Discrete Applied Mathematics 116 (2002) 179–191

DISCRETE APPLIED MATHEMATICS

# Partitioning problems in dense hypergraphs

A. Czygrinow

*Department of Mathematics, Arizona State University, Tempe, AZ 85287-1804, USA*

Received 27 April 1999; revised 10 April 2000; accepted 19 June 2000

**Abstract**

We study the general partitioning problem and the discrepancy problem in dense hypergraphs. Using the regularity lemma (Szemerédi, Problemes Combinatories et Theorie des Graphes (1978), pp. 399–402) and its algorithmic version proved in Czygrinow and Rödl (SIAM J. Comput., to appear), we give polynomial-time approximation schemes for the general partitioning problem and for the discrepancy problem. © 2002 Elsevier Science B.V. All rights reserved.

*Keywords*: Partitioning problem; Approximation algorithm; The regularity lemma

## 1. Introduction

For some NP-complete problems good approximation algorithms were developed for the dense instances of the problems. A typical example is the Max-Cut problem in which one tries to partition a vertex set of a graph into two subsets in such a way that the number of edges that have the endpoints in different sets is maximized. The Max-Cut problem is known to be NP-complete [12] and, moreover, hardness results for the corresponding approximation problem have been recently proved [3]. In contrast to the difficulty of the general case, for the dense instances of the problem (graph $(V,E)$ has at least $c|V|^2$ edges for some positive constant $c$) polynomial-time approximation schemes were developed. Arora et al. [2] used the sampling method to design an approximation algorithm that finds in $O(|V|^{O(1/\varepsilon^2)})$ time a cut of value which is within $1-\varepsilon$ factor of the optimal. Fernandez de la Vega [7] presented a $O(2^{1/\varepsilon^2+o(1)})$ algorithm which finds a cut of value within $1-\varepsilon$ factor of the optimal. Using a version of the regularity lemma Frieze and Kannan [10] gave an algorithm which finds in $O(n^{2.376})$ time, a cut with value which is within $1-\varepsilon$ of the optimal. Other partitioning problems were also considered in [10]. In addition, both sampling method and a version of

*E-mail address:* andrzej@math.la.asu.edu (A. Czygrinow)


PII: S0166-218X(00)00323-1

regularity combined with sampling lead to fast randomized algorithms [10,11]. It should be noted then in most of the applications of the regularity lemma a constant hidden in big O depends on $\varepsilon$ and is enormous. However, version proposed by Frieze and Kannan has constant which is relatively small. Many of the problems for which the algorithmic version of the regularity lemma was successfully applied have interesting generalizations to hypergraphs. The situation in hypergraphs is however much more complicated and the obvious generalization of the concept of graph regularity (weak regularity) does not always lead to the extensions of applications for graphs. Therefore, different approaches to measure the regularity of hypergraphs were developed (strong regularity) [8,9].

We should also mention that the original proof of Szemerédi [14] was not constructive, in a sense that it did not provide a polynomial-time algorithm that would find a desired partition. The algorithmic version of graph regularity lemma was given by Alon et al. in [1] and was applied to various problems [1,6,4,10,13] The algorithm from [1] is based on the characterization of regularity which roughly states that a pair of sets is $\varepsilon$-regular with density $d$ if and only if "most" of the vertices have degree around $dn$ and "most" of the pairs of vertices have co-degree around $d^2n$. This characterization does not have a natural generalization to hypergraphs, the algorithmic version of the weak regularity lemma for hypergraphs was however proved recently in [5] (see also [11]).

In this paper, we will be interested in applications of the algorithmic version of hypergraph regularity lemma. We will discuss two applications, first we will concentrate on the general partitioning problem and then we will discuss the discrepancy problem. Let us start with definitions and notation.

An $l$-uniform hypergraph is a pair $H=(V,E)$ where $V$ is a nonempty finite set of vertices and $E$ is a set of $l$-element subsets of $V$ called edges. We will use $[k]=\{1,\dots,k\}$.

**Definition 1.** Let $V_1,V_2,\dots,V_k$ be pairwise disjoint nonempty subsets of the vertex set $V$. An edge $e\in E$ is called crossing in $V_1,V_2,\dots,V_k$ if for every $i\in[k]$, $|e\cap V_i|\leqslant 1$. The number of crossing edges in $V_1,V_2,\dots,V_k$ will be denoted by $e(V_1,V_2,\dots,V_k)$.

For an $l$-uniform hypergraph $H=(V,E)$, we will be interested in finding a partition of $V$ into $k$ (where $k\geqslant l$) nonempty sets that maximizes the number of crossing hyperedges.

For a partition $V_1\cup V_2\cup\cdots\cup V_k$ of $V$ define

$$f(V_1,V_2,\dots,V_k)=|\{e\in H\colon\ |e\cap V_i|\leqslant 1,\ i=1,\dots,k\}| \tag{1}$$

and let

$$\bar{f}(H)=\max f(V_1,V_2,\dots,V_k), \tag{2}$$

where the maximum is taken over all partitions of $V$ into $k$ nonempty sets ($k$ is fixed and independent of the size of $H$). Finding the value of $\bar{f}(H)$ is NP-complete as even in the case $l = k = 2$, we get the well-studied Max-Cut problem. Using the regularity lemma, we will device an approximation algorithm which finds a partition $V_1 \cup V_2 \cup \cdots \cup V_k$, the value of which is within $\eta n^l$ of $\bar{f}(H)$.

**Theorem 2.** *For $l$ and $k \geqslant l$ there is an algorithm which for every $l$-uniform hypergraph $H = (V, E)$ with $|V| = n$ and every $0 < \eta < 1$ finds in $O(n^{2l-1} \log^2 n)$ time a partition $V_1, \ldots, V_k$ of $V$ such that*

$$f(V_1, V_2, \ldots, V_k) \geqslant \bar{f}(H) - \eta n^l.$$

The algorithm of Theorem 2 can be used to approximate $\bar{f}(H)$ in case when the number of edges in $H$ is at least $cn^l$ for some positive constant $c$.

Our second application concerns the discrepancy of a coloring of $l$-element subsets of $[n]$. Let $\chi : \binom{[n]}{l} \to \{-1, +1\}$ be a $\{-1, +1\}$-coloring of $l$-element subsets of $[n]$. Define the discrepancy of a set $S \subset [n]$ as

$$d(S) = \left| \sum_{e \subset S} \chi(e) \right|, \tag{3}$$

where the sum is taken over all $e \in \binom{[n]}{l}$ which are contained in $S$ and the discrepancy of the coloring $\chi$

$$disc(\chi) = \max_S d(S). \tag{4}$$

Our problem will be to find a set $S^* \subset [n]$ that maximizes the value of $d$. Using the algorithmic version of the regularity lemma for hypergraphs, we will show the following hypergraph analog of the result in [4].

**Theorem 3.** *For every $l$ there is an algorithm which for a coloring of the $l$-element subsets of $[n]$, $\chi$: $\binom{[n]}{l} \to \{-1, +1\}$ and $0 < \eta < 1$ finds in $O(n^{2l-1} \log^2 n)$ time a set $S^* \subset [n]$ such that*

$$d(S^*) \geqslant disc(\chi) - \eta n^l.$$

Clearly, the algorithm of Theorem 3 will give meaningful results only if $\mathrm{disc}(\chi) \geqslant cn^l$ for some positive constant $c$. If in addition $c$ is known in advance, the algorithm can be easily modified to a polynomial-time approximation scheme.

## 2. Preliminaries

In this section, we will introduce the necessary notation and we will formulate the regularity lemma. Let $H = (V, E)$ be an $l$-uniform hypergraph.

For an $l$-tuple of pairwise disjoint nonempty sets $V_1, V_2, \ldots, V_l$, we define the density of $(V_1, V_2, \ldots, V_l)$ as

$$d(V_1, V_2, \ldots, V_l) = \frac{e(V_1, V_2, \ldots, V_l)}{|V_1||V_2| \cdots |V_l|}.$$

An $l$-tuple $(V_1, V_2, \ldots, V_l)$ is called $\varepsilon$-regular if for every $V_i' \subset V_i$ with $|V_i'| \geqslant \varepsilon |V_i|$, where $i = 1, \ldots, l$,

$$|d(V_1', V_2', \ldots, V_l') - d(V_1, V_2, \ldots, V_l)| \leqslant \varepsilon.$$

**Definition 4.** A partition $V_0 \cup V_1 \cup \cdots \cup V_t$ of vertex set is called $\varepsilon$-regular if the following conditions are satisfied.
1. $|V_0| \leqslant \varepsilon |V|$.
2. For every $i, j \in [t]$, $|V_i| = |V_j|$.
3. All but at most $\varepsilon t^l$ of $l$-tuples $(V_{i_1}, \ldots, V_{i_l})$ $(i_1, \ldots, i_l \in [t])$ are $\varepsilon$-regular.

The powerful lemma of Szemerédi [12] states that for every $0 < \varepsilon < 1$ every hypergraph which is large enough admits an $\varepsilon$-regular partition into a constant number of classes.

**Lemma 5.** *For every* $\varepsilon \in (0,1)$ *and every positive integers* $l$ *and* $m$ *there exist two integers* $N = N(\varepsilon, m, l)$ *and* $M = M(\varepsilon, m, l)$ *such that every* $l$*-uniform hypergraph with at least* $N$ *vertices admits an* $\varepsilon$*-regular partition* $V_0 \cup V_1 \cup \cdots \cup V_t$ *with* $m \leqslant t \leqslant M$.

It is essential for our algorithms (as well as for most of other applications of the regularity lemma) that the number of partition classes $t + 1$ depends on $\varepsilon$ but not on the size of the hypergraph. Alon et al. [1] gave a $O(|V|^{2.376})$ algorithm which finds an $\varepsilon$-regular partition of graphs, Czygrinow and Rödl [5] gave a $O(|V|^{2l-1} \log^2 |V|)$ algorithm that finds an $\varepsilon$-regular partition of an $l$-uniform hypergraph. The result from [5] can be stated as follows.

**Lemma 6.** *For every* $l$, $m$, *and* $\varepsilon$ *there exist* $N$, $M$ *and an algorithm which for any* $l$*-uniform hypergraph* $H = (V, E)$ *with* $|V| = n \geqslant N$ *finds in* $O(n^{2l-1} \log^2 n)$ *time an* $\varepsilon$*-regular partition* $V_0 \cup V_1 \cup \cdots \cup V_t$ *with* $m \leqslant t \leqslant M$.

We will use the algorithmic version of the lemma in both our approximation algorithms.

## 3. The general partitioning problem

In this section, we prove Theorem 2. Let $H = (V, E)$ be an $l$-uniform hypergraph. Before the formal description let us give an idea of the algorithm. The algorithm will proceed in two steps. In the first step, it finds an $\varepsilon$-regular partition $U_0 \cup U_1 \cup \cdots \cup U_t$ of the hypergraph. In the second step, it checks exhaustively all partitions into $k$ sets which are the unions of $U_i$'s and chooses one that maximizes a function $f^*$ which is an "approximation" of $f$.

It will be convenient to assume that we are searching for a partition with at most $k$ sets rather than exactly $k$. In case, a partition with less than $k$ classes is found, one can increase the number of classes to $k$ (arbitrarily) without loosing any of the crossing edges.

**Algorithm 1.**
1. Set $\varepsilon = \eta/2(k+4)$. Find an $\varepsilon$-regular partition $U_0 \cup U_1 \cup \cdots \cup U_t$ of $H$ with $t \geqslant \max\{1/\varepsilon, k\}$.
2. Check all partitions $V_1 \cup V_2 \cup \cdots \cup V_k$ that satisfy the following property: for every $j = 0, \ldots, t$ there is exactly one $i \in [k]$ such that $U_j \subset V_i$. Choose within these partitions a partition $V_1 \cup V_2 \cup \cdots \cup V_k$ that maximizes

$$f^*(V_1, \ldots, V_k) = \sum_{1 \leqslant i_1 < \cdots < i_l \leqslant k} \sum_{1 \leqslant j_1 < \cdots < j_l \leqslant t} d(U_{j_1}, \ldots, U_{j_l}) |V_{i_1} \cap U_{j_1}| \cdots |V_{i_l} \cap U_{j_l}|.$$

Note that in the second step we check less than $k^t$ partitions, where both $k$ and $t$ are constants independent of $n$. Thus, the complexity of the algorithm is determined by the first step and is $O(n^{2l-1} \log^2 n)$.

First, we will show that $f^*$ is maximized for partitions considered in the second step of the algorithm.

**Fact 7.** *For every partition $V_1 \cup \cdots \cup V_k$ and for every $j \in [t]$ there exists $i \in [k]$ such that*

$$f^*(V_1, \ldots, V_k) \leqslant f^*(V_1 \setminus U_j, \ldots, V_{i-1} \setminus U_j, V_i \cup U_j, V_{i+1} \setminus U_j, \ldots, V_k \setminus U_j).$$

**Proof.** By definition,

$$f^*(V_1, \ldots, V_k) = \sum_{1 \leqslant i_1 < \cdots < i_l \leqslant k} \sum_{1 \leqslant j_1 < \cdots < j_l \leqslant t} d(U_{j_1}, \ldots, U_{j_l}) |V_{i_1} \cap U_{j_1}| \cdots |V_{i_l} \cap U_{j_l}|. \tag{5}$$

Divide the summation on the right-hand side of (5) into two sums.

$$f^*(V_1, \ldots, V_k) = f_1^*(V_1, \ldots, V_k) + f_2^*(V_1, \ldots, V_k) \tag{6}$$

where $f_1^*$ and $f_2^*$ are defined as

$$f_1^*(V_1, \ldots, V_k) = \sum_{i=1}^{k} \sum d(U_j, U_{j_1}, \ldots, U_{j_{l-1}}) |V_i \cap U_j| |V_{i_1} \cap U_{j_1}| \cdots |V_{i_{l-1}} \cap U_{j_{l-1}}|, \tag{7}$$

where the second sum is taken over all $(i_1, \ldots, i_{l-1})$ such that $1 \leqslant i_1 < \cdots < i_{l-1} \leqslant k$ with $i_\alpha \neq i$ and all $(j_1, \ldots, j_{l-1})$ such that $1 \leqslant j_1 < \cdots < j_{l-1} \leqslant t$ with $j_\alpha \neq j$.

$$f_2^*(V_1, \ldots, V_k) = \sum_{1 \leqslant i_1 < \cdots < i_l \leqslant k} \sum_{1 \leqslant j_1 < \cdots < j_l \leqslant t, j_\alpha \neq j} d(U_{j_1}, \ldots, U_{j_l}) |V_{i_1} \cap U_{j_1}| \cdots |V_{i_l} \cap U_{j_l}|. \tag{8}$$

Then (7) can be written as

$$f_1^*(V_1,\dots,V_k)=\sum_{i=1}^{k}|V_i\cap U_j|\sum d(U_j,U_{j_1},\dots,U_{j_{l-1}})|V_{i_1}\cap U_{j_1}|\cdots|V_{i_{l-1}}\cap U_{j_{l-1}}|. \tag{9}$$

Since $\sum_{i=1}^{k}|V_i\cap U_j|=|U_j|$ there exists $i\in[k]$ such that

$$f_1^*(V_1,\dots,V_k)\leqslant |U_j|\sum_{1\leqslant i_1<\cdots<i_{l-1}\leqslant k, i_\alpha\neq i}\ \sum_{1\leqslant j_1<\cdots<j_{l-1}\leqslant t, j_\alpha\neq j} d(U_j,U_{j_1},\dots,U_{j_{l-1}})\, |V_{i_1}\cap U_{j_1}|\cdots|V_{i_{l-1}}\cap U_{j_{l-1}}|, \tag{10}$$

and the right-hand side of (10) is equal to $f_1^*(V_1\backslash U_j,\dots,V_{i-1}\backslash U_j,V_i\cup U_j,V_{i+1}\backslash U_j,\dots,V_k\backslash U_j)$. Since the right-hand side (8) does not involve $U_j$ at all, we have

$$f_2^*(V_1,\dots,V_k)=f_2^*(V_1\backslash U_j,\dots,V_{i-1}\backslash U_j,V_i\cup U_j,V_{i+1}\backslash U_j,\dots,V_k\backslash U_j).$$

Therefore, by (5),

$$f^*(V_1,\dots,V_k)\leqslant f^*(V_1\backslash U_j,\dots,V_{i-1}\backslash U_j,V_i\cup U_j,V_{i+1}\backslash U_j,\dots,V_k\backslash U_j). \qquad \square$$

In our next lemma, we will show that $f^*(V_1,\dots,V_k)$ is a "good" approximation of $f(V_1,\dots,V_k)$.

**Lemma 8.** *For every partition $V_1\cup V_2\cup\cdots\cup V_k$ of $V$, we have*

$$f(V_1,\dots,V_k)\leqslant f^*(V_1,\dots,V_k)+\varepsilon(4+k)n^l.$$

**Proof.** Let $F$ be the set of crossing edges in $V_1\cup V_2\cup\cdots\cup V_k$, that is

$$F=\{e\in E\colon |e\cap V_i|\leqslant 1,\ i=1,\dots,k\} \tag{11}$$

To show that $f^*$ is an $\varepsilon$-approximation of $f$ we will use a five step process. In each step, we define the set $F_i$ and $f_i=|\{e\in F_i\}|$ and show that $f_i$ is a "good" approximation of $f_{i-1}$ (with $f_0=f$).

Let $F_1$ be the subset of $F$ of the edges that are crossing in $U_1,U_2,\dots,U_t$ and let $f_1(V_1,\dots,V_k)=|F_1|$.

**Fact 9.** $f(V_1,\dots,V_k)\leqslant f_1(V_1,\dots,V_k)+\varepsilon n^l$.

**Proof.** Clearly, the number of edges that are non-crossing in $U_1,\dots,U_t$ is at most $t\cdot t^{l-2}(n/t)^l\leqslant n^l/t\leqslant\varepsilon n^l$. $\square$

Let $F_2\subset F_1$ be the set of the edges that are not adjacent to the exceptional class $U_0$ and let $f_2(V_1,\dots,V_k)=|\{e\in F_2\}|$.

**Fact 10.** $f_1(V_1,\dots,V_k)\leqslant f_2(V_1,\dots,V_k)+\varepsilon n^l$.

**Proof.** The total number of edges that are adjacent to $U_0$ is at most $|U_0|n^{l-1} \leqslant \varepsilon n^l$. $\square$

Let $F_3 \subset F_2$ be the set of edges that are the crossing edges of $\varepsilon$-regular $l$-tuples and let $f_3(V_1,\dots,V_k) = |\{e \in F_3\}|$.

**Fact 11.** $f_2(V_1,\dots,V_k) \leqslant f_3(V_1,\dots,V_k) + \varepsilon n^l$.

**Proof.** Indeed, the number of edges that occur in $\varepsilon$-irregular $l$-tuples is at most $\varepsilon t^l (n/t)^l \leqslant \varepsilon n^l$, as we have at most $\varepsilon t^l$ irregular $l$-tuples. $\square$

Let $R$ be the set of $(j_1,\dots,j_l)$ which are such that $(U_{j_1},\dots,U_{j_l})$ is $\varepsilon$-regular and $1 \leqslant j_1 < \cdots < j_l \leqslant t$. Note that

$$f_3(V_1,\dots,V_k) = \sum_{1 \leqslant i_1 < \cdots < i_l \leqslant k} \sum_{(j_1,\dots,j_l) \in R} e(V_{i_1} \cap U_{j_1},\dots,V_{i_l} \cap U_{j_l}). \tag{12}$$

In the next fact, we exclude the edges that are adjacent to "small intersections", that is adjacent to $V_{i_\alpha} \cap U_{j_\alpha}$ such that $|V_{i_\alpha} \cap U_{j_\alpha}| < \varepsilon |U_{j_\alpha}|$. Let $F_4 = \{e \in F_3\colon \forall_{i_\alpha, j_\alpha} (e \cap (V_{i_\alpha} \cap U_{j_\alpha}) \neq \emptyset \to |V_{i_\alpha} \cap U_{j_\alpha}| \geqslant \varepsilon |U_{j_\alpha}|)\}$ and $f_4(V_1,\dots,V_k) = |\{e \in F_4\}|$.

**Fact 12.** $f_3(V_1,\dots,V_k) \leqslant f_4(V_1,\dots,V_k) + \varepsilon k n^l$.

**Proof.** If $|V_{i_\alpha} \cap U_{j_\alpha}| < \varepsilon |U_{j\alpha}| \leqslant \varepsilon n/t$ then the number of crossing edges adjacent to $V_{i_\alpha} \cap U_{j_\alpha}$ is at most $|V_{i_\alpha} \cap U_{j_\alpha}| n^{l-1} \leqslant \varepsilon n^l/t$. Since we have $t \cdot k$ intersections $V_{i_\alpha} \cap U_{j_\alpha}$ the number of edges $e \in F_3 \backslash F_4$ is at most $t \cdot k \cdot \varepsilon (n^l/t) = \varepsilon k n^l$. $\square$

In the next fact, we will approximate $f_4(V_1,\dots,V_k)$ with $f^*(V_1,\dots,V_k)$.

**Fact 13.** $f_4(V_1,\dots,V_k) \leqslant f^*(V_1,\dots,V_k) + \varepsilon n^l$.

**Proof.** Let $(U_{j_1},\dots,U_{j_l})$ be an $\varepsilon$-regular $l$-tuple. If for $(V_{i_1},\dots,V_{i_l})$ we have

$$|U_{j_\alpha} \cap V_{i_\alpha}| \geqslant \varepsilon |U_{j_\alpha}| \tag{13}$$

then by the $\varepsilon$-regularity,

$$\begin{aligned} e(V_{i_1} \cap U_{j_1},\dots,V_{i_l} \cap U_{j_l}) &= d(V_{i_1} \cap U_{j_1},\dots,V_{i_l} \cap U_{j_l}) |V_{i_1} \cap U_{j_1}| \cdots |V_{i_l} \cap U_{j_l}| \\ &\leqslant (d(U_{j_1},\dots,U_{j_l}) + \varepsilon) |V_{i_1} \cap U_{j_1}| \cdots |V_{i_l} \cap U_{j_l}|. \end{aligned}$$

Thus by (12) and the fact that for the terms in $f_4(V_1,\dots,V_k)$ condition (13) is always satisfied, we have

$$\begin{aligned} &f_4(V_1,\dots,V_k) \\ &\quad\leqslant \sum\sum d(U_{j_1},\dots,U_{j_l}) |V_{i_1} \cap U_{j_1}| \cdots |V_{i_l} \cap U_{j_l}| \\ &\qquad + \varepsilon \sum\sum |V_{i_1} \cap U_{j_1}| \cdots |V_{i_l} \cap U_{j_l}| \end{aligned}$$

$$\leqslant \sum_{1\leqslant i_1<\cdots<i_l\leqslant k}\ \sum_{1\leqslant j_1<\cdots<j_l\leqslant t} d(U_{j_1},\dots,U_{j_l})|V_{i_1}\cap U_{j_1}|\cdots|V_{i_l}\cap U_{j_l}|+\varepsilon n^l$$

$$=f^*(V_1,\dots,V_k)+\varepsilon n^l.$$

Combining Facts 9–13 gives

$$f(V_1,\dots,V_k)\leqslant f^*(V_1,\dots,V_k)+\varepsilon(4+k)n^l. \qquad \square$$

Similarly, one can show the following lemma.

**Lemma 14.** *For every partition* $V_1\cup V_2\cup\cdots\cup V_k$

$$f^*(V_1,\dots,V_k)\leqslant f(V_1,\dots,V_k)+\varepsilon(4+k)n^l.$$

**Proof of Theorem 2.** Let $\bar{V}_1\cup\cdots\cup\bar{V}_k$ be an optimal partition, that is

$$\bar{f}(H)=f(\bar{V}_1,\dots,\bar{V}_k),$$

and let $V_1\cup\cdots\cup V_k$ be the partition found by Algorithm 1. Then by Lemma 8

$$f(\bar{V}_1,\dots,\bar{V}_k)\leqslant f^*(\bar{V}_1,\dots,\bar{V}_k)+\varepsilon(4+k)n^l. \tag{14}$$

Using Fact 7, we have

$$f^*(\bar{V}_1,\dots,\bar{V}_k)\leqslant f^*(V_1,\dots,V_k) \tag{15}$$

and by Lemma 14

$$f^*(V_1,\dots,V_k)\leqslant f(V_1,\dots,V_k)+\varepsilon(4+k)n^l. \tag{16}$$

Combining (14) with (15) and (16) gives

$$f(\bar{V}_1,\dots,\bar{V}_k)\leqslant f(V_1,\dots,V_k)+2\varepsilon(4+k)n^l.$$

Since $\varepsilon=\eta/(2(4+k))$, we have

$$f(V_1,\dots,V_k)\geqslant\bar{f}(H)-\eta n^l. \qquad \square$$

## 4. Discrepancy

In this section, we consider the algorithmic approach to the discrepancy problem for hypergraphs. Our main aim is to prove Theorem 3, that is, we want to present an algorithm which for every $0<\eta<1$ finds $S^*$ such that

$$d(S^*)\geqslant disc(\chi)-\eta n^l. \tag{17}$$

**Proof of Theorem 3.** Let us define the following hypergraphs.

$$H_1=([n],\chi^{-1}(1)), \tag{18}$$

$$H_{-1} = ([n], \chi^{-1}(-1)). \tag{19}$$

Clearly $\chi^{-1}(-1) \cup \chi^{-1}(1) = \binom{[n]}{l}$.

It will be convenient to introduce some additional notation. For $S \subset [n]$ let $h_1(S)$ be the number of $e \in \binom{[n]}{l}$, such that $e \subset S$ and $\chi(e) = 1$, and similarly let $h_{-1}(S)$ be the number of $e \in \binom{[n]}{l}$ such that $e \subset S$ and $\chi(e) = -1$. Then

$$d(S) = |h_1(S) - h_{-1}(S)|. \tag{20}$$

For an $l$-tuple of pairwise disjoint sets $(V_1, V_2, \dots, V_l)$ we will denote by $h_1(V_1, \dots, V_l)$ the number of crossing edges $e$ in $V_1, \dots, V_l$ such that $\chi(e) = 1$, $h_{-1}(V_1, \dots, V_l)$ is defined in the analogous way. Similarly as in Section 3, the algorithm proceeds in two steps. In the first step, it finds an $\varepsilon$-regular partition (where $\varepsilon$ depends on $\eta$) of $H_1$ and in the second step it finds a set $S^*$ that maximizes an appropriately defined approximation of $d$.

**Algorithm 2.**

1. Set $\varepsilon = \eta/18$. Find an $\varepsilon$-regular partition $U_0, U_1, \dots, U_t$ with $t \geqslant 1/\varepsilon$ of $H_1$.
2. Check all of the sets $S = \bigcup_{j \in L} U_j$ where $L \subset [t]$ and choose one that maximizes

$$d^*(S) = \left| \sum_{1 \leqslant j_1 < \cdots < j_l \leqslant t} (2d_{j_1, j_2, \dots, j_l} - 1)|S \cap U_{j_1}||S \cap U_{j_2}| \cdots |S \cap U_{j_l}| \right|,$$

where $d_{j_1, j_2, \dots, j_l}$ is the density of $(U_{j_1}, \dots, U_{j_l})$ in $H_1$.

Note that the complexity of the procedure is $O(n^{2l-1} \log^2 n)$ as in the second step, we check $2^t$ sets $S$ and $t$ is a constant which depends only on $\varepsilon$. The proof of the correctness of the algorithm will be divided into two steps. Similarly as in [4], we will first establish a combinatorial fact which shows there is a set $S^*$ which is the union of $U_i$'s and for which $d^*(S^*) = \max_{S \subset [n]} d^*(S)$ and then we will show that $d^*$ is a "good" approximation of $d$.

**Fact 15.** *For any $1 \leqslant j \leqslant t$, and set $S \subset [n]$,*

$$d^*(S) \leqslant \max(d^*(S \cup U_j), d^*(S \backslash U_j)).$$

**Proof.** First assume that

$$\sum_{1 \leqslant j_1 < \cdots < j_l \leqslant t} (2d_{j_1, j_2, \dots, j_l} - 1)|S \cap U_{j_1}||S \cap U_{j_2}| \cdots |S \cap U_{j_l}| \geqslant 0. \tag{21}$$

Then

$$\sum_{1 \leqslant j_1 < \cdots < j_l \leqslant t} (2d_{j_1, j_2, \dots, j_l} - 1)|S \cap U_{j_1}||S \cap U_{j_2}| \cdots |S \cap U_{j_l}| = d_1^* + d_2^*,$$

where

$$d_1^* = \sum_{1 \leqslant j_1 < \cdots < j_{l-1} \leqslant t, j_\alpha \neq j} (2d_{j, j_1, \dots, j_{l-1}} - 1)|S \cap U_j||S \cap U_{j_1}| \cdots |S \cap U_{j_{l-1}}| \tag{22}$$

and

$$d_2^* = \sum_{1 \leqslant j_1 < \cdots < j_l \leqslant t, j_\alpha \neq j} (2d_{j_1,j_2,\ldots,j_l} - 1)|S \cap U_{j_1}||S \cap U_{j_2}| \cdots |S \cap U_{j_l}|. \tag{23}$$

Then

$$d_1^* = |S \cap U_j| \sum_{1 \leqslant j_1 < \cdots < j_{l-1} \leqslant t, j_\alpha \neq j} (2d_{j,j_1,\ldots,j_{l-1}} - 1)|S \cap U_{j_1}| \cdots |S \cap U_{j_{l-1}}|. \tag{24}$$

If

$$\sum_{1 \leqslant j_1 < \cdots < j_{l-1} \leqslant t, j_\alpha \neq j} (2d_{j,j_1,\ldots,j_{l-1}} - 1)|S \cap U_{j_1}| \cdots |S \cap U_{j_{l-1}}| \geqslant 0$$

then

$$\begin{aligned} &|S \cap U_j| \sum_{1 \leqslant j_1 < \cdots < j_{l-1} \leqslant t, j_\alpha \neq j} (2d_{j,j_1,\ldots,j_{l-1}} - 1)|S \cap U_{j_1}| \cdots |S \cap U_{j_{l-1}}| \\ &\quad \leqslant |U_j| \sum_{1 \leqslant j_1 < \cdots < j_{l-1} \leqslant t, j_\alpha \neq j} (2d_{j,j_1,\ldots,j_{l-1}} - 1)|S \cap U_{j_1}| \cdots |S \cap U_{j_{l-1}}|. \end{aligned}$$

In this case,

$$d^*(S) \leqslant d^*(S \cup U_j). \tag{25}$$

If

$$\sum_{1 \leqslant j_1 < \cdots < j_{l-1} \leqslant t, j_\alpha \neq j} (2d_{j,j_1,\ldots,j_{l-1}} - 1)|S \cap U_{j_1}| \cdots |S \cap U_{j_{l-1}}| < 0,$$

then

$$|S \cap U_j| \sum_{1 \leqslant j_1 < \cdots < j_{l-1} \leqslant t, j_\alpha \neq j} (2d_{j,j_1,\ldots,j_{l-1}} - 1)|S \cap U_{j_1}| \cdots |S \cap U_{j_{l-1}}| \leqslant 0, \tag{26}$$

and so

$$d^*(S) \leqslant d^*(S \backslash U_j). \tag{27}$$

Combining (25) with (27) gives

$$d^*(S) \leqslant \max(d^*(S \cup U_j), d^*(S \backslash U)).$$

In case, the term on the right-hand side of (21) is negative the proof can be repeated with minor changes. □

We will next show that $d^*$ is a "good" approximation of $d$.

**Lemma 16.** *For every set $S \subset [n]$, we have*

$$|d(S) - d^*(S)| \leqslant 9\varepsilon n^l.$$

**Proof.** Let us first estimate the number of edges incident to $U_0$ and the number of edges that are not crossing in $U_1,\dots,U_t$. Since $d(S)=|h_1(S)-h_{-1}(S)|$, we have

$$\left|d(S)-\left|\sum_{1\leqslant j_1<j_2<\cdots<j_l\leqslant t}(h_1(S\cap U_{j_1},\dots,S\cap U_{j_l})-h_{-1}(S\cap U_{j_1},\dots,S\cap U_{j_l}))\right|\right|$$

$$\leqslant |U_0|n^{l-1}+t\begin{pmatrix}\frac{n}{t}\\ c2\end{pmatrix}n^{l-2}. \tag{28}$$

Since $1/t\leqslant\varepsilon$ and $|U_0|\leqslant\varepsilon n$, we can further estimate the right-hand side of (28)

$$\left|d(S)-\left|\sum_{1\leqslant j_1<j_2<\cdots<j_l\leqslant t}(h_1(S\cap U_{j_1},\dots,S\cap U_{j_l})-h_{-1}(S\cap U_{j_1},\dots,S\cap U_{j_l}))\right|\right|$$

$$\leqslant\frac{3\varepsilon}{2}n^l. \tag{29}$$

Now, we will disregard edges in irregular $l$-tuples and edges adjacent to "small intersections". We have

$$|h_1(S\cap U_{j_1},\dots,S\cap U_{j_l})-h_{-1}(S\cap U_{j_1},\dots,S\cap U_{j_l})|\leqslant|S\cap U_{j_1}|\cdots|S\cap U_{j_l}|,$$

which in case when for some $j_\alpha$, $|S\cap U_{j_\alpha}|\leqslant\varepsilon|U_{j_\alpha}|$ gives

$$|h_1(S\cap U_{j_1},\dots,S\cap U_{j_l})-h_{-1}(S\cap U_{j_1},\dots,S\cap U_{j_l})|\leqslant\varepsilon\frac{n^l}{t^l}. \tag{30}$$

Since we have at most $\varepsilon t^l$, $\varepsilon$-irregular $l$-tuples $(U_{j_1},\dots,U_{j_l})$, the right-hand side of (29) can be further estimated as follows.

$$\left|d(S)-\left|\sum(h_1(S\cap U_{j_1},\dots,S\cap U_{j_l})-h_{-1}(S\cap U_{j_1},\dots,S\cap U_{j_l}))\right|\right|$$

$$\leqslant\varepsilon t^l\frac{n^l}{t^l}+t^l\frac{\varepsilon n^l}{t^l}+\frac{3\varepsilon}{2}n^l, \tag{31}$$

where the summation is taken over set $R$ of all $l$-sets $\{j_1,\dots,j_l\}$ such that
1. $(U_{j_1},\dots,U_{j_l})$ is $\varepsilon$-regular, and
2. $|S\cap U_{j_1}|\geqslant\varepsilon|U_{j_1}|,\dots,|S\cap U_{j_l}|\geqslant\varepsilon|U_{j_l}|$.

Thus, we have

$$\left|d(S)-\left|\sum_R(h_1(S\cap U_{j_1},\dots,S\cap U_{j_l})-h_{-1}(S\cap U_{j_1},\dots,S\cap U_{j_l}))\right|\right|\leqslant\frac{7\varepsilon}{2}n^l. \tag{32}$$

In the similar way, we can show that

$$\left|d^*(S)-\sum_R(2d_{j_1,\dots,j_l}-1)|S\cap U_{j_1}|\cdots|S\cap U_{j_l}|\right|\leqslant\frac{7\varepsilon}{2}n^l. \tag{33}$$

Note that if $(U_{j_1},\dots,U_{j_l})$ is $\varepsilon$-regular in $H_1$ with density $d_{j_1,\dots,j_l}$ then it is also $\varepsilon$-regular in $H_{-1}$ with density $1-d_{j_1,\dots,j_l}$. Thus,

$$\begin{aligned}
&h_1(S\cap U_{j_1},\dots,S\cap U_{j_l})-h_{-1}(S\cap U_{j_1},\dots,S\cap U_{j_l})\\
&\quad=d(S\cap U_{j_1},\dots,S\cap U_{j_l})|S\cap U_{j_1}|\cdots|S\cap U_{j_l}|\\
&\qquad-(1-d(S\cap U_{j_1},\dots,S\cap U_{j_l}))|S\cap U_{j_1}|\cdots|S\cap U_{j_l}|\\
&\quad=(2d(S\cap U_{j_1},\dots,S\cap U_{j_l})-1)|S\cap U_{j_1}|\cdots|S\cap U_{j_l}|.
\end{aligned}$$

In case $\{j_1,\dots,j_l\}\in R$,

$$|d(S\cap U_{j_1},\dots,S\cap U_{j_l})-d_{j_1,\dots,j_l}|\leqslant\varepsilon.$$

Therefore,

$$\begin{aligned}
&\left|\left|\sum_R(2d(S\cap U_{j_1},\dots,S\cap U_{j_l})-1)|S\cap U_{j_1}|\cdots|S\cap U_{j_l}|\right|\right.\\
&\quad\left.-\left|\sum_R(2d_{j_1,\dots,j_l}-1)|S\cap U_{j_1}|\cdots|S\cap U_{j_l}|\right|\right|\\
&\quad\leqslant\left|\sum_R 2(d(S\cap U_{j_1},\dots,S\cap U_{j_l})-d_{j_1,\dots,j_l})|S\cap U_{j_1}|\cdots|S\cap U_{j_l}|\right|\\
&\quad\leqslant 2\varepsilon t^l\left(\frac{n}{t}\right)^l=2\varepsilon n^l.
\end{aligned}\tag{34}$$

Combining (32), (33), with (34) yields

$$|d(S)-d^*(S)|\leqslant 9\varepsilon n^l.\qquad\square$$

To finish the proof of Theorem 3, we observe that if $S^*$ is the set found by Algorithm 2 then by Fact 15, $d^*(S^*)\geqslant d^*(S)$ for every set $S\subset[n]$. We denote by $\bar{S}$ an optimal set, that is

$$d(\bar{S})=disc(\chi).\tag{35}$$

Then

$$d(S^*)\geqslant d^*(S^*)-|d^*(S^*)-d(S^*)|,$$

which by Lemma 16 gives

$$d(S^*)\geqslant d^*(S^*)-9\varepsilon n^l.\tag{36}$$

Using Fact 15, we have

$$d(S^*)\geqslant d^*(\bar{S})-9\varepsilon n^l,\tag{37}$$

and so by applying Lemma 16 again, we get

$$d(S^*)\geqslant d(\bar{S})-18\varepsilon n^l=disc(\chi)-\eta n^l.\qquad\square$$

## References


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

[2] S. Arora, D. Krager, M. Karpinski, Polynomial time approximation schemes for dense instances of NP-hard problems, Proceedings of the 27th Annual ACM Symposium on Theory of Computing, 1995, pp. 284–293.

[3] S. Arora, C. Lund, R. Motwani, M. Sudan, M. Szegedy, Proof verification and hardness of approximation problems, Proceedings of the 33rd IEEE Symposium on Foundations of Computing, 1992, pp. 14–23.

[4] A. Czygrinow, S. Poljak, V. Rödl, Constructive Quasi–Ramsey numbers and tournament ranking, SIAM J. Discrete Math. 12 (1) (1999) 48–63.

[5] A. Czygrinow, V, Rödl, An algorithmic regularity lemma for hypergraphs, SIAM J. Comput., to appear.

[6] R.A. Duke, H. Lefmann, V. Rödl, A fast algorithm for computing frequencies in a given graph, SIAM J. Comput. 24 (3) (1995) 598–620.

[7] W. Fernandez de la Vega, MAX-CUT has an approximation scheme in Dense graphs, Random Structures and Algorithms 8(3), (1996) 187–198.

[8] P. Frankl, V. Rödl, The uniformity lemma for hypergraphs, Graphs Combin. 8 (1992) 309–312.

[9] P. Frankl, V. Rödl, Extremal problems on set systems, preliminary version, 1996.

[10] A. Frieze, R. Kannan, The Regularity Lemma and Approximation schemes for dense problems, Proc. 37th FOCS, 1996, pp. 12–20.

[11] A. Frieze, R. Kannan, Quick Approximation to Matrices and Applications, Combinatorica 19 (1999) 175–220.

[12] M.R. Garey, D.S. Johnson, Computers and Intractability. A Guide to the Theory of NP-Completeness, W.H. Freeman, San Francisco, 1979.

[13] P. Haxell, V. Rödl, Integer and Fractional Packing in Dense Graphs, Preliminary Version, 1998.

[14] E. Szemerédi, Regular Partitions of Graphs, Colloques Internationaux C.N.R.S., Problemes Combinatories et Theorie des Graphes, 1978, pp. 399–402.