Lecture 11

Testing dense graph properties via SRL:

Δ- freeness

the lower bound ...

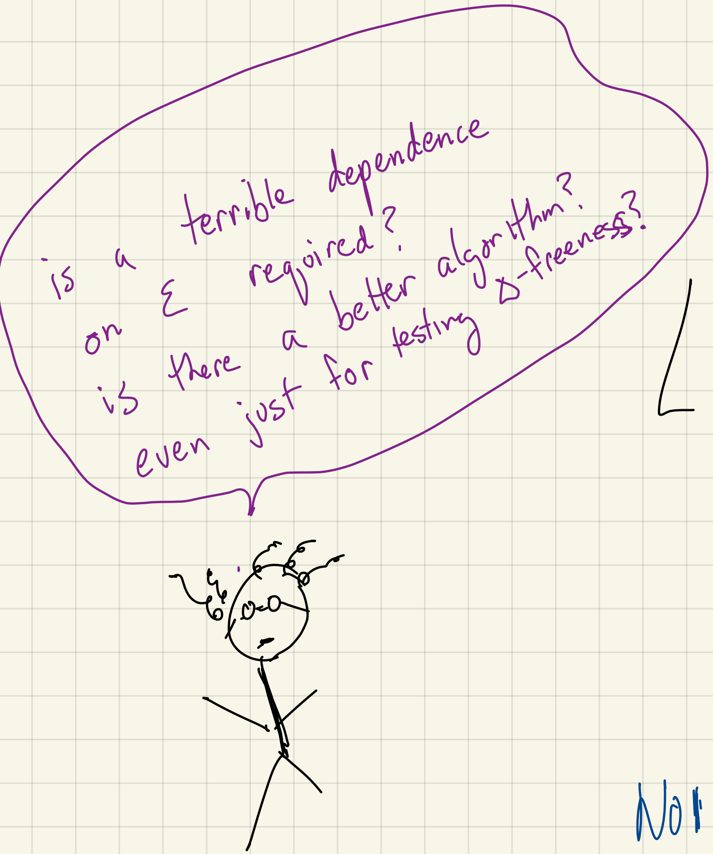
An intriguing Characterization of bipartite graphs:

For graphs in adjacency matrix model:

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

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

↑ we will prove for $H = \triangle$



Lower bounds for testing Δ-freeness:

No! superpoly dependence on ε required!

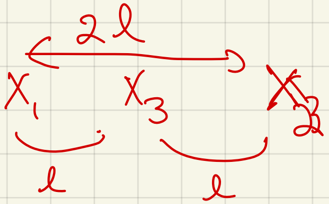
i.e. $\geq \left(\frac{c}{\varepsilon}\right)^{c \log\left(\frac{c}{\varepsilon}\right)}$ for some const c

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 ⟶ $$x_1 + x_2 = 2x_3$$



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

↑

will use to construct graphs which are (1) far from $\triangle$-free

(2) any algorithm needs lots (in terms of ε) queries to find $\triangle$

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

$\{1, 2, 4, 8, 16, 32, \dots\} \leftarrow$ 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. $|X_B| \geq$

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 $e^{10\sqrt{\log m}}$)

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

(so $K \approx \frac{\log m}{10\sqrt{\log m}} \approx \frac{\sqrt{\log m}}{10}$)

define

$$X_B = \Big\{ \underbrace{\sum_{i=0}^{K} x_i \cdot d^i}_{\substack{\text{view } x\in M \text{ in base } d\\ \text{representation } x=(x_0,x_1,\dots x_K)}} \;\Big|\; \underbrace{x_i < \frac{d}{2} \text{ for } 0 \leq i \leq k}_{①} \;+\; \underbrace{\sum_{i=0}^{k} x_i^2}_{②} = B \Big\}$$

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? $B \leq (K+1)\left(\frac{d}{2}\right)^2 < K \cdot d^2$ ← bound on x_i's

how small can $\sum |X_B|$ be ? $\left|\bigcup_B X_B\right| \geq \left(\frac{d}{2}\right)^{K+1} > \left(\frac{d}{2}\right)^K$ but X_B's are disjoint so this lower bnd sum

So $\exists$ B s.t. $|X_B| \geq \frac{\left(\frac{d}{2}\right)^K}{Kd^2}$, using settings get $\exists B$ 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 $< \frac{d}{2} \Rightarrow$

- will use both to show X_B is sum-free

Proof that X_B is sum-free:

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

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 $e^{10\sqrt{\log m}}$)

$k \leftarrow \left\lfloor \frac{\log m}{\log d} \right\rfloor - 1$ (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 ②

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 $X_1 + X_2 = 2X_3$

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

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

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

① ②

Why the constraints?

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

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

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 $x_i = y_i = z_i$

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

so use Jensens $\neq$: $\frac{1}{2}(f(a_1) + f(a_2)) \geq f\left(\frac{a_1 + a_2}{2}\right)$ with equality only if all a_n's are $=$

$$\Rightarrow \frac{x_i^2 + y_i^2}{2} \geq \left(\frac{2z_i}{2}\right)^2 = z_i^2$$ & equal only if $x_i = y_i = z_i$

Why the constraints?

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

- x_i's $< \frac{d}{2} \implies$ 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 $x, y, z \in X_B$: $x + y = 2z \iff \sum_{i=0}^{k} x_i d^i + \sum_{i=0}^{k} y_i d^i = 2 \cdot \sum_{i=0}^{k} z_i d^i$

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

<u>but</u> $\forall i \ x_i + y_i = z_i \implies \forall i \ x_i^2 + y_i^2 \geq 2 \cdot z_i^2$ with equality only if $x_i + y_i = z_i$

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 $k \approx \frac{\log m}{10\sqrt{\log m}} \approx \frac{\sqrt{\log m}}{10}$)

$$X_B = \left\{ \sum_{i=0}^{k} x_i \cdot d^i \ \middle| \ \underset{①}{x_i < \frac{d}{2} \text{ for } 0 \leq i \leq k} + \underset{②}{\sum_{i=0}^{k} x_i^2 = B} \right\}$$

Why the constraints?

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

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

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 $\exists \ i$ s.t. $\underbrace{\text{not}(x_i = y_i = z_i)}$

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

So we have this lemma:

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

$$X_1 + X_2 = 2X_3$$

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

$\Longrightarrow \exists$ tester T': "Natural tester"

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

$O(q^2)$ queries

WE CAN CALCULATE THIS

Prob pass $= \dfrac{\#\text{ triangles}}{\#\text{ node triples}}$

Consequences:

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

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

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

$\Rightarrow$ Only need to find class of graphs on which natural tester does badly & distance $\geq \varepsilon$

Graphs on which natural tester needs lots of queries:

"Natural tester"

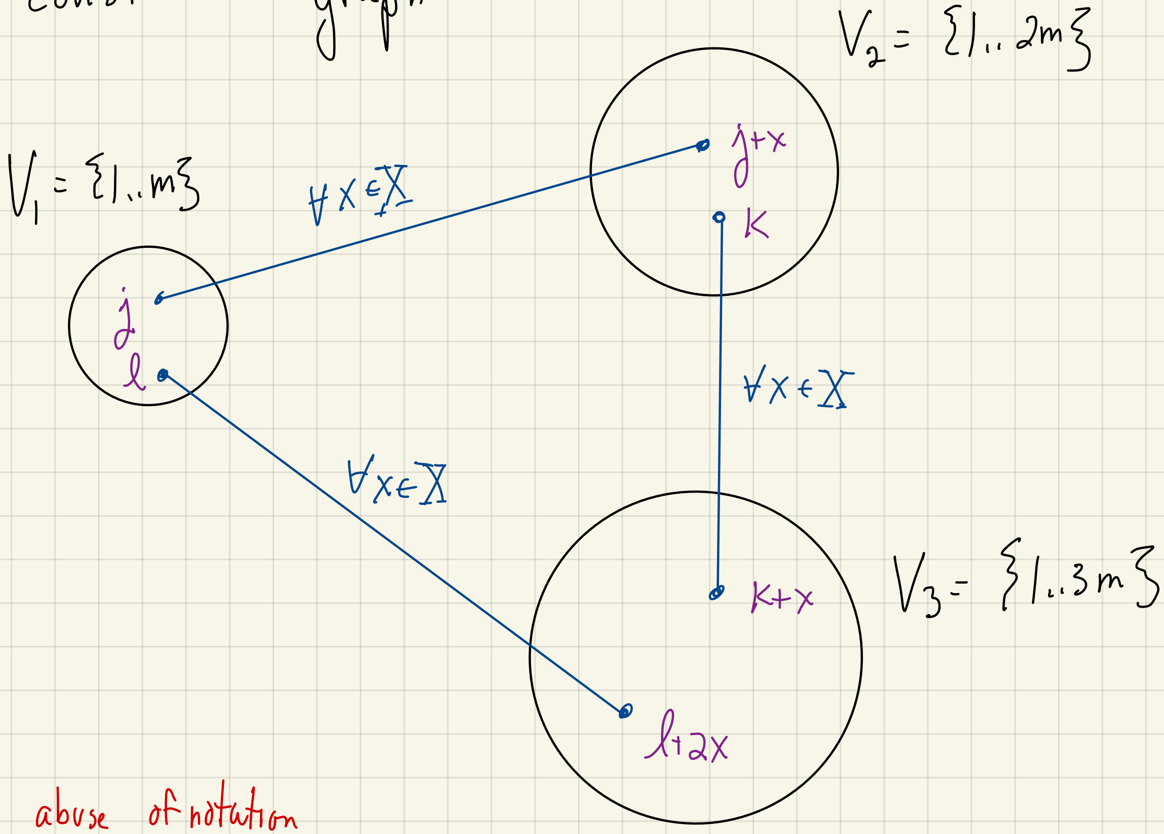
pick $q(n,\varepsilon)$ nodes ran

query submatrix

decide

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

construct graph



slight abuse of notation

should be $(\underset{\in\{1,2,3\}}{i}, \underset{\in\{1..3m\}}{j})$

drop i

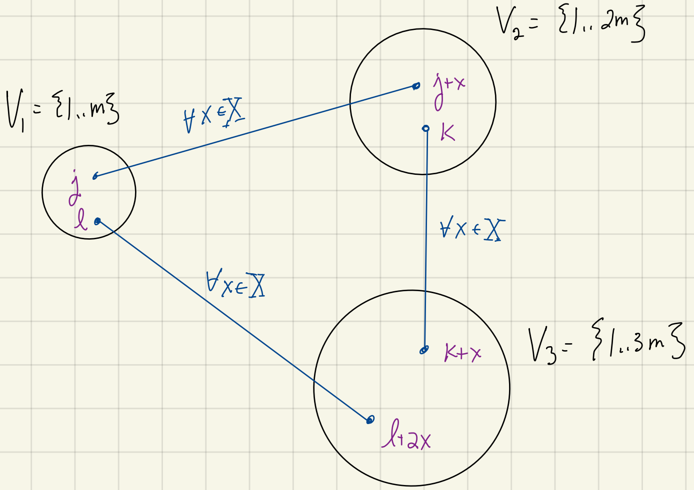
$\#$ nodes $= 6m$ so $m = \theta(n)$

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

$= \underbrace{\theta\left(n^2 / e^{10-\sqrt{\log n}}\right)}_{}$

not exactly dense enough

need to fix this to be $\theta(n^2)$



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

$$x_1 + x_2 = 2x_3$$

"Natural tester"

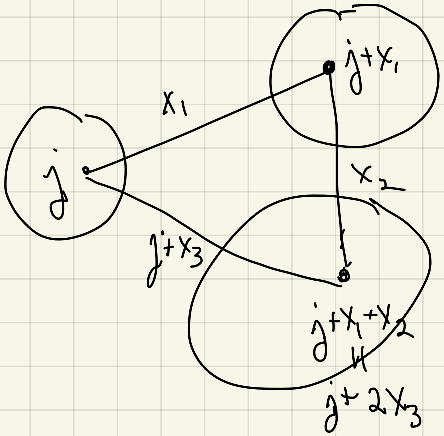
pick $f(n,\varepsilon)$ nodes randomly
query submatrix
decide

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

$$\# \quad m \cdot |x| = \Theta\left(n^2 / e^{10\sqrt{\lg n}}\right)$$

unintended Δ's

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



$\Rightarrow x_1 + x_2 = 2x_3$

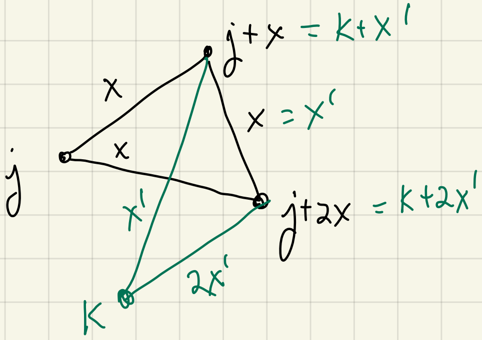
$\Rightarrow x_1 = x_2 = x_3$ since X is sum-free
but these are intended Δ's

$\Big\} \Rightarrow$ no non intended Δ's

disjoint cycles:

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

suppose not:



since $x=x'$, $k=j$

→ ←

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

$$X_1 + X_2 = 2X_3$$

"Natural tester"

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

distance to Δ-free:

Δ's edge disjoint ⇒ must remove ≥1 edge from each Δ

⇓

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

$= \Theta(m|X|)$

<u>Problem</u> need $\Theta(n^2)$ distance

Idea for fix

"s-blow up" $G \Rightarrow G^{(s)}$

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

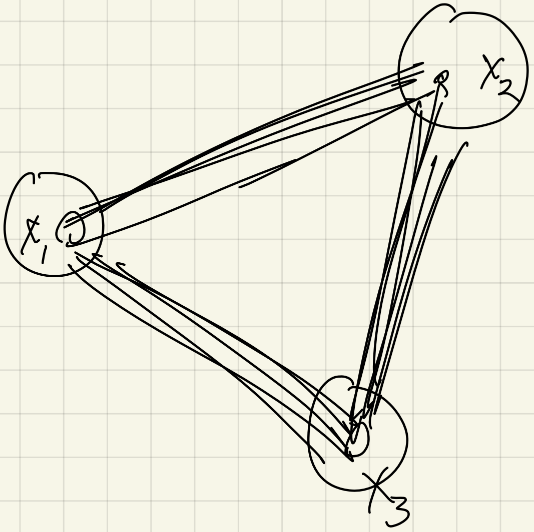
$\sim m|X|s^2$ — edge in G $\Rightarrow$ complete bipartite graph in $G^{(s)}$

$\sim m|X|s^3$ — Δ in G $\Rightarrow$ s^3 Δ's in $G^{(s)}$

↑ prove a lemma

so likely to find one?

but not necessarily edge disjoint now



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

m = largest int s.t.

$\varepsilon \le \frac{1}{e}^{10\sqrt{\log m}}$

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

$\ge m|X|s^2$

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

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

$= \dfrac{|X|}{m} \geq \dfrac{1}{e^{10\sqrt{\log m}}} \geq \varepsilon$ by choice of m + lemma on $|X|$

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

$= \dfrac{m^2}{e^{10\sqrt{\log m}}} \cdot \left(\dfrac{n}{6m}\right)^3 \approx \left(\dfrac{\varepsilon}{c'}\right)^{c' \log \frac{c'}{\varepsilon}} n^3$

If take sample of size q

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

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

so by Markov's $\neq$ $\Rightarrow$ $\Pr[\text{see } \Delta] << 1$

Since 1-sided error, must find Δ to output fail $\blacksquare$