The Regularity Lemma and Approximation schemes for dense problems

Alan Frieze
Department of Mathematical Sciences,
Carnegie Mellon University.
af1p@andrew.cmu.edu*

Ravi Kannan
Department of Computer Science,
Carnegie Mellon University.
ravi.kannan@cs.cmu.edu†

October 8, 1996

1 Introduction

There has been some recent success in designing polynomial time approximation schemes for certain graph problems (like the Max Cut problem) on *dense graphs* – see for example Arora, Karger and Karpinski [5], Fernandez de la Vega [11], Arora, Frieze and Kaplan [4]. This is in contrast to the fact that the existence of such schemes for general graphs would imply that NP equals P by the powerful results of Arora, Lund, Motwani, Sudan and Szegedy [6]. This mirrors the situation in approximate counting where dense problems have sometimes been easier to attack – Annan [3], Broder [7], Jerrum and Sinclair [16], Dyer, Frieze and Jerrum [10] and Alon, Frieze and Welsh [2].

This raises the question of why dense problems should be particularly "easy", at least in theory. A plausible answer is that dense graphs are like random graphs. This answer on one level seems superficial but it does have a significant level of truth to it. The reason being Szemerédi's remarkable Regularity Lemma [20] – "one of the most powerful tools of (extremal) graph theory", Komlós and Simonovits [17]. This theorem shows in a very strong sense that large dense graphs have many properties of random graphs. The lemma promises a partition of every graph into a *bounded* number of pieces such that the edges between pieces are nicely dispersed – we make this precise in Section 2. Szemerédi's original proof was non-costructive but recently Alon, Duke, Lefmann, Rödl and Yuster [1] have proved a constructive version.

There are two main contributions of the paper. The first is to develop a new algorithmic proof of the Regularity Lemma. Our algorithm is a Monte Carlo algorithm which runs in **constant** (depending on the error parameter ϵ) time to produce (implicitly) a partition proving the Lemma. With the partition, we argue that we can solve the Max-Cut, Graph Bisection, Min multi-way cuts and Graph Seperator problems approximately in dense graphs. The running time is again constant to implicitly produce a solution and $O(\alpha(1/\epsilon)n + \beta(1/\epsilon))$ to explicitly produce one; here $\alpha(1/\epsilon) = \tilde{O}((1/\epsilon^2))$ and $\beta = 2^{\tilde{O}(1/\epsilon^2)}$. Arora, Karger and Karpinski [5] gave the first

*Supported in part by NSF grant CCR9225008.
†Supported in part by NSF grant CCR9528973.

PTAS's for these problems whose running time is $O(n^{O(1/\epsilon^2)})$. Fernandez de la Vega [11] gave an $O(2^{1/\epsilon^{2+o(1)}}n^2)$ time algorithm for the unweighted Max Cut and Maximum Acyclic Subgraph problems. Algorithms with similar running times to ours for the above problems have also been obtained by Goldwasser, Goldreich and Ron [14]. Sampling plays an important role in all of these papers. In particular the approach in [14] caused us to incorporate sampling and thereby significantly improve on earlier versions of this paper.

Perhaps a central point of our paper is that the Regularity Lemma provides a uniform explanation of why these Max-SNP hard problems turn out to be easy in dense graphs. In addition to the above problems, we also give a simple PTAS for dense versions of a special case of the **Quadratic Assignment Problem** (QAP) [8, 19] and the **Maximum Acyclic Subgraph** problem. The first PTAS for the QAP was given in [4] where the running time is $O(n^{O(1/\epsilon^2)})$. In the algorithm here this becomes $O(\alpha(1/\epsilon)n + \hat{\beta}(1/\epsilon))$.

We also prove a constructive version of the Regularity Lemma for s-uniform hypergraphs for fixed s. [An $s-$uniform hypergraph has a finite vertex set and each edge consists of s vertices.] This yields PTAS's for dense versions of all Max-SNP problems where density is as defined in [5]. (The class Max-SNP was introduced by Papadimitriou and Yanakakis [18]. We will briefly explain the class in section 6.)

Our second contribution is that the version of the Regularity Lemma we prove constructively has considerably better constants (but has a weaker conclusion that is still sufficient for our purposes) than the original one of Sźemeredi's. This may be of use in other contexts where the Regularity Lemma is used. (The β of the last paragraph we would get from Szemerédi's original version is extremely large - only $\log^*(\beta)$ is a polynomial).

For concreteness, we focus first on the $\ell-$ way cut problem : We are given a graph $G = (V, E)$ and weights $w(e) \in \boldsymbol{R}$ for each $e \in E$. Let $\mathcal{S} = S_1, S_2, \ldots, S_\ell$ be a partition of V. Let $E(\mathcal{S})$ be the set of edges $e = (u, v)$ such that u, v are in different subsets of partition $\mathcal{S}$. Then let

$$w(\mathcal{S}) = \sum_{e \in E(\mathcal{S})} w(e).$$

The ℓ-way cut problem is to

Maximise $w(\mathcal{S})$ over all partitions $\mathcal{S}$ of V into ℓ subsets.

It is assumed that the value of ℓ is fixed and $|w(e)| \leq 1$ for all $e \in E$ on problem instances.

Theorem 1 *There is a randomised algorithm $A_1(\epsilon)$ which given an n-vertex graph G, with probability at least 3/4, computes a partition $\mathcal{S}_\epsilon$ such that*

$$w(\mathcal{S}_\epsilon) \geq w(\mathcal{S}^*) - \epsilon n^2. \tag{1}$$

Here $\mathcal{S}^*$ is the maximum weight partition.

The algorithm requires $2^{\tilde{O}(\epsilon^{-2})}$[1] time to construct an implicit *description of the partition and at most $\alpha(\epsilon)n + \beta(\epsilon)$ time to construct the partition itself. $\alpha(\epsilon) = \tilde{O}(\epsilon^{-2})$ and $\log \beta = \tilde{O}(\epsilon^{-2})$.*

We explain the notion of implicit description later in the paper.

[1]The $\tilde{O}$ notations hides factors of order polylog$(1/\epsilon)$

Remark 1 If we restrict our attention to cases where we know $w(\mathcal{S}^*) = \Omega(n^2)$ then $(A_1(\epsilon))$ is a PTAS.

For example if G is γ-dense i.e. $|E| \geq \gamma n^2$ and $w(e) = 1$ for $e \in E$ then

$$w(\mathcal{S}^*) \geq (1 - 1/\ell)|E| \geq \gamma(1 - 1/\ell)n^2.$$

We focus next on the Koopmans-Beckmann version of the QAP. Here one is given a set of n items V which have to be assigned to a set of n locations X, one per location. We are given two $n \times n$ *symmetric* non-negative matrices T, D. Here $T_{i,i'}$ is the amount of *traffic* between item i and i' and $D_{x,x'}$ is the *distance* between location x and x'. Then if item i is assigned to location $\pi(i)$ for $i \in [n]$ the total cost $c(\pi)$ is defined by

$$c(\pi) = \sum_{i=1}^{n} \sum_{i'=1}^{n} T_{i,i'} D_{\pi(i),\pi(i')}.$$

The problem is to minimise $c(\pi)$ over all bijections $\pi : V \to X$.

A typical example is where a location is a room in a building (e.g. hospital) and each item is a facility of some sort (e.g. operating theatre, intensive care unit etc.) and the total cost is the expected sum over pairs of facilities of the product of traffic intensity and distance.

We will restrict our attention to the case where the n locations are the points of a finite metric space X with metric D. We assume that

1. diam(X)=1 i.e. $\max_{x,y} D_{x,y} = 1$.
2. For all $\epsilon > 0$ there exists a partition $X = X_1 \cup X_2 \cup \cdots \cup X_\ell$, $\ell = \ell(\epsilon)$, such that $\text{diam}(X_j) \leq \epsilon$, for $1 \leq j \leq \ell$.
   We call this an ϵ-refinement of X.
   We can therefore define an $\ell \times \ell$ matrix $\bar{D}$ such that if $x \in X_j$ and $x' \in X_{j'}$ then $|D_{x,x'} - \bar{D}_{j,j'}| \leq 2\epsilon$.
   Furthermore this partition is computable in time polynomial in n and $1/\epsilon$ – for the cases we have in mind, this will be insignificant compared with that required by the rest of the algorithm.

We call this the *metric* QAP.

The **Minimum Linear Arrangement** problem [13] where $X = \{0, 1/n, 2/n, \ldots, 1\}$ is a special case. We will also assume that $T_{i,i'} \leq 1$ for all i, i' and this can be achieved by scaling. Let π^* denote the permutation which minimises c.

Theorem 2 *There is a randomised algorithm algorithm $A_3(\epsilon)$ for the metric QAP which, with probability at least 3/4, produces a permutation π_ϵ such that $c(\pi_\epsilon) \leq c(\pi^*) + \epsilon n^2$ and which runs in time at most $\alpha_2(1/\epsilon)n + \beta_3(1/\epsilon)$. The exact expression for β_3 will be exposed in the proof of the Theorem.*

The paper is organised as follows: Section 2 summarises the Regularity Lemma. Section 3 states the new version and explains how it can be used to prove Theorem 1. Section 4 describes our QAP algorithm. Section 5 describes the proof of our version of the Regularity lemma. Section 6 describes the construcitve Regularity Lemma for hypergraphs and its use for Max-SNP problems.

2 Szemerédi's Regularity Lemma

Let $G = (V, E)$ be a graph with n vertices. For disjoint sets $A, B \subseteq V$ let $e(A, B)$ denote the number of edges between A and B. The *density* $d(A, B)$ is defined by

$$d(A,B) = \frac{e(A,B)}{|A|\,|B|}.$$

A disjoint pair $A, B \subseteq V$ is said to be $\epsilon - regular$ if for *every* $X \subseteq A$ with $|X| \geq \epsilon|A|$ and $Y \subseteq B$ with $|Y| \geq \epsilon|B|$, we have

$$|d(X,Y) - d(A,B)\,| < \epsilon.$$

Theorem 3 *(**Regularity Lemma**) For every $\epsilon > 0$ and integer $m > 0$ there are integers $P(\epsilon, m), Q(\epsilon, m)$ with the following property: for* every *graph $G = (V, E)$ with $n \geq P(\epsilon, m)$ vertices there is a partition of V into k classes $V_1, \ldots, V_k$ such that*

- *$m \leq k \leq Q(\epsilon, m)$.*
- *$|V_i| \in \{\nu - 1, \nu\}$ for $1 \leq i \leq k$ where $\nu = \lceil n/k \rceil$.*
- *All but ϵk^2 of the pairs (V_i, V_j) are ϵ-regular.*

The partition alluded to in the theorem will be referred to as an $\epsilon - RL$ partition.

As mentioned previously, Szemerédi's proof is non-constructive but Alon et al show how to construct an $\epsilon - RL$ partition (with different values of P, Q) in time $O(\alpha_1(\epsilon)M(n))$.

$Q(\epsilon, m)$ is huge - only $\log^*(Q)$ is a polynomial in $1/\epsilon$ and m (of degree about 20).

3 The Algorithm

Proof (of Theorem 1) We will see that we do not need the full strength of the Regularity Lemma in order to prove Theorem 1.

Let $V_1, \ldots, V_k$ be a partition of V. Let $d_{i,j} = d(V_i, V_j)$. For $X \subseteq V$ and $I \subseteq K = \{1, 2, \ldots, k\}$ we let $X_I = \bigcup_{i \in I} X_i$ where $X_i = X \cap V_i$. Let S, T be disjoint subsets of V. Let

$$\Delta(S,T) = e(S,T) - \sum_{i \in K} \sum_{j \in K} d_{i,j}|S_i||T_j|.$$

The term $d_{i,j}|S_i||T_j|$ would be (approximately) $e(S_i, T_j)$ if the pair V_i, V_j is ϵ-regular. So $\Delta(S, T)$ measures the deviation from regularity, but the important difference is that we only look at the total of such deviations over all i, j.

A partition is ϵ-*sufficient* if

$$|\Delta(S,T)| \leq \epsilon n^2 \text{ for all disjoint subsets } S, T \text{ of } V.$$

Notice that we do not insist on the subsets being of (almost) the same size. This can easily be enforced, at a small extra cost, see Lemma 5 below.

Our version of the Regularity Lemma to be described in Section 5 will produce an ϵ-sufficient partition with $\log k = O(\epsilon^{-2})$. But, in this section, we will see how to use it to get an algorithm. But first, the following simple lemma (which we do not use) shows that the partition produced by the original Regularity Lemma is in fact an ϵ sufficient partition.

Lemma 1 *An ϵ-RL partition, with $k \geq \epsilon^{-1}$, is 4ϵ-sufficient.*

Proof Let $L_2 = \{(i,j) \in K \times K : |S_i| \leq \epsilon\nu \text{ or } |T_j| \leq \epsilon\nu\}$ and $L = \{(i,j) \in K \times K : i \neq j$ and (V_i, V_j) is an ϵ-regular pair$\}$. Then

$$\Delta(S,T) = \Delta_1 + \Delta_2 + \Delta_3 + D_4$$

where

$$\Delta_i = \sum_{(i,j)\in L_i} (e(S_i, T_j) - d_{i,j}|S_i||T_j|).$$

Here $L_1 = L \setminus L_2$, $L_3 = (K \times K) \setminus (L_1 \cup L_2)$ and $L_4 = \{(i,i) : i \in K\}$. Now one can easily show that $|\Delta_i| \leq \epsilon\nu^2 k^2$ for $i = 1, 2, 3, 4$ and the lemma follows. □

We now consider the unweighted case of Theorem 1 where $w(e) = 1$ for $e \in E$. Assume that now we have an ϵ-sufficient partition $V_1, V_2, \ldots V_k$.

Consider any ℓ-way cut $\mathcal{S} = S_1, S_2, \ldots, S_\ell$. Let $S_{i,r} = V_i \cap S_r$, $T_{i,r} = V_i \setminus S_{i,r}$. Then

$$\begin{aligned} 2w(\mathcal{S}) &= \sum_{r=1}^{\ell} e(S_r, V \setminus S_r) \\ &= \sum_{r=1}^{\ell} \sum_{i\in K} \sum_{j\in K} d_{i,j}|S_{i,r}||T_{j,r}| + \theta, \end{aligned}$$

where $|\theta| \leq \ell\epsilon n^2$.

There is one more simplification we can make. Let $\nu_j = |V_j|$, $1 \leq j \leq k$, $\rho = \lfloor \epsilon n/k \rfloor$, $\bar{\nu}_j = \lfloor \nu_j/\rho \rfloor$, $n_{i,r} = |S_{i,r}|$ and $\bar{n}_{i,r} = \lfloor n_{i,r}/\rho \rfloor$ for $i, r \geq 1$. We observe that

$$|n_{i,r}(\nu_j - n_{j,r}) - \rho^2 \bar{n}_{i,r}(\bar{\nu}_j - \bar{n}_{j,r})| \leq \rho(\nu_i + \nu_j),$$

and so

$$\left| \sum_{r=1}^{\ell} \sum_{i\in K} \sum_{j\in K} d_{i,j} n_{i,r}(\nu_j - n_{j,r}) - \rho^2 \sum_{r=1}^{\ell} \sum_{i\in K} \sum_{j\in K} d_{i,j} \bar{n}_{i,r}(\bar{\nu}_j - \bar{n}_{j,r}) \right| \leq 2\epsilon\ell n^2.$$

Thus,

$$\left| 2w(\mathcal{S}) - \rho^2 \sum_{r=1}^{\ell} \sum_{i=1}^{k} \sum_{j=1}^{k} d_{i,j} \bar{n}_{i,r}(\bar{\nu}_j - \bar{n}_{j,r}) \right| \leq 3\ell\epsilon n^2. \tag{2}$$

Thus to find a cut which maximises $w(\mathcal{S})$ to within $3\ell\epsilon n^2/2$ we need only find one which maximises

$$\sum_{r=1}^{\ell}\sum_{i=1}^{k}\sum_{j=1}^{k} d_{i,j}\bar{n}_{i,r}(\bar{\nu}_j - \bar{n}_{j,r}). \tag{3}$$

This is a simple matter as $\bar{n}_{i,r}$ takes one of at most $2/\epsilon$ values. It is then easy to show that we now need to solve an ℓ-way cut problem on k/ϵ vertices. For this we use the algorithm of [5]. Once we have maximised (3) we can take $n_{i,r} = \rho\bar{n}_{i,r}$ and then choose *any* $n_{i,r}$-subset of V_i to be our $S_{i,r}$, for $1 \le i \le k$. This proves Theorem 1 for this special case, after replacing ϵ by $\epsilon/(3\ell)$.

Now let us consider the general weighted case. We first replace $w(e)$ by $\hat{w}(e) = \epsilon\lfloor w(e)/\epsilon\rfloor$. Since $|\hat{w}(e) - w(e)| \le \epsilon$ we see that $|w(\mathcal{S}) - \hat{w}(\mathcal{S})| \le \epsilon n^2$ for any partition $\mathcal{S}$. Now each edge has a weight $p\epsilon$ where $p \in P = \{0, \pm 1, \pm 2, \ldots, \pm W = \lfloor 1/\epsilon\rfloor\}$. Let $E_p = \{e \in E : \hat{w}(e) = p\epsilon\}$ for $p \in P$ and let $G_p = (V, E_p)$. We prove a "multicoloured" version of the Regularity Lemma which gives a *single* partition $V_1, V_2, \ldots, V_k$, $\log k = O(|P|\epsilon^{-2}) = O(\epsilon^{-3})$ which is *simultaneously* an ϵ-sufficient partition for all G_p. The algorithm of [1] can be modified to construct the partition as can that of Section 5. We apply the latter algorithm to compute a partition which is simultaneously ϵ^2-sufficient for all graphs G_p. Applying the argument for the unweighted case we see that find a cut which maximises $\hat{w}(\mathcal{S})$ to within $(2W+1)3\ell\epsilon n^2/2 \le 6\epsilon\ell n^2$ we need only find one which maximises

$$\sum_{p=-W}^{W}\sum_{r=1}^{\ell}\sum_{i\in K}\sum_{j\in K} d_{i,j,p}\bar{n}_{i,r}(\bar{\nu}_j - \bar{n}_{j,r})p\epsilon.$$

where $d_{i,j,p}$ is the density of V_i, V_j in G_p.

This proves Theorem 1 for the general case, after replacing ϵ by $\epsilon/(12\ell)$. □

Remark 2 Note that we do not need to know the $d_{i,j}$ exactly in order to get a good approximation. All we need are values $\hat{d}_{i,j}$ such that $|d_{i,j} - \hat{d}_{i,j}| = O(\epsilon)$ where the hidden constant is suitably small, uniformly in i, j. To compute the $d_{i,j}$ exactly would require order n^2 time and we have only allowed ourselves order n in the theorem.

Similar ideas work for graph-bisection and separator problems.

4 Metric QAP

We use the notation introduced in section 1 for this problem. We replace T by $\hat{T}$ where $\hat{T}_{i,i'} = \epsilon_1\lfloor T_{i,i'}/\epsilon_1\rfloor$, $\epsilon_1 = \epsilon/2$. This replaces c by $\hat{c}$ where

$$|c(\pi) - \hat{c}(\pi)| \le \epsilon n^2/2 \tag{4}$$

for all bijections π. We can therefore attempt to minimise $\hat{c}$.

Now each $\hat{T}_{i,i'} = p\epsilon_1$ for some $p \in P = \{0, \pm 1, \pm 2, \ldots, \pm M = \lfloor 1/\epsilon_1\rfloor\}$. Let $E_p = \{(i,i') : \hat{T}_{i,i'} = p\epsilon_1\}$ for $p \in P$ and let $G_p = (V, E_p)$. The G_p are graphs since we have assumed that T is symmetric.

We compute an ϵ_2-refinement of X, $\epsilon_2 = \epsilon^2/36$. (The idea for doing this comes from [4]).

We compute a partition $V_1, V_2, \ldots, V_k$, $\log k = O(\epsilon_3^{-2})$ of V which is simultaneously ϵ_3-sufficient, $\epsilon_3 = \epsilon^2/(36\ell(\epsilon_2)^2)$ for every G_p.

Now let π be a fixed bijection from V to X and

$$S_{i,j} = \{v \in V_i : \ \pi(v) \in X_j\}.$$

Next let $\rho = \lfloor \epsilon_4 n/k \rfloor$, $\epsilon_4 = \epsilon^2/(120\ell)$, $\nu = \lceil n/k \rceil$ and $s_{i,j} = \lfloor |S_{i,j}|/\rho \rfloor$. We claim that if $d_{i,i',p}$ denotes the density of $V_i, V_{i'}$ in G_p and if

$$\Delta = \rho^2 \sum_{i,i',j,j',p} s_{i,j} s_{i',j'} d_{i,i',p} \bar{D}_{j,j'} p \epsilon_1$$

then

$$\hat{c}(\pi) = \Delta + \mathrm{err}(\pi), \tag{5}$$

where $|\mathrm{err}(\pi)| \le \epsilon n^2/2$.

This and (4) implies that we can minimise c up to error $\le \epsilon n^2$ by choosing values of $s_{i,j}$ which minimise Δ and then arbitrarily choosing a π consistent with these values. The number of distinct values for $s_{i,j}$ is at most $2/\epsilon_4$ and the number of distinct i, j is k, ℓ. Thus the number of possibilities for the $s_{i,j}$ is significantly smaller than $(2/\epsilon_4)^{k\ell}$ and Theorem 2 follows with $\beta_3 = (2/\epsilon_4)^{k\ell}$.

Proof of (5) Let

$$\Delta_1 = \sum_{i,i',j,j',p} |S_{i,j}||S_{i',j'}| d_{i,i',p} \bar{D}_{j,j'} p \epsilon_1.$$

Then as

$$\rho s_{i,j} - |S_{i,j}|| \le \epsilon_4 n/k$$

we have

$$\Delta_1 = \Delta + \mathrm{err}_1(\pi) \tag{6}$$

where

$$|\mathrm{err}_1(\pi)| \le \frac{10\epsilon_4 \ell}{\epsilon_1} n^2.$$

Next let

$$\Delta_2 = \sum_{i,i',j,j',p} e_p(S_{i,j}, S_{i',j'}) \bar{D}_{j,j'} p \epsilon_1,$$

where $e_p(A, B)$ is the number of G_p edges joining sets A and B.

For a fixed $j \neq j'$ we have (by ϵ_3-sufficiency)

$$\left| \sum_{i,i'} (|S_{i,j}||S_{i',j'}| d_{i,i',p} - e_p(S_{i,j}, S_{i',j'})) \right| \le \epsilon_3 n^2.$$

This then implies that

$$\Delta_2 = \Delta_1 + \mathrm{err}_2(\pi) \tag{7}$$

where

$$|\mathrm{err}_2(\pi)| \le \left(\frac{3\epsilon_3 \ell^2}{\epsilon_1} + \frac{3\epsilon_2}{\epsilon_1} \right) n^2.$$

We finally approximate

$$\hat{c}(\pi) = \Delta_2 + \mathrm{err}_3(\pi) \tag{8}$$

where

$$|\mathrm{err}_3(\pi)| \leq \frac{5\epsilon_2}{\epsilon_1} n^2.$$

Equation (5) follows from (6), (7) and (8).

Maximum Acyclic Subgraph Problem: Here we are given a (weighted) digraph and the problem is to find the maximum (weight) subset of the edges which induces an acyclic digraph. This similar to the QAP but we need to use an ϵ-sufficient partition for digraphs. This requires simple changes to the undirected case – see also Section 6. We assume also that each $|V_i| \leq \epsilon n$, see Remark 5. We re-formulate the problem as one of finding a bijection $f : V \to [n]$ which maximises $|\{e = (x, y) \in E : f(x) < f(y)\}$. This is relaxed to finding $\phi : V \to [\lceil 1/\epsilon \rceil]$ such that (i) $|\phi^{-1}(i)| \approx \epsilon n$ for $j \in [\lceil 1/\epsilon \rceil]$ and (ii) $|\{e = (x, y) \in E : \phi(x) < \phi(y)\}$ is maximised. With an ϵ-sufficient partion we can find ϕ by (approximately) optimising over the the choice of the number $n_{i,j}$ of V_i which are mapped into j by ϕ. Details are left to the final paper.

5 A new version of the Regularity Lemma

The main Theorem we prove in this section is the following :

Theorem 4 *For any graph $G(V, E)$ with $|V| = n$ sufficiently large, and $\epsilon > 0$, we can construct in time $\tilde{O}(\epsilon^{-2})n$ an ϵ-sufficient partition of V; the partition has at most k parts, where $\log k = O(\epsilon^{-2})$.*

This will follow from the two lemmas below. Earlier versions of the paper used an approach based on that given in [1]. It now seems that one obtains more efficient algorithms and (conceptually) simpler proofs by using sampling.

For a partition $\mathcal{P}$ we use, following [20], a number called the *index* of $\mathcal{P}$. It has a slightly different form to that of [20], reflecting the fact that we do not keep the subsets of the partition (nearly) equal in size.

$$\mathrm{ind}(\mathcal{P}) = \frac{1}{n^2} \sum_{1 \leq r < s \leq k} d_{r,s}^2 |V_r||V_s|.$$

Note that $\mathrm{ind}(\mathcal{P}) \leq 1/2$.

We start with an arbitrary partition e.g the trivial partition consisting of a single set V. We then refine it until it is ϵ-sufficient. To keep the running time within $O(n)$ we only keep a representation of the current partition. This will be a tree structure where each node x represents a subset $V_x \subseteq V$. If $x_1, x_2, \ldots,$ are the children of x then $V_{x_1}, V_{x_2}, \ldots,$ is a partition of V_x. An exact description is left until after we have proved Lemmas 2 and 3 below.

Lemma 2 *Fix $0 < \delta < 1$. If $\mathcal{P}$ is not ϵ-sufficient then in time $O(2^{O(1/\epsilon^2)}\delta^{-1}\log\delta^{-1})$ we can with probability at least $1 - \delta$ construct a disjoint pair $A, B \subseteq V$ such that*

$$|\Delta(A, B)| \geq \frac{\epsilon n^2}{40}.$$

Remark 3 Since $\mathcal{P}$ is not ϵ-sufficient, there **exists** a pair A, B such that $|\Delta(A,B)| \geq \epsilon n^2$. In the above, we **find** a pair A, B, but with a slightly weaker conclusion.

Lemma 3 *Given a partition $\mathcal{P}$, a pair of sets disjoint A, B with $|\Delta(A,B)| \geq \gamma n^2$ we can construct a new partition $\mathcal{P}'$ with*

- $k' \leq 3k$.
- $\mathrm{ind}(\mathcal{P}') \geq \mathrm{ind}(\mathcal{P}) + 4\gamma^2$.

It follows that at most $200\epsilon^{-2}$ constructions from Lemma 2 suffice to determine an ϵ-sufficient partition. Also, at the end of the process $k = 2^{O(1/\epsilon^2)})$. The argument can be extended to the multi-coloured case so that general weights can be handled as they were for Theorem 1.

Before proving Lemma 2 we consider the following computational problem which arises in its proof. We are given a complete graph (W_1, W_2, E) where each $e \in E$ has a weight $w(e) \in [-1, 1]$. For $S \subseteq W_1$, $T \subseteq W_2$ we define $w(S,T) = \sum_{e \in S \times T} w(e)$. Let $n_1 = |W_1|$, $n_2 = |W_2|$ and $n = n_1 + n_2$. For $A \subseteq W_1$ we define $P(A) = \{x \in W_2 : \ w(A,x) \geq 0\}$ and similarly define $P(B) \subseteq W_1$ for $B \subseteq W_2$. Let

$$p = \frac{32}{\gamma^2} \ln \frac{32}{\gamma\delta} \text{ and } m = 36\frac{p}{\gamma^2} + \frac{18}{\gamma^2} \ln \frac{4}{\delta}. \tag{9}$$

Lemma 4 *Suppose there exist $S \subseteq W_1$, $T \subseteq W_2$ with $w(S,T) \geq \gamma n^2$. Fix $0 < \delta < 1$. There is a randomised algorithm* SAMPLE *which with probability at least $1 - \delta$ produces sets $Z_1 \subseteq W_1$, $Z_2 \subseteq W_2$ with $|Z_1|, |Z_2| \leq p$ such that $A = P(Z_2), B = P(Z_1)$ satisfy $w(A,B) \geq \gamma n^2/4$. The running time of the algorithm is $O(m^2 2^p)$ assuming that we can pick a random vertex in unit time and we can check the adjacency of two vertices in unit time.*

Proof We describe an algorithm which has similarities to one described in [14].

SAMPLE

- Independently choose random subsets $U_1, R_1 \subseteq W_1$, $U_2, R_2 \subseteq W_2$ with $|U_1| = |U_2| = p$ and $|R_1| = |R_2| = m$
- **For all** $U_1' \subseteq U_1$ and $U_2' \subseteq U_2$ **do**
  begin
  if $w(P(U_2') \cap R_1, P(U_1') \cap R_2) \geq \gamma m^2 \left(\frac{n^2}{n_1 n_2} - 2\right)$ **then**
  output $A = P(U_2')$ and $B = P(U_1')$ and **terminate.**
  end

For each $u \in W_1$ we can write

$$w(u, U_2 \cap T) = \sum_{v \in U_2} a_v$$

where for $v \in W_2$

$$a_v = \begin{cases} 0 & v \notin T \\ w(u,v) & v \in T \end{cases}$$

Thus

$$\mathbf{E}(w(u, U_2 \cap T) = \frac{p}{n_2} w(u,T).$$

Applying Theorems 2 and 4 (sampling with replacement) of Hoeffding [15] we see that

$$\mathbf{Pr}\left(\left|w(u,U_2\cap T)-\frac{p}{n_2}w(v,T)\right|\geq\frac{\gamma p}{4}\right)\leq 2e^{-\gamma^2p/32}=\delta\gamma/16.$$

Let

$$BAD=\left\{u\in V_I:\left|w(u,U_2\cap T)-\frac{p}{n_2}w(u,T)\right|\geq\gamma p/4\right\}.$$

So, we have $\mathbf{E}(|BAD|)\leq\delta\gamma n_1/16$ and so

$$\mathbf{Pr}_{U_2}\left(|BAD|\geq\epsilon n_1/4\right)\leq\delta/4.$$

Thus we have that with probability at leat $1-\delta/4$,

$$\begin{aligned}|w(P(T),T)-w((P(U_2\cap T),T)| &\leq |BAD|n_2+(n_1-|BAD|)\gamma n_2/4\\ &\leq \gamma n_1n_2/2.\end{aligned}$$

Thus if $S'=P(U_2\cap T)$ we see that with probability at leat $1-\delta/4$,

$$\begin{aligned}w(S',T) &\geq w(P(T),T)-\gamma n_1n_2/2\\ &\geq w(S,T)-\gamma n_1n_2/2.\end{aligned}$$

A similar argument applied to U_1 and S' shows that if $T'=P(U_1\cap S')$ then with probability at leat $1-\delta/2$,

$$\begin{aligned}w(S',T') &\geq w(S',T)-\gamma n_1n_2/2\\ &\geq w(S,T)-\gamma n_1n_2.\end{aligned}$$

Summarising the above computation we see that with probability at leat $1-\delta/2$,

$$w(P(U_1\cap S'),P(U_2\cap T))\geq\gamma n^2-\gamma n_1n_2. \tag{10}$$

We see then that the sets $U_1\cap S',U_2\cap T$ will likely suffice as U_1',U_2'. It would be too expensive for us to actually find them by computing $w(P(U_2'),P(U_1'))$ for each $U_1'\subseteq U_1$ and $U_2'\subseteq U_2$. Instead we use n_1n_2/m^2 times $w(R_1\cap P(U_2'),\ R_2\cap P(U_1'))$ as an estimate. We prove that for any $X\subseteq W_1$ and $Y\subseteq W_2$ we have

$$\mathbf{Pr}\left(\left|w(X,Y)-\frac{n_1n_2}{m^2}w(X\cap R_1,Y\cap R_2)\right|\geq\gamma n_1n_2\right)\leq e^{-2p}\delta. \tag{11}$$

Before proving this, let us see that it suffices to complete the proof of the lemma. We first condition on the values of U_1,U_2. Then

$$\begin{aligned}\mathbf{Pr}\left(\exists U_1'\subseteq U_1,U_2'\subseteq U_2:\ \left|w(P(U_2'),P(U_1'))-\frac{n_1n_2}{m^2}w(P(U_2')\cap R_1,P(U_1')\cap R_2)\right|\geq\gamma n_1n_2\right)&\\ \leq 2^{2p}e^{-2p}\delta\leq\delta/2.&\end{aligned} \tag{12}$$

It follows from (10) and (12) that with probability at least $1-\delta$

$$w(P(U_2\cap T)\cap R_1,P(U_1\cap S')\cap R_2)\geq\gamma m^2\left(\frac{n^2}{n_1n_2}-2\right).$$

Hence with probability at least $1-\delta$ SAMPLE will output two sets U_1', U_2'. Furthermore, another application of (11) shows that with $A = P(U_2')$, $B = P(U_1')$

$$w(A,B) \geq \gamma n^2 - 3\gamma n_1 n_2 \geq \gamma n^2/4$$

proving the lemma.

Proof of (12)

If

$$b_u = \begin{cases} 0 & u \notin X \\ w(u,Y) & u \in X \end{cases}$$

then

$$w(X \cap R_1, Y) = \sum_{u \in R_1} b_u.$$

In particular

$$\mathbf{E}(w(X \cap R_1, Y)) = \frac{m}{n_1} w(X,Y).$$

Now $b_u \in [-|Y|, |Y|]$ and so applying Theorems 2 and 4 of [15] we see that

$$\mathbf{Pr}\left(\left|w(X \cap R_1, Y) - \frac{m}{n_1} w(X,Y)\right| \geq \gamma m|Y|/3\right) \leq e^{-2p}\delta/2. \tag{13}$$

Applying the symmetric argument, we get

$$\mathbf{Pr}\left(\left|w(X \cap R_1, Y \cap R_2) - \frac{m}{n_2} w(X \cap R_1, Y)\right| \geq \gamma m|X \cap R_1|/3\right) \leq e^{-2p}\delta/2. \tag{14}$$

Inequality (12) and the lemma follows by combining (13) and (14). □

Proof of Lemma 2 Suppose there exist $S, T \subseteq V$ such that $|\Delta(S,T)| \geq \epsilon n^2$. First we consider a random partition of K into two sets I, J. Let

$$X = X(I,J) = \sum_{i \in I} \sum_{j \in J} (e(S_i, T_j) - d_{i,j}|S_i||T_j|).$$

Then

(i) $|X| \leq n^2$.

(ii) $\mathbf{E}(|X|) \geq |\mathbf{E}(X)| = \left|\frac{\Delta(S,T)}{2}\right| \geq \frac{\epsilon n^2}{2}$.

It follows that

$$\mathbf{Pr}\left(|X| \geq \frac{\epsilon n^2}{4}\right) \geq \frac{\epsilon}{4-\epsilon}.$$

Suppose we repeat the whole refinement procedure r times at each attempt. Let say we have an X-failure if during some attempt at refinement we never achieve $|X| \geq \epsilon n^2/4$. The probability of an X-failure in the whole construction is $O(\epsilon^{-2}(1-\epsilon/4)^r)$. This can be made as small as necessary by taking $r = O(\epsilon^{-1} \log \epsilon^{-1})$.

Assume then that we have a pair I, J with $X(I,J) \geq \epsilon n^2/4$ (the case of $X(I,J) \leq -\epsilon n^2/4$ is similar). At this point it would be natural to apply Lemma 4 with $w = \Delta$ and $\gamma = \epsilon/4$. This would require exact knowledge of the $d_{i,j}$. This turns out to be expensive and unnecessary. For each i,j we randomly choose $q = 1200\epsilon^{-2}\log(k^2/\delta)$ pairs of vertices in $V_i \times V_j$ (see Remark 4 below). Let $X_{i,j}$ be the number of times that a pair of adjacent vertices is chosen. Then $X_{i,j}$ has binomial distribution $B(q, d_{i,j})$ and so if $\hat{d}_{i,j} = X_{i,j}/q$

$$\mathbf{Pr}(|\hat{d}_{i,j} - d_{i,j}| \geq \epsilon/20) \leq 2e^{-\epsilon^2 q/1200} \leq 2\delta/k^2. \tag{15}$$

For $e = (x,y) \in V_i \times V_j$ we let

$$w(e) = \begin{cases} 1 - \hat{d}_{i,j} & e \text{ is an edge of } G \\ -\hat{d}_{i,j} & \text{otherwise} \end{cases}$$

It follows that with probability at least $1 - \delta$

$$|w(e) - \Delta(e)| \leq \epsilon/20 \qquad \text{for all } e \in V_I \times V_J.$$

It follows that $w(S,T) \geq \epsilon n^2/5$. Applying Lemma 4 we obtain that with probability at least $1 - \delta$

$$\begin{aligned} \Delta(A,B) &\geq w(A,B) - \frac{\epsilon n^2}{40} \\ &\geq \frac{\epsilon n^2}{20} - \frac{\epsilon n^2}{40} = \frac{\epsilon n^2}{40}. \end{aligned}$$

Replacing δ by $\delta/2$ completes the proof of Lemma 2. □

Remark 4 We need not actually pick q vertices for each $V_i \times V_j$. Instead we pick $O(k^2 q)$ vertices from V and then let $\hat{d}_{i,j} = X_{i,j}/q_i q_j$ where q_i is the number of vertices chosen from V_i. Details are left to the final paper.

Proof of Lemma 3 Without loss of generality, assume that $\Delta(A,B) \geq \gamma n^2$. (The case $\Delta(A,B) \leq -\gamma n^2$ is handles exactly symmetrically.)

We obtain our new partition $\mathcal{P}'$ by replacing each V_i by the sets $V_i \cap A, V_i \cap B, V_i \setminus (A \cup B)$, empty sets not included. Thus the number of sets $k' \leq 3k$ as required.

Consider a fixed pair $1 \leq i < j \leq k$. For $X \subseteq V_i \times V_j$, $X \neq \emptyset$ let $e(X) = |X \cap E(G)|$, $d(X) = e(X)/|X|$ and define

$$\phi(X) = \frac{e(X)^2}{|X|} = d(X)^2|X|.$$

Thus

$$d_{i,j}^2 |V_i|\,|V_j| = \phi(V_i \times V_j).$$

If X is partitioned into non-empty sets $X_1 \cup X_2$ then

$$\begin{aligned} \phi(X_1) + \phi(X_2) - \phi(X) &= \frac{1}{|X|}\left(e(X_1)\sqrt{\frac{|X_2|}{|X_1|}} - e(X_2)\sqrt{\frac{|X_1|}{|X_2|}}\right)^2 \\ &= \frac{|X_1|\,|X_2|}{|X|}(d(X_1) - d(X_2))^2 \\ &= \frac{|X|\,|X_1|}{|X_2|}(d(X_1) - d(X))^2. \end{aligned} \tag{16}$$

It follows that if $C_{i,j} = A_i \times B_j$ and $\bar{C}_{i,j} = (V_i \times V_j) \setminus C_{i,j}$ then

$$\mathrm{ind}(\mathcal{P}') - \mathrm{ind}(\mathcal{P}) \geq \frac{1}{n^2} \sum_{i,j} (\phi(C_{i,j}) + \phi(\bar{C}_{i,j}) - \phi(V_i \times V_j))$$

(Interpret the summand as zero if $C_{i,j} = \emptyset$ or $V_i \times V_j$.)

$$\begin{aligned} &= \frac{1}{n^2} \sum_{i,j} \frac{|V_i|\,|V_j|\,|A_i|\,|B_j|}{|V_i|\,|V_j| - |A_i|\,|B_j|} (d(C_{i,j}) - d_{i,j})^2 \\ &= \frac{1}{n^2} \sum_{i,j} \frac{|V_i|\,|V_j|}{|A_i|\,|B_j|(|V_i|\,|V_j| - |A_i|\,|B_j|)} \Delta(A_i, B_j)^2 \\ &\geq \frac{4}{n^2} \sum_{i,j} \frac{\Delta(A_i, B_j)^2}{|V_i|\,|V_j|}. \end{aligned} \tag{17}$$

Now $\sum_{i,j} \Delta(A_i, B_j) \geq \gamma n^2$ and the minimum of (17) subject to this inequality is obtained by putting $\Delta(A_i, B_j) = \lambda |V_i|\,|V_j|$ where $\lambda = \gamma n^2 / \sum_{i,j} |V_i|\,|V_j|$. Substituting in (17) yields the lemma. □

5.1 Representing partitions

We can now explain our tree structure $\mathcal{T}$ which represents partitions. Objects at level l of $\mathcal{T}$ are indexed by superscript l. Thus at level l we keep $U_1'^{(l)}$, $U_2'^{(l)}$, $I^{(l)}$, $J^{(l)}$, $\sigma^{(l)}$ and $\hat{d}_{i,j}^{(l)}$ for $1 \leq i < j \leq k^{(l)}$. $\sigma^{(l)} = \pm$ indicating whether or not $\Delta(S, T)$ was positive in Lemma 2. We need to explain here, how, given l and knowledge that $v \in V_i^{(l)}$ we can determine which of $S^{(l)}$, $T^{(l)}$ or $\bar{S}^{(l)} \cap \bar{T}^{(l)}$ v now lies. Assume that $\sigma^{(l)} = +$ and let w be defined in terms of $\hat{d}_{i,j}^{(l)}$ as is done in Lemma 3. Then if $i \in I^{(l)}$ we compute $w(v, , U_2'^{(l)})$ and place v in $S^{(l)}$ if this is non-negative. We deal with $i \in J^{(l)}$ by computing $w(v, , U_1'^{(l)})$ etc. The depth of the tree is $O(\epsilon^{-2})$ and it take $O(p)$ time – p defined in (9) – to go down one level. It therefore takes $O(2^{O(1/\epsilon^{-2})})$ time to produce a new level of the tree. Thus on termination we have a data structure which answers the question: which subset contains vertex v? Each query takes $\tilde{O}(\epsilon^{-4})$ time. This is what we mean by an implicit description. An explicit description can then be obtained in $\tilde{O}(\epsilon^{-4} n)$ time. □

These times can be reduced to $\tilde{O}(\epsilon^{-2} n$ and $\tilde{O}(\epsilon^{-2} n)$ if we only estimate the $w(v, , U_j'^{(l)})$ – by sampling, of course – in the final construction. We leave the details to the final paper.

Remark 5 In some cases we need the subsets of the partition to be small e.g. $|V_i| \leq \nu$. This is achieved by starting with an arbitrary partion of V into sets of size ν or $\nu - 1$.

5.2 Multicoloured Version

The extension of Lemma 3 to the multicolored case is straightforward. We use the suffix i to extend our notation from G to G_i. The usual form of the Cauchy Schwartz inequality implies that if $\mathcal{P}'$ is a refinement of $\mathcal{P}$ then $\mathrm{ind}(\mathcal{P}') \geq \mathrm{ind}(\mathcal{P})$. So if we have a partition $\mathcal{P}$ and a pair of sets A, B and a graph G_i for which $|\Delta_i(A, B)| \geq \gamma n^2$, $\gamma = \epsilon/40$ then we can find a refinement

$\mathcal{P}'$ of $\mathcal{P}$ such that $\sum_i \text{ind}_i(\mathcal{P}') \geq \sum_i \text{ind}_i(\mathcal{P}) + \gamma n^2$ and so at most $(W+1)/\gamma^2$ iterations are needed to get an ϵ-sufficient partition w.r.t. all the G_i.

5.3 Equitable Partitions

Let a partition of V be equitable if $||V_i| - |V_j|| \leq 1$ for all i, j. The decomposition in Szemerédi's theorem is equitable. We show that equitability can be achieved at a small extra cost.

Lemma 5 *An equitable partition $\mathcal{P}$, with $\log k = \tilde{O}(\epsilon^{-2})$ can be obtain after $O(\epsilon^{-2})$ iterations.*

Proof After finding an ϵ-sufficient partition $\mathcal{P}$ as described above we take each V_i and partition it into $V_{i,j}$, $1 \leq j \leq s_i$ where $|V_{i,j}| = \lfloor \epsilon n/k \rfloor$ for $1 \leq j < s_i$ and $|V_{i,s_i}| < \epsilon n/k$ to obtained a partition $\mathcal{P}'$. Note that $|\bigcup_{i \in K} V_{i,s_i}| \leq \epsilon n$. Now (16) shows that $\text{ind}(\mathcal{P}') \geq \text{ind}(\mathcal{P})$. $\mathcal{P}'$ may not be ϵ-sufficient and so we repeat the whole process and make it so. After at most $200\epsilon^{-2}$ iterations we will have an ϵ-sufficient partition in which all but at most ϵn vertices lie in a collection of sets which are of equal size. These latter vertices are then equitably spread over the equal sized sets. The new partition is 2ϵ-sufficent. $\square$

6 Hypergraphs and other Max-SNP problems

A version of the Regularity Lemma has been proved for hypergraphs by Frankl and Rödl [12] and Chung [9]. This is however, non- constructive. We prove a constructive version and use this in a natural way to obtain PTAS's for all dense max-SNP problems. Again, [5] already gives PTAS's for these problems. Our algorithms follow in the same way from the Regularity Lemma for hypergraphs as the algorithms for graph partitioning problems followed from the Regularity Lemma for graphs. Unfortunately, we seem to need to consider *directed* hypergraphs.

Let $H = (V, E)$ be a s-uniform directed hypergraph i.e. $E \subseteq V^s$. Let $V_1, V_2, \ldots, V_k$ be a partition of V and for $I = (i_1, i_2, \ldots, i_s) \in K^s$ let

$$d_I = e(V_{i_1}, V_{i_2}, \ldots, V_{i_s}) / \prod_{t=1}^{s} |V_{i_t}|,$$

where for disjoint sets $A_1, A_2, \ldots, A_k$, $e(A_1, A_2, \ldots, A_s) = \{e = (a_1, a_2, \ldots, a_s) \in E : \ a_i \in A_i,\ 1 \leq i \leq s\}$. Let

$$\Delta(A_1, A_2, \ldots, A_s) = e(A_1, A_2, \ldots, A_s) - \sum_I d_I \prod_{t=1}^{s} |V_{i_t} \cap A_t|.$$

We say that $V_1, V_2, \ldots, V_k$ is ϵ-sufficient if

$$|\Delta(A_1, A_2, \ldots, A_s)| \leq \epsilon n^s \qquad \text{for all disjoint sets } A_1, A_2, \ldots, A_s \subseteq V.$$

Lemma 6 *There is a polynomial time algorithm for computing an ϵ-sufficient partition of a s-uniform hypergraph, assuming ϵ and s are fixed.*

Proof (sketch)

If $\mathcal{P}$ is not ϵ-sufficient then by considering a random ordered partition of K into s subsets $I_1, I_2, \ldots, I_s$ we can assume that there are $A_1, A_2, \ldots, A_s,\ A_t \subseteq V_{i_t}$ with $|\Delta(A_1, A_2, \ldots, A_s)| \geq \frac{\epsilon n^s}{s^s} = \gamma n^s$.

We then randomly choose $U_{2,1} \subseteq V_{i_2}, U_{3,1} \subseteq V_{i_3} \ldots, U_{s,1} \subseteq V_{i_s}$ of size $\tilde{O}(\gamma^{-2})$ and try all subsets of them to find $U'_{2,1}, U'_{3,1} \ldots, U'_{s,1}$ such that

$$\Delta(P(U'_{2,1}, U'_{3,1}) \ldots, U'_{s,1}), A_2, \ldots, A_s) \geq \Delta(A_1, A_2, \ldots, A_s) - \frac{\gamma n^s}{2s}$$

where P is defined analogously to its definition in Lemma 2. Putting $A' = P(U'_{2,1}, U'_{3,1}) \ldots, U'_{s,1})$ we then randomly choose $U_{1,2} \subseteq V_{i_1}, U_{3,2} \subseteq V_{i_3} \ldots, U_{s,2} \subseteq V_{i_s}$ and try all subsets of them to find $U'_{1,2}, U'_{3,2} \ldots, U'_{s,2}$ such that

$$\Delta(A'_1, P(U'_{1,2}, U'_{3,2}) \ldots, U'_{s,2}), A_3, \ldots, A_s) \geq \Delta(A'_1, A_2, \ldots, A_s) - \frac{\gamma n^s}{2s}.$$

Continuing we end up with $U'_{i,j},\ 1 \leq i \neq j \leq s$ which *define* our partition.

The index is defined analogously to the (undirected) case $s = 2$ and the equivalent of Lemma 3 can be shown. Details will be provided in the full paper. □

Let MAX-s-FUNCTION-SAT be the problem where the input consists of m Boolean functions $f_1, f_2, \ldots f_m$ in n variables, but where each f_i depends on only s variables (s fixed). The aim is to assign truth values to the n variables, so as to satisfy as many of the f_i as possible. It is well-known [18] that a Max-SNP problem can be viewed as a MAX-q-FUNCTION-SAT problem for a fixed s.

We may formulate the MAX-s-FUNCTION-SAT problem above in terms of "cuts" in a multi-coloured s-uniform hypergraph $H(V, E)$ as follows : H will have n vertices, one for each variable and m edges - one edge e_i corresponding to each function f_i. e_i will be the ordered $k-$tuple consisting of the arguments of f_i in (natural) order. There are at most $l = 2^{2^s}$ possible Boolean functions of s variables; we number them $1, 2, \ldots l$. Edge e_i will be coloured with colour p if f_i is the pth of these functions on its arguments. [So, the edges will be coloured with one of l colours.]

Suppose now we have a truth assignment $T : V \to \{0, 1\}$.

We may express each function in Disjunctive Normal Form. So based only on the colour p we can determine a subset Q_p of $\{0, 1\}^s$ such that the corresponding function $f_i(u_{i,1}, u_{i,2}, \ldots u_{i,s})$ is TRUE iff

$$(T(u_{i,1}), T(u_{i,2}), \ldots T(u_{i,s})) \in Q_p.$$

For an edge $e = (u_1, u_2, \ldots u_s)$, we let $T(e)$ denote the s-tuple $(T(u_1), T(u_2), \ldots T(u_s))$. Then we have

$$|\{i : f_i = 1 \text{ under } T\}| = \sum_p \sum_{a \in Q_p} |\{e : \text{colour}(e) = p;\ T(e) = a\}|.$$

We will use the above expression, i.e., we will approximately maximize the right hand side of the above expression using Lemma 6.

We note that

$$\begin{aligned} |\{e : \text{colour}(e) = p;\ T(e) = a\}| &= \sum_I |\{e : e \in V^I, \text{colour}(e) = p;\ T(e) = a\} \\ &\approx \sum_I d_{I,p} \prod_{t=1}^{s} |V_{i_t} \cap T^{-1}(a_t)|. \end{aligned}$$

Now we note that the last expression depends only upon **the number** of variables n_i set to true among those corresponding to vertices each V_i. Now we also note that the n_i need only be known approximately to evaluate the expression approximately. So, for each n_i we need to try only $1/\epsilon$ values and then choose the optimal values. The details are left to the final paper.

7 Open Problems

It will be interesting to see if our version of the Regularity Lemma can derive some of the consequences of the original version; if so, the constants will be substantially better.

It would be especially interesting (and we do not know if this is possible) if we could apply it to the situation that Szemerédi originally used it for - namely to give an asymptotic upper bound on the cardinality of any subset of $\{1, 2, \ldots n\}$ which is free of arithmetic progressions of 3 terms (and more generally any fixed number of terms.)

Acknowledgement: We thank David Karger for a helpful conversation.

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) 80-109.

[2] N.Alon, A.M.Frieze and D.Welsh, *Polynomial time randomised approximation schemes for the Tutte polynomial of dense graphs*, Random Structures and Algorithms 6 (1995) 459-478.

[3] J.Annan, *A randomised approximation algorithm for counting the number of forests in dense graphs*, to appear.

[4] S.Arora, A.M.Frieze and H.Kaplan, *A new rounding procedure for the assignment problem with applications to dense graph arrangement problems*, to appear.

[5] S.Arora, D.Karger and 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) 284-293.

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

[7] A.Z.Broder, *How hard is it to marry at random? (On the approximation of the permanent)*, Proceedings of the 18th Annual ACM Symposium on Theory of Computing (1986) 50-58. Erratum in Proceedings of the 20th Annual ACM Symposium on Theory of Computing, (1988) 551.

[8] R.E.Burkard and E.Cela, *Quadratic and 3-dimensional assignment problems*, In M Dell'Amico, F.Maffioli and S.Martello, editors, Annotated Bibliographies in Combinatorial Optimization, to appear.

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

[10] M.E.Dyer, A.M.Frieze and M.R.Jerrum, *Approximately counting Hamilton cycles in dense graphs*, Proceedings of the 4th Annual ACM/SIAM Symposium on Discrete Algorithms (1994) 336-343.

[11] W.Fernandez-de-la-Vega, *MAX-CUT has a Randomized Approximation Scheme in Dense Graphs*, Random Structures and Algorithms 8 (1996) 187-199.

[12] P. Frankl and V. Rödl, *The Uniformity Lemma for hypergraphs*, Graphs and Combinatorics 8 (1992) pp 309-312.

[13] M.R.Garey, D.S.Johnson and L.Stockmeyer, *Some simplified NP-Complete graph problems*, Theoretical Computer Science 1 (1976) 237-267.

[14] O.Goldreich, S.Goldwasser and D.Ron

[15] W.Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association

[16] M.R.Jerrum and A.Sinclair, *Approximating the permanent*, SIAM Journal on Computing 18 (1989) 1149-1178.

[17] J.Komlós and M.Simonovits, *Szemerédi's regularity Lemma and its applications in graph theory*, to appear.

[18] C. H. Papadimitriou and M. Yanakakis, *Optimization, approximation and complexity classes*, JCSS, 43 : pp 425-440 (1991)

[19] P.M.Pardalos, F.Rendl and H.Wolkowicz, *The quadratic assignment problem: a survey of recent developments*, In P.Pardalos and H.Wolkowicz editiors, *Quadratic assignment and related problems.* DIMACS Series in Discrete Mathematics and Theoretical Computer Science Volume 16 (1994) 1-42.

[20] E.Szemerédi, *Regular partitions of graphs*, Proceedings, Colloque Inter. CNRS (J.-C. Bermond, J.-C.Fournier, M.Las Vergnas and D.Sotteau, Eds.) (1978) 399-401.