Lecture 11

Testing dense graph properties via SRL:

Δ-freeness

the lower bound ...

An intriguing Characterization of bipartite graphs:

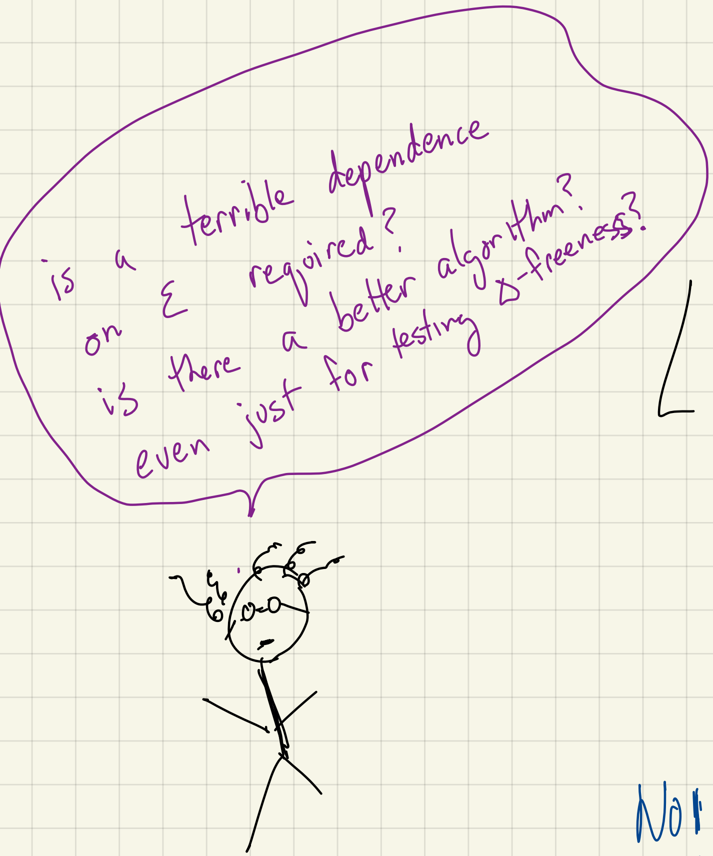
For graphs in adjacency matrix model:

<u>Thm</u> Complexity of testing H-freeness property,

[Noga Alon]

- if H bipartite, $\text{poly}(\frac{1}{\varepsilon})$ is enough
- if H not bipartite, no $\text{poly}(\frac{1}{\varepsilon})$ 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 $\forall m$, $\exists X \subset M = \{1, 2, \dots, m\}$ of size $\geq \dfrac{m}{e^{10\sqrt{\lg m}}}$

with no nontrivial soln to

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

$\overbrace{\underbrace{x_1 \quad x_3}_{\ell} \underbrace{\quad x_2}_{\ell}}^{2\ell}$ $\qquad x_3 = \dfrac{x_1 + x_2}{2}$

$\Uparrow$

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

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

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

$$K \leftarrow \left\lfloor \frac{\log m}{\log d} \right\rfloor - 1 \quad \left(\text{so } K \approx \frac{\log m}{10\sqrt{\log m}} \approx \frac{\sqrt{\log m}}{10}\right)$$

can be convex any fctn of x_i's

define $$X_B = \Big\{ \underbrace{\sum_{i=0}^{K} x_i \cdot d^i}_{\text{view } x \in M \text{ in base } d \text{ representation } x = (x_0 \cdots x_K)} \;\Big|\; \underbrace{x_i < \frac{d}{2} \text{ for } 0 \leq i \leq k}_{1} \;\&\; \underbrace{\sum_{i=0}^{k} x_i^2 = B}_{2} \Big\}$$ $\forall B > 0$

(2) partitions $M \cap \{x \mid x \text{ satisfies } (1)\}$

Claim $X_B \subseteq M$ why? largest number in $X_B \leq d^{K+1} \leq d^{\lfloor \frac{\log m}{\log d} \rfloor - 1 + 1} \leq d^{\log_d m} = m^{\log_d d} = m$

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

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

how small can $\sum |X_B|$ be? $\left|\bigcup_B X_B\right| = \sum_B |X_B| \geq \left(\frac{d}{2}\right)^{K+1} > \left(\frac{d}{2}\right)^K$

So $\exists B$ s.t. $|X_B| \geq \frac{\left(\frac{d}{2}\right)^K}{K d^2}$, so $\exists B$ s.t. $|X_B| \geq \frac{m}{e^{10\sqrt{\log m}}}$

Then if B is sum-free, we have the lemma!

Why the constraints?

- x_i's $< \frac{d}{2} \Rightarrow$ sum pairs of elts in X_B doesn't generate carries
- will use both to show X_B is sum-free

Lemma $\forall m, \exists X \subset M = \{1, 2, \dots, m\}$ of size $\geq \frac{m}{e^{10\sqrt{\log m}}}$ with no nontrivial soln to $x_1 + x_2 = 2x_3$

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

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

$X_B = \{ \sum_{i=0}^{k} x_i \cdot d^i \mid x_i < \frac{d}{2}$ for $0 \leq i \leq k$ $+ \sum_{i=0}^{k} x_i^2 = B \}$

① ②

Proof that X_B is sum-free:

for $x, y, z \in X_B$: $x + y = 2 \cdot z \Leftrightarrow \sum_{i=0}^{k} x_i \cdot d^i + \sum_{i=0}^{k} y_i \cdot d^i = 2 \cdot \sum_{i=0}^{k} z_i \cdot d^i$

$$\Leftrightarrow \left.\begin{aligned} x_0 + y_0 &= 2 \cdot z_0 \\ x_1 + y_1 &= 2 \cdot z_1 \\ &\vdots \\ x_k + y_k &= 2 \cdot z_k \end{aligned}\right\} \text{since no carries}$$

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

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

using Jensen's ≠: $\frac{1}{2}(f(a_1) + f(a_2)) \geq f\left(\frac{a_1 + a_2}{2}\right)$ with equality only if $a_1 = a_2 = \frac{a_1 + a_2}{2}$

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

Why the constraints?

Lemma: $\forall m$, $\exists X \subset M = \{1, 2, \dots, m\}$ of size $\geq \frac{m}{e^{10\sqrt{\log m}}}$ with no nontrivial soln to $x_1 + x_2 = 2x_3$

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

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

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

1 — $x_i < \frac{d}{2}$; 2 — $\sum_{i=0}^{k} x_i^2 = B$

- x_i's $< \frac{d}{2}$ $\Rightarrow$ sum pairs of elts in X_B doesn't generate any carries!
- will use both to show X_B is sum-free

Proof that X_B is sum-free:

for $x, y, z \in X_B$: $\quad x + y = 2z \iff \sum_{i=0}^{k} x_i d^i + \sum_{i=0}^{k} y_i d^i = 2 \cdot \sum_{i=0}^{k} z_i d^i$

$$\iff \left.\begin{aligned} x_0 + y_0 &= 2z_0 \\ x_1 + y_1 &= 2z_1 \\ &\vdots \\ x_k + y_k &= 2z_k \end{aligned}\right\} \text{since no carries}$$

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

so if $\exists \; x, y, z \in X_B$ s.t. $x + y = 2z$ & $\text{not}(x = y = z)$ then $\underbrace{\exists i \text{ s.t. } \text{not}(x_i = y_i = z_i)}$

for this i $\quad x_i^2 + y_i^2 > 2z_i^2$

for all other j $\quad x_j^2 + y_j^2 \geq 2z_j^2$

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

So we have this 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 soln to

$$X_1 + X_2 = 2X_3$$

How do we use it?

- Characterize form of "nearly best" property testers
- Use lemma to build class of graphs which make property testers do the wrong thing

Main tool #2: Characterization of "best" algorithms for property testing

Homework 2

G in adj matrix model
Property P
Tester T using $q(n,\varepsilon)$ queries

$\Rightarrow \exists$ tester T'; "Natural tester"
pick $q(n,\varepsilon)$ nodes randomly
query submatrix
decide

$\underbrace{}_{\Theta(q^2) \text{ queries}}$

In our case:
Prob pass $\approx \dfrac{\#\ \Delta\text{'s in } G}{\#\ \text{node triples}}$

We will calcclate $\#\Delta$'s for our lower bound family of graphs

Consequences:

l.b. for natural tester of $\Omega(q)$

$\Rightarrow$ l.b. for any tester of $\Omega(\sqrt{q})$

* in q queries where $q \approx \left(\frac{1}{\varepsilon}\right)^{\log \frac{1}{\varepsilon}}$

reduction preserves 1-sidedness
so l.b. implication preserved too

Only need to find class of graphs on which natural tester can't find a Δ + distance is big ($\geq \varepsilon$)

"Natural tester"

pick $q(n,\varepsilon)$ nodes ran[dom]

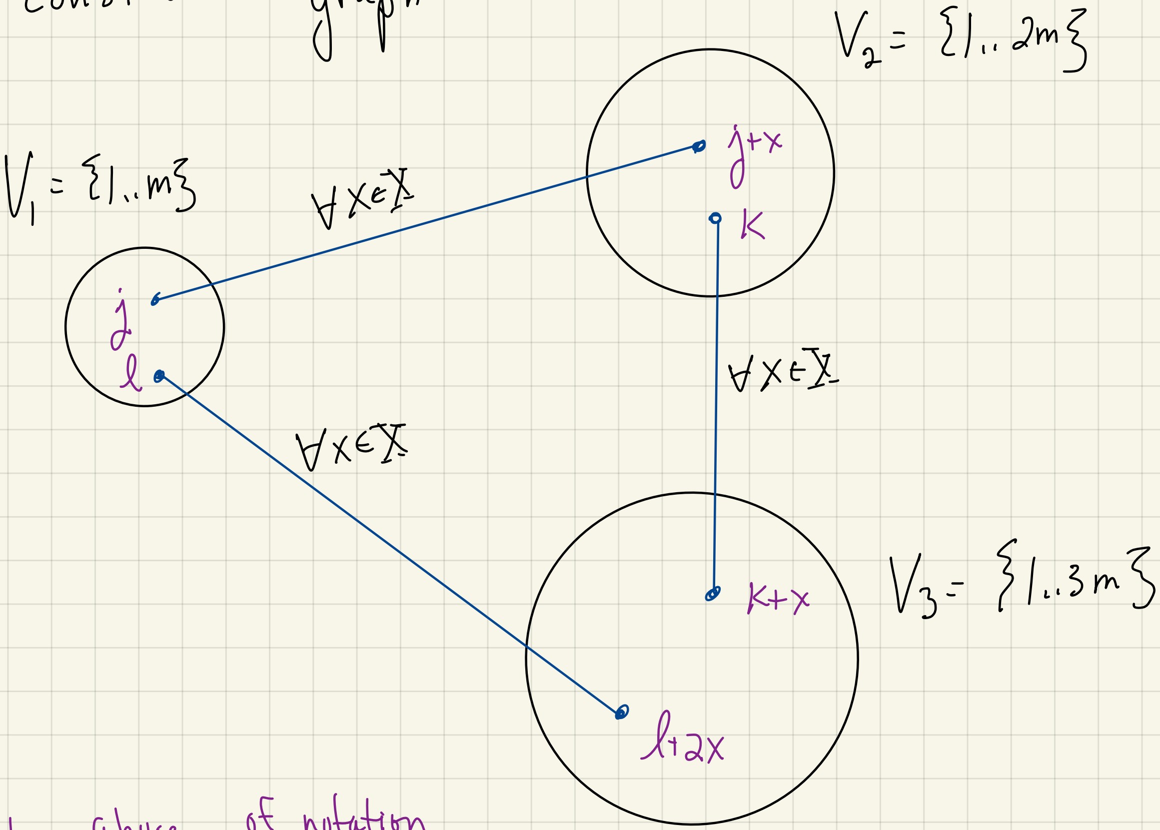
query submatrix

decide

Graphs on which natural tester needs lots of queries:

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

construct graph



$\#$ nodes $= 6m$ so $m = \Theta(n)$

$\#$ edges $= \Theta(m \cdot |X|)$

$= \Theta(n^2 / e^{10\sqrt{\log n}})$

not dense enough

need dist to be $\varepsilon \cdot n^2$

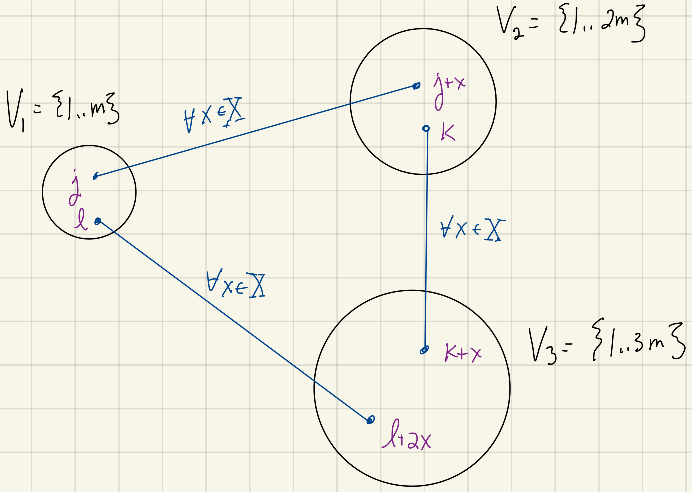
slight abuse of notation

should be (i, j)

$i \in \{1,2,3\}$

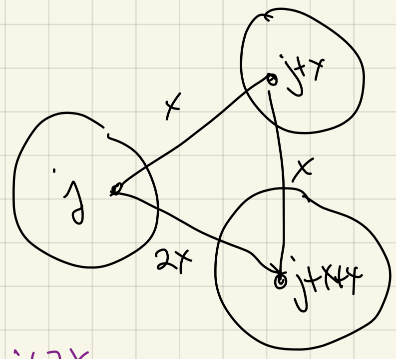
$j \in \{1,..m\}$, $\{1..2m\}$, $\{1..3m\}$

but well assume i from context



Lemma $\forall m, \exists X \subset M = \{1, 2, \ldots, m\}$ of size $\geq \frac{m}{e^{10\sqrt{\log m}}}$ with no nontrivial soln to

$$X_1 + X_2 = 2X_3$$



"Natural tester"

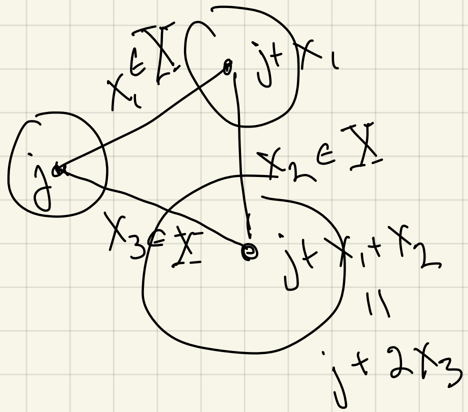
pick $f(n, \varepsilon)$ nodes randomly
query submatrix
decide

cycles: intended Δ's of form $j, j+x, j+2X$

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

unintended Δ's

V_1, V_2, V_3 have no internal edges $\Rightarrow$ any Δ has $v_1 \in V_1$, $v_2 \in V_2$ & $v_3 \in V_3$



$x_1 + x_2 = 2x_3$ but $x_1, x_2, x_3 \in X$

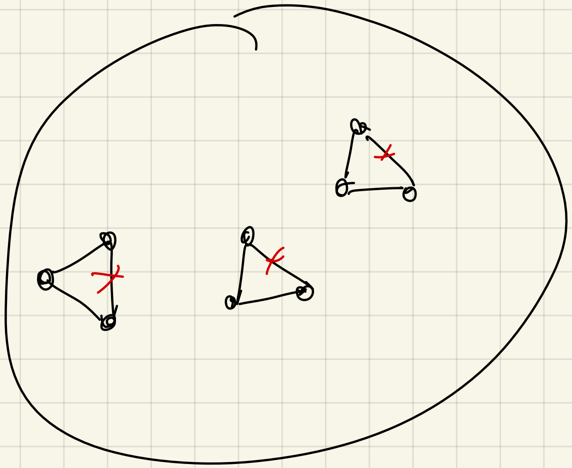
$\Rightarrow x_1 = x_2 = x_3$

since X has no nontrivial solns to $x_1 + x_2 = x_3$

$\Bigg\}$ no unintended Δ's

distance to Δ-free $\neq$ # Δ's

G

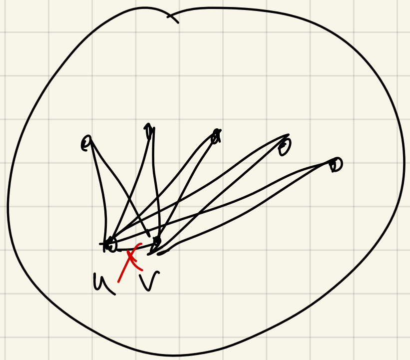


all Δ's disjoint

distance to Δ-free:
need to delete edge in each Δ

$\Rightarrow$ dist $\geq$ # disjoint Δ's

G'



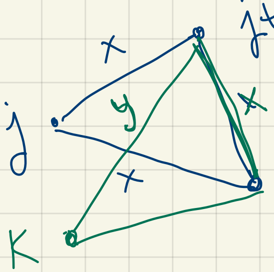
all Δ's share edge (u,v)

dist to Δ-free: 1 edge

disjoint cycles:

all intended Δ's are disjoint (share no edges)

suppose not:



$j+x = K+y$

$x = y$

$j+2x = K+2y$

since $x = y$

$\Downarrow$

$j = K$

$\Downarrow$

$(j, j+x, j+2x) = (K, K+y, K+2y)$

this shows can't share 2^{nd} edge
repeat argument for 1^{st} & 3^{rd} edges

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

$X_1 + X_2 = 2X_3$

"Natural tester"

pick $q(n, \varepsilon)$ nodes randomly
query submatrix
decide

distance to Δ-free:

Δ's edge disjoint $\Rightarrow$ must remove ≥ 1 edge from each disjoint Δ

$\Downarrow$

must remove $\theta(\#\ \Delta's) = \theta\left(\frac{n^2}{e^{10\sqrt{\log n}}}\right)$

$= \theta(m \cdot |X|)$

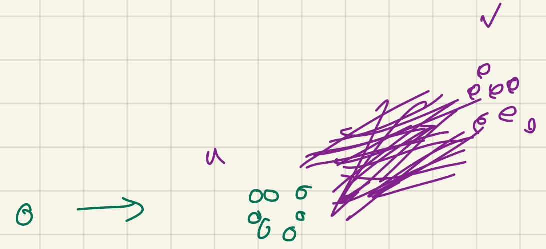
Problem need ε distance

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

Idea for fix

"s-blow up"

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

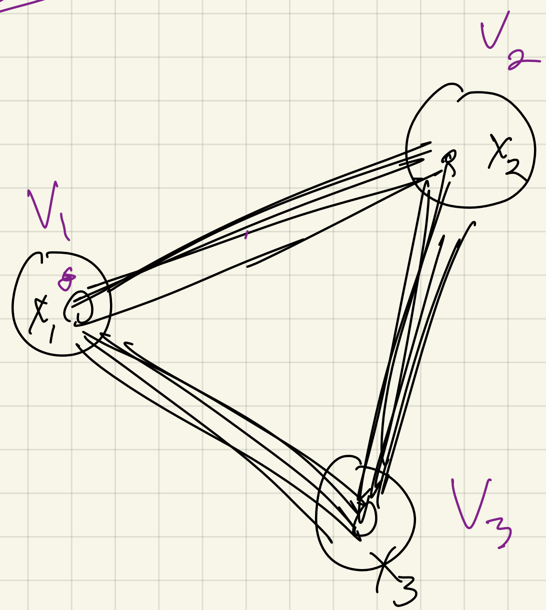


node in G $\Longrightarrow$ size s independent set in $G^{(s)}$

$\sim m \cdot |s|$

edge in G $\Longrightarrow$ complete bipartite graph in $G^{(s)}$

$\sim m \cdot |X| \cdot |s|^2$



Δ in G $\Longrightarrow$ s^3 Δ's in $G^{(s)}$

$\sim m \cdot |X| \cdot |S|^3$

↑ prove a lemma

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

pick m largest int $\varepsilon \le e^{-\frac{1}{10}\sqrt{\log m}}$

so $m \ge \left(\frac{c}{\varepsilon}\right)^{c\log c/\varepsilon}$

lemma: dist of $G^{(s)}$ from Δ free $\ge$ # edge disjoint Δ's

$\ge m|X| \cdot s^2$

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

so distance $\approx \dfrac{m\,|X|\cdot s^2}{m^2 \cdot s^2}$ $\leftarrow$ size adj matrix

$$= \frac{|X|}{m} \geq \frac{1}{e^{10\sqrt{\log m}}} \geq \varepsilon \quad \text{by choice of } m$$

$$\#\Delta\text{'s} \approx m \cdot |X| \cdot |S|^3$$

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

If take sample of size q (& run natural algorithm)

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

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

q is cube root of

Then by Markov's $\neq \Rightarrow \Pr[\text{any } \Delta \text{ in sample}] \ll 1$

since 1-sided error, must see Δ to output fail ∎