AN ALGORITHMIC REGULARITY LEMMA FOR HYPERGRAPHS

A. CZYGRINOW, V. RÖDL

Abstract. In this paper, we will consider the problem of designing an efficient algorithm that finds an ϵ-regular partition of an l-uniform hypergraph.

1. Introduction

The regularity lemma of Szemerédi [12] is a powerful tool used in extremal combinatorics and graph theory. The lemma states that the vertex set of any graph can be partitioned into subsets that meet certain regularity conditions. The original proof of the lemma is not constructive but recently, Alon, Duke, Lefmann, Rödl, and Yuster [1] found a way to convert it into an efficient algorithm. The algorithm is based on the characterization of regularity which states that a pair of subsets of vertices is irregular if and only if either the degrees of "many" vertices are "far from average" or the cardinality of the intersection of neighborhoods of many pairs of vertices is "far from the average case". The algorithmic version of the lemma has already been applied to design the algorithms for various combinatorial problems. Applications include the Max-Cut problem [7], the tournament ranking problem [3], or the fast algorithm for computing the frequency of a subgraph [4]. Many of these problems have appealing generalizations to hypergraphs, which leads to a natural question: Can the regularity lemma be extended to hypergraphs?

In the hypergraph case regularity can be measured in a few different ways. The most straightforward and perhaps natural approach defines the density and regularity in the same way as for graphs. The corresponding hypergraph regularity lemma can be then proved along the lines of Szemerédi's proof for graphs [11]. In this paper we consider an algorithmic version of this regularity lemma. Note that other versions of regularity lemma were considered in [2], [5], [6], and recently in [10]. In a proof of the regularity lemma it is necessary to distinguish between bipartite graphs which have uniformly distributed edges (say, which are ϵ-regular) from these which are not. For an algorithmic proof we need to have an efficient algorithm. It was proved however in [1] that it is co-NP complete to decide if a bipartite graph is ϵ-regular. Authors of [1] got around this difficulty by finding another polynomial checkable characterization which distinguishes between bipartite graphs which fail to be $\epsilon^{1/5}$-regular and those which are ϵ-regular. Note that in the hypergraph case we were unable to find even such a simple characterization.

We found it convenient to consider a slightly more general weighted version ($l = 2$ and $\omega : [V]^l \to \{0, 1\}$ gives the original Szemerédi's version). Let $H = (V, \omega)$ be an l-uniform hypergraph with nonnegative weights $\omega : [V]^l \to Z_+ \cup \{0\}$, and let $K = \max_{\{v_1,\dots,v_l\} \in [V]^l} \omega(v_1, \dots, v_l) + 1$ (for technical reason we require that K is a strictly larger than $\max_{\{v_1,\ldots,v_l\}\in[V]^l}\omega(v_1,\ldots,v_l)$). For l subsets $V_1,\ldots,V_l$ of V such that $V_i\cap V_j=\emptyset$ if $i\neq j$ define

$$d_\omega(V_1,\ldots,V_l)=\frac{\sum\{\omega(v_1,\ldots,v_l):(v_1,\ldots,v_l)\in V_1\times\cdots\times V_l\}}{K|V_1|\ldots|V_l|}. \tag{1}$$

An l-tuple $(V_1,\ldots,V_l)$ of subsets of V with $V_i\cap V_j=\emptyset$ is called (ϵ,ω)-regular if for every $W_i\subset V_i$, $i=1,\ldots,l$, with $|W_i|\geq\epsilon|V_i|$ we have

$$|d_\omega(V_1,\ldots,V_l)-d_\omega(W_1,\ldots,W_l)|<\epsilon. \tag{2}$$

An l-tuple $(W_1,\ldots,W_l)$ that violates (2) is called a witness. A partition $V_0\cup V_1\cup\cdots\cup V_t$ of V is called (ϵ,ω)-regular if

1. $|V_0|\leq\epsilon|V|$,
2. $|V_i|=|V_j|$ for all $i,j\in[t]$,
3. all but at most ϵt^l l-tuples $(V_{i_1},\ldots,V_{i_l})$ with $\{i_1,\ldots,i_l\}\subset[t]^l$ are (ϵ,ω)-regular.

The regularity lemma for hypergraphs states that for every $\epsilon>0$ and every integer m there exist $M=M(\epsilon,m)$ and $N=N(\epsilon,m)$ such that every hypergraph $H=(V,\omega)$ with $|V|\geq N$ admits an (ϵ,ω)-regular partition $V_0\cup V_1\cup\cdots\cup V_t$ with $m\leq t\leq M$. We show the following:

Theorem 1. *For every l, K, m, and ϵ there exist M, L and an algorithm which for any l-uniform weighted hypergraph $H=(V,\omega)$ with $K=\max\omega(v_1,\ldots,v_l)+1$ and $|V|=n\geq L$ finds in $O(n^{2l-1}\log^2 n)$ time an (ϵ,ω)-regular partition $V_0\cup V_1\cup\cdots\cup V_t$ of H with $m\leq t\leq M$.*

Similar results were obtained in Frieze and Kannan [8] (compare also [7]). In particular, [8] contains a randomized algorithm which for every $\epsilon>0$ and every $\delta>0$ finds a subset which with probability $1-\delta$ contains all of the information necessary to construct an ϵ-regular partition. Also [7] and [9] contain many applications of the constructive graph and hypergraph regularity lemma. The approach taken by Frieze and Kannan in [7], [8], and [9] is different from ours. In addition to (1), the following density function will be used. Let $(V_1,\ldots,V_l)$ be an l-tuple of subsets with $V_i\cap V_j=\emptyset$, and let $1\leq k\leq l-1$. For $x\in V_{k+1}\times\cdots\times V_l$ define

$$d_\omega(x,V_1,\ldots,V_k)=\frac{\sum\{\omega(v_1,\ldots,v_k,x):(v_1,\ldots,v_k)\in V_1\times\cdots\times V_k\}}{K|V_1|\ldots|V_k|}.$$

For an l-tuple $(V_1,\ldots,V_l)$ of pairwise disjoint sets and for $1\leq k\leq l-1$ define

$$ind_k(V_1,\ldots,V_l)=\sum_{x\in V_{k+1}\times\cdots\times V_l}\frac{(d_\omega(x,V_1,\ldots,V_k))^2}{|V_{k+1}|\ldots|V_l|} \tag{3}$$

and

$$ind_l(V_1,\ldots,V_l)=\sum_{x\in V_{l-1}}\frac{(d_\omega(x,V_1,\ldots,V_{l-2},V_l))^2}{|V_{l-1}|}. \tag{4}$$

Note that

$$ind_{l-1}(V_1,\ldots,V_{l-2},V_l,V_{l-1})=ind_l(V_1,\ldots,V_{l-1},V_l). \tag{5}$$

Finally, define an index of a partition $P = V_0 \cup V_1 \cup \cdots \cup V_t$ of V, as follows.

$$ind(P) = \frac{1}{lt^l} \sum_{(V_{i_1},\ldots,V_{i_l})} (\sum_{k=1}^{l} ind_k(V_{i_1},\ldots,V_{i_l}))$$

Clearly, $ind(P)$ is always less than or equal to 1. In the same way as in [1] and [12], if a partition of V is (ϵ,ω)- irregular then one can construct a subpartition P' such that $ind(P') \geq ind(P) + \frac{\delta^{17}}{16 \cdot l}$, where $\delta = \frac{1}{96}\frac{\epsilon^{2l+1}}{2^{(2l+1)(l-1)}}$. Then iterating our "improvement" process $\frac{16 \cdot l}{\delta^{17}} + 1$ times we obtain an (ϵ,ω)-regular partition of the hypergraph. The rest of the paper is organized as follows. In Section 2 we prove some facts about hypergraph densities. Section 3 and Section 4 contain two procedures that construct "witness sets" for irregular l-tuples. Both procedures are applied in Section 5 where we give the description of the algorithm that constructs an (ϵ,ω)-regular partition of a hypergraph. Section 6 contains the analysis of the algorithm. In Section 7, we outline two applications of Theorem 1. Finally, it should be noted that there was no attempt made to optimize the constants.

2. Preliminary Facts

Let us first observe the following property of densities, compare with [12]. Note that l is a fixed constant independent of n and $1 \leq i \leq l$.

Fact 2. *(Continuity of densities)*

1. *Let $G = (X, Y, \omega)$ be a weighted bipartite graph. For $\delta \in (0,1)$, let $X' \subset X$, $Y' \subset Y$ be such that $|X'| \geq (1-\delta)|X|$ and $|Y'| \geq (1-\delta)|Y|$. Then*

$$|d_\omega(X', Y') - d_\omega(X, Y)| \leq 2\delta$$

and

$$|(d_\omega(X', Y'))^2 - (d_\omega(X, Y))^2| \leq 4\delta.$$

2. *Let $(V_1 \cup \cdots \cup V_l, \omega)$ be an l-uniform hypergraph with $V_i \cap V_j = \emptyset$. For $\delta \in (0,1)$, let $\bar{V}_1 \subset V_1, \ldots, \bar{V}_i \subset V_i$ be such that $|\bar{V}_j| \geq (1-\delta)|V_j|$, where $j = 1,\ldots,i$. Then for $x \in V_{i+1} \times \cdots \times V_l$,*

$$|d_\omega(x, \bar{V}_1, \ldots, \bar{V}_i) - d_\omega(x, V_1, \ldots, V_i)| \leq i\delta$$

and

$$|(d_\omega(x, \bar{V}_1, \ldots, \bar{V}_i))^2 - (d_\omega(x, V_1, \ldots, V_i))^2| \leq 2i\delta.$$

In many places of the proof, we will use the following defect form of the Schwarz inequality, see [12].

Fact 3. *If for some $m < n$, $\sum_{k=1}^{m} X_k = \frac{m}{n}\sum_{k=1}^{n} X_k + \rho$ then*

$$\sum_{k=1}^{n} X_k^2 \geq \frac{1}{n}(\sum_{k=1}^{n} X_k)^2 + \frac{\rho^2 n}{m(n-m)}.$$

Let $X = \{x_1, x_2, \dots, x_N\}$ and $U = \{u_1, u_2, \dots, u_M\}$ be two disjoint sets. Let $G = (X, U, \omega)$ be a bipartite graph with nonnegative weights on edges $\omega(x, u) < K$. For $i = 1, 2, \dots, M$, $j = 1, 2, \dots, N$ set

$$deg(x_j) = \sum_{i=1}^{M} \omega(x_j, u_i),$$

$$d_j = \frac{deg(x_j)}{MK},$$

$$deg(u_i) = \sum_{j=1}^{N} \omega(x_j, u_i),$$

$$\Delta_i = \frac{deg(u_i)}{NK}.$$

We also set

$$d = d_\omega(X, U) = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} \omega(x_j, u_i)}{MNK} = \frac{\sum_{j=1}^{N} d_j}{N} = \frac{\sum_{i=1}^{M} \Delta_i}{M}$$

and $T = dMNK$ for the total weight of all edges. For $x_i \in X$ consider the vector

$$\overrightarrow{x_i} = (\omega(x_i, u_k))_{k=1}^{M}.$$

Definition 4. A graph $G = (X, U, \omega)$ is called δ -vector regular if $|\langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle - d_i d_j MK^2| < \delta MK^2$ for all but at most $\delta \binom{N}{2}$ pairs $\{x_i, x_j\}$.

Lemma 5. *For $\epsilon, \delta \in (0,1)$, suppose $G = (X, U, \omega)$ is δ-vector regular and $|X| \geq 1/\delta$. Then for every $U' \subset U$ with $|U'| \geq \epsilon |U|$*

$$|d_\omega(X, U') - d_\omega(X, U)| < \epsilon,$$

provided $\epsilon^3 \geq 3\delta$.

Proof. Suppose the lemma is false and consider $U' \subset U$ such that $m = |U'| \geq \epsilon |U|$ and

$$|d_\omega(X, U') - d_\omega(X, U)| \geq \epsilon. \tag{6}$$

Without loss of generality, assume that $U' = \{u_1, \dots, u_m\}$, where $m \geq \epsilon M$. Condition (6) is equivalent to

$$|\sum_{i=1}^{m} \Delta_i - dm| \geq \epsilon m. \tag{7}$$

From Fact 3 (applied with $\rho = \epsilon m$ and $n = M$) we infer that

$$\sum_{i=1}^{M} \Delta_i^2 \geq M(d^2 + \frac{\epsilon^3}{1-\epsilon}) \tag{8}$$

which implies

$$\sum_{i=1}^{M} \binom{deg(u_i)}{2} = \sum_{i=1}^{M} \binom{\Delta_i NK}{2} = \sum_{i=1}^{M} \frac{\Delta_i NK(\Delta_i NK - 1)}{2} \geq$$

$$\frac{1}{2}N^2MK^2(d^2+\frac{\epsilon^3}{1-\epsilon})-\frac{NK}{2}\sum_{i=1}^{M}\Delta_i=\frac{1}{2}N^2MK^2(d^2+\frac{\epsilon^3}{1-\epsilon})-\frac{T}{2}. \tag{9}$$

On the other hand, we have

$$\sum_{i=1}^{M}\binom{deg(u_i)}{2}=\sum_{i=1}^{M}\binom{\sum_{j=1}^{N}\omega(x_j,u_i)}{2}=\frac{1}{2}\sum_{i=1}^{M}(\sum_{j=1}^{N}\omega(x_j,u_i)(\sum_{k=1}^{N}\omega(x_k,u_i)-1))$$

$$\leq\frac{1}{2}(\sum_{j,k\in[N]}(\sum_{i=1}^{M}\omega(x_k,u_i)\omega(x_j,u_i))-T)=\frac{1}{2}(\sum_{j,k\in[N]}\langle\overrightarrow{x_j},\overrightarrow{x_k}\rangle-T). \tag{10}$$

By assumption, we know that for all but at most $\delta\binom{N}{2}$ of pairs $\{x_i,x_j\}$, $\langle\overrightarrow{x_i},\overrightarrow{x_j}\rangle\leq(d_id_j+\delta)MK^2$ and we can bound the scalar product for the remaining pairs: $\langle\overrightarrow{x_i},\overrightarrow{x_j}\rangle<MK^2$. Therefore,

$$\sum_{j,k\in[N]}\langle\overrightarrow{x_j},\overrightarrow{x_k}\rangle<\sum_{j,k\in[N]}(d_jd_k+\delta)MK^2+\sum_{j=1}^{N}\langle\overrightarrow{x_j},\overrightarrow{x_j}\rangle+\delta N^2MK^2$$

$$\leq\sum_{j,k\in[N]}(d_jd_k+\delta)MK^2+2\delta N^2MK^2$$

$$\leq(\sum_{i=1}^{N}d_i)^2MK^2+3\delta N^2MK^2\leq(d^2+3\delta)N^2MK^2. \tag{11}$$

By combining (10) and (11), we see that

$$\sum_{i=1}^{M}\binom{deg(u_i)}{2}<\frac{1}{2}((d^2+3\delta)N^2MK^2-T). \tag{12}$$

Comparing (9) and (12) gives

$$(d^2+3\delta)N^2MK^2>N^2MK^2(d^2+\frac{\epsilon^3}{1-\epsilon})$$

which gives $3\delta>\frac{\epsilon^3}{1-\epsilon}$. This contradicts our assumption that $\epsilon^3\geq3\delta$. □

Let $V_1,\ldots,V_k$ be pairwise disjoint sets and let $(V_1\cup\cdots\cup V_k,\omega)$ be a weighted k-uniform hypergraph. We consider a bipartite graph (U,X,ω) where $U=V_1$, $X=\{x_1,\ldots,x_N\}=V_2\times\cdots\times V_k$ and where $\omega(u,x)=\omega(v_1,v_2,\ldots,v_k)$ if $u=v_1$ and $x=(v_2,\ldots,v_k)$. The next lemma shows that irregularity of an l-tuple can be reduced either to vector irregularity or to irregularity of an $(l-1)$-tuple, with weights appropriately defined.

Lemma 6. *If (X,U) be the bipartite graph defined above which moreover satisfies the following conditions:*

1. $|\langle\overrightarrow{x_i},\overrightarrow{x_j}\rangle-d_id_jMK^2|<\delta MK^2$ *for all but at most* $\delta\binom{N}{2}$ *pairs* $\{x_i,x_j\}$.
2. $|d_\omega(V_1,V_2',\ldots,V_k')-d_\omega(V_1,V_2,\ldots,V_k)|<\frac{\epsilon}{2}$ *for every* $V_i'\subset V_i$ *with* $|V_i'|\geq\frac{\epsilon}{2}|V_i|$.

then for every $V_i' \subset V_i$ *with* $|V_i'| \geq \epsilon|V_i|$

$$|d_\omega(V_1', V_2', \ldots, V_k') - d_\omega(V_1, V_2, \ldots, V_k)| < \epsilon$$

provided $48\delta \leq \epsilon^{2k+1}$.

Proof. The triangle inequality and the second condition imply that for every $V_i' \subset V_i$ with $|V_i'| \geq \epsilon|V_i|$

$$|d_\omega(V_1', V_2', \ldots, V_k') - d_\omega(V_1, V_2, \ldots, V_k)| \leq \frac{\epsilon}{2} + |d_\omega(V_1', V_2', \ldots, V_k') - d_\omega(V_1, V_2', \ldots, V_k')|.$$

Let $X' = V_2' \times \cdots \times V_k'$. Observe that for at most $\delta\binom{N}{2} \leq \delta\frac{N^2}{2} \leq \frac{\delta}{\epsilon^{2(k-1)}}\frac{|X'|^2}{2} \leq \frac{2\delta}{\epsilon^{2(k-1)}}\binom{|X'|}{2}$ of the pairs $\{x_i, x_j\}$

$$|\langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle - d_i d_j MK^2| \geq \delta MK^2$$

and so Lemma 5 (applied to X') implies that

$$|d_\omega(V_1', X') - d_\omega(V_1, X')| < \frac{\epsilon}{2}$$

as $6\frac{\delta}{\epsilon^{2(k-1)}} \leq (\frac{\epsilon}{2})^3$. Clearly $d_\omega(Y, X') = d_\omega(Y, V_2', \ldots, V_k')$ which shows that

$$|d_\omega(V_1', V_2', \ldots, V_k') - d_\omega(V_1, V_2, \ldots, V_k)| < \epsilon.$$

□

Denote by $\omega'(v_2, \ldots, v_k) = \sum_{v_1 \in V_1} \omega(v_1, v_2, \ldots, v_k)$ and by $K' = K|V_1|$. Then from Lemma 6 we see that if a k-tuple $(V_1, V_2, \ldots, V_k)$ is (ϵ, ω)- irregular then either $|\langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle - d_i d_j MK^2| > \delta MK^2$ for at least δN^2 pairs $\{x_i, x_j\}$, with $\delta = \epsilon^{2k+1}/48$, or the $(k-1)$-tuple $(V_2, \ldots, V_k)$ is $(\frac{\epsilon}{2}, \omega')$-irregular.

3. Finding witnesses of vector irregularity

Let (X, U, ω) be a weighted "vector irregular" bipartite graph with $X = \{x_1, \ldots, x_N\}$ and $U = \{u_1, \ldots, u_M\}$. In this section we will show how to construct sets $X' \subset X$ and $U' \subset U$ such that $d_\omega(X', U')$ essentially differs from $d_\omega(X', U)$.

Theorem 7. *Let* $\delta \leq \frac{1}{3}$ *and assume that for at least* $\delta\binom{|X|}{2}$ *of the pairs* $\{x_i, x_j\}$ *we have* $|\langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle - d_i d_j MK^2| \geq \delta MK^2$. *Then there is a* $O(N^2 M \log^2 K)$ *algorithm that finds sets* $X' \subset X$ *with* $|X'| > \frac{\delta^7}{2}|X|$ *and* $U' \subset U$ *with* $|U'| > \delta^6|U|$ *such that*

$$|d_\omega(X', U') - d_\omega(X', U)| > \delta^2.$$

Proof. Set $\epsilon = \delta^3$ (we will assume that $1/\epsilon$ is an integer), and partition the set X into sets X_r in the following way:

$$X_r = \{x \in X : \epsilon rKM \leq deg(x) < \epsilon(r+1)KM\}, \tag{13}$$

for $r = 0, \ldots, \frac{1}{\epsilon} - 1$. Since at least $\delta\binom{|X|}{2}$ of the pairs $\{x_i, x_j\}$ satisfy $|\langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle - d_i d_j MK^2| \geq \delta MK^2$, we can find $x_0 \in X$ such that for at least $\frac{\delta}{2}N$ of the x_j's

$$\langle \overrightarrow{x_0}, \overrightarrow{x_j} \rangle - d_0 d_j MK^2 \geq \delta MK^2 \tag{14}$$

or

$$-(\langle \overrightarrow{x_0}, \overrightarrow{x_j} \rangle - d_0 d_j M K^2) \geq \delta M K^2. \tag{15}$$

We assume (14), and note that in the case of (15), the proof can be repeated with minor changes. Then, we can find $r \in \{0, \ldots, \frac{1}{\epsilon} - 1\}$ such that for at least $\frac{\epsilon\delta}{2}N$ of x_j's in X_r, $\langle \overrightarrow{x_0}, \overrightarrow{x_j} \rangle - d_0 d_j M K^2 \geq \delta M K^2$. Set $\bar{X} = \{x_j \in X_r : \langle \overrightarrow{x_j}, \overrightarrow{x_0} \rangle \geq (d_0 d_j + \delta) M K^2\}$, and observe that $|\bar{X}| \geq \frac{\epsilon\delta}{2}N$. In addition to the partition $X = \bigcup X_r$, we compute a partition of $U = \bigcup U_s$ where

$$U_s = \{u \in U : \epsilon s K \leq \omega(x_0, u) < \epsilon(s+1)K\} \tag{16}$$

with $s = 0, \ldots, \frac{1}{\epsilon} - 1$.

Claim 8. $\sum_{s=0}^{1/\epsilon - 1} (s+1)\epsilon K |U_s| \leq (d_0 + \epsilon) M K$.

Proof.

$$\sum_{s=0}^{1/\epsilon-1} (s+1)\epsilon K |U_s| = \sum s\epsilon K |U_s| + \epsilon K \sum |U_s| \leq deg(x_0) + \epsilon M K = (d_0 + \epsilon) M K.$$

□

Claim 9. *For every $x_j \in \bar{X}$*

$$\sum_{s; |U_s| > \epsilon^2 M} (s+1)\epsilon K \sum_{u_i \in U_s} \omega(x_j, u_i) \geq (d_0 d_j + \delta - \epsilon) M K^2.$$

Proof. In view of the definition of $\bar{X}$, for every $x_j \in \bar{X}$ the following holds.

$$(d_0 d_j + \delta) M K^2 \leq \langle \overrightarrow{x_0}, \overrightarrow{x_j} \rangle \leq \sum_{s=0}^{1/\epsilon-1} (s+1)\epsilon K \sum_{u_i \in U_s} \omega(x_j, u_i) \leq$$

$$\sum_{s; |U_s| > \epsilon^2 M} (s+1)\epsilon K \sum_{u_i \in U_s} \omega(x_j, u_i) + \epsilon^3 K^2 M \sum_s (s+1) \leq$$

$$\sum_{s; |U_s| > \epsilon^2 M} (s+1)\epsilon K \sum_{u_i \in U_s} \omega(x_j, u_i) + \epsilon K^2 M.$$

□

Next, we will show the following claim.

Claim 10. *Fix $x_{j_0} \in \bar{X}$ (arbitrarily). Then there exists $\bar{s} \in \{0, \ldots, \frac{1}{\epsilon} - 1\}$ such that*

(i) $|U_{\bar{s}}| \geq \epsilon^2 M$ *and*
(ii) $\sum_{u_i \in U_{\bar{s}}} \omega(x_j, u_i) \geq |U_{\bar{s}}|(d_{j_0} + \delta^2 + \epsilon) K$ *holds for at least $\epsilon|\bar{X}|$ of the $x_j \in \bar{X}$.*

Proof. First, we show that for every $x_j \in \bar{X}$ there is $s \in \{0\frac{1}{\epsilon} - 1\}$ such that

$$|U_s| > \epsilon^2 M \tag{17}$$

and

$$\sum_{u_i \in U_s} \omega(x_j, u_i) \geq |U_s| \frac{(d_0 d_j + \delta - \epsilon)K}{d_0 + \epsilon}. \tag{18}$$

Indeed, assume that there exists j such that for every $|U_s| > \epsilon^2 M$ we have

$$\sum_{u_i \in U_s} \omega(x_j, u_i) < |U_s| \frac{(d_0 d_j + \delta - \epsilon)K}{d_0 + \epsilon}.$$

Then

$$\sum_{s; |U_s| > \epsilon^2 M} (s+1)\epsilon K \sum_{u_i \in U_s} \omega(x_j, u_i) < \sum_s (s+1)\epsilon K |U_s| \frac{(d_0 d_j + \delta - \epsilon)K}{d_0 + \epsilon}$$

which by Claim 8 is less than or equal to $(d_0 d_j + \delta - \epsilon)MK^2$. This however contradicts Claim 9.

Since $\epsilon = \delta^3$ and $\delta \leq \frac{1}{3}$, one can further simplify the right-hand side of (18) to infer that for every $x_j \in \bar{X}$ there exists s such that

$$|U_s| > \epsilon^2 M$$

and

$$\sum_{u_i \in U_s} \omega(x_j, u_i) \geq |U_s| K(d_j + \delta^2 + 2\epsilon). \tag{19}$$

It follows from the definition of X_r that for every $x_{j_1}, x_{j_2} \in \bar{X}$, $|d_{j_1} - d_{j_2}| \leq \epsilon$, which implies that

$$\sum_{u_i \in U_s} \omega(x_j, u_i) \geq |U_s|(d_{j_0} + \delta^2 + \epsilon)K, \tag{20}$$

if x_{j_0} is chosen arbitrarily from $\bar{X}$. Therefore, for every $x_{j_0} \in \bar{X}$ and every $x_j \in \bar{X}$ there is an $s \in \{0, \ldots, \frac{1}{\epsilon} - 1\}$ such that

$$|U_s| > \epsilon^2 M$$

and

$$\sum_{u_i \in U_s} \omega(x_j, u_i) \geq |U_s|(d_{j_0} + \delta^2 + \epsilon)K.$$

In order to prove (i) and (ii) we need to "reverse" the quantifiers of j and s. We know that for every j there is a "big" set U_s such that (20) holds. Since there are at most $\frac{1}{\epsilon}$ choices of s, there exists $\bar{s} \in \{0, \ldots, \frac{1}{\epsilon} - 1\}$ such that $U_{\bar{s}}$ is "big" and for at least $\epsilon|\bar{X}|$ of the x_j's (20) holds. More precisely, there exists $\bar{s} \in \{0, \ldots, \frac{1}{\epsilon} - 1\}$ such that

(i) $|U_{\bar{s}}| \geq \epsilon^2 M$ and
(ii) $\sum_{u_i \in U_{\bar{s}}} \omega(x_j, u_i) \geq |U_{\bar{s}}|(d_{j_0} + \delta^2 + \epsilon)K$ holds for at least $\epsilon|\bar{X}|$ of the $x_j \in \bar{X}$,

which proves the claim. □

Let $U' = U_{\bar{s}}$ and let X' be the set of those $\epsilon|\bar{X}|$ vertices from $\bar{X}$ that satisfy (20). Observe that $|U'| \geq \delta^6|U|$ and $|X'| \geq \frac{\delta^7}{2}|X|$. We have

$$d_\omega(X', U') = \frac{\sum_{x_j \in X', u_i \in U'} \omega(x_j, u_i)}{K|X'||U'|} \geq \frac{|U'||X'|(d_{j_0} + \delta^2 + \epsilon)K}{K|X'||U'|} = (d_{j_0} + \delta^2 + \epsilon)$$

and

$$d_\omega(X', U) = \frac{\sum_{x \in X'} deg(x)}{K|X'||U|} \leq \frac{|X'|(deg(x_{j_0}) + \epsilon KM)}{K|X'||U|} = d_{j_0} + \epsilon.$$

Therefore,

$$d_\omega(X', U') - d_\omega(X', U) \geq (d_{j_0} + \delta^2 + \epsilon) - d_{j_0} - \epsilon = \delta^2. \tag{21}$$

The above proof gives an efficient algorithm: First compute scalar products to find x_0 and partitions of X and U. Then check all sets U_s to find one that satisfies Claim 10. The main computational task is to compute $O(N^2)$ scalar products $\langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle = \sum_{l=1}^{M} \omega(x_i, u_l)\omega(x_j, u_l)$. Since $\omega(x_j, u_l) < K$ the multiplication $\omega(x_i, u_l)\omega(x_j, u_l)$ can be done in $O(\log^2 K)$ time. Thus the total number of steps is $O(N^2 M \log^2 K)$. □

4. Finding a witness of irregularity in a weighted bipartite graph

Let $G = (X, Y, \omega)$ be a bipartite graph with nonnegative weights on edges $\omega(x, y) < K$. In this section, we will show how to find $X' \subset X$ and $Y' \subset Y$ such that

$$|d_\omega(X', Y') - d_\omega(X, Y)| > \delta^2.$$

Unlike in the previous section neither X nor Y will be products of other sets. Note that the algorithm of this section is a generalization of the algorithm of [1] to weighted graphs. Let $M = |X| = |Y|$ and denote by $d = d_\omega(X, Y)$. We first observe the following fact.

Fact 11. *For $\rho \in (0,1)$ let $X^* = \{x \in X : |deg(x) - dMK| \geq \rho MK\}$. If $|X^*| \geq \rho M$ then there is $X^{**} \subset X^*$ with $|X^{**}| \geq \frac{\rho}{2} M$ such that*

$$|d_\omega(X^{**}, Y) - d_\omega(X, Y)| \geq \rho.$$

Proof. Let $X^{**} = \{x \in X^* : deg(x) - dMK \geq \rho MK\}$. Without loss of generality we can assume that $|X^{**}| \geq \frac{|X^*|}{2}$. Then,

$$d_\omega(X^{**}, Y) - d_\omega(X, Y) = \frac{\sum_{x \in X^{**}} deg(x) - d|X^{**}|MK}{|X^{**}|MK} \geq \rho.$$

□

Lemma 12. *For $\epsilon, \delta \in (0,1)$, $\delta^2 < \epsilon$, $M \geq \frac{1}{\delta}$, let $X^* = \{x \in X : |deg(x) - dMK| > \delta^2 MK\}$. If both of the following conditions are satisfied*

1. *$|X^*| < \delta^2 M$,*
2. *for all but at most $\delta\binom{M}{2}$ of the pairs $\{x_i, x_j\}$ of vertices from X, $|\langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle - d^2K^2M| < \delta K^2 M$,*

then for every $X' \subset X$ with $|X'| \geq \epsilon M$ and every $Y' \subset Y$ with $|Y'| \geq \epsilon M$ we have

$$|d_\omega(X', Y') - d_\omega(X, Y)| \leq 2\frac{\delta^2}{\epsilon} + \frac{\sqrt{5\delta}}{\epsilon^2 - \epsilon\delta^2}.$$

Proof. Fix $X' \subset X$ and $Y' \subset Y$ with $|X'| \geq \epsilon|X|$ and $|Y'| \geq \epsilon|Y|$. Let $X'' = X' \setminus X^*$. Note that $|X''| > (1 - \frac{\delta^2}{\epsilon})|X'|$. Without loss of generality, we can assume that $X'' = \{x_1, \dots, x_m\}$ and $Y' = \{y_1, \dots, y_n\}$. For $i = 1, \dots, m$, we consider vectors $\overrightarrow{a_i} = \overrightarrow{x_i} - (dK, \dots, dK)$ and hence $\overrightarrow{a_i} = (a_{i1}, \dots, a_{iM})$ where

$$a_{ij} = \omega(x_i, y_j) - dK. \tag{22}$$

Then, due to the fact that

$$||\overrightarrow{a_1} + \cdots + \overrightarrow{a_m}||^2 = \sum_{i=1}^{m} ||\overrightarrow{a_i}||^2 + \sum_{i \neq j} \langle \overrightarrow{a_i}, \overrightarrow{a_j} \rangle$$

and from the form of a_i's, we see that

$$\langle \overrightarrow{a_i}, \overrightarrow{a_j} \rangle = \langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle - dK \sum_{l=1}^{M} \omega(x_i, y_l) - dK \sum_{l=1}^{M} \omega(x_j, y_l) + d^2K^2M$$

$$\leq \langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle - 2dK(dKM - \delta^2 KM) + d^2K^2M \leq \langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle + 2\delta^2 dK^2M - d^2K^2M.$$

Therefore, for all but at most $\delta\binom{M}{2}$ of $\{x_i, x_j\}$

$$\langle \overrightarrow{a_i}, \overrightarrow{a_j} \rangle \leq \delta K^2 M + 2\delta^2 dK^2 M \leq 3\delta K^2 M$$

and always

$$\langle \overrightarrow{a_i}, \overrightarrow{a_j} \rangle \leq K^2 M.$$

We infer that

$$||\overrightarrow{a_1} + \cdots + \overrightarrow{a_m}||^2 \leq \sum_{i=1}^{m} ||\overrightarrow{a_i}||^2 + 3\delta K^2 M^3 + \delta K^2 M^3$$

$$\leq K^2M^2 + 4\delta K^2 M^3 \leq 5\delta K^2 M^3,$$

as $M \geq 1/\delta$. On the other hand, we have $||\overrightarrow{a_1} + \cdots + \overrightarrow{a_m}||^2 = \xi_1^2 + \cdots + \xi_M^2$, where $\xi_i = a_{1i} + \cdots + a_{mi}$. Then $\xi_1^2 + \cdots + {\xi_M}^2 \geq \xi_1^2 + \cdots + \xi_n^2 \geq \frac{1}{n}(\xi_1 + \cdots + \xi_n)^2$ which implies

$$(\xi_1 + \cdots + \xi_n)^2 \leq 5\delta K^2 M^3 n \leq 5\delta K^2 M^4.$$

We infer that

$$|\xi_1 + \cdots + \xi_n| \leq \sqrt{5\delta} K M^2 \tag{23}$$

and by (22) we can write (23) as follows

$$\Big| \sum_{x \in X'', y \in Y'} \omega(x, y) - |X''||Y'|dK \ \Big| \leq \sqrt{5\delta} K M^2.$$

Therefore,

$$|d_\omega(X'', Y') - d| \leq \frac{\sqrt{5\delta} M^2}{|X''||Y'|}$$

and since $|X''| \geq (1 - \frac{\delta^2}{\epsilon})|X'|$ we infer by continuity of density (Fact 2) that

$$|d_\omega(X', Y') - d| \leq |d_\omega(X'', Y') - d| + |d_\omega(X'', Y') - d_\omega(X', Y')| \leq 2\frac{\delta^2}{\epsilon} + \frac{\sqrt{5\delta} M^2}{|X''||Y'|}. \tag{24}$$

Now, since $|Y'| \geq \epsilon M$ and $|X''| \geq (1 - \frac{\delta^2}{\epsilon})\epsilon M$ we can further estimate the right-hand side of (24) to obtain

$$|d_\omega(X',Y') - d| \leq 2\frac{\delta^2}{\epsilon} + \frac{\sqrt{5\delta}}{\epsilon^2 - \epsilon\delta^2}.$$

□

Theorem 13. *Let $G = (X, Y, \omega)$ be a weighted bipartite graph with $M = |X| = |Y|$ and $0 \leq \omega(x,y) < K$. For $\epsilon < 1/10$ and $\delta = \epsilon^7$, if $M > \frac{1}{\delta}$ then there exists an $O(M^3 \log^2 K)$ algorithm such that if (X,Y) is (ϵ,ω)-irregular then it finds $X' \subset X$, with $|X'| > \delta^4|X|$ and $Y' \subset Y$, with $|Y'| > \delta^4|Y|$ such that*

$$|d_\omega(X',Y') - d_\omega(X,Y)| > \delta^2.$$

Proof. First observe that due to the assumptions about δ and ϵ, $2\frac{\delta^2}{\epsilon} + \frac{\sqrt{5\delta}}{\epsilon^2-\epsilon\delta^2} \leq \epsilon$. Since (X,Y) is (ϵ,ω)-irregular we can infer from Lemma 12 that either

(i) $|X^*| \geq \delta^2 M$, or
(ii) for at least $\delta\binom{M}{2}$ pairs $\{x_i, x_j\}$, $|\langle \overrightarrow{x_i}, \overrightarrow{x_j}\rangle - d^2K^2M| \geq \delta K^2 M$.

In case of (i) set $X' = X^{**}$ from Fact 11, and $Y' = Y$. Then by Fact 11

$$|d_\omega(X',Y') - d_\omega(X,Y)| \geq \delta^2.$$

Observe that constructing the set X^{**} requires computing the degrees of $x_j \in X$ which can be done in $O(M^2 \log K)$ time. Hence, in this case we are done. Next, we will show how to construct the witness sets X' and Y' if (ii) holds while (i) is false. Assume that for at least $\delta\binom{M}{2}$ of pairs $\{x_i, x_j\}$, $|\langle \overrightarrow{x_i}, \overrightarrow{x_j}\rangle - d^2K^2M| > \delta K^2 M$ and that $|X^*| < \delta^2 M$. Then for at least $(\delta - 2\delta^2)\binom{M}{2}$ of pairs $\{x_i, x_j\}$, where $x_i, x_j \in X \setminus X^*$ we have $|\langle \overrightarrow{x_i}, \overrightarrow{x_j}\rangle - d^2K^2M| > \delta K^2 M$. Therefore, we can find $x_0 \in X \setminus X^*$ such that for at least $\frac{\delta - 2\delta^2}{2}M$ of $x_j \in X$ either

$$\langle \overrightarrow{x_0}, \overrightarrow{x_j}\rangle - d^2K^2M > \delta K^2 M \tag{25}$$

or

$$-(\langle \overrightarrow{x_0}, \overrightarrow{x_j}\rangle - d^2K^2M) > \delta K^2 M. \tag{26}$$

We assume (25) and set $\rho = \delta^2$. Let $\bar{X} = \{x_j \in X \setminus X^* : \langle \overrightarrow{x_0}, \overrightarrow{x_j}\rangle - d^2K^2M > \delta K^2 M\}$. We partition the set Y as follows:

$$Y_s = \{y \in Y : s\rho K \leq \omega(x_0, y) \leq (s+1)\rho K\} \tag{27}$$

where $s = 0, \ldots, \frac{1}{\rho} - 1$.

Fact 14. 1. $\sum_{s=0}^{1/\rho-1}(s+1)\rho K|Y_s| \leq (d + 2\rho)KM$.
2. *For every $x_j \in \bar{X}$*

$$\sum_{s;|Y_s| \geq \rho^2 M} (s+1)\rho K \sum_{y \in Y_s} \omega(x_j, y) \geq (d^2 + \delta - \rho)K^2 M. \tag{28}$$

Proof. From (27) and the fact that for any $x_0 \in X \setminus X^*$, $deg(x_0) \leq (d+\rho)KM$ we infer that

$$\sum_{s=0}^{1/\rho-1} (s+1)\rho K|Y_s| = \sum s\rho K|Y_s| + \rho K \sum |Y_s| \leq deg(x_0) + \rho KM \leq dKM + 2\rho KM$$

which shows the first part. To prove the second part, we observe that

$$\langle \vec{x_0}, \vec{x_j} \rangle = \sum_{y\in Y} \omega(x_0, y)\omega(x_j, y) \leq \sum_{s=0}^{1/\rho-1} (s+1)\rho K \sum_{y\in Y_s} \omega(x_j, y)$$

$$\leq \sum_{s;|Y_s|\geq\rho^2 M} (s+1)\rho K \sum_{y\in Y_s} \omega(x_j, y) + \rho^3 MK^2 \sum (s+1)$$

$$\leq \sum_{s;|Y_s|\geq\rho^2 M} (s+1)\rho K \sum_{y\in Y_s} \omega(x_j, y) + \rho MK^2. \tag{29}$$

Comparing (25) and (29) we infer the inequality (28). □

We will first show that for every $x_j \in \bar{X}$ there is s such that

$$|Y_s| \geq \rho^2 M \tag{30}$$

and

$$\sum_{y\in Y_s} \omega(x_j, y) \geq |Y_s| \frac{(d^2+\delta-\rho)K}{(d+2\rho)}. \tag{31}$$

Indeed, assume that there exists j such that for every s such that $|Y_s| \geq \rho^2 M$ we have $\sum_{y\in Y_s} \omega(x_j, y) < |Y_s| \frac{(d^2+\delta-\rho)K}{(d+2\rho)}$. Then averaging over all "big" Y_s we see that

$$\sum_{s;|Y_s|\geq\rho^2 M} (s+1)K\rho \sum_{y\in Y_s} \omega(x_j, y) < \sum_{s;|Y_s|\geq\rho^2 M} (s+1)\rho K|Y_s| \frac{(d^2+\delta-\rho)K}{(d+2\rho)},$$

which by Fact 14 part (1) is less than or equal to $(d^2+\delta-\rho)K^2M$. This however contradicts part (2) of Fact 14.

Since $\delta = \epsilon^7 < 1/10^7$, the right-hand side of (31) can be simplified:

$$\sum_{y\in Y_s} \omega(x_j, y) \geq |Y_s|K(d + \frac{\delta}{2}). \tag{32}$$

Thus for every $x_j \in \bar{X}$ there is $s \in \{0, \ldots, \frac{1}{\rho} - 1\}$ such that $|Y_s| \geq \rho^2 M$ and (32) holds. Similarly as in Section 3 we use the simple pigeonhole principle to "reverse" the quantifiers of j and s to obtain the existence of $\bar{s} \in \{0, \ldots, \frac{1}{\rho} - 1\}$ that satisfies

(i) $|Y_{\bar{s}}| \geq \rho^2 M$ and
(ii) for at least $\rho|\bar{X}|$ of x_j's from $\bar{X}$, $\sum_{y\in Y_{\bar{s}}} \omega(x_j, y) \geq |Y_{\bar{s}}|K(d + \frac{\delta}{2})$.

Let $Y' = Y_{\bar{s}}$ and let X' be the set of those x_j's that satisfy (ii). Then

$$d_\omega(X',Y') - d = \frac{\sum_{y\in Y', x_j\in X'} \omega(x_j,y)}{K|X'||Y'|} - d \geq \frac{|X'||Y_{\bar{s}}|K(d+\frac{\delta}{2})}{K|X'||Y'|} - d = \frac{\delta}{2} > \delta^2.$$

In order to construct X' and Y' we must compute the scalar products to check (25), compute the partition (27), and check condition (ii). Since $\omega(x_j,y) < K$ this can be done in $O(M^3 \log^2 K)$ time. □

5. The main algorithm

In this section, we will describe the main algorithm which finds a refinement of an (ϵ,ω)-irregular partition such that the value of the index of the refined partition is closer to one. Let us first outline the idea in the case $l = 3$. For each (ϵ,ω)-irregular triple (V_1, V_2, V_3) we either find (using the algorithm from Theorem 7) $V_1' \subset V_1$ and $X' \subset V_2 \times V_3$ such that

$$|d_{\omega_1}(X', V_1') - d_{\omega_1}(X', V_1)| \geq \delta_1,$$

or using the algorithm from Theorem 13, we find $V_2' \subset V_2$ and $V_3' \subset V_3$ such that

$$|d_{\omega_2}(V_2', V_3') - d_{\omega_2}(V_2, V_3)| \geq \delta_2,$$

where ω_1, ω_2, δ_1, δ_2 are defined below. In the first case, only V_1' is used to refine a given partition, in the second case both V_2' and V_3' are used.

Consider an l-uniform weighted hypergraph (V, H, ω) with $K = \max|\omega(v_1, v_2, \ldots, v_l)| + 1$. We introduce some additional notation: For an l-tuple $(V_1, \ldots, V_l)$

$$\omega_1(v_1,\ldots,v_l) = \omega(v_1,\ldots,v_l), \qquad K_1 = K,$$
$$\omega_2(v_2,\ldots,v_l) = \sum_{v_1\in V_1} \omega_1(v_1,v_2,\ldots,v_l), \qquad K_2 = |V_1|K_1.$$

In general,

$$\omega_i(v_i,\ldots,v_l) = \sum_{v_{i-1}\in V_{i-1}} \omega_{i-1}(v_{i-1},v_i,\ldots,v_l), \qquad K_i = |V_{i-1}|K_{i-1}. \tag{33}$$

Also, we set

$$\epsilon_k = \frac{\epsilon}{2^{k-1}}$$

and

$$\delta_i = \frac{1}{48}\epsilon_i^{2(l-i+1)+1}.$$

Also, denote by $X_k = V_{k+1} \times \cdots \times V_l$. Note that for every $i \in [l]$,

$$\delta_i \geq \frac{1}{48}\frac{\epsilon^{2l+1}}{2^{(2l+1)(l-1)}}. \tag{34}$$

We can now describe the procedure that "improves" a given partition P. For each (ϵ_1, ω_1)-irregular l-tuple, $(V_1, \ldots, V_l)$ consider the weighted bipartite graph (V_1, X_1, ω_1) with $X_1 = V_2 \times \cdots \times V_l$ and $\omega_1(v,x) = \omega_1(v, v_2, \ldots, v_l)$ if $x = (v_2, \ldots, v_l)$. Denote by $M = |V_1|$ and $N = |X_1|$. The algorithm is illustrated in Figure 1: Lemma 6 implies that

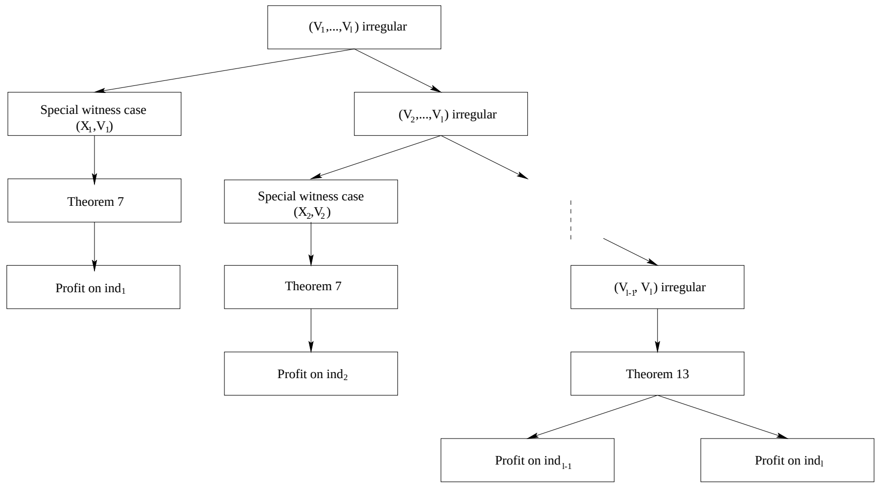

FIGURE 1. Proof of Theorem 1

if $(V_1, \ldots, V_l)$ is (ϵ_1, ω_1)-irregular then either (X_1, V_1) is vector irregular and for at least $\delta_1 \binom{N}{2}$ of pairs $\{x_i, x_j\}$ of vertices in X_1

$$|\langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle - K_1^2 M d_i d_j| > \delta_1 K_1^2 M, \tag{35}$$

or the $(l-1)$-tuple $(V_2, \ldots, V_l)$ is (ϵ_2, ω_2)-irregular. If (35) holds then we can use the algorithm from Theorem 7 to find $X_1' \subset X_1$ with $|X_1'| \geq \frac{\delta_1^7}{2}|X_1|$ and $V_1' \subset V_1$ with $|V_1'| \geq \delta_1^6 |V_1|$ such that

$$|d_{\omega_1}(X_1', V_1') - d_{\omega_1}(X_1', V_1)| > \delta_1^2. \tag{36}$$

Note that X_1' will not be used as a witness when improving the partition P, only V_1' will. In the case (35) does not hold we continue to apply Lemma 6 to the $(l-i)$- tuples $(V_i, \ldots, V_l)$ until $i = l-2$. Finally, if we haven't found a witness set so far, we apply Theorem 13 to the pair (V_{l-1}, V_l), in this case we find two subsets $V_{l-1}' \subset V_{l-1}$ and $V_l' \subset V_l$ with $|V_{l-1}'| \geq \delta_{l-1}^4 |V_{l-1}|$ and $|V_l'| \geq \delta_{l-1}^4 |V_l|$ such that

$$|d_{\omega_{l-1}}(V_{l-1}', V_l') - d_{\omega_{l-1}}(V_{l-1}, V_l)| > \delta_{l-1}^2.$$

Both V_{l-1}' and V_l' will be used as witness sets to improve P. More precisely, the following algorithm can be used to improve the partition P.

Algorithm Improve

1. For every l-tuple $(V_1, \ldots, V_l)$ in the partition P do:
 2. For $i = 1$ to $l-2$ successively
 3. Apply the procedure from Theorem 7 to search for $X_i' \subset X_i = V_{i+1} \times \cdots \times V_l$ and $V_i' \subset V_i$ such that

$$|d_{\omega_i}(X_i', V_i') - d_{\omega_i}(X_i', V_i)| > \delta_i^2. \tag{37}$$

 4. If (37) is satisfied for some $i \leq l-2$ then V_i' is a witness set and we move to the next l-tuple.

5. If there is no $i \leq l-2$ for which (37) holds then we apply the procedure from Theorem 13 to search for $V'_{l-1} \subset V_{l-1}$ and $V'_l \subset V_l$ such that

$$|d_{\omega_{l-1}}(V'_{l-1}, V'_l) - d_{\omega_{l-1}}(V_{l-1}, V_l)| > \delta^2_{l-1}. \tag{38}$$

6. In case of (38) both V'_{l-1} and V'_l are witness sets for the l-tuple.

7. If the number of l-tuples for which witness sets were found is at least ϵt^l then compute a subpartition P' (described below) of P that respects all the witness sets found in steps 3 and 5. Otherwise the partition P is (ϵ, ω)-regular.

Let us conclude this section with the following fact which shows what sets were found by the algorithm **Improve**.

Fact 15. *If an l-tuple $(V_1, \ldots, V_l)$ is (ϵ, ω)-irregular then* **Improve** *either finds $X'_k \subset X_k$ and $V'_k \subset V_k$ with $|X'_k| \geq \frac{\delta_k^7}{2}|X_k|$ and $|V'_k| \geq \delta_k^6|V_k|$ such that (37) holds for some $1 \leq k \leq l-2$, or it finds $V'_{l-1} \subset V_{l-1}$ and $V'_l \subset V_l$ with $|V'_{l-1}| \geq \delta^4_{l-1}|V_{l-1}|$ and $|V'_l| \geq \delta^4_l|V_l|$ such that (38) holds.*

Proof. For every $k = 1, \ldots, l-2$, Lemma 6 implies that if an $(l-k+1)$-tuple $(V_k, \ldots, V_l)$ is (ϵ_k, ω_k)-irregular then either the $(l-k)$-tuple $(V_{k+1}, \ldots, V_l)$ is $(\epsilon_{k+1}, \omega_{k+1})$-irregular or at least $\delta_k|X_k|^2$ of the pairs $\{x_i, x_j\}$ of vertices from X_k satisfy

$$|\langle \overrightarrow{x_i}, \overrightarrow{x_j} \rangle - K_k^2 M d_i d_j| > \delta_k K_k^2 M. \tag{39}$$

If (39) is satisfied then the algorithm of Theorem 7 finds $X'_k \subset X_k$ and $V'_k \subset V_k$ such that (37) holds. If (39) does not hold for any $1 \leq k \leq l-2$ then a pair (V_{l-1}, V_l) is $(\epsilon_{l-1}, \omega_{l-1})$-irregular and the algorithm of Theorem 13 finds $V'_{l-1} \subset V_{l-1}$ and $V'_l \subset V_l$ such that (38) holds. □

6. The analysis of the main algorithm

In this section, we will analyze the algorithm **Improve**. Although a little bit technical, the philosophy of the analysis is the same as the Szemerédi's proof of the regularity lemma [12]. We will show that for a subpartition P' computed by the algorithm the value of the index $ind(P')$ is bigger than the value of the index of the original partition $ind(P)$. In addition, we will show that the size of the exceptional class does not increase in any significant way. The proof is divided into five rather technical facts. In Fact 16, we will show that the size of the exceptional class does not increase too much. Fact 17 shows that the value of the index associated with each (ϵ, ω)-regular l-tuple will remain about the same after the refinement of the original partition. In Fact 18, we will show that we get a "profit" on the index if an (ϵ, ω)-irregular l-tuple is reduced to the vector irregularity. Fact 19 shows that the index will increase in case an (ϵ, ω)-irregular l-tuple is reduced to the irregularity of the weighted bipartite graph. Lemma 20 combines Fact 18 and Fact 19 to show that the value of the index of the refined partition is greater than the value of the original one.

Similarly as in the original proof of the regularity lemma [12], we consider a subpartition P' of P into atoms of size

$$m = \lfloor \frac{|V_i|}{2^{2t^l}} \rfloor$$

that respects the Venn diagram of witness sets found for each (ϵ, ω)-irregular l-tuple. For $j = 1, \ldots, p$, denote by $W_i(j)$ the jth atom in V_i (that is the jth subset in the partition of V_i), and let

$$\bar{V_i} = \bigcup_{j=1}^{p} W_i(j).$$

Observe that for every i

$$||\bar{V_i}| - |V_i|| \leq \frac{|V_i|}{2^{t^l}}. \tag{40}$$

We first observe:

Fact 16. *The size of the exceptional class V_0' increases by at most $\frac{n}{2^{t^l}}$ from the size of V_0.*

Proof. Indeed, there are at most 2^{t^l} equivalence classes of atoms and so in the process of refining P we increase $|V_0|$ by at most

$$t 2^{t^l} \frac{|V_i|}{2^{2t^l}} \leq \frac{n}{2^{t^l}}.$$

□

Fact 17. *For every l-tuple $(V_1, \ldots, V_l)$ and every $i \in [l]$*

$$\frac{1}{p^i} \sum_{j_1,\ldots,j_i} \sum_{x \in \bar{V}_{i+1} \times \cdots \times \bar{V_l}} \frac{(d_\omega(x, W_1(j_1), \ldots, W_i(j_i)))^2}{|\bar{V}_{i+1}| \ldots |\bar{V_l}|} \geq ind_i(V_1, \ldots, V_l) - \frac{3l}{2^{t^l}}.$$

Proof. For every $x \in V_{i+1} \times \cdots \times V_l$,

$$\frac{1}{p^i} \sum_{j_1,\ldots,j_i} (d_\omega(x, W_1(j_1), \ldots, W_i(j_i)))^2 \geq (\frac{1}{p^i} \sum_{j_1,\ldots,j_i} d_\omega(x, W_1(j_1), \ldots, W_i(j_i)))^2 =$$

$$(d_\omega(x, \bar{V_1}, \ldots, \bar{V_i}))^2.$$

Since $|\bar{V_i}| \geq (1 - \frac{1}{2^{t^l}})|V_i|$, by the continuity of density (Fact 2) we have

$$(d_\omega(x, \bar{V_1}, \ldots, \bar{V_i}))^2 \geq (d_\omega(x, V_1, \ldots, V_i))^2 - \frac{2l}{2^{t^l}}, \tag{41}$$

and so

$$\frac{1}{p^i} \sum_{j_1,\ldots,j_i} \sum_{x \in \bar{V}_{i+1} \times \cdots \times \bar{V_l}} \frac{(d_\omega(x, W_1(j_1), \ldots, W_i(j_i)))^2}{|\bar{V}_{i+1}| \ldots |\bar{V_l}|} \geq \sum_{x \in \bar{V}_{i+1} \times \cdots \times \bar{V_l}} \frac{(d_\omega(x, V_1, \ldots, V_i))^2 - \frac{2l}{2^{t^l}}}{|\bar{V}_{i+1}| \ldots |\bar{V_l}|}$$

$$\geq \sum_{x \in \bar{V}_{i+1} \times \cdots \times \bar{V_l}} \frac{(d_\omega(x, V_1, \ldots, V_i))^2}{|V_{i+1}| \ldots |V_l|} - \frac{2l}{2^{t^l}}. \tag{42}$$

Using (40) and the fact that $d_\omega(x, V_1, \dots, V_i) \le 1$

$$\sum_{x \in V_{i+1} \times \dots \times V_l} (d_\omega(x, V_1, \dots, V_i))^2 - \sum_{x \in \bar{V}_{i+1} \times \dots \times \bar{V}_l} (d_\omega(x, V_1, \dots, V_i))^2$$

$$\le \sum_{k=i+1}^{l} \sum \{ (d_\omega(x, V_1, \dots, V_i))^2, x \in V_{i+1} \times \dots V_{k-1} \times (V_k \setminus \bar{V}_k) \times V_{k+1} \times \dots \times V_l \}$$

$$\le \frac{l}{2^{t^l}} |V_{i+1}| \dots |V_l|. \tag{43}$$

By (42) and (43),

$$\frac{1}{p^i} \sum_{j_1, \dots, j_i} \sum_{x \in \bar{V}_{i+1} \times \dots \times \bar{V}_l} \frac{(d_\omega(x, W_1(j_1), \dots, W_i(j_i)))^2}{|\bar{V}_{i+1}| \dots |\bar{V}_l|} \ge \sum_{x \in V_{i+1} \times \dots \times V_l} \frac{(d_\omega(x, V_1, \dots, V_i))^2}{|V_{i+1}| \dots |V_l|} - \frac{3l}{2^{t^l}}$$

$$= ind_i(V_1, \dots, V_l) - \frac{3l}{2^{t^l}}.$$

□

Suppose that an l-tuple $(V_1, \dots, V_l)$ is (ϵ, ω)-irregular and that for some $1 \le i \le l-2$, we found two sets $V_i' \subset V_i$ with $|V_i'| > \delta_i^6 |V_i|$ and $X_i' \subset X_i$ with $|X_i'| > \frac{\delta_i^7}{2} |X_i|$ such that

$$|d_{\omega_i}(X_i', V_i') - d_{\omega_i}(X_i', V_i)| > \delta_i^2. \tag{44}$$

Then we have the following fact.

Fact 18. *Assume that an l-tuple $(V_1, \dots, V_l)$ is (ϵ, ω)-irregular and for some $1 \le i \le l-1$, we found two sets $V_i' \subset V_i$ with $|V_i'| > \delta_i^6 |V_i|$ and $X_i' \subset X_i$ with $|X_i'| > \frac{\delta_i^7}{2} |X_i|$ such that $|d_{\omega_i}(X_i', V_i') - d_{\omega_i}(X_i', V_i)| > \delta_i^2$. If $\delta_i^8 \ge \frac{l}{2^{t^{l-2}}}$, then*

$$\frac{1}{p^i} \sum_{j_1, \dots, j_i} \sum_{x \in \bar{V}_{i+1} \times \dots \times \bar{V}_l} \frac{(d_\omega(x, W_1(j_1), \dots, W_i(j_i)))^2}{|\bar{V}_{i+1}| \dots |\bar{V}_l|} \ge ind_i(V_1, \dots, V_l) + \frac{\delta_i^{17}}{8(1-\delta_i^6)} - \frac{4l}{2^{t^l}}.$$

Proof. First observe that for $Z \subset X_i = V_{i+1} \times \dots \times V_l$, $Y \subset V_i$

$$d_{\omega_i}(Z, Y) = \frac{\sum_{z \in Z, y \in Y} \omega_i(y, z)}{K_i |Z||Y|} = \frac{1}{|Z|} \sum_{z \in Z} \frac{\sum_{y \in Y} \sum_{v_1, \dots, v_{i-1}} \omega(v_1, \dots, v_{i-1}, y, z)}{K |V_1| \dots |V_{i-1}||Y|}$$

$$= \frac{1}{|Z|} \sum_{z \in Z} d_\omega(z, V_1, \dots, V_{i-1}, Y). \tag{45}$$

We may assume that for the witness V_i',

$$\bar{V}_i' = \bigcup_{j_i=1}^{q} W_i(j_i).$$

Then

$$||V_i'| - |\bar{V}_i'|| \le \frac{|V_i'|}{2^{t^l} \delta_i^6}.$$

For every $x \in X_i$, define

$$\alpha_x = \frac{1}{p^i} \sum_{j_1,\dots,j_i \in [p]} d_\omega(x, W_1(j_1), \dots, W_i(j_i)) - \frac{1}{qp^{i-1}} \sum_{j_1\dots,j_{i-1} \in [p]} \sum_{j_i \in [q]} d_\omega(x, W_1(j_1), \dots, W_i(j_i)).$$

Observe that

$$\sum_{j_1\dots,j_{i-1} \in [p]} \sum_{j_i \in [q]} d_\omega(x, W_1(j_1), \dots, W_i(j_i)) = \frac{q}{p} \sum_{j_1,\dots,j_i \in [p]} d_\omega(x, W_1(j_1), \dots, W_i(j_i)) - qp^{i-1}\alpha_x$$

and using the defect form of Schwarz inequality (Fact 3), we infer that

$$\frac{1}{p^i} \sum_{j_1,\dots,j_i} (d_\omega(x, W_1(j_1), \dots, W_i(j_i)))^2$$

$$\geq \left(\frac{1}{p^i} \sum_{j_1\dots,j_i} d_\omega(x, W_1(j_1), \dots, W_i(j_i))\right)^2 + \frac{(\alpha_x qp^{i-1})^2 p^i}{p^i qp^{i-1}(p^i - qp^{i-1})}$$

$$= (d_\omega(x, \bar{V}_1, \dots, \bar{V}_i))^2 + \frac{\alpha_x^2 q}{p-q},$$

due to the fact that $\sum_{j_1,\dots,j_i} d_\omega(x, W_1(j_1), \dots, W_i(j_i)) = d_\omega(x, \bar{V}_1, \dots, \bar{V}_i)$ and since $q \geq \delta_i^6 p$, we have

$$\frac{1}{p^i} \sum_{j_1,\dots,j_i} (d_\omega(x, W_1(j_1), \dots, W_i(j_i)))^2 \geq (d_\omega(x, \bar{V}_1, \dots, \bar{V}_i))^2 + \alpha_x^2 \frac{\delta_i^6}{1-\delta_i^6}. \tag{46}$$

Since $|\bar{V}_i| \geq (1 - \frac{1}{2^{t^l}})|V_i|$ and $|\bar{V}_i'| \geq (1 - \frac{1}{2^{t^l}\delta_i^6})|V_i'|$, by the continuity of density (Fact 2), we have

$$d_\omega(x, V_1, \dots, V_i) - d_\omega(x, V_1, \dots, V_{i-1}, V_i') \leq d_\omega(x, \bar{V}_1, \dots, \bar{V}_i) - d_\omega(x, \bar{V}_1, \dots, \bar{V}_{i-1}, \bar{V}_i') +$$

$$|d_\omega(x, V_1, \dots, V_i) - d_\omega(x, \bar{V}_1, \dots, \bar{V}_i)| + |d_\omega(x, \bar{V}_1, \dots, \bar{V}_{i-1}, \bar{V}_i') - d_\omega(x, V_1, \dots, V_{i-1}, V_i')|$$

$$\leq d_\omega(x, \bar{V}_1, \dots, \bar{V}_i) - d_\omega(x, \bar{V}_1, \dots, \bar{V}_{i-1}, \bar{V}_i') + \frac{2l}{2^{t^l}\delta_i^6} = \alpha_x + \frac{2l}{2^{t^l}\delta_i^6}, \tag{47}$$

as $\alpha_x = d_\omega(x, \bar{V}_1, \dots, \bar{V}_i) - d_\omega(x, \bar{V}_1, \dots, \bar{V}_{i-1}, \bar{V}_i')$.

Applying (45) with $Z = X_i'$, $Y = V_i$ (or V_i' respectively) combined with (47) yields

$$|d_{\omega_i}(X_i', V_i) - d_{\omega_i}(X_i', V_i')| = \frac{1}{|X_i'|} \left|\sum_{x \in X_i'} (d_\omega(x, V_1, \dots, V_i) - d_\omega(x, V_1, \dots, V_{i-1}, V_i')\right|$$

$$\leq \frac{1}{|X_i'|} \left|\sum_{x \in X_i'} \alpha_x\right| + \frac{2l}{2^{t^l}\delta_i^6}.$$

Since $|d_{\omega_i}(X_i', V_i) - d_{\omega_i}(X_i', V_i')| \geq \delta_i^2$, we have

$$\frac{1}{|X_i'|} \left|\sum_{x \in X_i'} \alpha_x\right| \geq \delta_i^2 - \frac{2l}{2^{t^l}\delta_i^6}. \tag{48}$$

Thus,

$$\frac{1}{|X_i|}\sum_{x\in X_i}\alpha_x^2 \geq \frac{1}{|X_i|}\sum_{x\in X_i'}\alpha_x^2 \geq \frac{1}{|X_i||X_i'|}(\sum_{x\in X_i'}\alpha_x)^2 \geq \frac{|X_i'|}{|X_i|}(\delta_i^2 - \frac{2l}{2^{t^l}\delta_i^6})^2. \tag{49}$$

Since $|X_i'| \geq \frac{\delta_i^7}{2}|X_i|$ and $\delta_i^2 - \frac{2l}{2^{t^l}\delta_i^6} \geq \frac{\delta_i^2}{2}$

$$\frac{1}{|X_i|}\sum_{x\in X_i}\alpha_x^2 \geq \frac{\delta_i^{11}}{8}. \tag{50}$$

Therefore, by (46),

$$\frac{1}{p^i}\sum_{j_1,\ldots,j_i}\sum_{x\in\bar{V}_{i+1}\times\cdots\times\bar{V}_l}\frac{(d_\omega(x,W_1(j_1),\ldots,W_i(j_i)))^2}{|\bar{V}_{i+1}|\ldots|\bar{V}_l|}$$

$$\geq \sum_{x\in\bar{V}_{i+1}\times\cdots\times\bar{V}_l}\frac{(d_\omega(x,\bar{V}_1,\ldots,\bar{V}_i))^2+\alpha_x^2\frac{\delta_i^6}{1-\delta_i^6}}{|\bar{V}_{i+1}|\ldots|\bar{V}_l|}.$$

By (41)

$$\sum_{x\in\bar{V}_{i+1}\times\cdots\times\bar{V}_l}\frac{(d_\omega(x,\bar{V}_1,\ldots,\bar{V}_i))^2+\alpha_x^2\frac{\delta_i^6}{1-\delta_i^6}}{|\bar{V}_{i+1}|\ldots|\bar{V}_l|}$$

$$\geq \sum_{x\in\bar{V}_{i+1}\times\cdots\times\bar{V}_l}\frac{(d_\omega(x,V_1,\ldots,V_i))^2+\alpha_x^2\frac{\delta_i^6}{1-\delta_i^6}-\frac{2l}{2^{t^l}}}{|\bar{V}_{i+1}|\ldots|\bar{V}_l|}.$$

Since $|\bar{V}_j| \leq |V_j|$, we have

$$\sum_{x\in\bar{V}_{i+1}\times\cdots\times\bar{V}_l}\frac{(d_\omega(x,V_1,\ldots,V_i))^2+\alpha_x^2\frac{\delta_i^6}{1-\delta_i^6}-\frac{2l}{2^{t^l}}}{|\bar{V}_{i+1}|\ldots|\bar{V}_l|}$$

$$\geq \sum_{x\in\bar{V}_{i+1}\times\cdots\times\bar{V}_l}\frac{(d_\omega(x,V_1,\ldots,V_i))^2+\alpha_x^2\frac{\delta_i^6}{1-\delta_i^6}-\frac{2l}{2^{t^l}}}{|V_{i+1}|\ldots|V_l|}$$

$$= \sum_{x\in\bar{V}_{i+1}\times\cdots\times\bar{V}_l}(\frac{(d_\omega(x,V_1,\ldots,V_i))^2}{|V_{i+1}|\ldots|V_l|}+\frac{\delta_i^6}{1-\delta_i^6}\frac{\alpha_x^2}{|X_i|})-\frac{2l}{2^{t^l}}.$$

Clearly $d_\omega(x,V_1,\ldots,V_i) \leq 1$ and $\frac{\delta_i^6}{1-\delta_i^6}\alpha_x \leq 1$. Thus, using the same argument as in (43) we have

$$\sum_{x\in\bar{V}_{i+1}\times\cdots\times\bar{V}_l}\frac{(d_\omega(x,V_1,\ldots,V_i))^2}{|V_{i+1}|\ldots|V_l|} \geq \sum_{x\in V_{i+1}\times\cdots\times V_l}\frac{(d_\omega(x,V_1,\ldots,V_i))^2}{|V_{i+1}|\ldots|V_l|}-\frac{l}{2^{t^l}}$$

and

$$\sum_{x\in\bar{V}_{i+1}\times\cdots\times\bar{V}_l}\frac{\delta_i^6}{1-\delta_i^6}\frac{\alpha_x^2}{|X_i|}\geq\sum_{x\in V_{i+1}\times\cdots\times V_l}\frac{\delta_i^6}{1-\delta_i^6}\frac{\alpha_x^2}{|X_i|}-\frac{l}{2^{t^l}}.$$

Therefore,

$$\sum_{x\in\bar{V}_{i+1}\times\cdots\times\bar{V}_l}\frac{(d_\omega(x,V_1,\ldots,V_i))^2+\alpha_x^2\frac{\delta_i^6}{1-\delta_i^6}-\frac{2l}{2^{t^l}}}{|\bar{V}_{i+1}|\ldots|\bar{V}_l|}\geq$$

$$\sum_{x\in V_{i+1}\times\cdots\times V_l}\left(\frac{(d_\omega(x,V_1,\ldots,V_i))^2}{|V_{i+1}|\ldots|V_l|}+\frac{\delta_i^6}{1-\delta_i^6}\frac{\alpha_x^2}{|X_i|}\right)-\frac{4l}{2^{t^l}}. \tag{51}$$

Using (50) we further estimate the right-hand side of (51) from below by

$$\sum_{x\in V_{i+1}\times\cdots\times V_l}\frac{(d_\omega(x,V_1,\ldots,V_i))^2}{|V_{i+1}|\ldots|V_l|}+\frac{\delta_i^{17}}{8(1-\delta_i^6)}-\frac{4l}{2^{t^l}}=ind_i(V_1,\ldots,V_l)+\frac{\delta_i^{17}}{8(1-\delta_i^6)}-\frac{4l}{2^{t^l}}.$$

□

Fact 19. *For an $0<\epsilon<1$ let $\delta=\frac{1}{96}\frac{\epsilon^{2l+1}}{2^{(2l+1)(l-1)}}$. Suppose that an l-tuple is (ϵ,ω)-irregular and we found sets $V'_{l-1}\subset V_{l-1}$ with $|V'_{l-1}|\geq\delta_{l-1}^4|V_{l-1}|$ and $V'_l\subset V_l$ with $|V'_l|\geq\delta_{l-1}^4|V_l|$ such that*

$$|d_{\omega_{l-1}}(V'_{l-1},V'_l)-d_{\omega_{l-1}}(V_{l-1},V_l)|>\delta_{l-1}^2.$$

Then either

$$\frac{1}{p^{l-1}}\sum_{j_1,\ldots,j_{l-1}}\sum_{x\in\bar{V}_l}\frac{(d_\omega(x,W_1(j_1),\ldots,W_{l-1}(j_{l-1})))^2}{|\bar{V}_l|}\geq ind_{l-1}(V_1,\ldots,V_l)+\frac{\delta^{17}}{8(1-\delta^6)}-\frac{4l}{2^{t^l}}$$

or

$$\frac{1}{p^{l-1}}\sum_{j_1,\ldots,j_{l-2},j_l}\sum_{x\in\bar{V}_{l-1}}\frac{(d_\omega(x,W_1(j_1),\ldots,W_{l-2}(j_{l-2}),W_l(j_l)))^2}{|\bar{V}_{l-1}|}$$

$$\geq ind_l(V_1,\ldots,V_l)+\frac{\delta^{17}}{8(1-\delta^6)}-\frac{4l}{2^{t^l}}.$$

Proof. If

$$|d_{\omega_{l-1}}(V'_{l-1},V'_l)-d_{\omega_{l-1}}(V_{l-1},V_l)|>\delta_{l-1}^2 \tag{52}$$

then either

$$|d_{\omega_{l-1}}(V'_{l-1},V_l)-d_{\omega_{l-1}}(V_{l-1},V_l)|>\frac{\delta_{l-1}^2}{2} \tag{53}$$

or

$$|d_{\omega_{l-1}}(V'_{l-1},V'_l)-d_{\omega_{l-1}}(V'_{l-1},V_l)|>|\frac{\delta_{l-1}^2}{2}. \tag{54}$$

Set $\delta=\frac{1}{96}\frac{\epsilon^{2l+1}}{2^{(2l+1)(l-1)}}$ and note that by (34), $\frac{\delta_{l-1}^2}{2}\geq\delta^2$.

In case of (53) we apply Fact 18 to $(V_1, \dots, V_{l-2}, V_{l-1}, V_l)$ with $i = l-1$, $X_i' = V_l$, and $V_i' = V_{l-1}'$ to conclude that

$$\frac{1}{p^{l-1}} \sum_{j_1,\dots,j_{l-1}} \sum_{x \in \bar{V}_l} \frac{(d_\omega(x, W_1(j_1), \dots, W_{l-1}(j_{l-1})))^2}{|\bar{V}_l|}$$

$$\geq ind_{l-1}(V_1, \dots, V_l) + \frac{\delta^{17}}{8(1-\delta^6)} - \frac{4l}{2^{t^l}}.$$

In case of (54) we apply Fact 18 to $(V_1, \dots, V_l, V_{l-1})$ with $i = l-1$, $X_i' = V_{l-1}'$, and $V_i' = V_l'$ to conclude that

$$\frac{1}{p^{l-1}} \sum_{j_1,\dots,j_{l-2},j_l} \sum_{x \in \bar{V}_{l-1}} \frac{(d_\omega(x, W_1(j_1), \dots, W_{l-2}(j_{l-2}), W_l(j_l)))^2}{|\bar{V}_{l-1}|}$$

$$\geq ind_{l-1}(V_1, \dots, V_{l-2}, V_l, V_{l-1}) + \frac{\delta^{17}}{8(1-\delta^6)} - \frac{4l}{2^{t^l}}$$

which by (5) is equal to

$$ind_l(V_1, \dots, V_{l-1}, V_l) + \frac{\delta^{17}}{8(1-\delta^6)} - \frac{4l}{2^{t^l}}.$$

□

We have:

Lemma 20. *For every $\epsilon \leq 1/2$ and $\delta = \frac{1}{96}\frac{\epsilon^{2l+1}}{2^{(2l+1)(l-1)}}$ let $P = V_0 \cup V_1 \cup \dots \cup V_t$ be an (ϵ, ω)-irregular partition of an l-uniform hypergraph and let $\frac{4l}{2^{t^l}} \leq \frac{\delta^{17}}{16 \cdot l}$. If we use algorithm* **Improve** *to construct the subpartition P' then*

$$ind(P') \geq ind(P) + \frac{\delta^{17}}{16 \cdot l}.$$

Proof. Assume that each V_i has been partitioned into $W_i(j)$ for $j = 1, \dots, p$ and let P' be the resulting subpartition. Then

$$P' = W_0 \cup W_1(1) \cup \dots \cup W_1(p) \cup \dots \cup W_t(1) \cup \dots \cup W_t(p).$$

To simplify the notation we will also write

$$P' = W_0 \cup W_1 \cup \dots \cup W_{t'}$$

where $t' = tp$. Then

$$ind(P') = \frac{1}{l(tp)^l} \sum_{(W_{i_1},\dots,W_{i_l})} (\sum_{k=1}^{l} ind_k(W_{i_1}, \dots, W_{i_l}))$$

and so

$$ind(P') \geq \frac{1}{lt^l} \sum_{(i_1,\dots,i_l)\in[t]^l} (\frac{1}{p^l} \sum_{j_1 \dots j_l \in [p]} (\sum_{k=1}^{l} ind_k(W_{i_1}(j_1), \dots, W_{i_l}(j_l)))).$$

For every $1 \le k \le l$,

$$\frac{1}{p^l} \sum_{j_1,\dots,j_l \in [p]} ind_k(W_{i_1}(j_1), \dots, W_{i_l}(j_l)) =$$

$$\frac{1}{p^l} \sum_{j_1,\dots,j_l \in [p]} \sum_{x \in W_{i_{k+1}}(j_{k+1}) \times \dots \times W_{i_l}(j_l)} \frac{(d_\omega(x, W_{i_1}(j_1), \dots, W_{i_k}(j_k)))^2}{|W_{i_{k+1}}(j_{k+1})| \dots |W_{i_l}(j_l)|} =$$

$$\frac{1}{p^l} \sum_{j_1,\dots,j_k} \sum_{x \in \bar{V}_{i_{k+1}} \times \dots \times \bar{V}_{i_l}} \frac{(d_\omega(x, W_{i_1}(j_1), \dots, W_{i_k}(j_k)))^2}{|W_{i_{k+1}}(j_{k+1})| \dots |W_{i_l}(j_l)|} =$$

$$\frac{1}{p^k} \sum_{j_1,\dots,j_k} \sum_{x \in \bar{V}_{i_{k+1}} \times \dots \times \bar{V}_{i_l}} \frac{(d_\omega(x, W_{i_1}(j_1), \dots, W_{i_k}(j_k)))^2}{|\bar{V}_{i_{k+1}}| \dots |\bar{V}_l|}, \tag{55}$$

as $|\bar{V}_{i_l}| = p|W_{i_l}(j_l)|$. Fact 17 implies that for every $1 \le k \le l$,

$$\frac{1}{p^l} \sum_{j_1,\dots,j_l \in [p]} ind_k(W_{i_1}(j_1), \dots, W_{i_l}(j_l)) \ge ind_k(V_{i_1}, \dots, V_{i_l}) - \frac{2l}{2^{t^l}}.$$

If an l-tuple $(V_{i_1}, \dots, V_{i_l})$ is (ϵ, ω)-irregular then by Fact 15, we can find either

- $V'_{k_0} \subset V_{k_0}$ and $X'_{k_0} \subset X_{k_0}$ that satisfy the assumptions of Fact 18, for some $1 \le k_0 \le l-2$, or
- $V'_{l-1} \subset V_{l-1}$ and $V'_l \subset V_l$ that satisfy the assumptions of Fact 19.

If the former holds we combine (55) and Fact 18 to infer that for some $1 \le k_0 \le l-2$,

$$\frac{1}{p^l} \sum_{j_1,\dots,j_l \in [p]} ind_{k_0}(W_{i_1}(j_1), \dots, W_{i_l}(j_l)) \ge ind_{k_0}(V_{i_1}, \dots, V_{i_l}) + \frac{\delta_{k_0}^{17}}{8(1-\delta_{k_0}^6)} - \frac{4l}{2^{t^l}}.$$

If the latter holds then by Fact 19 for $k_0 = l-1$, or for $k_0 = l$

$$\frac{1}{p^l} \sum_{j_1,\dots,j_l \in [p]} ind_{k_0}(W_{i_1}(j_1), \dots, W_{i_l}(j_l)) \ge ind_{k_0}(V_{i_1}, \dots, V_{i_l}) + \frac{\delta^{17}}{8(1-\delta^6)} - \frac{4l}{2^{t^l}}.$$

Since the partition P is (ϵ, ω)-irregular at least ϵt^l of l-tuples $(V_{i_1}, \dots, V_{i_l})$ are (ϵ, ω)-irregular. Thus, (by (34) $\delta_{k_0} \ge \frac{1}{48} \frac{\epsilon^{2l+1}}{2^{(2l+1)(l-1)}} > \delta$, $1 - \delta \le 1$)

$$ind(P') \ge ind(P) + \frac{\delta^{17}}{8 \cdot l} - \frac{4l}{2^{t^l}}$$

and since t satisfies $\frac{4l}{2^{t^l}} \le \frac{\delta^{17}}{16 \cdot l}$ by assumption, we have

$$ind(P') \ge ind(P) + \frac{\delta^{17}}{16 \cdot l}.$$

□

Proof of Theorem 1. Set $\delta = \frac{1}{96}\frac{\epsilon^{2l+1}}{2^{(2l+1)(l-1)}}$ and partition the vertex set of a hypergraph into t subsets (arbitrarily), but such that

$$\frac{1}{2^{t^l}} \leq \frac{\epsilon\delta^{17}}{16 \cdot l + \delta^{17}}.$$

Invoke the procedure **Improve** $\frac{16 \cdot l}{\delta^{17}} + 1$ times. By Lemma 20 (note that $\frac{1}{2^{t^l}} \leq \frac{\delta^{17}}{16 \cdot l}$) after at most $\frac{16 \cdot l}{\delta^{17}} + 1$ iterations we find a partition Q with less than ϵt^l (ϵ, ω)-irregular l-tuples, otherwise $ind(Q) > 1$ which is not possible. Also, by Fact 16 the size of the exceptional class

$$|V_0| \leq (\frac{16 \cdot l}{\delta^{17}} + 1)\frac{n}{2^{t^l}} \leq \epsilon n.$$

Next we will argue that the complexity of the algorithm is $O(n^{2l-1}\log^2 n)$. First observe that we iterate the algorithm **Improve** a constant number of times. Since l is constant and the number of partition classes is constant the complexity of **Improve** depends only on the algorithms of Theorem 7 and Theorem 13. Recall that the complexity of the algorithm of Theorem 7 is $O(N^2 M \log^2 K_k)$, where $N = |V_{k+1} \times \cdots \times V_l|$ and $M = |V_k|$. Since $1 \leq k \leq l-1$ and by (33), $K_k \leq Kn^l$ (K is constant by an assumption), the complexity of the algorithm from Theorem 7 is $O(n^{2l-1}\log^2 n)$. The complexity of the algorithm of Theorem 13 is $O(M^3 \log^2 K_{l-1})$ where $M = |V_{l-1}| = |V_l|$. Therefore, in this case, the complexity is $O(n^3 \log^2 n)$. The total complexity of **Improve** is $O(n^{2l-1}\log^2 n)$. □

7. Applications

In this section, we outline the applications of Theorem 1 to the Max-Cut problem for hypergraphs and to the problem of estimating the chromatic number of a hypergraph. Let $H = (V, E)$ be an l-uniform hypergraph and let $n = |V|$. We consider the unweighted case $\omega : [V]^l \to \{0, 1\}$ and to simplify the notation we write $d(V_1, \ldots, V_l)$ instead of $d_\omega(V_1, \ldots, V_l)$. In the Max-Cut problem one wants to find a partition of V into l subsets which is such that the number of hyperedges that intersect each partition class (have nonempty intersection) is maximized. Case $l = 2$ gives the Max-Cut problem for graphs and its "dense case" was considered in [7]. Let $OPT(H) = max|\{e \in E : |e \cap V_i| = 1; i = 1, \ldots, l\}|$, where the maximum is taken over all partitions $V_1 \cup \cdots \cup V_l$ of V.

Theorem 21. *Let $H = (V, E)$ be an l-uniform hypergraph and let $n = |V|$. For every $\epsilon > 0$, there is an $O(n^{2l-1}\log^2 n)$ algorithm that finds a partition $V_1 \cup \cdots \cup V_l$ of V which is such that the number of hyperedges that intersect each V_i, $i = 1, \ldots, l$ is at least $OPT(H) - \epsilon n^l$.*

Proof sketch. The following algorithm finds the postulated partition. The constant ϵ' depends on ϵ and can be computed explicitly.

1. Find an ϵ'-regular partition of H: $W_0, \ldots, W_t$.

2. Check exhaustively all partitions $V_1,\dots,V_l$ in which for every $i\in[l]$ and every $j\in[t]$we have if $W_j\cap V_i\neq\emptyset$ then $W_j\subset V_i$. Choose a partition $V_1,\dots,V_l$ that maximizes
$$\sum_{W_{j_k}\subset V_k} d(W_{k_1},W_{k_2},\dots,W_{k_l})|W_{k_1}|\dots|W_{k_l}|.$$

Note that since there are l^t partitions that are checked in the second step of the algorithm, the complexity of the procedure is $O(n^{2l-1}\log^2 n)$.

For a partition $U_1,\dots,U_l$ of V define
$$f(U_1,\dots,U_l)=max|\{e\in E:|e\cap U_i|=1,i=1,\dots,l\}|,$$
and
$$f^*(U_1,\dots,U_l)=\sum_{W_{j_k}\subset U_k} d(W_{j_1},W_{j_2},\dots,W_{j_l})|U_1\cap W_{k_1}|\dots|U_l\cap W_{k_l}|.$$
One can verify that f^* is maximized for a partition $U_1,\dots,U_l$ which is of the form considered in the second step of the algorithm, i.e. if $W_j\cap U_i\neq\emptyset$ then $W_j\subset U_i$. Also, choosing ϵ' appropriately one can show that for every partition $U_1,\dots,U_l$
$$|f(U_1,\dots,U_l)-f^*(U_1,\dots,U_l)|\le\frac{\epsilon}{2}n^l.$$
Let $V_1,\dots,V_l$ be a partition found by the algorithm and let $U_1,\dots,U_l$ be an optimal partition. Then
$$f(V_1,\dots,V_l)\ge f^*(V_1,\dots,V_l)-\frac{\epsilon}{2}n^l\ge f^*(U_1,\dots,U_l)-\frac{\epsilon}{2}n^l$$
$$\ge f(U_1,\dots,U_l)-\epsilon n^l=OPT(H)-\epsilon n^l.$$
□

Our second application concerns the chromatic number of a hypergraph. The chromatic number $\chi(H)$ is defined as the minimum number of colors needed to color the vertices of H so that there is no hyperedge of H that contains more than one vertex of the same color. Define
$$\chi_\epsilon(H)=min\{\chi(H\setminus E'):E'\subset E;|E'|\le\epsilon n^l\}.$$
We define a hyperedge $\{v_1,\dots,v_l\}$ to be crossing in an l-tuple $(V_1,\dots,V_l)$ if for every $i=1,\dots,l$, $v_i\in V_i$.

Theorem 22. *Let $H=(V,E)$ be an l-uniform hypergraph and let $n=|V|$. For every $\epsilon>0$ there is $O(n^{2l-1}\log^2 n)$ algorithm that finds a number k satisfying*
$$\chi_\epsilon(H)\le k\le\chi(H).$$

Proof sketch. Set $\epsilon'=\frac{\epsilon}{4}$ and find an ϵ'-regular partition $W_0,\dots,W_t$. Construct a subhypergraph H' of H by deleting all the hyperedges adjacent to W_0, hyperedges that are crossing in ϵ'-irregular l-tuples, and the hyperedges that are crossing in the ϵ'-regular l-tuples that have densities not greater than ϵ'. In this process we delete at most $3\epsilon' n^l$ hyperedges. Construct a subhypergraph H'' of H' as follows. Group (arbitrarily) classes $W_0,W_1,\dots,W_t$ into $t'=\frac{1}{\epsilon'}$ sets $V_1,\dots,V_{t'}$ (say $V_i=\bigcup_{j=(i-1)\epsilon' t}^{i\epsilon' t}W_j$) and delete the hyperedges that contain at least two vertices from the same V_i. In this process we delete at most $\epsilon' n^l$ hyperedges. Let $Aux(H'')$ be an l-uniform hypergraph with vertex set $\{W_0, \ldots, W_t\}$ and with $\{W_{i_1}, \ldots, W_{i_l}\} \in E(Aux(H''))$ if and only if there is at least one hyperedge of H'' contained in $W_{i_1} \cup \cdots \cup W_{i_l}$.

Claim 23. $\chi(Aux(H'')) = \chi(H'')$.

Proof. Clearly $\chi(Aux(H'')) \geq \chi(H'')$, as a proper coloring of $Aux(H'')$ induces the proper coloring of H''. Assume that $\chi(H'') < \chi(Aux(H''))$. Let $\bar{W}_i \subset W_i$ be a set of vertices colored in the most frequent color of W_i in a $\chi(H'')$-coloring of H'' (ties are resolved arbitrarily). Consider the coloring of $Aux(H'')$ induced by these "most frequent" colors. Then there exists a hyperedge of $Aux(H'')$, $\{W_{i_1}, \ldots, W_{i_l}\}$ such that at least two of W_{i_j} have the same color. We next show that there must be a crossing hyperedge in $(\bar{W}_{i_1}, \ldots, \bar{W}_{i_l})$. From the construction of H'' it follows that $\chi(Aux(H'')) \leq \frac{1}{\epsilon'}$ and so for every $i = 1, \ldots, t$

$$|\bar{W}_i| \geq \epsilon' |W_i|. \tag{56}$$

Since $\{W_{i_1}, \ldots, W_{i_l}\}$ is a hyperedge of $Aux(H'')$ we have $d(W_{i_1}, \ldots, W_{i_l}) > \epsilon'$, and by the ϵ'-regularity of $(W_{i_1}, \ldots, W_{i_l})$

$$d(\bar{W}_{i_1}, \ldots, \bar{W}_{i_l}) > 0. \tag{57}$$

Therefore, there is at least one crossing hyperedge in $(\bar{W}_{i_1}, \ldots, \bar{W}_{i_l})$ which contradicts the fact that H'' was properly colored. □

Since $Aux(H'')$ has t vertices, $\chi(Aux(H''))$ can be found in a constant time by exhaustive search and so we found $k = \chi(Aux(H'')) = \chi(H'')$ satisfying $k \geq \chi_\epsilon(H)$. □

Acknowledgments

We would like to thank referees for helpful comments and suggestions.

References

[1] N. ALON, R.A. DUKE, H. LEFMANN, V. RÖDL and R. YUSTER, The algorithmic aspects of the Regularity Lemma, Journal of Algorithms 16, (1994), pp. 80-109.

[2] F.R.K. CHUNG, Regularity Lemmas for hypergraphs and quasi-randomness, Random Structures and Algorithms 2, (1991), pp. 241-252

[3] A. CZYGRINOW S. POLJAK and V. RÖDL, Constructive Quasi-Ramsey numbers and tournament ranking, SIAM J. Discrete Math. 12 (1), (1999), pp. 48-63.

[4] R.A. DUKE, H. LEFMANN and V. RÖDL, A fast algorithm for computing frequencies in a given graph, SIAM J. Computing 24 (3), (1995), pp. 598-620.

[5] P. FRANKL and V. RÖDL, The Uniformity Lemma for hypergraphs, Graphs and Combinatorics 8, (1992), pp. 309-312.

[6] P. FRANKL and V. RÖDL, Extremal problems with no exponent, submitted.

[7] A. FRIEZE and R. KANNAN, The Regularity Lemma and Approximation schemes for dense problems, Proc 37th FOCS, (1996), pp. 12-20.

[8] A. FRIEZE and R. KANNAN, Quick approximation to matrices and applications, Combinatorica 19, 175-220.

[9] A. FRIEZE and R. KANNAN, A simple algorithm for constructing Szemerédi's Regularity Partition,Electron. J. Combin.,6 (1), 1999, pp. (electronic).

[10] B. NAGLE and V. RÖDL, The asymptotic number of triple systems not containing a fixed one, to appear in 5th Czech-Slovak International Symp. on Combinatorics, Graph Theory, Algorithms and Applications, Discrete Mathematics.

[11] H. J. PRÖMEL and A. STEGER, Excluding induced subgraphs III; a general asymptotic, Random Structures and Algorithms 3, 1994, 19-31.

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

DEPT. OF MATHEMATICS AND COMPUTER SCIENCE, EMORY UNIVERSITY, ATLANTA, GA 30322