

# Szemerédi Regularity Lemma

Aakash Ghosh(19MS129)

October 15, 2022

Let  $G$  be a graph and let  $A$  and  $B$  be disjoint set of vertices of  $G$  and let  $e_G(A, B)$  be the number of edges from  $A$  to  $B$  in  $G$ . Define a density function  $d(A, B) = \frac{e_G(A, B)}{|A||B|}$

**Remark:** Note that  $d(A, B)$  is a somewhat normalised measure of how well two vertex set is connected with each other. The value of  $d$  will range from 0(i.e no connection) to 1(i.e. all possible edges are present and very well-connected)

**Lemma 1.** If  $y \in [a, b]$  then

$$|x - y| \leq c \forall y \in [a, b] \Leftrightarrow |x - a| \leq c, |x - b| \leq c \quad (1)$$

*Proof.* There are three possible cases as shown in Fig 1. The blue line shows the values  $y$  takes.  $\square$

**Theorem 2.** If  $X \subset A$  and  $Y \subset B$  with  $\frac{|X|}{|A|} \geq 1 - \delta$  and  $\frac{|Y|}{|B|} \geq 1 - \delta$  where  $0 \leq \delta \leq \frac{1}{2}$ . Then:

$$|d(X, Y) - d(A, B)| \leq 2\delta \quad (2)$$

$$|d^2(X, Y) - d^2(A, B)| \leq 4\delta \quad (3)$$

**Proof outline:** We try to go the other way around: we fix edges of  $X, Y$  and try to build edges of  $A, B$  around them. We note that since  $d(A, B)$  increases with the number of edges added in the extension of the old graph, we just need to make sure that inequality holds for the maximum and minimum values only(lemma 1). We find bounds for this maximum and minimum values to complete the proof. Eqn3 follows as a consequence of Eqn2 and bounds on  $d(A, B), d(X, Y)$ .

*Proof.* Let  $d(X, Y) = d'$ . Note that:

$$e_G(A, B) = e_G(X, Y) + e_G(X, B \setminus Y) + e_G(A \setminus X, Y) + e_G(A \setminus X, B \setminus Y) \quad (4)$$

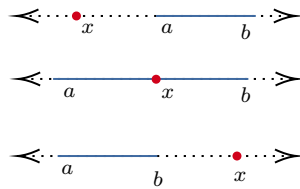


Figure 1: The three cases required to prove lemma 1

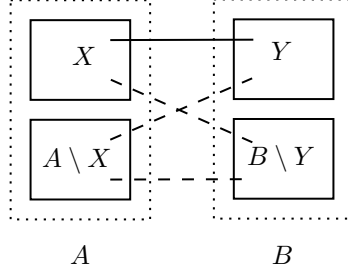


Figure 2: Each line represents all edges between the sets it joins. Edges represented by the solid line is given. Maximum is attained when all edges represented by the dotted line is present. Minimum is attained when no edge of represented by dotted lines are given.

We shall first try to prove the inequality when  $d(A, B) = d_1$  is maximum. It is easy to see this occurs when each term of eqn4 attains their maximum value. Note only  $e_G(X, Y)$  is fixed. For the other terms we have:

$$e_G(X, B \setminus Y) \leq (|X|)(|B| - |Y|) \quad (5)$$

$$e_G(A \setminus X, Y) \leq (|A| - |X|)(|Y|) \quad (6)$$

$$e_G(A \setminus X, B \setminus Y) \leq (|A| - |X|)(|B| - |Y|) \quad (7)$$

Where maximum is attained when all the respective edges are drawn. Putting in eqn4 and simplifying we get:

$$d_1 |A| |B| = e_G(X, Y) + |A| |B| - |X| |Y| \quad (8)$$

$$\Rightarrow d_1 |A| |B| = d' |X| |Y| + |A| |B| - |X| |Y| = |A| |B| - (1 - d') |X| |Y| \quad (9)$$

$$\Rightarrow d_1 = 1 - (1 - d') \left( \frac{|X|}{|A|} \right) \left( \frac{|Y|}{|B|} \right) \quad (10)$$

We know  $1 \geq \left( \frac{|X|}{|A|} \right), \left( \frac{|Y|}{|B|} \right) \geq 1 - \delta$ . From this we get  $d_1 \geq d'$ . Therefore,

$$\begin{aligned} |d_1 - d'| &= d_1 - d' \\ &\leq 1 - (1 - d') \left( \frac{|X|}{|A|} \right) \left( \frac{|Y|}{|B|} \right) - d' \\ &\leq 1 - (1 - d')(1 - \delta)^2 - d' \\ &\leq (1 - d')(1 - (1 - \delta)^2) \\ &\leq (1 - (1 - \delta)^2) \\ &\leq 2\delta - \delta^2 \leq 2\delta \end{aligned} \quad (11)$$

Now we prove it when  $d(A, B) = d_2$  is minimum. This occurs when each term of eqn4 attains minimum value.  $e_G(X, Y)$  is fixed. For the other terms the minimum value attained is 0. Therefore,

$$d_2 |A| |B| = e_G(X, Y) = d' |X| |Y| \quad (12)$$

$$\Rightarrow d_2 = d' \left( \frac{|X|}{|A|} \right) \left( \frac{|Y|}{|B|} \right) \quad (13)$$

From the bounds of  $\left(\frac{|X|}{|A|}\right), \left(\frac{|Y|}{|B|}\right)$  we get  $d_2 \leq d'$ . Therefore,

$$\begin{aligned}
|d_2 - d'| &= d' - d_2 \\
&= d' - d' \left(\frac{|X|}{|A|}\right) \left(\frac{|Y|}{|B|}\right) \\
&= d' \left(1 - \left(\frac{|X|}{|A|}\right) \left(\frac{|Y|}{|B|}\right)\right) \\
&\leq d' (1 - (1 - \delta)^2) \\
&\leq 2\delta - \delta^2 \leq 2\delta
\end{aligned} \tag{14}$$

Since the inequality is proved for maximum and minimum values of  $d(A, B)$ , we can use lemma 1 to conclude

$$|d(A, B) - d(X, Y)| \leq 2\delta \tag{15}$$

And from this it follows that:

$$|d(A, B)^2 - d(X, Y)^2| \leq |d(A, B) - d(X, Y)| \times |d(A, B) + d(X, Y)| \leq 2\delta(1 + 1) = 4\delta \tag{16}$$

□

**Remark:** Theorem 2 essentially tells that large subsets of two graphs carries density information of parent graph

**Definition 1** ( $\epsilon$ -Regularity). We say  $A, B$  are epsilon regular if for  $X \subseteq A$  and  $Y \subseteq B$  with  $|X| \geq \epsilon|A|$  and  $|Y| \geq \epsilon|B|$  we have  $|d(A, B) - d(X, Y)| \leq \epsilon$

**Theorem 3.** Let  $\epsilon \geq 0$  and  $(A, B)$  be a disjoint pair of subsets of vertices. If  $d(A, B) \leq \epsilon^3$ , then show that  $(A, B)$  is an  $\epsilon$ -regular pair.

*Proof.* Let  $X \subseteq A$  and  $Y \subseteq B$  such that  $|X| \geq \epsilon|A|$  and  $|Y| \geq \epsilon|B|$ . Then we note that:

$$d(X, Y) = \frac{e_G(X, Y)}{|X||Y|} = \frac{e_G(X, Y)}{|A||B|} \frac{|A||B|}{|X||Y|} \leq \frac{e_G(A, B)}{|A||B|} \frac{|A||B|}{|X||Y|} = d(A, B) \frac{|A|}{|X|} \frac{|B|}{|Y|} \leq \epsilon^3 \frac{1}{\epsilon^2} \leq \epsilon \tag{17}$$

As  $0 \leq d(A, B), d(X, Y) \leq \epsilon$ ,

$$|d(X, Y) - d(A, B)| \leq \epsilon \tag{18}$$

Therefore,  $A, B$  is an epsilon regular pair. □

**Theorem 4.** Let  $\epsilon' \geq \epsilon$ . If  $(A, B)$  is epsilon regular, then  $(A, B)$  is  $\epsilon'$  regular too.

*Proof.* For any  $X \subseteq A, Y \subseteq B$  such that  $|X| \geq \epsilon'|A| \geq \epsilon|A|$  and  $|Y| \geq \epsilon'|B| \geq \epsilon|B|$  we have:

$$|d(X, Y) - d(A, B)| \leq \epsilon \leq \epsilon' \tag{19}$$

Therefore,  $A, B$  are  $\epsilon'$  regular too. □

**Notation:** For  $v \in G$ , we define  $\tau(v, S) = \{y | y \in S, (v, y) \in e_G G\}$

**Theorem 5.** Let  $(A, B)$  be an  $\epsilon$  regular pair with density  $d$  and  $Y \subset B$  with  $\frac{|Y|}{|B|} \geq \epsilon$ . Then:

$$|\{x \in A \mid |\tau(x, Y)| \leq (d - \epsilon)|Y|\}| \leq \epsilon|A| \quad (20)$$

$$|\{x \in A \mid |\tau(x, Y)| \geq (d + \epsilon)|Y|\}| \leq \epsilon|A| \quad (21)$$

**Proof outline:** We shall exploit the fact that if a large enough subset of vertices has more neighbors than others, then the graph is in a sense more dense for this subset with respect to other parts of the graphs. This disrupts the regularity of the graph. Similarly, if a large subset of vertices has fewer neighbors than others, then the graph is more sparse for this subset.

*Proof.* We prove by contradiction. Assume that the first inequality doesn't hold. Let  $X_1 = \{x \in A \mid |\tau(x, Y)| \leq (d - \epsilon)|Y|\}$ . Then:

$$d(X_1, Y) = \frac{e_G(X_1, Y)}{|X_1||Y|} = \frac{\sum_{x \in X_1} |\tau(x, Y)|}{|X_1||Y|} \leq \frac{\sum_{x \in X_1} (d - \epsilon)|Y|}{|X_1||Y|} = \frac{(d - \epsilon)|X_1||Y|}{|X_1||Y|} = d - \epsilon \quad (22)$$

But then  $|d(A, B) - d(X, Y)| \geq d(A, B) - d(X, Y) \geq \epsilon$ , which contradicts the  $\epsilon$ -regularity of  $(A, B)$ . Therefore,  $|X_1| \leq \epsilon|A|$ .

Similarly, define  $X_2 = \{x \in A \mid |\tau(x, Y)| \geq (d + \epsilon)|Y|\}$ . Then:

$$d(X_2, Y) = \frac{e_G(X_2, Y)}{|X_2||Y|} = \frac{\sum_{x \in X_2} |\tau(x, Y)|}{|X_2||Y|} \geq \frac{\sum_{x \in X_2} (d + \epsilon)|Y|}{|X_2||Y|} = \frac{(d + \epsilon)|X_2||Y|}{|X_2||Y|} = d + \epsilon \quad (23)$$

This results in a contradiction as  $|d(X, Y) - d(A, B)| \geq d(X, Y) - d(A, B) = \epsilon$ . Therefore,  $|X_2| \leq \epsilon|A|$ .  $\square$

**Theorem 6.** More than  $(1 - \epsilon)|A|$  vertices in  $A$  have less than  $(d + \epsilon)|B|$  neighbours in  $B$ . More than  $(1 - \epsilon)|A|$  vertices in  $A$  have more than  $(d - \epsilon)|B|$  neighbours in  $B$ .

*Proof.* Set  $Y = B$  in theorem 5 to get:

$$\begin{aligned} |A| &= |\{x \in A \mid |\tau(x, B)| \geq (d + \epsilon)|B|\}| + |\{x \in A \mid |\tau(x, B)| \leq (d + \epsilon)|B|\}| \\ &\leq \epsilon|A| + |\{x \in A \mid |\tau(x, B)| \leq (d + \epsilon)|B|\}| \\ \Rightarrow (1 - \epsilon)|A| &\leq |\{x \in A \mid |\tau(x, B)| \leq (d + \epsilon)|B|\}| \end{aligned} \quad (24)$$

Similarly,

$$\begin{aligned} |A| &= |\{x \in A \mid |\tau(x, B)| \geq (d + \epsilon)|B|\}| + |\{x \in A \mid |\tau(x, B)| \leq (d + \epsilon)|B|\}| \\ &\leq \epsilon|A| + |\{x \in A \mid |\tau(x, B)| \geq (d + \epsilon)|B|\}| \\ \Rightarrow (1 - \epsilon)|A| &\leq |\{x \in A \mid |\tau(x, B)| \geq (d + \epsilon)|B|\}| \end{aligned} \quad (25)$$

$\square$

**Theorem 7.** Let  $(A, B)$  be an  $\epsilon$  regular pair of  $G$ . Define

$$A' = \{x \in A \mid (d - \epsilon)|B| \leq |\tau(a, B)| \leq (d + \epsilon)|B|\} \quad (26)$$

Then  $|A'| \geq (1 - 2\epsilon)|A|$

*Proof.* Define:

$$A'_1 = \{x \in A \mid |\tau(x, B)| \leq (d - \epsilon)|B|\} \quad (27)$$

$$A'_2 = \{x \in A \mid |\tau(x, B)| \geq (d + \epsilon)|B|\} \quad (28)$$

Set  $Y = B$  in theorem 5 to get  $|A'_1|, |A'_2| \leq \epsilon|A|$ . Note that  $A' = A - (A'_1 \cup A'_2)$ . Therefore,

$$|A'| = |A| - |A'_1 \cup A'_2| \geq |A| - |A'_1| - |A'_2| \geq |A|(1 - 2\epsilon) \quad (29)$$

□

**Theorem 8.** Let  $(A, B)$  be an  $\epsilon$  regular pair with density  $d$  and  $Y \subseteq B$  with  $(d - \epsilon)^{m-1}|Y| \geq |B|$  where  $m \geq 1$  is an integer. Then:

$$\left| \left\{ (x_1, x_2, x_3 \dots x_m) \in A^m : \bigcup_{i=1}^m \tau(x_i, Y) \leq (d - \epsilon)|Y| \right\} \right| \leq m\epsilon|A|^m \quad (30)$$

**Theorem 9.** Let  $(A, B)$  be an  $\epsilon$  regular pair with density  $d$  where  $0 \leq \epsilon \leq d \leq 1$ . If  $d - \epsilon \geq \epsilon$  then:

$$\left| \left\{ (x_1, x_2) \in A^2 : \bigcup_{i=1}^2 \tau(x_i, B) \leq (d - \epsilon)|B| \right\} \right| \leq 2\epsilon|A|^m \quad (31)$$

*Proof.* Set  $Y = B$  and  $m = 2$ . As  $d - \epsilon \geq \epsilon$ ,  $(d - \epsilon)|B| \geq \epsilon|B|$ . Now apply the theorem above to get the result. □

**Theorem 10.** Let  $(A, B)$  be an  $\epsilon$  regular pair with density  $d$  and  $Y \subseteq B$  with  $(d - \epsilon)^{m-1}|Y| \geq |B|$  where  $m \geq 1$  is an integer. Then:

$$\left| \left\{ (x_1, x_2, x_3 \dots x_m) \in A^m : \bigcup_{i=1}^m \tau(x_i, Y) \geq (d + \epsilon)|Y| \right\} \right| \leq m\epsilon|A|^m \quad (32)$$

**Theorem 11.** Let  $(A, B)$  be an  $\epsilon$ -regular pair with density  $d$  in a graph  $G$ . Let  $\overline{G}[A, B]$  be a graph with vertex set  $A \sqcup B$  and  $\{x, y\}$ , where  $x \in A, y \in B$ , forms an edge in  $\overline{G}[A, B]$  if and only if  $\{x, y\}$  is not an edge in  $G$ . Show that if  $G$  is  $\epsilon$ -regular with density  $d$ , then  $\overline{G}[A, B]$  is  $\epsilon$ -regular with density  $(1 - d)$ .

*Proof.* Let  $X \subseteq A$  and  $Y \subseteq B$ . Let  $d$  be the density function in  $G$  and  $\overline{d}$  be the density function in  $\overline{G}$ . Then note:

$$\overline{d}(X, Y) = \frac{e_{\overline{G}}(X, Y)}{|X||Y|} = \frac{|X||Y| - e_G(X, Y)}{|X||Y|} = 1 - \frac{e_G(X, Y)}{|X||Y|} = 1 - d(X, Y) \quad (33)$$

Set  $X = A$  and  $Y = B$  to get the second part of the theorem. Let  $|A'| \geq \epsilon|A|$  and  $|B'| \geq \epsilon|B|$ . Then

$$|\overline{d}(A', B') - \overline{d}(A, B)| = |1 - d(A', B') - 1 + (A, B)| = |d(A, B) - d(A', B')| \leq \epsilon \quad (34)$$

Therefore,  $A, B$  are  $\epsilon$  regular in  $\overline{G}$ . □

**Theorem 12** (Generalised Intersection Property). Let  $(A, B)$  be an  $\epsilon$  regular pair with density  $d$ , where  $0 \leq \epsilon \leq d \leq 1$ . Let  $m \geq 1$  and  $\mu$  be integer with  $0 \leq \mu \leq m$ . If  $Y \subseteq B$  and  $[\min(d, 1 - d - \epsilon)]^{m-1} \geq \epsilon|B|$ , then  $|S_1 \sqcup S_2| \leq 2m\epsilon|A|^m$ , where for  $(x_1, x_2 \dots x_m) \in A^m$ ,

$$Y'(x_1, x_2 \dots x_m) = \left( \bigcap_{i=1}^{\mu} \{y \in Y \mid \{x_i, y\} \in E(G)\} \right) \left( \bigcap_{i=\mu+1}^m \{y \in Y \mid \{x_i, y\} \notin E(G)\} \right) \quad (35)$$

**Theorem 13.** Let  $k, k'$  be two positive integers. If  $(A, B)$  is  $\epsilon$  regular pair with density  $d$  and  $d \gg \epsilon$ ,  $n = |A| = |B|$ , where  $n$  is sufficiently large then,  $G$  contains a copy of  $K_{k,k'}$

**Proof outline:** We shall use the fact that if there is some  $X = \{x_1, x_2, x_3 \dots x_k\}$  and  $Y = \bigcap_{x \in X} \tau(x, B)$  such that  $|Y| \geq k'$  then the graph formed by taking  $X$  and a  $k'$  subset  $Y'$  of  $Y$  as independent vertex sets is a complete bipartite graph embedded in  $G$ .

*Proof.* By theorem 8, we have:

$$\left| \left\{ (x_1, x_2, x_3 \dots x_k) \in A^k : \bigcup_{i=1}^k \tau(x_i, B) \leq (d - \epsilon)|B| \right\} \right| \leq k\epsilon|A|^k \quad (36)$$

$$\Rightarrow \left| \left\{ (x_1, x_2, x_3 \dots x_k) \in A^k : \bigcup_{i=1}^k \tau(x_i, B) \geq (d - \epsilon)|B| \right\} \right| \geq \binom{|A|}{k} - k\epsilon|A|^k \quad (37)$$

Now if  $n$  is sufficiently large, we use Stirling's approximation to get  $\binom{|A|}{k} \sim \frac{n^k}{k!}$ . As  $1 \geq d \gg \epsilon > 0$ ,  $\frac{1}{k!} - k\epsilon > 0$ . Again take  $n$  large enough so that

$$\binom{|A|}{k} - k\epsilon|A|^k \sim \frac{n^k}{k!} - k\epsilon n^k = n^k \left( \frac{1}{k!} - k\epsilon \right) \geq 1 \quad (38)$$

and

$$(d - \epsilon)|B| = (d - \epsilon)n \geq k' \quad (39)$$

holds. As  $\left\{ (x_1, x_2, x_3 \dots x_k) \in A^k : \bigcup_{i=1}^k \tau(x_i, B) \leq (d - \epsilon)|B| \right\}$  is non-empty, choose an element  $X = \{x_1, x_2 \dots x_k\}$  from it. Choose a  $k'$  subset  $Y$  of  $\bigcup_{i=1}^k \tau(x_i, Y)$ . Then the subgraph spanned by  $X \sqcup Y$  contains  $K_{k,k'}$  as subgraph.  $\square$

**Theorem 14** (Slicing Lemma). Let  $A, B$  be  $\epsilon$  regular pair and for some  $\alpha \geq \epsilon$ , let  $\frac{|X|}{|A|}, \frac{|Y|}{|B|} \geq \alpha$ . Then  $X, Y$  is a  $\epsilon'$  regular pair where  $\epsilon' = \max\{\epsilon/\alpha, 2\epsilon\}$ . Moreover,  $|d(X, Y) - d(A, B)| \leq \epsilon$

*Proof.* Let  $\tilde{X}, \tilde{Y}$  be subsets of  $X, Y$  such that the following holds

$$\frac{|\tilde{X}|}{|X|} \geq \epsilon' \quad \frac{|\tilde{Y}|}{|Y|} \geq \epsilon' \quad (40)$$

We note that  $\epsilon' \geq \epsilon/\alpha \Rightarrow \alpha\epsilon' \geq \epsilon$ . Therefore:

$$\frac{|\tilde{X}|}{|A|} = \frac{|\tilde{X}|}{|X|} \frac{|X|}{|A|} \geq \epsilon'\alpha \geq \epsilon \quad \frac{|\tilde{Y}|}{|B|} = \frac{|\tilde{Y}|}{|Y|} \frac{|Y|}{|B|} \geq \epsilon'\alpha \geq \epsilon \quad (41)$$

Now as  $A, B$  is  $\epsilon$  regular and  $\alpha \geq \epsilon$ ,  $|d(X, Y) - d(A, B)| \leq \epsilon$ . Therefore,

$$|d(\tilde{X}, \tilde{Y}) - d(X, Y)| \leq |d(X, Y) - d(A, B)| + |d(\tilde{X}, \tilde{Y}) - d(A, B)| \leq 2\epsilon \leq \epsilon' \quad (42)$$

$\square$

**Theorem 15.** Let  $(A, B)$  be an  $\epsilon$ -regular pair and for some  $\alpha > \epsilon$  and  $\frac{|X|}{|A|}, \frac{|Y|}{|B|} \geq \delta \geq \epsilon$  where  $X \subset A$ . Show that  $(X, B)$  is an  $\epsilon(1 + \frac{1}{\delta})$  regular pair with density at least  $d(A, B) - \epsilon$

*Proof.* Note that  $\epsilon(1 + \frac{1}{\delta}) \geq \epsilon/\delta$ . As  $\delta \leq 1$ ,  $1 + \frac{1}{\delta} < 2$  and  $\epsilon(1 + \frac{1}{\delta}) \geq 2\epsilon$ . Therefore,  $\epsilon(1 + \frac{1}{\delta}) \geq \max\{\epsilon/\delta, 2\epsilon\}$ . Now use theorem 14 and 4 to get the first part of the theorem. The second part follows from regularity condition:

$$|d(A, B) - d(X, B)| \leq \epsilon \Rightarrow d(A, B) - d(X, B) \leq \epsilon \Rightarrow d(X, B) \geq d(A, B) - \epsilon \quad (43)$$

□