (78)$$

509 Next we consider the average degree of  $\mathcal{S}^*$ . Owing to  $\mathcal{S}^* \subset \mathcal{S}[[n]_p]$ , (78) and Lemma 5.4, the  
510 average degree  $4|\mathcal{S}^*|/|V(\mathcal{S}^*)|$  of  $\mathcal{S}^*$  is such that, w.o.p.,  $4|\mathcal{S}^*|/|V(\mathcal{S}^*)| \leq 4X/|V(\mathcal{S}^*)| \leq n^2p^3$ .

511 We now are ready to apply Lemma 7.2. In view of our average degree estimate above, we set  $t =$   
512  $(n^2p^3)^{1/3}$ . Given (78), Lemma 7.2 implies that, w.o.p., the hypergraph  $\mathcal{S}^*$ , and thus  $\mathcal{S}[[n]_p]$ , has  
513 an independent set of size

$$c \frac{(\log t)^{1/3}}{t} |V(\mathcal{S}^*)| \geq c \frac{[(1/3) \log(n^2p^3)]^{1/3}}{(n^2p^3)^{1/3}} np(1 + o(1)) \geq d(n \log(n^2p^3))^{1/3}, \quad (79)$$

514 for, say,  $d = c/2$ . This completes the proof of Lemma 7.5.  $\square$

515 In order to finish the proof of Lemma 7.5, it remains to prove Facts 7.6 and 7.7.

516 *Proof of Fact 7.6.* Lemma 5.3(ii) tells us that, w.o.p.,  $X_3 = O(\max\{n^2p^3, n^\delta\})$ . From the assump-  
517 tion  $n^{-1+\delta} \ll p \ll n^{-1/2}$ , we have both  $n^2p^3 \ll np$  and  $n^\delta \ll np$ , whence, w.o.p.,  $X_3 = o(np)$ .  $\square$

518 *Proof of Fact 7.7.* We give a sketch of the proof. Let  $\mathcal{P}$  be the family of the pairs  $\{E_1, E_2\}$  of  
519 distinct members  $E_1$  and  $E_2$  of  $\mathcal{S}_4[[n]_p]$  with  $|E_1 \cap E_2| \geq 2$ . Observe that

$$|\tilde{\mathcal{S}}_4| \leq 2|\mathcal{P}|. \quad (80)$$

520 An argument similar to one in the proof of Lemma 5.3(ii), based on the Kim–Vu polynomial  
521 concentration result, tells us that  $|\mathcal{P}| = O(\max\{\mathbb{E}[|\mathcal{P}|], n^\delta\}) = O(\max\{n^4p^6, n^\delta\})$  holds w.o.p.  
522 From the assumption  $n^{-1+\delta} \ll p \ll n^{-2/3+1/15} = n^{-3/5}$ , we have both  $n^4p^6 \ll np$  and  $n^\delta \ll np$ ,  
523 and hence  $|\mathcal{P}| = o(np)$  holds w.o.p. Given (80), we have, w.o.p.,  $|\tilde{\mathcal{S}}_4| = o(np)$ .  $\square$

524 In order to establish Lemma 7.3, we need to expand the range of  $p$  in Lemma 7.5 from  $2n^{-2/3} \leq$   
525  $p \ll n^{-2/3+1/15} = n^{-3/5}$  to  $p \geq 2n^{-2/3}$ .

526 *Proof of Lemma 7.3.* To complement the range of  $p$  covered by Lemma 7.5, it is enough to show  
527 that, say, for  $p \geq n^{-2/3+1/16}$ , we have, w.o.p.,  $F([n]_p) \geq d' (n \log(n^2p^3))^{1/3}$  for some absolute  
528 constant  $d' > 0$ . Lemma 7.5 implies that, for  $p = n^{-2/3+1/16}$ , we have, w.o.p.,

$$\begin{aligned} F([n]_p) &\geq d[n \log(n^2n^{-2+3/16})]^{1/3} = d[n \log(n^{3/16})]^{1/3} \\ &= d[n(3/16) \log n]^{1/3} > d(1/16)^{1/3} [n(2 \log n)]^{1/3} = d' [n \log n^2]^{1/3}, \end{aligned}$$

529 where  $d' = d(1/16)^{1/3}$ . By Fact 7.1, we infer that, for  $p \geq n^{-2/3+1/16}$ , we have, w.o.p.,  $F([n]_p) \geq$   
530  $d' [n \log n^2]^{1/3} \geq d' [n \log(n^2p^3)]^{1/3}$ , completing the proof of Lemma 7.3.  $\square$

531 **Acknowledgement.** The fourth author is indebted to Tomasz Schoen for drawing his attention  
532 to the problem of counting Sidon sets.

## REFERENCES

- 533
- [1] M. Ajtai, J. Komlós, J. Pintz, J. Spencer, and E. Szemerédi, *Extremal uncrowded hypergraphs*, J. Combin. Theory Ser. A **32** (1982), no. 3, 321–335. 7.1
- 534 [2] N. Alon, J. Balogh, R. Morris, and W. Samotij, *Counting sum-free sets in Abelian groups*, submitted. 3
- 535 [3] N. Alon and J. H. Spencer, *The probabilistic method*, second ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience [John Wiley & Sons], New York, 2000, With an appendix on the  
536 life and work of Paul Erdős. 5.2
- 537 [4] J. Balogh and W. Samotij, *The number of  $K_{m,m}$ -free graphs*, Combinatorica **31** (2011), no. 2, 131–150. 3
- 538 [5] ———, *The number of  $K_{s,t}$ -free graphs*, J. Lond. Math. Soc. (2) **83** (2011), no. 2, 368–388. 3
- 539 [6] B. Bollobás, *Random graphs*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1985. 2.2,  
540 2.2
- 541 [7] P. J. Cameron and P. Erdős, *On the number of sets of integers with various properties*, Number theory (Banff, AB, 1988), de Gruyter, Berlin, 1990, pp. 61–79. 1, 1.1, 1.1
- 542 [8] S. Chowla, *Solution of a problem of Erdős and Turán in additive-number theory*, Proc. Nat. Acad. Sci. India. Sect. A. **14** (1944), 1–2. 1, 1.2, 7.2
- 543 [9] D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, submitted, 70pp, 2010. 1.2, 1.2
- 544 [10] R. A. Duke, H. Lefmann, and V. Rödl, *On uncrowded hypergraphs*, Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, “Random Graphs ’93”  
545 (Poznań, 1993), vol. 6, 1995, pp. 209–212. 7.1
- 546 [11] P. Erdős, *On a problem of Sidon in additive number theory and on some related problems. Addendum*, J. London Math. Soc. **19** (1944), 208. 1, 1.2, 7.2
- 547 [12] P. Erdős and P. Turán, *On a problem of Sidon in additive number theory, and on some related problems*, J. London Math. Soc. **16** (1941), 212–215. 1, 1.2
- 548 [13] Z. Füredi, *Random Ramsey graphs for the four-cycle*, Discrete Math. **126** (1994), no. 1-3, 407–410. 3
- 549 [14] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008),  
550 no. 2, 481–547. 1.2
- 551 [15] H. Halberstam and K. F. Roth, *Sequences*, second ed., Springer-Verlag, New York, 1983. 1
- 552 [16] S. Janson, T. Łuczak, and A. Ruciński, *An exponential bound for the probability of nonexistence of a specified subgraph in a random graph*, Random graphs ’87 (Poznań, 1987), Wiley, Chichester, 1990, pp. 73–87. 1.2
- 553 [17] ———, *Random graphs*, Wiley-Interscience, New York, 2000. 2.2, 2.2, 7
- 554 [18] J. H. Kim and V. H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20**  
555 (2000), no. 3, 417–434. 5.1, 5.1, 5.2
- 556 [19] D. J. Kleitman and K. J. Winston, *On the number of graphs without 4-cycles*, Discrete Math. **41** (1982), no. 2,  
557 167–172. 3
- 558 [20] Y. Kohayakawa, *Szemerédi’s regularity lemma for sparse graphs*, Foundations of computational mathematics (Rio de Janeiro, 1997), Springer, Berlin, 1997, pp. 216–230. 1.2
- 559 [21] Y. Kohayakawa, B. Kreuter, and A. Steger, *An extremal problem for random graphs and the number of graphs with large even-girth*, Combinatorica **18** (1998), no. 1, 101–120. 3
- 560 [22] Y. Kohayakawa, T. Łuczak, and V. Rödl, *Arithmetic progressions of length three in subsets of a random set*, Acta Arith. **75** (1996), no. 2, 133–163. 1.2
- 561 [23] J. Komlós, M. Sulyok, and E. Szemerédi, *Linear problems in combinatorial number theory*, Acta Math. Acad. Sci. Hungar. **26** (1975), 113–121. 7.2
- 562 [24] K. O’Bryant, *A complete annotated bibliography of work related to Sidon sequences*, Electron. J. Combin. (2004),  
563 Dynamic surveys 11, 39 pp. (electronic). 1
- 564
- 565
- 566
- 567
- 568
- 569
- 570
- 571
- 572
- 573
- 574
- 575
- 576

- 577 [25] O. Reingold, L. Trevisan, M. Tulsiani, and S. P. Vadhan, *Dense subsets of pseudorandom sets*, FOCS, 2008,  
578 pp. 76–85. 1.2
- 579 [26] ———, *Dense subsets of pseudorandom sets*, Electronic Colloquium on Computational Complexity (ECCC) **15**  
580 (2008), no. 045. 1.2
- 581 [27] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109. 1.2
- 582 [28] M. Schacht, *Extremal results for random discrete structures*, submitted, 27pp, 2009. 1.2, 1.2
- 583 [29] J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Transactions of the  
584 American Mathematical Society **43** (1938), 377–385. 1
- 585 [30] E. Szemerédi, *On sets of integers containing no  $k$  elements in arithmetic progression*, Acta Arith. **27** (1975),  
586 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik. 1.2
- 587 [31] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge  
588 University Press, Cambridge, 2006. 1, 1.2
- 589 [32] L. Trevisan, *Guest column: additive combinatorics and theoretical computer science*, SIGACT News **40** (2009),  
590 no. 2, 50–66. 1

591 INSTITUTO DE MATEMÁTICA E ESTATÍSTICA, UNIVERSIDADE DE SÃO PAULO, RUA DO MATÃO 1010, 05508–090 SÃO  
592 PAULO, BRAZIL (Y. Kohayakawa)

593 *E-mail address:* `yoshi@ime.usp.br`

594 DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EMORY UNIVERSITY, ATLANTA, GA 30322, USA (Y. Ko-  
595 hayakawa, S. J. Lee and V. Rödl)

596 *E-mail address:* `slee242@emory.edu`, `sjlee242@gmail.com`, `rodl@mathcs.emory.edu`

597 SCHOOL OF MATHEMATICAL SCIENCES, TEL AVIV UNIVERSITY, TEL AVIV 69978, ISRAEL, AND TRINITY COLLEGE,  
598 CAMBRIDGE CB2 1TQ, UK (W. Samotij)

599 *E-mail address:* `ws299@cam.ac.uk`