Lecture 10:

Lower bounds for testing

Δ- freeness:

superpoly dependence on ε

is required!

A lower bound for testing Δ-freeness

In a previous lecture:

- saw property test for Δ-freeness
  - const time in terms of n
  - $\underbrace{\text{dependence on } \varepsilon \text{ horrible}}_{\text{is this required?}}$ - tower of 2's

Today:

- answer this question partially (for 1-sided testers)
- When testing H-freeness property,

interesting characterization of bipartiteness

- if H bipartite, poly(1/ε) is enough
- if H not bipartite no poly(1/ε) suffices

(We'll actually prove special case of H = Δ only)

Thm (adj matrix model)

∃ 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.

Main Tools:

(1) Goldreich-Trevisan Thm: (homework)

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
decide

} $O(q^2)$ queries nonadaptive

Consequences:

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

Main tools (cont.):

(2) Additive Number theory lemma

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

of size $\geq \frac{m}{e^{10\sqrt{\log m}}}$

with no non trivial soln to $x_1 + x_2 = 2x_3$

i.e. $x_1 = x_2 = x_3$ is the trivial soln.

no x_3 is average of any x_1, x_2

Will use to construct graphs st.

- far from Δ-free
- natural algorithm needs $\Omega\left(\left(\frac{c}{\varepsilon}\right)^{c \log c/\varepsilon}\right)$ queries

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 $\log m$

Proof of lemma

- let d be integer (later, set to $e^{10\sqrt{\log m}}$)

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

Proof of lemma (cont.)

the two constraints will be used later in a nice way

$$\text{define } X_B = \left\{ \sum_{i=0}^{k} x_i d^i \;\middle|\; \underbrace{x_i < \frac{d}{2}}_{①} \text{ for } 0 \le i \le k \ \& \ \underbrace{\sum_{i=0}^{k} x_i^2 = B}_{②} \right\}$$

view each $x \in M$ as represented in base d

where $x = (\underbrace{x_0 \dots x_k}_{\text{"digits" of } x})$

Claim $X_B \subseteq M$

Why? largest number in X_B

$$\le d^{k+1} \le d^{(\lfloor \frac{\log m}{\log d} \rfloor - 1)+1} \le d^{\log_d m} = m^{\log_d d} = m$$

What is B? Pick s.t. $|X_B|$ maximized

Why the constraints?

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

we'll see why this is useful soon

(2) will use (along with (1)) to show that X_B is "sum-free"

Claim X_B is "sum free" i.e. $\nexists\, x, y, z \in X_B$ s.t.
$x + y = 2z$

Pf of claim assume to contrary
for $x, y, z \in X_B$

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

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

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

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

use Jensens $\neq$: $\dfrac{\sum_{i=1}^{n} f(a_i)}{n} \geq f\left(\dfrac{\sum_{i=1}^{n} a_i}{n}\right)$ with equality only if a_i's are all =

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

■ (proof of note)

finishing proof of claim:

if x, y, z s.t. not$(x=y=z)$

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

then note $\Rightarrow$ $x_i^2 + y_i^2 > 2z_i^2$

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

but then:

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

$\rightarrow\!\leftarrow$

but how do we know that X_B is big?

- $B \leq (k+1)\left(\frac{d}{2}\right)^2 < kd^2$

  $\uparrow$ bound on digits of B

- $\left|\bigcup_B X_B\right| \geq \left(\frac{d}{2}\right)^{k+1} > \left(\frac{d}{2}\right)^k$

  since disjoint $\rightarrow$ $\| \quad \sum_B |X_B|$

- $\exists\, B$ s.t. $|X_B| \geq \dfrac{\left(\frac{d}{2}\right)^k}{kd^2}$

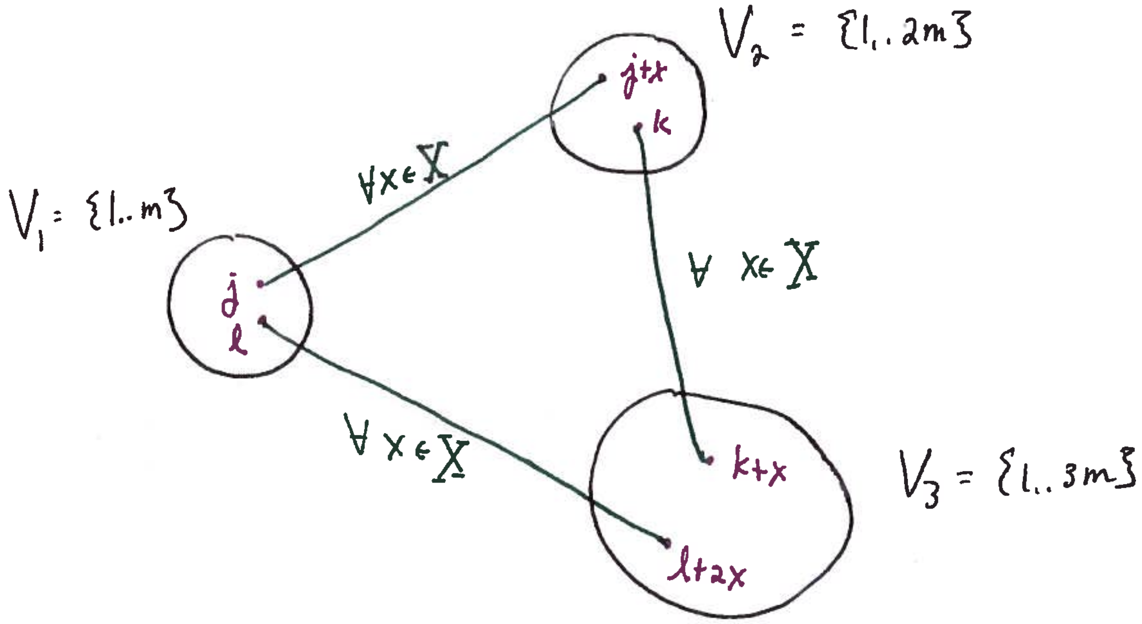
- Use settings of d, k, get $|X_B| \geq \dfrac{m}{e^{10\sqrt{\log m}}}$

For l.b.: Not enough! need another idea, but won't do it here

Proof of Thm (prop testing bound)

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

construct a graph:



- will abuse notation:

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

  will drop i if easy to see from context

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

• # cycles :

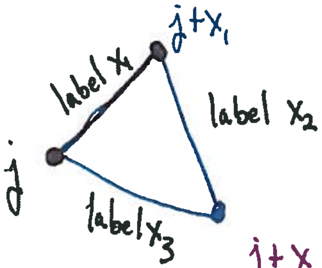
intended Δ's : $j,\ j+x,\ j+2x$

intended Δ's is $m|X| = \Theta(n^2/e^{10\sqrt{\log n}})$

nonintended Δ's :

• no edges internal to V_1, V_2 or V_3

∴ any Δ has
$u \in V_1$
$v \in V_2$
$w \in V_3$



$$\left.\begin{array}{l} j+x_1+x_2 \\ = j+2x_3 \end{array}\right\} \Rightarrow x_1+x_2 = 2x_3$$

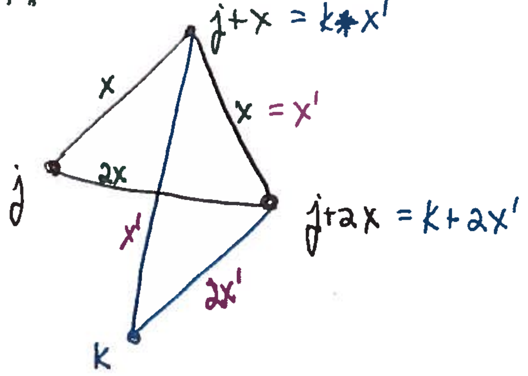
$$\Rightarrow \underbrace{x_1 = x_2 = x_3}_{\text{but these are intended!}} \quad \text{since } X \text{ is sum-free}$$

∴ no nonintended Δ's

• # disjoint cycles:

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

suppose not:



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

• distance to Δ-free:

must remove ≥ 1 edge from each Δ.

$\Downarrow$

"absolute" distance from Δ-free $= \Theta(\#\Delta\text{'s})$

$$= \Theta\left(\frac{n^2}{e^{10\sqrt{\log n}}}\right)$$

$$= \Theta(m/|X|)$$

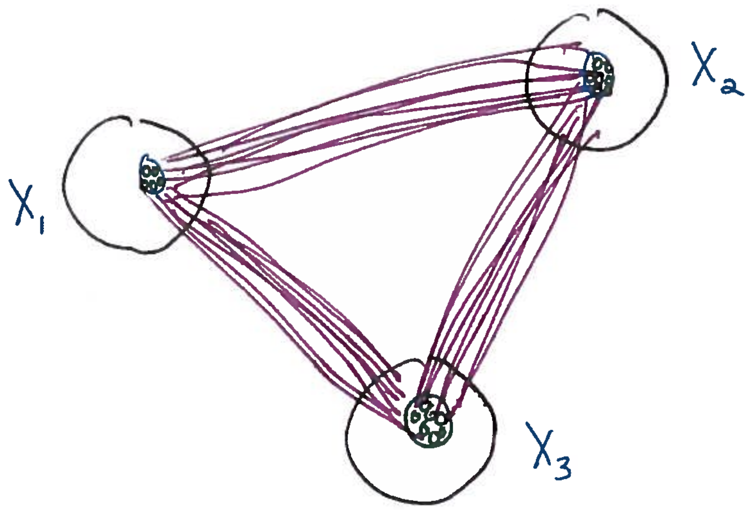
Problem need $\Omega(\varepsilon n^2)$ distance

<u>Idea for fix</u>

s-blow-up $\quad G \to G^{(s)}$

vertex in G $\;\to\;$ size s independent set in $G^{(s)}$

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



Note: Δ in $G \Rightarrow s^3$ Δ's in $G^{(s)}$ $\Rightarrow$ likely to find one!

\# nodes in $G^{(s)}$ $\sim m \cdot s$ (actually $6ms$)

\# edges " " $\sim m|X| \cdot s^2$

\# triangles " " $\sim m|X| s^3$ (no longer disjoint)

<u>Lemma</u> dist of $G^{(s)}$ from Δ-free

$\geq$ \# edge disjoint Δ's

$\geq m|X| s^2$

<u>Proof</u> show each triangle in $G \Rightarrow s^2$ disjoint Δ's in $G^{(s)}$

Given ε, pick m to be largest int satisfying

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

this m satisfies

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

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

$\Rightarrow$ #edges $\sim$ distance $\sim \varepsilon n^2$ (since $\approx \frac{m|X|s^2}{m^2 s^2}$ ← size of adj matrix $= \frac{|X|}{m} \ge \frac{1}{e^{10\sqrt{\log m}}} \ge \varepsilon$) ↑ $|X| = \frac{m}{e^{10\sqrt{\log m}}}$

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

↑

$$m|X| \cdot s^3 = \frac{m^2}{e^{10\sqrt{\log m}}} s^3$$

$$= \frac{1}{\varepsilon}\left(\left(\frac{c}{\varepsilon}\right)^{(\log c/\varepsilon)^2} \cdot \left(\frac{\varepsilon}{c'}\right)^{\log c/\varepsilon} n^3\right.$$

Finally if take sample of size q

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

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

by Markov's ≠ $\Rightarrow$ $\Pr[\text{see } \Delta] \ll 1$

But since 1-sided error, must find Δ in order to fail ∎