A SHARP THRESHOLD FOR VAN DER WAERDEN'S THEOREM IN RANDOM SUBSETS

EHUD FRIEDGUT, HIỆP HÀN, YURY PERSON, AND MATHIAS SCHACHT

Abstract. We establish the sharpness for the threshold of van der Waerden's theorem in random subsets of $\mathbb{Z}/n\mathbb{Z}$. More precisely, for $k \geqslant 3$ and $Z \subseteq \mathbb{Z}/n\mathbb{Z}$ we say Z has the van der Waerden property if any two-colouring of Z yields a monochromatic arithmetic progression of length k. Rödl and Ruciński (1995) determined the threshold for this property for any k and we show that this threshold is sharp.

The proof is based on Friedgut's criteria (1999) for sharp thresholds, and on the recently developed container method for independent sets in hypergraphs by Saxton and Thomason, and Balogh, Morris and Samotij.

1. Introduction

A main research direction in extremal and probabilistic combinatorics in the last two decades is the extension of classical results for discrete structures to the sparse random setting. Prime examples include Ramsey's theorem for graphs and hypergraphs [2,8,9], Turán's theorem in extremal graph theory and Szemerédi's theorem on arithmetic progressions [2,12] (see also [1,3,11]). Results of that form establish the *threshold* for the classical result in the random setting. The threshold is given by a function $\hat{p} = \hat{p}(n)$ such that for every $p_0 \ll \hat{p}$ the random graph $G(n, p_0)$ (or a random binomial subset of $[n] = \{1, 2, \ldots, n\}$) with parameter p_0, asymptotically almost surely does not posses a given property, whereas if p_0 is replaced by some $p_1 \gg \hat{p}$ the property does hold asymptotically almost surely. The two statements involving p_0 and p_1 are referred to as "0-statement" and "1-statement". For the properties mentioned above it could be even shown, that optimal parameters p_0 and p_1, for which the 0-statement and the 1-statement hold, only differ by a multiplicative constant.

Date: December 21, 2015.

2010 *Mathematics Subject Classification.* 05A18 (primary), 05C80, 05D10 (secondary).

Key words and phrases. sharp thresholds, van der Waerden theorem.

The first author was supported by ISF grant 0398246.

The fourth author was supported through the *Heisenberg-Programme* of the DFG.

The cooperation of the authors was supported by the *German-Israeli Foundation for Scientific Research and Development*.

The threshold for van der Waerden's theorem is such an example and was obtained by Rödl and Ruciński in [9, 10]. We denote by $[n]_p$ the binomial random subset of $[n]$, where every element of $[n]$ is included independently with probability $p = p(n)$. Furthermore, for a subset $A \subseteq [n]$ we write $A \to (k\text{-AP})_r$ to denote the fact that no matter how one colors the elements of A with r colors there is always a monochromatic arithmetic progression with k elements in A.

Theorem 1 (Rödl & Ruciński)**.** *For every $k \geqslant 3$ and $r \geqslant 2$ there exist positive constants c_0, $c_1 > 0$ such that*

$$\lim_{n\to\infty} \mathbb{P}([n]_p \to (k\text{-}AP)_r) = \begin{cases} 0, & \text{if } p \leqslant c_0 n^{-\frac{1}{k-1}}, \\ 1, & \text{if } p \geqslant c_1 n^{-\frac{1}{k-1}}. \end{cases}$$

For the corresponding result in $\mathbb{Z}/n\mathbb{Z}$ and for two colors we close the gap between c_0 and c_1. More precisely, we show that there exist bounded sequences $c_0(n)$ and $c_1(n)$ with ratio tending to 1 as n tends to infinity such that the statement holds (see Theorem 2 below). In other words, we establish a *sharp threshold* for van der Waerden's theorem for two colors in $\mathbb{Z}/n\mathbb{Z}$.

Similarly as above for subsets of $[n]$ we write $A \to (k\text{-AP})_r$ for subsets $A \subseteq \mathbb{Z}/n\mathbb{Z}$ if any r-coloring of A yields a monochromatic arithmetic progression with k-elements in $\mathbb{Z}/n\mathbb{Z}$ and we denote by $\mathbb{Z}_{n,p}$ the binomial random subset of $\mathbb{Z}/n\mathbb{Z}$ with parameter p. With this notation at hand we can state our main result.

Theorem 2. *For all $k \geqslant 3$ there exist constants $c_1 > c_0 > 0$ and a function $c(n)$ with $c_0 \leqslant c(n) \leqslant c_1$ such that for every $\varepsilon > 0$ we have*

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{Z}_{n,p} \to (k\text{-}AP)_2) = \begin{cases} 0, & \text{if } p \leqslant (1-\varepsilon)c(n) n^{-\frac{1}{k-1}}, \\ 1, & \text{if } p \geqslant (1+\varepsilon)c(n) n^{-\frac{1}{k-1}}. \end{cases}$$

We have to insist on the setting of $\mathbb{Z}/n\mathbb{Z}$ (instead of $[n]$) since the symmetry will play a small but crucial rôle in our proof. Another shortcoming is the restriction to two colors $r = 2$ and we believe it would be very interesting to extend the result for arbitrary r. We remark that only few sharp thresholds for Ramsey properties are known (see, e.g., [6, 7]) so far.

Among other tools our proof heavily relies on the criteria for sharp thresholds of Friedgut and its extension due to Bourgain [4]. Another crucial tool is the recent *container theorem* for independent sets in hypergraphs due to Balogh, Morris and Samotij [1] and Thomason and Saxton [11]. The proof extends to other Ramsey properties for two colors, as long as the corresponding extremal problem is *degenerate*, i.e., positive density yields many copies of the target structure and the target structure is strictly balanced with respect to its so-called 2-*density*. For example, even cycles in graphs, complete k-partite, k-uniform hypergraphs, and strictly balanced, density regular Rado systems (see [10]) satisfy these assumptions. Moreover, Schacht and Schulenburg [13] noted that the approach undertaken here can be refined to give a shorter proof for the sharp threshold of the Ramsey property for triangles and two colors from [7] and, more generally, for arbitrary odd cycles.

2. Locality of coarse thresholds

In [4] Friedgut gave a necessary condition for a graph property to have a coarse threshold, namely, that it is approximable by a "local" property. In the appendix to this work Bourgain proved a similar result for more general discrete structures. Here we state the special case applicable for properties in $\mathbb{Z}/n\mathbb{Z}$.

Theorem 3 (Bourgain)**.** *There exist functions $\delta(C,\tau)$ and $K(C,\tau)$ such that the following holds. Let $p = o(1)$ as n goes to infinity, let $\mathcal{A}$ be a monotone family of subsets of $\mathbb{Z}/n\mathbb{Z}$, with*

$$\tau < \mu(p,\mathcal{A}) := \mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{A}) < 1-\tau\,,$$

and assume also $p \cdot \frac{d\mu(p,\mathcal{A})}{dp} \leqslant C$. Then there exists some $B \subseteq \mathbb{Z}/n\mathbb{Z}$ with $|B| \leqslant K$ such that

$$\mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{A} \,|\, B \subseteq \mathbb{Z}_{n,p}) > \mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{A}) + \delta\,. \tag{1}$$

Note that whenever a property $\mathcal{A}$ (or rather, a series of properties $\mathcal{A}_n$) has a coarse threshold there exist constants C and τ such that for infinitely many values of n the hypothesis of the theorem holds. For applications, it would be problematic if there exists a B with $|B| \leqslant K$ and $B \in \mathcal{A}$, since this would trivialise the conclusion (1). However, as observed in [5], the above theorem can be strengthened, without modifying the original proof, to deduce that the set of B's for which the assertion holds has non-negligible measure, i.e., there exists a family $\mathcal{B}$ such that

$$\mathbb{P}(B \subseteq \mathbb{Z}_{n,p} \text{ for some } B \in \mathcal{B}) > \eta,$$

where $\eta > 0$ depends only on C and τ but not on n, and every $B \in \mathcal{B}$ satisfies the conclusion of Theorem 3, i.e., $|B| \leqslant K$ and

$$\mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{A} \,|\, B \subseteq \mathbb{Z}_{n,p}) > \mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{A}) + \delta\,.$$

This allows us to make assumptions about B in the application below, as long as the set of B's violating the assumptions has negligible measure. In particular, Lemma 4 below asserts that any B of bounded size fails to have the van der Waerden property, i.e., $B \nrightarrow (3\text{-AP})_2$.

Lemma 4. *Let $\mathcal{B}$ be a family of subsets of $\mathbb{Z}/n\mathbb{Z}$ with the property that every $B \in \mathcal{B}$ satisfies $|B| \leqslant \log n$ and $B \to (k\text{-}AP)_2$. Then for every function $p = p(n) = \Theta(n^{-\frac{1}{k-1}})$ we have $\mathbb{P}(B \subseteq \mathbb{Z}_{n,p}$ for some $B \in \mathcal{B}) = o(1)$.*

Lemma 4 was implicitly proved in [10, Section 7] (see Deterministic and Probabilistic Lemma there). However, there are two differences which we discuss below. Firstly, in [10] subsets of $[n]$ were considered, but it is easy to check that the same argument can be applied in $\mathbb{Z}/n\mathbb{Z}$. Secondly, in [10, Section 7] the 0-statement of Theorem 1 (in fact, for general Rado-system) is given. As a consequence, the Probabilistic Lemma in [10] concerns configurations $B \subseteq [n]$ of arbitrary size with $B \to (k\text{-AP})_2$. In order to rule out those of size larger than $\log n$ it is required that $p < c_0 n^{-\frac{1}{k-1}}$ for some sufficiently small $c_0 > 0$. However, in Lemma 4 we only consider such configurations B of size at most $\log n$ and one can check that in the proof in [10] these can be excluded as long as $p = O(n^{-\frac{1}{k-1}})$.

We summarise the discussion above in the following corollary of Theorem 3, which is tailored for our proof of Theorem 2.

Corollary 5. *Assume that the property $\{Z \subseteq \mathbb{Z}/n\mathbb{Z} \colon Z \to (k\text{-}AP)_2\}$ does not have a sharp threshold. Then there exist constants c_1, c_0, α, ε, $\mu > 0$, and K and a function $c(n) \colon \mathbb{N} \to \mathbb{R}$ with $c_0 < c(n) < c_1$ so that for infinitely many values of n and $p = c(n)n^{-\frac{1}{k-1}}$ the following holds.*

There exists a subset B of $\mathbb{Z}/n\mathbb{Z}$ of size at most K with $B \nrightarrow (k\text{-}AP)_2$ such that for every family $\mathcal{Z}$ of subsets from $\mathbb{Z}/n\mathbb{Z}$ satisfying $\mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{Z}) > 1 - \mu$ there exists a $Z \in \mathcal{Z}$ so that

- *(a) $\mathbb{P}(Z \cup (B + x) \to (k\text{-}AP)_2) > \alpha$, where $x \in \mathbb{Z}/n\mathbb{Z}$ is chosen uniformly at random, and*
- *(b) $\mathbb{P}(Z \cup \mathbb{Z}_{n,\varepsilon p} \to (k\text{-}AP)_2) < \alpha/2$.*

We remark that the $\mathbb{P}(\cdot)$ in Corollary 5 concern different probability spaces. While the assumption $\mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{Z}) > 1 - \mu$ concerns the binomial random subset $\mathbb{Z}_{n,p}$, we consider x chosen uniformly at random from $\mathbb{Z}/n\mathbb{Z}$ in *(a)* and the binomial random subset $\mathbb{Z}_{n,\varepsilon p}$ in *(b)*. We close this section with a short sketch of the proof of Corollary 5.

Proof (Sketch). For $k \geqslant 3$ we consider the property $\mathcal{A} = \{Z \subseteq \mathbb{Z}/n\mathbb{Z} \colon Z \to (k\text{-AP})_2\}$ and assume that it does not have a sharp threshold. Consequently, there exists a function $p = p(n)$ such that for infinitely many n the assumptions of Theorem 3 hold, which implicitly yields constants C, τ, δ, and K. Let $\bar{\mathcal{A}} = \mathcal{P}(\mathbb{Z}/n\mathbb{Z}) \setminus \mathcal{A}$ be the family of subsets of $\mathbb{Z}/n\mathbb{Z}$ that fail to have the van der Waerden property. Since we assume that the threshold for $\mathcal{A}$ is not sharp, we may fix $\varepsilon > 0$ sufficiently small, such that there must be some α with $\delta/2 > \alpha > 0$ so that if we let $\mathcal{Z}' \subseteq \bar{\mathcal{A}}$ be the sets $Z \in \bar{\mathcal{A}}$ for which

$$\mathbb{P}(Z \cup \mathbb{Z}_{n,\varepsilon p} \rightarrow (k\text{-AP})_2) < \alpha/2 \tag{2}$$

then $\mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{Z}' \mid \mathbb{Z}_{n,p} \in \bar{\mathcal{A}}) \geqslant 1 - \delta/4$.

Also for $p = p(n)$ we have $\tau < \mathbb{P}(\mathbb{Z}_{n,p} \rightarrow (k\text{-AP})_2) = \mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{A}) < 1 - \tau$, so by Theorem 1 there exist some constants $c_1 \geqslant c_0 > 0$ such that $p = p(n) = c(n)n^{-\frac{1}{k-1}}$ for some function $c(n)\colon \mathbb{N} \rightarrow \mathbb{R}$ satisfying $c_0 \leqslant c(n) \leqslant c_1$. Strictly speaking, we should use the version of Theorem 1 for $\mathbb{Z}/n\mathbb{Z}$ instead of $[n]$. However, it is easy to see that the 1-statement for random subsets of $[n]$ implies the 1-statement for random subsets of $\mathbb{Z}/n\mathbb{Z}$ (up to a different constant c_1) and, as we discussed in the paragraph following Lemma 4, the proof of the 0-statement can be straightforwardly adjusted for subsets of $\mathbb{Z}/n\mathbb{Z}$.

Moreover, for any such n Theorem 3 yields a family $\mathcal{B}$ of subsets of $\mathbb{Z}/n\mathbb{Z}$ each of size at most K such that (1) holds and an element of $\mathcal{B}$ appears as a subset of $\mathbb{Z}_{n,p}$ with probability at least η. Consequently, Lemma 4 asserts that at least one such $B \in \mathcal{B}$ fails to have the van der Waerden property itself, i.e., $B \nrightarrow (k\text{-AP})_2$. By symmetry it follows from (1), that the same holds for every translate $B + x$ with $x \in \mathbb{Z}/n\mathbb{Z}$. In particular, consider the family $\mathcal{Z}'' \subseteq \bar{\mathcal{A}}$ of all sets $Z \in \bar{\mathcal{A}}$ such that for at least $(\delta/2)n$ translates $B + x$ we have $Z \cup (B + x) \rightarrow (k\text{-AP})_2$, i.e.,

$$\mathbb{P}(Z \cup (B + x) \rightarrow (k\text{-AP})_2) > \delta/2 > \alpha$$

for x chosen uniformly at random from $\mathbb{Z}/n\mathbb{Z}$. Then $\mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{Z}'' \mid \mathbb{Z}_{n,p} \in \bar{\mathcal{A}}) \geqslant \delta/2$. So, taking $\mu < \delta \cdot \mathbb{P}(\mathbb{Z}_{n,p} \in \bar{\mathcal{A}})/8$ we have that if $\mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{Z}) \geqslant 1 - \mu$ then $\mathcal{Z} \cap \mathcal{Z}' \cap \mathcal{Z}'' \neq \varnothing$. Any Z in this non-empty family has the desired properties. □

3. Lemmas and the proof of the main theorem

In this section we state all the necessary notation and lemmas to give the proof of Theorem 2. We start with an outline of this proof.

3.1. **Outline of the proof.** The point of departure is Corollary 5 and we will derive a contradiction to its second property. To this end, we consider an appropriate set Z as given by Corollary 5 and let Φ denote the set of all colorings of Z without a monochromatic k-AP. The main obstacle is to find a partition of Φ into $i_0 = 2^{o(pn)}$ classes $\Phi_1, \ldots, \Phi_{i_0}$, such that any two colorings φ, φ' from any partition class Φ_i agree on a relatively dense subset C_i of Z, i.e. $\varphi(z) = \varphi'(z)$ for all $z \in C_i$. Let B_i denote the larger monochromatic subset of C_i, say of color blue. Note that Corollary 5 allows us to impose further conditions on Z as long as $\mathbb{Z}_{n,p}$ satisfies them almost surely. One of these properties is that the "focus" of B_i, denoted by $F(B_i)$, is of a linear size, i.e. there are a linear number of elements each of which forms a k-AP with some $(k-1)$-tuple of elements in B_i. By a quantitative version of Szemerédi's theorem we know that the number of k-APs in the focus of B_i is $\Omega(n^2)$. Consider $U_i = (\mathbb{Z}/n\mathbb{Z})_{\varepsilon p} \cap F(B_i)$ and note that if any element of U_i is colored blue then this induces a blue k-AP with Z under any coloring $\varphi \in \Phi_i$. Hence, to extend any $\varphi \in \Phi_i$ to a coloring of $Z \cup U_i$ without a monochromatic k-AP it is necessary that all elements in U_i are colored red. Consequently, the probability of a successful extension of any coloring in Φ_i is bounded from above by the probability that U_i does not contain a k-AP. This, however, is at most $\exp(-\Omega(p^k n^2)) = \exp(-\Omega(pn))$ by Janson's inequality. We conclude by the union bound that after the second round, i.e. $\mathbb{Z}_{n,\varepsilon p}$, the probability that any k-AP-free coloring of Z survives is $i_0 \exp(-\Omega(pn)) = o(1)$ with contradiction to (*b*).

To establish the above mentioned partition of Φ we will define an auxiliary hypergraph H in such a way that every $\varphi \in \Phi$ can be associated with a hitting set of H. As the complements of hitting sets are independent and as H will be "wellbehaved" we can apply a structural result of Balogh, Morris and Samotij [1] on independent sets in uniform hypergraphs (see Theorem 15) to "capture" the hitting sets of H and hence a partition of Φ with the properties mentioned above (see Lemma 8). Next we will introduce the necessary concepts along with the lemmas needed to give the proof of the main theorem. The proof of Theorem 2 will then be given in Section 3.3.

3.2. **Lemmas.** We call (Z, B) an *interacting pair* if $Z \nrightarrow (k\text{-AP})_2$ and $B \nrightarrow (k\text{-AP})_2$ but $Z \cup B \rightarrow (k\text{-AP})_2$. Further, (Z, B, X) is called an *interacting triple* if $(Z, B + x)$ is interacting for all $x \in X$. Note that Corollary 5 asserts that there is an interacting triple (Z, B, X) with $|X| > \alpha n$. In the following we shall concentrate on elements which are decisive for interactions. Given a (not necessarily interacting) pair (A, B) we say that an element $a \in A$ *focuses* on $b \in B + x$ if there are $k-2$ further elements in $A \cup B$ forming a k-AP with a and b.

The set of vertices of particular interest, given a pair (A, B), is

$$M(A, B) = \{a \in A\colon \text{there is a } b \in B \text{ such that } a \text{ focuses on } b\}$$

and for a triple (A, B, X) we define the hypergraph $H = H(A, B, X)$ with the vertex set A and the edge set consisting of all $M(A, B + x)$ with $x \in X$. We are interested in the hypergraph $H(Z, B, X)$ with an interacting triple (Z, B, X). We will make use of the fact that Corollary 5 allows us to put further restrictions on Z as long these events occur a.a.s.

for $\mathbb{Z}_{n,p}$. The requirement we want to make is that the maximum degree and co-degree of $H(Z, B, X)$ are well behaved.

Lemma 6. *For given C_1, k, K and all $B \subset \mathbb{Z}/n\mathbb{Z}$ of size $|B| \leqslant K$ the following holds a.a.s. for $2\log n \leqslant p \leqslant C_1 n^{-1/(k-1)}$: There is a set $Y \subset \mathbb{Z}/n\mathbb{Z}$ of size at most $n^{1-1/(k-1)}\log n$ such that the hypergraph $H = H(\mathbb{Z}_{n,p}, B, (\mathbb{Z}/n\mathbb{Z})\backslash Y)$ satisfies*

(1) $2pn \geqslant v(H) \geqslant pn/2$,
(2) $\Delta(H) \leqslant 10k^3Kp^{k-2}n$
(3) $\Delta_2(H) \leqslant 8\log n$.

We postpone the proof of Lemma 6. It can be found in Section 4.

A set of vertices of a hypergraph is called a hitting set if it intersects every edge of this hypergraph. The conditions in Lemma 6 will be used to control the hitting sets of $H(Z, B, X)$ which play an important rôle as explained in the following. A coloring of a set is called k-AP free if it does not exhibit a monochromatic k-AP. For an interacting triple (Z, B, X) we fix a k-AP free coloring of B and the same coloring for all its translates $B + x$. More precisely, let $\sigma\colon [|B|] \to \{\text{red,blue}\}$ be such that for each $x \in X$ the coloring $\sigma_x\colon (B + x) \to \{\text{red,blue}\}$ is k-AP free, where σ_x is defined via $\sigma_x(b_i) = \sigma(i)$ with $b_1 < \cdots < b_{|B|}$ being the elements of $B + x$. Such a function σ exists due to $B \not\rightarrow (k\text{-AP})_2$.

For any k-AP-free coloring φ of Z and any $x \in X$ the coloring of $Z \cup (B+x)$ induced by σ_x and φ must exhibit a monochromatic k-AP (intersecting both Z and $B+x$) since $(Z, B+x)$ is interacting. Hence, for each $x \in X$ the edge $M(Z, B + x)$ contains an element z focussing on an element $b \in B + x$ such that $\varphi(z) = \sigma_x(b)$. Such a vertex $z \in M(Z, B + x)$ we call *activated* by σ_x and φ and we define the set of activated vertices

$$A_\varphi^{\sigma_x}(Z, B + x) = \{z \in Z\colon z \text{ is activated by } \sigma_x \text{ and } \varphi\}$$

which is a non-empty subset of $M(Z, B + x)$.

Observation 7. *Given an interacting triple (Z, B, X), a k-AP free coloring $\sigma\colon [|B|] \to \{red,blue\}$ of B and the same for all its translates $B + x$, $x \in X$. Further, let φ be a k-AP free coloring of Z. Then the set of activated vertices*

$$A_\varphi = A_\varphi^\sigma(Z, B, X) = \bigcup_{x \in X} A_\varphi^{\sigma_x}(Z, B + x)$$

is a hitting set of $H(Z, B, X)$.

The following lemma shows that the hitting sets of well-behaved uniform hypergraphs can be "captured" by a small number of sets of big size called cores.

Lemma 8. *For every natural $k \geqslant 3$, $\ell \geqslant 2$ and all positive c_0, c_1 there are c_2 and $\beta > 0$ such that the following holds.*

If H if an ℓ-uniform hypergraph with m vertices and $c_0 m^{1+1/(k-2)}$ edges such that $\Delta_1(H) \leqslant c_1 m^{1/(k-2)}$ and $\Delta_2(H) \leqslant c_1 \log m$ then there is a family $\mathcal{C}$ of subsets of $V(H)$, which we shall call cores, *such that*

(*i*) *for $t = 1 - 1/(k-2)(\ell-1)$ we have*

$$|\mathcal{C}| \leqslant \sum_{i \in [c_2 m^t]} \binom{m}{i}$$

(*ii*) *$|C| \geqslant \beta m$ for every $C \in \mathcal{C}$, and*

(*iii*) *every hitting set of H contains some C from $\mathcal{C}$.*

Lemma 8 will follow from the main result from [1]. The proof can be found in Section 5.

As it turns out, we can insist that the interacting triple (Z, B, X) guaranteed by Corollary 5 has the additional property that X contains a suitable subset $X' \subset X$ so that the hypergraph $H(Z, B, X')$ is uniform. In this case Lemma 8 allows us to partition the sets of all k-AP free colorings of Z into a small number of partition classes $\{\Phi_C\}_{C \in \mathcal{C}}$, each represented by a big core $C \in \mathcal{C}$.

However, we wish to refine the partition classes further so that every two colorings $\varphi, \varphi' \in \Phi_C$ from the same partition class agree on a large vertex set. This can be accomplished by applying Lemma 8 to $H(Z, B, X')$ for a more refined subset $X' \subset X$. Indeed, we will make sure that there is a set which guarantees that the colors of the activated vertices A_φ under φ as defined in Observation 7 is already "determined" by σ. This particularly implies that any two colorings $\varphi, \varphi' \in \Phi_C$ agree on $A_\varphi \cap A_{\varphi'}$, hence, on the core C representing them, i.e. $\varphi(z) = \varphi'(z)$ for all $z \in C$. To make this formal we need the following definitions. A triple (Z, B, X) is called *regular* if for all $x \in X$ every element of Z focuses on at most one element in $B + x$. Given a regular triple (Z, B, X) and an $x \in X$ let $z_1 < \cdots < z_\ell$ denote the elements of $M_x = M(Z, B + x) \in H(Z, B, X)$. We say that $z \in M(Z, B + x)$ has *index* i if $z = z_i$ and the triple (Z, B, X) is called *index consistent* if for any element $z \in Z$ and any two edges M_x, $M_{x'}$ containing z the indices of z in M_x and $M_{x'}$ are the same.

Further, let $b_1 < \cdots < b_{|B|}$ be the elements of $B + x$. We associate to the edge M_x its *profile* which is the function $\pi \colon [\ell] \to [|B|]$ indicating which z_i is focussed on which b_j, formally:

$\pi(i) = j$ if z_i focusses on b_j. Since (Z, B, X) is regular, each $z \in M_x$ focuses on exactly one $b \in B + x$, thus, the profile of M_x is well-defined and unique. We call ℓ the *length* of the profile and we say that the triple (Z, B, X) has profile π with length ℓ if all edges of $H(Z, B, X)$ does. We summarize the desired properties for the hitting sets of $H(Z, B, X)$ associated to k-AP free colorings of Z.

Observation 9. *Suppose the triple (Z, B, X) in Observation 7 is index consistent and has profile π. Let $A_\varphi = A^\sigma_\varphi(Z, B, X)$ be the vertex set activated by φ and σ as defined in Observation 7. Then for any vertex $z \in A_\varphi$ the color $\varphi(z)$ of z is already determined by σ and the (unique) index i of z, indeed, $\varphi(z) = \sigma(i)$. In particular, any two k-AP free coloring φ and φ' of Z agree on $A_\varphi \cap A_{\varphi'}$, i.e. $\varphi(z) = \varphi'(z)$ for all $z \in A_\varphi \cap A_{\varphi'}$.*

The following lemma will allow us to restrict the consideration to index consistent triples with a bounded length profile.

Lemma 10. *For all $C_1 > 0$, k, K and $\alpha > 0$ there exist L and $\alpha' > 0$ such that for all $B \subset \mathbb{Z}/n\mathbb{Z}$ of size $|B| \leqslant K$ and $p \leqslant C_1 n^{-1/(k-1)}$ the following holds a.a.s. There is a set $Y_n \subset \mathbb{Z}/n\mathbb{Z}$ of size at most $n^{1-1/(k-1)} \log n$ such that for every set $X \subset \mathbb{Z}/n\mathbb{Z}$ of size $X \geqslant \alpha n$ there is a set $X' \subset X \backslash Y_n$ of size $|X'| \geqslant \alpha' n$ and a profile π of length at most L such that $(\mathbb{Z}_{n,p}, B, X')$ is index consistent and has profile π.*

The proof of Lemma 10 can be found in Section 4. Lastly, we put another restriction on Z as to make sure that any relatively dense subset of any core create many k-AP's for the second round. The proof of Lemma 11 can be found in Section 6.

Lemma 11. *For every $c_0 > 0$ and $\gamma > 0$ there is a $\delta > 0$ such that for $p \geqslant c_0 n^{-1/(k-1)}$ a.a.s. the following holds. The size of $\mathbb{Z}_{n,p}$ is at most $2pn$ and for every subset $S \subset \mathbb{Z}_{n,p}$ of size $|S| > \gamma pn$ the set*

$$F(S) = \{z \in \mathbb{Z}/n\mathbb{Z} \colon \text{there are } a_1, \ldots, a_{k-1} \in S \text{ which forms a } k\text{-AP with } z\}$$

has size at least δn.

The proof of Lemma 11 can be found in Section 6. We are now in the position to prove the main theorem.

3.3. **Proof of the main theorem.** The proof of the main theorem uses the lemmas introduced in the previous section and follows the scheme described.

Proof of Theorem 2. For a given $k \geqslant 3$ assume for a contradiction that the property $\mathbb{Z}_{n,p} \to (k\text{-AP})$ does not have a sharp threshold. By Corollary 5 there exist $c_0 \leqslant c(n) \leqslant c_1$, $p = c(n)n^{-1/(k-1)}$, α, ε, $\mu > 0$ and K such that there exist infinitely many n for which there exists a subset $B \subset \mathbb{Z}/n\mathbb{Z}$ of size at most K such that $B \twoheadrightarrow (k\text{-AP})_2$.

We apply Lemma 10 with C_1, k, K and α to obtain L and $\alpha' > 0$. For each $2 \leqslant \ell \leqslant L$ we apply Lemma 8 with the constants k, ℓ, $c_0 = \alpha'/(2C_1)$ and $c_1 = 2(10k^3K)^{k-2}C_1^{(k-2)^2-1}$ to obtain $c_2(\ell)$ and $\beta(\ell) > 0$. Let $c_2 = \max\{c_2(\ell)\colon 2 \leqslant \ell \leqslant L\}$ and $\beta = \min\{\beta(\ell)\colon 2 \leqslant \ell \leqslant L\}$ and we apply Lemma 11 with C_0 and $\gamma = \beta/2$ to obtain $\delta > 0$.

For each n we define $\mathcal{Z}_n$ to be the sets of subsets $Z \subset \mathbb{Z}/n\mathbb{Z}$ which satisfies the conclusions of Lemma 6, Lemma 10 and Lemma 11 (with $\mathbb{Z}_{n,p}$ replaced by Z) with the constants given and chosen from above. As the conclusions of these lemmas holds a.a.s. for $\mathbb{Z}_{n,p}$ we know for sufficiently large n that $\mathbb{P}(\mathbb{Z}_{n,p} \in \mathcal{Z}_n) > 1 - \mu$ hence, by Corollary 5 there is an interacting triple (Z, B, X) such that $|B| < K$, $|X| \geqslant \alpha n$ and $Z \in \mathcal{Z}_n$. In particular, since Z satisfies the conclusion of Lemma 6 and Lemma 10 there exists a set Y_n of size at most $2n^{1-1/(k-1)}\log n$ and a profile π of length $1 \leqslant \ell \leqslant L$ and a set $X' \subset X \backslash Y_n$ such that

- $|X'| \geqslant \alpha' n$
- the (interacting) triple (Z, B, X') is index consistent and has profile π,
- the hypergraph $H = H(Z, B, X')$ satisfies $2pn \geqslant v(H) = |Z| \geqslant pn/2$, the maximum degree of H satisfies $\Delta(H) \leqslant 10k^3Kp^{k-2}n$ and $\Delta_2(H) \leqslant 8\log n$.

As (Z, B, X') is particularly regular we have $\ell \geqslant k-1 \geqslant 2$ by definition. Further, H is an ℓ-uniform hypergraph on $m = |Z|$ vertices which satisfies the presumptions of Lemma 8 with the constants chosen above. Hence, by the conclusions of Lemma 8 we obtain a family $\mathcal{C}$ of cores, such that

(*i*) for $t = 1 - 1/(k-2)(\ell-1)$ we have

$$|\mathcal{C}| = \sum_{i \in [c_2 m^t]} \binom{m}{i}$$

(*ii*) $|C| \geqslant \beta m$ for every $C \in \mathcal{C}$, and

(*iii*) every hitting set of H contains some C from $\mathcal{C}$.

Let Φ be the set of all k-AP free colorings of Z. By Observation 7 and Observation 9 we can associate to each $\varphi \in \Phi$ a hitting set A_φ of H such that any two colorings $\varphi, \varphi' \in \Phi$ agree on $A_\varphi \cap A_{\varphi'}$, i.e. $\varphi(z) = \varphi'(z)$ for all $z \in A_\varphi \cap A_{\varphi'}$. For any $C \in \mathcal{C}$ we define Φ_C to be the set of $\varphi \in \Phi$ such that $C \subset A_\varphi$ and obtain $\Phi = \bigcup_{C \in \mathcal{C}} \Phi_C$. Clearly, for any $C \in \mathcal{C}$, any two $\varphi, \varphi' \in \Phi_C$ agree on $C \subset A_\varphi \cap A_{\varphi'}$. Let $B_C \subset C$ be the larger monochromatic subset of C under (any)

$\varphi \in \Phi_C$, say of color blue. Then B_C has size $|B_C| \geqslant |C|/2 \geqslant \gamma pn$ and as $Z \in \mathcal{Z}_n$ we know by Lemma 11 that $|F(B_C)\backslash Z| > \delta n/2$. Let $\mathcal{P}(C)$ denote the set of all k-AP contained in $F(B_C)$. By the quantitavive version of Szemerédi's theorem (see [14]) we know that there is an $\eta > 0$ such that for sufficiently large n we have $|\mathcal{P}(C)| \geqslant \eta n^2$. Consider the second round exposure $U_C = \mathbb{Z}_{n,\varepsilon p} \cap F(B_C)$ and let t_i be the indicator random variable for the event $i \in U_C$. We are interested in the probability that there is a $\varphi \in \Phi_C$ which can be extended to a k-AP free coloring of $Z \cup U_C$. To extend any coloring $\varphi \in \Phi_C$ of Z to a k-AP free coloring of $Z \cup U_C$, however, it is necessary that $U_C \subset F(B_C)$ is completely colored red, i.e. that U_C does not contain any k-AP. This probability can be bounded using Janson's inequality for $X = \sum_{P\in\mathcal{P}(C)} \prod_{i\in P} t_i$ given by

$$\mathbb{P}\left(X = 0\right) \leqslant \exp\left\{-\frac{\mathbb{E}\left(X\right)^2}{2\Delta}\right\}$$

where

$$\Delta = \sum_{A,B\in\mathcal{P}(C)\colon A\cap B\neq\varnothing} \mathbb{E}\left(\prod_{i\in A\cup B} t_i\right) \leqslant p^{2k-1}n^3 + p^{k+1}k^2n^2 \leqslant 2p^{2k-1}n^3$$

for large enough n. We obtain

$$\begin{aligned}\mathbb{P}\left(\exists\varphi \in \Phi_C\colon \varphi \text{ can be extended to a } k\text{-AP free coloring of } Z \cup U_C\right)\\ \leqslant \mathbb{P}\left(U_C \text{ does not contain a } k\text{-AP}\right) < \exp\{-\eta^2 pn/4\}.\end{aligned}$$

Taking union bound we conclude

$$\mathbb{P}\left((Z \cup \mathbb{Z}_{n,\varepsilon p}) \nrightarrow (k\text{-AP})_2\right) \leqslant |\mathcal{C}| \exp\{-\eta^2 pn/4\}$$

which goes to zero as n goes to infinity. This, however, is a contradiction to the property ((*b*)) of Corollary 5. ☐

4. Proofs of the Lemmas 6 and 10

In this section we prove the lemmas introduced in the previous section. We start with some technical observations. Given $B \subset \mathbb{Z}/n\mathbb{Z}$ and an element $z \in (\mathbb{Z}/n\mathbb{Z})\backslash B$ let

$$\mathcal{P}(z, B) = \left\{P \subset \mathbb{Z}/n\mathbb{Z}\colon \text{ There is a } b \in B \text{ such that } P \cup \{z, b\} \text{ forms a } k\text{-AP}\right\}$$

and let $\mathcal{P}(z, z', B) = \mathcal{P}(z, B) \times \mathcal{P}(z', B)$ where z and z' need not to be distinct. Further, let $\mathcal{P}(z, B, \mathbb{Z}/n\mathbb{Z}) = \bigcup_{x\in\mathbb{Z}/n\mathbb{Z}} \mathcal{P}(z, B + x)$ and in the same manner define $\mathcal{P}(z, z', B, \mathbb{Z}/n\mathbb{Z})$.

Fact 12. *Let $z, z' \in \mathbb{Z}/n\mathbb{Z}$ and $a \in \mathbb{Z}/n\mathbb{Z}$ be given. Then*

(1) *the number of* $P \in \mathcal{P}(z, B, \mathbb{Z}/n\mathbb{Z})$ *such that* $a \in P$ *is at most* $k^3|B|$.
(2) *the number of pairs* $(P, P') \in \mathcal{P}(z, z', B, \mathbb{Z}/n\mathbb{Z})$ *such that* $a \in P \cup P'$ *is at most* $2k^5|B|^2$.

Proof. We only prove the second property. For that we count the number of pairs $(P, P') \in \mathcal{P}(z, z', B, \mathbb{Z}/n\mathbb{Z})$ such that a is in, say, P. Recall that there must exist $x \in \mathbb{Z}/n\mathbb{Z}$ and $b, b' \in B + x$ such that $P \cup \{z, b\}$ and $P' \cup \{z', b'\}$ are both k-APs. Choosing the positions of z and a uniquely determines the first k-AP. There are at most $(k-2)$ choices for b to be contained in the k-AP and at most $|B|$ choices of x such that $b \in B + x$. Each such choice determines P and moreover, gives rise to at most $|B|$ choices for b'. Choosing the positions of b' and z' then determines the second k-AP, hence also P'. □

We define

$$\mathcal{P}_0(z, B) = \big\{P \in \mathcal{P}(z, B) \colon P \text{ and } B \text{ are disjoint}\big\}$$

and

$$\mathcal{P}_0(z, z', B) = \big\{(P, P') \in \mathcal{P}(z, z', B) \colon P \cup \{z\}, P' \cup \{z'\} \text{ and } B \text{ are pairwise disjoint}\big\}.$$

Further, let

$$\mathcal{P}_1(z, B) = \mathcal{P}(z, B) \backslash \mathcal{P}_0(z, B) \quad \text{and} \quad \mathcal{P}_1(z, z', B) = \mathcal{P}(z, z', B) \backslash \mathcal{P}_0(z, z', B).$$

For given sets $A, B \subset \mathbb{Z}/n\mathbb{Z}$ we call $x \in \mathbb{Z}/n\mathbb{Z}$ *bad* (with respect to A and B) if

(1) there are $z \in A \backslash B + x$ and $P \in \mathcal{P}_1(z, B + x)$ such that $P \cup \{z\} \subset A \cup B + x$ or
(2) there are $z, z' \in A \backslash B + x$ and $(P, P') \in \mathcal{P}_1(z, z', B + x)$ such that $P \cup P' \cup \{z, z'\} \subset A \cup B + x$.

Fact 13. *Let* $p \leqslant C_1 n^{1-1/(k-1)}$ *and let* B *be a set of constant size. Let* Y_n *be the set of bad elements with repect to* $\mathbb{Z}_{n,p}$ *and* B*. Then a.a.s.* $|Y_n| < pn \log n$.

Proof. We will show that the expected size of Y_n is of order pn so that the statement follows from Markov's inequality.

For a fixed x we first deal with the case that x is bad due to the first property, i.e. there is a k-AP $P \cup \{z, b\}$ with $b \in B + x$, P intersecting $B + x$ and $z \in \mathbb{Z}_{n,p}$ which does not belong to $B + x$. Note that after choosing b, one common element of P and $B + x$ and their positions the k-AP is uniquely determined. Then there are at most $(k-2)$ choices for z each of which uniquely determines one P. Hence the probability that x is bad due to the first property is at most $|B|k^3p$ and summing over all x we conclude that the expected bad elements due to the first property is at most $|B|k^3pn$.

If x is bad due to the second property then there are two k-APs $P \cup \{z, b\}$ and $P' \cup \{z', b'\}$ such that two of the three sets $P \cup \{z\}$, $P' \cup \{z'\}$, $B+x$ intersect and $P \cup P' \cup \{z, z'\} \subset \mathbb{Z}_{n,p} \cup B+x$ where z and z' are not in $B + x$.

We distinguish two cases and first consider all tuples (P, P', z, z', b, b') with the above mentioned properties such that P (or P' respectively) does not intersect $B + x$. Note that with this additional property the probability that x is bad due to (P, P', z, z', b, b') is at most p^k since z, z' and P all need to be in $\mathbb{Z}_{n,p}$. First, we count the number of such tuples with the additional property that P' (or P respectively) also has empty intersection with $B + x$. This implies that P and P' must intersect and in this case, choosing $b, b' \in B + x$ and one common element $a \in P \cap P'$ and the positions of b, b', a in the k-AP's uniquely determines both k-APs. After these choices there are at most k^2 choices for z, z'. Hence, there are at most $|B|^2 k^5 n$ such tuples for a fixed x.

Next, we count the number of tuples (P, P', z, z', b, b') with the property that P' and $B + x$ intersect. In this case choosing $b' \in B + x$ and one element in $P' \cap B + x$ and their positions in the k-AP uniquely determines the second k-AP. Choosing $b \in B + x$, another element $a \in \mathbb{Z}/n\mathbb{Z}$, and their positions determines the first k-AP. After these choices there are at most k^2 choices for z, z' hence in total there are at most $|B|^3 k^6 n$ such tuples for a fixed x. We conclude that the expected number of bad x due to tuples (P, P', z, z', b, b') such that P (or P' respectively) does not intersect $B + x$ is at most $p^k 2n(|B|^2 k^5 n + |B|^3 k^6 n)$.

It is left to consider the tuples (P, P', z, z', b, b') such that $(P, P') \in \mathcal{P}_1(z, z', B + x)$ and P and P' both intersect $B + x$. In this case choosing b, b', the element(s) in $P \cap B + x$ and $P' \cap B + x$ and their positions uniquely determine the two k-APs. Since $z, z' \in \mathbb{Z}_{n,p}$ with probability p^2 the expected number of bad x due to tuples (P, P', z, z', b, b') with the above mentioned property is at most $|B|^4 k^6 p^2 n$.

Hence the expected size of Y_n is

$$|Y_n| < 2p^k n(|B|^2 k^5 n + |B|^3 k^6 n) + |B|^4 k^6 p^2 n < 5|B|^3 k^6 p^k n^2,$$

as claimed. □

Fact 14. *Let $\ell \geqslant 1$ be an integer and let F be an ℓ-uniform hypergraph on the vertex set $\mathbb{Z}/n\mathbb{Z}$ which has maximum vertex degree at most D. Let $U = \mathbb{Z}_{n,p}$ with $p = cn^{-1/(k-1)}$. Then with probability at most $2Dn^{-4}$ we have $e(F[U]) > 5D(p^\ell n/\ell + \log n)$.*

Proof. For $\ell = 1$ the bound directly follows from Chernoff's bound

$$\mathbb{P}\left(|X - \mathbb{E}\left(X\right)| > t\right) \leqslant 2\exp\left\{-\frac{t^2}{2(\mathbb{E}\left(X\right) + t/3)}\right\} \tag{3}$$

for a binomial distributed random variable X. For $\ell > 1$ we split the edges of F into $i_0 \leqslant D$ matchings $M_1, \dots, M_{i_0}$, each of size at most n/ℓ. For an edge $e \in E(F)$ let t_e denote the random variable indicating that $e \in E(F[U])$. Then $\mathbb{P}\left(t_e = 1\right) = p^\ell$ and we set $s = 4\max\{p^\ell n/\ell, \log n\}$. By Chernoff's bound we have

$$\mathbb{P}\left(\exists i \leqslant i_0 \colon \sum_{e \in M_i} t_e > s\right) \leqslant 2Dn^{-4}.$$

This finishes the proof since $e(F[U]) = \sum_{i \in [i_0]} \sum_{e \in M_i} t_e$ which exceeds Ds with probability at most $2Dn^{-4}$. □

Proof of Lemma 6 and Lemma 10. Based on the preparation from above we give the proof of Lemma 6 and Lemma 10 in this section.

Proof of Lemma 6. Let t_i be the indicator random variable for the event $i \in \mathbb{Z}_{n,p}$. Since $v(H) = \sum_{i \in \mathbb{Z}/n\mathbb{Z}} t_i$ is binomially distributed the first property directly follows from Chernoff's bound (3).

For the second and third properties we first consider all elements from $\mathbb{Z}/n\mathbb{Z}$ which are bad with respect to $\mathbb{Z}_{n,p}$ and B. By Fact 13 we know that a.a.s. the set Y_n of bad elements has size at most $n^{1-1/(k-1)} \log n$ and in the following we will condition on this event.

We consider the degree and co-degree in the hypergraph $H(\mathbb{Z}_{n,p}, B, (\mathbb{Z}/n\mathbb{Z}) \setminus Y_n)$. Let $z, z' \in \mathbb{Z}/n\mathbb{Z}$ be given. If z is contained in an edge $M_x = M(\mathbb{Z}_{n,p}, B + x)$ then there is an element $P \in \mathcal{P}(z, B + x)$ such that $P \subset \mathbb{Z}_{n,p} \cup B + x$. It is sufficient to focus on those $P \in \mathcal{P}_0(z, B + x)$ since after removing Y_n those $P \in \mathcal{P}_1(z, B + x)$ will have no contribution. Hence we can assume $P \subset \mathbb{Z}_{n,p}$ or equivalently $\prod_{i \in P} t_i = 1$. Further, there are at most three different values of y such that P is contained in $\mathcal{P}_0(z, B + y)$. Hence, letting $\mathcal{P}_0(z, B, \mathbb{Z}/n\mathbb{Z}) = \bigcup_{x \in \mathbb{Z}/n\mathbb{Z}} \mathcal{P}_0(z, B + x)$ we can bound the degree of z by

$$\deg(z) \leqslant 3 \sum_{P \in \mathcal{P}_0(z, B, \mathbb{Z}/n\mathbb{Z})} \prod_{i \in P} t_i. \tag{4}$$

Similarly, if z, z' are contained in an edge M_x then there is a pair $(P, P') \in \mathcal{P}_0(z, z', B + x)$ such that $P \cup P' \subset \mathbb{Z}_{n,p} \cup B + x$, i.e. $\prod_{i \in P \cup P'} t_i = 1$. We obtain

$$\operatorname{codeg}(z_1, z_2) \leqslant 9 \sum_{(P, P') \in \mathcal{P}_0(z, z', B, \mathbb{Z}/n\mathbb{Z})} \prod_{i \in P \cup P'} t_i \tag{5}$$

We consider $\mathcal{P}_0(z, B, \mathbb{Z}/n\mathbb{Z})$ (resp., $\mathcal{P}_0(z, z', B, \mathbb{Z}/n\mathbb{Z})$) as an $(k-2)$-uniform (resp., $(2k-4)$-uniform) hypergraph on the vertex set $\mathbb{Z}/n\mathbb{Z}$. By Fact 12 we know that the maximum degree of the hypergraph is at most k^3K (resp. $5k^5K^2$). By Fact 14 the probability that $\deg(z) > 10k^3Kp^{k-2}n$ or $\operatorname{codeg}(z, z') > 8\log n$ is at most $2k^3Kn^{-4}$. Taking union bound over all elements and all pairs of $\mathbb{Z}/n\mathbb{Z}$ we obtain the desired property. □

Proof of Lemma 10. For given k, c, K and α set $L = 20c^{k-1}k^2/\alpha$ and $\alpha' = \alpha/2(LK)^L$. First we choose Y_n as to guarantee regularity of the triple $(\mathbb{Z}_{n,p}, B, (\mathbb{Z}/n\mathbb{Z})\backslash Y_n)$. Note that there are two sources of irregularity: k-APs containing two elements from $B + x$ for some $x \in \mathbb{Z}/n\mathbb{Z}$ and pairs of k-APs with one common element in $\mathbb{Z}_{n,p}$ and each containing one element in $B + x$ for some $x \in \mathbb{Z}/n\mathbb{Z}$. These are ruled out by removing all x which are bad with respect to $\mathbb{Z}_{n,p}$ and B. By Fact 13 a.a.s. the set Y_n of bad elements has size at most $n^{1-1/(k-1)}\log n$.

Further, we count the number of $(k-1)$-element sets in $\mathbb{Z}_{n,p}$ which arise from k-APs with one element removed. The expected number of such sets is at most $p^{k-1}kn^2$ and the variance is of order at most $p^{k-2}n^2$. Hence, by Chebyshev's inequality the number of such sets in $\mathbb{Z}_{n,p}$ is at most $2c^{k-1}kn$ asymptotically almost surely.

Consider any set $A \subset \mathbb{Z}/n\mathbb{Z}$ which posseses the two properties mentioned above: there is a set Y_n of size at most $n^{1-1/(k-1)}\log n$ such that $(A, B, (\mathbb{Z}/n\mathbb{Z})\backslash Y_n)$ is regular and the number of $(k-1)$-element sets in A which arise from k-APs with one element removed is at most $2c^{k-1}kn$. Let $X \subset \mathbb{Z}/n\mathbb{Z}$ of size $|X| \geqslant \alpha n$ be given. For every $x \in X\backslash Y_n$ let ℓ_x denote the size of $M_x = M(A, B, X)$. Then there are at least $\ell_x/(k-1)$ sets of size $(k-1)$ contained in M_x each forming a k-AP with an element in $B + x$. Further, each such set is contained in $B + x'$ for at most three $x' \in \mathbb{Z}/n\mathbb{Z}$, hence, $\sum_{x\in X\backslash Y_n} \ell_x/(k-1) \leqslant 6c^{k-1}kn$ and we conclude that the number of $x \in X\backslash Y_n$ such that $\ell_x > 20c^{k-1}k^2/\alpha = L$ is at most $\alpha n/3$. Moreover, there are at most $|B|^\ell$ distinct profiles of length ℓ, hence there is a profile π of length $\ell \leqslant L$ and a set $U \subset X\backslash Y_n$ of size $|U| \geqslant \alpha n/2K^L$ such that (A, B, U) has profile π.

To obtain a set $X' \subset U$ such that (A, B, X') is index consistent we consider a random partition of A into ℓ classes $(V_1, \dots, V_\ell)$. We say that an edge $M_x \in H(A, B, U)$ with the elements $z_1 < \dots < z_\ell$ survives if $z_i \in V_i$ for all $i \in [\ell]$. The probability of survival is ℓ^ℓ, hence there is a partition such that at least $|U|/\ell^\ell$ edges survives. Choosing the corresponding set $X' \subset U$ yields a set with the desired properties. □

5. Proof of Lemma 8

In this section we prove Lemma 8. The proof relies crucially on a structural theorem of Balogh, Morris and Samotij [1] which we state in the following. Let H be a uniform hypergraph with vertex set V and let $\mathcal{F}$ be an increasing family of subsets of V and $\varepsilon \in (0,1]$. The hypergraph H is called $(\mathcal{F},\varepsilon)$-dense if for every $A \in \mathcal{F}$

$$e(H[A]) \geqslant \varepsilon e(H).$$

Further, let $\mathcal{I}(H)$ denote the set of all independent sets of H. The theorem by Balogh, Morris and Samotij [1] reads as follows.

Theorem 15. [1, Theorem 2.2] *For every $\ell \in \mathbb{N}$ and all positive c and ε, there exists a positive constant c' such that the following holds. Let H be an ℓ-uniform hypergraph and let $\mathcal{F}$ be an increasing family of subsets of V such that $|A| \geqslant \varepsilon v(H)$ for all $A \in \mathcal{F}$. Suppose that $\mathcal{H}$ is $(\mathcal{F},\varepsilon)$-dense and $p \in (0,1)$ is such that for every $t \in [\ell]$,*

$$\Delta_t(H) \leqslant cp^{t-1}\frac{e(H)}{v(H)}\,.$$

Then there is a family $\mathcal{S} \subseteq \binom{V(H)}{\leqslant c'pv(H)}$ and functions $f\colon \mathcal{S} \to \overline{\mathcal{F}}$ and $g\colon \mathcal{I}(H) \to \mathcal{S}$ such that for every $I \in \mathcal{I}(H)$,

$$g(I) \subseteq I \quad \text{and} \quad I\backslash g(I) \subseteq f(g(I)).$$

Theorem 15 roughly says that if an uniform hypergraph H satisfies certain conditions then the set of independent sets $\mathcal{I}(H)$ of H can be "captured" by a family $\mathcal{S}$ consisting of small sets. Indeed, every independent set $I \in \mathcal{I}(H)$ contains a (small) set $g(I) \in \mathcal{S}$ and the remaining elements of I must come from a set determined by $g(I)$.

We now in the position to derive Lemma 8 from Theorem 15.

Proof of Lemma 8. Given the constants k, ℓ, c_0 and c_1 we apply Theorem 15 with $\varepsilon = 1/2$, and $c = \max\{1, c_1/c_0\}$ to obtain c'. Let $c_2 = c'$ and, further, let $\beta = \min\{1/4, c_0/(4c_1)\} > 0$.

We define the increasing family $\mathcal{F}$ by

$$\mathcal{F} = \{A\colon A \subset V(H),\ \text{s.t. } |A| \geqslant m/2 \text{ and } e(H[A]) \geqslant e(H)/2\}.$$

Clearly, H is $(\mathcal{F}, 1/2)$-dense and, moreover, any set $A \in \overline{\mathcal{F}}$ has size less than $m/2$ or $e(H[A]) < e(H)/2$. Thus, at least $e(H)/2$ edges of H are incident to the vertices outside of A. Therefore, $\Delta_1(H)(m - |A|) \geqslant e(H)/2 \geqslant c_0 m^{1+1/(k-2)}/2$ and with $\Delta_1(H) \leqslant c_1 m^{1/(k-2)}$ we conclude:

$$|A| \leqslant (1 - 2\beta)m.$$

We define $p = m^{-1/(k-2)(\ell-1)}$ so that $\Delta_t(H) \leqslant cp^{t-1}\frac{e(H)}{v(H)}$ for all $t \in [\ell]$. In fact, for $t = 1$ this follows directly from the bound $\Delta_1(H) \leqslant c_1 m^{\frac{1}{k-2}}$ given by the assumption of Lemma 8. For $t = 2, \dots, \ell - 1$ we use $\Delta_t(H) \leqslant \Delta_2(H)$ and on $\Delta_2(H)$ given by the assumption of Lemma 8. Finally, for $t = \ell$ we note that $\Delta_\ell(H) = 1$ and the desired bound follows again from the choices of p and c.

Thus there exist a family $\mathcal{S}$ and functions $f : \mathcal{S} \to \overline{\mathcal{F}}$ and $g : \mathcal{I}(H) \to \mathcal{S}$ with the properties described in Theorem 15. We define

$$\mathcal{A} = \{S \cup f(S) : S \in \mathrm{Im}(g)\},$$

where $\mathrm{Im}(g)$ is the image of g. Our cores will be the complements of the elements of $\mathcal{A}$,

$$\mathcal{C} = \{V(H) \backslash A : A \in \mathcal{A}\}.$$

Since $|\mathcal{C}| = |\mathcal{A}| \leqslant |\mathcal{S}|$, we infer (i) of Lemma 8. Further, every $A \in \mathcal{A}$ has size at most $(1 - 2\beta)m + c'pm \leqslant (1 - \beta)m$ which yields the property (ii) of Lemma 8. Finally, by the properties of the functions f and g, every independent set I is contained in $A = g(I) \cup f(g(I))$, so, by taking complements, every hitting set contains an element of $\mathcal{C}$ which completes the property (iii) of Lemma 8. □

6. Proof of Lemma 11

In this section we prove Lemma 11 which relies on the following result by the last author [12] (see also [1, 2, 11]).

Theorem 16. *For every integer $k \geqslant 3$ and every $\gamma \in (0, 1)$ there exists C and $\xi > 0$ such that for every sequence $p = p_n \geqslant Cn^{-1/(k-1)}$ the following holds a.a.s. Every subset of $\mathbb{Z}_{n,p}$ of size at least γpn contains at least $\xi p^k n^2$ arithmetic progression of length k.*

With this result at hand we now prove Lemma 11.

Proof of Lemma 11. The upper bound on the size of $\mathbb{Z}_{n,p}$ follows from Chernoff's bound (3). For the second property let $k \geqslant 3$ and γ be given. For $k \geqslant 4$ we apply Theorem 16 with $k - 1$ and γ to obtain C and ξ. We may assume that $\xi \leqslant \gamma^2/4$ and we note that in the case $k \geqslant 4$ we have $p \geqslant C_0 n^{-1/(k-1)} > Cn^{-1/(k-2)}$ for sufficiently large n. We choose $\delta = \xi^2/20$.

Let $S \subset \mathbb{Z}_{n,p}$ be a set of size at least γpn. For a given $i \in \mathbb{Z}/n\mathbb{Z}$ let $\deg(i)$ denote the number of $(k - 1)$-APs in S which form a k-AP with i. Note that $i \in F(S)$ if $\deg(i) \neq 0$.

Then a.s.s we have

$$\sum_{i\in\mathbb{Z}/n\mathbb{Z}} \deg(i) \geqslant \xi p^{k-1}n^2$$

which holds trivially for the case $k = 3$ due to $\xi \leqslant \gamma^2/4$ and which is a consequence of Theorem 16 for larger k.

Further, let $W = \sum_{i\in\mathbb{Z}/n\mathbb{Z}} \binom{\deg(i)}{2}$. Then, as $S \subset \mathbb{Z}_{n,p}$

$$\mathbb{E}\,(W) \leqslant n(p^{2(k-1)}n^2 + k^2p^{k-1-\lfloor\frac{k-1}{2}\rfloor}n) \leqslant 2p^{2(k-1)}n^3$$

and its variance is of order at most $p^{2(k-1)}n^3$. Hence, by Chebychev's inequality we have $\sum_{i\in\mathbb{Z}/n\mathbb{Z}} \binom{\deg(i)}{2} \leqslant 4p^{2(k-1)}n^3$ asymptotically almost surely.

Altogether we obtain

$$\binom{\xi p^{k-1}n^2}{2} \leqslant \binom{\sum_{i\in F(S)}\deg(i)}{2} \leqslant |F(S)| \sum_{i\in F(S)} \binom{\deg(i)}{2} \leqslant |F(S)|4p^{2(k-1)}n^3$$

and we conclude that $|F(S)| \geqslant \xi^2 n/20 = \delta n$. □

References

[1] J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc. **28** (2015), no. 3, 669–709, DOI 10.1090/S0894-0347-2014-00816-X. MR3327533 ↑1, 1, 3.1, 3.2, 5, 15, 6

[2] D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*. submitted. ↑1, 6

[3] D. Conlon, W. T. Gowers, W. Samotij, and M. Schacht, *On the KŁR conjecture in random graphs*, Israel J. Math. **203** (2014), no. 1, 535–580, DOI 10.1007/s11856-014-1120-1. MR3273450 ↑1

[4] E. Friedgut, *Sharp thresholds of graph properties, and the k-sat problem*, J. Amer. Math. Soc. **12** (1999), no. 4, 1017–1054, DOI 10.1090/S0894-0347-99-00305-7. With an appendix by Jean Bourgain. MR1678031 (2000a:05183) ↑1, 2

[5] ———, *Hunting for sharp thresholds*, Random Structures Algorithms **26** (2005), no. 1-2, 37–51, DOI 10.1002/rsa.20042. MR2116574 (2005k:05216) ↑2

[6] E. Friedgut and M. Krivelevich, *Sharp thresholds for certain Ramsey properties of random graphs*, Random Structures Algorithms **17** (2000), no. 1, 1–19, DOI 10.1002/1098-2418(200008)17:1<1::AID-RSA1>3.0.CO;2-4. MR1768845 (2001i:05136) ↑1

[7] E. Friedgut, V. Rödl, A. Ruciński, and P. Tetali, *A sharp threshold for random graphs with a monochromatic triangle in every edge coloring*, Mem. Amer. Math. Soc. **179** (2006), no. 845, vi+66, DOI 10.1090/memo/0845. MR2183532 (2006h:05208) ↑1

[8] E. Friedgut, V. Rödl, and M. Schacht, *Ramsey properties of random discrete structures*, Random Structures Algorithms **37** (2010), no. 4, 407–436, DOI 10.1002/rsa.20352. MR2760356 (2012a:05274) ↑1

[9] V. Rödl and A. Ruciński, *Threshold functions for Ramsey properties*, J. Amer. Math. Soc. **8** (1995), no. 4, 917–942, DOI 10.2307/2152833. MR1276825 (96h:05141) ↑1

[10] V. Rödl and A. Ruciński, *Rado partition theorem for random subsets of integers*, Proc. London Math. Soc. (3) **74** (1997), no. 3, 481–502, DOI 10.1112/S0024611597000178. MR1434440 (98d:05147) ↑1, 1, 2

[11] D. Saxton and A. Thomason, *Hypergraph containers*, Invent. Math. **201** (2015), no. 3, 925–992, DOI 10.1007/s00222-014-0562-8. MR3385638 ↑1, 1, 6

[12] M. Schacht, *Extremal results for random discrete structures*. submitted. ↑1, 6

[13] M. Schacht and F. Schulenburg, *A sharp threshold for a Ramsey property of strictly balanced nearly bipartite graphs*, 34 pages. preprint. ↑1

[14] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **34** (1959), 358–360. MR0106865 (21 #5595) ↑3.3

Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel

E-mail address: ehud.friedgut@weizmann.ac.il

Departamento de Ingeniería Matemática, Universidad de Chile

E-mail address: han.hiep@googlemail.com

Goethe-Universität, Institut für Mathematik, Robert-Mayer-Str. 10, 60325 Frankfurt am Main, Germany

E-mail address: person@math.uni-frankfurt.de

Fachbereich Mathematik, Universität Hamburg, Bundesstrasse 55, D-20146 Hamburg, Germany

E-mail address: schacht@math.uni-hamburg.de