Hypergraph removal with polynomial bounds

Lior Gishboliner * Asaf Shapira †

Abstract

Given a fixed $k$-uniform hypergraph $F$, the $F$-removal lemma states that every hypergraph with few copies of $F$ can be made $F$-free by the removal of few edges. Unfortunately, for general $F$, the constants involved are given by incredibly fast growing Ackermann-type functions. It is thus natural to ask for which $F$ can one prove removal lemmas with polynomial bounds. One trivial case where such bounds can be obtained is when $F$ is $k$-partite. Alon proved that when $k = 2$ (i.e. when dealing with graphs), only bipartite graphs have a polynomial removal lemma. Kohayakawa, Nagle and Rödl conjectured in 2002 that Alon’s result can be extended to all $k > 2$, namely, that the only $k$-graphs $F$ for which the hypergraph removal lemma has polynomial bounds are the trivial cases when $F$ is $k$-partite. In this paper we prove this conjecture.

1 Introduction

The hypergraph removal lemma is one of the most important results of extremal combinatorics. It states that for every fixed integer $k$, $k$-uniform hypergraph ($k$-graph for short) $F$ and positive $\varepsilon$, there is $\delta = \delta(F, \varepsilon) > 0$ so that if $G$ is an $n$-vertex $k$-graph with at least $\varepsilon n^k$ edge-disjoint copies of $F$, then $G$ contains $\delta n^{v(F)}$ copies of $F$. This lemma was first conjectured by Erdős, Frankl and Rödl [5] as an alternative approach for proving Szemerédi’s theorem [15]. The quest to proving this lemma, which involved the development of the hypergraph extension of Szemerédi’s regularity lemma [16], took more than two decades, culminating in several proofs, first by Gowers [8] and Rödl–Skokan–Nagle–Schacht [11, 13] and later by Tao [17]. For the sake of brevity, we refer the reader to [12] for more background and references on the subject.

While the hypergraph removal lemma has far-reaching qualitative applications, its main drawback is that it supplies very weak quantitative bounds. Specifically, for a general $k$-graph $F$, the function $1/\delta(F, \varepsilon)$ grows like the $k^{th}$ Ackermann function. It is thus natural to ask for which $k$-graphs $F$ one can obtain more sensible bounds. Further motivation for studying such questions comes from the area of graph property testing [7], where graph and hypergraph removal lemmas are used to design fast randomized algorithms.

Suppose first that $k = 2$. In this case it is easy to see that if $F$ is bipartite then $\delta(F, \varepsilon)$ grows polynomially with $\varepsilon$. Indeed, if $G$ has $\varepsilon n^2$ edge-disjoint copies of $F$ then it must have at least $\varepsilon n^2$ edges, which implies by the well-known Kövári–Sós–Turán theorem [10], that $G$ has at least $\text{poly}(\varepsilon) n^{v(F)}$ copies of $F$. In the seminal paper of Ruzsa and Szemerédi [14] in which they proved

---

*Department of Mathematics, ETH, Zürich, Switzerland. Email: lior.gishboliner@math.ethz.ch. Research supported by SNSF grant 200021_196965.
†School of Mathematics, Tel Aviv University, Tel Aviv 69978, Israel. Email: asafico@tau.ac.il. Supported in part by ISF Grant 1028/16, ERC Consolidator Grant 863438 and NSF-BSF Grant 20196.

1The lemma’s assumption is sometimes stated as $G$ being $\varepsilon$-far from $F$-freeness, meaning that one should remove at least $\varepsilon n^k$ edges to turn $G$ into an $F$-free hypergraph. It is easy to see that up to constant factors, this notion is equivalent to having $\varepsilon n^k$ edge-disjoint copies of $F$. 

1
the first version of the graph removal lemma, they also proved that when $F$ is the triangle $K_3$, the removal lemma has a super-polynomial dependence on $\varepsilon$. A highly influential result of Alon [1] completed the picture by extending the result of [14] to all non-bipartite graphs $F$.

Moving now to general $k > 2$, it is natural to ask for which $k$-graphs the function $\delta(F, \varepsilon)$ depends polynomially on $\varepsilon$. Let us say that in this case the $F$-removal lemma is polynomial. It is easy to see that like in the case of graphs, the $F$-removal lemma is polynomial whenever $F$ is $k$-partite. This follows from Erdős’s [4] well-known hypergraph extension of the Kővári–Sós–Turán theorem. Motivated by Alon’s result [1] mentioned above, Kohayakawa, Nagle and Rödl [9] conjectured in 2002 that the $F$-removal lemma is polynomial if and only if $F$ is $k$-partite. They further proved that the $F$-removal lemma is not polynomial when $F$ is the complete $k$-graph on $k + 1$ vertices. Alon and the second author [2] proved that a more general condition guarantees that the $F$-removal lemma is not polynomial, but fell short of covering all non-$k$-partite $k$-graphs. In the present paper we complete the picture, by fully resolving the problem of Kohayakawa, Nagle and Rödl [9].

**Theorem 1.** For every $k$-graph $F$, the $F$-removal lemma is polynomial if and only if $F$ is $k$-partite.

As a related remark, we note that for $k \geq 3$, the analogous problem for the induced $F$-removal lemma (that is, a characterization of $k$-graphs for which the induced $F$-removal lemma has polynomial bounds) was recently settled in [6], following a nearly-complete characterization given in [2].

Before proceeding, let us recall the notion of a core, which plays an important role in the proof of Theorem 1. Recall that for a pair of $k$-graphs $F_1, F_2$, a homomorphism from $F_1$ to $F_2$ is a map $\varphi : V(F_1) \to V(F_2)$ such that for every $e \in E(F_1)$ it holds that $\{\varphi(x) : x \in e\} \in E(F_2)$. The core of a $k$-graph $F$ is the smallest (with respect to the number of edges) subgraph of $F$ to which there is a homomorphism from $F$. It is not hard to show that the core of $F$ is unique up to isomorphism. Also, note that the core of a $k$-graph $F$ is a single edge if and only if $F$ is $k$-partite. In particular, if a $k$-graph is not $k$-partite, then neither is its core. We say that $F$ is a core if it is the core of itself.

Alon’s [1] approach relies on the fact that the core of every non-bipartite graph has a cycle. It is then natural to try and prove Theorem 1 by finding analogous sub-structures in the core of every non-$k$-partite $k$-graphs. Indeed, this was the approach taken in [2, 9]. The main novelty in this paper, and what allows us to handle all cases of Theorem 1, is that instead of directly inspecting the $k$-graph $F$, we study the properties of a certain graph associated with $F$. More precisely, given a $k$-graph $F = (V,E)$, we consider its 2-shadow, which is the graph on the same vertex set $V$ in which $\{u,v\}$ is an edge if and only if $u,v$ belong to some $e \in E$. The proof of Theorem 1 relies on the two lemmas described below.

**Lemma 1.1.** Suppose a $k$-graph $F$ is a core and its 2-shadow contains a cycle $C$ such that $|V(C) \cap e| \leq 2$ for every $e \in E(F)$. Then the $F$-removal lemma is not polynomial. In particular, if the 2-shadow of $F$ contains an induced cycle of length at least 4, then the $F$-removal lemma is not polynomial.

Note that this is a generalization of Alon’s result mentioned above since the 2-shadow of every non-bipartite graph $F$ (which is of course $F$ itself in this case) must contain a cycle. Our second lemma is the following.

**Lemma 1.2.** Suppose a $k$-graph $F$ is a core and its 2-shadow contains a clique of size $k + 1$. Then the $F$-removal lemma is not polynomial.

Note that this is a generalization of the result of Kohayakawa, Nagle and Rödl [9] mentioned above since the 2-shadow of the complete $k$-graph on $k + 1$ vertices is a clique of size $k + 1$.

The proofs of Lemmas 1.1 and 1.2 appear in Section 2, but let us first see why they together allow us to handle all non-$k$-partite $k$-graphs, thus proving Theorem 1.
Proof of Theorem 1. The if part was discussed above. As to the only if part, suppose \( F \) is a \( k \)-graph which is not \( k \)-partite and assume first that \( F \) is a core. Let \( G \) denote the 2-shadow of \( F \). If \( G \) contains an induced cycle of length at least 4, then the result follows from Lemma 1.1. Suppose then that \( G \) contains no such cycle, implying that \( G \) is chordal. Since \( F \) is not \( k \)-partite, \( G \) is not \( k \)-colorable. Since \( G \) is assumed to be chordal, and chordal graphs are well-known to be perfect, this means that \( G \) has a clique of size \( k + 1 \). Hence, the result follows from Lemma 1.1.

To prove the result when \( F \) is not necessarily a core, one just needs to observe that if \( F' \) is the core of \( F \), then (i) as noted earlier, \( F' \) is not \( k \)-partite, and (ii) since the \( F' \) removal lemma is not polynomial (by the previous paragraph), then neither is the \( F \) removal-lemma (see Claim 2.1 for the short proof of this fact).

\[\square\]

2 Proofs of Lemmas 1.1 and 1.2

We start by introducing some recurring notions. Recall that the \( b \)-blowup of a \( k \)-graph \( H = (V, E) \) is the \( k \)-graph obtained by replacing every vertex \( v \in V \) with a \( b \)-tuple of vertices \( S_v \), and then replacing every edge \( e = \{v_1, \ldots, v_k\} \in E \) with all possible \( b^k \) edges \( S_{v_1} \times S_{v_2} \times \cdots \times S_{v_k} \). Note that if \( H' \) is the \( b \)-blowup of \( H \), then the map sending \( S_v \) to \( v \) is a homomorphism from \( H' \) to \( H \). We will frequently refer to this as the natural homomorphism from \( H' \) to \( H \). We say that a \( k \)-graph \( H \) is homomorphic to a \( k \)-graph \( F \) if there is a homomorphism from the former to the latter. We first prove the following assertion, which was used in the proof of Theorem 1.

Claim 2.1. Let \( F \) be a \( k \)-graph and let \( C \) be a subgraph of \( F \) so that \( F \) is homomorphic to \( C \). Then, if the \( C \)-removal lemma is not polynomial, then neither is the \( F \)-removal lemma.

Proof. Since the \( C \)-removal lemma is not polynomial, there is a function \( \delta : (0, 1) \to (0, 1) \) such that \( 1/\delta(\varepsilon) \) grows faster than any polynomial in \( 1/\varepsilon \), and such that for every \( \varepsilon > 0 \) and large enough \( n \) there is an \( n \)-vertex \( k \)-graph \( H_1 \) which contains a collection \( \mathcal{C} \) of \( \varepsilon n^k \) edge-disjoint copies of \( C \) but only \( \delta n^v(C) \) copies of \( C \) altogether. Let \( H \) be the \( \nu(F) \)-blowup of \( H_1 \). Note that the \( \nu(F) \)-blowup of \( C \) contains a copy of \( F \). Also, copies of \( F \) corresponding to different copies of \( C \) from \( \mathcal{C} \) are edge-disjoint. Hence, \( H \) has a collection of \( \varepsilon n^k = \varepsilon (\nu(H)/\nu(F))^k = \Omega(\varepsilon \cdot \nu(H)^k) = \varepsilon' \nu(H)^k \) edge-disjoint copies of \( F \), for a suitable \( \varepsilon' = \Omega(\varepsilon) \). Let us bound the total number of copies of \( F \) in \( H \). Since \( C \) is a subgraph of \( F \), each copy of \( F \) must contain a copy of \( C \). Let \( \varphi : V(H) \to V(H_1) \) be the natural homomorphism from \( H \) to \( H_1 \) (as defined above). For each copy \( C' \) of \( C \) in \( H \), consider the subgraph \( \varphi(C') \) of \( H_1 \). The number of copies \( C' \) of \( C \) with \( \nu(\varphi(C')) < \nu(C) \) is at most \( \nu(F)^{\nu(C)} \cdot O(\nu(C)^{-1}) \leq \delta n^v(C) \), provided that \( n \) is large enough. The number of copies \( C' \) of \( C \) with \( \nu(\varphi(C')) \geq \nu(C) \) is at most \( \nu(F)^{\nu(C)} \cdot \delta n^v(C) = O(\delta n^v(C)) \), because \( H_1 \) contains at most \( \delta n^v(C) \) copies of \( C \). So in total, \( H \) contains at most \( O(\delta n^v(C)) \) copies of \( C \). This means that \( H \) contains at most \( O(\delta n^v(C)) \cdot \nu(F)^{\nu(C)} = O(\delta \cdot \nu(C)^{\nu(F)}) = \delta' \nu(C)^{\nu(F)} \) copies of \( F \), for a suitable \( \delta' = O(\delta) \). Note that \( 1/\delta' \) is super-polynomial in \( 1/\varepsilon' \). This shows that the \( F \)-removal lemma is not polynomial. \[\square\]

Since the core of \( F \) satisfies the properties of \( C \) in the above claim, it indeed establishes the assertion which we used when proving Theorem 1, namely that it suffices to prove the theorem when \( F \) is a core.

It thus remains to prove Lemmas 1.1 and 1.2. We begin preparing these proofs with some auxiliary lemmas. Throughout the rest of this section we will assume that \( F \) in Theorem 1 has no isolated vertices since removing isolated vertices does not make the removal lemma easier/harder. The following is a key property of cores that we will use in this section.
Claim 2.2. Let $F$ be a core $k$-graph, let $H$ be a $k$-graph, and let $\varphi : H \to F$ be a homomorphism. Then for every copy $F'$ of $F$ in $H$, the map $\varphi |_{V(F')}$ is an isomorphism\(^2\) from $F'$ to $F$.

Proof. We first claim that every homomorphism $\varphi$ from a core $F$ to itself is an isomorphism. Indeed, first note that since we assume that $F$ has no isolated vertices, then if $\varphi$ is not injective then $\varphi$’s image has less than $E(F)$ edges induced on it, which contradicts the minimality of $F$. Now, since $\varphi$ is an injection, and since it maps edges to edges, it must map non-edges to non-edges, and is therefore an isomorphism. The assertion of the claim now follows from the fact that $\varphi |_{V(F')}$ is a homomorphism from $F'$ to $F$. ■

We now describe our approach for proving Lemma 1.1 (the approach for Lemma 1.2 is analogous). Let $I \subseteq V(F)$ be a set of vertices so that the 2-shadow of $F$ induces on $I$ a graph containing a cycle, and so that $|e \cap I| \leq 2$ for every $e \in E(F)$. Let $S$ be the graph induced on $I$ by the 2-shadow of $F$. We first use the approach of [1] in order to construct a graph consisting of many edge-disjoint copies of $S$ yet containing few copies of $S$ altogether. The second step is then to extend the graph thus constructed into a $k$-graph containing many edge-disjoint copies of $F$ yet few copies of $F$. The following lemma will help us in performing this extension. For $\ell \geq 1$, two sets are called $\ell$-disjoint if their intersection has size at most $\ell - 1$. Two subgraphs of a hypergraph are called $\ell$-disjoint if their vertex-sets are $\ell$-disjoint.

Lemma 2.3. Let $r, s, k, \ell \geq 0$ satisfy $k \geq \ell$ and $r \geq k - \ell$. Let $V_1, \ldots, V_s, V_{s+1}, \ldots, V_{s+r}$ be pairwise-disjoint sets of size $n$ each. Let $S \subseteq V_1 \times \cdots \times V_s$ be a family of $\ell$-disjoint sets. Then there is a family $\mathcal{F} \subseteq V_1 \times \cdots \times V_{s+r}$ with the following properties:

1. For every $F \in \mathcal{F}$ it holds that $F|_{V_1 \times \cdots \times V_s} \in S$.
2. $|\mathcal{F}| \geq \Omega_{r,s,k}(\beta n^{k-\ell})$.
3. For every pair of distinct $F_1, F_2 \in \mathcal{F}$, if $|F_1 \cap F_2| \geq k$ then
   \[
   \#\{s + 1 \leq i \leq s + r : F_1(i) = F_2(i)\} \leq k - \ell - 1
   \]

Proof. We construct the family $\mathcal{F}$ as follows. For each $S \in S$ and each $r$-tuple $A \in V_{s+1} \times \cdots \times V_{s+r}$, add $S \cup A$ to $\mathcal{F}$ with probability $\frac{1}{Cn^{-k+\ell}}$, where $C$ is a large constant to be chosen later. Item 1 is satisfied by definition. Let us estimate the number of pairs $F_1, F_2 \in \mathcal{F}$ violating Item 3; denote this number by $B$. Suppose that $F_1 = S_1 \cup A_1$ and $F_2 = S_2 \cup A_2$ violate Item 3. Then $d := |A_1 \cap A_2| \geq k - \ell$ and $|S_1 \cap S_2| \geq k - d$. The number of choices of $A_1, A_2 \in V_{s+1} \times \cdots \times V_{s+r}$ with $|A_1 \cap A_2| = d$ is at most $n^r \cdot \binom{s}{d} \cdot n^{r-d}$. Also, for $0 \leq t \leq \ell$, the number of choices of $S_1, S_2 \in S$ with $|S_1 \cap S_2| \geq t$ is at most $|S| \cdot \binom{s}{t} \cdot n^{\ell-t}$, because the sets in $S$ are pairwise $\ell$-disjoint. Note that $k - d \leq \ell$. We can also allow $t$ to be negative by replacing $t$ with $\max\{0, t\}$ in the above formula. Finally, the probability that $S_1 \cup A_1, S_2 \cup A_2 \in \mathcal{F}$ is $\left(\frac{1}{Cn^{-k+\ell}}\right)^2$. Hence, the number $B$ of violations to Item 3 is, in expectation, at most
\[
E[B] \leq \sum_{d=k-\ell}^{r} n^r \cdot \binom{s}{d} \cdot n^{r-d} \cdot |S| \cdot \left(\max\{0, k - d\}\right) \cdot n^{\ell-\max\{0, k - d\}} \cdot \left(\frac{1}{Cn^{-k+\ell}}\right)^2
\]
\[
= O_{s,r,k}\left(\frac{1}{C^2}\right) \cdot |S| \cdot n^{k-\ell}.
\]
\(^2\)Just to clarify, we do not claim that $\varphi |_{V(F')}$ is an isomorphism between $F$ and the graph induced by $H$ on $V(F')$. Rather, $\varphi |_{V(F')}$ is an isomorphism between $F$ and the graph $(V(F'), E(F'))$.  

4
On the other hand, the expected size of $F$ is $|S| \cdot n^r \cdot \frac{1}{Cn^{1-\epsilon}} = \frac{1}{C} \cdot |S| \cdot n^{k-\epsilon}$. So by choosing $C$ to be large enough (as a function of $s, r, k$), we can guarantee that $\mathbb{E}(|F| - B) \geq \frac{1}{2C} \cdot |S| \cdot n^{k-\epsilon}$. By fixing such a choice of $F$ and deleting one set $F \in F$ from each violation, we get the required conclusion. 

The following well-known fact is an easy corollary of Lemma 2.3.

**Lemma 2.4.** Let $1 \leq k \leq r$, and let $V_1, \ldots, V_r$ be pairwise-disjoint sets of size $n$ each. Then there is $F \subseteq V_1 \times \cdots \times V_r$, $|F| \geq \Omega(n^k)$, such that the sets in $F$ are $k$-disjoint.

**Proof.** Apply Lemma 2.3 with $s = \ell = 0$ and $S = \emptyset$. ■

The next lemma shows why constructing a $k$-graph with a sublinear number of edge disjoint copies of $F$ can be boosted to prove Lemmas 1.1 and 1.2. The lemma makes crucial use of the fact that $F$ is a core.

**Lemma 2.5.** Let $F$ be a core $k$-graph, and suppose that for a constant $C$ and for every large enough $n$, there is a $k$-graph $H$ which is homomorphic to $F$, has a collection of $n^k / e^{C \sqrt{\log n}}$ edge-disjoint copies of $F$, but has at most $n^{v(F)-1}$ copies of $F$ altogether. Then the $F$-removal lemma is not polynomial.

**Proof.** Let $\varepsilon > 0$ and let $n$ be large enough. Let $m$ be the largest integer satisfying $e^{C \sqrt{\log m}} \leq 1/\varepsilon$. It is easy to check that $m \geq (1/\varepsilon)^{O(\log(1/\varepsilon))}$. Let $H$ be the $k$-graph guaranteed to exist by the assumption of the lemma, but with $m$ in place of $n$. So $H$ has $m$ vertices, contains a collection $F$ of $m^k / e^{C \sqrt{\log m}} \geq \varepsilon n^k$ edge-disjoint copies of $F$, but has at most $m^{v(F)-1}$ copies of $F$ altogether.

Let $G$ be the $\frac{n}{m}$-blowup of $H$. Each $F' \in F$ gives rise to $\Omega((\frac{n}{m})^k)$ $k$-disjoint (and hence also edge-disjoint) copies of $F$ in $G$, by Lemma 2.4 applied with $r = v(F)$ and with $\frac{n}{m}$ in place of $n$. Copies arising from different $F'_1, F'_2 \in F$ are edge-disjoint, because the copies in $F$ are edge-disjoint. Altogether, this gives a collection of $\varepsilon n^k \cdot \Omega((\frac{n}{m})^k) = \Omega(\varepsilon n^k)$ edge-disjoint copies of $F$ in $G$.

Let us upper-bound the total number of copies of $F$ in $G$. By assumption, there is a homomorphism $\phi$ from $H$ to $F$. Let $\psi$ be the “natural” homomorphism from $G$ to $H$ (as described in the beginning of the section). Then $\phi \circ \psi$ is a homomorphism from $G$ to $F$. By Claim 2.2, for every copy $F'$ of $F$ in $G$ the map $\phi \circ \psi_{|F'(F)}$ is an isomorphism between $F'$ and $F$. We claim that this means that $\psi$ maps every copy $F'$ of $F$ in $G$ onto a copy of $F$ in $H$. Indeed, $\psi_{|F'(F)}$ must be injective (otherwise $\phi \circ \psi_{|F'(F)}$ would not be an isomorphism), and since $\psi_{|F'(F)}$ must map edges to edges (on account of being a homomorphism) its image must contain a copy of $\phi$. We thus see that every copy of $F$ in $G$ must come from the blown-up copies of $F$ in $H$. But each copy of $F$ in $H$ gives rise to $(\frac{n}{m})^{v(F)}$ copies of $F$ in $G$. Hence, the total number of copies of $F$ in $G$ is at most

$$m^{v(F)-1} \cdot \left((n/m)^{v(F)}\right) \leq \varepsilon \cdot \Omega(\log(1/\varepsilon)) \cdot n^{v(F)}.$$ 

This shows that the $F$-removal lemma is not polynomial. ■

Let $S$ be a $k$-graph on $[s]$ and let $G$ be an $s$-partite $k$-graph with sides $V_1, \ldots, V_s$. A canonical copy of $S$ in $G$ is a copy consisting of vertices $v_1 \in V_1, \ldots, v_s \in V_s$ in which $v_i$ plays the role of $i \in V(S)$ for each $i = 1, \ldots, s$. The following result appears implicitly in [1]. For the sake of completeness, we include a proof.

**Lemma 2.6.** Let $S$ be a graph on $[s]$ containing a cycle. Then for every large enough $n$, there is an $s$-partite graph $G$ with sides $V_1, \ldots, V_s$, each of size $n$, such that $G$ has a collection of $n^2 / e^{O(\sqrt{\log n})}$ $2$-disjoint canonical copies of $S$, but at most $n^{s-1}$ canonical copies of $S$ altogether.
Proof. Without loss of generality, suppose that \((1, 2, \ldots, t, 1)\) is a cycle in \(S\) (otherwise permute the coordinates) where \(t \geq 3\). Take a set \(B \subseteq [n/s]\), \(|B| \geq n/e^{O(\log n)}\), with no non-trivial solution to the linear equation \(y_1 + \cdots + y_{t-1} = (t-1)y_t\) with \(y_1, \ldots, y_t \in B\) (where a solution is trivial if \(y_1 = y_2 = \ldots = y_t\)). The existence of such a set \(B\) is by a simple generalization of Behrend’s construction [3] of sets avoiding 3-term arithmetic progressions, see [1, Lemma 3.1]. Take pairwise-disjoint sets \(V_1, \ldots, V_s\) of size \(n\) each, and identify each \(V_i\) with \([n]\). For each \(x \in [n/s]\) and \(y \in B\), add to \(G\) a canonical copy \(S_{x,y}\) of \(S\) on the vertices \(v_i = x + (i-1)y \in V_i\), \(i = 1, \ldots, s\). Note that \(x+(i-1)y \leq x+(s-1)y \leq n\), so \(v_i\) indeed “fits” into \(V_i = [n]\). The copies \(S_{x,y}\) (where \(x \in [n/s]\), \(y \in B\)) are 2-disjoint. Indeed, if \(S_{x_1,y_1}, S_{x_2,y_2}\) intersect in \(V_i\) and in \(V_j\), then \(x_1 + (i-1)y_1 = x_2 + (i-1)y_2\) and \(x_1 + (j-1)y_1 = x_2 + (j-1)y_2\), and solving this system of equations gives \(x_1 = x_2, y_1 = y_2\). The number of copies \(S_{x,y}\) is \(\frac{n}{s} \cdot |B| \geq n^2/e^{O(\sqrt{\log n})}\).

Let us bound the total number of canonical copies of \(S\) in \(G\). Fix a canonical copy with vertices \(v_1, \ldots, v_s, v_t \in V_t\). Then \(v_1, \ldots, v_t\) is a cycle in \(G\). For \(1 \leq j \leq t - 1\), let \(x_j \in [n/s]\), \(y_j \in B\) such that \(v_j, v_{j+1} \in S_{x_j,y_j}\). Similarly, let \(x_t \in [n/s]\), \(y_t \in B\) such that \(v_t, v_1 \in S_{x_t,y_t}\). Then we have \(v_{i+1} - v_i = y_j\) for every \(1 \leq j \leq t - 1\), and \(v_t - v_1 = (t-1)y_t\). So \(y_1 + \cdots + y_{t-1} = (t-1)y_t\). By our choice of \(B\), we have \(y_1 = \cdots = y_t = y\). Now, for each \(1 \leq j \leq t - 1\) we have \(x_j = v_{i+j} - j \cdot y = v_{i+j+1}\), so \(x_1 = \cdots = x_t = x\). So we see that for each canonical copy \(v_1, \ldots, v_s\) of \(S\), there are \(x \in [n/s]\), \(y \in B\) such that \(v_1, \ldots, v_t \in S_{x,y}\). The number of choices for \(x, y\) is \((n/s)|B| \leq n^2\). Hence, the number of canonical copies of \(S\) is at most \(n^2 \cdot n^{s-1} \leq n^{s-1}\).

Recall that \(K_s^{(s-1)}\) is the \((s-1)\)-graph with vertices \(1, \ldots, s\) and all \(s\) possible edges. The following construction appears implicitly in [9] (see also [2]). Again, for completeness, we include a proof.

Lemma 2.7. Let \(s \geq 3\). For every large enough \(n\), there is an \(s\)-partite \((s-1)\)-graph \(G\) with sides \(V_1, \ldots, V_s\), each of size \(n\), such that \(G\) has a collection of \(n^{s-1}/e^{O(\sqrt{\log n})}\) \((s-1)\)-disjoint canonical copies of \(K_s^{(s-1)}\), but at most \(n^{s-1}\) copies of \(K_s^{(s-1)}\) altogether.

Proof. Take a set \(B \subseteq [n/s]\), \(|B| \geq n/e^{O(\sqrt{\log n})}\), with no non-trivial solution to \(y_1 + y_2 = 2y_3\), \(y_1, y_2, y_3 \in B\). Take pairwise-disjoint sets \(V_1, \ldots, V_s\) of size \(n\) each, and identify each \(V_i\) with \([n]\). For each \(x_1, \ldots, x_{s-2} \in [n/s]\) and \(y \in B\), add to \(G\) a copy \(K_{x_1, \ldots, x_{s-2}, y}\) of \(K_s^{(s-1)}\) on the vertices

\[
x_1 \in V_1, \quad x_2 \in V_2, \quad \ldots \quad x_{s-2} \in V_{s-2}, \quad y + \sum_{i=1}^{s-2} x_i \in V_{s-1}, \quad 2y + \sum_{i=1}^{s-2} x_i \in V_s
\]

It is easy to see that these copies are \((s-1)\)-disjoint, because fixing any \(s-1\) of the \(s\) coordinates allows to solve for \(x_1, \ldots, x_{s-2}, y\). Also, the number of such copies plus thus places is \((n/s)^{s-2} \cdot |B| \geq n^{s-1}/e^{O(\sqrt{\log n})}\).

Let us show that the are no other copies of \(K_s^{(s-1)}\) in \(G\). This would imply that the total number of copies of \(K_s^{(s-1)}\) in \(G\) is \((n/s)^{s-2} \cdot |B| \leq n^{s-1}\). So suppose that \(v_1 \in V_1, \ldots, v_s \in V_s\) form a copy of \(K_s^{(s-1)}\). Let \(x^{(i)} = (x_1^{(i)}, \ldots, x_{s-2}^{(i)}) \in [n/s]^{s-2}\) and \(y^{(i)} \in B\), \(i = 1, 2, 3\), be such that \(\{v_2, \ldots, v_s\} \subseteq K_{x_1^{(i)}, y^{(i)}}\), \(\{v_1, \ldots, v_{s-2}, v_s\} \subseteq K_{x_2^{(i)}, y^{(i)}}\) and \(\{v_1, \ldots, v_{s-2}\} \subseteq K_{x_3^{(i)}, y^{(i)}}\). Then \(x_1^{(2)} = x_1^{(3)} = v_1\) and

\[
x_j^{(1)} = x_j^{(2)} = x_j^{(3)} = v_j \text{ for every } 2 \leq j \leq s - 2.
\]

Also, \(v_s - v_{s-1} = y_1, v_{s-1} - v_1 = x_2^{(2)} + \cdots + x_{s-2}^{(2)} + 2y_2\) and \(v_s - v_1 = x_2^{(3)} + \cdots + x_{s-2}^{(3)} + 2y_3\). Combining these three equations and using (1), we get \(y_1 + y_2 = 2y_3\), and so \(y_1 = y_2 = y_3 =: y\) by our choice of \(B\). Also, \(x_1^{(1)} = v_{s-1} - (v_2 + \cdots + v_{s-2} + y) = x_1^{(2)}\). So \(x^{(1)} = x^{(2)} = x^{(3)}\).
We now prove two lemmas, 2.8 and 2.9, which imply Lemmas 1.1 and 1.2, respectively. Recall that for a $k$-graph $F$ and $2 \leq \ell \leq k$, the $\ell$-shadow of $F$, denoted $\partial_{\ell}F$, is the $\ell$-graph consisting of all $f \in (V(F))^{\ell}$ such that there is $e \in E(F)$ with $f \subseteq e$.

**Lemma 2.8.** Let $k \geq 2$, let $F$ be a core $k$-graph and suppose that there is a set $I \subseteq V(F)$ such that $(\partial_{2}F)[I]$ contains a cycle and $|e \cap I| \leq 2$ for every $e \in E(F)$. Then for every large enough $n$ there is a $k$-graph $H$ which is homomorphic to $F$, has a collection of $n^{k}/e^{O(\sqrt{\log n})}$ edge-disjoint copies of $F$, but has at most $n^{(F)-1}$ copies of $F$ altogether.

**Proof.** It will be convenient to write $|I| = s$, $|V(F)| = s + r$, and to assume that $I = [s]$ and $V(F) = [s + r]$. Let $S := (\partial_{2}F)[I]$, that is, the graph induced by $F$'s 2-shadow on $I$. By assumption, $S$ contains a cycle. Take disjoint sets $V_{1}, \ldots, V_{r+s}$ of size $n$ each. Let $G$ be the $s$-partite graph with sides $V_{1}, \ldots, V_{s}$ given by Lemma 2.6. Let $S$ be a collection of $n^{2}/e^{O(\sqrt{\log n})}$ 2-disjoint canonical copies of $S$ in $G$. Apply Lemma 2.3 to $S$ with $\ell = 2$ to obtain a family $F \subseteq V_{1} \times \cdots \times V_{s+r}$ satisfying Items 1-3 in that lemma. Note that $r \geq k - 2 = k - \ell$ (because each edge of $F$ contains at most 2 vertices from $I = [s]$), so the conditions of Lemma 2.3 are satisfied. Define the hypergraph $H$ by placing a canonical copy of $F$ on each $F' \in F$. We claim that these copies of $F$ are edge-disjoint. Indeed, suppose by contradiction that the copies on $F_{1}, F_{2} \in F$ share an edge $e$. Then $|F_{1} \cap F_{2}| \geq k$. By Item 3 of Lemma 2.3, we have $\#\{s + 1 \leq i \leq s + r : F_{1}(i) = F_{2}(i)\} \leq k - 3$. This implies that $\#\{1 \leq i \leq s : e \cap V_{i} \neq \emptyset\} \geq 3$. But this means that in $F$ there is an edge which intersects $I = [s]$ in at least 3 vertices, in contradiction to the assumption of the lemma. So the copies in $F$ are indeed edge-disjoint. Their number is $|F| \geq \Omega(|S|n^{s-2}) \geq n^{k}/e^{O(\sqrt{\log n})}$, by Item 2 of Lemma 2.3.

To complete the proof, it remains to show that $H$ has at most $n^{s+r-1}$ copies of $F$. Observe that $H$ is homomorphic to $F$; indeed, the map $\varphi$ which sends $V_{j} \mapsto j$, $j = 1, \ldots, s + r$, is such a homomorphism. Let $F^{*}$ be a copy of $F$ in $H$. Since $F$ is a core and $\varphi$ is a homomorphism from $H$ to $F$, we can apply Claim 2.2 to conclude that $F^{*}$ must have the form $v_{1}, \ldots, v_{s+r}$, with $v_{i} \in V_{i}$ playing the role of $i$ for each $i = 1, \ldots, s + r$. We claim that $v_{1}, \ldots, v_{s}$ form a canonical copy of $S$ in $G$. To see this, fix any $(i, j) \in E(S)$ and let us show that $\{v_{i}, v_{j}\} \in E(G)$. Since $S = (\partial_{2}F)[I]$, there must be an edge $e \in E(F)$ containing $i, j$. Then $\{v_{a} : a \in e\} \in E(F^{*}) \subseteq E(H) = \bigcup_{F' \in F} E(F')$. Let $F' \in F$ such that $\{v_{a} : a \in e\} \in E(F')$. By Item 1 of Lemma 2.3, we have $S' := F'|_{V_{1} \times \cdots \times V_{s}} \subseteq S$. Now, $S'$ is the vertex set of a canonical copy of $S$ in $G$, and hence $\{v_{i}, v_{j}\} \in E(G)$, as required. This proves our claim that $v_{1}, \ldots, v_{s}$ form a canonical copy of $S$ in $G$. Summarizing, every copy of $F$ in $H$ contains the vertices of a canonical copy of $S$ in $G$. By the guarantees of Lemma 2.6, the number of canonical copies of $S$ in $G$ is at most $n^{s-1}$. Hence, the number of copies of $F$ in $H$ is at most $n^{s-1} \cdot n^{r} = n^{s+r-1}$, as required.

**Lemma 2.9.** Let $F$ be a core $k$-graph and suppose that there are $3 \leq s \leq k + 1$ and a set $I \subseteq V(F)$ such that $(\delta_{s-1}F)[I] \cong K_{s}^{(s-1)}$ and $|e \cap I| \leq s - 1$ for every $e \in E(F)$. Then for every large enough $n$ there is a $k$-graph $H$ which is homomorphic to $F$, has a collection of $n^{k}/e^{O(\sqrt{\log n})}$ edge-disjoint copies of $F$, but has at most $n^{(F)-1}$ copies of $F$ altogether.

**Proof.** The proof is very similar to that of Lemma 2.8. Assume that $I = [s]$, $V(F) = [s + r]$. Take disjoint sets $V_{1}, \ldots, V_{r+s}$ of size $n$ each. Let $G$ be the $s$-partite $(s - 1)$-graph with sides $V_{1}, \ldots, V_{s}$ given by Lemma 2.7. Let $S$ be a collection of $n^{s-1}/e^{O(\sqrt{\log n})}$ $(s - 1)$-disjoint copies of $K_{s}^{(s-1)}$ in $G$.

---

3Strictly speaking we apply Lemma 2.3 to the vertex sets of the copies of $S$.

4Note that by definition of $S$, the 2-shadow of $F^{*}$ creates a copy of $S$ in the 2-shadow of $H$. The first key point is that this copy of $S$ must appear in $G$. Also, note that this fact is trivial if $F^{*}$ is one of the canonical copies of $F$ we placed in $H$ when defining it. The second key point is that this holds for every copy $F^{*}$ of $F$ in $H$. 

7
Apply Lemma 2.3 to $S$ with $\ell = s - 1$ to obtain a family $F \subseteq V_1 \times \cdots \times V_{s+r}$ satisfying Items 1-3 in that lemma. Define the hypergraph $H$ by placing a canonical copy of $F$ on each $F' \in F$. These copies of $F$ are edge-disjoint. Indeed, suppose by contradiction that the copies on $F_1, F_2 \in F$ share an edge $e$. Then $|F_1 \cap F_2| \geq k$, and hence $\#\{s + 1 \leq i \leq s + r : F_1(i) = F_2(i)\} \leq k - \ell - 1 = k - s$ by Item 3 of Lemma 2.3. But then $\#\{1 \leq i \leq s : e \cap V_i \neq \emptyset\} = s$, meaning that there is an edge in $F$ which contains $I = [s]$, a contradiction to the assumption of the lemma. We have $|F| \geq \Omega(|S|n^{k-s+1}) \geq n^k/e^{O(\sqrt{\log n})}$, using Item 2 of Lemma 2.3.

The map $V_j \mapsto j$, $j = 1, \ldots, s + r$ is a homomorphism from $H$ to $F$. Let us bound the number of copies of $F$ in $H$. By Claim 2.2, every copy $F'$ of $F$ must be of the form $v_1, \ldots, v_{s+r}$, with $v_i \in V_i$ playing the role of $i$ for each $i = 1, \ldots, s + r$. We claim that $v_1, \ldots, v_s$ span a copy of $K_s^{(s-1)}$ in $G$. So let $J \in \binom{[s]}{s-1}$. Since $(\partial_{s-1}F)[I] \cong K_s^{(s-1)}$, there is an edge $e \in E(F)$ with $J \subseteq e$. Since $F'$ is a canonical copy of $F$, we have $\{v_i : i \in e\} \in E(F) \subseteq E(H) = \bigcup_{F' \in F} E(F')$. Let $F' \in F$ such that $\{v_i : i \in e\} \in E(F')$. By Item 1 of Lemma 2.3, we have $S' := F'|_{V_1 \times \cdots \times V_s} \in S$. Now, $S'$ is a canonical copy of $K_s^{(s-1)}$ in $G$, and hence $\{v_i : i \in J\} \in E(G)$, as required. So we see that every copy of $F$ in $H$ contains the vertices of a copy of $K_s^{(s-1)}$ in $G$. By the guarantees of Lemma 2.6, $G$ has at most $n^{s-1}$ copies of $K_s^{(s-1)}$. Hence, $H$ has at most $n^s \cdot n^r = n^{s+r-1}$ copies of $F$, as required.

Observe that Lemma 1.1 follows by combining Lemmas 2.5 and 2.8. Let us prove Lemma 1.2.

**Proof of Lemma 1.2.** Let $X$ be a clique of size $k + 1$ in $\partial_2 F$. Let $I$ be a smallest set in $X$ which is not contained in an edge of $F$. Note that $I$ is well-defined (because $X$ itself is not contained in any edge of $F$, as $|X| = k + 1$). Also, $|I| \geq 3$ because every pair of vertices in $X$ is contained in some edge, as $X$ is a clique in $\partial_2 F$. Put $s = |I|$. Then $(\partial_{s-1}F)[I] \cong K_s^{(s-1)}$ and $|e \cap I| \leq s - 1$ for every $e \in E(F)$, by the choice of $I$. Now the assertion of Lemma 1.2 follows by combining Lemmas 2.5 and 2.9.

**References**

[1] N. Alon, Testing subgraphs in large graphs, Random Structures Algorithms 21 (2002), 359–370.

[2] N. Alon and A. Shapira, Linear equations, arithmetic progressions and hypergraph property testing, Theory of Computing Vol 1 (2005), 177–216.

[3] F. A. Behrend, On sets of integers which contain no three terms in arithmetic progression, Proc. Natl. Acad. Sci. U.S.A. 32 (1946), 331–332.

[4] P. Erdős, On extremal problems of graphs and generalized graphs, Israel J. Math. 2 (1964), 183–190.

[5] P. Erdős, P. Frankl and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, Graphs Combin. 2 (1986), 113–121.

[6] L. Gishboliner and I. Tomon, On 3-graphs with no four vertices spanning exactly two edges, arXiv preprint arXiv:2109.04944, 2021.

[7] O. Goldreich, *Introduction to Property Testing*, Cambridge University Press, 2017.
[8] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, Ann. of Math. 166 (2007), 897–946.

[9] Y. Kohayakawa, B. Nagle and V. Rödl, Efficient testing of hypergraphs, Proc. of the International Colloquium on Automata, Languages, and Programming (ICALP) 2002, 1017–1028.

[10] T. Kovári, V. Sós and P. Turán, On a problem of K. Zarankiewicz. In Colloquium Mathematicum, 1954, 50–57.

[11] B. Nagle, V. Rödl and M. Schacht, The counting lemma for regular $k$-uniform hypergraphs, Random Structures Algorithms 28 (2006), 113–179.

[12] V. Rödl, Quasi-randomness and the regularity method in hypergraphs, Proceedings of the International Congress of Mathematicians (ICM) 1 (2015), 571–599.

[13] V. Rödl and J. Skokan, Regularity lemma for $k$-uniform hypergraphs, Random Structures Algorithms 25 (2004), 1–42.

[14] I. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai 18, Volume II, 939–945.

[15] E. Szemerédi, On sets of integers containing no $k$ elements in arithmetic progression, Acta Arith. 27 (1975), 199–245.

[16] E. Szemerédi, Regular partitions of graphs, In: Proc. Colloque Inter. CNRS, 1978, 399–401.

[17] T. Tao, A variant of the hypergraph removal lemma, J. Combin. Theory Ser. A 113 (2006), 1257–1280.