Lecture 24:

Szemerédi's Regularity lemma (SRL)

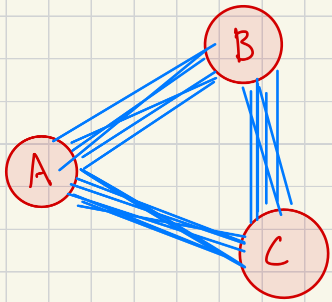
Testing dense graph properties via the SRL:

Δ-freeness

Graphs with "random properties":

example question:

how many triangles in a random tripartite graph?



density η

$\forall u \in A, v \in B, w \in C$:

$$\Pr[u \sim v \sim w] = \eta^3$$

$$\sigma_{u,v,w} = \begin{cases} 1 & \text{if } u \sim v \sim w \\ 0 & \text{o.w} \end{cases}$$

$$E[\sigma_{u,v,w}] = \eta^3$$

$$E[\#\text{triangles}] = E\Big[\sum_{\substack{u \in A \\ v \in B \\ w \in C}} \sigma_{u,v,w}\Big] = \eta^3 |A| \cdot |B| \cdot |C|$$

Can we make weaker assumptions + still get reasonable bounds?

Density + Regularity of set pairs:

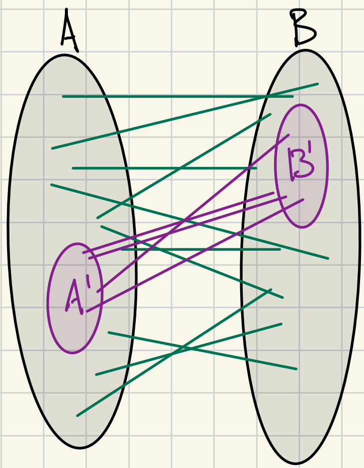
<u>def</u> for $A, B \subseteq V$ s.t.

(1) $A \cap B = \emptyset$

(2) $|A|, |B| > 1$

Let $e(A,B)$ = # edges between $A+B$

+ density $d(A,B) = \dfrac{e(A,B)}{|A|\cdot|B|}$



Say A, B is γ-regular if

$\forall\ A' \subseteq A,\ B' \subseteq B$

s.t. $|A'| \geq \gamma |A|$
$|B'| \geq \gamma |B|$

using same γ in both places not necessary (reduces # of symbols)

$$\underbrace{|d(A',B') - d(A,B)| < \gamma}_{\text{behaves like random graph}}$$

Regularity $\Rightarrow$ lots of Δ's:

Lemma

regularity parameter, depends only on η

$\forall \eta > 0 \quad \exists \gamma = \frac{1}{2}\eta \equiv \gamma^{\Delta}(\eta)$

$\delta = (1-\eta)\frac{\eta^3}{8} \geq \frac{\eta^3}{16} \equiv \delta^{\Delta}(\eta)$

Δ's depends only on η

if $\eta < \frac{1}{2}$

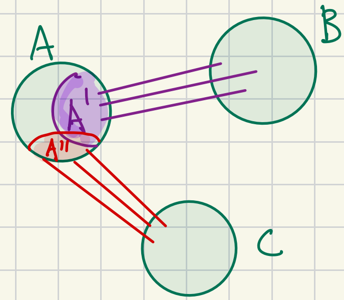
note that this is on the order of what you expect in a random graph!

s.t. if A, B, C disjoint subsets of V s.t. each pair is γ-regular with density $> \eta$ then G contains $\geq \delta \cdot |A| \cdot |B| \cdot |C|$ distinct Δ's with node in each of A, B, C.

Proof

$A^* \leftarrow$ nodes in A with $\geq |\eta - \gamma| \cdot |B|$ nbrs in B

$\geq |\eta - \gamma| \cdot |C|$ nbrs in C



<u>Claim</u> $|A^*| \geq (1-2\gamma)|A|$

<u>Why?</u> (pf of claim)

$A' \leftarrow$ "bad" nodes wrt B $(< |\eta - \gamma| \cdot |B|$ nbrs in $B)$

$A'' \leftarrow$ "bad" nodes wrt C $(< |\eta - \gamma| \cdot |C|$ nbrs in $C)$

then $|A'| \leq \gamma |A|$

$|A''| \leq \gamma |A|$

Why? consider pair A', B

$$d(A', B) < \frac{|A'| \cdot \overbrace{|\eta - \gamma| \cdot |B|}^{\text{def of } A'}}{|A'| \cdot |B|} = \eta - \gamma$$

but $d(A, B) > \eta$

so $|d(A', B) - d(A, B)| > \gamma$

& we know $|B| \geq \gamma |B|$

so if $|A'| \geq \gamma |A|$ then (A, B) <u>not</u> γ-regular

contradiction!

Let $A^* = A \setminus (A' \cup A'')$ then $|A^*| \geq |A| - |A'| - |A''|$

$\geq |A| - 2\gamma |A| = (1-2\gamma)|A|$ $\blacksquare$

$A^* \leftarrow$ nodes in A with $\geq |\eta - \gamma| \cdot |B|$ nbrs in B
$\geq |\eta - \gamma| \cdot |C|$ nbrs in C

Claim $|A^*| \geq (1 - 2\gamma)|A|$

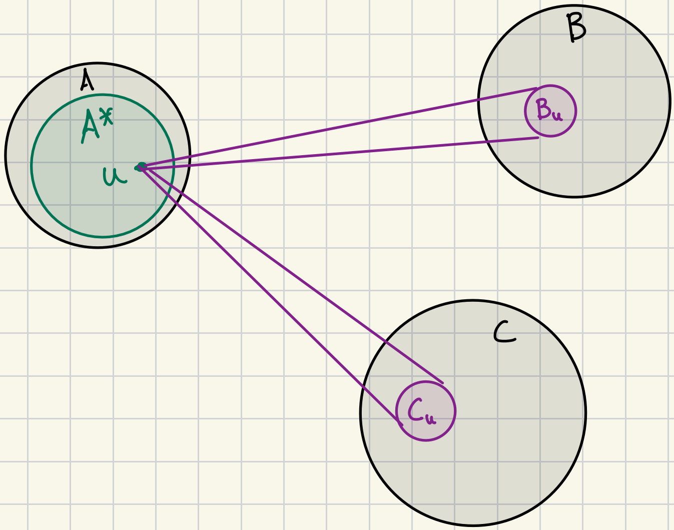
back to proof of lemma ...

For each $u \in A^*$:

Count # Δ's that u participates in

define $B_u \equiv$ nbrs of u in B

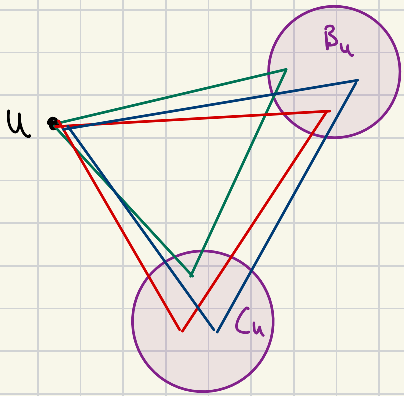
$C_u \equiv$ nbrs of u in C



$B_u + C_u$ can't be too small:

since $\gamma = \frac{\eta}{2}$, $|B_u| \geq (\eta - \gamma)|B| \geq \gamma|B|$

$|C_u| \geq (\eta - \gamma)|C| \geq \gamma|C|$



\# edges between B_u + C_u is lower bound on # Δ's in which u participates

$$d(B,C) \geq \eta \Rightarrow d(B_u, C_u) \geq \eta - \gamma$$

since B_u, C_u big enough & (B,C) is γ-regular

$$\Rightarrow e(B_u, C_u) \geq (\eta - \gamma)|B_u||C_u|$$

def of $d(\cdot)$

$$\Rightarrow (\eta - \gamma)^3 |B||C|$$ triangles in which u participates

so total # Δ's is (sum over all u in A^* to get lower bound)

$$\geq (1-2\gamma)|A| \cdot (\eta - \gamma)^3 |B||C| \geq (1-\eta)(\eta/2)^3 |A||B||C|$$

$\gamma = \eta/2$

$\blacksquare$

Do interesting graphs have regularity properties?

Yes in some sense, all graphs do

"can be approximated as small collection of random graphs"

Szemerédi's Regularity Lemma

would like it to say:

"can always equipartition nodes of V into $V_1 \cdots V_k$

sometimes useful to have lower bound on k to make V_i's small

(for constant k) such that all pairs (V_i, V_j)

independent of n

are ε-regular"

most ($\geq 1-\varepsilon$) is good enough

note $k=1$, $k=n$ trivial

Szemerédi's Regularity Lemma

no dependence on n

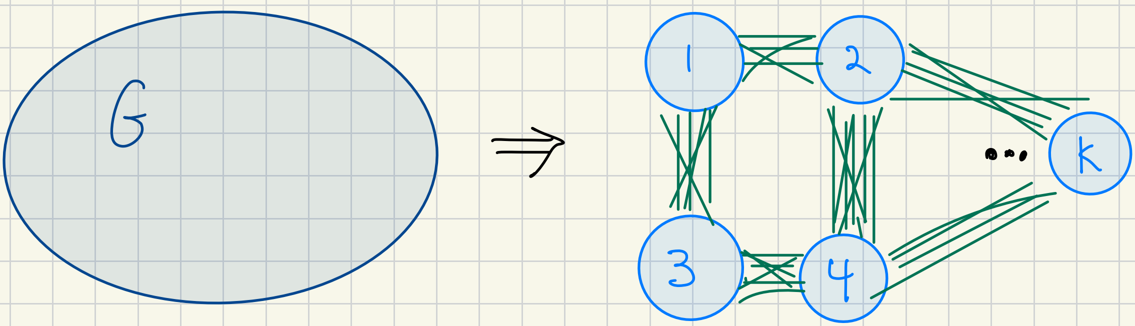
$\forall\, m, \varepsilon > 0 \quad \exists\ T = T(m, \varepsilon)$ s.t. given $G = (V, E)$

s.t. $|V| > T$ + $\mathcal{A}$ an equipartion of V into m sets

then $\exists$ equipartition $\mathcal{B}$ into k sets refining $\mathcal{A}$

s.t. $m \le k \le T$

+ $\le \varepsilon \binom{k}{2}$ set pairs <u>not</u> ε-regular



First studied to prove that sequences of integers have long arithmetic progressions.

Application of SRL to property testing:

Given: G adjacency matrix format

Question: is G Δ-free?

desired behavior: 1-sided error

if G is Δ-free, output PASS

if G $\underbrace{\varepsilon\text{-far}}_{\text{must delete } \varepsilon n^2 \text{ edges}}$ from Δ-free, output FAIL

Algorithm:

Do $O(1/\delta)$ times

pick $v_1, v_2, v_3 \in_r V$

if Δ, reject + halt

Accept

fctn of ε only

Thm $\forall \varepsilon \;\; \exists \delta$ s.t. $\forall G$ s.t. $|V|=n$

 + s.t. G is ε-far from Δ-free

 then G has $\geq \delta \binom{n}{3}$ distinct Δ's

Corr Algorithm has desired behavior

Why?
- if Δ-free: never reject ✓
- if ε-far:

$$\geq \delta \binom{n}{3} \; \Delta\text{'s}$$

$\Rightarrow$ each loop passes with prob $\leq 1-\delta$

so $\Pr[\text{don't see } \Delta \text{ in any loop}]$

$$\leq (1-\delta)^{c/\delta}$$

$$\leq e^{-c} \underset{\uparrow}{<} 1/3$$

choice of c

so reject with prob $\geq 2/3$

$\square$

Thm $\forall \varepsilon\ \exists \delta$ s.t. $\forall G$ s.t. $|V|=n$
+ s.t. G is ε-far from Δ-free
then G has $\geq \delta\binom{n}{3}$ distinct Δ's

Proof of Theorem:

use regularity lemma to get equipartition $\{V_1 \cdots V_k\}$ s.t.

$$\frac{5}{\varepsilon} \leq k \leq T\left(\frac{5}{\varepsilon}, \varepsilon'\right)$$

$\leftarrow$ need $\geq \frac{5}{\varepsilon}$ sets in partition so that no set has $\geq \frac{\varepsilon}{5}$ fraction of nodes

equivalent: $\frac{\varepsilon n}{5} \geq \frac{n}{k} \geq \frac{n}{T(\frac{5}{\varepsilon}, \varepsilon')}$

how? start with arbitrary equipartition $\mathcal{A}$ into $\frac{5}{\varepsilon}$ sets

for $\varepsilon' = \min\left\{\frac{\varepsilon}{5}, \gamma^{\Delta}\left(\frac{\varepsilon}{5}\right)\right\}$ s.t. $\leq \varepsilon'\binom{k}{2}$ pairs not ε'-regular

Assume $\frac{n}{k}$ is integer

"Clean up" G:

define $G' \equiv$ take G and

1) $\forall i$, delete V_i's internal edges

(if $|V_i|$ small, few such edges)

how many? $\le \underbrace{\frac{n}{k}}_{\text{deg w/in } V_i} \cdot \underbrace{n}_{\text{sum over all nodes}} \le \frac{\varepsilon n^2}{5}$

2) delete edges between non regular pairs

how many? $\le \underbrace{\varepsilon'\binom{k}{2}}_{\text{\# nonregular pairs}} \underbrace{\left(\frac{n}{k}\right)^2}_{\text{max \# edges per pair since } |V_i| \approx |V_j| = \frac{n}{k}} \le \frac{\varepsilon}{5} \cdot \frac{\cancel{k^2}}{2} \cdot \frac{n^2}{\cancel{k^2}} = \frac{\varepsilon n^2}{10}$

3) delete edges between $\underbrace{\text{low density}}_{<\varepsilon/5}$ pairs

how many? $\le \sum_{\substack{\text{low} \\ \text{density}}} \left(\frac{\varepsilon}{5}\right) \underbrace{\left(\frac{n}{k}\right)^2}_{\text{upper bounded by total \# edge slots for } k \ge 2} \le \frac{\varepsilon}{5}\binom{n}{2} \approx \frac{\varepsilon}{10} n^2$

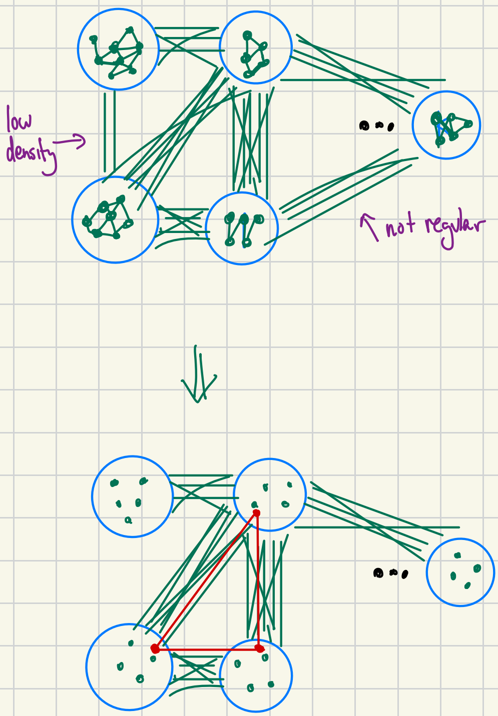
Total deleted edges: $< \varepsilon n^2$

But G is ε-far from Δ-free

(must delete εn^2 edges to make Δ-free)

so G' must still have a Δ!!!

Main point: In this cleaned up graph, one triangle implies many triangles!



Δ in G' must connect:

1) nodes in 3 distinct $V_i V_j V_k$
(erased internal edges)

2) regular pairs
(erased edges in nonregular pairs)

3) high density pairs
(erased edges in low density pairs)

$\therefore \exists\, i, j, k$ distinct s.t. $x \in V_i, y \in V_j, z \in V_k$

$V_i V_j V_k$ all $\geq \frac{\varepsilon}{5} = \eta$ density pairs

& $\geq \gamma^{\Delta}(\varepsilon/5)$-regular $\geq \frac{\eta}{2} = \frac{\varepsilon}{10}$

three partitions which are pairwise regular! (note, each partition has $\sim \varepsilon' n$ nodes)

Δ-counting lemma $\Rightarrow$

$$\geq \delta^{\Delta}\left(\tfrac{\varepsilon}{5}\right) |V_i|\,|V_j|\,|V_k| \quad \Delta\text{'s in } G'$$

where $\delta^{\Delta} = (1-\eta)\frac{\eta^3}{8}$
$\geq \frac{1}{2}\frac{\varepsilon^3}{8000} = \frac{\varepsilon^3}{16000}$

$$\geq \delta' \binom{n}{3} \quad \Delta\text{'s in } G'$$
(and thus G)

for $\delta' = 6\,\delta^{\Delta}\left(\frac{\varepsilon}{5}\right)\left(T\left(\frac{5}{\varepsilon}, \varepsilon'\right)\right)^{-3}$

$\blacksquare$

(–) runtime of property tester is $O(1/\delta) \sim O(\underbrace{T(\tfrac{5}{\varepsilon}, \varepsilon')}_{2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}\,\}\,\log 1/\varepsilon}{}^3)$

(+) Powerful technique!

- similar lemma to Δ-counting for all constant sized subgraphs
- almost "as is" can use same method for all "1st order" graph properties:

$$\exists\, u_1, u_2, u_3 \ldots u_k \;\; \forall\, v_1 \cdots v_\ell \;\; R(u_1 \cdots u_k\, v_1 \cdots v_\ell)$$

nodes

define via $\wedge, \vee, \neg$ + nbr queries to adjacency matrix ie. $\forall u_1 u_2 u_3 \quad \neg(\underbrace{u_1 \sim u_2,\ u_2 \sim u_3,\ u_3 \sim u_1}_{\Delta})$

more generally, H-freeness for const size H

Dense graph testable properties:

- 1-sided error const time $\approx$ hereditary graph properties
  (closed under vertex removal: chordal, perfect, interval)

  difficulty: infinite set of forbidden subgraphs

- 2-sided error const time $\approx$ any property that can be reduced to testing if satisfies one of finite # of Szemeredi partitions

Are there faster testers (in terms of ε) for specific properties?

maybe the reason the dependence on ε is so bad is that the technique is too "general purpose"?

still, Δ-free can't be tested in time poly$(1/\varepsilon)$
(see next lecture)