Property Testing in Hypergraphs and the Removal Lemma

[Extended Abstract] *

Vojtěch Rödl†
Department of Mathematics and Computer Science
Emory University
Atlanta, GA 30322, USA
rodl@mathcs.emory.edu

Mathias Schacht‡
Institut für Informatik
Humboldt-Universität zu Berlin
Unter den Linden 6
D-10099 Berlin, Germany
schacht@informatik.hu-berlin.de

ABSTRACT

Property testers are efficient, randomized algorithms which recognize if an input graph (or other combinatorial structure) satisfies a given property or if it is "far" from exhibiting it. Generalizing several earlier results, Alon and Shapira showed that *hereditary* graph properties are testable (with one-sided error). In this paper we prove the analogous result for hypergraphs. This result is an immediate consequence of a (hyper)graph theoretic statement, which is an extension of the so-called *removal lemma*. The proof of this generalization relies on the *regularity method for hypergraphs*.

Categories and Subject Descriptors

G.3 [**Mathematics of Computing**]: Probabilistic algorithms; G.2.2 [**Discrete Mathematics**]: Graph Theory—*Graph algorithms, Hypergraphs*

General Terms

Algorithms, Theory

Keywords

property testing, hypergraphs, removal lemma, regularity lemma, hereditary properties

1. INTRODUCTION AND RESULTS

1.1 Property testing

The general notion of *property testing* was introduced by Rubinfeld and Sudan [31]. Roughly speaking, the typical question in property testing concerns the existence of efficient algorithms, which distinguish between objects G having a given property $\mathscr{P}$ and those being "far away" (in an appropriate metric) from $\mathscr{P}$. Problems of that type appeared already earlier in the literature (see, e.g., [9]). However, the systematic study of property testing for discrete structures, such as graphs, digraphs, hypergraphs, discrete functions and sets of integers, was initiated in the seminal work of Golderich, Goldwasser, and Ron [19].

In this paper we exclusively focus on properties of graphs and k-uniform hypergraphs (k-graphs). A k-uniform hypergraph $\mathcal{H}^{(k)}$ on the vertex set V is some family of k-element subsets of V, i.e., $\mathcal{H}^{(k)} \subseteq \binom{V}{k}$. Note that we identify hypergraphs with their edge set and we write $V(\mathcal{H}^{(k)})$ for the vertex set. We only consider uniform hypergraphs, where the uniformity is some fixed number independent of the size of the hypergraph. We usually indicate the uniformity by a superscript and call a k-uniform hypergraph simply a hypergraph, when the uniformity is clear from the context.

A property $\mathscr{P}$ of hypergraphs is a family of hypergraphs, which is closed under isomorphisms. We only consider **decidable** properties. These are those properties $\mathscr{P}$ for which the membership problem is solvable, i.e., there exists an algorithm which for every hypergraph $\mathcal{G}^{(k)}$ decides whether $\mathcal{G}^{(k)} \in \mathscr{P}$ in finite time. We say a k-uniform hypergraph $\mathcal{G}^{(k)}$ is **η-far from $\mathscr{P}$** if every hypergraph $\mathcal{H}^{(k)}$ on the same vertex set V with $|\mathcal{G}^{(k)} \triangle \mathcal{H}^{(k)}| \leq \eta\binom{V}{k}$ does not satisfy $\mathscr{P}$. In other words, we have to change $\mathcal{G}^{(k)}$ on at least an η-fraction of all possible edges in order to obtain a hypergraph $\mathcal{H}^{(k)}$, which satisfies $\mathscr{P}$.

Roughly speaking, a property $\mathscr{P}$ is testable, if for every $\eta > 0$ there exists a constant time, randomized algorithm which distinguishes for an input hypergraph $\mathcal{H}^{(k)}$ between the cases $\mathcal{H}^{(k)} \in \mathscr{P}$ and $\mathcal{H}^{(k)}$ is η-far from $\mathscr{P}$. More precisely, we say a property $\mathscr{P}$ of k-uniform hypergraphs is **testable with one-sided error** if for every $\eta > 0$ there exist a constant $q = q(\mathscr{P}, \eta)$ and a randomized algorithm $\mathscr{A}$ which does the following:

For a given hypergraph $\mathcal{H}^{(k)}$ the algorithm $\mathscr{A}$ can query some oracle whether a k-tuple K of $V(\mathcal{H}^{(k)})$ spans and edge in $\mathcal{H}^{(k)}$ or not. After at most q queries the algorithm outputs:

- *$\mathcal{H}^{(k)} \in \mathscr{P}$ with probability 1 if $\mathcal{H}^{(k)} \in \mathscr{P}$ and*
- *$\mathcal{H}^{(k)} \notin \mathscr{P}$ with probability at least 2/3 if $\mathcal{H}^{(k)}$ is η-far from $\mathscr{P}$.*

---

*A full version of this paper is available as *Generalizations of the removal lemma* (submitted for publication).

†Supported by NSF grant DMS 0300529.

‡Supported by DFG grant SCHA 1263/1-1.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
STOC'07, June 11–13, 2007, San Diego, California, USA.
Copyright 2007 ACM 978-1-59593-631-8/07/0006 ...$5.00.

If $\mathcal{H}^{(k)} \notin \mathscr{P}$ and $\mathcal{H}^{(k)}$ is not η-far from $\mathscr{P}$, then there are no guarantees for the output of $\mathscr{A}$.

In [19] it was proved that many "natural (hyper)graph properties" such as having chromatic number smaller than ℓ (which essentially appeared already in [13]), having a large clique or having a large cut are testable and the question of characterizing testable graph properties in general was raised.

In a series of papers Alon (jointly with several collaborators) studied that question. In [2] Alon et al. found a logical characterization for a general family of testable graph properties. According to this characterization every first order graph property of type "$\forall\exists$" is testable with one-sided error, while there are first order graph properties of type "$\exists\forall$" that are not testable even with two-sided error. For the proof of the positive result the authors developed a powerful strengthening of Szemerédi's regularity lemma [34]. Alon and Shapira continued that line of research and showed in [4] that in fact every decidable, hereditary property of graphs is testable with one-sided error. A property of k-uniform hypergraphs $\mathscr{P}$ is **hereditary** if it is closed under induced sub-hypergraphs, i.e., $\mathcal{H}^{(k)} \in \mathscr{P}$ implies that also all induced sub-hypergraphs $\mathcal{H}^{(k)}[U] = \mathcal{H}^{(k)} \cap \binom{U}{k}$ are in $\mathscr{P}$ for every $U \subseteq V(\mathcal{H}^{(k)})$. An alternative proof is due to Lovász and Szegedy [24] (see also [11]). For more results concerning testable and non-testable combinatorial properties see, e.g., [6, 16, 18, 30] and the references therein. In this paper we consider testable properties of k-uniform hypergraphs. Our main result is a generalization of the recent work of Alon and Shapira.

THEOREM 1 (MAIN RESULT). *Every decidable, hereditary k-uniform hypergraph property is testable with one-sided error.*

The proof of Theorem 1 is a direct consequence of Theorem 6 below. This is a result from *extremal (hyper)graph theory*, which is a common generalization of an analogous result of Alon and Shapira obtained for graphs [4] and the *removal lemma* for hypergraphs.

1.2 Removal lemma and its generalizations

Answering a question of Brown, T. Sós, and Erdős [12, 33], Ruzsa and Szemerédi [32] established the *triangle removal lemma* for graphs. They proved that every graph which does not contain many triangles "can be easily made" triangle free.

THEOREM 2 (RUZSA & SZEMERÉDI [32]). *For all $\eta > 0$ there exists $c > 0$ and n_0 so that every graph G on $n \geq n_0$ vertices, which contains at most cn^3 triangles can be made triangle free by removing at most $\eta\binom{n}{2}$ edges.* □

A possible generalization of Theorem 2 to hypergraphs was suggested in [15, Problem 6.1]. The first result in this direction was obtained by Frankl and Rödl [17] who extended Theorem 2 to 3-uniform hypergraphs with the triangle replaced by the complete 3-uniform hypergraph on 4 vertices. The general result, which settles the conjecture from [15] was recently obtained independently by Gowers [20] and Nagle, Skokan and the authors [25, 28, 29] and subsequently by Tao in [35].

THEOREM 3 (GOWERS [20], RÖDL ET AL. [25, 28]). *For all k-uniform hypergraphs $\mathcal{F}^{(k)}$ on ℓ vertices and and every $\eta > 0$ there exist $c > 0$ and n_0 so that the following holds. Suppose $\mathcal{H}^{(k)}$ is a k-uniform hypergraph on $n \geq n_0$ vertices. If $\mathcal{H}^{(k)}$ contains at most cn^ℓ copies of $\mathcal{F}^{(k)}$, then one can delete $\eta\binom{n}{k}$ edges from $\mathcal{H}^{(k)}$ so that the resulting sub-hypergraph contains no copy of $\mathcal{F}^{(k)}$.* □

One possible generalization of Theorem 3 is to replace the single hypergraph $\mathcal{F}^{(k)}$ by a possibly infinite family $\mathscr{F}$ of k-uniform hypergraphs. Such a result was first proved for graphs by Alon and Shapira [5] in the context of property testing. For a family of graphs $\mathscr{F}$ consider the class $\mathrm{Forb}(\mathscr{F})$ of all graphs H containing no member of $\mathscr{F}$ as a (not necessarily induced) subgraph.

THEOREM 4 (ALON & SHAPIRA [5]). *For every (possibly infinite) family of graphs $\mathscr{F}$ and every $\eta > 0$ there exist constants $c > 0$, $C > 0$, and n_0 such that the following holds. Suppose H is a graph on $n \geq n_0$ vertices. If for every $\ell = 1, \ldots, C$ and every $F \in \mathscr{F}$ on ℓ vertices, H contains at most at most cn^ℓ copies of F, then one can delete $\eta\binom{n}{2}$ edges from H such that the resulting subgraph H' contains no copy of any member of $\mathscr{F}$, i.e., $H' \in \mathrm{Forb}(\mathscr{F})$.* □

Theorem 3 for graphs is equivalent to the special case of Theorem 4 when $\mathscr{F}$ consists of only one graph. While for finite families $\mathscr{F}$ Theorem 4 can be proved along the lines of the proof of Theorem 3 (or be deduced with $|\mathscr{F}|$ applications of Theorem 3), for infinite families $\mathscr{F}$ the proof of Theorem 4 is more sophisticated. Perhaps one of the earliest results of this nature was obtained by Bollobás et al. [10], who essentially proved Theorem 4 for the special family $\mathscr{F}$ of blow-up's of odd cycles. In [13] answering a question of Erdős (see, e.g., [14]) Duke and Rödl generalized the result from [10] and proved Theorem 4 for the special case of families $\mathscr{F}$ consisting of all $(r+1)$-chromatic graphs $r \geq 2$.

The proof of Theorem 4 for arbitrary families $\mathscr{F}$ relies on a strengthened version of Szemerédi's regularity lemma, which was obtained by Alon et al. [2] by iterating the regularity lemma for graphs. Recently, Theorem 4 was extended by Avart and the authors in [8] from graphs to hypergraphs. The proof in [8] follows the approach of Alon and Shapira and is based on two successive applications of the hypergraph regularity lemma from [27].

Another natural variant of Theorem 3 is an *induced* version. For graphs this was first considered by Alon et al. [2]. Note that in this case in order to obtain an induced F-free graph, we may need not only to remove, but also to add edges.

THEOREM 5 (ALON ET AL. [2]). *For all graphs F on ℓ vertices and every $\eta > 0$ there exist $c > 0$ and n_0 so that the following holds. Suppose H is a graph on $n \geq n_0$ vertices. If H contains at most cn^ℓ induced copies of F, then one can change $\eta\binom{n}{2}$ pairs from $V(H)$ (deleting or adding the edge) so that the resulting graph H' contains no induced copy of F.* □

An extension of Theorem 5 to 3-uniform hypergraphs was obtained by Kohayakawa et al. in [22]. In [4] Alon and Shapira proved a common generalization of Theorem 4 and Theorem 5, extending Theorem 5 from one forbidden induced graph F to a forbidden family of induced graphs $\mathscr{F}$. In this paper we generalize their result to k-uniform hypergraphs.

1.3 New generalization of the removal lemma

For a family of k-uniform hypergraphs $\mathscr{F}$, we denote by $\mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$ the family of all hypergraphs $\mathcal{H}^{(k)}$ which contain no induced copy of any member of $\mathscr{F}$. Clearly, $\mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$ is a *hereditary* family (or *hereditary property*) of hypergraphs.

THEOREM 6. *For every (possibly infinite) family $\mathscr{F}$ of k-uniform hypergraphs and every $\eta > 0$ there exist constants $c > 0$, $C > 0$, and n_0 such that the following holds. Suppose $\mathcal{H}^{(k)}$ is a k-uniform hypergraph on $n \geq n_0$ vertices. If for every $\ell = 1, \ldots, C$ and every $\mathcal{F}^{(k)} \in \mathscr{F}$ on ℓ vertices, $\mathcal{H}^{(k)}$ contains at most cn^{ℓ} induced copies of $\mathcal{F}^{(k)}$, then $\mathcal{H}^{(k)}$ is not η-far from* $\mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$.

For graphs Theorem 6 was obtained by Alon and Shapira [4]. The proof in [4] is again based on the strong version of Szemerédi's regularity lemma from [2]. Another proof for graphs was found by Lovász and Szegedy [24] (see also [11]). Below we discuss the relation of Theorem 6 and Theorem 1.

Recall that for every hereditary property $\mathscr{P}$ of k-uniform hypergraphs, there exists a family of k-uniform hypergraphs $\mathscr{F}$ such that $\mathscr{P} = \mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$. Consequently, Theorem 6 states that if $\mathcal{H}^{(k)}$ is η-far from some hereditary property $\mathscr{P} = \mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$, then it contains many $(cn^{|V(\mathcal{F}^{(k)})|})$ induced copies of some "forbidden" hypergraph $\mathcal{F}^{(k)} \in \mathscr{F}$ of size at most C, which "proves" that $\mathcal{H}^{(k)}$ is not in $\mathscr{P}$. In other words, if $\mathcal{H}^{(k)}$ is η-far from some given hereditary property $\mathscr{P}$, then it is "easy" to detect that $\mathcal{H}^{(k)} \notin \mathscr{P}$. This implies Theorem 1.

PROOF OF THEOREM 1. Let a decidable and hereditary property $\mathscr{P} = \mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$ of k-uniform hypergraphs and a constant $\eta > 0$ be given. By Theorem 6, there is some $c > 0$ and there are integers C and $n_0 \in \mathbb{N}$ such that any k-uniform hypergraph on $n \geq n_0$ vertices, which is η-far from exhibiting $\mathscr{P}$ contains at least cn^{ℓ} induced copies of some $\mathcal{F}_0^{(k)} \in \mathscr{F}$ with $\ell = |V(\mathcal{F}_0^{(k)})| \leq C$.

Let $s \in \mathbb{N}$ be such that $(1-c)^{s/C} < 1/3$ and set $m_0 = \max\{s, n_0\}$. We claim that there exists a one-sided tester with query complexity $q = \binom{m_0}{k}$ for $\mathscr{P}$. For that let $\mathcal{H}^{(k)}$ be a k-uniform hypergraph on n vertices. First observe that we may assume that $n > m_0$. Indeed if $n \leq m_0$, then the tester simply queries all edges of $\mathcal{H}^{(k)}$ and since $\mathscr{P}$ is decidable, there is an exact algorithm with running time only depending on the fixed m_0, which determines correctly if $\mathcal{H}^{(k)} \in \mathscr{P}$ or not.

Consequently, let $n > m_0$. Then we choose uniformly at random a set S of s vertices from $\mathcal{H}^{(k)}$. We consider the hypergraph $\mathcal{H}^{(k)}[S] = \mathcal{H}^{(k)} \cap \binom{S}{k}$ induced on S, i.e., the algorithm queries all $\binom{s}{k}$ k-tuples. If $\mathcal{H}^{(k)}[S]$ has $\mathscr{P}$, then the tester outputs "$\mathcal{H}^{(k)} \in \mathscr{P}$" and otherwise "$\mathcal{H}^{(k)} \notin \mathscr{P}$." Since $\mathscr{P}$ is decidable and s is fixed the algorithm decides whether or not $\mathcal{H}[S]$ is in $\mathscr{P}$ in constant time (constant only depending on s and $\mathscr{P}$). Note that such an algorithm exists, since s is a constant depending only on $\mathscr{A}$ and η and since $\mathscr{A}$ is decidable.

Clearly, if $\mathcal{H}^{(k)} \in \mathscr{P} = \mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$ or $n \leq m_0$, then this tester outputs correctly and hence it is one-sided. On the other hand, if $\mathcal{H}^{(k)}$ is η-far from $\mathscr{P}$ and $n > m_0$, then due to Theorem 6 the random set S spans a copy of $\mathcal{F}_0^{(k)}$ for some $\mathcal{F}_0^{(k)} \in \mathscr{F}$ on $\ell \leq C$ vertices, with probability at least $cn^{\ell}/\binom{n}{\ell} \geq c$. Hence the probability that S does not span any copy of $\mathcal{F}_0^{(k)}$ is at most $(1-c)^{s/\ell} \leq (1-c)^{s/C} < 1/3$. In other words, S spans a copy of $\mathcal{F}_0^{(k)}$ with probability at least $2/3$, which shows that the tester works as specified. □

1.4 Related results

We conclude this introduction with a few remarks on recent results related to Theorem 1.

In [4] Alon and Shapira used the the graph version of Theorem 1 to give a full charecterization of graph properties which admit an *oblivious, one-sided* tester. Such a tester has no access to the size of the input hypergraph and it only queries an induced sub-hypergraph of order $q = q(\mathscr{P}, \eta)$ uniformly at random from all labeled, induced sub-hypergraphs of size q. The main result in [4] states that a graph property admits an oblivious, one-sided tester if and only if it is semi-hereditary. Here a property $\mathscr{Q}$ is **semi-hereditary** if it is contained in an hereditary property $\mathscr{P}$ and every hypergraph $\mathcal{H}^{(k)}$ η-far from $\mathscr{Q}$ contains an induced sub-hypergraph of order at most $M(\eta)$, which is not in $\mathscr{P}$. The proof of this charecterization works verbatim for hypergraphs and we omit the details here.

In [3] Alon et al. gave a charecterization of graph properties testable with two-sided error, where **two-sided** means the tester is not required to recognize that $\mathcal{H}^{(k)} \in \mathscr{P}$ with probability 1, but only with probability stricly bigger than $1/2$. Roughly speaking, the main result in [3] states that a graph property $\mathscr{P}$ is testable with two-sided error if and only if testing $\mathscr{P}$ can be reduced to testing the property of satisfying one of finitely many "regular partitions" coming from Szemerédi's regularity lemma. In other words, a property $\mathscr{P}$ is testable if and only if distinguishing between graphs having the property $\mathscr{P}$ and being η-far is equivalent to admitting one of finitely many "cluster graphs" (see Section 2.3) after an application of Szemerédi's regularity lemma. We believe the methods developed in this paper, can be used to generalize this charecterization from graphs to hypergraphs and intend to come back to this in the near future.

The last remark concerns a result from [7]. Based on techniques developed in [5] it was shown there, that for every monotone graph property $\mathscr{P}$ and $\varepsilon > 0$ there exists a deterministic algorithm, which in polynomial time approximates up to εn^2 the *distance* from an input n-vertex graph $\mathcal{G}$ to $\mathscr{P}$. Here the distance is the minimum number of edges must be deleted from G to make it satisfy $\mathscr{P}$. We believe it would be interesting to derive an analogous result for hypergraphs. However, in the design of the above graph algorithm the algorithmic version of Szemerédi regularity lemma [1] played crucial rôle. For hypergraphs an algorithmic regularity lemma for 3-uniform hypergraphs was recently developed by Haxell et al. in [21].

Organization. In the rest of the paper we outline the proof of Theorem 6. For that we introduce the necessary definitions of the *regularity method for hypergraphs* in Section 2. In Section 3 we introduce the main lemma, Lemma 15, and deduce Theorem 6 from it. For the proof of Lemma 15 we refer to the full version of this paper [26].

2. REGULARITY METHOD

2.1 Basic definitions

In this paper ℓ-partite, j-uniform hypergraphs play a special rôle, where $j \leq \ell$. Given vertex sets $V_1, \ldots, V_\ell$, we denote by $K_\ell^{(j)}(V_1,\dots,V_\ell)$ the **complete** ℓ-partite, j-uniform hypergraph (i.e., the family of all j-element subsets $J \subseteq \bigcup_{i\in[\ell]} V_i$ satisfying $|V_i \cap J| \le 1$ for every $i \in [\ell]$). If $|V_i| = m$ for every $i \in [\ell]$, then an (m,ℓ,j)**-hypergraph** $\mathcal{H}^{(j)}$ on $V_1 \cup \dots \cup V_\ell$ is any subset of $K_\ell^{(j)}(V_1,\dots,V_\ell)$. The vertex partition $V_1 \cup \dots \cup V_\ell$ is an $(m,\ell,1)$-hypergraph $\mathcal{H}^{(1)}$. For $j \le i \le \ell$ and set $\Lambda_i \in [\ell]^i$, we denote by $\mathcal{H}^{(j)}[\Lambda_i] = \mathcal{H}^{(j)}\left[\bigcup_{\lambda\in\Lambda_i} V_\lambda\right]$ the sub-hypergraph of the (m,ℓ,j)-hypergraph $\mathcal{H}^{(j)}$ induced on $\bigcup_{\lambda\in\Lambda_i} V_\lambda$. For an (m,ℓ,j)-hypergraph $\mathcal{H}^{(j)}$ and an integer $2 \le j \le i \le \ell$, we denote by $\mathcal{K}_i(\mathcal{H}^{(j)})$ the family of all i-element subsets of $V(\mathcal{H}^{(j)})$ which span complete sub-hypergraphs in $\mathcal{H}^{(j)}$ of order i. For $1 \le i \le \ell$, we denote by $\mathcal{K}_i(\mathcal{H}^{(1)})$ the family of all i-element subsets of $V(\mathcal{H}^{(1)})$ which 'cross' the partition $V_1 \cup \dots \cup V_\ell$, i.e., $I \in \mathcal{K}_i(\mathcal{H}^{(1)})$ if, and only if, $|I \cap V_s| \le 1$ for all $1 \le s \le \ell$. For $2 \le j \le i \le \ell$, $|\mathcal{K}_i(\mathcal{H}^{(j)})|$ is the number of all copies of $K_i^{(j)}$ in $\mathcal{H}^{(j)}$. Given an $(m,\ell,j-1)$-hypergraph $\mathcal{H}^{(j-1)}$ and an (m,ℓ,j)-hypergraph $\mathcal{H}^{(j)}$, we say $\mathcal{H}^{(j-1)}$ **underlies** $\mathcal{H}^{(j)}$ if $\mathcal{H}^{(j)} \subseteq \mathcal{K}_j(\mathcal{H}^{(j-1)})$. This brings us to one of the main concepts of this paper, the notion of a complex.

Definition 7. *Let $m \ge 1$ and $\ell \ge h \ge 1$ be integers. An (m,ℓ,h)-**complex** $\boldsymbol{\mathcal{H}}$ is a collection of (m,ℓ,j)-hypergraphs $\{\mathcal{H}^{(j)}\}_{j=1}^h$ such that*

(a) $\mathcal{H}^{(1)}$ is an $(m,\ell,1)$-hypergraph, i.e., $\mathcal{H}^{(1)} = V_1 \cup \dots \cup V_\ell$ with $|V_i| = m$ for $i \in [\ell]$, and

(b) $\mathcal{H}^{(j-1)}$ underlies $\mathcal{H}^{(j)}$ for $2 \le j \le h$.

We sometimes shorten the notation and write (ℓ,h)-complex and (ℓ,h)-hypergraph for (m,ℓ,h)-complex and (m,ℓ,h)-hypergraph, when the cardinality $m = |V_1| = \dots = |V_s|$ isn't of primary concern.

We define the relative **density of a j-uniform hypergraph $\mathcal{H}^{(j)}$ w.r.t. $(j-1)$-uniform hypergraph $\mathcal{H}^{(j-1)}$** on the same vertex set by

$$d\big(\mathcal{H}^{(j)}\big|\mathcal{H}^{(j-1)}\big) = \begin{cases} \frac{|\mathcal{H}^{(j)} \cap \mathcal{K}_j(\mathcal{H}^{(j-1)})|}{|\mathcal{K}_j(\mathcal{H}^{(j-1)})|} & \text{if } \left|\mathcal{K}_j(\mathcal{H}^{(j-1)})\right| > 0\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

We also define a notion of regularity for (m,j,j)-hypergraphs w.r.t. some underlying $(m,j,j-1)$-hypergraphs.

Definition 8. *Let constants $\delta > 0$ and $d \ge 0$ and a positive integer r be given along with an (m,j,j)-hypergraph $\mathcal{H}^{(j)}$ and an $(m,j,j-1)$-hypergraph $\mathcal{H}^{(j-1)}$ on the same vertex set. We say $\mathcal{H}^{(j)}$ is (δ,d,r)-**regular** w.r.t. $\mathcal{H}^{(j-1)}$ if for every collection $\boldsymbol{\mathcal{X}} = \{\mathcal{X}_1^{(j-1)},\dots,\mathcal{X}_r^{(j-1)}\}$ of not necessarily disjoint sub-hypergraphs of $\mathcal{H}^{(j-1)}$ satisfying*

$$\left|\bigcup_{i\in[r]} \mathcal{K}_j(\mathcal{X}_i^{(j-1)})\right| > \delta \left|\mathcal{K}_j(\mathcal{H}^{(j-1)})\right|\,,$$

we have

$$d\big(\mathcal{H}^{(j)}\big|\boldsymbol{\mathcal{X}}\big) = \frac{\left|\mathcal{H}^{(j)} \cap \bigcup_{i\in[r]} \mathcal{K}_j(\mathcal{X}_i^{(j-1)})\right|}{\left|\bigcup_{i\in[r]} \mathcal{K}_j(\mathcal{X}_i^{(j-1)})\right|} = d \pm \delta\,.$$

*We write $(\delta,*,r)$-**regular** to mean $\left(\delta, d\big(\mathcal{H}^{(k)}\big|\mathcal{H}^{(k-1)}\big), r\right)$-regular. Moreover, we say $\mathcal{H}^{(j)}$ is $(\delta,\ge d,r)$-**regular** with respect to $\mathcal{H}^{(j-1)}$ if $d\big(\mathcal{H}^{(k)}\big|\mathcal{H}^{(k-1)}\big) \ge d$ and $\mathcal{H}^{(j)}$ is $(\delta,*,r)$-regular w.r.t. $\mathcal{H}^{(j-1)}$.*

Next we extend the notion of regular (m,j,j)-hypergraph to (m,ℓ,j)-hypergraphs and to complexes.

Definition 9. *For positive integers m, $\ell \ge j$ we say an (m,ℓ,j)-hypergraph $\mathcal{H}^{(j)}$ is (δ,d,r)-**regular** (resp. $(\delta,\ge d,r)$-**regular**) w.r.t. an $(m,\ell,j-1)$-hypergraph $\mathcal{H}^{(j-1)}$ if for every $\Lambda_j \in [\ell]^j$, the restriction $\mathcal{H}^{(j)}[\Lambda_j] = \mathcal{H}^{(j)}\left[\bigcup_{\lambda\in\Lambda_j} V_\lambda\right]$ is (δ,d,r)-regular (resp. $(\delta,\ge d,r)$-regular) w.r.t. the restriction $\mathcal{H}^{(j-1)}[\Lambda_j] = \mathcal{H}^{(j-1)}\left[\bigcup_{\lambda\in\Lambda_j} V_\lambda\right]$.*

Definition 10. *For $h \ge 2$ let $\boldsymbol{\delta} = (\delta_2,\dots,\delta_h)$ be a vector of positive reals and let $\boldsymbol{d} = (d_2,\dots,d_h)$ be a vector of non-negative reals. We say an (m,ℓ,h)-complex $\boldsymbol{\mathcal{H}} = \{\mathcal{H}^{(j)}\}_{j=1}^h$ is $(\boldsymbol{\delta},\boldsymbol{d},r)$-**regular** (resp. $(\boldsymbol{\delta},\ge\boldsymbol{d},r)$-**regular**) if*

(i) $\mathcal{H}^{(2)}$ is $(\delta_2,d_2,1)$-regular (resp. $(\delta_2,\ge d_2,1)$-regular) w.r.t. $\mathcal{H}^{(1)}$ and

(ii) $\mathcal{H}^{(j)}$ is (δ_j,d_j,r)-regular (resp. $(\delta_j,\ge d_j,1)$-regular) w.r.t. $\mathcal{H}^{(j-1)}$ for every $j = 3,\dots,h$.

2.2 Partitions

The regularity lemma for k-uniform hypergraphs provides a structured family of partitions $\mathscr{P} = \{\mathscr{P}^{(1)},\dots,\mathscr{P}^{(k-1)}\}$ of vertices, pairs, …, and $(k-1)$-tuples of the vertex set. We now discuss the structure of these partitions recursively. Here the partition classes of $\mathscr{P}^{(j)}$ will be (j,j)-hypergraphs, i.e., j-uniform, j-partite hypergraphs.

Let k be a fixed integer and V be a set of vertices. Let $\mathscr{P}^{(1)} = \{V_1,\dots,V_{|\mathscr{P}^{(1)}|}\}$ be a partition of V. For every $1 \le j \le |\mathscr{P}^{(1)}|$, let $\mathrm{Cross}_j(\mathscr{P}^{(1)})$ be the family of all crossing j-tuples J, i.e., the set of j-tuples which satisfy $|J \cap V_i| \le 1$ for every $V_i \in \mathscr{P}^{(1)}$.

Suppose for $1 \le i \le j-1$ partitions $\mathscr{P}^{(i)}$ of $\mathrm{Cross}_i(\mathscr{P}^{(1)})$ into (i,i)-hypergraphs are given. Then for every $(j-1)$-tuple I in $\mathrm{Cross}_{j-1}(\mathscr{P}^{(1)})$, there exists a unique $(j-1,j-1)$-hypergraph $\mathcal{P}^{(j-1)} = \mathcal{P}^{(j-1)}(I) \in \mathscr{P}^{(j-1)}$ so that $I \in \mathcal{P}^{(j-1)}$. For every j-tuple J in $\mathrm{Cross}_j(\mathscr{P}^{(1)})$, we define the **polyad** of J

$$\hat{\mathcal{P}}^{(j-1)}(J) = \bigcup \left\{\mathcal{P}^{(j-1)}(I)\colon\ I \in [J]^{j-1}\right\}\,.$$

In other words, $\hat{\mathcal{P}}^{(j-1)}(J)$ is the unique set of j partition classes (or $(j-1,j-1)$-hypergraphs) of $\mathscr{P}^{(j-1)}$ each containing a $(j-1)$-subset of J. Observe that $\hat{\mathcal{P}}^{(j-1)}(J)$ we view as a $(j,j-1)$-hypergraph. More generally, for $1 \le i < j$, we set

$$\begin{aligned} \hat{\mathcal{P}}^{(i)}(J) &= \bigcup \left\{\mathcal{P}^{(i)}(I)\colon\ I \in [J]^i\right\} \\ \text{and}\quad \boldsymbol{\mathcal{P}}(J) &= \left\{\hat{\mathcal{P}}^{(i)}(J)\right\}_{i=1}^{j-1}\,. \end{aligned} \tag{1}$$

Next, we define $\hat{\mathscr{P}}^{(j-1)}$, the family of all polyads

$$\hat{\mathscr{P}}^{(j-1)} = \left\{\hat{\mathcal{P}}^{(j-1)}(J)\colon\ J \in \mathrm{Cross}_j(\mathscr{P}^{(1)})\right\}\,.$$

Note that $\hat{\mathcal{P}}^{(j-1)}(J)$ and $\hat{\mathcal{P}}^{(j-1)}(J')$ are not necessarily distinct for different j-tuples J and J'.

The requirement on the partition $\mathscr{P}^{(j)}$ of $\mathrm{Cross}_j(\mathscr{P}^{(1)})$ is

$$\mathscr{P}^{(j)} \prec \{\mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)})\colon\ \hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}\}\,. \tag{2}$$

In other words, we require that the set of cliques spanned by a polyad in $\hat{\mathscr{P}}^{(j-1)}$ is sub-partitioned in $\mathscr{P}^{(j)}$ and every partition class in $\mathscr{P}^{(j)}$ belongs to precisely one polyad in $\hat{\mathscr{P}}^{(j-1)}$. Note that (2) implies (inductively) that $\boldsymbol{\mathcal{P}}(J)$ defined in (1) is a $(j, j-1)$-complex.

Throughout this paper, we want to have an upper bound on the number of partition classes in $\mathscr{P}^{(j)}$, and more specifically, over the number of classes contained in $\mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)})$ for a fixed polyad $\hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}$. We make this precise in the following definition.

DEFINITION 11. *Suppose V is a set of vertices, $k \geq 2$ is an integer and $\boldsymbol{a} = (a_1, \ldots, a_{k-1})$ is a vector of positive integers. We say $\mathscr{P} = \mathscr{P}(k-1, \boldsymbol{a}) = \{\mathscr{P}^{(1)}, \ldots, \mathscr{P}^{(k-1)}\}$* ***is a family of partitions on*** *V, if it satisfies the following:*

(i) $\mathscr{P}^{(1)}$ is a partition of V into a_1 classes,

(ii) $\mathscr{P}^{(j)}$ is a partition of $\mathrm{Cross}_j(\mathscr{P}^{(1)})$ satisfying:

$$\mathscr{P}^{(j)} \prec \{\mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)}) \colon \hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}\}$$

and for every $\hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}$ we have

$$\left|\left\{\mathcal{P}^{(j)} \in \mathscr{P}^{(j)} \colon \mathcal{P}^{(j)} \subseteq \mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)})\right\}\right| = a_j\,.$$

Moreover, we say $\mathscr{P} = \mathscr{P}(k-1, \boldsymbol{a})$ is T-bounded, if

$$\max\{a_1, \ldots, a_{k-1}\} \leq T\,.$$

2.3 Similarity of hypergraphs

An important part of the argument in the proof of Theorem 6 will be to compare hypergraphs of very different sizes to find two of "similar structure." For that we will use the hypergraph regularity lemma. This lemma provides "regular" families of partitions. Similarly as in the graph case, if the hypergraph regularity lemma is applied to different hypergraphs with the *same* input parameters, then the sizes of the families of partitions corresponding to each of the "regularized" hypergraphs are bounded by the same T_0. Let us assume for now that all the partitions have the same size or more precisely have the same vector $\boldsymbol{a}$. Then we would like to say that two hypergraphs have the same structure, if there densities are similar on "every pair of corresponding polyads," for an appropriate bijection between the polyads of two partitions. The similar idea of comparing "cluster graphs" corresponding to graphs of various sizes was used by Lovász and Szegedy [24].

Let $\mathscr{P}(k-1, \boldsymbol{a})$ be a family of partitions on V (see Definition 11). Consider an arbitrary numbering of the vertex classes of $\mathscr{P}^{(1)}$, i.e., $\mathscr{P}^{(1)} = \{V_i \colon i \in [a_1]\}$. For $j = 2, \ldots, k-1$ let $\varphi^{(j)} \colon \mathscr{P}^{(j)} \to [a_j]$ be a labeling such that for every polyad $\hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}$ the members of $\{\mathcal{P}^{(j)} \in \mathscr{P}^{(j)} \colon \mathcal{P}^{(j)} \subseteq \mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)})\}$ are numbered from 1 to a_j in an arbitrary way. This way, we obtain for every k-tuple $K = \{v_1, \ldots, v_k\} \in \mathrm{Cross}_k(\mathscr{P}^{(1)})$ an integer vector $\hat{\boldsymbol{x}}_K = (\boldsymbol{x}_K^{(1)}, \ldots, \boldsymbol{x}_K^{(k-1)})$, where

$$\boldsymbol{x}_K^{(1)} = (\alpha_1 < \cdots < \alpha_k) \text{ so that w.l.o.g. } K \cap V_{\alpha_i} = \{v_i\} \tag{3}$$

and for $j = 2, \ldots, k-1$ we set

$$\boldsymbol{x}_K^{(j)} = \Big(\varphi^{(j)}(\mathcal{P}^{(j)}) \colon \{v_\lambda \colon \lambda \in \Lambda\} \in \mathcal{P}^{(j)}\Big)_{\Lambda \in \binom{[k]}{j}} \tag{4}$$

Let $\binom{[a_1]}{k}_< = \{(\alpha_1, \ldots, \alpha_k) \colon 1 \leq \alpha_1 < \cdots < \alpha_k \leq a_1\}$ be the set of all "naturally" ordered k-element subsets of $[a_1]$ and set

$$\hat{A}(k-1, \boldsymbol{a}) = \binom{[a_1]}{k}_< \times \prod_{j=2}^{k-1} \underbrace{[a_j] \times \cdots \times [a_j]}_{\binom{k}{j}\text{-times}} \tag{5}$$

for the address space of all k-tuples $K \in \mathrm{Cross}_k(\mathscr{P}^{(1)})$. The definitions above yield $\hat{\boldsymbol{x}}_K \in \hat{A}(k-1, \boldsymbol{a})$ for every $K \in \mathrm{Cross}_k(\mathscr{P}^{(1)})$. Moreover, for every $\hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}$ we have

$$\hat{\boldsymbol{x}}_K = \hat{\boldsymbol{x}}_{K'} \text{ for all } K, K' \in \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}) \tag{6}$$

hence, for every $\hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}$ with $\mathcal{K}_k(\hat{\mathscr{P}}^{(k-1)}) \neq \emptyset$ we may set

$$\hat{\boldsymbol{x}}(\hat{\mathcal{P}}^{(k-1)}) = \hat{\boldsymbol{x}}_K \text{ for some } K \in \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)})\,. \tag{7}$$

Let

$$\hat{\mathscr{P}}^{(k-1)}_{\neq\emptyset} = \left\{\hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)} \colon \mathcal{K}_k(\hat{\mathscr{P}}^{(k-1)}) \neq \emptyset\right\}$$

and

$$\hat{A}_{\neq\emptyset} = \big\{\hat{\boldsymbol{x}} \in \hat{A}(k-1, \boldsymbol{a}) \colon \exists\, \hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}_{\neq\emptyset} \text{ s.t. } \hat{\boldsymbol{x}}(\hat{\mathcal{P}}^{(k-1)}) = \hat{\boldsymbol{x}}\big\}\,.$$

It is easy to see that the definition in (7) establishes a bijection between $\hat{\mathscr{P}}^{(k-1)}_{\neq\emptyset}$ and $\hat{A}_{\neq\emptyset}$.

Moreover, since $|\hat{\mathscr{P}}^{(k-1)}| = |\hat{A}(k-1, \boldsymbol{a})|$ this bijection can be extended to a bijection between $\hat{\mathscr{P}}^{(k-1)}$ and $\hat{A}(k-1, \boldsymbol{a})$. The inverse bijection maps $\hat{\boldsymbol{x}} \mapsto \hat{\mathcal{P}}^{(k-1)}(\hat{\boldsymbol{x}})$ and in the case $\mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(\hat{\boldsymbol{x}})) \neq \emptyset$, i.e., $\hat{\boldsymbol{x}} \in \hat{A}_{\neq\emptyset}$ then

$$\boldsymbol{\mathcal{P}}(\hat{\boldsymbol{x}}) = \boldsymbol{\mathcal{P}}(K) \text{ for some } K \in \hat{\mathcal{P}}^{(k-1)}(\hat{\boldsymbol{x}})\,,$$

is well defined due to (6). Note that $\boldsymbol{\mathcal{P}}(\hat{\boldsymbol{x}}) = \{\mathcal{P}^{(j)}\}_{j=1}^{k-1}$ is a $(k, k-1)$-complex with $\mathcal{P}^{(k-1)} = \hat{\mathcal{P}}^{(k-1)}(\hat{\boldsymbol{x}})$. For later reference we summarize the above.

DEFINITION 12. *Suppose $k \geq 2$ is an integer and $\boldsymbol{a} = (a_1, \ldots, a_{k-1})$ is a vector of positive integers. We say*

$$\hat{A}(k-1, \boldsymbol{a}) = \binom{[a_1]}{k}_< \times \prod_{j=2}^{k-1} \underbrace{[a_j] \times \cdots \times [a_j]}_{\binom{k}{j}\text{-times}},$$

is the ***address space****.*

For a family of partitions $\mathscr{P}(k-1, \boldsymbol{a})$ on $V = V_1 \cup \cdots \cup V_{a_1}$ we say a set of mappings $\boldsymbol{\varphi} = \{\varphi^{(2)}, \ldots, \varphi^{(k-1)}\}$ with $\varphi^{(j)} \colon \mathscr{P}^{(j)} \to [a_j]$ for every $j = 2, \ldots, k-1$ is an ***$\boldsymbol{a}$-labeling*** *if for every $\hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}$ we have*

$$\varphi^{(j)}\big(\{\mathcal{P}^{(j)} \in \mathscr{P}^{(j)} \colon \mathcal{P}^{(j)} \subseteq \mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)})\}\big) = [a_j]\,.$$

Then $\hat{\boldsymbol{x}}_K = (\boldsymbol{x}_K^{(1)}, \ldots, \boldsymbol{x}_K^{(k-1)}) \in \hat{A}(k-1, \boldsymbol{a})$ defined in (3) and (4) defines an equivalence relation on $\mathrm{Cross}_k(\mathscr{P}^{(1)})$. Consequently, such a labeling $\boldsymbol{\varphi}$ defines a bijection between $\hat{A}_{\neq\emptyset}$ and $\hat{\mathscr{P}}^{(k-1)}_{\neq\emptyset}$ (see (7) and below) which can be extended to a bijection between $\hat{A}(k-1, \boldsymbol{a})$ and $\hat{\mathscr{P}}^{(k-1)}$ such that

(a) $\hat{\boldsymbol{x}} \in \hat{A}(k-1, \boldsymbol{a}) \mapsto \hat{\mathcal{P}}^{(k-1)}(\hat{\boldsymbol{x}}) \in \hat{\mathscr{P}}^{(k-1)}$ and

(b) if $\mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(\hat{\boldsymbol{x}})) \neq \emptyset$, then $\boldsymbol{\mathcal{P}}(\hat{\boldsymbol{x}}) = \boldsymbol{\mathcal{P}}(K)$ for some $K \in \hat{\mathcal{P}}^{(k-1)}(\hat{\boldsymbol{x}})$ is well defined,

(c) $K \in \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(\hat{\boldsymbol{x}}_K))$ *for every* $K \in \mathrm{Cross}_k(\mathscr{P}^{(1)})$, *and*

(d) $\boldsymbol{\mathcal{P}}(\hat{\boldsymbol{x}}) = \{\mathcal{P}^{(j)}\}_{j=1}^{k-1}$ *is a* $(k, k-1)$*-complex with* $\mathcal{P}^{(k-1)} = \hat{\mathcal{P}}^{(k-1)}(\hat{\boldsymbol{x}})$.

The following definition will enable us to compare hypergraphs of different sizes. Roughly speaking, we will think of two hypergraphs of being "similar" if there exists an integer vector $\boldsymbol{a}$ so that for each of them there exists an $\boldsymbol{a}$-labeled family of partitions on there respective vertex sets such that for every $\hat{\boldsymbol{x}} \in \hat{A}(k-1, \boldsymbol{a})$ the hypergraphs have the similar density on the respective polyad with address $\hat{\boldsymbol{x}}$.

Definition 13. *Suppose* $\varepsilon > 0$, $\boldsymbol{a} = (a_1, \dots, a_{k-1})$ *is a vector of positive integers,* $\hat{A}(k-1, \boldsymbol{a})$ *is an address space,* $d_{\boldsymbol{a},k}\colon \hat{A}(k-1, \boldsymbol{a}) \to [0, 1]$ *is a* ***density function****, and* $\mathcal{H}^{(k)}$ *is a* k*-uniform hypergraph. We say an* $\boldsymbol{a}$*-labeled family of partitions* $\mathscr{P} = \mathscr{P}(k-1, \boldsymbol{a})$ *on* $V(\mathcal{H}^{(k)})$ *is a* $(d_{\boldsymbol{a},k}, \varepsilon)$***-partition of*** $\mathcal{H}^{(k)}$ *if* $d\big(\mathcal{H}^{(k)} \big| \hat{\mathcal{P}}^{(k-1)}(\hat{\boldsymbol{x}})\big) = d_{\boldsymbol{a},k}(\hat{\boldsymbol{x}}) \pm \varepsilon$ *for every* $\hat{\boldsymbol{x}} \in \hat{A}(k-1, \boldsymbol{a})$.

The concepts above allow to define an object similar to the *cluster graph* in the context of Szemerédi's regularity lemma. For a given $\delta > 0$ Szemerédi's regularity lemma provides a partition of the vertex set $V = V_1 \cup \dots \cup V_t$ of a given graph G, so that all but δt^2 pairs (V_i, V_j) are $(\delta, *, 1)$-regular. For many applications of that lemma it suffices to "reduce" the whole graph to a weighted graph on $[t]$, where the weight of the edge ij corresponds to the density of the bipartite subgraph of G induced on (V_i, V_j) (usually it will also be useful to mark those edges which correspond to irregular pairs). With that notion of cluster graph, one may say that two graphs G_1 and G_2 have the same structure if they admit a regular partition in the same number of parts so that the weights (densities) of the cluster graphs are essentially equal or deviate by at most ε.

The notion of address space extends the concept of the vertex labeling of the cluster graph in the context of the hypergraph regularity lemma. This way the function $d_{\boldsymbol{a},k}$ plays the rôle of the edge weights of the cluster graph. As we considered two graphs to be similar if they admit a regular partition with essentially the same cluster graph, we will view hypergraphs $\mathcal{H}_1^{(k)}$ and $\mathcal{H}_2^{(k)}$ to be ε-similar if there exists an integer vector $\boldsymbol{a}$ (and hence an address space $\hat{A}(k-1, \boldsymbol{a})$) and a density function function $d_{\boldsymbol{a},k}$ such that there is a "regular" $(d_{\boldsymbol{a},k}, \varepsilon)$-partition $\mathscr{P}_1(k-1, \boldsymbol{a})$ of $\mathcal{H}_1^{(k)}$ and a "regular" $(d_{\boldsymbol{a},k}, \varepsilon)$-partition $\mathscr{P}_2(k-1, \boldsymbol{a})$ of $\mathcal{H}_2^{(k)}$.

3. PROOF OF MAIN RESULT

In our argument we will assume that Theorem 6 fails. This means that there exists a family of k-uniform hypergraphs $\mathscr{F}$ and a constant $\eta > 0$ such that for every c, C, and n_0 there exists a hypergraph $\mathcal{H}^{(k)}$ on $n \ge n_0$ vertices which is η-far from $\mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$ and which for every $\ell \le C$ contains at most cn^ℓ induced copies of $\mathcal{F}^{(k)}$ for every $\mathcal{F}^{(k)} \in \mathscr{F}$ on ℓ vertices. Applying this assumption successively with $c = 1/i$ and $C = i$ for $i \in \mathbb{N}$ yields the following fact.

Fact 14. *If Theorem 6 fails for a family of* k*-uniform hypergraphs* $\mathscr{F}$ *and* $\eta > 0$*, then there exists a sequence of hypergraphs* $(\mathcal{H}_i^{(k)})_{i=1}^\infty$ *with* $n_i = |V(\mathcal{H}_i^{(k)})| \to \infty$ *such that for every* $i \in \mathbb{N}$

(i) $\mathcal{H}_i^{(k)}$ *is* η*-far from* $\mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$ *and*

(ii) $\mathcal{H}_i^{(k)}$ *contains less than* n_i^ℓ / i *induced copies of every* $\mathcal{F}^{(k)} \in \mathscr{F}$ *with* $|V(\mathcal{F}^{(k)})| = \ell \le i$.

The same assumption (for graphs) was considered by Lovász and Szegedy [24]. While they derived a contradiction based on the properties of a "limit object" of a sub-sequence of $(\mathcal{H}_i^{(k)})_{i=1}^\infty$ the existence of which was established in [23], here we will only consider hypergraphs of the sequence $(\mathcal{H}_i^{(k)})_{i=1}^\infty$. More precisely, the following, main lemma in the proof of Theorem 6, will locate two special hypergraphs $\mathcal{I}^{(k)} = \mathcal{H}_i^{(k)}$ and $\mathcal{J}^{(k)} = \mathcal{H}_j^{(k)}$ in the sequence from which we derive a contradiction.

Lemma 15. *Suppose Theorem 6 fails for a family* $\mathscr{F}$ *and* $\eta > 0$*. Then there exist*

(i) *a* k*-uniform hypergraphs* $\mathcal{I} = \mathcal{I}^{(k)}$ *on* ℓ *vertices and* $\mathcal{J} = \mathcal{J}^{(k)}$ *on* $n \ge \ell$ *vertices,*

(ii) *integer vectors* $\boldsymbol{a} = (a_1, \dots, a_{k-1}) \in \mathbb{N}^{k-1}$ *and* $\boldsymbol{b} = (b_1, \dots, b_{k-1}) \in \mathbb{N}^{k-1}$,

(iii) *family of partitions* $\mathscr{Q}_\mathcal{I} = \mathscr{Q}_\mathcal{I}(k-1, \boldsymbol{a})$ *on* $V(\mathcal{I}^{(k)})$ *and* $\mathscr{Q}_\mathcal{J} = \mathscr{Q}_\mathcal{J}(k-1, \boldsymbol{a})$ *and* $\mathscr{P}_\mathcal{J} = \mathscr{P}_\mathcal{J}(k-1, \boldsymbol{b})$ *on* $V(\mathcal{J}^{(k)})$,

(iv) *and a density function* $d_{\boldsymbol{a},k}\colon \hat{A}(k-1, \boldsymbol{a}) \to [0, 1]$

such that

(I1) $\mathscr{Q}_\mathcal{I}$ *is a* $(d_{\boldsymbol{a},k}, \eta/24)$*-partition of* $\mathcal{I}^{(k)}$,

(I2) $|\mathrm{Cross}_k(\mathscr{Q}_\mathcal{I}^{(1)})| \ge (1 - \frac{\eta}{24})\binom{\ell}{k}$*, and*

(I3) $\mathcal{I}^{(k)}$ *is* η*-far from* $\mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$.

and

(J1) $\mathscr{Q}_\mathcal{J}$ *is a* $(d_{\boldsymbol{a},k}, \eta/24)$*-partition of* $\mathcal{J}^{(k)}$ *and*

(J2) $\mathscr{P}_\mathcal{J} \prec \mathscr{Q}_\mathcal{J}$*, i.e.,* $\mathscr{P}_\mathcal{J}^{(j)}$ *refines* $\mathscr{Q}_\mathcal{J}^{(j)}$ *for every* $j = 1, \dots, k-1$.

Moreover, there exists an ℓ*-set* $L \in \mathrm{Cross}_\ell(\mathscr{P}_\mathcal{J}^{(1)})$ *such that*

(L1) $|L \cap V_i| = |U_i|$ *where* $\mathscr{Q}_\mathcal{I}^{(1)} = \{U_i\colon i \in [a_1]\}$ *and* $\mathscr{Q}_\mathcal{J}^{(1)} = \{V_i\colon i \in [a_1]\}$,

(L2)

$$\Big|\big\{K \in \tbinom{L}{k} \cap \mathrm{Cross}_k(\mathscr{Q}_\mathcal{J}^{(1)})\colon |d(\mathcal{J}^{(k)}|\hat{\mathcal{P}}_J^{(k-1)}(K)) - d(\mathcal{J}^{(k)}|\hat{\mathcal{Q}}_J^{(k-1)}(K))| > \tfrac{\eta}{12}\big\}\Big| \le \tfrac{4\eta}{9}\tbinom{\ell}{k}$$

(L3) *any* k*-uniform hypergraph* $\mathcal{G}^{(k)}$ *with vertex set* L *and with the property*

$$K \in \mathcal{G}^{(k)} \Rightarrow d(\mathcal{J}^{(k)}|\hat{\mathcal{P}}_J^{(k-1)}(K)) \ge \tfrac{\eta}{12}$$
$$\text{and} \quad K \notin \mathcal{G}^{(k)} \Rightarrow d(\mathcal{J}^{(k)}|\hat{\mathcal{P}}_J^{(k-1)}(K)) \le 1 - \tfrac{\eta}{12}\,,$$

belongs to $\mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$.

The proof of Lemma 15 relies on the hypergraph regularity lemma from [28] and the corresponding counting lemma from [25]. We give the proof of Lemma 15 in [26] and below we derive the main result of this paper from Lemma 15.

Proof of Theorem 6. The proof is by contradiction. Suppose there exists a family of k-uniform hypergraphs $\mathscr{F}$ and some $\eta > 0$ so that Theorem 6 fails. We apply Lemma 15 which yields hypergraphs $\mathcal{I}^{(k)}$ (on ℓ vertices) and $\mathcal{J}^{(k)}$ (on n vertices) and an ℓ-set $L \subseteq V(\mathcal{J}^{(k)})$. In view of property (*L3*) we will define a hypergraph $\mathcal{G}^{(k)}$ on the vertex set L. In order to obtain the desired contradiction we will "compare" the ℓ-vertex hypergraph $\mathcal{G}^{(k)}$ with the ℓ-vertex hypergraph $\mathcal{I}^{(k)}$. For that we need some bijection ψ from L to $V(\mathcal{I}^{(k)})$. We choose some bijection ψ which "agrees" with the labellings of $\mathscr{Q}_{\mathcal{J}}$ and $\mathscr{Q}_{\mathcal{I}}$, i.e., we require that for any k-tuple $K \in \mathrm{Cross}_k(\mathscr{Q}_{\mathcal{J}}^{(1)})$ the address $\hat{\boldsymbol{x}}_K$ (see Definition 12) of K w.r.t. the $\boldsymbol{a}$-labeled partition $\mathscr{Q}_{\mathcal{J}}$ coincides with the address $\hat{\boldsymbol{x}}_{\psi(K)}$ of $\psi(K)$ w.r.t. the $\boldsymbol{a}$-labeled partition $\mathscr{Q}_{\mathcal{I}}$. More precisely, fix a bijection $\psi\colon L \to V(\mathcal{I}^{(k)})$ such that for every $K \in \binom{L}{k}$ the following holds: if $K \in \mathrm{Cross}_k(\mathscr{Q}_{\mathcal{J}}^{(1)})$ then

$$\psi(K) \in \mathrm{Cross}_k(\mathscr{Q}_{\mathcal{I}}^{(1)}) \quad \text{and} \quad \hat{\boldsymbol{x}}_K = \hat{\boldsymbol{x}}_{\psi(K)}\,. \tag{8}$$

For a subset of $E \subseteq \binom{L}{k}$ we set $\psi(E) = \{\psi(K)\colon K \in E\}$. We then define the hypergraph $\mathcal{G}^{(k)}$ on L by

$$K \in \mathcal{G}^{(k)} \Leftrightarrow \begin{cases} \text{either } d(\mathcal{J}^{(k)}|\hat{\mathcal{P}}_J^{(k-1)}(K)) \geq \frac{\eta}{12} \text{ and } \psi(K) \in \mathcal{I}^{(k)} \\ \text{or} \quad\ \ d(\mathcal{J}^{(k)}|\hat{\mathcal{P}}_J^{(k-1)}(K)) > 1 - \frac{\eta}{12}\,. \end{cases} \tag{9}$$

for every $K \in \binom{L}{k}$. Consequently, by (*L3*) of Lemma 15 $\mathcal{G}^{(k)} \in \mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$. It is left to show

$$\left|\mathcal{I}^{(k)} \triangle \psi(\mathcal{G}^{(k)})\right| \leq \eta\tbinom{\ell}{k}\,, \tag{10}$$

which then contradicts (*I3*) of Lemma 15, i.e., (10) contradicts that $\mathcal{I}^{(k)}$ is η-far from $\mathrm{Forb}_{\mathrm{ind}}(\mathscr{F})$. We cover the symmetric difference $\mathcal{I}^{(k)} \triangle \psi(\mathcal{G}^{(k)})$ by four sets $D_1, \dots, D_4$ defined by

$$\begin{aligned} D_1 &= \binom{V(\mathcal{I}^{(k)})}{k} \setminus \mathrm{Cross}_k(\mathscr{Q}_{\mathcal{I}}^{(1)})\,, \\ D_2 &= \psi\Big(\Big\{K \in \tbinom{L}{k} \cap \mathrm{Cross}_k(\mathscr{Q}_{\mathcal{J}}^{(1)})\colon \\ &\qquad \left|d(\mathcal{J}^{(k)}|\hat{\mathcal{P}}_J^{(k-1)}(K)) - d(\mathcal{J}^{(k)}|\hat{\mathcal{Q}}_J^{(k-1)}(K))\right| > \tfrac{\eta}{12}\Big\}\Big), \\ D_3 &= \mathcal{I}^{(k)} \cap \bigcup\big\{\mathcal{K}_k(\hat{\mathcal{Q}}_I^{(k-1)})\colon d(\mathcal{I}^{(k)}|\hat{\mathcal{Q}}_I^{(k-1)}) < \tfrac{\eta}{4}\big\}, \\ D_4 &= \binom{L}{k} \setminus \Big(\mathcal{I}^{(k)} \cap \bigcup\big\{\mathcal{K}_k(\hat{\mathcal{Q}}_I^{(k-1)})\colon \\ &\qquad d(\mathcal{I}^{(k)}|\hat{\mathcal{Q}}_I^{(k-1)}) > 1 - \tfrac{\eta}{4}\big\}\Big)\,. \end{aligned}$$

We first show that indeed $\mathcal{I}^{(k)} \triangle \psi(\mathcal{G}^{(k)}) \subseteq D_1 \cup \dots \cup D_4$. For that first consider some $K' \in \mathcal{I}^{(k)} \setminus \psi(\mathcal{G}^{(k)})$ and set $K = \psi^{-1}(K')$. By the definition of $\mathcal{G}^{(k)}$ in (9) we have $d(\mathcal{J}^{(k)}|\hat{\mathcal{P}}_J^{(k-1)}(K)) < \frac{\eta}{12}$. Then it is easy to show that if $K' \notin D_1 \cup D_2$ then $K' \in D_3$. Indeed, we have:

$$\begin{aligned} K' &\in \mathcal{I}^{(k)} \setminus \big(\psi(\mathcal{G}^{(k)}) \cup D_1 \cup D_2\big) \\ &\overset{(9)}{\Rightarrow} d(\mathcal{J}^{(k)}|\hat{\mathcal{P}}_J^{(k-1)}(K)) < \frac{\eta}{12} \\ &\overset{K' \notin D_1 \cup D_2}{\Rightarrow} d(\mathcal{J}^{(k)}|\hat{\mathcal{Q}}_J^{(k-1)}(K)) < \frac{\eta}{6}. \end{aligned} \tag{11}$$

Due to (*J1*) and (*I1*) of Lemma 15, both $\mathscr{Q}_{\mathcal{J}}$ and $\mathscr{Q}_{\mathcal{I}}$ are $(d_{\boldsymbol{a},k}, \eta/24)$-partitions with the same $\hat{A}(k-1, \boldsymbol{a})$ and the same density function $d_{\boldsymbol{a},k}\colon \hat{A}(k-1,\boldsymbol{a}) \to [0,1]$.

Hence, on the one hand, we infer $d(\mathcal{J}^{(k)}|\hat{\mathcal{Q}}_J^{(k-1)}(K)) = d_{\boldsymbol{a},k}(\hat{\boldsymbol{x}}_K) \pm \eta/24$ and, on the other hand, due to (8) and $K = \psi^{-1}(K')$, we have $d(\mathcal{I}^{(k)}|\hat{\mathcal{Q}}_I^{(k-1)}(K')) = d_{\boldsymbol{a},k}(\hat{\boldsymbol{x}}_K) \pm \eta/24$. Thus, $|d(\mathcal{J}^{(k)}|\hat{\mathcal{Q}}_J^{(k-1)}(K)) - d(\mathcal{I}^{(k)}|\hat{\mathcal{Q}}_I^{(k-1)}(K'))| \leq \eta/12$ and the right-hand side of (11) implies

$$\begin{aligned} K' &\in \mathcal{I}^{(k)} \setminus \big(\psi(\mathcal{G}^{(k)}) \cup D_1 \cup D_2\big) \\ &\overset{(11)}{\Rightarrow} d(\mathcal{J}^{(k)}|\hat{\mathcal{Q}}_J^{(k-1)}(K)) < \frac{\eta}{6} \\ &\overset{(J1)\&(I1)}{\Rightarrow} d(\mathcal{I}^{(k)}|\hat{\mathcal{Q}}_I^{(k-1)}(K')) < \frac{\eta}{4} \\ &\Rightarrow K' \in D_3\,. \end{aligned}$$

Similarly, for $K' \in \psi(\mathcal{G}^{(k)}) \setminus \mathcal{I}^{(k)}$ and $K = \psi^{-1}(K')$ we infer by similar arguments as above:

$$\begin{aligned} K' &\in \psi(\mathcal{G}^{(k)}) \setminus \big(\mathcal{I}^{(k)} \cup D_1 \cup D_2\big) \\ &\overset{(9)}{\Rightarrow} d(\mathcal{J}^{(k)}|\hat{\mathcal{P}}_J^{(k-1)}(K)) > 1 - \frac{\eta}{12} \\ &\overset{K' \notin D_1 \cup D_2}{\Rightarrow} d(\mathcal{J}^{(k)}|\hat{\mathcal{Q}}_J^{(k-1)}(K)) > 1 - \frac{\eta}{6} \\ &\overset{(J1)\&(I1)}{\Rightarrow} d(\mathcal{I}^{(k)}|\hat{\mathcal{Q}}_I^{(k-1)}(K')) > 1 - \frac{\eta}{4} \\ &\Rightarrow K' \in D_4\,. \end{aligned}$$

Consequently, $\mathcal{I}^{(k)} \triangle \psi(\mathcal{G}^{(k)}) \subseteq D_1 \cup \dots \cup D_4$. Moreover, from (*I2*) of Lemma 15 we infer

$$|D_1| = \left|\binom{V(\mathcal{I}^{(k)})}{k} \setminus \mathrm{Cross}_k(\mathscr{Q}_{\mathcal{I}}^{(1)})\right| \leq \eta\binom{\ell}{k}/24$$

and (*L2*) implies $|D_2| \leq 4\eta\binom{\ell}{k}/9$. Finally, the definitions of D_3 and D_4 yield $|D_3| \leq \eta\binom{\ell}{k}/4$ and $|D_3| \leq \eta\binom{\ell}{k}/4$. Summarizing the above, we obtain

$$\begin{aligned} \left|\mathcal{I}^{(k)} \triangle \psi(\mathcal{G}^{(k)})\right| &\leq |D_1| + |D_2| + |D_3| + |D_4| \\ &\leq \left(\frac{\eta}{24} + \frac{4\eta}{9} + \frac{\eta}{4} + \frac{\eta}{4}\right)\binom{\ell}{k} < \eta\binom{\ell}{k}\,. \end{aligned}$$

Thus we proved (10), which by definition of $\mathcal{G}^{(k)}$ and (*L3*) contradicts (*I3*) of Lemma 15. $\square$

4. REFERENCES

[1] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster. The algorithmic aspects of the regularity lemma. *J. Algorithms*, 16(1):80–109, 1994.

[2] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[3] N. Alon, E. Fischer, I. Newman, and A. Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. In *Proceedings of the thirty-eighth annual ACM Symposium on Theory of Computing*, pages 251–260, New York, NY, USA, 2006. ACM Press.

[4] N. Alon and A. Shapira. A characterization of the (natural) graph properties testable with one-sided error. In *Proceedings of the fourty-sixth annual IEEE*

Symposium on Foundations of Computer Science, pages 429–438. IEEE Computer Society, 2005.

[5] N. Alon and A. Shapira. Every monotone graph property is testable. In *Proceedings of the thirty-seventh annual ACM Symposium on Theory of Computing*, pages 128–137, New York, NY, USA, 2005. ACM Press.

[6] N. Alon and A. Shapira. Homomorphisms in graph property testing. In *Topics in discrete mathematics*, volume 26 of *Algorithms Combin.*, pages 281–313. Springer, Berlin, 2006.

[7] N. Alon, A. Shapira, and B. Sudakov. Additive approximation for edge-deletion problems. In *Proceedings of the fourty-sixth annual IEEE Symposium on Foundations of Computer Science*, pages 419–428. IEEE Computer Society, 2005.

[8] C. Avart, V. Rödl, and M. Schacht. Every monotone 3-graph property is testable. *SIAM J. Discrete Math.* to appear.

[9] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (Baltimore, MD, 1990)*, volume 47, pages 549–595, 1993.

[10] B. Bollobás, P. Erdős, M. Simonovits, and E. Szemerédi. Extremal graphs without large forbidden subgraphs. *Ann. Discrete Math.*, 3:29–41, 1978. Advances in graph theory (Cambridge Combinatorial Conf., Trinity Coll., Cambridge, 1977).

[11] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, B. Szegedy, and K. Vesztergombi. Graph limits and parameter testing. In *Proceedings of the thirty-eighth annual ACM Symposium on Theory of Computing*, pages 261–270, New York, NY, USA, 2006. ACM Press.

[12] W. G. Brown, P. Erdős, and V. T. Sós. Some extremal problems on r-graphs. In *New directions in the theory of graphs (Proc. Third Ann Arbor Conf., Univ. Michigan, Ann Arbor, Mich, 1971)*, pages 53–63. Academic Press, New York, 1973.

[13] R. A. Duke and V. Rödl. On graphs with small subgraphs of large chromatic number. *Graphs Combin.*, 1(1):91–96, 1985.

[14] P. Erdős. Problems and results on graphs and hypergraphs: similarities and differences. In J. Nešetřil and V. Rödl, editors, *Mathematics of Ramsey theory*, volume 5 of *Algorithms Combin.*, pages 12–28. Springer, Berlin, 1990.

[15] 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(2):113–121, 1986.

[16] E. Fischer. The art of uninformed decisions: a primer to property testing. *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, (75):97–126, 2001.

[17] P. Frankl and V. Rödl. Extremal problems on set systems. *Random Structures Algorithms*, 20(2):131–164, 2002.

[18] O. Goldreich. Combinatorial property testing (a survey). In *Randomization methods in algorithm design (Princeton, NJ, 1997)*, volume 43 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 45–59. Amer. Math. Soc., Providence, RI, 1999.

[19] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[20] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. submitted.

[21] P. Haxell, B. Nagle, and V. Rödl. An algorithmic version of the hypergraph regularity method. In *Proceedings of the fourty-sixth annual IEEE Symposium on Foundations of Computer Science*, pages 439–448. IEEE Computer Society, 2005.

[22] Y. Kohayakawa, B. Nagle, and V. Rödl. Efficient testing of hypergraphs (extended abstract). In *Automata, languages and programming*, volume 2380 of *Lecture Notes in Comput. Sci.*, pages 1017–1028. Springer, Berlin, 2002.

[23] L. Lovász and B. Szegedy. Limits of dense graph sequences. Technical Report MSR-TR-2004-79, Microsoft Research, 2004.

[24] L. Lovász and B. Szegedy. Graph limits and testing hereditary graph properties. Technical Report MSR-TR-2005-110, Microsoft Research, 2005.

[25] B. Nagle, V. Rödl, and M. Schacht. The counting lemma for regular k-uniform hypergraphs. *Random Structures Algorithms*, 28(2):113–179, 2006.

[26] V. Rödl and M. Schacht. Generalizations of the removal lemma. submitted.

[27] V. Rödl and M. Schacht. Regular partitions of hypergraph. *Combin. Probab. Comput.* to appear.

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

[29] V. Rödl and J. Skokan. Applications of the regularity lemma for uniform hypergraphs. *Random Structures Algorithms*, 28(2):180–194, 2006.

[30] D. Ron. Property testing. In *Handbook of randomized computing, Vol. I, II*, volume 9 of *Comb. Optim.*, pages 597–649. Kluwer Acad. Publ., Dordrecht, 2001.

[31] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.

[32] I. Z. Ruzsa and E. Szemerédi. Triple systems with no six points carrying three triangles. In *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II*, volume 18 of *Colloq. Math. Soc. János Bolyai*, pages 939–945. North-Holland, Amsterdam, 1978.

[33] V. T. Sós, P. Erdős, and W. G. Brown. On the existence of triangulated spheres in 3-graphs, and related problems. *Period. Math. Hungar.*, 3(3-4):221–228, 1973.

[34] E. Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[35] T. Tao. A variant of the hypergraph removal lemma. *J. Combin. Theory Ser. A*, 113(7):1257–1280, 2006.