UGP II Report: The Regularity lemma

April 24, 2025

Sanjna S | 230918

Contents

1 Introduction

This is a UGP Report on Regularity lemma.It is a summary of what I have learnt during the semester. Szemerédi's Regularity Lemma is a foundational result in graph theory that shows how large graphs can be partitioned in a structured way such that the edges distribution among these parts is similar to that in a random graph. This result is not very helpful in approximating sparse graphs as the more sparse a graph, the more the edge density approaches zero . In approximations involving dense graphs, the regularity lemma serves as a strong tool for various estimations. Szemerédi's original proof is not fully constructive. Later work by Alon et al. provided a polynomial time algorithm to construct such a partition of a graph. We will look at details of this construction broadly. We will continue by looking at few improved randomized algorithms developed by Frieze and Kannan and finally look at regularity conditions for hypergraphs given by Fischer et al.

2 The Regularity lemma

2.1 Preliminaries

- **Density Between Disjoint Sets:**Let $G = (V, E)$ be a graph, and let $X, Y \subseteq V$ be disjoint subsets. We denote by $\|X, Y\|$ the number of edges between X and Y, and define the **density** of the pair (X, Y) as:
$$d(X, Y) := \frac{\|X, Y\|}{|X||Y|}.$$
The value of density lies between 0 and 1.

- **ε-Regular Pairs:** Given $\varepsilon > 0$, a pair (A, B) of disjoint subsets $A, B \subseteq V$ is called **ε-regular** if for all subsets $X \subseteq A$ and $Y \subseteq B$ with:
$$|X| \geq \varepsilon|A| \quad \text{and} \quad |Y| \geq \varepsilon|B|,$$
it holds that:
$$|d(X, Y) - d(A, B)| \leq \varepsilon.$$
In an ε-Regular pair the edges are uniformly distributed in any between subsets of the pairs of a threshold size.

- **ε-Regular Partition:** Let $\{V_0, V_1, \dots, V_k\}$ be a partition of the vertex set V, where V_0 is an exceptional set (possibly empty). This partition is said to be an **ε-regular partition** if the following conditions are satisfied:

  1. $|V_0| \leq \varepsilon|V|$,
  2. $|V_1| = |V_2| = \cdots = |V_k|$,
  3. All but at most εk^2 of the pairs (V_i, V_j) with $1 \leq i < j \leq k$ are ε-regular.

The role of V_0 is of pure convenience. The vertices of V_0 are disregarded in the regularity lemma, but the size of V_0 is small. In an ϵ-regular partition, most of the pairs will be ϵ-regular.

2.2 Regularity lemma

Theorem: *For every $\epsilon > 0$ and every integer $m \geq 1$ there exists an integer M such that every graph of order at least m admits an ϵ-regular partition $\{V_0, V_1, \dots, V_k\}$ with $m \leq k \leq M$.*

The Regularity Lemma asserts that for any $\epsilon > 0$, every graph can be partitioned into a bounded number of parts such that the partition is ϵ-regular. This means that the graph's vertex set can be divided into a finite number of subsets where most pairs of these subsets behave in a uniform (or regular) manner with respect to edge distribution. The lemma guarantees the existence of an upper bound M on the number of parts, depending only on ϵ, which implies that the individual parts can be chosen large if the graph itself is large. It is important to note that ϵ-regularity is trivially satisfied when the partition consists of singletons, but the condition becomes significantly more meaningful and powerful when the subsets are large. Moreover, the lemma allows the specification of a lower bound m on the number of parts in the partition. This flexibility is useful because having more parts can increase the relative number of edges between different parts compared to edges within parts. This lemma is particularly useful in deriving results for dense graphs. For sparse graphs, all the densities just tend to zero

2.2.1 Outline of the proof

The partition is constructed using the following iterative algorithm:

- Start with the trivial partition (a single part).
- While the partition is not ε-regular:
  - For each pair (V_i, V_j) that is not ε-regular, identify subsets $A^{i,j} \subset V_i$ and $A^{j,i} \subset V_j$ that demonstrate the irregularity of (V_i, V_j).
  - Refine the partition simultaneously using all such subsets $A^{i,j}$.

If this process concludes after a finite number of steps, the regularity lemma is successfully established. To prove that the process halts in a bounded number of steps, we use a method known as the *energy increment argument*.

2.2.2 The Energy factor or the Index:

Let $U, W \subseteq V(G)$ and $n = |V(G)|$. We define the Energy factor or the index q(U, W), of two disjoint subsets as follows

$$q(U, W) = \frac{|U||W|}{n^2} d(U, W)^2.$$

For partitions $\mathcal{P}_U = \{U_1, \dots, U_k\}$ of U and $\mathcal{Q}_W = \{W_1, \dots, W_l\}$ of W, we define the energy as

$$q(\mathcal{P}_U, \mathcal{Q}_W) = \sum_{i=1}^{k} \sum_{j=1}^{l} q(U_i, W_j).$$

Finally, for a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of $V(G)$, define the *energy* of $\mathcal{P}$ to be $q(\mathcal{P}, \mathcal{P})$. Specifically,

$$q(\mathcal{P}) = \sum_{i=1}^{k} \sum_{j=1}^{k} q(V_i, V_j) = \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2.$$

2.3 Szemeredi's proof

. **Lemma 2.1:** Let $C, D \subseteq V$ be disjoint. If $\mathcal{C}$ is a partition of C and $\mathcal{D}$ is a partition of D, then $q(\mathcal{C}, \mathcal{D}) \geq q(C, D)$.

Proof. Let $\mathcal{C} = \{C_1, \dots, C_k\}$ and $\mathcal{D} = \{D_1, \dots, D_\ell\}$. Then

$$\begin{aligned} q(\mathcal{C}, \mathcal{D}) &= \sum_{i,j} q(C_i, D_j) \\ &= \frac{1}{n^2} \sum_{i,j} \frac{\|C_i, D_j\|^2}{|C_i||D_j|} \\ &\geq \frac{1}{n^2} \left(\frac{\sum_{i,j} \|C_i, D_j\|}{\sum_{i,j} |C_i||D_j|} \right)^2 \\ 8pt] &= \frac{1}{n^2} \cdot \frac{\|C, D\|^2}{\left(\sum_i |C_i|\right) \left(\sum_j |D_j|\right)} \\ &= \frac{1}{n^2} \cdot \frac{\|C, D\|^2}{|C||D|} \\ &= q(C, D) \end{aligned}$$

Lemma 2.2: If $\mathcal{P}, \mathcal{P}'$ are partitions of V and $\mathcal{P}'$ refines $\mathcal{P}$, then $q(\mathcal{P}') \geq q(\mathcal{P})$.

Proof. Let $\mathcal{P} = \{C_1, \dots, C_k\}$, and for each $i = 1, \dots, k$, let $\mathcal{P}'_i$ denote the restriction of the finer partition $\mathcal{P}'$ to C_i, i.e., a partition of C_i into smaller parts.

Then we can write:

$$q(\mathcal{P}) = \sum_{1 \leq i,j \leq k} q(C_i, C_j)$$

Since $\mathcal{P}'$ refines each C_i, we have:

$$q(C_i, C_j) \leq q(\mathcal{P}'_i, \mathcal{P}'_j)$$

$$q(\mathcal{P}) = \sum_{i,j} q(C_i, C_j) \leq \sum_{i,j} q(\mathcal{P}'_i, \mathcal{P}'_j) = q(\mathcal{P}').$$

This lemma establishes that the energy of a pair of partitions defined on the same vertex set is always at least as large as the energy computed over their coarser counterparts. In other words, refining a partition.That is, subdividing the vertex classes of an existing partition into smaller, more granular subsets cannot decrease the total energy. Formally, a partition P' is said to be a refinement of a partition P if every set in P' is contained within some set of P. The energy function, defined as a weighted sum of squared edge densities between pairs of subsets, increases or remains unchanged under such refinement.

Lemma 2.3: Let $\epsilon > 0$, and let $C, D \subseteq V$ be disjoint. If (C, D) is not ϵ-regular, then there are partitions

$$\mathcal{C} = \{C_1, C_2\} \text{ of } C \quad \text{and} \quad \mathcal{D} = \{D_1, D_2\} \text{ of } D$$

such that

$$q(\mathcal{C}, \mathcal{D}) \geq q(C, D) + \epsilon^4 \frac{|C||D|}{n^2}.$$

Proof

Since (C, D) is not ϵ-regular, there exist subsets $C_1 \subseteq C$, $D_1 \subseteq D$ with:

$$|C_1| \geq \epsilon|C|, \quad |D_1| \geq \epsilon|D|, \quad \text{and} \quad |d(C_1, D_1) - d(C, D)| > \epsilon.$$

Let us define the following subsets as shown

$$\mathcal{C} = \{C_1, C_2 = C \setminus C_1\}, \quad \mathcal{D} = \{D_1, D_2 = D \setminus D_1\}.$$

$$q(\mathcal{C}, \mathcal{D}) = \frac{1}{n^2} \sum_{i,j} \frac{e_{ij}^2}{|C_i||D_j|}.$$

This gives the bound:

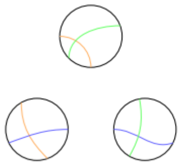
$$q(\mathcal{C}, \mathcal{D}) \geq q(C, D) + \epsilon^4 \frac{|C||D|}{n^2}.$$

Lemma 2.4 Let $0 < \epsilon \leq 1/4$, and let $\mathcal{P} = \{C_0, C_1, \ldots, C_k\}$ be a partition of V, with exceptional set C_0 of size $|C_0| \leq \epsilon n$ and $|C_1| = \ldots = |C_k| =: c$. If $\mathcal{P}$ is not ϵ-regular, then there is a partition

$$\mathcal{P}' = \{C_0', C_1', \ldots, C_\ell'\}$$

of V with exceptional set C_0', where $k \leq \ell \leq k^{4k+1}$, such that $|C_0'| \leq |C_0| + n/2^k$, all other sets C_i' have equal size, and either $\mathcal{P}'$ is ϵ-regular or

$$q(\mathcal{P}') \geq q(\mathcal{P}) + \epsilon^5/2.$$



The boundaries of irregular witnesses refine each part of the partition.

Proof: For all $1 \leq i < j \leq k$, let us define a partition C_{ij} of C_i and a partition C_{ji} of C_j, as follows. If the pair (C_i, C_j) is ϵ-regular, we let

$$C_{ij} := \{C_i\} \quad \text{and} \quad C_{ji} := \{C_j\}.$$

If not, then by Lemma 2.3 there are partitions C_{ij} of C_i and C_{ji} of C_j with $|C_{ij}| = |C_{ji}| = 2$ and

$$q(C_{ij}, C_{ji}) \geq q(C_i, C_j) + \epsilon^4 \frac{|C_i||C_j|}{n^2} = q(C_i, C_j) + \frac{\epsilon^4 c^2}{n^2}.$$

For each $i = 1, \dots, k$, let C_i' be the unique minimal partition of C_i that refines every partition C_{ij} with $j \neq i$.

Thus, $|C_i| \leq 2^{k-1}$. Now consider the partition

$$\mathcal{C} := \{C_0\} \cup \bigcup_{i=1}^{k} C_i'$$

of V, with C_0 as exceptional set. Then $\mathcal{C}$ refines $\mathcal{P}$ and $|\mathcal{C} \setminus \{C_0\}| \leq k2^{k-1}$, so

$$k \leq |\mathcal{C}| \leq k2^{k-1}.$$

Let $C_0 := \{\{v\} \mid v \in C_0\}$. If $\mathcal{P}$ is not ϵ-regular, then for more than ϵk^2 of the pairs (C_i, C_j) with $1 \leq i < j \leq k$, the partition C_{ij} is non-trivial.

Hence,

$$\begin{aligned}
q(\mathcal{C}) &= \sum_{1 \leq i < j} q(C_i', C_j') + \sum_{1 \leq i} q(C_0, C_i') + \sum_{0 \leq i} q(C_i') \\
&\geq \sum_{1 \leq i < j} q(C_{ij}, C_{ji}) + \sum_{1 \leq i} q(C_0, \{C_i\}) + q(C_0) \\
&\geq \sum_{1 \leq i < j} q(C_i, C_j) + \epsilon k^2 \frac{\epsilon^4 c^2}{n^2} + \sum_{1 \leq i} q(C_0, \{C_i\}) + q(C_0) \\
&= q(\mathcal{P}) + \epsilon^5 \left(\frac{kc}{n}\right)^2 \geq q(\mathcal{P}) + \epsilon^5/2.
\end{aligned}$$

To form a proper partition $\mathcal{P}'$, we split sets in $\mathcal{C}$ into equal-sized parts.

If $c < 4^k$, let $\mathcal{P}'$ be the ϵ-regular partition into $C_0' := C_0$ and singletons. Else, suppose $c \geq 4^k$. Let $C_1', \ldots, C_\ell'$ be a maximal collection of disjoint sets of size $d := \lfloor c/4^k \rfloor \geq 1$, such that each C_i' is in some $C \in \mathcal{C} \setminus \{C_0\}$. Then define

$$C_0' := V \setminus \bigcup C_i', \quad \mathcal{P}' := \{C_0', C_1', \ldots, C_\ell'\}.$$

Then $\mathcal{P}'$ refines $\mathcal{C}$, and

$$q(\mathcal{P}') \geq q(\mathcal{C}) \geq q(\mathcal{P}) + \epsilon^5/2.$$

Also,

$$|C_0'| \leq |C_0| + d|\mathcal{C}| \leq |C_0| + \frac{c}{4^k} k 2^k.$$

Proof of regularity lemma: Let $\epsilon > 0$ and $m \geq 1$ be given, and assume wlog that $\epsilon \leq 1/4$. Let $s := 2/\epsilon^5$.

A key requirement: the partition $\{C_0, C_1, \ldots, C_k\}$ with $|C_1| = \cdots = |C_k|$ must satisfy Lemma 2.4's precondition:

$$|C_0| \leq \epsilon n.$$

Each application of the lemma increases $|C_0|$ by at most $n/2^k$. Thus we must choose k large enough so that s such increments plus the initial size remain $\leq \epsilon n$.

This leads to:

$$k + \frac{sn}{2^k} \leq \epsilon n.$$

Choose $k \geq m$ such that:

$$2^{k-1} \geq \frac{s}{\epsilon}.$$

Then:

$$\frac{s}{2^k} \leq \frac{\epsilon}{2}, \quad \text{and} \quad k + \frac{sn}{2^k} \leq \epsilon n$$

Now select M to bound the number of non-exceptional sets after up to s iterations.

Define $f(r) = 4r^4 + 1$ and let:

$$M := \max\{f^s(k), 2k/\epsilon\}.$$

This ensures $n \geq M$ is large enough for the process.

For any graph $G = (V, E)$ with $|V| = n \geq m$, if $n \leq M$, partition V into singletons: $V = \{V_1, \ldots, V_n\}$ with $V_0 = \emptyset$. This is trivially ϵ-regular.

If $n > M$, choose $C_0 \subseteq V$ minimal so k divides $|V \setminus C_0|$, and partition the remainder evenly.

Applying the lemma2.4 repeatedly on $\{C_0, C_1, \ldots, C_k\}$.We can clearly see that the exceptional set size remains $\leq \epsilon n$, and the lemma can be re-applied at most s times until the partition is ϵ-regular.

3 Applications of Regularity lemma

In this section, we explore The Blow up Lemma and Szemeredi's Theorem on arithmetic progressions. Notably, the Regularity Lemma was originally discovered by Szemeredi in the course of his efforts to prove his celebrated theorem on arithmetic progressions.

3.1 Regularity Graph and Its Blow-up

The Regularity Graph: Let G be a graph with an ϵ-regular partition $\{V_0, V_1, \dots, V_k\}$, where:

- V_0 is an exceptional set with $|V_0| \le \epsilon|V|$,
- $|V_1| = \cdots = |V_k| = \ell$.

Given $d \in [0, 1]$, define a graph R on vertex set $\{1, \dots, k\}$, where $i \sim j$ if:

- the pair (V_i, V_j) is ϵ-regular, and
- the edge density between V_i and V_j is at least d.

We call R the **regularity graph** of G with parameters ϵ, ℓ, d.

Blow-up of the Regularity Graph: For any $s \in \mathbb{N}$, construct R_s by:

- Replacing each vertex i of R with a set V_i^s of s vertices,
- Replacing each edge by a complete bipartite graph between corresponding s-sets.

Lemma 3.1 Let (A, B) be an ϵ-regular pair, of density d say, and let $Y \subseteq B$ have size $|Y| \ge \epsilon|B|$. Then all but fewer than $\epsilon|A|$ of the vertices in A have (each) at least $(d - \epsilon)|Y|$ neighbours in Y.

The proof of the above stated lemma is elementary and we will skip this.

Lemma 3.2 (Blow-up Lemma) For all $d \in (0, 1]$ and $\Delta \ge 1$ there exists an $\epsilon_0 > 0$ with the following property: if G is any graph, H is a graph with $\Delta(H) \le \Delta$, $s \in \mathbb{N}$, and R is any regularity graph of G with parameters $\epsilon \le \epsilon_0$, $\ell \ge 2s/d^\Delta$ and d, then

$$H \subseteq R_s \quad \Rightarrow \quad H \subseteq G.$$

Outline of the proof : We prove the above lemma by showing that there exists a possible embedding of graph H in G by inductively mapping the vertices of H to those in G. During this process using lemma 3.1, we conclude that there are enough vertices as options for a successful mapping to happen.

Setup of the Embedding Argument: Given d and Δ, choose $\epsilon_0 > 0$ small enough so that:

$$(d - \epsilon_0)^\Delta - \Delta\epsilon_0 \ge \frac{1}{2}d^\Delta.$$

Clearly this is always possible as when ϵ_0 approaches zero we approach the desired result.

Let G, H, s, and R be given. Let $\{V_0, V_1, \dots, V_k\}$ be an ϵ-regular partition of G corresponding to R, with $\epsilon < \epsilon_0$ and $V(R) = \{1, \dots, k\}$. Assume H is a subgraph of R_s with vertices $u_1, \dots, u_h$.

Our goal is to embed H into G by mapping $u_i \mapsto v_i \in V_{\sigma(i)}$ such that $v_i v_j$ is an edge in G whenever $u_i u_j$ is an edge in H.

Embedding Strategy and Target Sets: We inductively choose $v_1, \dots, v_h$. At each step i, maintain a target set $Y_i \subseteq V_{\sigma(i)}$ of valid candidates for v_i.

Initially, $Y_i = V_{\sigma(i)}$. As we embed, we update Y_i by removing vertices not adjacent to previously embedded v_j's:

$$V_{\sigma(i)} = Y_i^0 \supseteq Y_i^1 \supseteq \cdots \supseteq Y_i^i = \{v_i\}.$$

The sets shrink, but we aim to show that despite the shrinkage each Y_i^j remains large enough for a successful embedding.

Controlling Target Set Shrinkage : To keep Y_i large, when embedding u_j with $j < i$, and $u_j u_i \in E(H)$, define:

$$Y_i^j = N(v_j) \cap Y_i^{j-1}.$$

We want:

$$|Y_i^j| \geq (d - \epsilon)|Y_i^{j-1}|.$$

By Lemma 3.1, all but $\ell\epsilon$ bad choices of v_j (from $Y_j^{i-1} \subseteq V_{\sigma(j)}$) satisfy this condition.

Thus, at each step we preserve most of Y_i provided we avoid bad choices.

Bounding Target Set Size: We iterate this shrinking over at most Δ neighbors:

$$|Y_i^{j-1}| - \Delta\ell\epsilon \geq (d - \epsilon)^\Delta \ell - \Delta\ell\epsilon \geq (d - \epsilon_0)^\Delta \ell - \Delta\epsilon_0\ell \geq \frac{1}{2} d^\Delta \ell \geq s.$$

This ensures each Y_i stays sufficiently large to select v_i, avoiding previously used vertices and maintaining adjacency.

All Y_i^j stay $\geq \ell$ or $> s$, enabling valid embedding of H into G. Thus proving the theorem.

3.2 Szemeredi's Theorem

Upper Density:

For a subset $A \subseteq \mathbb{N}$, the **upper density** is defined as:

$$\overline{d}(A) = \limsup_{n \to \infty} \frac{|A \cap \{1, 2, \dots, n\}|}{n}$$

- Measures how "dense" A is in the set of natural numbers.

- A set A has positive upper density if $\overline{d}(A) > 0$.
- In any partition $\{A_1, \dots, A_p\}$ of $\mathbb{N}$, one A_i will satisfy $|A_i \cap n| \geq \delta n$ for arbitrarily large n and some $\delta > 0$.

Theorem: (Szemerédi 1975) *Every set $A \subseteq \mathbb{N}$ of positive upper density contains arbitrarily long arithmetic progressions.*
The proof of the above stated theorem is beyond the scope of this report.

4 Constructive version of the Regularity lemma

In This section, We will go over an approximate construction of a regular partition for a graph given by Alon et al.
Theorem : For every $\varepsilon > 0$ and every positive integer t, there exists an integer $Q = Q(\varepsilon, t)$ such that every graph with $n > Q$ vertices has an ε-regular partition into $k + 1$ classes, where $t \leq k \leq Q$.

Furthermore, for every fixed $\varepsilon > 0$ and $t \geq 1$, such a partition can be found:

in $O(M(n))$ sequential time, where $M(n)$ is the time to multiply two $n \times n$ 0–1 matrices over the integers; or in $O(\log n)$ time on an EREW PRAM using a polynomial number of parallel processors.

Algorithm Behavior for Approximating Regularity: Verifying if two disjoint sets form an εregular pair is a Co-NP complete problem. But we can find an εregular partition for a given graph using approximation algorithms that run in polynomial time. Given a graph G and a parameter $\varepsilon > 0$, the algorithm computes a refined threshold $\varepsilon_0 < \varepsilon$, where ε_0 depends on ε.

- If G is not ε-regular, the algorithm correctly detects this and produces a certificate showing that G is not ε_0-regular.
- If G is ε_0-regular (and hence ε-regular), the algorithm declares that G is ε-regular.
- If G is ε-regular but not ε_0-regular, the algorithm may behave in either of the above two ways, and this behavior is non-deterministic.

Neighbourhood deviation: Let H be a bipartite graph with equal color classes $|A| = |B| = n$. Let d be the average degree of H.

For two distinct vertices $y_1, y_2 \in B$, define the *neighbourhood deviation* of y_1 and y_2 by

$$\sigma(y_1, y_2) = |N(y_1) \cap N(y_2)| - \frac{d^2}{n}.$$

For a subset $Y \subset B$, denote the *deviation* of Y by

$$\sigma(Y) = \frac{\sum_{y_1, y_2 \in Y} \sigma(y_1, y_2)}{|Y|^2}.$$

Intuitively, we can see that the term $\frac{d^2}{n}$ represents the average number of neighbours a vertex will have in common with another vertex belonging to the same partition of the bipartite graph. Sigma function here measures the deviation from the norm.

Lemma 4.1*Let H be a bipartite graph with equal classes $|A| = |B| = n$, and let d denote the average degree of H. Let $0 < \varepsilon < \frac{1}{16}$. If there exists $Y \subset B$, $|Y| \geq \varepsilon n$ such that $\sigma(Y) \geq \frac{\varepsilon^3}{2}n$, then at least one of the following cases occurs.*

1. $d < \varepsilon^3 n$.
2. There exists in B a set of more than $\frac{1}{8}\varepsilon^4 n$ vertices whose degrees deviate from d by at least $\varepsilon^4 n$.
3. There are subsets $A' \subset A$, $B' \subset B$, $|A'| \geq \frac{\varepsilon^4}{4}n$, $|B'| \geq \frac{\varepsilon^4}{4}n$ and $|d(A', B') - d(A, B)| \geq \varepsilon^4$.

Moreover, there is an algorithm whose input is a graph H with a set $Y \subset B$ as above that outputs either

(i) The fact that 1 holds, or

(ii) The fact that 2 holds and a subset of more than $\frac{1}{8}\varepsilon^4 n$ members of B demonstrating this fact, or

(iii) The fact that 3 holds and two subsets A' and B' as in 3 demonstrating this fact.

The algorithm runs in sequential time $O(M(n))$, where $M(n) = O(n^{2.376})$ is the time needed to multiply two $n \times n$ 0,1 matrices over the integers.

Proof. We begin by assuming that cases 1 and 2 do not occur, and we aim to show that case 3 must therefore hold.

Define the set $Y' = \{y \in Y \mid |\deg(y) - d| < \epsilon^4 n\}$. Since case 2 is ruled out, Y' is non-empty.

Select a vertex $y_0 \in Y'$ that maximizes the sum $\sum_{y \in Y} \sigma(y_0, y)$. We estimate this sum as follows. Clearly, we have:

$$\sum_{y' \in Y'} \sum_{\substack{y \in Y' \\ y \neq y'}} \sigma(y', y) = \sigma(Y', Y')|Y'|^2 - \sum_{y' \in Y'} \sum_{\substack{y \in Y' \\ y \neq y'}} \sigma(y', y) \geq \frac{\epsilon^3}{2} n|Y|^2 - \frac{\epsilon^4}{8} n|Y|n.$$

Since $|Y'| \leq |Y|$, it follows that:

$$\sum_{y \in Y} \sigma(y_0, y) \geq \frac{\epsilon^3}{2} n|Y| - \frac{\epsilon^4}{8} n^2 \geq \frac{3}{8}\epsilon^3 n|Y|.$$

There must be at least $\frac{\epsilon^4}{2}n$ vertices $y \in Y$ such that the neighborhood deviation from y_0 is more than $2\epsilon^4 n$. Otherwise, the total deviation would be bounded above by:

$$\sum_{y \in Y} \sigma(y_0, y) \leq \frac{\epsilon^4}{4} n^2 + |Y| \cdot 2\epsilon^4 n \leq \frac{\epsilon^3}{4} n|Y| + 2\epsilon^4 n|Y| < \frac{3}{8}\epsilon^3 n|Y|,$$

which contradicts the previous lower bound.

Hence there is a set $B' \subset Y$, $|B'| \geq \frac{\epsilon^4}{4}n$, $y_0 \notin B'$, and for every vertex $b \in B'$ we have $|N(b) \cap N(y_0)| > \frac{d^2}{n} + 2\epsilon^4 n$. Define $A' = N(y_0)$. Clearly,

$$|A'| \geq d - \epsilon^4 n \geq \epsilon^3 n - \epsilon^4 n \geq 15\epsilon^4 n \geq \frac{\epsilon^4}{4}n.$$

We will show that $|d(A', B') - d(A, B)| \geq \epsilon^4$. Indeed,

$$e(A', B') = \sum_{b \in B'} |N(y_0) \cap N(b)| > \frac{|B'|d^2}{n} + 2\epsilon^4 n|B'|.$$

Therefore,

$$\begin{aligned} d(A', B') - d(A, B) &> \frac{d^2}{n|A'|} + \frac{2\epsilon^4 n}{|A'|} - \frac{d}{n} \\ &\geq \frac{d^2}{n(d + \epsilon^4 n)} + 2\epsilon^4 - \frac{d}{n} \\ &= 2\epsilon^4 - \frac{d\epsilon^4}{d + \epsilon^4 n} \geq \epsilon^4. \end{aligned}$$

Existence of Sequential Algorithm The existence of the required sequential algorithm is simple. One can clearly check if 1 holds in time $O(n^2)$. Similarly, it is trivial to check if 2 holds in $O(n^2)$ time, and in case it holds to exhibit the required subset of B establishing this fact. If both cases above fail we continue as follows.

For each $y_0 \in B$ with $|\deg(y_0) - d| < \epsilon^4 n$ we find the set of vertices $B_{y_0} = \{y \in B \mid \sigma(y_0, y) \geq 2\epsilon^4 n\}$.

The last proof guarantees the existence of at least one such y_0 for which $|B_{y_0}| \geq \frac{\epsilon^4}{4}n$. The subsets $B' = B_{y_0}$ and $A' = N(y_0)$ are the required ones. Since the computation of all the quantities $\sigma(y, y')$ for $y, y' \in B$ can be done by squaring the adjacency matrix of H, the claimed sequential running time follows. The parallelization is obvious.

Lemma 4.2 Let H be a bipartite graph with equal classes $|A| = |B| = n$. Let $2n^{-1/4} < \epsilon < \frac{1}{16}$.

Lemma 4.3 Let H be a bipartite graph with equal classes $|A| = |B| = n$. Let $2n^{-1/4} < \epsilon < \frac{1}{16}$,There is an $O(M(n))$ algorithm that verifies that H is ϵ-regular or finds two subsets $A' \subset A$, $B' \subset B$, $|A'| \geq \frac{\epsilon^4}{16}n$, $|B'| \geq \frac{\epsilon^4}{16}n$, such that $|d(A, B) - d(A', B')| \geq \epsilon^4$. The algorithm can be parallelized and implemented in NC^1.

Proof. We begin by computing d, the average degree of H. If $d < \epsilon^4 n$, H is ϵ-regular, and we are done.

Next, we count the number of vertices in B whose degrees deviate from d by at least $\epsilon^4 n$. If there are more than $\frac{\epsilon^4}{8}n$ such vertices, then the degrees of at least half of them deviate in the same direction and if we let B' be such a set of vertices, then $|B'| \geq \frac{\epsilon^4}{16}n$. A simple computation yields that

$$|d(B', A) - d(B, A)| \geq \epsilon^4,$$

and we are done.

By Lemma 4.2, it is now sufficient to show that if there exists $Y \subset B$ with $|Y| \geq \epsilon n$ and $\sigma(Y) \geq \frac{\epsilon^3 n}{2}$, then one can find in $O(M(n))$ time the required subsets A' and B'. But this follows from the assertion of Lemma 4.1. The parallelization is immediate.

Lemma 4.4 Fix k and γ and let $G = (V, E)$ be a graph with n vertices. Let P be an equitable partition of V into classes $C_0, C_1, \ldots, C_k$. Assume $|C_1| > 42^k$ and $4^k > 600\gamma^{-5}$. Given proofs that more than γk^2 pairs (C_r, C_s) are not γ-regular (where by proofs we mean subsets $X = X(r,s) \subset C_r$, $Y = Y(r,s) \subset C_s$ that violate the condition of γ-regularity of (C_r, C_s)), then one can find in $O(n)$ time a partition P' (which is a refinement of P) into $1 + k4^k$ classes, with an exceptional class of cardinality at most

$$|C_0| + \frac{n}{4^k}$$

and such that

$$\text{ind}(P') \geq \text{ind}(P) + \frac{\gamma^5}{20}.$$

The proof will closely resemble lemma 2.4, so we are omitting it here.

Regularity Partitioning Algorithm

1. Arbitrarily divide the vertices of G into an equitable partition P_1 with classes $C_0, C_1, \ldots, C_b$, where $|C_1| = \lfloor n/b \rfloor$ and hence $|C_0| < b$. Denote $k_1 = b$.
2. For every pair (C_r, C_s) of P_i, verify if it is ϵ-regular or find $X \subset C_r$, $Y \subset C_s$, $|X| \geq \frac{\epsilon^4}{16}|C_1|$, $|Y| \geq \frac{\epsilon^4}{16}|C_1|$, such that $|d(X, Y) - d(C_s, C_t)| \geq \epsilon^4$.
3. If there are at most $\epsilon\binom{k_i}{2}$ pairs that are not verified as ϵ-regular, then halt. P_i is an ϵ-regular partition.
4. Apply Lemma 3.4 where $P = P_i$, $k = k_i$, $\gamma = \frac{\epsilon^4}{16}$, and obtain a partition P' with $1 + k_i 4^{k_i}$ classes.
5. Let $k_{i+1} = k_i 4^{k_i}$, $P_{i+1} = P'$, $i = i + 1$, and go to step 2.

Runtime and Iterations Finally, we note that the number of iterations is constant (does not depend on n), and that the running time of each iteration is bounded by the $O(M(c))$ bound of Corollary 3.3, where c is the size of the classes in the equitable partition, which is less than n.

5 Applications in Computer Science

5.1 Maximum Weight Partition

The **Maximum Weight Partition Problem** is a fundamental combinatorial optimization task that arises in the analysis of large-scale data and structures such as graphs and matrices. Given a set of elements and a weight function that assigns a real-valued weight to each pair of elements, the objective is to partition the set into disjoint subsets so as to maximize the sum of weights of the intra-part interactions.

Formally, let V be a set of n elements, and let $w : V \times V \to \mathbb{R}$ be a symmetric weight function (i.e., $w(u,v) = w(v,u)$). The goal is to partition V into subsets $V_1, V_2, \dots, V_k$ such that the total intra-cluster weight

$$\sum_{i=1}^{k} \sum_{u,v \in V_i} w(u,v)$$

is maximized. This problem is NP-hard in general, and finding optimal partitions requires exploring exponentially many possibilities. However, for certain classes of graphs,especially dense graphs approximation schemes such as those derived from the Regularity Lemma can be used to produce nearly optimal partitions in polynomial time.

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

$$w(S_\epsilon) \geq w(S^*) - \epsilon n^2.$$

Here S^* is the maximum weight partition.

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

$$\alpha(\epsilon) = \tilde{O}(\epsilon^{-2}) \quad \text{and} \quad \log \beta = \tilde{O}(\epsilon^{-2}).$$

5.2 Metric Quadratic Assignment Problem (Metric QAP)

The **Metric Quadratic Assignment Problem** is a well-studied variant of the Quadratic Assignment Problem (QAP), which is one of the most challenging problems in combinatorial optimization. The classical QAP involves assigning n facilities to n locations with the goal of minimizing the total cost, where the cost is a quadratic function dependent on both the distance between locations and the flow between facilities.

In the *metric version*, the distance function $d(i,j)$ satisfies the triangle inequality (i.e., it is a metric), which allows for leveraging geometric properties in designing approximation algorithms.

Formally, given:

- A flow matrix $F = (f_{ij})$, where f_{ij} represents the flow between facilities i and j,

- A distance matrix $D = (d_{kl})$, where d_{kl} is the distance between locations k and l, and
- A bijection $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$, representing an assignment of facilities to locations,

the goal is to minimize the cost:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \cdot d_{\pi(i)\pi(j)}.$$

When the distance matrix D defines a metric space, it is known that polynomial-time approximation schemes (PTAS) can be constructed for this problem in specific settings, especially when coupled with the structure imposed by the Regularity Lemma or other decomposition techniques.

Theorem : There is a randomised algorithm $A_3(\epsilon)$ for the metric QAP which, with probability at least 3/4, produces a permutation π_ϵ such that

$$c(\pi_\epsilon) \le c(\pi^*) + \epsilon n^2$$

and which runs in time at most $\alpha_2(1/\epsilon)n + \beta_3(1/\epsilon)$.

6 Hypergraph regularity

ϵ-Edge density in Hypergraphs: Let H be an r-uniform hypergraph. Let $(U_1, \dots, U_r)$ be an r-tuple of vertex sets of H. Define the edge density:

$$d(U_1, \dots, U_r) = \frac{\text{number of edges with one vertex in each } U_i}{\prod_{i=1}^{r} |U_i|}$$

ϵ-Regularity of a Tuple: The tuple $(U_1, \dots, U_r)$ is *ϵ-regular* if for every subtuple $(U'_1, \dots, U'_r)$ such that:

$$U'_i \subseteq U_i \quad \text{and} \quad |U'_i| \ge \epsilon |U_i| \quad \forall i,$$

$$d(U'_1, \dots, U'_r) = d(U_1, \dots, U_r) \pm \epsilon.]$$

ϵ-Regular Partition: An equipartition $V_1, \dots, V_k$ of the vertices of H is ϵ-regular if all but at most $\epsilon\binom{k}{r}$ of the r-tuples $(V_{i_1}, \dots, V_{i_r})$ are ϵ-regular.

Index of an Equipartition: Let $A = (V_1, \dots, V_k)$ be an equipartition of the vertex set of an r-uniform hypergraph G. We Define the index of A as:

$$\text{ind}(A) = k^{-r} \sum_{1 \le i_1 < \cdots < i_r \le k} \left(d(V_{i_1}, \dots, V_{i_r}) \right)^2$$

Theorem : For any fixed r and $\varepsilon > 0$, there exists an $O(n)$-time randomized algorithm that: Finds an ε-regular partition of an r-uniform hypergraph with n vertices.

The lemma states that for any fixed r and $\varepsilon > 0$, there exists a randomized algorithm that can find an ε-regular partition of an r-uniform hypergraph in expected $O(n)$ time. An ε-regular partition ensures that the edge distribution between most parts of the hypergraph approximates randomness, capturing the essence of pseudorandom structure. This result is significant because it not only guarantees the existence of such partitions but also provides an efficient way to construct them.

7 Conclusion

Szemerédi's Regularity Lemma is a powerful idea in graph theory that helps us understand large, complicated graphs by breaking them into smaller, more uniform pieces. This kind of structure, which looks almost random, plays a key role in many important results in combinatorics, including Szemerédi's Theorem on arithmetic progressions.Although the original proof wasn't constructive, later work by Alon, Frieze, and Kannan gave us ways to actually build these regular partitions using algorithms. These methods make the lemma more useful, especially in computer science, where they help solve problems that would otherwise be too complex.

In this report, we looked at applications like the Maximum Weight Partition and the Metric QAP, which show how the Regularity Lemma can lead to good approximations for tough problems. We also explored how the idea extends to hypergraphs, showing just how far this concept can go.All in all, the Regularity Lemma is an amazing example of how deep mathematical ideas can help us make sense of large systems, connect randomness and structure, and find new solutions in both math and computer science.

References

[1] Noga Alon, Richard A Duke, Hans Lefmann, Vojtěch Rödl, and Raphael Yuster. The algorithmic aspects of the regularity lemma. *Journal of Algorithms*, 16(1):80–109, 1994.

[2] Reinhard Diestel. *Graph Theory*. Graduate Texts in Mathematics. Springer, 5th edition, 2017. ISBN 9783662536216.

[3] Eldar Fischer, Arie Matsliah, and Raphael Yuster. Approximate hypergraph partitioning and applications. *SIAM Journal on Computing*, 36(4):889–922, 2006.

[4] Alan Frieze and Ravi Kannan. The regularity lemma and approximation schemes for dense problems. *Proceedings of the 37th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 12–20, 1996.

[2] [1] [4] [3]