Szemerédi Regularity Lemma

For $X, Y \subset V(G)$, $X \cap Y = \emptyset$ define <u>density</u> as

$$d_G(X,Y) = \frac{e_G(X,Y)}{|X||Y|}, \text{ where } e_G(X,Y) = |\{uv \in E(G) : u \in X, v \in Y\}|$$

Given $\varepsilon > 0$, the pair (X,Y) is <u>ε-regular</u> if

$\forall X' \subseteq X, Y' \subseteq Y, |X'| \geq \varepsilon|X|, |Y'| \geq \varepsilon|Y|$ we have

$$|d_G(X',Y') - d_G(X,Y)| \leq \varepsilon.$$

also <u>pseudo-random</u>

A partition $V(G) = V_0 \cup \dots \cup V_k$ is <u>ε-regular</u> if

(i) $|V_0| \leq \varepsilon|G|$, (ii) $|V_1| = \dots = |V_k|$, (iii) $|\{1 \leq i < j \leq k : (V_i, V_j) \text{ is not } \varepsilon\text{-reg.}\}| \leq \varepsilon k^2$

trash or garbage (rubbish) set.

<u>Lemma 4 (Szemerédi, 1978)</u> $\forall \varepsilon > 0\ \forall m \in \mathbb{N}\ \exists M \in \mathbb{N}$:

$\forall G, |G| \geq m\ \exists\ \varepsilon$-reg. partition $(V_0, \dots, V_k)$ for some $m \leq k \leq M$.

<u>Remarks</u>: <u>Condition</u> $k \leq M$ guarantees that V_i, $1 \leq i \leq k$, are <u>large</u>, b/c $|V_i| \geq \frac{(1-\varepsilon)|G|}{M}$

<u>condition</u> $k \geq m$ allows to bound # edges within sets V_i, $1 \leq i \leq k$:

$$\sum_{i=1}^{k} \|G[V_i]\| \leq k \frac{|G|^2}{2k^2} \leq \frac{|G|^2}{2m}.$$

Thus, L says that "most edges of any graph can be partitioned into small number of pseudo-random bipartite subgraphs.

<u>Proof idea</u>: An index $q(\mathcal{P})$ of a partition $\mathcal{P}$ is defined. By def., $q(\mathcal{P}) \leq \frac{1}{2}$. Moreover, if $\mathcal{P}'$ is a sub-partition of $\mathcal{P}$, then $q(\mathcal{P}') \geq q(\mathcal{P})$. Main step is to show that if $\mathcal{P}$ is <u>not</u> ε-reg., then $\exists$ a sub-partition $\mathcal{P}'$ of $\mathcal{P}$ into at most $k4^k$ parts and

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

Thus, after at most $1/\varepsilon^5$ steps we are bound to get an ε-reg. partition since otherwise we would get some $\mathcal{P}'$ with $q(\mathcal{P}') > \frac{1}{2}$ – $\lightning$.

Proof of L4: Let $n=|G|$. For any pair $A, B \subseteq V(G)$, $A\cap B=\emptyset$ define $q(A,B)=\frac{|A||B|}{n^2}(d(A,B))^2$. For a pair of partitions $(\mathcal{A},\mathcal{B})$, $\mathcal{A}$ – a partition of A, $\mathcal{B}$ – a partition of B, let

$$q(\mathcal{A},\mathcal{B})=\sum_{A'\in\mathcal{A},\,B'\in\mathcal{B}} q(A',B').$$

Finally, for a partition $\mathcal{P}=\{C_1,\dots,C_k\}$ of $V(G)$, let

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

Note that $d(A,B)\le 1$, so $q(A,B)\le \frac{|A||B|}{n^2}$.

Consequently, if $|C_i|\le \frac{n}{k}$, then $q(\mathcal{P})\le \binom{k}{2}\frac{1}{k^2}<\frac{1}{2}$.

Prop 7 (i) $q(\mathcal{A},\mathcal{B})\ge q(A,B)$

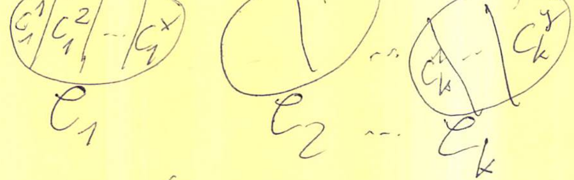
(ii) If $\mathcal{P}'\prec\mathcal{P}$, then $q(\mathcal{P}')\ge q(\mathcal{P})$ (Here $\mathcal{P}'\prec\mathcal{P}$ means: $\mathcal{P}'$ is a sub-part. of $\mathcal{P}$)

Proof (i) $$q(\mathcal{A},\mathcal{B})=\frac{1}{n^2}\sum_{i,j}\frac{e^2(A_i,B_j)}{|A_i||B_j|}\ \underset{\text{by Cauchy-Schwarz}}{\ge}\ \frac{1}{n^2}\frac{\left(\sum_{i,j}e(A_i,B_j)\right)^2}{\sum_{i,j}|A_i||B_j|}=\frac{e^2(A,B)}{n^2|A||B|}=q(A,B)$$ □

(ii) $\mathcal{P}=\{C_1,\dots,C_k\}$, $\mathcal{C}_i=\mathcal{P}'[C_i]$. By (i),

$$q(\mathcal{P})=\sum_{1\le i<j\le k} q(C_i,C_j)\le\sum q(\mathcal{C}_i,\mathcal{C}_j)\le\sum q(\mathcal{C}_i,\mathcal{C}_j)+\sum_i q(\mathcal{C}_i)=q(\mathcal{P}').$$ □

$\mathcal{P}$:



Prop. 8 If a pair (C,D) is not ε-reg. (in G), then $\exists\ \mathcal{C}=\{C_1,C_2\}$ & $\mathcal{D}=\{D_1,D_2\}$, $C=C_1\cup C_2$, $D=D_1\cup D_2$, : $q(\mathcal{C},\mathcal{D})\ge q(C,D)+\frac{\varepsilon^4|C||D|}{n^2}$

Proof As (C,D) is not ε-reg, $\exists\ C_1\subseteq C$, $D_1\subseteq D$, $|C_1|\ge\varepsilon|C|$, $|D_1|\ge\varepsilon|D|$: $|d(C_1,D_1)-d(C,D)|>\varepsilon$. Let $\eta=d(C_1,D_1)-d(C,D)$, $c=|C|$, $d=|D|$, $e=e(C,D)$, $C_2=C\setminus C_1$, $D_2=D\setminus D_1$, $c_i=|C_i|$, $d_i=|D_i|$, $e_{ij}=e(C_i,D_j)$, $i,j=1,2$.

Then $$n^2q(\mathcal{C},\mathcal{D})=\frac{e_{11}^2}{c_1d_1}+\sum_{i+j>2}\frac{e_{ij}^2}{c_id_j}\ge\frac{e_{11}^2}{c_1d_1}+\frac{(e-e_{11})^2}{cd-c_1d_1}$$ by C-Schw.

(Substituting) ~~Setting~~ $e_{11} = \dfrac{c_1 d_1 e}{cd} + \eta c_1 d_1$ ~~& squaring side-wise~~, $\boxed{S2^3}$

The R-H-S becomes:

$$\frac{c_1 d_1 e^2}{c^2 d^2} + \cancel{2\eta \frac{c_1 d_1 e}{cd}} + \eta^2 c_1 d_1 + \frac{e^2(cd - c_1 d_1)}{c^2 d^2} - \cancel{2\eta \frac{c_1 d_1 e}{cd}} + \underbrace{\eta^2 \frac{c_1^2 d_1^2}{(cd - c_1 d_1)^2}}_{} \geq$$

$$\frac{e^2}{cd} + \eta^2 c_1 d_1 > \frac{e^2}{cd} + \varepsilon^4 cd = n^2 q(C, D) + \varepsilon^4 |C||D| \quad \square$$

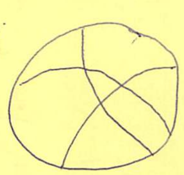
$$\boxed{\text{b/c } q(C,D) = \frac{cd}{n^2}\left(\frac{e}{cd}\right)^2 = \frac{e^2}{n^2 cd}}$$

<u>Prop 9</u> Let $0 < \varepsilon < \frac{1}{4}$ & $\mathcal{P} = \{C_0, \dots, C_k\}$ be an ε-irreg. partition of $V(G)$. Then $\exists\, \ell$, $k \le \ell \le k4^k$ & $\mathcal{P}' = \{C_0', \dots, C_\ell'\}$ – a subpartition of $\mathcal{P}$ s.t. $|C_0'| \le |C_0| + n2^{-k}$, $|C_1'| = \dots = |C_\ell'|$ & $q(\mathcal{P}') \ge q(\mathcal{P}) + \varepsilon^5/2$.

<u>Proof</u> For clarity of presentation, assume that $C_0 = \phi$ and that $k|n$. Set $c = |C_i| = n/k$, $i = 1, \dots, k$. For every ε-irr. pair (C_i, C_j), let $\mathcal{C}_{ij}$ be a partition of C_i & $\mathcal{C}_{ji}$ – of C_j, guaranteed by Prop 8. Then $q(\mathcal{C}_{ij}, \mathcal{C}_{ji}) \ge q(C_i, C_j) + \varepsilon^4 \frac{c^2}{n^2}$.

For other (ε-reg.) pairs we set $\mathcal{C}_{ij} = \{C_i\}$, $\mathcal{C}_{ji} = \{C_j\}$.

Note that for a given i, the partitions $\mathcal{C}_{ij}$, $j \ne i$, "chop" the set C_i into at most 2^{k-1} disjoint subsets.

C_i  Let $\mathcal{C}_i$ be be the partition of C_i obtained that way. In other words, $\mathcal{C}_i$ is a minimal partition of C_i (min. in the # of parts) which is a subpart. of each $\mathcal{C}_{ij}$, $j \ne i$.

Finally, let $\mathcal{C} = \bigcup_{i=1}^{k} \mathcal{C}_i$. Note that $k \le |\mathcal{C}| \le k2^{k-1} < k2^k$.

We'll show that $q(\mathcal{C}) \ge q(\mathcal{P}) + \varepsilon^5/2$ [$\mathcal{C}$ is <u>not</u> a final part. yet]

By Props. and , $q(\mathcal{C}) \ge \sum_{1 \le i < j} q(\mathcal{C}_i, \mathcal{C}_j) \ge \sum q(\mathcal{C}_{ij}, \mathcal{C}_{ji}) \ge$

$\ge q(C_i, C_j) + \varepsilon k^2 \frac{\varepsilon^4 c^2}{n^2} = q(\mathcal{P}) + \varepsilon^5$ [In fact, if $C_0 \ne \phi$, then $c \ge \frac{n - \varepsilon n}{k} \ge \frac{3}{4}\frac{n}{k}$, so we get $\ge q(\mathcal{P}) + \frac{1}{2}\varepsilon^5$]

Now we refine $\mathcal{C}$ by spliting each part into sets of size $\lfloor c4^{-k} \rfloor$ and throwing out the remainders to ~~trash set~~ C_0'.
New (final!) partition $\mathcal{S}'$ has $l \le k4^k$ parts (why?).
The trash set will grow by $\le \lfloor c4^{-k} \rfloor |\mathcal{C}| < c4^{-k} k 2^k \le n2^{-k}$.

$\square$

The final part of the proof of Lemma.

Given $0 < \varepsilon < \frac{1}{4}$, $m \ge 1$, set $s = \varepsilon^{-5}$ for the max. # of iterations of the main step – which is an application of Prop.
Let k be the smallest integer s.t. $k \ge m$ & $2^{k-1} \ge \frac{s}{\varepsilon}$.
Define $M = \max\{f^{(s)}(k), 2k/\varepsilon\}$, where $f(x) = x4^x$
E.g. $f^{(2)}(x) = x4^x 4^{x4^x}$, $f^{(3)}(x) = x4^{x+x4^x + x4^{x+x4^x}}$, ...
Let G be a graph with $|G| = n \ge m$. For $n \le M$, the conclusion is trivial (take singletons as the desired partition).
Assume thus that $n > M$ and split $V(G) = C_0 \cup \ldots \cup C_k$ arbitrarily but such that $|C_1| = \ldots = |C_k|$, $|C_0| < k \le \frac{\varepsilon M}{2} < \frac{\varepsilon n}{2}$.
Suppose this partition is ε-irreg.
Apply Prop. 9 obtaining a new, refined part. with $l < f(k)$ parts and trash set of size $< k + n2^{-k}$. Repeat, if necessary, at most ~~M~~ s times obtaining a final ε-reg partition with $\le M$ parts and with $|C_0'| < k + sn2^{-k} \le \frac{\varepsilon n}{2} + \frac{\varepsilon n}{2} = \varepsilon n$.

$\square$

Cauchy - Schwarz Ineq.

$\forall\, u_1, \dots, u_n,\ v_1, \dots, v_n$

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \leq \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right)$$

<u>Proof</u> $0 \leq (u_1 x + v_1)^2 + \dots + (u_n x + v_n)^2 = (\sum u_i^2)x^2 + 2(\sum u_i v_i)x + \sum v_i^2$

As nonnegative, this quadratic f. must have non-positive discriminant

$$\Delta = 4(\sum u_i v_i)^2 - 4(\sum u_i^2)(\sum v_i^2) \leq 0 \ /:4 \quad \square$$

<u>Corollary (Sedrakyan's Lemma)</u>

$$\frac{(\sum u_i)^2}{\sum v_i} \leq \sum \frac{u_i^2}{v_i}$$

<u>Proof</u> In C-S ineq., substitute

$u_i' = u_i v_i,\ v_i' = v_i^2 \iff u_i = \frac{u_i'}{\sqrt{v_i'}},\ v_i = \sqrt{v_i'}$ to get

$$\left(\sum u_i'\right)^2 \leq \left(\sum \frac{(u_i')^2}{v_i'}\right)\left(\sum v_i'\right)$$

Now, drop "$'$". $\square$

<u>In the proof of L4</u>, we apply the cor. with

$u_{ij} = e^2(A_i, B_j)$ and $v_{ij} = |A_i||B_j|$, $1 \leq i \leq |A|$, $1 \leq j \leq |B|$.

Then also with $n=3$, $u_1 = e_{12}$, $u_2 = e_{21}$, $u_3 = e_{22}$, $v_1 = c_1 d_2$, $v_2 = c_2 d_1$, $v_3 = c_2 d_2$: $\sum_{i=1}^{3} \frac{u_i^2}{v_i} = \frac{e_{12}^2}{c_1 d_2} + \frac{e_{21}^2}{c_2 d_1} + \frac{e_{22}^2}{c_2 d_2} \geq \frac{(\sum u_i)^2}{\sum v_i} = \frac{(e_{12}+e_{21}+e_{22})^2}{c_1 d_2 + c_2 d_1 + c_2 d_2} = \frac{(e-e_{11})^2}{cd - c_1 d_1}$