Recap:

<u>Goal:</u> Any streaming algo. that computes $(1+\varepsilon)$ approx of F_0 needs $\Omega(\frac{1}{\varepsilon^2})$ space

$\varepsilon = \frac{1}{\sqrt{n}}$

$$\text{GAP-HAMMING}(t) = \begin{cases} 1 & d_H(x,y) < t - c\sqrt{n} \\ 0 & d_H(x,y) > t + c\sqrt{n} \end{cases}$$

Pick $t = \frac{n}{2}$

<u>Thm:</u> Randomized one-way comm. complexity of GAP-HAMMING is $\Omega(n)$

<u>Pf</u> Assume wlog n odd, large enough

(Clever) randomized redⁿ from INDEX

Alice: string $x \in \{0,1\}^n$

Bob: index $i \in [n]$

Alice & Bob generate (without communication) an input (x', y') to GAP-HAMMING

(High level) Alice & Bob generate sequence of random bits (x'_j, y'_j) slightly correlated if $x_i = 1$

anti-correlated if $x_i = 0$

Interpret first n public coins as random string r

Bob: $b = r_i$

Alice: $a = 1$ if $d_H(x,r) < \frac{n}{2}$

$a = 0$ if $d_H(x,r) > \frac{n}{2}$

Key point: a & b correlated!

Condition on $n-1$ bits of r other than i^{th} bit

2 cases

① $d_H(x_{-i}, r_{-i}) < \frac{n-1}{2}$

or $> \frac{n-1}{2}$

then a already determined (independent of $r_i = b$)

$Pr[a=b] = \frac{1}{2}$

② If $d_H(x_{-i}, r_{-i}) = \frac{n-1}{2}$

Happens with prob. $\binom{n-1}{\frac{n-1}{2}} \frac{1}{2^{n-1}}$ [redacted] $\frac{c'}{\sqrt{n}}$

$a = 1$ iff $x_i = r_i = b$

If $X_i = 1$ a, b agree

$X_i = 0$ a, b disagree

$$Pr[a=b] = \begin{cases} \frac{1}{2}\left(1 - \frac{c'}{\sqrt{n}}\right) + 1 \cdot \frac{c'}{\sqrt{n}} = \frac{1}{2}\left(1 + \frac{c'}{\sqrt{n}}\right) & \text{if } X_i = 1 \\ = \frac{1}{2}\left(1 - \frac{c'}{\sqrt{n}}\right) & \text{if } X_i = 0 \end{cases}$$

Repeat this process m times independently $m = qn$

q ↑ large const.

$$\mathbb{E}[d_H(x', y')] = m \cdot \frac{1}{2}\left(1 - \frac{c'}{\sqrt{n}}\right) = \frac{m}{2} - c'\sqrt{q}\sqrt{m} \quad \text{if } X_i = 1$$

$$\frac{m}{2} + c'\sqrt{q}\sqrt{m} \quad \text{if } X_i = 0$$

Using Chernoff bounds, for sufficiently const. C

w. prob. at least $\frac{8}{9}$

$$d_H(x', y') \cdot \begin{cases} < \frac{m}{2} - C\sqrt{m} & X_i = 1 \\ > \frac{m}{2} + C\sqrt{m} & X_i = 0 \end{cases}$$

When red^n is correct, Alice & Bob can solve INDEX(x, i)
by invoking protocol P for GAP-HAMMING on (x', y')
Comm. cost is that of P on inputs of size $m = \Theta(n)$

error $\leq$ error ~~of~~ red^n + error of protocol P

If there is randomized protocol for GAP-HAMMING
w. error $\frac{1}{3}$ and sublinear comm., then
there is randomized protocol for INDEX w. error $\frac{1}{3} + \frac{1}{9} = \frac{4}{9}$
& sublinear comm.

<u>Thm</u>: $\exists$ constant $c > 0$ s.t. any randomized
streaming algo that computes F_0 within $\left(1 + \frac{c}{\sqrt{n}}\right)$
w. prob $\geq \frac{2}{3}$ needs space $\Omega(n)$

Extend to larger ε?
Apply thm to inputs of size $m = \Theta\left(\frac{1}{\varepsilon^2}\right)$

$$\frac{c}{\sqrt{m}} = \varepsilon$$

computing F_0 within $(1+\varepsilon)$ needs space $\Omega\left(\frac{1}{\varepsilon^2}\right)$

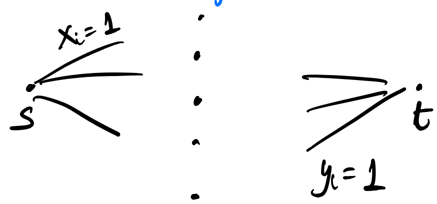
We have universe of size n, but we only use $m = \Theta\left(\frac{1}{\varepsilon^2}\right)$
elements $\Rightarrow$ computing F_0 within $(1+\varepsilon)$ needs
space $\Omega\left(\frac{1}{\varepsilon^2}\right)$

Lower bound for graph connectivity:
Given graph in streaming model, check if 2 vertices
s & t in same connected component or not
Needs $\Omega(n)$ bits of space!

Given $x, y \in \{0,1\}^n$ construct graph $G(V,E)$

Nodes : $V = [n] \cup \{s, t\}$

Edges : $\underbrace{\{(s,i) \mid \forall i : x_i = 1\}}_{\text{Alice holds these edges}} \cup \underbrace{\{(i,t) \mid \forall i : y_i = 1\}}_{\text{Bob holds these edges}}$

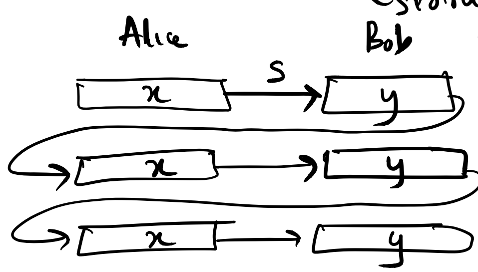


s, t connected iff x, y not disjoint

What about multiple passes ?

Thm: Randomized communication complexity of DISJ is $\Omega(n)$ (even with multiple rounds)

p-passes $\equiv (2p-1)$ messages between Alice & Bob of S bits each
$\uparrow$ storage



$\Omega\left(\frac{n}{p}\right)$ bits space needed for disjointness in p passes

Can have interesting dependence on #passes for some problems, e.g.
Computing median of n elements in p passes needs $\Theta(n^{1/p})$ space [Munro, Paterson]

Lower bound for Higher freq. moments:

Unique Disjointness for t-party communication

$UDISJ_n^t$

t players $A_1 \ldots A_t$

private string $x^{(i)} \in \{0,1\}^n$ (equivalently subset $S_i \subseteq [n]$)

for all $i \in [t-1]$, A_i sends message to A_{i+1} (in seq.)

A_t outputs result

$$UDISJ_n^t(x^{(1)} \ldots x^{(t)}) = \begin{cases} 0 & \text{if } DISJ(x^{(i)}, x^{(j)}) = 0 \ \forall i \neq j \in [t] \\ 1 & \text{if } k \in [n] \ \ x_k^{(i)} = 1 \ \forall i \in [t] \\ \text{don't care otherwise} \end{cases}$$

Randomized comm. complexity of $UDISJ_n^t$ is $\Omega(\frac{n}{t})$

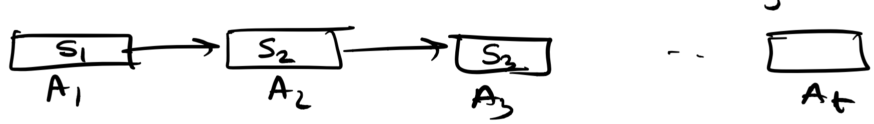
earlier [AMS] $\Omega(\frac{n}{t^3})$

<u>Thm</u>: For $p > 2$, any randomized streaming algo that approximates F_p within factor better than 2 needs space $\Omega(n^{1-\frac{2}{p}})$

<u>Pf</u>: On input $x^{(1)} \ldots x^{(t)}$ for $UDISJ_n^t$

run streaming algo for approximating $F_p(x^{(1)} \ldots x^{(t)})$

indices j where $x_j^{(1)} = 1$



$t = (4n)^{1/p}$

If $x^{(1)} \ldots x^{(t)}$ disjoint $F_p(S_1 \ldots S_t) \leq n$

If $x^{(1)} \ldots x^{(t)}$ intersect, then

$$F_p(S_1 \ldots S_t) \geq f_k^p = t^p = 4n$$

Any approximation within factor 2 can distinguish between 2 cases and solve $UDISJ_n^t$ with at most $(t-1)S$ bits

$$(t-1)S \text{ is } \Omega\left(\frac{n}{t}\right)$$

$$S \text{ is } \Omega\left(\frac{n}{t^2}\right) = \Omega\left(\frac{n}{n^{2/p}}\right) = \Omega\left(n^{1-\frac{2}{p}}\right)$$