Hardness of Approximation of (Multi-)LCS over small Alphabet

Diptarka Chakraborty
(joint work with Amey Bhangale and Rajendra Kumar)

ISI Kolkata

21st February, 2020

Longest Common Subsequence

Subsequence:

A C A G C T A C

A A G T C

LCS (between two strings):

Given two strings $x, y \in \Sigma^n$, find the length of the largest subsequence that is common to both x and y.

LCS between two strings

- Basic dynamic programming solves in $\underbrace{O(n^2)}$ time

"almost" tight assuming SETH

3-SAT on n-variables can't be solved in $2^{(1-\delta)n}$ time for any $\delta > 0$

LCS between more than two strings

- When # input strings is unrestricted, known to be NP-complete (even for $\Sigma = \{0,1\}$)
- For m strings any $O(n^{m-\epsilon})$ for any $\epsilon > 0$, time algo refutes SETH

What about approximation?

c-approx. : Output ℓ such that

$c \leq 1$

$$c \cdot |LCS| \leq \ell \leq |LCS|$$

Rand

- For $\Sigma = \{0,1\}$, $O(n^{-\epsilon})$ approx. in $O(n^{2-2\epsilon})$ time
- Recently, Hajiaghayi et. al. get $O(n^{-0.497956})$-approx. in $O(n)$ time

(break $O(n^{-\frac{1}{2}})$ barrier)

Approximating Multi-LCS

- Jiang and Li ['95] showed: There exists $\delta > 0$, s.t. polytime $n^{-\delta}$-approx. algo implies P=NP

Approximating Multi-LCS

- Jiang and Li ['95] showed: There exists $\delta > 0$, s.t. polytime $n^{-\delta}$- approx. algo implies P=NP
- Above reduction only works for $|\Sigma| = \Omega(n)$
- What can we say for small Σ?

Approximating Multi-LCS

- For any Σ, $\frac{1}{|\Sigma|}$ - approx. is easy

Is it the best possible in poly time?

- No hardness is known even for $|\Sigma| = o(n)$

Our Result

Thm: There exists a growing $f(n) = n^{o(1)}$ s.t. assuming ETH, no polytime $\frac{1}{f(n)}$ - approx. algo for Multi-LCS over alphabet of size $n^{o(1)}$.

No $2^{o(n)}$ time algo for 3-SAT

$n^{O\left(\frac{1}{(\log \log n)^{\frac{1}{4}}}\right)}$

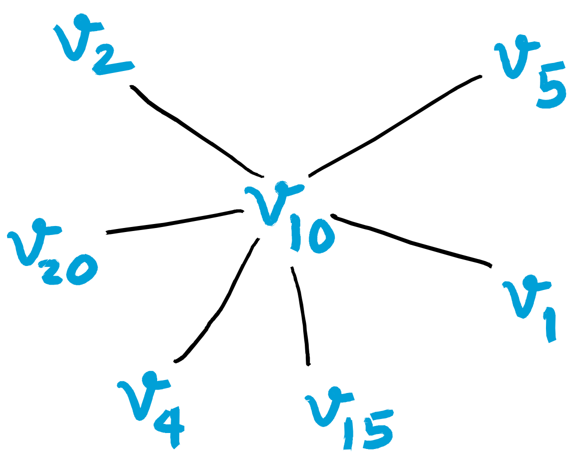
Reduction (Starting Point)

Max-Clique to Multi-LCS

- Given $G = (V, E)$ with n nodes create $2n$ strings
- $V = \{v_1, v_2, \ldots, v_n\} = \Sigma$

Reduction (Starting Point)

Max-Clique to Multi-LCS



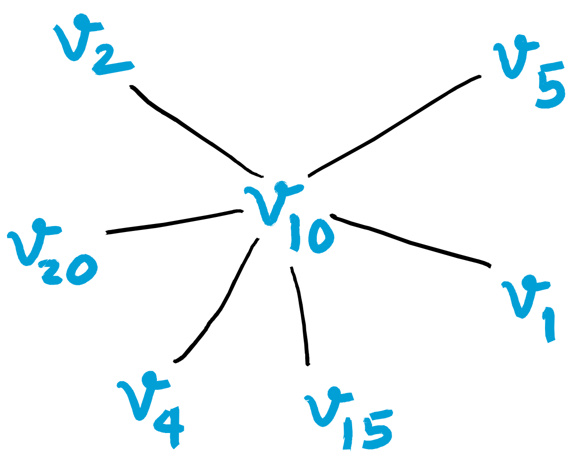
x_{10} : $v_1\ v_2 \cdots v_9\ v_{11} \cdots v_n\ v_{10}\ v_{15}\ v_{20}$

x'_{10} : $v_1\ v_2\ v_5\ v_{10}\ v_1 \cdots v_9\ v_{11} \cdots v_n$

(For each v_i create x_i and x'_i)

Reduction (Starting Point)

$$\boxed{\text{Max-Clique to Multi-LCS}}$$



$$x_{10} : v_1 \; v_2 \cdots v_9 \; v_{11} \cdots v_n \; v_{10} \; \underbrace{v_{15} \; v_{20}}_{N_>(v_{10})}$$

$$x'_{10} : \underbrace{v_1 \; v_2 \; v_5}_{N_<(v_{10})} \; v_{10} \; v_1 \cdots v_9 \; v_{11} \cdots v_n$$

(For each v_i create x_i and x'_i)

Reduction (Starting Point)

Max-Clique to Multi-LCS

1) A clique of size k
   $\Rightarrow$ LCS of size k

2) All v_i's in any common subsequence form a clique

x_{10} : $v_1\ v_2 \cdots v_9\ v_{11} \cdots v_n\ v_{10}\ v_{15}\ v_{20}$

x'_{10} : $v_1\ v_2\ v_5\ v_{10}\ v_1 \cdots v_9\ v_{11} \cdots v_n$

(For each v_i create x_i and x'_i)

Reduction (Starting Point)

$\boxed{\text{Max-Clique to Multi-LCS}}$

1) A clique of size c
$\Rightarrow$ LCS of size c

$\boxed{\text{2) All } v_i\text{'s in any common subsequence forms a clique}}$

e.g. $v_1 \; v_2 \; v_{10} \; v_{15}$

$$x_{10} : v_1 \; v_2 \; \cdots \; v_9 \; v_{11} \cdots \; v_n \; v_{10} \; \overbrace{v_{15} \; v_{20}}^{N_>(v_i)}$$

$$x'_{10} : \underbrace{v_1 \; v_2 \; v_5}_{N_<(v_i)} \; v_{10} \; v_1 \cdots v_9 \; v_{11} \; O \; v_n$$

(For each v_i create x_i and x'_i)

Alphabet Reduction

- Replace each v_i by some $S_i \in \Sigma'^{\,m}$

  $|\Sigma'| << |\Sigma|$

  $\quad\quad\quad\quad\;\; \| $

  $\quad\quad\quad\quad\;\; n$

- Completeness still holds, i.e.,

  Clique of size k $\Rightarrow$ LCS of size **at least** km

Alphabet Reduction

- Replace each v_i by some $S_i \in \Sigma'^{m}$

  $|\Sigma'| << |\Sigma|$ $= n$

- Completeness still holds, i.e.,

  Clique of size k $\implies$ LCS of size at least km

- What about Soundness?
  - Now S_i may not map to S_i
  - S_i can be partially map to S_j's

Alphabet Reduction

Natural idea: Can we select s_i's s.t.

$\forall_{i \neq j}$ $|LCS(s_i, s_j)|$ is small

Alphabet Reduction

Natural idea: Can we select s_i's s.t.
$\forall_{i \neq j} \; |LCS(s_i, s_j)|$ is **small**

- Let us pick $s_1, \ldots, s_n$ uniformly at random

Alphabet Reduction

Natural idea: Can we select S_i's s.t.
$\forall_{i \neq j}$ $|LCS(S_i, S_j)|$ is **small**

- Let us pick $S_1, \ldots, S_n$ uniformly at random

Guarantee: W.h.p. $\forall_{i \neq j}$, $|LCS(S_i, S_j)| \leq \alpha m$,
for $|\Sigma| = O(\frac{1}{\alpha^2})$, $m = O(\frac{1}{\alpha^2} \log n)$

Randomized Reduction

- Use previous reduction
- Pick $s_1, \ldots, s_n$ uniformly at random
- Replace v_i by s_i

Y_i : $s_1 \cdots s_{i-1} \; s_{i+1} \cdots s_n \; s_i \; \mathcal{N}_{>i}$

Y_i' : $\mathcal{N}_{<i} \; s_i \; s_1 \cdots s_{i-1} \; s_{i+1} \cdots s_n$

Randomized Reduction

Soundness: Large LCS $\Rightarrow$ Large clique?

Randomized Reduction

Soundness: Large LCS $\Rightarrow$ Large ~~clique~~ density

Densest k-Subgraph (with perfect completeness)

γ-DkS(k, n): Given a graph G with n nodes

Gap version
- Yes : $\exists$ a clique of size k
- No : All subgraph of size k have density $\leq \gamma$

\# vertices

Reduction from γ-DkS(k, n)

Thm: Let $k = \frac{\beta(n)}{\gamma(n)} \cdot n$, $\beta < \gamma \leq 1$.

No polytime algo for $\frac{\gamma^2}{4}$-DkS(k, n)

$\Downarrow$

No polytime algo for 2γ-approx. Multi-LCS over Σ of size $O\left(\frac{1}{\beta^6}\right)$.

Reduction from γ-DkS(k,n)

Thm: Let $k = \frac{\beta(n)}{\gamma(n)} \cdot n$, $\beta < \gamma \leq 1$.

No polytime algo for $\frac{\gamma^2}{4}$-DkS(k,n)

$\Downarrow$

No polytime algo for 2γ-approx. Multi-LCS over Σ of size $O\left(\frac{1}{\beta^6}\right)$.

Manurangsi '17: Under ETH, for regime

$$\gamma = n^{-O\left(\frac{1}{(\log\log n)^{1/4}}\right)},$$

$$k \in \left[n^{1-\epsilon},\ n^{1-\Omega\left(\frac{1}{\log\log n}\right)}\right]$$

Reduction from γ-DkS(k,n)

Thm: Let $k = \frac{\beta(n)}{\gamma(n)} \cdot n$, $\beta < \gamma \le 1$

No polytime algo for $\frac{\gamma^2}{4}$-DkS(k,n)

$\Downarrow$

No polytime algo for $\underbrace{2\gamma\text{-approx.}}_{n^{-o(1)}}$ Multi-LCS over Σ of size $\underbrace{O\left(\frac{1}{\beta^6}\right)}$. $\leftarrow n^{o(1)}$

Manurangsi '17: Under ETH, for regime

$\rightarrow \gamma = n^{-O\left(\frac{1}{(\log\log n)^{1/4}}\right)}$

$\rightarrow k \in \left[n^{1-\epsilon}, \underbrace{n^{1-\Omega\left(\frac{1}{\log\log n}\right)}}\right]$

Proof of Soundness

Y_1

Y_1'

Y_2

Y_2'

$\vdots$

Y_n

Y_n'

$\xrightarrow{\text{LCS}}$ L

To show: If G is a NO instance of $\frac{\gamma^2}{4}$-DkS, then $|L| \le 2\beta m n$

all k-sized subgraph has density $\le \gamma^2/4$

$\beta = \sqrt{8\alpha}$

$\left(|LCS(s_i, s_j)| \le \alpha m\right)$

Proof of Soundness

Y_1

Y_1'

$Y_2 \xrightarrow{LCS}$

Y_2'

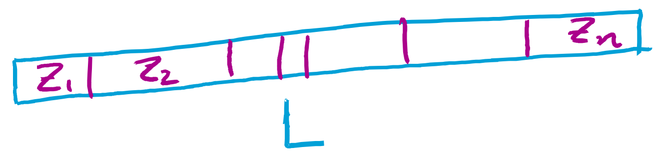
$\vdots$

$Y_n = S_1 \cdots\cdots S_n$

Y_n'

$\mathcal{N}_{>n} = \phi$

Z_i is a subseq. of S_i



Proof of Soundness

Y_1

Y_1'

Y_2

Y_2'

$\vdots$

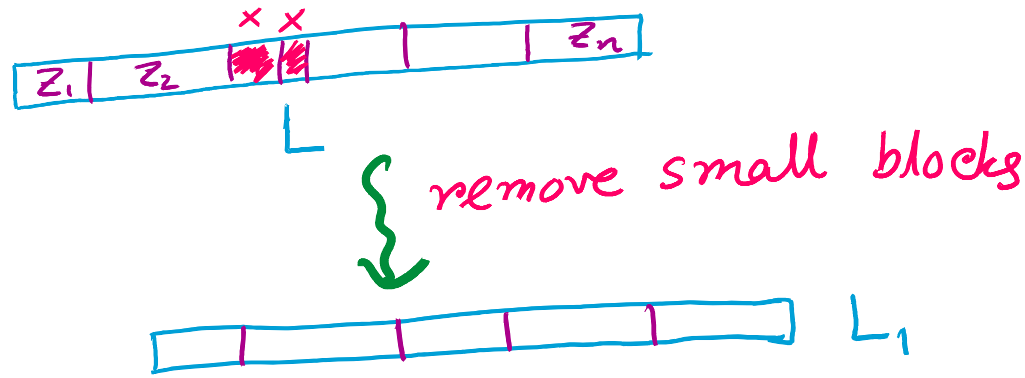
$Y_n = S_1 \cdots S_n$

Y_n'

$N_{>n} = \phi$

LCS

Z_i is a subseq. of S_i



$|Z_i| > \beta m$

Proof of Soundness

L_1  $\longrightarrow V_H := \{ v_i \mid z_i \in L_1 \}$

$$|V_H| \geq \frac{|L| - \beta m n}{m} \geq \beta n$$

Proof of Soundness

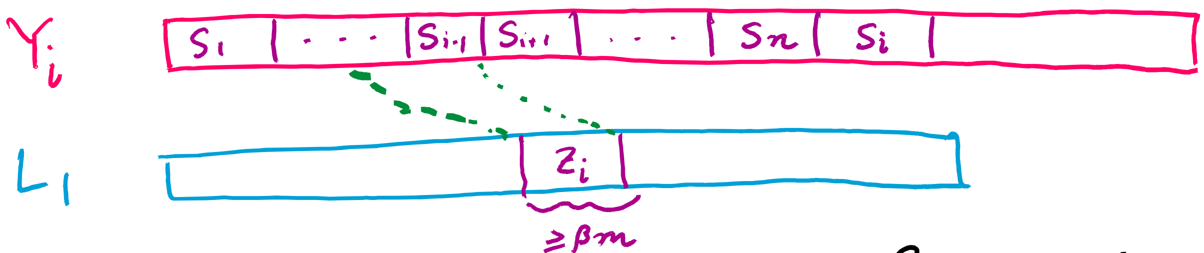
L_1  $\longrightarrow V_H := \{ v_i \mid z_i \in L_1 \}$

$$|V_H| \geq \frac{|L| - \beta m n}{m} \geq \beta n$$

<u>Notation:</u> $\underbrace{C[i,j]}_{\substack{\uparrow \\ \text{to capture} \\ \#\text{ blocks from } [i,j]}} :=$ set of $z_t \in L_1$ s.t. $i \leq t \leq j$

Sparse Local LCS

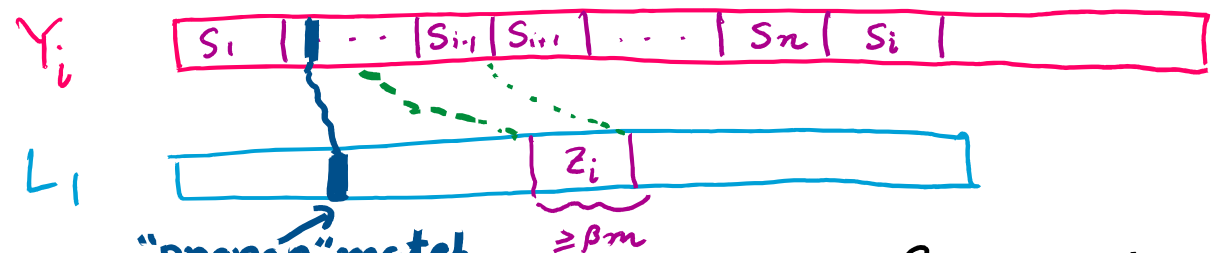
Case-I



- To map z_i we need $\geq \frac{\beta}{\alpha}$ blocks

Sparse Local LCS

Case - I



say, in Z_j

- To map Z_i we need $\geq \beta/\alpha$ blocks
- All blocks in $C[j+1, i-1]$ needs $\geq \beta/\alpha$ blocks

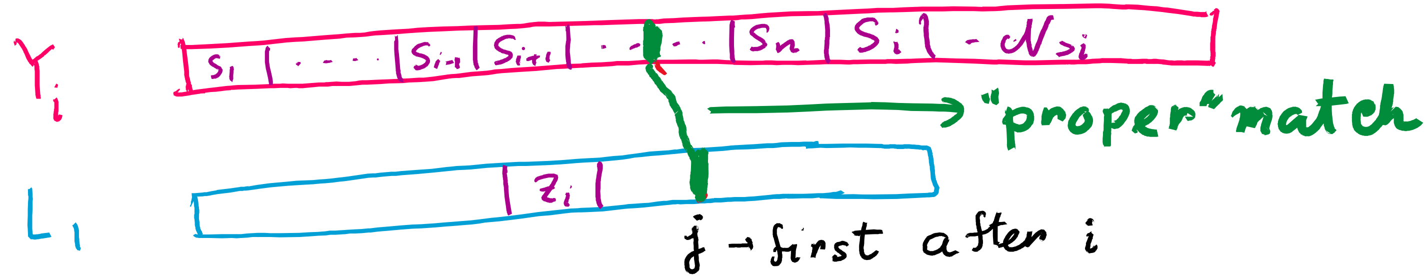
$$\Rightarrow \frac{\beta}{2\alpha} |C[j+1, i]| \leq j-i$$

$$\Rightarrow \frac{|C[j+1, i]|}{j-i} \leq \frac{2\alpha}{\beta}$$

very small

Sparse Local LCS

Case – II



- As before,

$$\frac{|C[i, j-1]|}{j-i} \leq \frac{2\alpha}{\beta}$$

Sparse Local LCS

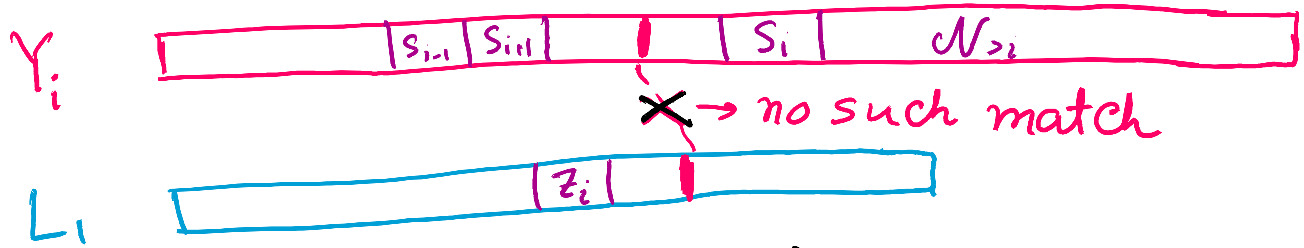
- Since LCS is large, only a "small" portion can be locally sparse

- Remove all such portions $C[i,j]$, and form $V_H' \subseteq V_H$

potentially dense

Dense Case

Case-III

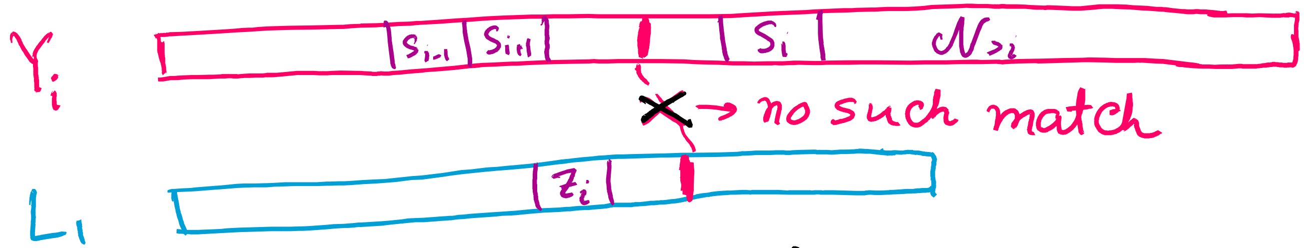


- $z_i \; z_{i+1} \dots \in L_1$ (i.e, $C[i, n]$) map to

  $\underbrace{S_{i+1} \dots S_i \; \mathcal{N}_{>i}}_{\leq 2(n-i)+1}$.

Dense Case

Case-III



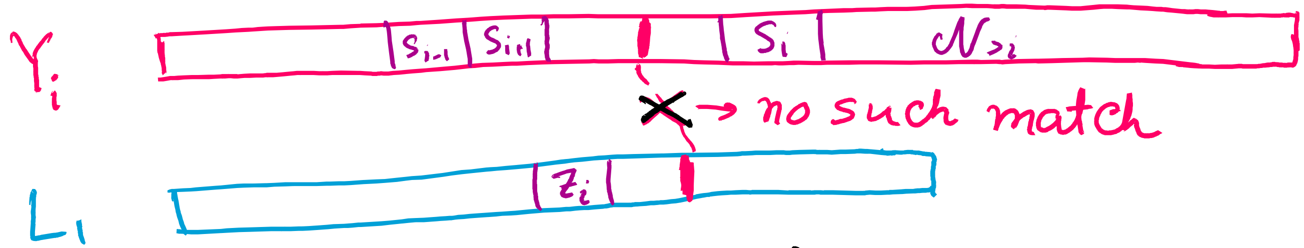
- $z_i\ z_{i+1} \ldots \in L_1$ (i.e, $C[i,n]$) map to $\underbrace{s_{i+1} \ldots s_i\, \mathcal{N}_{>i}}_{\leq 2(n-i)+1}$.

- Each block in $V_H^{>i} \setminus \mathcal{N}_{>i}$ needs $\geq \frac{\beta}{\alpha}$ blocks

vertices $>i$ in V_H

Dense Case

Case-III



- $z_i \; z_{i+1} \dots \in L_1$ (i.e, $C[i,n]$) map to $\underbrace{S_{i+1} \dots S_i \, \mathcal{N}_{>i}}_{\leq 2(n-i)+1}$.

vertices $>i$ in V_H

- Each block in $V_H^{>i} \setminus \mathcal{N}_{>i}$ needs $\geq \frac{\beta}{\alpha}$ blocks

$$|V_H^{>i} \cap \mathcal{N}_{>i}| + \frac{\beta}{2\alpha} |V_H^{>i} \setminus \mathcal{N}_{>i}| \leq 2(n-i)+1$$

Dense Case

Symmetric case: $\left|V_H^{<i} \cap \mathcal{N}_{<i}\right| + \frac{\beta}{2\alpha}\left|V_H^{<i} \setminus \mathcal{N}_{<i}\right| \leq 2i-1$

Dense Case

Symmetric case: $\left|V_H^{<i} \cap \mathcal{N}_{<i}\right| + \frac{\beta}{2\alpha}\left|V_H^{<i} \setminus \mathcal{N}_{<i}\right| \leq 2i-1$

deg in H

combine

$$\Longrightarrow \left|V_H \cap \mathcal{N}(v_i)\right| + \frac{\beta}{2\alpha}\left|V_H \setminus \mathcal{N}(v_i)\right| \leq 2n$$

Dense Case

Symmetric case: $|V_H^{<i} \cap \mathcal{N}_{<i}| + \frac{\beta}{2\alpha}|V_H^{<i} \setminus \mathcal{N}_{<i}| \leq 2i-1$

combine

deg in H

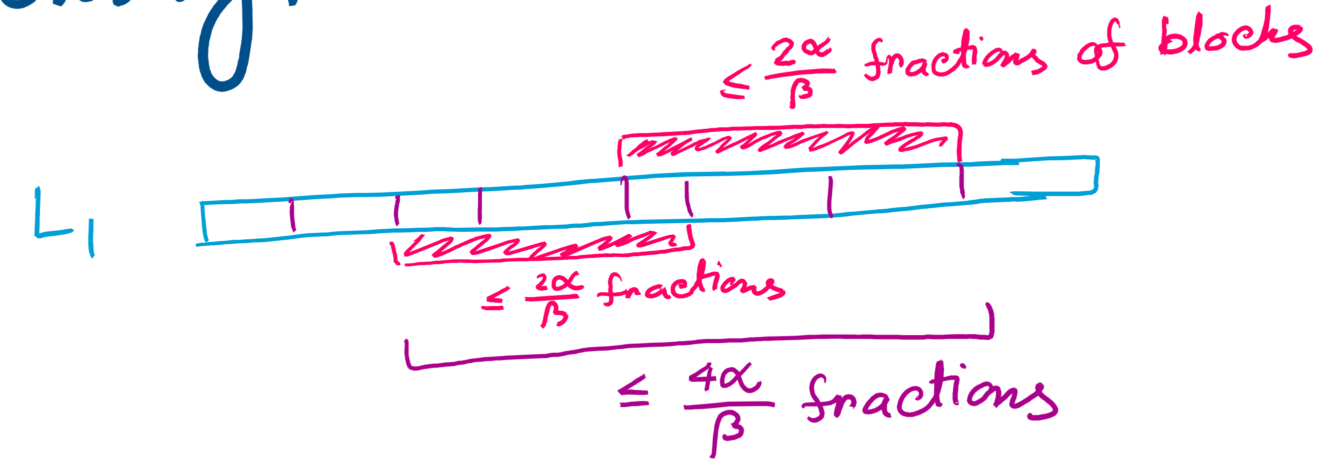
$$\Longrightarrow |V_H \cap \mathcal{N}(v_i)| + \frac{\beta}{2\alpha}|V_H \setminus \mathcal{N}(v_i)| \leq 2n$$

$$\Rightarrow \boxed{|V_H \cap \mathcal{N}(v_i)| \geq |V_H| - \frac{4\alpha}{\beta} n}$$

true for all i s.t. case I, II do not hold

i.e., all $v_i \in V_H'$ (after removal of sparse cases)

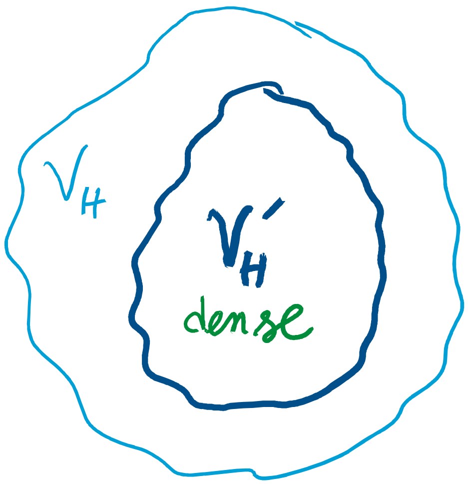
Large Density



$$|V_H'| \geq |V_H| - \frac{4\alpha}{\beta} n$$

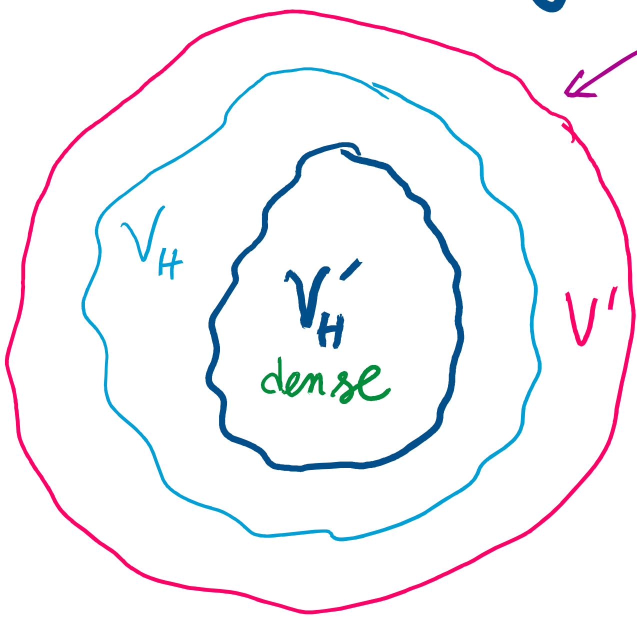
still a large subgraph

Large Density



if H itself is of size k, we are done

Large Density



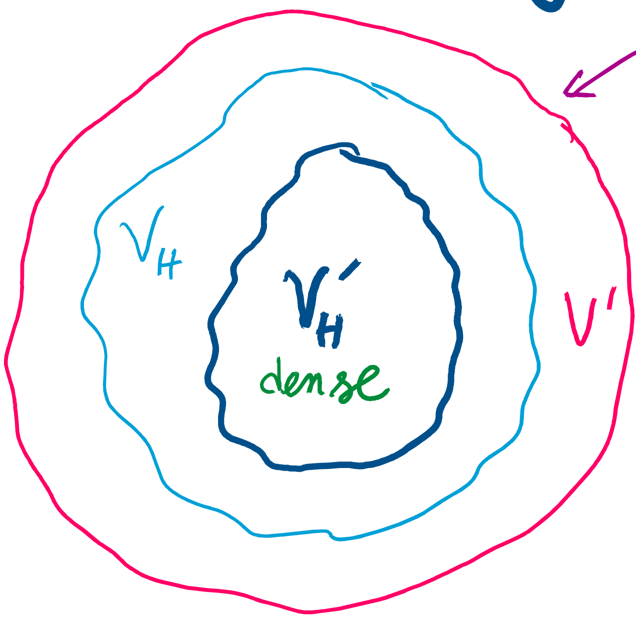
if H itself is of size k, we are done

otherwise

$$|V'| = k = \frac{\beta}{\gamma} n$$

$$\text{density}(V') \geq \frac{\frac{1}{2} \sum_{v \in V'_H} \left(|V_H| - \frac{4\alpha}{\beta} n \right)}{\binom{|V'|}{2}}$$

Large Density



if H itself is of size k, we are done

otherwise

$$|V'| = k = \frac{\beta}{\gamma} n$$

$$\text{density}(V') \geq \frac{\frac{1}{2} \sum_{v \in V_H'} \left(|V_H| - \frac{4\alpha}{\beta} n \right)}{\binom{|V'|}{2}}$$

$$\geq \left(\frac{\gamma}{2}\right)^2$$

(for the choices of α, β, γ)

Derandomization

- Use constructions of long distance synchronization strings
- Gives exactly same guarantee as random strings, but with "slightly large" alphabet

Conclusion

- No hardness is known for r-DkS when $k = \Theta\left(\frac{n}{\text{poly}\log n}\right)$, $r = (\log n)^{-c}$

  (this would give hardness for Multi-LCS over polylog sized Σ)
- Our reduction can't give hardness for $O(1)$-sized Σ, since r-DkS is easy for those r, k
- Open: Prove hardness for $O(1)$-sized Σ

Thank You!