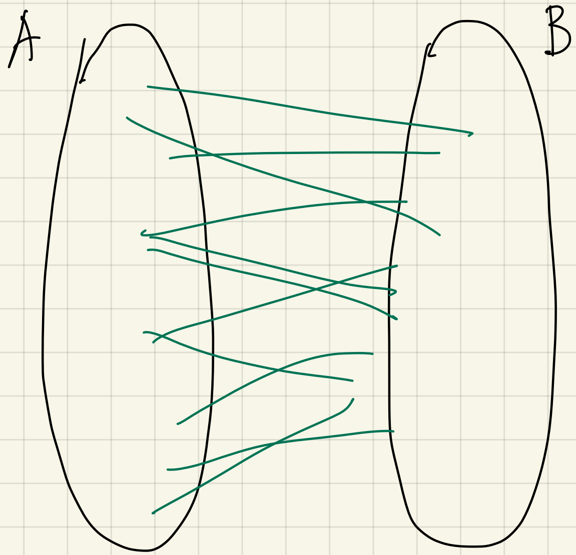
Lecture 10

Testing dense graph properties via SRL:

$\triangle$- freeness

Begin lower bound

Density & Regularity of set pairs:



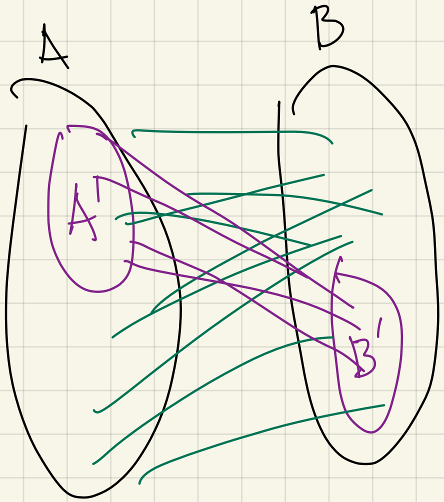
def. For $A, B \subseteq V$ s.t.

(1) $A \cap B = \emptyset$

(2) $|A|, |B| \geq 1$

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"

Lemma

density

$\forall \eta > 0$

today assume $\eta < 1/2$

regularity parameter, depends only on η

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

#triangles $\longrightarrow$ depends only on η

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

if $\eta < 1/2$

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

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

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

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

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

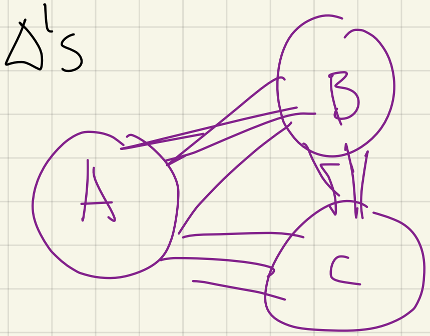
s.t. if A, B, C disjoint subsets of V s.t. each pair

is γ-regular with density $> \eta$

then G contains $\geq \delta \cdot |A|\cdot|B|\cdot|C|$ distinct Δ's

$\geq \delta^3/16 \cdot |A||B||C|$

with node in each of A, B, C.



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

$\delta(n)$

Do interesting graphs have regularity properties?

Yes in some sense <u>all</u> 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 $V_1 \ldots V_k$ (for const k) s.t.

<u>all</u> pairs (V_i, V_j) are ε-regular"

will get only "most"

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

to be useful sometimes need $k \ge m$ 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 sets ← # is const ind of n $<< T$

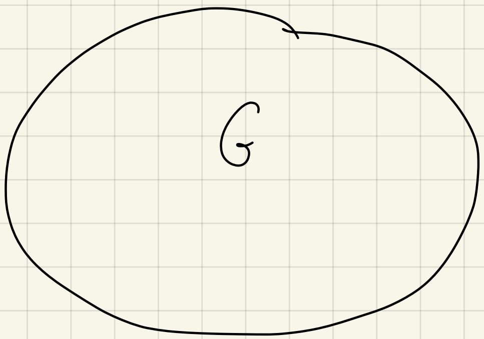
then exists equipartition $\mathcal{B}$ into k sets which refines $\mathcal{A}$

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

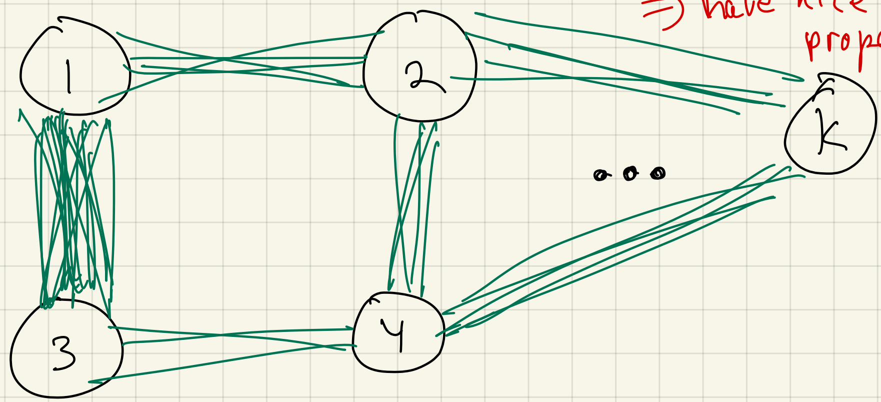
& $\le \varepsilon \binom{k}{2}$ set pairs not ε-regular

const # partitions
each pairs behaves like random graph
⇒ have nice properties

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



⇒



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

1-sided error

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

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

Corr Algorithm has desired behavior

Why?

- 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 $\leq 1 - \delta$

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

  $$\leq e^{-c} < 1/3$$

  for proper choice of c

  $\Rightarrow$ reject with prob $\geq 2/3$

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.

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

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

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

how? start with arbitrary equipartition into $5/\varepsilon$ sets $\leftarrow$ this is why we need ability to refine any partition

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

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

how many? $\leq \frac{n}{K} \cdot n \leq \frac{\varepsilon n^2}{5}$

deg w/in V_i

sum over all nodes

2) delete edges between ε'-non regular pairs

how many?

$$\leq \underbrace{\varepsilon' \binom{K}{2}}_{\#\text{ non regular pairs}} \cdot \underbrace{\left(\frac{n}{K}\right)^2}_{\text{max \# edges per pair}} \leq \frac{\varepsilon}{5} \cdot \frac{K^2}{2} \cdot \frac{n^2}{K^2} \leq \frac{\varepsilon}{10} n^2$$

since $|V_i| \approx |V_j| = \frac{n}{K}(+1)$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

<u>But</u> 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!!!

Δ in G' must connect:

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

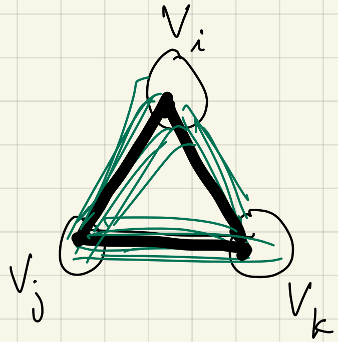
since no edges internal to partition in G'

2) regular pairs

since nonregular pair edges deleted in G'

3) high density pairs

since removed low density pairs in G'



$\therefore\ \exists\ i, j, k$ distinct s.t. $x \in V_i,\ y \in V_j,\ z \in V_k$

V_i, V_j, V_k all $\geq \underbrace{\frac{\varepsilon}{5}}_{\geq \eta}$ density pairs

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

Δ-counting lemma $\Rightarrow$

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

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

$$\geq \frac{1}{2} \cdot \frac{\varepsilon^3}{8000} = \frac{\varepsilon^3}{16000}$$

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

$$\geq \delta' \cdot \binom{n}{3} \quad \Delta\text{'s in } G' \text{ (and thus in } G)$$

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

$\blacksquare$

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:

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

queries to adj matrix

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

more generally,
H-freeness for all const sized H

For dense graphs, testable properties

- 1-sided error const time ≈ hereditary graph properties (closed under vertex removal: chordal, perfect, interval)

  difficulty: infinite set of forbidden subgraphs

- 2-sided error const time ≈ 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?

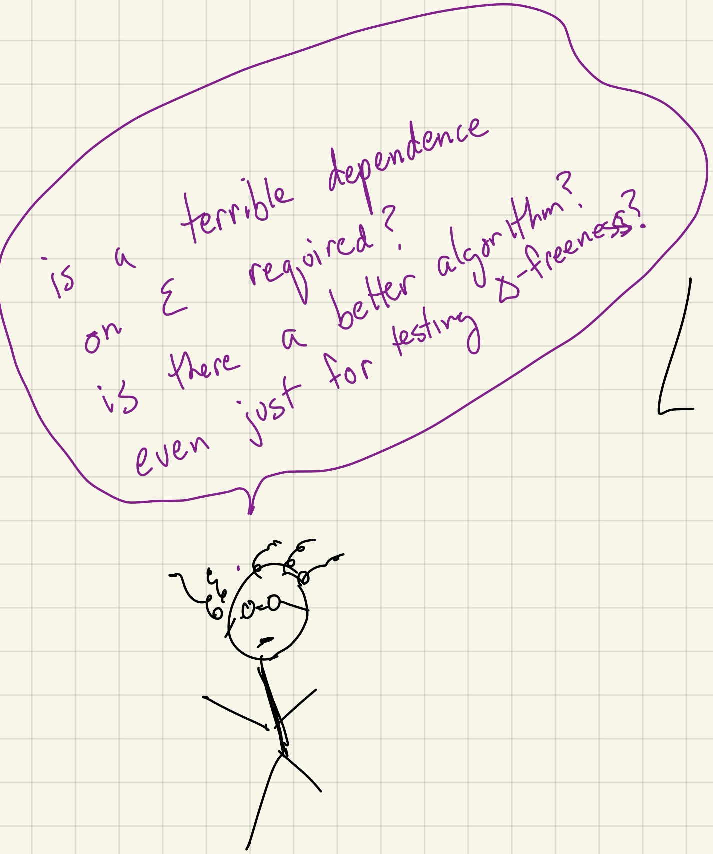
An intriguing characterization of bipartite graphs:

For graphs in adjacency matrix model:

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

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

we will prove this
for $H = \Delta$ only



Lower bounds for testing Δ-freeness:

Superpoly dependence on ε is 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

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

with no nontrivial soln to

no 3 evenly spaced points $\rightarrow$

$$X_1 + X_2 = 2X_3$$

$\Uparrow$

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

(2) any algorithm needs lots of 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?

$\{1, 2, 4, 5, 6, 7, 8, 9, 10, \dots\}$

$\{1, 2, 4, 8, 16, 32, \dots\}$

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

with no nontrivial soln to

$$X_1 + X_2 = 2X_3$$

<u>examples</u>:

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

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

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 $X_1 + X_2 = 2X_3$

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

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 $< \frac{d}{2}$ $\Rightarrow$ sum pairs of elts in X_B doesn't generate any 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 $x, y, z \in X_B$: $x + y = 2z \iff \sum_{i=0}^{k} x_i d^i + \sum_{i=0}^{k} y_i d^i = 2 \cdot \sum_{i=0}^{k} z_i d^i$

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

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 $x_i + y_i = z_i$

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_i'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 $x_i = y_i = z_i$

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

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

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

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

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 $\exists\, x,y,z$ s.t. 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 $\underbrace{\sum x_i^2}_{=B} + \underbrace{\sum y_i^2}_{=B} > \underbrace{\sum 2z_i^2}_{=2B}$ CONTRADICTION