Testing Δ-freeness in dense graphs: a lower bound

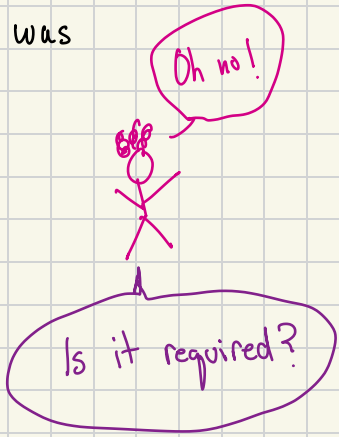
Superpoly dependence on ε is required!

Last time:

Algorithm for testing Δ-freeness with no dependence on n.

but dependence on ε was

$$2^{2^{\cdots^{2}}} \Big\} \log 1/\varepsilon$$



Intriguing characterization of bipartite graphs:

Thm [Alon] In adjacency matrix model, complexity of 1-sided testing H-freeness property is

- if H bipartite: $\text{poly}(1/\varepsilon)$
- if H not bipartite: no $\text{poly}(1/\varepsilon)$ suffices

↑ today will show for Δ's

More specifically:

Thm $\exists$ const c s.t. any 1-sided tester for whether graph G is Δ-free needs

$$\geq \left(\frac{c}{\varepsilon}\right)^{c \log 1/\varepsilon}$$ queries

A first main tool:

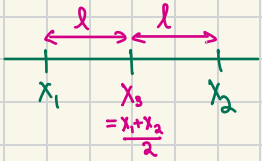
Additive number theory lemma

Lemma $\forall\ m,\ \exists\ X \subset M = \{1, 2, \dots m\}$

of size $\geq \frac{m}{e^{10\sqrt{\log m}}}$ with no nontrivial

solution to $x_1 + x_2 = 2x_3$

no 3 evenly spaced points



will use X to construct graphs which are

(1) far from Δ-free

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

examples:

Bad X: $\{1,2,3\}$

$\{5,9,13\}$

Good X:

how big? → $\{1,2,\cancel{3},4,5,\cancel{6},\cancel{7},\cancel{8},\cancel{9},10,\ldots\}$

only size $O(\log m)$ → $\{1,2,4,8,16,32,\ldots\}$

Proof of lemma:

Let d be integer (will set to $e^{10\sqrt{\log m}}$)

$k \leftarrow \lfloor \frac{\log m}{\log d} \rfloor - 1$ (so $k \approx \frac{\log m}{10\sqrt{\log m}} \approx \frac{\sqrt{\log m}}{10}$)

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

Can be any convex fctn of x_i's

view $x \in M$ in base d representation $x = (x_0 \cdots x_k)$

(1) no carries on sums!

(2) partitions the x's into groups

Why the constraints?

① x_i's $< \frac{d}{2}$ $\Rightarrow$ summing pairs of elements in X_B doesn't generate a carry in any location

(we'll see why useful soon...)

② use (along with ①) to show "sum-free"

Claim $X_B \subseteq M$ (so X_B's partition M)

why? largest int in $X_B \le d^{k+1} \le d^{\lfloor \frac{\log m}{\log d} \rfloor - 1 + 1} \le d^{\log_d m}$

$= m^{\log_d d} = m$ ∎

Pick B s.t. $|X_B|$ maximized: ave size of $X_B = \frac{|\bigcup_B X_B|}{\#B}$

how big can B be? $B \le (k+1)\left(\frac{d}{2}\right)^2 < k \cdot d^2$

using that $\sum x_i^2 = B$

equivalently, how many B's are there?

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

$$\left|\bigcup_B X_B\right| \overset{\text{disjoint}}{=} \sum_B |X_B| \geq \left(\frac{d}{2}\right)^{k+1} > \left(\frac{d}{2}\right)^k$$

so $\exists$ B s.t. $|X_B| \geq \frac{\left(\frac{d}{2}\right)^k}{k \cdot d^2}$

using settings of d, k, get

$$|X_B| \geq \frac{n}{e^{10\sqrt{\log n}}}$$

So if this B is also sum-free, we have the lemma!

(actually all X_B's are sum-free, so we will be good!)

Claim $\forall B$, X_B is sum-free

ie. $\nexists\ x, y, z \in X_B$ s.t. $x + y = 2z$

Proof of Claim:

for $x, y, z \in X_B$:

$$x + y = 2 \cdot z \iff \sum_{i=0}^{k} x_i \cdot d^i + \sum_{i=0}^{k} y_i \cdot d^i = 2 \cdot \sum_{i=0}^{k} z_i \cdot d^i$$

$$\iff \begin{aligned} x_0 + y_0 &= 2 z_0 \\ x_1 + y_1 &= 2 z_1 \\ &\vdots \\ x_k + y_k &= 2 z_k \end{aligned}$$

since no carries

this is where we use constraint ①

nok $\forall i \quad x_i + y_i = 2 z_i$

$\Rightarrow \forall i \quad x_i^2 + y_i^2 \geq 2 z_i^2$ with equality only if $x_i = y_i = z_i$

this is where we use constraint ②

why? $f(a) = a^2$ is convex

using Jensen's ≠:

$\frac{1}{2}\left(f(a_1) + f(a_2)\right) \geq f\left(\frac{a_1+a_2}{2}\right)$ with equality only if $a_1 = a_2 = \frac{a_1+a_2}{2}$

$\Rightarrow \frac{1}{2}\left(x_i^2 + y_i^2\right) \geq \left(\frac{2 z_i}{2}\right)^2$ " " " " $x_i = y_i = 2 z_i$

■ (proof of nok)

So if $x, y, z \in X_B$ s.t. not $(x=y=z)$

then $\exists$ i s.t. not $(x_i = y_i = z_i)$

For this i: $x_i^2 + y_i^2 > 2z_i^2$

For all other j: $x_j^2 + y_j^2 \geq 2z_j^2$

So $\underbrace{\sum_i x_i^2}_{=B} + \underbrace{\sum_i y_i^2}_{=B} > \underbrace{2 \cdot \sum_i z_i^2}_{=2B}$ CONTRADICTION! ■

So we have sum-free X_B s.t. $|X_B| \geq \frac{m}{e^{10\sqrt{\log m}}}$

for 1.6. on Δ-testing, not enough, will need another idea,

(but won't do it here.)

A second main tool:

Goldreich - Trevisan Thm: (home work?)

Adj matrix model
Property P
Tester T with $q(n,\varepsilon)$ queries } possibly adaptive

$\Rightarrow$ Tester T': "Natural tester"

pick $q(n,\varepsilon)$ nodes
query submatrix } $O(q^2)$ queries, nonadaptive
decide

Consequences:

- l.b. for natural tester of $\Omega(q')$ queries
  $\Rightarrow$ l.b. for <u>any</u> tester of $\Omega(\sqrt{q'})$ queries
- reduction preserves 1-sidedness, so l.b. implication does too.

In our case:

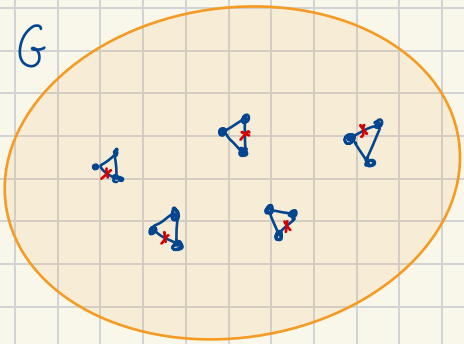
$$\text{prob}[\text{fail}] \approx \frac{\#\Delta\text{'s in } G}{\#\text{ node triples}}$$

we will calculate $\#\Delta$'s for our lower bound family of graphs

need to find a class of graphs on which natural tester doesn't find Δ (in $q \approx (\frac{1}{\varepsilon})^{\log \frac{1}{\varepsilon}}$ queries) & distance is big ($> \varepsilon$)

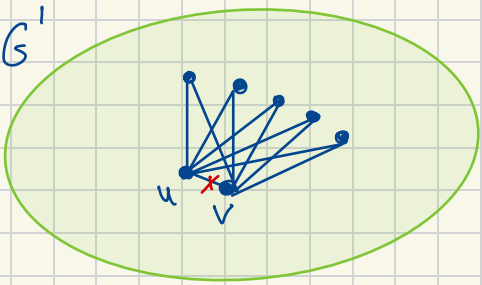
A difficulty:

distance to Δ-free $\not\approx$ # Δ's



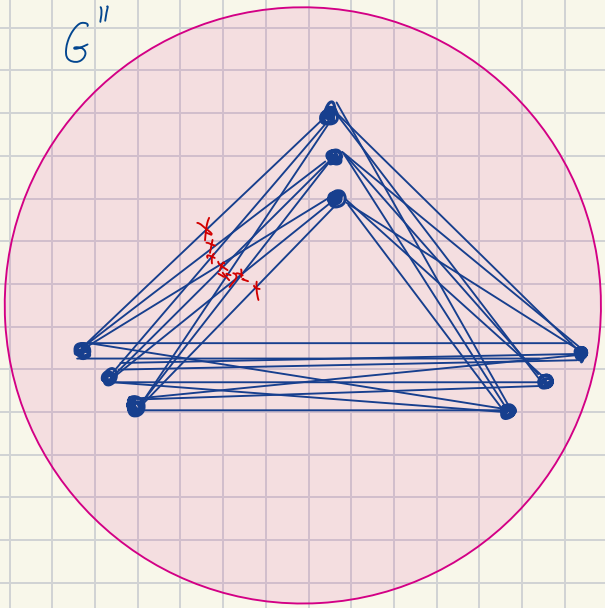
5 Δ's s.t. all Δ's disjoint

distance to Δ-free:
need to delete ≥ 1 edge in each Δ
$\Rightarrow$ dist $\geq$ # disjoint Δ's



5 Δ's s.t. all Δ's share edge (u, v)

dist to Δ-free $= 1$



edges: $\Theta(s^2)$

triangles: $\Theta(s^3)$

disjoint Δ's:

$\Omega(s^2)$

↑ homework?

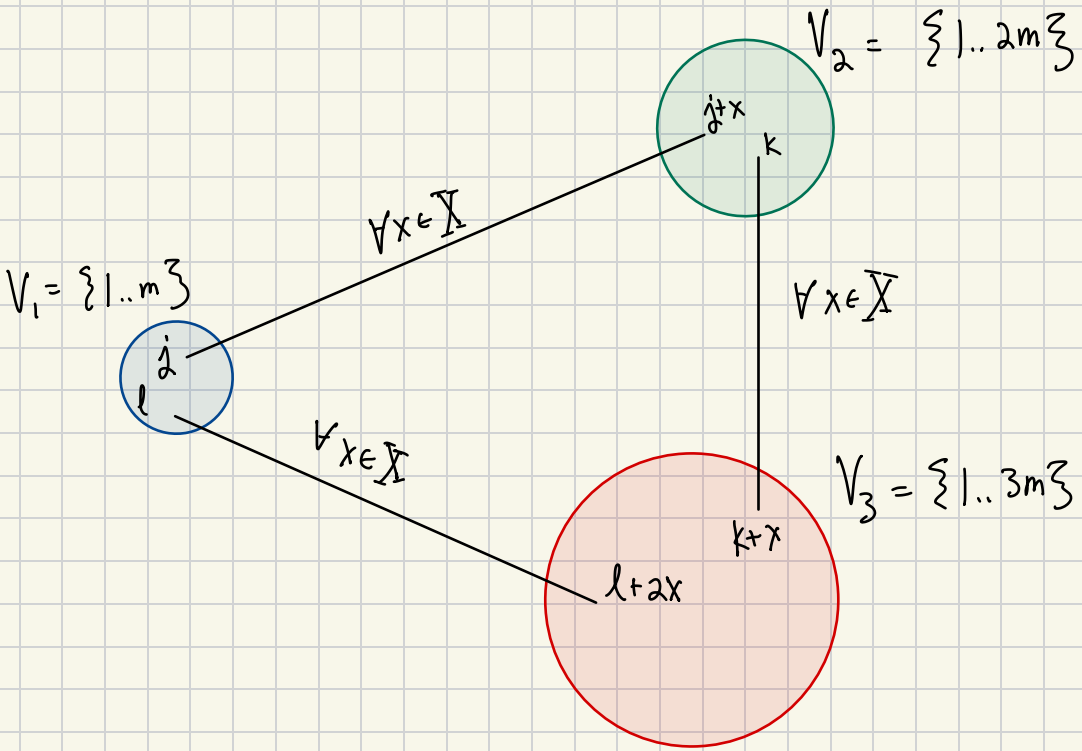
3 groups of vertices of size s connected by bipartite cliques

can remove all Δ's by deleting a bipartite clique

$\Rightarrow$ dist to Δ-free is $O(s^2)$

Graphs on which natural tester needs lots of queries:

given sum-free $X \subseteq \{1..m\}$



- will abuse notation:

  node should be (i,j)

  $i \in \{1,2,3\}$ $j \in \{1..m\}$ $\{1..2m\}$ $\{1..3m\}$

  will drop i if easy to see from context

- # nodes = $6m$
- # edges = $\Theta(m \cdot |X|) = \Theta(n^2 / e^{10\sqrt{\ln n}})$ ← not exactly dense

- # Δ's: [important for determining detection probability of algorithm]

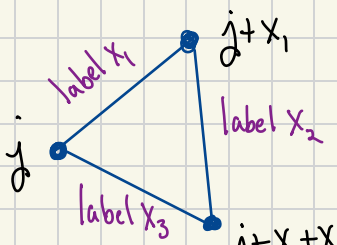
"intended": of form $j, j+x, j+2x$

$$m \cdot |X| = \Theta\left(n^2 / e^{10\sqrt{\ln n}}\right)$$

"unintended":

$V_1 V_2 V_3$ have no internal edges $\Rightarrow$

any Δ has $v_1 \in V_1$ $v_2 \in V_2$ $v_3 \in V_3$



$\Rightarrow x_1 + x_2 = 2x_3$

$\Rightarrow x_1 = x_2 = x_3$ since X is sum-free

but these are intended!

Conclusion:

no unintended Δ's

$\Pr[\text{pick } \Delta] \approx \left(n^2 / e^{10\sqrt{\ln n}}\right) / \binom{n}{3}$ ← very small!!

- # disjoint Δ's

important for determining distance to Δ-free:

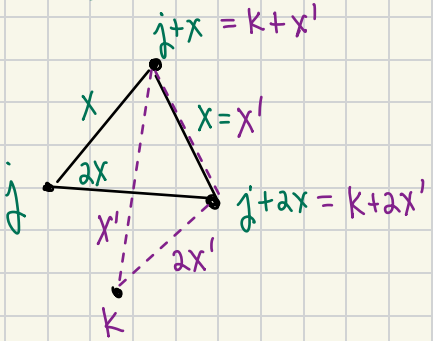
given set of disjoint Δ's, must remove ≥ 1 edge in each

$\Downarrow$

Claim "absolute" distance from Δ-free $= \Omega$(# disjoint Δ's)

Claim: All intended Δ's are disjoint (share no edges at all!)

suppose not:



(argument for why they can't share 2nd edge, similar argument holds for other edges)

since $x = x'$, $j = k$ $\rightarrow\leftarrow$

so distance to Δ-free $= \Theta(\#\Delta\text{'s}) = \Theta\left(\frac{n^2}{e^{10\sqrt{\log n}}}\right) = \Theta(m|X|)$

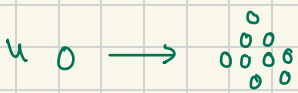
Problem need ε distance,

we only have $\theta\left(\frac{1}{e^{10\sqrt{\log n}}}\right)$ distance

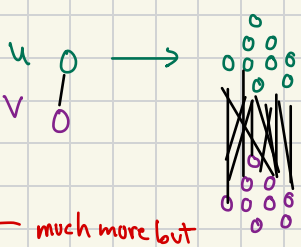
Idea for fix: "s-blow up"

$$G \Rightarrow G^{(s)}$$

$\sim m \cdot s$ node u in G $\Rightarrow$ size s independent set in $G^{(s)}$



$\sim m \cdot |X| \cdot s^2$ edge (u,v) in G $\Rightarrow$ complete bipartite graph in $G^{(s)}$



$\sim m \cdot |X| \cdot s^3$ Δ in G $\Rightarrow$ s^3 Δ's in $G^{(s)}$ ← much more but no longer disjoint

how many edge disjoint Δ's?

Lemma absolute dist of $G^{(s)}$ from Δ-free $\geq$ # edge disjoint Δ's $\geq m \cdot |X| \cdot s^2$

Proof show Δ in $G \Rightarrow s^2$ disjoint Δ's in $G^{(s)}$ ∎

use $s \approx \frac{n}{6m} \approx n \cdot \left(\frac{\varepsilon}{c}\right)^{c \log \varepsilon/c}$

pick m largest int $\varepsilon \le \frac{1}{e^{10\sqrt{\log m}}}$ so $m \ge \left(\frac{c}{\varepsilon}\right)^{c \log 1/\varepsilon}$

So <u>relative</u> distance of $G^{(s)} \approx \frac{m|X|s^2}{(ms)^2}$ ← size of adj matrix

(recall $|X| = \frac{m}{e^{10\sqrt{\log m}}}$) $= \frac{|X|}{m} \ge \frac{1}{e^{10\sqrt{\log m}}} \ge \varepsilon$

↑ choice of m

$\#\Delta$'s $\approx m \cdot |X| \cdot s^3$

$\approx \left(\frac{\varepsilon}{c'}\right)^{c' \log(c'/\varepsilon)} \cdot n^3$

If take sample of size q + run natural algorithm

$E[\#\Delta\text{'s in sample}] < \binom{q}{3} \left(\frac{\varepsilon}{c'}\right)^{c' \log(c'/\varepsilon)}$

$<< 1$ unless $q > \left(\frac{c''}{\varepsilon}\right)^{c'' \log(c''/\varepsilon)}$

so by Markov's $\neq$, $\Pr[\Delta \text{ in sample}] << 1$

since 1-sided error, most see Δ in sample to fail

