

2009

## A Hypergraph Regularity Method for Linear Hypergraphs

Shoaib Amjad Khan  
*University of South Florida*

Follow this and additional works at: <https://scholarcommons.usf.edu/etd>



Part of the [American Studies Commons](#)

---

### Scholar Commons Citation

Khan, Shoaib Amjad, "A Hypergraph Regularity Method for Linear Hypergraphs" (2009). *Graduate Theses and Dissertations*.

<https://scholarcommons.usf.edu/etd/2042>

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact [scholarcommons@usf.edu](mailto:scholarcommons@usf.edu).

A Hypergraph Regularity Method for Linear Hypergraphs

by

Shoaib Amjad Khan

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Arts  
Department of Mathematics  
College of Arts and Sciences  
University of South Florida

Major Professor: Brendan Nagle, Ph.D.  
Brian Curtin, Ph.D.  
Stephen Suen, Ph.D.

Date of Approval:  
May 11, 2009

Keywords: Regularity Lemma, Counting Lemma, Forbidden Families,  $\mathcal{F}$ -Counting  
Algorithm, Constructive Removal Lemma

©Copyright 2009, Shoaib Amjad Khan

## **Dedication**

To my parents whose loving prayers have been my anchor in turbulent times.

## Acknowledgements

I would like to thank all people who have helped and inspired me during my thesis research.

I especially wish to thank my advisor, Dr. Brendan Nagle, for his continuous guidance and encouragement throughout my thesis study. His perpetual energy and enthusiasm in research have been a constant source of inspiration for me. In addition, he has always been accessible and willing to help me out of any pitfalls that I stumbled into during the course of my research. As a result, I never felt isolated and research life became smooth and rewarding for me. This thesis could not have been possible without his person.

I am grateful to both Dr. Brian Curtin and Dr. Stephen Suen for taking their time to go over my thesis manuscript and for their constructive critique.

My sincere thanks are due to Prof. Nataša Jonoska for her understanding support and encouragement during my final semester. I am indebted to her for opening up new vistas for me by teaching what has become one of the most memorable classes that I have had the fortune to take.

My deepest gratitude goes to my family and friends for all their help. I dedicate my work to my parents whose unwavering love and prayers have been my anchor in turbulent waters.

For everyone aforementioned, and for a chance to thank them, I thank You my Kind and Loving Companion. May Your name be exalted, honored and glorified.

## Table of Contents

Abstract	ii
1 Introduction	1
1.1 Tools . . . . .	3
1.2 Results . . . . .	5
2 Regularity Lemma For $k$ -Uniform Hypergraphs	7
3 Counting Lemma For Simple $k$ -Uniform Hypergraphs	16
3.1 Main Theorem . . . . .	16
3.2 A Generalization . . . . .	20
4 Bounding $ \text{Forb}(n, \mathbb{F}) $	22
5 Counting Algorithm For Simple Hypergraphs	28
6 Constructive Removal Lemma	33
References	36

# A Hypergraph Regularity Method for Linear Hypergraphs

Shoaib A. Khan

## ABSTRACT

Szemerédi's Regularity Lemma is a powerful tool in Graph Theory, yielding many applications in areas such as Extremal Graph Theory, Combinatorial Number Theory and Theoretical Computer Science. Strong hypergraph extensions of graph regularity techniques were recently given by Nagle, Rödl, Schacht and Skokan, by W.T. Gowers, and subsequently, by T. Tao. These extensions have yielded quite a few non-trivial applications to Extremal Hypergraph Theory, Combinatorial Number Theory and Theoretical Computer Science.

A main drawback to the hypergraph regularity techniques above is that they are highly technical. In this thesis, we consider a less technical version of hypergraph regularity which more directly generalizes Szemerédi's regularity lemma for graphs. The tools we discuss won't yield all applications of their stronger relatives, but yield still several applications in extremal hypergraph theory (for so-called *linear* or *simple* hypergraphs), including algorithmic ones. This thesis surveys these lighter regularity techniques, and develops three applications of them.

## 1 Introduction

Szemerédi's regularity lemma [28, 29] is one of the most important tools in combinatorics, with numerous applications ranging across combinatorial number theory, extremal graph theory and theoretical computer science (see [17, 18] for excellent surveys). Roughly speaking, the lemma asserts that every graph can be decomposed into a bounded number of random-like parts, or more formally,  $\epsilon$ -regular pairs, which we now define.

For a graph  $G = (V, E)$  and  $\epsilon > 0$ , we say two non-empty disjoint subsets  $X, Y \subseteq V$  are  $\epsilon$ -regular if for all  $X' \subseteq X, |X'| > \epsilon|X|$  and  $Y' \subseteq Y, |Y'| > \epsilon|Y|$ , we have  $|d_G(X, Y) - d_G(X', Y')| < \epsilon$ , where  $d_G(X', Y') = |G[X', Y']|/(|X'||Y'|)$  is the *density* of the bipartite subgraph  $G[X', Y']$  of  $G$  (consisting of all edges  $\{x, y\} \in E$  with  $x \in X'$  and  $y \in Y'$ ). Szemerédi's regularity lemma is formally stated as follows.

**Theorem 1.1 (Szemerédi Regularity Lemma [28, 29])** *For all  $\epsilon > 0$  and all integers  $t_0$ , there exists an integer  $T_0$  such that every graph  $G = (V, E)$  of order at least  $t_0$  admits an equitable and  $\epsilon$ -regular partition  $V(G) = V_0 \cup V_1 \cup \dots \cup V_t$ ,  $t_0 \leq t \leq T_0$ , meaning*

- i)  $|V_1| = \dots = |V_t|$ ,  $|V_0| < \epsilon|V(G)|$ ;*
- ii) all but  $\epsilon \binom{t}{2}$  pairs  $V_i, V_j$ ,  $1 \leq i < j \leq t$ , are  $\epsilon$ -regular.*

Many applications of Szemerédi's regularity lemma depend on the fact that within an appropriately given  $\epsilon$ -regular partition  $V = V_0 \cup V_1 \cup \dots \cup V_t$  one may enumerate small subgraphs of a fixed isomorphism type. This result is formally due to the 'Counting Lemma' for graphs. While we state a precise version momentarily, roughly the

graph counting lemma would say for example, that if  $V_i, V_j, V_k$  have all pairs  $(V_i, V_j)$ ,  $(V_i, V_k)$ ,  $(V_j, V_k)$   $\epsilon$ -regular with respective densities  $d_{ij}, d_{ik}, d_{jk} > 0$ , then  $G[V_i, V_j, V_k]$  contains  $(1 \pm f(\epsilon))d_{ij}d_{ik}d_{jk}|V_i||V_j||V_k|$  triangles  $K_3$ , where  $f(\epsilon) \rightarrow 0$  as  $\epsilon \rightarrow 0$ . Joint application of Theorem 1.1 and the Counting Lemma for graphs is known as the *Graph Regularity Method* [26].

The great importance of Szemerédi’s Regularity Lemma led to a search for extensions to  $k$ -uniform hypergraphs, see for example [5, 6, 11, 12]. While these early generalizations did lead to some interesting applications, they did not seem to capture the full power of Szemerédi’s lemma for graphs. In particular, they did not allow for the embedding of small subsystems within a regular structure i.e., they did not admit a companion counting lemma for hypergraphs. The first generalized regularity lemma that did have this property was the lemma of Frankl and Rödl [10] for 3-uniform hypergraphs. Extending Frankl and Rödl’s lemma [10], regularity lemmas and counting lemmas for  $k$ -uniform hypergraphs, also allowing the embedding of small substructures, were developed later by Nagle, Rödl, Schacht and Skokan [20, 27], Gowers [13, 14], and subsequently Tao [30]. The combined use of a hypergraph regularity lemma and a hypergraph counting lemma is known as the *Hypergraph Regularity Method*.

A main drawback in applying the hypergraph regularity and counting lemmas discussed above is their significantly technical formulations. For one, they do not transparently generalize Szemerédi’s regularity lemma. In particular, Szemerédi regularity lemma ‘regularizes’ graph edges w.r.t. vertices. The  $k$ -uniform hypergraph regularity lemmas above ‘regularize’  $k$ -tuples w.r.t.  $(k - 1)$ -tuples, which are then regularized w.r.t.  $(k - 2)$ -tuples, and so on. The resulting ‘regular parts’ in these lemmas are, in fact, a family of hypergraphs which, in addition, become increasingly sparse over the lower uniformities.

While the technicality of the hypergraph regularity tools discussed above is necessary to achieve some of their applications, it is not necessary to achieve all desired applications. In this thesis, we explore several hypergraph problems using considerably simpler regularity tools. These tools, in a very direct sense, generalize the



original Szemerédi graph regularity lemma and graph counting lemma.

## 1.1 Tools

We begin discussing our work by stating a hypergraph regularity lemma for  $k$ -uniform hypergraphs which generalizes Szemerédi regularity lemma for graphs. To that end, we need the following concepts. For a  $k$ -uniform hypergraph  $\mathcal{H}$  on an  $n$  vertex set  $V$ , and  $\epsilon > 0$ , we call a  $k$ -tuple of nonempty pairwise disjoint sets  $X_1, X_2, \dots, X_k \subseteq V$   $\epsilon$ -regular if for all  $X'_i \subseteq X_i$ ,  $i \in [k]$ , satisfying  $|X'_i| \geq \epsilon|X_i|$ , we have  $|d_{\mathcal{H}}(X'_1, X'_2, \dots, X'_k) - d_{\mathcal{H}}(X_1, X_2, \dots, X_k)| \leq \epsilon$ , where

$$d_{\mathcal{H}}(X'_1, X'_2, \dots, X'_k) = \frac{|\mathcal{H}[X'_1, X'_2, \dots, X'_k]|}{|X'_1||X'_2| \dots |X'_k|}$$

is the *density* of the  $k$ -partite subhypergraph  $\mathcal{H}[X'_1, X'_2, \dots, X'_k]$  of  $\mathcal{H}$  (consisting of all hyperedges  $\{x_1, \dots, x_k\} \in \mathcal{H}$  with  $x_i \in X'_i$ ). Sometimes we write  $\|X'_1, X'_2, \dots, X'_k\|$  to denote  $\mathcal{H}[X'_1, X'_2, \dots, X'_k]$ . Further, we say a partition  $V(\mathcal{H}) = V_0 \cup V_1 \cup \dots \cup V_t$  is *equitable* if  $|V_1| = \dots = |V_t|$  and  $|V_0| < \epsilon n$ . Moreover, we say that the given partition is  *$t$ -equitable* if  $|V_1| = \dots = |V_t| = \lfloor \frac{n}{t} \rfloor$  and  $|V_0| < t$ . We now state the hypergraph version of Szemerédi's regularity lemma and give a rigorous proof in Chapter 2.

**Theorem 1.2 (Regularity Lemma)** *For every  $\epsilon > 0$ , every integer  $k \geq 2$  and every  $t_0 \geq 1$ , there exists an integer  $T_0$  s.t. every  $k$ -uniform hypergraph  $\mathcal{H}$  of order at least  $t_0$  admits an equitable and  $\epsilon$ -regular partition  $\{V_0, V_1, \dots, V_t\}$  with  $t_0 \leq t \leq T_0$ , meaning*

i)  $|V_1| = \dots = |V_t|$ ,  $|V_0| < \epsilon|V(\mathcal{H})|$ ;

ii) *all but at most  $\epsilon \binom{t}{k}$  of the  $k$ -tuples  $(V_{i_1}, V_{i_2}, \dots, V_{i_k})$ ,  $1 \leq i_1 < i_2 < \dots < i_k \leq t$ , are  $\epsilon$ -regular.*

In some of our applications, we need an algorithmic version of Theorem 1.2. Note, in particular, Theorem 1.2 guarantees the existence of an  $\epsilon$ -regular partition for a given hypergraph but does not show how to construct one. The following version of

Theorem 1.2, due to Czygrinow and Rödl, does just this. (For graphs, i.e.,  $k = 2$ , this was done by Alon et al. [2].)

**Theorem 1.3 (Czygrinow, Rödl [6])** *For all integers  $k \geq 2$  and  $t_0 \geq 1$  and all  $\epsilon > 0$ , there exists an integer  $T_0$  so that for every  $k$ -uniform hypergraph  $\mathcal{H}$  on  $n$  vertices, one may construct in time  $O(n^{2k-1} \log^2 n)$ , an  $\epsilon$ -regular and  $t$ -equitable partition  $V(\mathcal{H}) = V_1 \cup \dots \cup V_t$ ,  $t_0 \leq t \leq T_0$ , whose every  $\epsilon$ -irregular  $k$ -tuple  $V_{i_1}, \dots, V_{i_k}$ ,  $1 \leq i_1 < \dots < i_k \leq t$ , is identified.*

We now state a hypergraph counting lemma compatible with the regularity lemmas above, due to Kohayakawa, Nagle, Rödl and Schacht [16]. This counting lemma estimates the frequency of a fixed *simple*  $k$ -uniform hypergraph (to be defined momentarily)  $\mathcal{F}$  in a suitable environment provided by the regularity lemmas above. Before we state the counting lemma precisely, let us emphasize that this lemma applies *only* when the sub-hypergraphs  $\mathcal{F}$  are *simple* (or *linear*), that is, whose every pair of distinct  $k$ -tuples meet in at most one point. That said, the counting lemma takes place in the following environment.

Let  $\mathcal{F}$  be a simple  $k$ -uniform hypergraph on vertex set  $[f] = \{1, \dots, f\}$ . Let  $\mathcal{H}$  be an  $f$ -partite  $k$ -uniform hypergraph with vertex partition  $V = V(\mathcal{H}) = V_1 \cup \dots \cup V_f$ , where  $|V_1| = \dots = |V_f| = n$ , satisfying the following property: for all  $\{i_1, \dots, i_k\} \in \binom{[f]}{k}$ ,  $\mathcal{H}[V_{i_1}, \dots, V_{i_k}]$  is  $(d_F, \epsilon)$ -regular if  $F = \{i_1, \dots, i_k\} \in \mathcal{F}$ , and  $\mathcal{H} = \emptyset$  otherwise, where  $d_F > d_0$  for each  $F \in \mathcal{F}$ . In this context, we write  $\mathcal{F}(\mathcal{H})$  for the collection of copies of  $\mathcal{F}$  in  $\mathcal{H}$  which “cross” the partition  $V_1 \cup \dots \cup V_f$ , i.e., have a vertex in each class  $V_i$ ,  $1 \leq i \leq f$ .

**Theorem 1.4 (Counting Lemma for Simple  $k$ -Uniform Hypergraphs)** *For all integers  $f \geq k \geq 2$ , for all constants  $d_0, \gamma \in (0, 1]$ , and for all simple  $k$ -uniform hypergraphs  $\mathcal{F}$  on vertex set  $[f]$ , there exists  $\epsilon > 0$  so that whenever  $\mathcal{F}$  and hypergraph  $\mathcal{H}$  are as in the preceding setup with constants  $f, k, d_0, \epsilon, n$ , where  $n \geq n_0(f, k, d_0, \gamma, \epsilon)$  is sufficiently large,*

$$|\mathcal{F}(\mathcal{H})| = (1 \pm \gamma)n^f \prod_{F \in \mathcal{F}} d_F.$$

Theorem 1.4 is proven in Chapter 3. We also discuss a slight extension of Theorem 1.4 in Chapter 3. We remark that Theorem 1.4 does not hold whenever the subhypergraph  $\mathcal{F}$  is not simple (see [16] for details). We now proceed to our new results.

## 1.2 Results

For our first problem, let  $\mathbb{F} = \{\mathcal{F}_i\}_{i \in I}$  be a possibly infinite family of simple  $k$ -uniform hypergraphs. Let  $\text{Forb}(n, \mathbb{F})$  denote the collection of  $k$ -uniform hypergraphs  $\mathcal{H}$  on vertex set  $\{1, \dots, n\}$  which contain no copy of  $\mathcal{F} \in \mathbb{F}$  as a subhypergraph. Let  $\text{ex}(n, \mathbb{F}) = \max\{|\mathcal{H}| : \mathcal{H} \in \text{Forb}(n, \mathbb{F})\}$  denote the classical *Turán number* (almost none of which are known for  $k \geq 3$ ). Observe that  $\log_2 |\text{Forb}(n, \mathbb{F})| \geq \text{ex}(n, \mathbb{F})$  because all subhypergraphs of a maximal  $\mathcal{H} \in \text{Forb}(n, \mathbb{F})$  also belong to  $\text{Forb}(n, \mathbb{F})$ . We show that, in a sense, this lower bound is best possible.

**Theorem 1.5** *Let  $\mathbb{F} = \{\mathcal{F}_i\}_{i \in I}$  be a possibly infinite family of simple  $k$ -uniform hypergraphs. For all  $\delta > 0$ , there exists  $n_0 = n_0(\delta) \in \mathbb{N}$  so that for all  $n > n_0$ ,*

$$|\text{Forb}(n, \mathbb{F})| \leq 2^{\text{ex}(n, \mathbb{F}) + \delta n^k}.$$

Theorem 1.5, and various versions of it, were studied by a wide variety of authors [1, 3, 4, 8, 9, 15, 19, 21–25]. We mention that Theorem 1.5 holds even when  $\mathbb{F}$  consists of not necessarily simple hypergraphs  $\mathcal{F}$  (see [21]), but this proof is quite technical and relies on the hypergraph regularity techniques of [13, 14, 20, 27, 30] mentioned earlier. In the case where all  $\mathcal{F} \in \mathbb{F}$  are simple, we are able to give an easier proof, and do so in Chapter 4.

For our next result, we consider the problem of estimating the frequency  $|\mathcal{F}(\mathcal{G})|$  of a subhypergraph  $\mathcal{F}$  on  $f$  vertices in a host hypergraph  $\mathcal{G}$  on  $n$  vertices. (We slightly abuse the notation  $|\mathcal{F}(\mathcal{G})|$  here, since there is no partition to “cross”.) Clearly, one can compute  $|\mathcal{F}(\mathcal{G})|$  precisely in time  $O(n^f)$ . We show that when  $\mathcal{F}$  is simple,  $|\mathcal{F}(\mathcal{G})|$  may be accurately approximated in considerably shorter time.

**Theorem 1.6** *Let  $\mathcal{F}$  be a simple  $k$ -uniform hypergraph on  $f \geq k \geq 2$  vertices, and let  $\zeta > 0$  be given. There exists an algorithm which, for a given  $k$ -uniform hypergraph  $\mathcal{G}$  on  $n$  vertices, computes a value  $\Phi(\mathcal{G})$  in time  $O(n^{2k-1} \log^2 n)$  for which  $|\mathcal{F}(\mathcal{G}) - \Phi(\mathcal{G})| < \zeta n^f$ .*

Theorem 1.6 for  $k = 2$  was proven by Duke, Lefmann and Rödl [7]. We prove it for  $k \geq 2$  in Chapter 5.

For our final result, we consider a well-known theorem of Nagle, Rödl, Schacht, Skokan [20, 27] and Gowers [13, 14] known as the Removal Lemma. This theorem roughly asserts that if a ‘large’ hypergraph  $\mathcal{H}$  contains ‘few’ copies of a fixed subhypergraph  $\mathcal{F}$ , then one may delete ‘few’ edges from  $\mathcal{H}$  to destroy all these copies. This statement has surprising connections to various problems in combinatorial number theory and theoretical computer science (see [26] for a survey). The proof relies on the hypergraph regularity and counting lemmas of [13, 14, 20, 27] and is, therefore, quite technical. In the case that  $\mathcal{F}$  is simple, we give an easier proof which is, in fact, constructive, whereas the original more general proof was not.

**Theorem 1.7 (Constructive Removal Lemma)** *For all integers  $f \geq k \geq 2$ ,  $\gamma > 0$  and simple  $k$ -uniform hypergraphs  $\mathcal{F}$  on  $f$  vertices, there exists  $\delta > 0$  and integer  $n_0 = n_0(f, k, \gamma, \mathcal{F}, \delta)$  so that the following holds:*

*Given a  $k$ -uniform hypergraph  $\mathcal{H}$  on  $n \geq n_0$  vertices which contains fewer than  $\delta n^f$  copies of  $\mathcal{F}$ , one may delete, in time  $O(n^{2k-1} \log^2 n)$ , at most  $\gamma n^k$  edges from  $\mathcal{H}$  to make it  $\mathcal{F}$ -free.*

We prove theorem 1.7 in Chapter 6.

## 2 Regularity Lemma For $k$ -Uniform Hypergraphs

In this chapter we state and prove the Hypergraph Regularity Lemma. For definitions of technical terms and notation, refer to the discussion on Theorem 1.2 given in the introduction.

**Theorem 2.1** *For every  $\epsilon > 0$ , for all integers  $k$  and every  $t_0 \geq 1$ , there exists an integer  $T_0$  s.t. every  $k$ -uniform hypergraph  $\mathcal{H}$  of order at least  $t_0$  admits an  $\epsilon$ -regular partition  $\{V_0, V_1, \dots, V_t\}$  with  $t_0 \leq t \leq T_0$ .*

The proof of Theorem 2.1 follows the original argument of Szemerédi for simple graphs [28, 29]. In particular, if an equitable partition  $\{V_0, V_1, \dots, V_t\}$  of  $V = V(\mathcal{H})$  is not  $\epsilon$ -regular, then we shall refine the classes  $V_1, V_1, \dots, V_t$  to form a new equitable partition  $\{V'_0, V'_1, \dots, V'_t\}$  of  $V = V(\mathcal{H})$ , where  $t' \leq t4^{\binom{t-1}{k-1}}$  and where the latter partition is ‘closer’ to being  $\epsilon$ -regular than the former. More precisely, for pairwise disjoint sets  $X_1, X_2, \dots, X_k \subseteq V$ , we define a measure of regularity as follows:

$$\begin{aligned} q(X_1, X_2, \dots, X_k) &:= \frac{|X_1||X_2| \cdots |X_k|}{n^k} d^2(X_1, X_2, \dots, X_k) \\ &= \frac{\|X_1, X_2, \dots, X_k\|^2}{|X_1||X_2| \cdots |X_k| n^k}, \end{aligned} \tag{2.1}$$

call  $q(X_1, X_2, \dots, X_k)$  the *index* of the  $k$ -tuple  $(X_1, X_2, \dots, X_k)$ . For partitions  $\chi_i$  of  $X_i$ , let

$$q(\chi_1, \chi_2, \dots, \chi_k) = \sum \{q(Y'_1, Y'_2, \dots, Y'_k) : Y'_1 \in \chi_1, Y'_2 \in \chi_2, \dots, Y'_k \in \chi_k\}. \tag{2.2}$$

For a partition  $\mathcal{P} = \{C_1, C_2, \dots, C_t\}$  of  $V$ , let

$$q(\mathcal{P}) := \sum_{i_1 < i_2 < \dots < i_k} q(C_{i_1}, C_{i_2}, \dots, C_{i_k}). \quad (2.3)$$

However, if  $\mathcal{P} = \{C_0, C_1, \dots, C_t\}$  is a partition of  $V$  with exceptional set  $C_0$ , we treat  $C_0$  as a set of singletons and define:

$$q(\mathcal{P}) := q(\tilde{\mathcal{P}}), \quad (2.4)$$

where  $\tilde{\mathcal{P}} := \{C_1, C_2, \dots, C_t\} \cup \{\{v\} : v \in C_0\}$ .

We now approach the crux of the argument. Note that for an arbitrary partition  $\mathcal{P} = \{C_0, C_1, \dots, C_t\}$  of  $V$ ,  $q(\mathcal{P}) \leq 1$ . Indeed,

$$\begin{aligned} q(\mathcal{P}) &= \sum_{i_1 < i_2 < \dots < i_k} q(C_{i_1}, C_{i_2}, \dots, C_{i_k}) \\ &= \sum_{i_1 < i_2 < \dots < i_k} \frac{|C_{i_1}| |C_{i_2}| \dots |C_{i_k}|}{n^k} d^2(C_{i_1}, C_{i_2}, \dots, C_{i_k}) \\ &\leq \frac{1}{n^k} \sum_{i_1 < i_2 < \dots < i_k} |C_{i_1}| |C_{i_2}| \dots |C_{i_k}| \\ &\leq 1. \end{aligned}$$

On the other hand, if a partition  $\mathcal{P}$  is not  $\epsilon$ -regular, then, the following lemma shows, it can be refined to produce a new partition  $\mathcal{P}'$  with larger index.

**Lemma 2.2** *Let  $0 < \epsilon \leq \frac{1}{4}$ , and let  $\mathcal{P} = \{C_0, C_1, \dots, C_t\}$  be a partition of  $V$ , with exceptional set  $C_0$  of size  $|C_0| \leq \epsilon n$  and  $|C_1| = |C_2| = \dots = |C_t| =: c$ . If  $\mathcal{P}$  is not  $\epsilon$ -regular, then there is a partition  $\mathcal{P}' = \{C'_0, C'_1, \dots, C'_{t'}\}$  of  $V$  with exceptional set  $C'_0$ , where  $t \leq t' \leq t4^{\binom{t-1}{k-1}}$ , s.t.  $|C'_0| \leq |C_0| + \frac{n}{2^{\binom{t-1}{k-1}}}$ , all other sets have equal size, and,*

$$q(\mathcal{P}') \geq q(\mathcal{P}) + \frac{\epsilon^{k+3}}{2^k}.$$

It is clear that starting with an arbitrary partition  $\mathcal{P}$  (of size  $t_0$ ) of  $V$ , Lemma 2.2 can only be iterated at most  $2^k / \epsilon^{k+3}$  times. The proof of Theorem 2.1 now follows.

*Proof of Theorem 2.1.* Let  $\epsilon > 0$  and  $t_0 \geq 1$  be given, and without loss of generality, let  $\epsilon \leq 1/4$ . Let  $s := 2^k/\epsilon^{k+3}$ . This number  $s$  is an upper bound on the number of iterations of Lemma 2.2 that can be applied to a partition of a hypergraph before it becomes  $\epsilon$ -regular. Recall that  $q(\mathcal{P}) \leq 1$  for all partitions  $\mathcal{P}$ .

There is one formal requirement which a partition  $\{C_0, C_1, \dots, C_t\}$  with  $|C_1| = |C_2| = \dots = |C_t|$  has to satisfy before Lemma 2.2 can be (re-)applied, viz., the size  $|C_0|$  of its exceptional set must not exceed  $\epsilon n$ . With each iteration of the lemma, however, the size of the exceptional set can grow by up to  $n/2^{\binom{t-1}{k-1}}$ . We thus want to choose  $t$  large enough so that even  $s$  increments of  $n/2^{\binom{t-1}{k-1}}$  add up to at most  $\frac{1}{2}\epsilon n$ , and  $n$  large enough that, for any initial value of  $|C_0| < t$ , we have  $|C_0| \leq \frac{1}{2}\epsilon n$ .

So let  $t \geq t_0$  be large enough that  $2^{\binom{t-1}{k-1}} \geq s/\epsilon$ . Then  $s/2^{\binom{t-1}{k-1}+1} \leq \epsilon/2$ , and hence

$$t + \frac{s}{2^{\binom{t-1}{k-1}+1}}n \leq \epsilon n \tag{2.5}$$

whenever  $t/n \leq \epsilon/2$  i.e. for all  $n \geq 2t/\epsilon$ .

Let us now choose  $T_0$ . This should be an upper bound on the number of (non-exceptional) sets in our partition after up to  $s$  iterations of Lemma 2.2, where in each iteration this number may grow from its current value  $r$  to at most  $r4^{\binom{r-1}{k-1}}$ . So let  $f$  be the function  $x \mapsto x4^{\binom{x-1}{k-1}}$ , and take  $T_0 := \max\{f^s(t), 2t/\epsilon\}$ ; the second term in the maximum ensures that any  $n \geq T_0$  is large enough to satisfy (2.5).

Finally, we have to show that every hypergraph  $\mathcal{H}$  of order at least  $t_0$  has an  $\epsilon$ -regular partition  $\{V_0, V_1, \dots, V_t\}$  with  $t_0 \leq t \leq T_0$ . So, let  $\mathcal{H}$  be given and let  $n := |V|$ . If  $n \leq T_0$ , we partition  $\mathcal{H}$  into  $t := n$  singletons, choosing  $V_0 = \phi$  and  $|V_1| = \dots = |V_t| = 1$ . This is clearly  $\epsilon$ -regular.

Suppose now  $n > T_0$ , let  $C_0 \subseteq V$  be minimal such that  $t$  divides  $|V \setminus C_0|$ , and let  $\{C_1, \dots, C_t\}$  be any partition of  $V \setminus C_0$  into sets of equal size. Then  $|C_0| < t$  and hence  $|C_0| \leq \epsilon n$  by (2.5). Starting with  $\{C_0, C_1, \dots, C_t\}$ , we reapply Lemma 2.2 again and again, until the partition of  $\mathcal{H}$  obtained is  $\epsilon$ -regular; this will happen after at most  $s$  iterations, since by (2.5) the size of the exceptional set in the partitions stays below  $\epsilon n$  so the lemma could indeed be reapplied up to the theoretical maximum of  $s$  times. ■

In what follows, we prove two further lemmas which are eventually needed in the proof of Lemma 2.2. We begin by showing that when we refine a partition, the value of  $q$  will not decrease.

**Lemma 2.3** *i) Let  $X_1, X_2, \dots, X_k \subseteq V$  be disjoint. If  $\chi_i$  is a partition of  $X_i$ , then*

$$q(\chi_1, \chi_2, \dots, \chi_k) \geq q(X_1, X_2, \dots, X_k).$$

*ii) If  $\mathcal{P}, \mathcal{P}'$  are partitions of  $V$  and  $\mathcal{P}'$  refines  $\mathcal{P}$ , then  $q(\mathcal{P}') \geq q(\mathcal{P})$ .*

*Proof.* *i)* Let  $\chi_i =: \{Y_{i1}, Y_{i2}, \dots, Y_{il_i}\}$ , where  $Y_{ij} \subseteq X_i$ , for all  $i \in [k]$ . Then,

$$\begin{aligned} q(\chi_1, \chi_2, \dots, \chi_k) &= \sum_{\substack{j_i \in [l_i] \\ i \in [k]}} q(Y_{1j_1}, Y_{2j_2}, \dots, Y_{kj_k}) \\ &= \frac{1}{n^k} \sum_{\substack{j_i \in [l_i] \\ i \in [k]}} \frac{\|Y_{1j_1}, Y_{2j_2}, \dots, Y_{kj_k}\|^2}{|Y_{1j_1}| |Y_{2j_2}| \dots |Y_{kj_k}|} \\ &\geq_{(C.S)} \frac{1}{n^k} \frac{(\sum_{\substack{j_i \in [l_i] \\ i \in [k]}} \|Y_{1j_1}, Y_{2j_2}, \dots, Y_{kj_k}\|)^2}{\sum_{\substack{j_i \in [l_i] \\ i \in [k]}} |Y_{1j_1}| |Y_{2j_2}| \dots |Y_{kj_k}|} \\ &= \frac{1}{n^k} \frac{\|X_1, X_2, \dots, X_k\|^2}{(\sum_{j_1 \in [l_1]} |Y_{1j_1}|)(\sum_{j_2 \in [l_2]} |Y_{2j_2}|) \dots (\sum_{j_k \in [l_k]} |Y_{kj_k}|)} \\ &= \frac{1}{n^k} \frac{\|X_1, X_2, \dots, X_k\|^2}{|X_1| |X_2| \dots |X_k|} \\ &= q(X_1, X_2, \dots, X_k), \end{aligned}$$

where the inequality  $\geq_{(C.S)}$  follows from the Cauchy-Schwarz inequality.



ii) Let  $\mathcal{P} =: \{C_1, C_2, \dots, C_t\}$  and for  $i \in [t]$ , let  $\mathcal{C}_i$  be the partition of  $C_i$  induced by  $\mathcal{P}'$ . Then,

$$\begin{aligned} q(\mathcal{P}) &= \sum_{i_1 < i_2 < \dots < i_k} q(C_{i_1}, C_{i_2}, \dots, C_{i_k}) \\ &\leq^{(i)} \sum_{i_1 < i_2 < \dots < i_k} q(\mathcal{C}_{i_1}, \mathcal{C}_{i_2}, \dots, \mathcal{C}_{i_k}) \\ &\leq q(\mathcal{P}'). \end{aligned}$$

■

Next, we show that refining a partition by sub-partitioning an irregular  $k$ -tuple of partition sets increases the value of  $q$  a little; since we are dealing here with a single  $k$ -tuple only, the amount of this increase will still be less than any constant.

**Lemma 2.4** *Let  $\epsilon > 0$ , and let  $X_1, X_2, \dots, X_k \subseteq V$  be disjoint. If  $(X_1, X_2, \dots, X_k)$  is not  $\epsilon$ -regular, then there are partitions  $\chi_i = (Y_{i1}, Y_{i2})$  of  $X_i$  such that*

$$q(\chi_1, \chi_2, \dots, \chi_k) \geq q(X_1, X_2, \dots, X_k) + \epsilon^{k+2} \frac{|X_1||X_2| \dots |X_k|}{n^k}.$$

*Proof.* Suppose  $(X_1, X_2, \dots, X_k)$  is not  $\epsilon$ -regular. Then there are sets  $Y_{i1} \subseteq X_i$ , with:

$$|Y_{i1}| \geq \epsilon |X_i|$$

such that

$$|\eta| > \epsilon$$

for  $\eta := d(Y_{11}, Y_{21}, \dots, Y_{k1}) - d(X_1, X_2, \dots, X_k)$ . Let  $\chi_i := \{Y_{i1}, Y_{i2}\}$  where  $Y_{i2} := X_i \setminus Y_{i1}$ . We now show that  $\chi_1, \chi_2, \dots, \chi_k$  satisfy the conclusion of the lemma. For brevity, we shall write  $y_{ij} := |Y_{ij}|$ ;  $e_{i_1, \dots, i_k} := \|Y_{1i_1}, Y_{2i_2}, \dots, Y_{ki_k}\|$ ;  $x_i := |X_i|$  and  $e := \|X_1, X_2, \dots, X_k\|$ . As in the proof of Lemma 2.3:

$$\begin{aligned}
q(\chi_1, \chi_2, \dots, \chi_k) &= \frac{1}{n^k} \sum_{i_1, \dots, i_k \in \{1, 2\}} \frac{(e_{i_1, \dots, i_k})^2}{y_{1i_1} \cdot y_{2i_2} \cdots y_{ki_k}} \\
&= \frac{1}{n^k} \left( \frac{(e_{1, 1, \dots, 1})^2}{y_{11} \cdot y_{21} \cdots y_{k1}} + \sum_{\substack{i_1, \dots, i_k \in \{1, 2\} \\ (i_1, i_2, \dots, i_k) \neq (1, 1, \dots, 1)}} \frac{(e_{i_1, \dots, i_k})^2}{y_{1i_1} \cdot y_{2i_2} \cdots y_{ki_k}} \right) \\
&\geq_{(C.S)} \frac{1}{n^k} \left( \frac{(e_{1, 1, \dots, 1})^2}{y_{11} \cdot y_{21} \cdots y_{k1}} + \frac{(e - (e_{1, 1, \dots, 1}))^2}{(x_1 \cdot x_2 \cdots x_k) - (y_{11} \cdot y_{21} \cdots y_{k1})} \right),
\end{aligned}$$

where the last inequality follows from Cauchy-Schwarz inequality. By definition of  $\eta$ , we have  $e_{1, 1, \dots, 1} = \frac{(y_{11} \cdot y_{21} \cdots y_{k1}) \cdot e}{x_1 \cdot x_2 \cdots x_k} + \eta \cdot (y_{11} \cdot y_{21} \cdots y_{k1})$ . So,

$$\begin{aligned}
n^k q(\chi_1, \chi_2, \dots, \chi_k) &\geq \frac{1}{\prod_i y_{i1}} \left( \frac{(\prod_i y_{i1}) \cdot e}{\prod_i x_i} + \eta \cdot (\prod_i y_{i1}) \right)^2 + \\
&\quad \frac{1}{(\prod_i x_i) - (\prod_i y_{i1})} \left( \frac{(\prod_i x_i) - (\prod_i y_{i1})}{\prod_i x_i} e - \eta \cdot (\prod_i y_{i1}) \right)^2 \\
&= \frac{(\prod_i y_{i1}) \cdot e^2}{(\prod_i x_i)^2} + \frac{2e\eta \cdot (\prod_i y_{i1})}{\prod_i x_i} + \eta^2 \cdot (\prod_i y_{i1}) + \\
&\quad \frac{\prod_i x_i - \prod_i y_{i1}}{(\prod_i x_i)^2} e^2 - \frac{2e\eta \cdot (\prod_i y_{i1})}{\prod_i x_i} + \frac{\eta^2 \cdot (\prod_i y_{i1}^2)}{\prod_i x_i - \prod_i y_{i1}} \\
&\geq \frac{e^2}{\prod_i x_i} + \eta^2 \cdot \prod_i y_{i1} \\
&\geq_{|\eta| > \epsilon} \frac{e^2}{\prod_i x_i} + \epsilon^{k+2} \cdot \prod_i x_i
\end{aligned}$$

since  $y_{i1} \geq \epsilon x_i$  by the choice of  $Y_{i1}$ . ■

Finally, we show that if a partition has enough irregular  $k$ -tuples of partition sets to fall short of the definition of an  $\epsilon$ -regular partition, then sub-partitioning all those  $k$ -tuples at once results in an increase of  $q$  by a constant.

*Proof of Lemma 2.2.* For all  $1 \leq i_1 < i_2 < \dots < i_k \leq t$ , let us define a partition  $\mathcal{C}_{i_1, i_2, \dots, i_k}$  of  $C_i$  as follows:

If the  $k$ -tuple  $(C_{i_1}, C_{i_2}, \dots, C_{i_k})$  is  $\epsilon$ -regular, let  $\mathcal{C}_{i_1, i_2, \dots, i_k} = \{C_{i_1}\}$ . If not, then by Lemma 2.4, there are partitions  $\mathcal{C}_{i_1, i_2, \dots, i_k}, \mathcal{C}_{i_2, \dots, i_k, i_1}, \dots, \mathcal{C}_{i_k, i_1, \dots, i_{k-1}}$  of  $C_{i_1}, C_{i_2}, \dots, C_{i_k}$

respectively with  $|\mathcal{C}_{i_1, i_2, \dots, i_k}| = \dots = |\mathcal{C}_{i_k, i_1, \dots, i_{k-1}}| = 2$  and,

$$\begin{aligned} q(\mathcal{C}_{i_1, i_2, \dots, i_k}, \dots, \mathcal{C}_{i_k, i_1, \dots, i_{k-1}}) &\geq q(C_{i_1}, C_{i_2}, \dots, C_{i_k}) + \epsilon^{k+2} \frac{|\mathcal{C}_{i_1}| |\mathcal{C}_{i_2}| \dots |\mathcal{C}_{i_k}|}{n^k} \\ &= q(C_{i_1}, C_{i_2}, \dots, C_{i_k}) + \frac{\epsilon^{k+2} c^k}{n^k}. \end{aligned} \quad (2.6)$$

For each  $i = 1, \dots, t$ , let  $\mathcal{C}_i$  be the unique minimal partition of  $C_i$  that refines every partition  $\mathcal{C}_{i_1, \dots, i_k}$  with  $i_1 = i; i_1 \neq i_j, \forall j \neq 1$ . (In other words, if we consider two elements of  $C_i$  as equivalent whenever they lie in the same partition set of  $\mathcal{C}_{i_1, \dots, i_k}$  with  $i_1 = i; i_1 \neq i_j, \forall j \neq 1$ ; then  $\mathcal{C}_i$  is the set of equivalence classes). Thus,  $|\mathcal{C}_i| \leq 2^{\binom{t-1}{k-1}}$ . Now consider the following partition of  $V$ :

$$\mathcal{C} := C_0 \cup \bigcup_{i=1}^t \mathcal{C}_i$$

with  $C_0$  as exceptional set. Then  $\mathcal{C}$  refines  $\mathcal{P}$  and:

$$t \leq |\mathcal{C}| \leq t 2^{\binom{t-1}{k-1}}. \quad (2.7)$$

Let  $\mathcal{C}_0 := \{\{v\} : v \in C_0\}$ . Now, if  $\mathcal{P}$  is not  $\epsilon$ -regular, then for more than  $\epsilon \binom{t}{k}$  of the  $k$ -tuples  $(C_{i_1}, \dots, C_{i_k})$  with  $1 \leq i_1 < \dots < i_k \leq t$ , the partition  $\mathcal{C}_{i_1, \dots, i_k}$  is non-trivial.

Hence, by our definition of  $q$  for partitions with exceptional set, and Lemma 2.3 (i):

$$\begin{aligned}
q(\mathcal{C}) &= \sum_{1 \leq i_1 < i_2 < \dots < i_k} q(\mathcal{C}_{i_1}, \mathcal{C}_{i_2}, \dots, \mathcal{C}_{i_k}) + \\
&\quad \sum_{1 \leq i_1 < i_2 < \dots < i_{k-1}} q(\mathcal{C}_{i_0}, \mathcal{C}_{i_1}, \dots, \mathcal{C}_{i_{k-1}}) + \sum_{0 \leq i} q(\mathcal{C}_i) \\
&\geq \sum_{1 \leq i_1 < i_2 < \dots < i_k} q(\mathcal{C}_{i_1, \dots, i_k}, \mathcal{C}_{i_2, \dots, i_k, i_1}, \dots, \mathcal{C}_{i_k, i_1, \dots, i_{k-1}}) + \\
&\quad \sum_{1 \leq i_1 < i_2 < \dots < i_{k-1}} q(\mathcal{C}_0, \{C_{i_1}\}, \dots, \{C_{i_{k-1}}\}) + q(\mathcal{C}_0) \\
&\stackrel{(2.6)}{\geq} \sum_{1 \leq i_1 < i_2 < \dots < i_k} q(\mathcal{C}_{i_1}, \mathcal{C}_{i_2}, \dots, \mathcal{C}_{i_k}) + \frac{\epsilon^{k+2} c^k}{n^k} \cdot \epsilon \binom{t}{k} + \\
&\quad \sum_{1 \leq i_1 < i_2 < \dots < i_{k-1}} q(\mathcal{C}_0, \{C_{i_1}\}, \dots, \{C_{i_{k-1}}\}) + q(\mathcal{C}_0) \\
&= q(\mathcal{P}) + \epsilon^{k+3} \cdot \left(\frac{c}{n}\right)^k \binom{t}{k} \\
&\geq q(\mathcal{P}) + \frac{\epsilon^{k+3}}{2^k}.
\end{aligned}$$

For the last inequality, recall that  $|C_0| \leq \epsilon n \leq \frac{1}{4}n$ , so  $tc \geq \frac{3}{4}n$ .

In order to turn  $\mathcal{C}$  into our desired partition  $\mathcal{P}$ , all that remains to do is to cut its sets up into pieces of some common size, small enough that all remaining vertices can be collected into the exceptional set without making this too large. Let  $C'_1, C'_2, \dots, C'_{t'}$  be a maximal collection of disjoint sets of size  $d := \lfloor c/4 \binom{t-1}{k-1} \rfloor$  s.t. each  $C'_i$  is contained in some  $C \in \mathcal{C} \setminus \{C_0\}$  and put  $C'_0 := V \setminus \bigcup C'_i$ . Then  $\mathcal{P}' = \{C'_0, C'_1, \dots, C'_{t'}\}$  is indeed a partition of  $V$ . Moreover,  $\mathcal{P}'$  refines  $\mathcal{C}$ , so

$$q(\mathcal{P}') \geq q(\mathcal{C}) \geq q(\mathcal{P}) + \epsilon^{k+3}/2^k$$

by Lemma 2.3 (ii), and the result proved above.

Since each set  $C'_i \neq C'_0$  is also contained in one of the sets  $C_1, C_2, \dots, C_t$ , but no more than  $4 \binom{t-1}{k-1}$  sets  $C'_i$  can lie inside the same  $C_j$  (by the choice of  $d$ ), we also have  $t \leq t' \leq t 4 \binom{t-1}{k-1}$  as required. Finally, the sets  $C'_1, \dots, C'_{t'}$  use all but at most  $d$  vertices

from each set  $C \neq C_0$  of  $\mathcal{C}$ . Hence,

$$\begin{aligned} |C'_0| &\leq |C_0| + d|\mathcal{C}| \\ &\leq_{(2.7)} |C_0| + \frac{c}{4^{\binom{t-1}{k-1}}} \cdot t2^{\binom{t-1}{k-1}} \\ &= |C_0| + ct/2^{\binom{t-1}{k-1}} \\ &\leq |C_0| + n/2^{\binom{t-1}{k-1}}. \end{aligned}$$

■

### 3 Counting Lemma For Simple $k$ -Uniform Hypergraphs

In this chapter we state and prove the Counting Lemma for simple  $k$ -uniform hypergraphs. We ask the reader to recall the hypothesis for Counting Lemma from the discussion preceding Theorem 1.4 in the introduction.

#### 3.1 Main Theorem

**Theorem 3.1 (Counting Lemma for Simple  $k$ -Uniform Hypergraphs)** *For all integers  $f \geq k \geq 2$ , for all constants  $d, \gamma \in (0, 1]$ , and for all simple  $k$ -uniform hypergraphs  $\mathcal{F}$  on vertex set  $[f]$ , there exists  $\epsilon > 0$  so that whenever  $\mathcal{F}$  and hypergraph  $\mathcal{H}$  are as in the hypothesis of Theorem 1.4, with constants  $f, k, d, \epsilon, n$ , where  $n \geq n_0(f, d, k, \epsilon)$  is sufficiently large, then*

$$|\mathcal{F}(\mathcal{H})| = (1 \pm \gamma)d^{|\mathcal{F}|}n^f.$$

We say a few words about the proof. First, note that the Counting Lemma stated here promises both an upper and a lower bound:  $(1 - \gamma)n^f \prod d_F \leq |\mathcal{F}(\mathcal{H})| \leq (1 + \gamma)n^f \prod d_F$ . In what follows we show the lower bound only since the corresponding upper bound follows by symmetric arguments. Second, for notational simplicity, we shall assume that all  $d_F = d_0 = d$  (recall the hypothesis of Theorem 1.4). The proof allowing the densities  $d_F \geq d_0$ ,  $F \in \mathcal{F}$ , is philosophically the same. (Our proof of Theorem 3.1 is already fairly heavy in notation.)

*Proof.* Let integers  $f \geq k \geq 2$  and  $d = d_0, \gamma \in (0, 1]$  be given along with simple  $k$ -uniform hypergraph  $\mathcal{F}$  on vertex set  $[f]$ . To define  $\epsilon > 0$  and give our proof of 3.1, we induct on  $|\mathcal{F}|$ . If  $|\mathcal{F}| = 0$ , then any  $\epsilon > 0$  will do and the result is trivial.

Indeed, in this case  $|\mathcal{H}| = 0$  and therefore  $|\mathcal{F}(\mathcal{H})| = n^f \geq (1 - \gamma)d^0 n^f$ . If  $|\mathcal{F}| = 1$ , set  $\epsilon = d\gamma$  and the result is trivial. Indeed, in this case, suppose  $\mathcal{F} = \{i_1, \dots, i_k\} \in \binom{[f]}{k}$ . Then  $\mathcal{H} = \mathcal{H}[V_{i_1}, \dots, V_{i_k}]$ . By hypothesis,  $|\mathcal{H}| = (d \pm \epsilon)n^k$ , in which case  $|\mathcal{F}(\mathcal{H})| \geq (d - \epsilon)n^k n^{f-k} = (1 - \gamma)d^1 n^f$ .

Now, let  $\mathcal{F}$  be given as in the hypothesis of Theorem 1.4 with  $|\mathcal{F}| \geq 1$  edges. Delete any edge  $F_1 \in \mathcal{F}$  from  $\mathcal{F}$  (w.l.o.g,  $F_1 = \{f - k + 1, \dots, f\}$ ). Set  $\tilde{\gamma} = \frac{\gamma}{4}$  and let  $\epsilon_1 = \epsilon_{Thm.3.1}(f, k, d, \tilde{\gamma}, F \setminus F_1)$  be the constant guaranteed by the induction hypothesis. Set

$$\epsilon = \min\{\tilde{\gamma}d^{\binom{f}{k}}, \epsilon_1\} \quad (3.1)$$

We prove that with this choice of  $\epsilon > 0$ ,  $|\mathcal{F}(\mathcal{H})| \geq (1 - \gamma)d^{|\mathcal{F}|} n^f$ . We consider the following subhypergraphs of  $\mathcal{F}$ . Set  $\mathcal{F}^- = \mathcal{F} \setminus F_1$  and  $\mathcal{F}^* = \mathcal{F}[\{1, \dots, f - k\}]$ . Then  $\mathcal{F}^-$  is a simple  $k$ -uniform hypergraph on  $f$  vertices and  $|\mathcal{F}^-| - 1$  edges and  $\mathcal{F}^*$  is a simple  $k$ -uniform hypergraph on  $f - k$  vertices and at most  $|\mathcal{F}^-| - 1$  edges. Similarly, define  $\mathcal{H}^- = \mathcal{H} \setminus \mathcal{H}[V_{f-k+1}, \dots, V_f]$  and  $\mathcal{H}^* = \mathcal{H}[V_1, \dots, V_{f-k}]$ . Note that  $(\mathcal{F}^-, \mathcal{H}^-)$  and  $(\mathcal{F}^*, \mathcal{H}^*)$  are each as in hypothesis of Theorem 1.4.

We now define some related concepts. For  $\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)$  and  $\mathcal{F}_0^- \in \mathcal{F}^-(\mathcal{H}^-)$  ( $\mathcal{F}_0 \in \mathcal{F}(\mathcal{H})$ ), we say that  $\mathcal{F}_0^-$ , resp.  $\mathcal{F}_0$ , *extends*  $\mathcal{F}_0^*$  if  $\mathcal{F}_0^* \subseteq \mathcal{F}_0^-$ , resp.  $\mathcal{F}_0^* \subseteq \mathcal{F}_0$ . Let  $\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)$ ,  $\mathcal{F}_0^- \in \mathcal{F}^-(\mathcal{H}^-)$  and  $\mathcal{F}_0 \in \mathcal{F}(\mathcal{H})$  be given. For  $\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)$ , let

$$\begin{aligned} \text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*) &= \{\mathcal{F}_0^- \in \mathcal{F}^-(\mathcal{H}^-) : \mathcal{F}_0^- \text{ extends } \mathcal{F}_0^*\}, \\ \text{ext}_{\mathcal{F}}(\mathcal{F}_0^*) &= \{\mathcal{F}_0 \in \mathcal{F}(\mathcal{H}) : \mathcal{F}_0 \text{ extends } \mathcal{F}_0^*\}. \end{aligned}$$

Observe that:

$$\begin{aligned} |\mathcal{F}(\mathcal{H})| &= \sum_{\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)} |\text{ext}_{\mathcal{F}}(\mathcal{F}_0^*)|, \\ |\mathcal{F}^-(\mathcal{H}^-)| &= \sum_{\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)} |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)|. \end{aligned} \quad (3.2)$$

We now provide evaluations of  $\text{ext}_{\mathcal{F}}(\mathcal{F}_0^*)$  and  $\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)$  for a fixed term  $\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)$ . To that end, fix  $\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)$ . Set:

$$V_{\mathcal{F}_0^*} = \bigcup \{ \mathcal{F}_0^- : \mathcal{F}_0^- \in \text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*) \}. \quad (3.3)$$

Note that  $V_{\mathcal{F}_0^*} \subseteq V$ . For each  $f - k + 1 \leq i \leq f$ , set:

$$V_i^{\mathcal{F}_0^*} = V_i \cap V_{\mathcal{F}_0^*}.$$

We now make the following claim.

**Claim 3.2** *For  $\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)$  fixed,*

$$\begin{aligned} |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| &= |V_{f-k+1}^{\mathcal{F}_0^*}| \cdots |V_f^{\mathcal{F}_0^*}|, \\ |\text{ext}_{\mathcal{F}}(\mathcal{F}_0^*)| &= |\mathcal{H}[V_{f-k+1}^{\mathcal{F}_0^*}, \dots, V_f^{\mathcal{F}_0^*}]|. \end{aligned}$$

We defer the proof of Claim 3.2 to the end of this section in favor of continuing the proof of Theorem 3.1. Call  $\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)$  *big* if  $|\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| > \epsilon n^k$  and *small* otherwise. Write  $\mathcal{F}_{big}^*(\mathcal{H}^*)$  ( $\mathcal{F}_{small}^*(\mathcal{H}^*)$ ) for the collection of all *big* (*small*) elements  $\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)$ . It follows from Claim 3.2 that for each  $\mathcal{F}_0^* \in \mathcal{F}_{big}^*(\mathcal{H}^*)$ , we have:

$$|\text{ext}_{\mathcal{F}}(\mathcal{F}_0^*)| = (d \pm \epsilon) |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)|. \quad (3.4)$$

Indeed, by the  $(d, \epsilon)$ -regularity of  $\mathcal{H}[V_{f-k+1}^{\mathcal{F}_0^*}, \dots, V_f^{\mathcal{F}_0^*}]$ ,

$$\begin{aligned} |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| &= |V_{f-k+1}^{\mathcal{F}_0^*}| \cdots |V_f^{\mathcal{F}_0^*}| > \epsilon n^k \\ &\Rightarrow |V_{f-k+1}^{\mathcal{F}_0^*}| \cdots |V_f^{\mathcal{F}_0^*}| > \epsilon n \\ &\Rightarrow |\text{ext}_{\mathcal{F}}(\mathcal{F}_0^*)| = |\mathcal{H}[V_{f-k+1}^{\mathcal{F}_0^*}, \dots, V_f^{\mathcal{F}_0^*}]| \\ &= (d \pm \epsilon) |V_{f-k+1}^{\mathcal{F}_0^*}| \cdots |V_f^{\mathcal{F}_0^*}| \\ &= (d \pm \epsilon) \text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*). \end{aligned} \quad (3.5)$$

We may now conclude the proof of Theorem 3.1 using the following claim.



**Claim 3.3**

$$\sum_{\mathcal{F}_0^* \in \mathcal{F}_{big}^*(\mathcal{H}^*)} |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| \geq (1 - 2\tilde{\gamma})d^{|\mathcal{F}^-| - 1}n^f.$$

Indeed, by (3.2), (3.4) and Claim 3.3, we have:

$$\begin{aligned} |\mathcal{F}(\mathcal{H})| &\stackrel{(3.2)}{\geq} \sum_{\mathcal{F}_0^* \in \mathcal{F}_{big}^*(\mathcal{H}^*)} |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| \\ &\stackrel{(3.4)}{\geq} \sum_{\mathcal{F}_0^* \in \mathcal{F}_{big}^*(\mathcal{H}^*)} (d - \epsilon) |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| \\ &\geq \text{Claim.3.3} (d - \epsilon)(1 - 2\tilde{\gamma})d^{|\mathcal{F}^-| - 1}n^f \\ &\geq (1 - 4\tilde{\gamma})d^{|\mathcal{F}^-|}n^f. \end{aligned}$$

It remains to prove Claims 3.2 and 3.3.

*Proof of Claim 3.2.* We first establish the first identity. That

$$|\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| \leq |V_{f-k+1}^{\mathcal{F}_0^*}| \dots |V_f^{\mathcal{F}_0^*}|$$

is clear, so we establish the lower bound. To that end, let  $v_{f-k+1} \in V_{f-k+1}^{\mathcal{F}_0^*}, \dots, v_f \in V_f^{\mathcal{F}_0^*}$  be given. We claim that  $\mathcal{F}_0^* \cup \{v_{f-k+1}, \dots, v_f\} \in \mathcal{F}^-(\mathcal{H}^-)$ . Assume, on the contrary, that  $\mathcal{F}_0^* \cup \{v_{f-k+1}, \dots, v_f\} \notin \mathcal{F}^-(\mathcal{H}^-)$  and write, for simplicity,  $\mathcal{F}_0^* = \{v_1, \dots, v_{f-k}\}$ , where  $v_1 \in V_1, \dots, v_{f-k} \in V_{f-k}$ . Since  $\{v_1, \dots, v_f\} \in \mathcal{F}^-(\mathcal{H}^-)$ , there exists  $F_2 = \{i_1, \dots, i_k\} \in \mathcal{F}^-$ ,  $1 \leq i_1 < \dots < i_k \leq f$ , so that  $\{v_{i_1}, \dots, v_{i_k}\} \notin \mathcal{H}^-$ . Recall  $F_1 = \{f-k+1, \dots, f\}$  and observe that  $|F_1 \cap F_2| = 1$ . Indeed, since  $\mathcal{F}$  is simple and  $F_1, F_2 \in \mathcal{F}$ ,  $|F_1 \cap F_2| \leq 1$ . If  $F_1 \cap F_2 = \emptyset$ , then  $F_2 \in \mathcal{F}^*$ , and since  $\mathcal{F}_0^* \in \mathcal{F}^*(\mathcal{H}^*)$ , we'd have  $\{v_{i_1}, \dots, v_{i_k}\} \in \mathcal{H}^* \subseteq \mathcal{H}^-$ . Now,  $F_1 \cap F_2 = \{v_{i_k}\}$  where  $1 \leq i_1 < \dots < i_{k-1} < f-k+1 \leq i_k \leq f$ . But  $v_{i_k} \in V_{i_k}^{\mathcal{F}_0^*}$  which means there exist  $u_{f-k+1} \in V_{f-k+1}^{\mathcal{F}_0^*}, \dots, u_{i_k-1} \in V_{i_k-1}^{\mathcal{F}_0^*}, u_{i_k+1} \in V_{i_k+1}^{\mathcal{F}_0^*}, \dots, u_f \in V_f^{\mathcal{F}_0^*}$  so that  $\mathcal{F}_0^* \cup \{u_{f-k+1}, \dots, u_{i_k-1}, v_{i_k}, u_{i_k+1}, \dots, u_f\} \in \mathcal{F}^-(\mathcal{H}^-)$ . In this case,  $\{i_1, \dots, i_{k-1}, i_k\} \in \mathcal{F}^-$  implies  $\{v_{i_1}, \dots, v_{i_{k-1}}, v_{i_k}\} \in \mathcal{H}^-$ , a contradiction!

The proof of the second equality now easily follows. We showed that for each  $v_{f-k+1} \in V_{f-k+1}^{\mathcal{F}_0^*}, \dots, v_f \in V_f^{\mathcal{F}_0^*}$ , we have  $\mathcal{F}_0^* \cup \{v_{f-k+1}, \dots, v_f\} \in \mathcal{F}^-(\mathcal{H}^-)$ . As such,

each  $\{v_{f-k+1}, \dots, v_f\} \in \mathcal{H}[V_{f-k+1}^{\mathcal{F}_0^*}, \dots, V_f^{\mathcal{F}_0^*}]$  satisfies  $\mathcal{F}_0^* \cup \{v_{f-k+1}, \dots, v_f\} \in \mathcal{F}(\mathcal{H})$ . This proves Claim 3.2. ■

*Proof of Claim 3.3.* By our induction hypothesis,

$$|\mathcal{F}^-(\mathcal{H}^-)| \geq (1 - \tilde{\gamma})d^{|\mathcal{F}^-| - 1}n^f. \quad (3.6)$$

On the other hand, by (3.2), we have

$$\begin{aligned} |\mathcal{F}^-(\mathcal{H}^-)| &= \sum_{\mathcal{F}_0^* \in \mathcal{F}_{big}^*(\mathcal{H}^*)} |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| + \sum_{\mathcal{F}_0^* \in \mathcal{F}_{small}^*(\mathcal{H}^*)} |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| \\ &\leq \sum_{\mathcal{F}_0^* \in \mathcal{F}_{big}^*(\mathcal{H}^*)} |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| + |\mathcal{F}_{small}^*(\mathcal{H}^*)| \cdot \epsilon n^k \\ &\leq \sum_{\mathcal{F}_0^* \in \mathcal{F}_{big}^*(\mathcal{H}^*)} |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| + \epsilon n^f. \end{aligned}$$

Returning to (3), we see:

$$\begin{aligned} \sum_{\mathcal{F}_0^* \in \mathcal{F}_{big}^*(\mathcal{H}^*)} |\text{ext}_{\mathcal{F}^-}(\mathcal{F}_0^*)| &\geq (1 - \tilde{\gamma})d^{|\mathcal{F}^-| - 1}n^f - \epsilon n^f \\ &\geq (1 - 2\tilde{\gamma})d^{|\mathcal{F}^-| - 1}n^f \end{aligned}$$

■

And this completes our proof of the Counting Lemma. ■

### 3.2 A Generalization

The proof of Theorem 3.1 can be easily modified to prove a slight generalization of the Counting Lemma. We use this generalization in Chapter 6, and so we state it now.

#### Setup

(Extended Counting Environment) Let  $d_0, \epsilon > 0$  be given and let  $\mathcal{F}$  be a  $p$ -partite

simple  $k$ -uniform hypergraph with vertex partition  $V(\mathcal{F}) = W_1 \cup \dots \cup W_p$ . Let  $\mathcal{H}$  be a  $p$ -partite  $k$ -uniform hypergraph with vertex partition  $V(\mathcal{H}) = V_1 \cup \dots \cup V_p$ ,  $|V_1| = \dots = |V_p| = n$ , satisfying the following property: for all  $K = \{i_1, \dots, i_k\} \in \binom{[p]}{k}$ ,  $\mathcal{H}[V_{i_1}, \dots, V_{i_k}]$  is  $(d_K, \epsilon)$ -regular, where  $d_K \geq d_0$ , if  $\mathcal{F}$  has an edge crossing  $W_{i_1} \cup \dots \cup W_{i_k}$ , and  $\mathcal{H}[V_{i_1}, \dots, V_{i_k}] = \emptyset$  otherwise.

For  $\mathcal{F}$  and  $\mathcal{H}$  as given above, vertices  $v_1, \dots, v_f \subseteq V(\mathcal{H})$  are said to span a *partite-isomorphic* copy of  $\mathcal{F}$  in  $\mathcal{H}$  if there exists a bijection  $\psi : V(\mathcal{F}) \rightarrow \{v_1, \dots, v_f\}$  where  $v_j \in V_i$ ,  $1 \leq j \leq f$ ,  $1 \leq i \leq p$ , if and only if  $\psi^{-1}(v_j) \in W_i$  and where, for each  $F \in \mathcal{F}$ ,  $\psi(F) \in \mathcal{H}$ . In this context, we shall write  $\mathcal{F}(\mathcal{H})$  (again abusing notation) for the set of all partite-isomorphic copies of  $\mathcal{F}$  in  $\mathcal{H}$ .

**Theorem 3.4 (Extended Counting Lemma)** *For all integers  $f \geq p \geq k \geq 2$ , constants  $d_0, \gamma \in (0, 1]$ , and all  $p$ -partite simple hypergraphs  $\mathcal{F}$  on  $f$  vertices, there exists  $\epsilon > 0$  so that whenever  $\mathcal{F}$  and  $k$ -uniform hypergraph  $\mathcal{H}$  are as in the preceding Setup with these constants and  $n$  sufficiently large, then*

$$|\mathcal{F}(\mathcal{H})| = (1 \pm \gamma)n^f \cdot \prod_{F \in \mathcal{F}} \left\{ d_{K(F)} : K(F) = \{i_1, \dots, i_k\} \in \binom{[p]}{k} \text{ s.t. } F \subseteq W_{i_1} \cup \dots \cup W_{i_k} \right\}.$$

## 4 Bounding $|\text{Forb}(n, \mathbb{F})|$

We are now ready to present a proof for our first application of the hypergraph regularity method. Recall the discussion preceding Theorem 1.5 in the introduction for the semantics of technical notation.

**Theorem 4.1** *Let  $\mathbb{F} = \{\mathcal{F}_i\}_{i \in I}$  be a possibly infinite family of simple  $k$ -uniform hypergraphs. For all  $\delta > 0$ , there exists  $n_0 = n_0(\delta) \in \mathbb{N}$  so that for all  $n > n_0$ ,*

$$|\text{Forb}(n, \mathbb{F})| \leq 2^{\text{ex}(n, \mathbb{F}) + \delta n^k}.$$

*Proof.* Let family  $\mathbb{F} = \{\mathcal{F}_i\}_{i \in I}$  and  $\delta > 0$  be given as in Theorem 4.1. Our proof of Theorem 4.1 begins with a description of auxiliary constants we use in the proof. First, let  $d_0 \in (0, 1)$  be small enough so that

$$20 d_0 \log \frac{e}{4d_0} < \delta. \tag{4.1}$$

It is well known that the sequence  $((\binom{s}{k})^{-1} \text{ex}(s, \mathbb{F}))_{s=1}^{\infty}$  is non-increasing, and therefore, the limits

$$\lim_{s \rightarrow \infty} \overline{\text{ex}}(s, \mathbb{F}) = \lim_{s \rightarrow \infty} \frac{\text{ex}(s, \mathbb{F})}{\binom{s}{k}} \text{ and } \lim_{s \rightarrow \infty} \tilde{\text{ex}}(s, \mathbb{F}) = \lim_{s \rightarrow \infty} \frac{\text{ex}(s, \mathbb{F})}{s^k} \tag{4.2}$$

exist. Let  $s_0 \in \mathbb{N}$  be large enough, so that for all  $s_1, s_2 \geq s_0$ ,

$$\max\{|\overline{\text{ex}}(s_1, \mathbb{F}) - \overline{\text{ex}}(s_2, \mathbb{F})|, |\tilde{\text{ex}}(s_1, \mathbb{F}) - \tilde{\text{ex}}(s_2, \mathbb{F})|\} < \frac{\delta}{20}, \tag{4.3}$$

and let

$$t_0 = \max\left\{s_0, \frac{1}{d_0}\right\}. \quad (4.4)$$

Finally, for a simple  $k$ -uniform hypergraph  $\mathcal{F}$  on  $s_0$  vertices, let  $\epsilon_{\mathcal{F}} = \epsilon_{Thm.3.1}(\mathcal{F}, d_0, \gamma)$  with  $\gamma = 1/2$ , be the positive constant guaranteed by Theorem 3.1, the Counting Lemma for Simple Hypergraphs. Set,

$$\epsilon = \min_{\mathcal{F}}\{\epsilon_{\mathcal{F}}\} \leq d_0, \quad (4.5)$$

where the minimum is taken over all simple  $k$ -uniform hypergraphs  $\mathcal{F}$  on  $s_0$  vertices. (Note that  $\epsilon \leq d_0$  follows from Theorem 3.1.) In all that follows we take the integer  $n$  sufficiently large whenever needed. This concludes our discussion of the constants.

To each  $\mathcal{G} \in \text{Forb}(n, \mathbb{F})$ , associate an  $\epsilon$ -regular  $t_{\mathcal{G}}$ -equitable partition  $\mathbb{P}_{\mathcal{G}} : V(\mathcal{G}) = [n] = V_0 \cup V_1 \cup \dots \cup V_{t_{\mathcal{G}}}$ ,  $t_0 \leq t_{\mathcal{G}} \leq T_0$ . Every  $\mathcal{G} \in \text{Forb}(n, \mathbb{F})$  admits at least one such partition by Theorem 2.1, and if some  $\mathcal{G} \in \text{Forb}(n, \mathbb{F})$  admits multiple such partitions, we choose one arbitrarily. Define an equivalence relation  $\sim$  on  $\text{Forb}(n, \mathbb{F})$  as follows. For each  $\mathcal{G}_1, \mathcal{G}_2 \in \text{Forb}(n, \mathbb{F})$ ,

$$\mathcal{G}_1 \sim \mathcal{G}_2 \iff \mathbb{P}_{\mathcal{G}_1} = \mathbb{P}_{\mathcal{G}_2}.$$

Let  $\Pi = \Pi_1 \cup \dots \cup \Pi_N$  be the corresponding partition of  $\text{Forb}(n, \mathbb{F})$ . Note that:

$$|\Pi| = N \leq \sum_{i=t_0}^{T_0} \binom{n}{\lfloor n/i \rfloor}^i = 2^{O(n)}. \quad (4.6)$$

Now fix an equivalence class  $\Pi_j$ ,  $1 \leq j \leq N$ , i.e. fix a  $t$ -equitable vertex partition  $\mathbb{P} : [n] = V_0 \cup V_1 \cup \dots \cup V_t$ , where  $t_0 \leq t \leq T_0$ , with respect to which every  $\mathcal{G} \in \Pi_j$  is  $\epsilon$ -regular. We now partition  $\Pi_j$  as follows. For a function  $\phi \in \{0, 1\}^{\binom{[t]}{k}}$ , let

$$\begin{aligned} \Pi_{j,\phi} = \left\{ \mathcal{G} \in \Pi_j : \forall \{i_1, \dots, i_k\} \in \binom{[t]}{k}, \right. \\ \left. \mathcal{G}[V_{i_1}, \dots, V_{i_k}] \text{ is } \epsilon\text{-reg and } d_{\mathcal{G}}(V_{i_1}, \dots, V_{i_k}) \geq d_0 \iff \phi(\{i_1, \dots, i_k\}) = 1 \right\}. \end{aligned} \quad (4.7)$$

Then  $\Pi_j = \bigcup \{\Pi_{j,\phi} : \phi \in \{0, 1\}^{\binom{[t]}{k}}\}$  is a partition with at most

$$2^{\binom{t}{k}} \leq 2^{T_0^k} = 2^{O(1)} \quad (4.8)$$

parts. The proof of Theorem 1 rests on the following proposition.

**Proposition 4.2** *For  $1 \leq j \leq N$  and  $\phi \in \{0, 1\}^{\binom{[t]}{k}}$  fixed,  $|\Pi_{j,\phi}| \leq 2^{\text{ex}(n, \mathbb{F}) + \frac{\delta}{2}n^k}$ .*

Indeed by Proposition 4.2 and (4.6) - (4.8), we have:

$$\begin{aligned} |\text{Forb}(n, \mathbb{F})| &= \sum_{j=1}^N |\Pi_j| = \sum_{j=1}^N \sum \{|\Pi_{j,\phi}| : \phi \in \{0, 1\}^{\binom{[t]}{k}}\}, \\ &2^{\text{ex}(n, \mathbb{F}) + \frac{\delta}{2}n^k + O(n) + O(1)} \leq 2^{\text{ex}(n, \mathbb{F}) + \delta n^k}. \end{aligned} \quad (4.9)$$

It remains to prove Proposition 4.2.

*Proof of Proposition 4.2.* Fix  $1 \leq j \leq N$  and  $\phi \in \{0, 1\}^{\binom{[t]}{k}}$ . Note that every  $\mathcal{G} \in \Pi_{j,\phi}$  can be written as the union,

$$\mathcal{G} = \mathcal{G}^{V_0} \cup \mathcal{G}^{V_1} \cup \dots \cup \mathcal{G}^{V_t} \cup \mathcal{G}_0 \cup \mathcal{G}_1$$

where  $\mathcal{G}^{V_0} = \{K \in \mathcal{G} : K \cap V_0 \neq \emptyset\}$ ,  $\mathcal{G}^{V_a} = \{K \in \mathcal{G} : |K \cap V_a| \geq 2\}$  for  $1 \leq a \leq t$ , and for  $i = 0, 1$ ,

$$\mathcal{G}_i = \bigcup \{\mathcal{G}[V_{i_1}, \dots, V_{i_k}] : \{i_1, \dots, i_k\} \in \phi^{-1}(i)\}. \quad (4.10)$$

Since  $\mathbb{P}$  is a  $t$ -equitable partition (shared by all of  $\Pi_{j,\phi}$ ),

$$\begin{aligned} |\mathcal{G}^{V_0} \cup \mathcal{G}^{V_1} \cup \dots \cup \mathcal{G}^{V_t}| &\leq |V_0|n^{k-1} + t \binom{\lfloor n/t \rfloor}{2} n^{k-2} \\ &\leq tn^{k-1} + \frac{1}{2t}n^k \\ &\leq \left(\frac{T_0}{n} + \frac{1}{2t_0}\right)n^k \\ &\leq (o(1) + \frac{1}{2t_0})n^k. \end{aligned} \quad (4.11)$$

As well, we have

$$|\mathcal{G}_0| \leq (\epsilon + d_0)n^k. \quad (4.12)$$

Indeed, the  $\epsilon$ -regularity of  $\mathbb{P}$  ensures at most  $\epsilon t^k$   $k$ -tuples  $\{i_1, \dots, i_k\} \in \phi^{-1}(0)$  could have  $\mathcal{G}[V_{i_1}, \dots, V_{i_k}]$  being  $\epsilon$ -irregular (giving rise to at most  $\epsilon t^k \lfloor n/t \rfloor^k$   $k$ -tuples  $K \in \mathcal{G}$ ). Otherwise, when  $d_{\mathcal{G}}(V_{i_1}, \dots, V_{i_k}) \leq d_0$ , we have  $|\mathcal{G}[V_{i_1}, \dots, V_{i_k}]| \leq d_0 \lfloor n/t \rfloor^k$  (over at most  $\binom{t}{k}$   $k$ -tuples  $\{i_1, \dots, i_k\} \in \phi^{-1}(0)$ ). In other words, combining (4.11) and (4.12), we may write every  $\mathcal{G} \in \Pi_{j,\phi}$  as a disjoint union:

$$\mathcal{G} = \mathcal{G}_* \cup \mathcal{G}_1, \quad (4.13)$$

where,

$$\begin{aligned} |\mathcal{G}_*| &\leq (o(1) + \frac{1}{2t_0} + \epsilon + d_0)n^k \\ &\leq 4d_0n^k. \end{aligned} \quad (4.14)$$

and  $\mathcal{G}_1$  is given in (4.10). We now use the decomposition in (4.13) to count all of  $|\Pi_{j,\phi}|$ . Indeed note that there are at most

$$\begin{aligned} \sum_{i=0}^{4d_0n^k} \binom{\binom{n}{k}}{i} &\leq n^k \binom{n^k}{4d_0n^k} \leq n^k \left(\frac{e}{4d_0}\right)^{4d_0n^k} \\ &= 2^{4d_0n^k \log \frac{e}{4d_0} + k \log n} \\ &\leq 2^{5d_0n^k \log \frac{e}{4d_0}} \\ &\stackrel{(4.1)}{\leq} 2^{\frac{\delta}{4}n^k} \end{aligned} \quad (4.15)$$

$k$ -graphs  $\mathcal{G}_*$  of the form in (4.13) and (4.14). Similarly, there are at most

$$2^{\lfloor n/t \rfloor^k |\phi^{-1}(1)|} \leq 2^{(n/t)^k |\phi^{-1}(1)|} \quad (4.16)$$

$k$ -graphs of the form in (4.13). We use the following claim.

**Claim 4.3**

$$|\phi^{-1}(1)| \leq \text{ex}(t, \mathbb{F}) + \frac{\delta}{5} \binom{t}{k}.$$

Using (4.13),(4.14),(4.15) and Claim 4.3, we see that (cf. (4.2))

$$\begin{aligned} \log_2 |\Pi_{j,\phi}| &\leq \frac{n^k}{t} (\text{ex}(t, \mathbb{F}) + \frac{\delta}{5} t^k) + \frac{\delta}{4} n^k \\ &= n^k \bar{\text{ex}}(t, \mathbb{F}) + \frac{9}{20} \delta n^k \\ &\stackrel{(4.3)}{\leq} n^k (\bar{\text{ex}}(n, \mathbb{F}) + \frac{1}{20} \delta) + \frac{9}{20} \delta n^k \\ &= \text{ex}(n, \mathbb{F}) + \frac{\delta}{2} n^k, \end{aligned} \tag{4.17}$$

as promised by Proposition 4.2. It remain to prove Claim 4.3. ■

*Proof of Claim 4.3.* Assume, on the contrary, that

$$|\phi^{-1}(1)| > \text{ex}(t, \mathbb{F}) + \frac{\delta}{5} \binom{t}{k}. \tag{4.18}$$

For consistency of notation, we shall write the  $k$ -uniform hypergraph  $\phi^{-1}(1)$  on vertex set  $[t]$  as  $\mathcal{J}$ , where our assumption above is that  $|\mathcal{J}| > \text{ex}(t, \mathbb{F}) + \frac{\delta}{5} \binom{t}{k}$ . Now with  $s_0 \stackrel{(4.3)}{\leq} t_0 \leq t$  given in (4.2), note that there must also exist  $S_0 \in \binom{[t]}{s_0}$  for which  $|\mathcal{J}[S_0]| \geq \text{ex}(s_0, \mathbb{F}) + 1$ . For if not we would have

$$\begin{aligned} (\text{ex}(t, \mathbb{F}) + \frac{\delta}{5} \binom{t}{k}) \binom{t-k}{s_0-k} &< |\mathcal{J}| \binom{t-k}{s_0-k} \\ &\leq \binom{t}{s_0} \text{ex}(s, \mathbb{F}), \end{aligned} \tag{4.19}$$

or equivalently,

$$\bar{\text{ex}}(t, \mathbb{F}) + \frac{\delta}{5} < \bar{\text{ex}}(s_0, \mathbb{F}), \tag{4.20}$$

contradicting (4.3). Now, fix  $S_0 \in \binom{[t]}{s_0}$  and, for simplicity of notation (and w.l.o.g), suppose  $S_0 = \{1, \dots, s_0\}$ . The hypergraph  $\mathcal{J}[S_0]$  has more than  $\text{ex}(s_0, \mathbb{F})$  many edges, and therefore, must contain a copy of some  $\mathcal{F}_0 \in \mathbb{F}$ . We show the same copy must



also appear in every  $\mathcal{G} \in \Pi_{j,\phi} \subseteq \text{Forb}(n, \mathbb{F})$ , a clear contradiction, establishing that (4.18) was false.

Indeed, fix  $\mathcal{G}_0 \in \Pi_{j,\phi}$  and consider  $\mathcal{G}_0[V_1, \dots, V_{s_0}]$ . By definition of  $\phi$ , we have, for each  $\{i_1, \dots, i_k\} \in \mathcal{F}_0 \subseteq \mathcal{J} = \phi^{-1}(1)$ , that  $\mathcal{G}_0[V_{i_1}, \dots, V_{i_k}]$  is  $\epsilon$ -regular with density  $d_{\mathcal{G}_0}[V_{i_1}, \dots, V_{i_k}] \geq d_0$ . By our choice of  $\epsilon > 0$  in (4.5), Theorem 3.1 (Counting Lemma) implies:

$$\begin{aligned} |\mathcal{F}_0(\mathcal{G}_0)| &\geq \left(1 - \frac{1}{2}\right) d_0^{|\mathcal{F}_0|} \left(\frac{n}{t}\right)^{s_0} \\ &\geq \frac{1}{2} d_0^{\binom{s_0}{k}} \frac{n^{s_0}}{T_0^{s_0}} \\ &= \Omega(n^{s_0}) > 0. \end{aligned} \tag{4.21}$$

as promised. ■

■

## 5 Counting Algorithm For Simple Hypergraphs

In what follows, we consider the problem of estimating the frequency  $|\mathcal{F}(\mathcal{G})|$  of a subhypergraph  $\mathcal{F}$  on  $f$  vertices in a host hypergraph  $\mathcal{G}$  on  $n$  vertices. Clearly, one can compute  $|\mathcal{F}(\mathcal{G})|$  precisely in time  $O(n^f)$ . We show that when  $\mathcal{F}$  is simple,  $|\mathcal{F}(\mathcal{G})|$  may be accurately approximated in considerably shorter time.

**Theorem 5.1** *Let  $\mathcal{F}$  be a simple  $k$ -uniform hypergraph on  $f \geq k \geq 2$  vertices, and let  $\zeta > 0$  be given. There exists an algorithm which, for a given  $k$ -uniform hypergraph  $\mathcal{G}$  on  $n$  vertices, computes a value  $\Phi(\mathcal{G})$  in time  $O(n^{2k-1} \log^2 n)$  for which  $||\mathcal{F}(\mathcal{G})| - \Phi(\mathcal{G})| < \zeta n^f$ .*

*Proof.* We begin the proof by first discussing some constants involved. Let  $\mathcal{F}$ ,  $f \geq k \geq 2$  and  $\zeta > 0$  be given as in the hypothesis of Theorem 5.1. Define auxiliary constants  $d_0$ ,  $\gamma > 0$  and  $t_0 \in \mathbb{N}$  by

$$d_0 = \gamma = \zeta/6 \quad \text{and} \quad \lceil 3/\zeta \rceil = t_0. \tag{5.1}$$

Let  $\epsilon' = \epsilon_{Thm.3.1}(\mathcal{F}, d_0, \gamma) > 0$  be the constant guaranteed by Theorem 3.1(Counting Lemma), and set

$$\epsilon = \min\{\zeta/6, \epsilon'\}. \tag{5.2}$$

Let

$$T_0 = T_0(t_0, \epsilon) \tag{5.3}$$

be the constant guaranteed by Theorem 1.3 (Algorithmic Regularity Lemma). This concludes our discussion of the constants.

We now list the steps of the algorithm of Theorem 5.1. Afterwards, we note the corresponding complexity, and establish the accuracy of the parameter  $\Phi(\mathcal{G})$ .

**Step 1.** With the constants  $\epsilon > 0$  and  $t_0$  chosen above, apply Theorem 1.3 (Algorithmic Regularity Lemma) to  $\mathcal{G}$  to construct an  $\epsilon$ -regular,  $t$ -equitable partition  $V(\mathcal{G}) = V_0 \cup V_1 \cup \dots \cup V_t$  where  $t_0 \leq t \leq T_0$ .

**Step 2.** Construct the following weighted *Cluster Hypergraph*  $(\mathcal{G}_*, \omega)$  on vertex set  $[t] = \{1, \dots, t\}$  whose edges are “weighted by density”.

**I:** Define  $\omega : \binom{[t]}{k} \rightarrow [0, 1]$  by:

$$\omega(\{i_1, \dots, i_k\}) = \begin{cases} d_{\mathcal{G}}(V_{i_1}, \dots, V_{i_k}) & \text{if } d_{\mathcal{G}}(V_{i_1}, \dots, V_{i_k}) \geq d_0 \\ 0 & \text{otherwise} \end{cases}$$

where  $d_0 > 0$  was chosen in (5.1).

**II:** Set  $\mathcal{G}_* = \omega^{-1}(0, 1]$  to be the collection of  $\{i_1, \dots, i_k\} \in \binom{[t]}{k}$  for which  $d_{\mathcal{G}}(V_{i_1}, \dots, V_{i_k}) \geq d_0$ .

**Step 3.** Construct the family  $\mathcal{F}(\mathcal{G}_*)$  of distinct copies  $\mathcal{F}_0$  of  $\mathcal{F}$  in  $\mathcal{G}_*$ .

**Step 4.** Compute and return

$$\Phi(\mathcal{G}) = \lfloor n/t \rfloor^f \sum_{\mathcal{F}_0 \in \mathcal{F}(\mathcal{G}_*)} \prod_{F \in \mathcal{F}_0} \omega(F). \quad (5.4)$$

This concludes the algorithm. We proceed to an analysis of its complexity and accuracy.

Note that the complexity of the algorithm above is determined by Step 1. Indeed, Theorem 2.1 constructs the partition  $V(\mathcal{G}) = V_0 \cup V_1 \cup \dots \cup V_t$  in time  $O(n^{2k-1} \log^2 n)$ .

The weighted hypergraph  $(\mathcal{G}_*, \omega)$  is constructed in time  $O(n^k)$ , since there densities are computed (and recorded) and  $t \leq T_0 = O(1)$ . Step 3 is greedily completed in constant time since, again,  $t \leq T_0 = O(1)$ .

We now prove:

$$||\mathcal{F}(\mathcal{G})| - \Phi(\mathcal{G})| < \zeta n^f. \quad (5.5)$$

We begin with some initial related estimates (see upcoming (5.6)-(5.8)). To that end, let  $\mathcal{F}_\times(\mathcal{G}) \subseteq \mathcal{F}(\mathcal{G})$  denote the collection of copies  $\mathcal{F}_0 \in \mathcal{F}(\mathcal{G})$  which “cross” the classes  $V_1 \cup \dots \cup V_t$ , i.e.  $|V(\mathcal{F}_0) \cap V_i| \leq 1$  for each  $1 \leq i \leq t$  and  $V(\mathcal{F}_0) \cap V_0 = \emptyset$ . Clearly,

$$\begin{aligned} |\mathcal{F}(\mathcal{G}) \setminus \mathcal{F}_\times(\mathcal{G})| &\leq |V_0|n^{f-1} + t \binom{\lfloor n/t \rfloor}{2} n^{f-2} \\ &< tn^{f-1} + \frac{n^f}{2t} \leq T_0 n^{f-1} + \frac{n^f}{2t_0} \\ &\stackrel{(5.1), (5.3)}{\leq} O(n^{f-1}) + \zeta n^f / 12 \leq \frac{\zeta n^f}{6}. \end{aligned} \quad (5.6)$$

Now, let  $\mathcal{F}_\times^-(\mathcal{G}) \subseteq \mathcal{F}_\times(\mathcal{G})$  denote the copies  $\mathcal{F}_0 \in \mathcal{F}_\times(\mathcal{G})$  for which there exists some  $F \in \mathcal{F}_0$  and some  $\{i_1, \dots, i_k\} \in \binom{[t]}{k}$  such that  $F \in \binom{V_{i_1} \cup \dots \cup V_{i_k}}{k}$  and  $d_{\mathcal{G}}(V_{i_1}, \dots, V_{i_k}) < d_0$ , i.e.  $\omega(\{i_1, \dots, i_k\}) = 0$ . Clearly,

$$|\mathcal{F}_\times^-(\mathcal{G})| \leq \binom{t}{k} d_0 \lfloor n/t \rfloor^k n^{f-k} \stackrel{(5.1)}{<} \frac{\zeta}{6} n^f. \quad (5.7)$$

Finally, let  $\mathcal{F}_\times^{irr}(\mathcal{G})$  denote the copies  $\mathcal{F}_0 \in \mathcal{F}_\times(\mathcal{G})$  for which, for some  $F \in \mathcal{F}_0$ , there exists  $\{i_1, \dots, i_k\} \in \binom{[t]}{k}$  for which  $F \in \binom{V_{i_1} \cup \dots \cup V_{i_k}}{k}$  and  $\mathcal{G}[V_{i_1}, \dots, V_{i_k}]$  is  $\epsilon$ -irregular. By Theorem 2.1:

$$|\mathcal{F}_\times^{irr}(\mathcal{G})| \leq \epsilon \binom{t}{k} \lfloor n/t \rfloor^k n^{f-k} < \epsilon n^f. \quad (5.8)$$

Now, let  $\mathcal{F}_\times^+(\mathcal{G}) = \mathcal{F}_\times(\mathcal{G}) \setminus (\mathcal{F}_\times^-(\mathcal{G}) \cup \mathcal{F}_\times^{irr}(\mathcal{G}))$ . Note that (5.6)-(5.8) combine to yield

$$||\mathcal{F}(\mathcal{G})| - |\mathcal{F}_\times^+(\mathcal{G})|| \leq (\epsilon + \frac{\zeta}{3})n^f \stackrel{(5.2)}{\leq} \frac{\zeta}{2}n^f. \quad (5.9)$$

Thus to prove (5.5), it suffices to prove the following claim.

**Claim 5.2**

$$|\mathcal{F}_\times^+(\mathcal{G})| - \Phi(\mathcal{G}) < \frac{\zeta}{2} n^f.$$

*Proof of Claim 5.2.* Consider the following subhypergraph  $\mathcal{G}'_* \subseteq \mathcal{G}_*$  of the earlier cluster hypergraph  $\mathcal{G}_*$ : for each  $\{i_1, \dots, i_k\} \in \mathcal{G}_*$ , let  $\{i_1, \dots, i_k\} \in \mathcal{G}'_*$  if, and only if,  $\mathcal{G}[V_{i_1}, \dots, V_{i_k}]$  is  $\epsilon$ -regular. Then  $\mathcal{G}'_*$  is a subhypergraph of  $\mathcal{G}_*$  with all but  $\epsilon \binom{t}{k}$  fewer edges. Now, for each  $\mathcal{F}_0 \in \mathcal{F}(\mathcal{G}'_*)$ , define the following subhypergraph  $\mathcal{H}_{\mathcal{F}_0} \subseteq \mathcal{G}$  of  $\mathcal{G}$ : For each  $\{i_1, \dots, i_k\} \in \binom{[t]}{k}$ , let

$$\mathcal{H}_{\mathcal{F}_0}[V_{i_1}, \dots, V_{i_k}] = \begin{cases} \mathcal{G}[V_{i_1}, \dots, V_{i_k}] & \text{if } \{i_1, \dots, i_k\} \in \mathcal{F}_0 \\ \emptyset & \text{otherwise.} \end{cases}$$

Then,

$$|\mathcal{F}_\times^+(\mathcal{G})| = \sum_{\mathcal{F}_0 \in \mathcal{F}(\mathcal{G}'_*)} |\mathcal{F}(\mathcal{H}_{\mathcal{F}_0})|.$$

Moreover, applying Theorem 3.1 (Counting Lemma) to each term in the sum above, we have:

$$\begin{aligned} |\mathcal{F}_\times^+(\mathcal{G})| &= (1 \pm \gamma) \lfloor n/t \rfloor^f \sum_{\mathcal{F}_0 \in \mathcal{F}(\mathcal{G}'_*)} \prod_{F \in \mathcal{F}_0} d_F \\ &= (1 \pm \gamma) \lfloor n/t \rfloor^f \sum_{\mathcal{F}_0 \in \mathcal{F}(\mathcal{G}'_*)} \prod_{F \in \mathcal{F}_0} \omega(F) \end{aligned}$$

where for each  $F = \{i_1, \dots, i_k\} \in \mathcal{F}_0 \in \mathcal{F}(\mathcal{G}'_*)$ ,  $d_F = d_{\mathcal{G}}(V_{i_1}, \dots, V_{i_k}) = \omega(F)$ . We therefore have

$$\left| |\mathcal{F}_\times^+(\mathcal{G})| - \lfloor n/t \rfloor^f \sum_{\mathcal{F}_0 \in \mathcal{F}(\mathcal{G}'_*)} \prod_{F \in \mathcal{F}_0} \omega(F) \right| \leq \gamma n^f. \quad (5.10)$$

Recalling (5.4), we see:

$$\begin{aligned}
& \left| \lfloor n/t \rfloor^f \sum_{\mathcal{F}_0 \in \mathcal{F}(\mathcal{G}'_*)} \prod_{F \in \mathcal{F}_0} \omega(F) - \Phi(\mathcal{G}) \right| \\
& \leq \lfloor n/t \rfloor^f \sum \left\{ \prod_{F \in \mathcal{F}_0} \omega(F) : \mathcal{F}_0 \in \mathcal{F}(\mathcal{G}_*) \setminus \mathcal{F}(\mathcal{G}'_*) \right\} \\
& \leq \lfloor n/t \rfloor^f |\mathcal{F}(\mathcal{G}_*) \setminus \mathcal{F}(\mathcal{G}'_*)| \\
& \leq \lfloor n/t \rfloor^f |\mathcal{G}_* \setminus \mathcal{G}'_*| t^{f-k} \\
& \leq \lfloor n/t \rfloor^f \epsilon \binom{t}{k} t^{f-k} < \epsilon n^f.
\end{aligned} \tag{5.11}$$

Combining (5.10) and (5.11) implies:

$$\left| |\mathcal{F}_\times^+(\mathcal{G})| - \Phi(\mathcal{G}) \right| < (\gamma + \epsilon) n^f \stackrel{(5.1), (5.2)}{<} \frac{\zeta}{2} n^f,$$

hence, proving Claim 5.2. ■

■

## 6 Constructive Removal Lemma

For our final result, we consider a well known Theorem of Nagle, Rödl, Shcacht, Skokan [20, 27] and Gowers [13, 14] known as the Removal Lemma. This theorem roughly asserts that if a ‘large’ hypergraph  $\mathcal{H}$  contains ‘few’ copies of a fixed subhypergraph  $\mathcal{F}$ , then one may delete ‘few’ edges from  $\mathcal{H}$  to destroy all these copies. The original proof relies on the hypergraph regularity and counting lemmas of [13, 14, 20, 27] and is highly technical. In the case that  $\mathcal{F}$  is simple, we give an easier proof which is, in fact, constructive, whereas the original more general proof was not.

**Theorem 6.1 (Constructive Removal Lemma)** *For all integers  $f \geq k \geq 2$ ,  $\gamma > 0$  and simple  $k$ -uniform hypergraphs  $\mathcal{F}$  on  $f$  vertices, there exists  $\delta > 0$  and integer  $n_0 = n_0(f, k, \gamma, \mathcal{F}, \delta)$  so that the following holds.*

*Given a  $k$ -uniform hypergraph  $\mathcal{H}$  on  $n \geq n_0$  vertices which contains fewer than  $\delta n^f$  copies of  $\mathcal{F}$ , one may delete, in time  $O(n^{2k-1} \log^2 n)$ , at most  $\gamma n^k$  edges from  $\mathcal{H}$  to make it  $\mathcal{F}$ -free.*

*Proof of Theorem 6.1.* Let integers  $f \geq k \geq 2$ ,  $\zeta > 0$  and simple  $k$ -uniform hypergraph  $\mathcal{F}$  on  $f$  vertices be given. To define the promised constants, we first consider several auxiliary constants. Set  $\gamma = 1/2$ ,  $d_0 = \zeta/3$  and  $t_0 = \lceil 4/\zeta \rceil$ . We now appeal to Theorem 3.4 (Generalized Counting Lemma). To that end, let  $\mathcal{F}$  be  $p$ -partite for some  $f \geq p \geq k \geq 2$ . Let  $\epsilon_{\text{Thm. 3.4}}(p) = \epsilon_{\text{Thm. 3.4}}(f, p, k, \gamma, d_0, \mathcal{F}) > 0$  be the constant guaranteed by Theorem 3.4, and let  $\epsilon_{\text{Thm. 3.4}} = \min_p \epsilon_{\text{Thm. 3.4}}(p)$  where the minimum is taken over all integers  $k \leq p \leq f$  for which  $\mathcal{F}$  is  $p$ -partite. Let  $\epsilon = \min\{\zeta/3, \epsilon_{\text{Thm. 3.4}}\}$ . Let  $T_0 = T_{\text{Thm. 1.3}}(t_0, \epsilon)$  be the integer guaranteed by The-

orem 1.3. We set  $\delta = d_0^{f^k}/(2T_0^f)$  and take  $n_0 = n_0(f, k, \gamma, d_0, \mathcal{F}, \delta)$  sufficiently large whenever needed.

Let  $k$ -uniform hypergraph  $\mathcal{H}$  on  $n \geq n_0$  vertices be given, where  $\mathcal{H}$  contains fewer than  $\delta n^f$  copies of the fixed hypergraph  $\mathcal{F}$ . We show that in time  $O(n^{2k-1} \log^2 n)$  we may locate fewer than  $\zeta n^k$  edges in  $\mathcal{H}$  whose removal makes  $\mathcal{H}$   $\mathcal{F}$ -free. To that end, with constants  $\epsilon > 0$  and  $t_0$  above, apply Theorem 1.3 (constructive regularity lemma) to  $\mathcal{H}$  to obtain, in time  $O(n^{2k-1} \log^2 n)$ , an  $\epsilon$ -regular and  $t$ -equitable partition  $V(\mathcal{H}) = V_1 \cup \dots \cup V_t$ ,  $t_0 \leq t \leq T_0$ , whose every  $\epsilon$ -irregular  $k$ -tuple  $V_{i_1}, \dots, V_{i_k}$  is identified.

We now delete the following edges  $H \in \mathcal{H}$ :

1. all  $H \in \mathcal{H}$  for which  $H \cap V_0 \neq \emptyset$  or  $|H \cap V_i| \geq 2$  for some  $1 \leq i \leq t$ ;
2. all  $H \in \mathcal{H}$  for which there exist  $1 \leq i_1 < \dots < i_k \leq t$  such that  $|H \cap V_{i_j}| = 1$ ,  $1 \leq j \leq k$ , where  $d_{\mathcal{H}}(V_{i_1}, \dots, V_{i_k}) < d_0$ ;
3. all  $H \in \mathcal{H}$  for which there exist  $1 \leq i_1 < \dots < i_k \leq t$  such that  $|H \cap V_{i_j}| = 1$ ,  $1 \leq j \leq k$ , where  $V_{i_1}, \dots, V_{i_k}$  is  $\epsilon$ -irregular.

We now enumerate the deleted edges (and use that  $n$  is sufficiently large). First, the edges  $H \in \mathcal{H}$  satisfying (1) total at most

$$tn^{k-1} + t \binom{\lfloor n/t \rfloor}{2} n^{k-2} < T_0 n^{k-1} + \frac{n^k}{t} \leq O(n^{k-1}) + \frac{n^k}{t_0} \leq \frac{\zeta}{3} n^k$$

and can be identified (for deletion) in time  $O(t(n/t)^2 n^{k-2}) = O(n^k)$ . Second, the edges  $H \in \mathcal{H}$  satisfying (2) total at most

$$\binom{t}{k} d_0 \lfloor n/t \rfloor^k \leq d_0 n^k \leq \frac{\zeta}{3} n^k$$

and can be identified in time  $O(t^k (n/t)^k) = O(n^k)$  since each  $d_{\mathcal{H}}(V_{i_1}, \dots, V_{i_k})$  (and in particular,  $|\mathcal{H}[V_{i_1}, \dots, V_{i_k}]|$ )  $1 \leq i_1 < \dots < i_k \leq t$ , can be computed in time



$O(\lfloor n/t \rfloor^k)$ ). Finally, the edges  $H \in \mathcal{H}$  satisfying (3) total at most

$$\epsilon \binom{t}{k} \lfloor n/t \rfloor^k \leq \epsilon n^k \leq \frac{\zeta}{3} n^k$$

and can be identified in time  $O(t^k(n/t)^k) = O(n^k)$  since Theorem 1.3 already identified all  $\epsilon$ -irregular  $k$ -tuples  $V_{i_1}, \dots, V_{i_k}$ ,  $1 \leq i_1 < \dots < i_k \leq t$ . Note that the deletion process above has removed at most  $\zeta n^k$  edges from  $\mathcal{H}$ . Note that the complexity of the deletion process is determined by the application of Theorem 1.3, and so all of the above is done in time  $O(n^{2k-1} \log^2 n)$ .

Let  $\mathcal{H}'$  be the hypergraph obtained after the edges above have been removed from  $\mathcal{H}$ . We now argue that  $\mathcal{H}'$  is  $\mathcal{F}$ -free. Indeed, assume on the contrary that  $\mathcal{H}'$  contains a copy  $\mathcal{F}'$  of  $\mathcal{F}$ . Let  $\mathcal{F}'$  be  $p$ -partite for some  $k \leq p \leq f$  and write  $V(\mathcal{F}') = W_{j_1} \cup \dots \cup W_{j_p}$  where  $W_{j_\ell} \subseteq V_{j_\ell}$ ,  $1 \leq \ell \leq p$ . By virtue of the deletion process above, every edge  $F \in \mathcal{F}'$  crosses the  $p$ -partition above and further satisfies the following property: if  $F \subseteq W_{j_{\ell_1}} \cup \dots \cup W_{j_{\ell_k}}$  for some  $1 \leq \ell_1 < \dots < \ell_k \leq p$ , then  $\mathcal{H}[V_{j_{\ell_1}}, \dots, V_{j_{\ell_k}}]$  is  $\epsilon$ -regular with density at least  $d_0$ . Theorem 3.4 then applies to say (with  $n$  sufficiently large)

$$|\mathcal{F}(\mathcal{H})| \geq |\mathcal{F}(\mathcal{H}[V_{j_{\ell_1}}, \dots, V_{j_{\ell_k}}])| \geq \frac{1}{2} d_0^{|F|} \lfloor n/t \rfloor^f > \frac{1}{2} d_0^{fk} \left( \frac{n}{T_0} \right)^f = \delta n^f,$$

which contradicts our hypothesis that  $|\mathcal{F}(\mathcal{H})| \leq \delta n^f$ . ■

## References

- [1] V.E. Alekseev, *On the entropy values of hereditary classes of graphs*, Discrete Mathematics and Applications **3** (1993), 191–199.
- [2] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, R. Yuster, *The algorithmic aspects of the regularity lemma*, J. Algorithms **16** (1994), no. 1, 80–109.
- [3] B. Bollobás, A. Thomason, *Projections of bodies and hereditary properties of hypergraphs*, Bull. London. Math. Soc. **27** (1995), 417–424.
- [4] B. Bollobás, A. Thomason, *Hereditary and monotone properties of graphs*, The Mathematics of Paul Erdős, II, Springer, Berlin (1997), 70–78.
- [5] F. Chung. *Regularity lemmas for hypergraphs and quasirandomness*. Random Structures and Algorithms, 2:241–252, 1991.
- [6] A. Czygrinow, V. Rödl, *An algorithmic regularity lemma for hypergraphs*, SIAM J. Comput. **30** (2000), no. 3–4, 293–332.
- [7] R. Duke, H. Lefmann, V. Rödl, *A fast approximation algorithm for computing the frequencies of subgraphs of a given graph*, SIAM Journal of Computing Vol. 24, pp. 598–620, 1995
- [8] P. Erdős, P. Frankl, V. Rödl, *Asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs Comb. **2** (1986), 113–121.

- [9] P. Erdős, D.J. Kleitman, B.L. Rothschild, *Asymptotic enumeration of  $K_n$ -free graphs*, Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II, pp. 19–27. Atti dei Convegna Lincei, No. 17, Accad. Naz. Lincei, Rome, 1976.
- [10] P. Frankl, V. Rödl, *Extremal problems on set systems*, Random Structures Algorithms **20** (2002), no. 2, 131–164.
- [11] P. Frankl, V. Rödl. *The uniformity lemma for hypergraphs*. Graphs and Combinatorics, 8:309–312, 1992.
- [12] A. Frieze, R. Kannan. *The regularity lemma and approximation schemes for dense problems*. In Proceedings of the Symposium on Foundations of Computer Science, pp. 12–20. IEEE, 1996.
- [13] W.T. Gowers, *Quasirandomness, counting and regularity for 3-uniform hypergraphs*, Combin. Probab. Comput. **15** (2006), no. 1–2, 143–184.
- [14] W.T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi Theorem*, Ann. of Math. (2) **166** (2007), no. 3–897–946.
- [15] Y. Kohayakawa, B. Nagle, V. Rödl, *Hereditary properties of triple systems*, Combin. Probab. Comput. **12** (2003), 155–189.
- [16] Y. Kohayakawa, B. Nagle, V. Rödl, M. Schacht, *Weak regularity and linear hypergraphs*, Journal of Combinatorial Theory (B). (to appear)
- [17] J. Komlós, A. Shokoufandeh, M. Simonovits, E. Szemerédi, *The regularity lemma and its applications in graph theory*, Theoretical aspects of computer science (Tehran, 2000), Lecture Notes in Comput. Sci., vol. 2292, Springer–Berlin, 2002, pp. 84–112.
- [18] J. Komlós, M. Simonovits, *Szemerédi’s Regularity Lemma and its applications in graph theory*, Combinatorics, Paul Erdős is eighty, Vol. 2, pp. 295–352, 1996.
- [19] B. Nagle, V. Rödl, *The asymptotic number of triple systems not containing a fixed one*, Disc. Math. **235** (2001), 271–290.

- [20] B. Nagle, V. Rödl, M. Schacht, *The counting lemma for regular  $k$ -uniform hypergraphs*, Random Structures Algorithms **28** (2006), no. 2, 113–179.
- [21] B. Nagle, V. Rödl, M. Schacht, *Extremal hypergraph problems and the regularity method*, Topics in discrete mathematics, Algorithms Combin., vol. 26, Springer, Berlin, 2006, pp. 247–278.
- [22] H.J. Prömel, A. Steger, *Excluding induced subgraphs: quadrilaterals*, Random Structures Algorithms **2** (1991), 55–71.
- [23] H.J. Prömel, A. Steger, *The asymptotic number of graphs not containing a fixed color-critical subgraph*, Combinatorica **12** (1992), 463–473.
- [24] H.J. Prömel, A. Steger, *Excluding induced subgraphs III. A general asymptotic*, Random Structures Algorithms **3** (1992), no. 1, 19–31.
- [25] H.J. Prömel, A. Steger, *The asymptotic structure of  $H$ -free graphs*, Graph structure theory (Seattle, WA, 1991), Contemp. Math., vol. 147, Amer. Math. Soc., Providence, RI, 1993, pp. 167–178.
- [26] V. Rödl, B. Nagle, J. Skokan, M. Schacht, Y. Kohayakawa, *The hypergraph regularity method and its applications*, Proc. Natl. Acad. Sci. USA **102** (2005), no. 23, 8109–8113.
- [27] V. Rödl, J. Skokan, *Regularity lemma for  $k$ -uniform hypergraphs*, Random Structures Algorithms **25** (2004), no. 1, 1–42.
- [28] E. Szemerédi, *On sets of integers containing no  $k$  elements in arithmetic progression*, Acta Arithmetica **27** (1975), 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik.
- [29] E. Szemerédi, *Regular partitions of graphs*, Problèmes en combinatoires et théorie des graphes, (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Colloq. Internat. CNRS, vol. 260, CNRS, Paris, 1978, pp. 399–401.

- [30] T. Tao, *A variant of the hypergraph removal lemma*, J. Combin. Theory Ser. A **113** (2006), no. 7, 1257–1280.