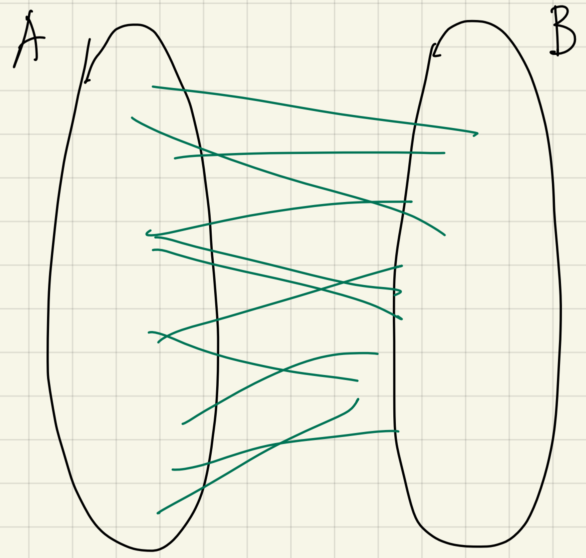
Lecture 10

Testing dense graph properties via SRL:

$\triangle$-freeness

Begin lower bound

Density & Regularity of set pairs:



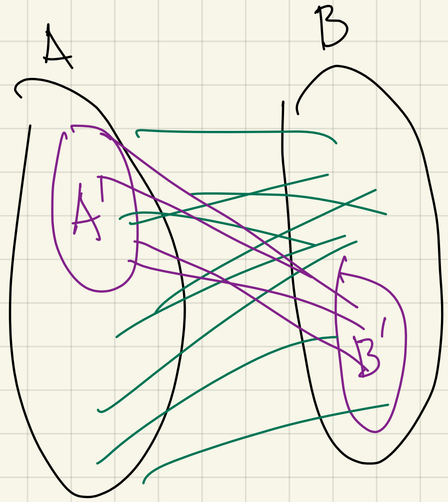
__def.__ For $A, B \subseteq V$ s.t.

(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. $|A'| \geq \gamma |A|$

$|B'| \geq \gamma |B|$

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

behaves like "random graph"

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

A,B is γ-regular if $\forall A' \subseteq A, B' \subseteq B$ s.t. $|A'| \geq \gamma|A|$, $|B'| \geq \gamma|B|$

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

Lemma

density

regularity parameter, depends only on η

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

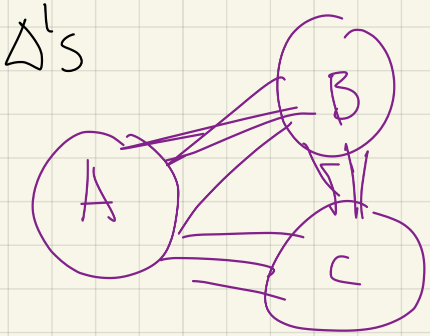
\# triangles depends only on η

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

if $\eta < 1/2$

assume $\eta < 1/2$

s.t. if A, B, C disjoint subsets of V s.t. each pair is γ-regular with density $> \eta$ then G contains $\geq \delta \cdot |A|\cdot|B|\cdot|C|$ $\geq \eta^3/16 \cdot |A||B||C|$ distinct Δ's with node in each of A, B, C.



compare | for random tripartite graphs: $\eta^3 \cdot |A||B||C|$

Do interesting graphs have regularity properties?

Yes in some sense all graphs do

can be approximated as small collection of random regular sets

Szemerédi's Regularity Lemma

would like it to say:

"one can equipartition nodes of V into $V_1 \dots V_k$ (for const k) s.t.

all pairs (V_i, V_j) are ε-regular"

will get only "most"

$\leq \varepsilon \binom{k}{2}$ are not regular

to be useful sometimes need $k \geq m$ for some m

$k=1$ & $k=n$ trivial

Szemerédi's Regularity Lemma: (especially useful version)

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

& $\mathcal{A}$ an equipartition of V into $\overset{m}{\vee}$ sets $\leftarrow$ # is const ind of n $\ll T$

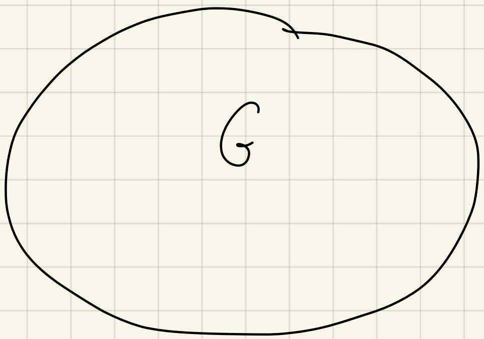
then exists equipartition $\mathcal{B}$ into k sets which refines $\mathcal{A}$

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

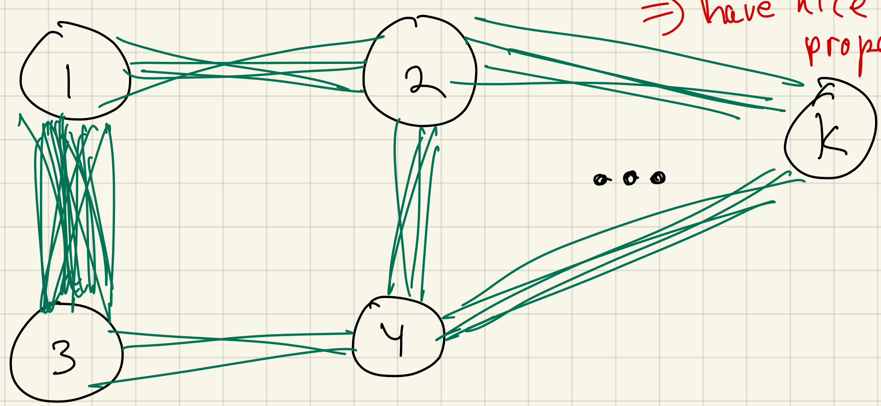
& $\le \varepsilon \binom{k}{2}$ set pairs not ε-regular

const # partitions
each pairs behaves like random graph
$\Rightarrow$ have nice properties

Note: T does not depend on $|V|$



$\Longrightarrow$



An application of the SRL:

Given G in adj matrix form

Is it Δ-free?

desired behavior: if G is Δ-free, output PASS

if G $\underbrace{\varepsilon\text{-far}}_{\text{must delete } \geq \varepsilon n^2 \text{ edges}}$ from Δ-free output FAIL with prob $\geq 3/4$

1-sided error

Just because you need to delete $\geq \varepsilon n^2$ edges, do we know that there are a lot of Δ's???

Algorithm: Do $O(\delta^{-1})$ times:

Pick $v_1, v_2, v_3 \in_r V$

if Δ reject & halt

Accept

fctn of ε only

<u>Thm</u> $\forall \varepsilon$, $\exists \delta$ 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

<u>Corr</u> Algorithm has desired behavior

<u>Why?</u>
- if Δ-free: we never reject ✓
- if ε-far from Δ-free:

$\geq \delta \binom{n}{3}$ Δ's

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

$$\Pr[\text{don't find } \Delta] \leq (1-\delta)^{c/\delta}$$

$$\leq e^{-c} < 1/4$$

for proper choice of c

$\Rightarrow$ reject with prob $\geq 3/4$ ✓

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

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

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

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

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

how? start with arbitrary equipartition into $5/\varepsilon$ sets

for $\varepsilon' \equiv \min\{\frac{\varepsilon}{5}, \gamma^{\Delta}(\frac{\varepsilon}{5})\}$

s.t. $\leq \varepsilon' \binom{k}{2}$ pairs not ε'-regular assume $\frac{n}{K}$ is integer

$G' =$ take G and

1) delete edges internal to any V_i

this is why the V_i's need to be small

(if #nodes per partition small, few internal edges)

$\leq \frac{n}{K}$

how many? $\leq \underbrace{\frac{n}{K}}_{\text{\#edges per node}} \cdot \underbrace{n}_{\text{total \# nodes in graph}} \leq \frac{n^2}{K} \leq \frac{\varepsilon n^2}{5}$

2) delete edges between ε'-non regular pairs

how many?

$\leq \underbrace{\varepsilon' \binom{K}{2}}_{\text{\#irregular pairs}} \left(\frac{n}{K}\right)^2 \leq \frac{\varepsilon}{5} \cdot \frac{\cancel{K^2}}{2} \cdot \frac{n^2}{\cancel{K^2}} \leq \frac{\varepsilon n^2}{10}$

edge slots per irregular pair

$\frac{\varepsilon n}{5} \geq \frac{n}{K} \geq \frac{n}{T(\frac{5}{\varepsilon}, \varepsilon')}$

#nodes in partition

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

A,B is γ-regular if $\forall A' \subseteq A, B' \subseteq B$

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

$|B'| \geq \gamma |B|$

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

$\gamma^{\Delta}(\eta) = \frac{1}{2}\eta$

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

$\varepsilon' = \min\{\frac{\varepsilon}{5}, \gamma^{\Delta}(\frac{\varepsilon}{5})\}$

$\Rightarrow \leq \varepsilon' \binom{K}{2}$ pairs not ε'-regular

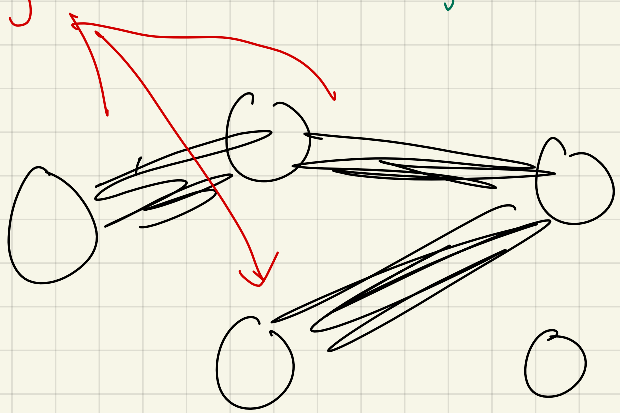
3) delete edges between $\underbrace{\text{low density}}_{<\varepsilon/5}$ pairs

how many? $\leq \underset{\substack{\text{low}\\\text{density}\\\text{pairs}}}{\sum} \left(\frac{\varepsilon}{5}\right)\left(\frac{n}{K}\right)^2$

note $\sum \left(\frac{n}{K}\right)^2 \leq \binom{n}{2}$

$\leq \frac{\varepsilon}{5}\binom{n}{2} \approx \frac{\varepsilon n^2}{10}$

regular & density $\geq \varepsilon/5$



Total deleted edges: $\leq \frac{\varepsilon n^2}{5} + \frac{\varepsilon n^2}{10} + \frac{\varepsilon n^2}{10} < \varepsilon n^2$

But G is ε-far from Δ-free (must delete $\geq \varepsilon n^2$ edges to remove all Δ's)

so G' must still have a triangle!

$$\frac{\varepsilon n}{5} \geq \frac{n}{K} \geq \frac{n}{T\left(\frac{5}{\varepsilon}, \varepsilon'\right)}$$

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

A,B is γ-regular if $\forall A' \subseteq A, B' \subseteq B$
s.t. $|A'| \geq \gamma|A|$
$|B'| \geq \gamma|B|$

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

$\varepsilon' = \min\{\frac{\varepsilon}{5}, \gamma^\delta(\frac{\varepsilon}{5})\}$

& $\leq \varepsilon'\binom{K}{2}$ pairs not ε'-regular

Δ in G' must connect:

1) nodes in 3 distinct $V_i\ V_j\ V_k$

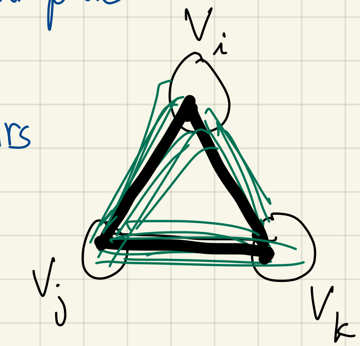
since deleted all internal edges

2) regular pairs

since deleted all edges between irregular pairs

3) high density pairs

since deleted all edges between 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 $\geq \frac{\varepsilon}{5} \overset{=\eta}{}$ density pairs

& $\geq \underbrace{\gamma^{\Delta}\left(\frac{\varepsilon}{5}\right)}_{\geq \frac{\eta}{2} \geq \frac{\varepsilon}{10}}$-regular

Δ-counting lemma $\Rightarrow$

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

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

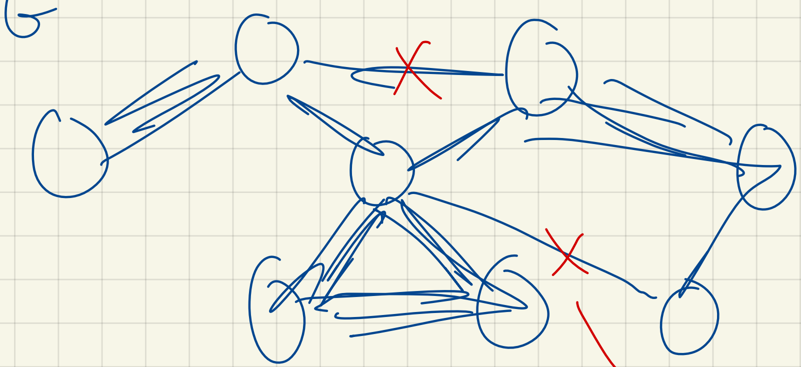
$\geq \Omega(\varepsilon^3)$

$$\geq \frac{\delta^{\Delta}\left(\frac{\varepsilon}{8}\right) n^3}{\left(T(5/\varepsilon, \varepsilon')\right)^3} \quad \Delta's$$

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

for $\delta' = 6\,\delta^{\Delta}\left(\frac{\varepsilon}{5}\right)\left(T\left(\frac{\varepsilon}{8}, \varepsilon'\right)\right)^3$ $\blacksquare$

$G \Rightarrow G'$



This is a powerful technique!

- similar lemma to $\triangle$ counting holds for all const sized subgraphs

- almost "as is" can use same method to test all

  "1st order" graph properties:

$$\exists\, u_1, u_2 u_3 \cdots u_k \quad \forall\, v_1 \cdots v_\ell \quad \underbrace{R(u_1 \cdots u_k\, v_1 \cdots v_\ell)}_{R \text{ defined via } \wedge, \vee, \neg \text{ + neighbors}}$$

nodes

queries to adj matrix

i.e. $\forall\, u_1, u_2, u_3 \quad \neg\,(\underbrace{u_1 \sim u_2,\ u_2 \sim u_3,\ u_3 \sim u_1}_{\text{triangle}})$

more generally,
H-freeness for all const sized H

independent of n

For dense graphs, 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

Maybe the reason that the dependence on ε is so bad is that this technique is too "general purpose"?
Maybe specific properties (e.g. Δ-freeness) have better testers?

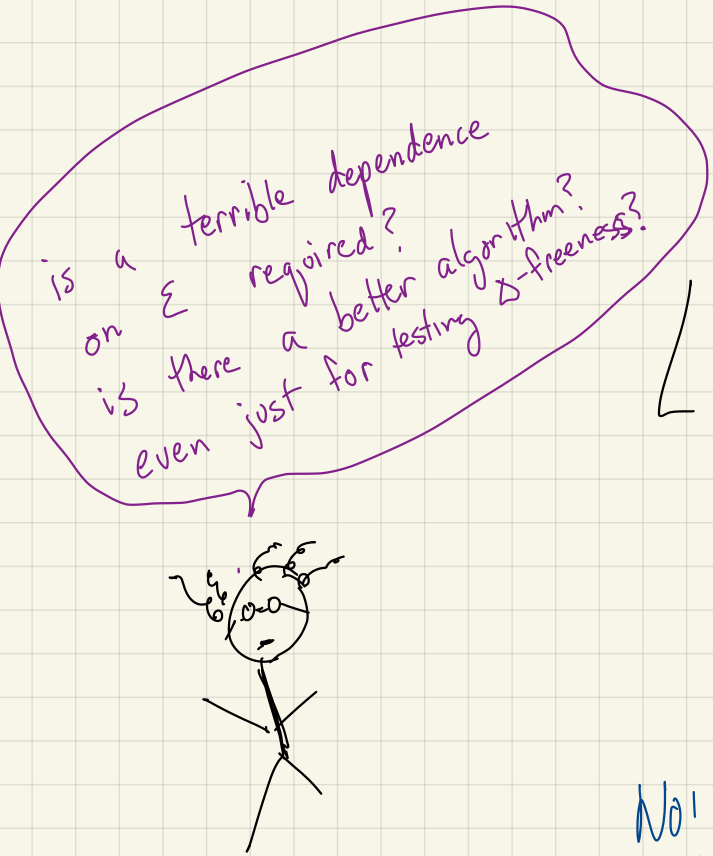
An intriguing characterization of bipartite graphs:

For graphs in adjacency matrix model:

Thm Complexity of testing H-freeness property,
[Noga Alon]

- if H bipartite, $\mathrm{poly}(\frac{1}{\varepsilon})$ is enough
- if H not bipartite, no $\mathrm{poly}(\frac{1}{\varepsilon})$ suffices

we will prove for $H = \triangle$



Lower bounds for testing Δ-freeness:

No! superpoly dependence on ε required!

i.e. $\geq \left(\frac{c}{\varepsilon}\right)^{c \log\left(\frac{c}{\varepsilon}\right)}$ for some const c

Main tool #1: Additive number theory lemma

Lemma $\forall m$, $\exists X \subset M = \{1, 2, \ldots, m\}$ of size $\geq \dfrac{m}{e^{10\sqrt{\lg m}}}$

with no nontrivial soln to

no three evenly spaced points $\longrightarrow$ $$x_1 + x_2 = 2x_3$$

$\overleftrightarrow{\quad 2\ell \quad}$

$\underbrace{x_1 \quad x_3}_{\ell} \underbrace{\quad x_2}_{\ell}$

$$x_3 = \frac{x_1 + x_2}{2}$$

$\Uparrow$

will use to construct graphs which are (1) far from Δ-free

(2) any algorithm needs lots (in terms of ε) queries to find Δ

Lemma $\forall m$, $\exists X < M = \{1, 2, \dots, m\}$ of size $\geq \frac{m}{e^{10\sqrt{\log m}}}$

with no nontrivial soln to

$$X_1 + X_2 = 2X_3$$

examples:

Bad X: $\{1, 2, 3\}$
$\{5, 9, 13\}$

Good X? $\{1, 2, 4, 5, \cancel{6}, \cancel{7}, \cancel{8}, \cancel{9}, 10, \dots\}$ ← how big?

$\{1, 2, 4, 8, 16, 32, \dots\}$ ← only size $\log m$

Lemma $\forall m,\ \exists X < M = \{1,2,\dots,m\}$ of size $\geq \frac{m}{e^{10\sqrt{\lg m}}}$ with no nontrivial soln to $X_1 + X_2 = 2X_3$

Proof Let d be integer

$$k \leftarrow \left\lfloor \frac{\log m}{\log d} \right\rfloor - 1$$

define $X_B = \left\{ \sum_{i=0}^{k} x_i \cdot d^i \;\middle|\; x_i < \frac{d}{2} \text{ for } 0 \leq i \leq k \ \& \ \sum_{i=0}^{k} x_i^2 = B \right\}$

view $x \in M$ in base d representation $x = (x_0, x_1, \dots, x_k)$

Claim $X_B \subseteq M$

Pick B s.t. $|X_B|$ maximized:

how big can B be?

how small can $\sum |X_B|$ be?

So $\exists B$ s.t. $|X_B| \geq$