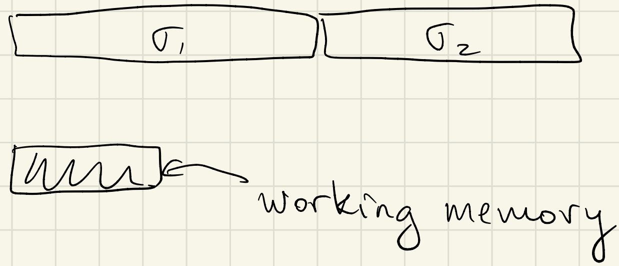
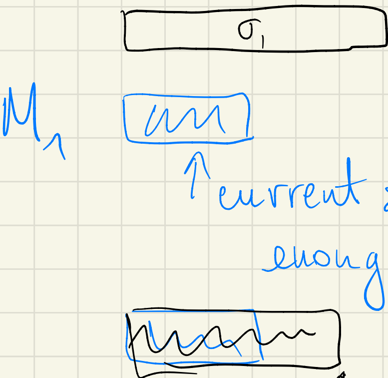
Today & next time: Data streaming lower bounds from communication complexity

Data stream

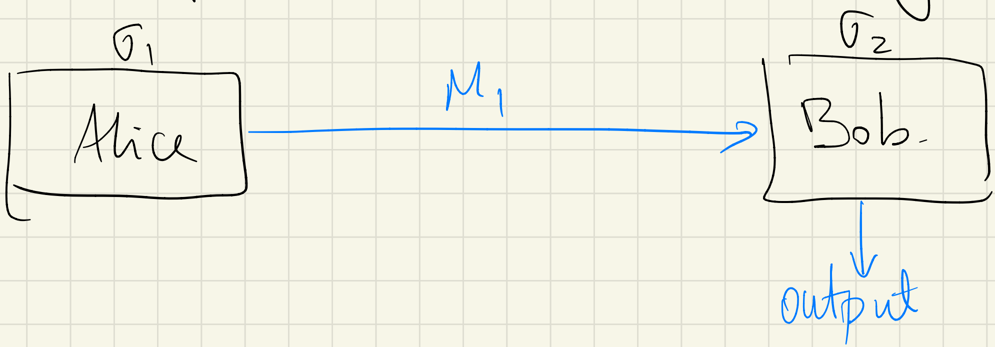




current state of working memory enough to start processing σ_2

final output / state

View problem as a communication game.



Alice & Bob each have an input
σ_1, σ_2 respectively
& want to compute $f(\sigma_1, \sigma_2)$
with min # of bits comunicated.

Intuitively, if we have a streaming alg
with space s bits, $\exists$ a communic.
protocol in which Alice sends the
s bits & Bob on s & σ_2 outputs
the answer.

So if can show a reduction from
a communication problem to a streaming
problem, & have a lower bound for
communic. problem then have lower bound
for the amount of space of streaming
problem.

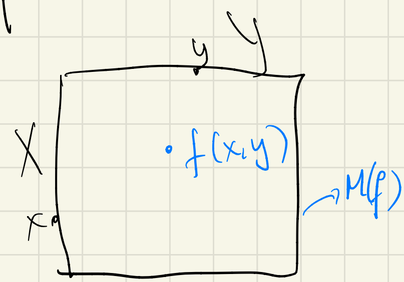
Some basic communication complexity:

2 (or more parties) cooperate to compute f on their inputs.

Say $f: X \times Y \to \{0,1\}$

Alice gets input $x \in X$

Bob $y \in Y$



Goal: output $f(x,y)$ with min amount of bits exchanged.

Alice $\longrightarrow$ Bob

x $\longleftarrow$ y

$\longrightarrow$

$\vdots$

$\longrightarrow$

$f(x,y)$

(communic is expensive)

Time / space don't matter.

Trivial: $\min\{\log|X|, \log|Y|\} + 1$

(send entire msg)

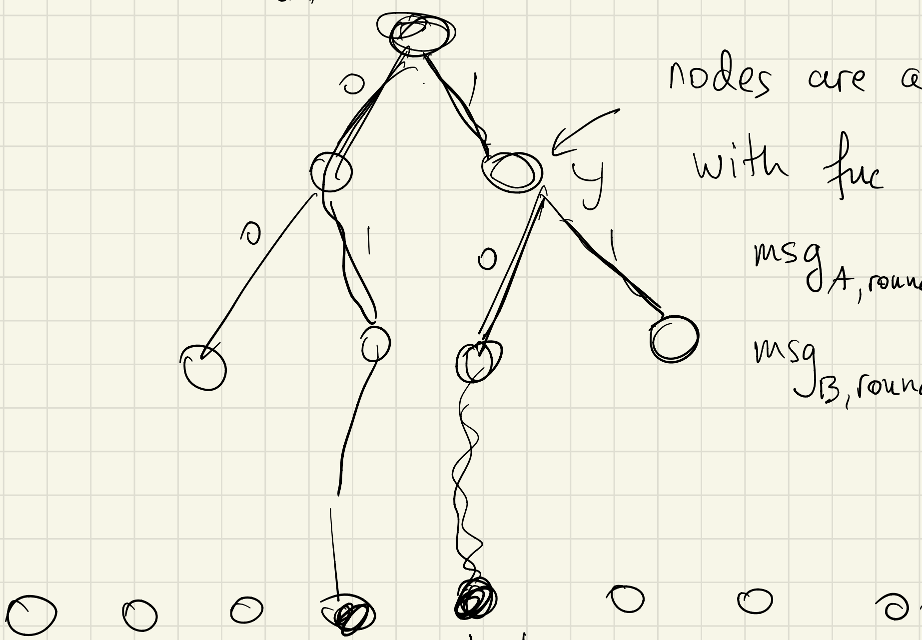
Want: sublinear in $\uparrow$

Parties follow a protocol.

↓

represented by a binary decision tree.

$msg_{A,0}: X \to \{0,1\}$.



nodes are associated with fuc

$msg_{A,\text{round } i}: X \times \{0,1\} \to \{0,1\}$

$msg_{B,\text{round } i}: Y \times \{0,1\} \to \{0,1\}$

output

$0/1 = f(x,y)$

↑ leaves

$\underline{out^{\pi}(x,y)} \to$ outcome of protocol π on x, y.

Def: Π computes f if $\underline{out^{\Pi}(x,y)} = \underline{f(x,y)}$ $\forall x, y \in X \times Y$

Cost of protocol Π : worst case # of bits communicated over all inputs.

Types of protocols :

- deterministic : $D(f)$: min cost of a protocol that is correct on every input.

- Randomized :
  - private coin - players have private rand bits
  - public coin - players use a public string of random bits

↑ stronger version
(usual notion of randomized protocol here

Def: Π computes f with error δ if

$\forall (x,y) \in X \times Y$

$$\Pr_R\left[\mathrm{out}^{\Pi}(x,y;R) \neq f(x,y)\right] \leq \delta.$$

Def: $R_\delta(f)$ is the min cost of a δ-error protocol for f.

Claim: $\forall \quad 0 < \delta < 1/3$

$$R_\delta(f) = \Theta\left(R_{1/3}(f)\right)$$

Def: $\vec{D}(f)$, $\vec{R}(f)$ — one way protocols. Alice only talks.

$D^k(f)$; $R^k(f)$ k-round communicat

Lecture 19

Basic Observations:

1. $R_{1/3}(f) \leq D(f) \leq \min \log|X|, \log|Y|$

2. $D(f) \leq D^{k}(f)$, $\forall k > 0$

3. $R(f) \leq R^{k}(f) \leq R^{\ell}(f)$
   $\forall \quad k \geq \ell$.

Specific com. problems:

- Alice, Bob

$$EQ_N(x, y) = \begin{cases} 1 & x = y \\ 0 & \text{ow} \end{cases}$$

$x, y \in \{0,1\}^N$

$$D(EQ_N) \geq N$$

$$R^{\rightarrow \text{priv}} = O(\log N)$$

- INDEX

$$\underset{\in\{0,1\}^N}{\overset{\text{Alice}}{\text{IND}_N(x}},\ \underset{\in[N]}{\overset{\text{Bob}}{i}}) = x_i$$

one way

$$\vec{D}(\text{IND}_N) \geq N$$

$$R^{\rightarrow}(\text{IND}_N) \geq \Omega(N) ;\ R(\text{IND}_N) = \Omega(N)$$

$$D^2(\text{IDX}) \leq \log N$$

- $\text{DISJ}_N(\overset{\text{Alice}}{x}, \overset{\text{Bob}}{y}) = \begin{cases} 1 & \text{if } x \cap y = \emptyset \\ 0 & \text{ow} \end{cases}$

$x, y \in \{0,1\}^N$; char vects of sets

$$R(\text{DISJ}_N) \geq \Omega(N)$$

Claim: $\vec{D}(\mathrm{IDX}_N) \geq N$

Pf: Alice $\xrightarrow{\quad\quad\quad}$ Bob.

x $\quad$ msg(x) $\quad$ j

? x_j

Deterministic: Bob, no matter what $j \in N$ he has, must recover x_j from $m(x)$, exactly.

Alice has 2^N many msgs. x

Bob needs to recover every bit of x from $m(x)$. So Bob can recover x from msg(x)

$$x \xrightarrow{\quad} msg(x) \xrightarrow{Bob(msg(x))} x$$

So both mappings are 1-1 & onto.

How many msg can be sent of length $< N$

$$2^{N-1} + 2^{N-2} + \cdots + 2 + 1 = 2^N - 1 < 2^N \Leftarrow$$

So if $\exists$ protocal of length $\leq N-1$

fuc. msg wouldn't be a bijection.

(contradiction)

Claim : $\vec{R}^{priv}(EQ_N) = O(\log N)$

$$\begin{array}{ccc} \text{Alice} & & \text{Bob} \\ x & \xrightarrow{\text{msg}(x,r)} & y \\ r & & \end{array}$$

$EQ_N(x,y) = \{x = y ?\}$

Pf: Alice, Bob can view their inputs as polys :

$x \longrightarrow P_x(t) = \sum_{i \in [N]} x_i t^i$

$y \longrightarrow P_y(t) = \sum y_i t^i$

$P_x, P_y \in \mathbb{F}[t] \leftarrow$ they agree on beforehand.

Alice picks rand $r \leftarrow \mathbb{F}$

send $r, P_x(r)$ to Bob.

Bob compares $P_y(r)$ with $P_x(r)$

If "=" output 1

else outp 0

bits communicated

$$|r| + |P_x(r)|$$

$$\parallel$$

$$\log|\mathbb{F}| + \log|\mathbb{F}| = 2\log|\mathbb{F}|$$

$$\uparrow$$

Pr[error = ?].

$$O(\log N)$$

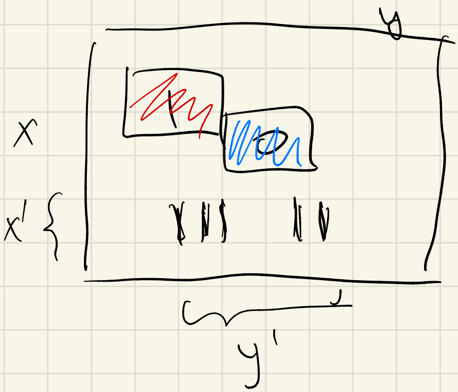
$x \neq y \quad \Pr_r [P_x(r) = P_y(r)] \leq \frac{N}{\|F\|}$

$|F| = 20N$

or $|F| = N^2$

Communication Matrix

Detour into Log Rank conjecture



Def: $C \subseteq X \times Y$ is a combinatorial rectangle if $C = X' \times Y'$

$X' \subseteq X$, $Y' \subseteq Y$

Def: C is monochromatic wrt f if

$f(x,y) = f(x',y') \quad \forall (x,y), (x',y') \in C.$

Def: Let $C^D(f)$ be min # of monochrom. rectangles that can partition M_f.

Thm: $D(f) \geq \log(C^D(f))$.

Thm: $f: X \times Y \to \{0,1\}$

$$D(f) \geq \log(\operatorname{rank}(M_f))$$

Claim: Let ℓ be a leaf of the decision tree protocol π

Let $C_\ell = \{(x,y)$ whose transcript ends at $\ell\}$

Then C_ℓ is a combinatorial rectangle:

$$\left.\begin{array}{l} \forall (x,y) \in C_\ell \\ (x',y') \in C_\ell \end{array}\right\} \begin{array}{l} (x',y) \\ (x,y') \end{array} \in C_\ell.$$

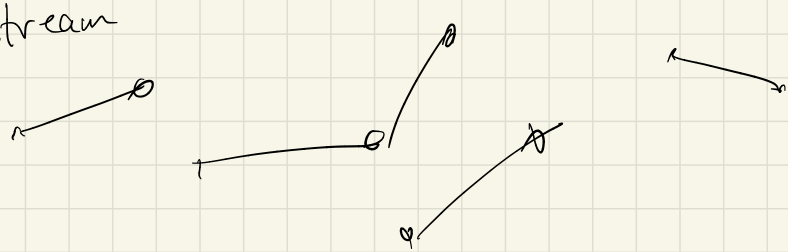
Thm • $D(f)$ NP hard to compute

• Big conj: $D(f) = O(\text{poly} \log \operatorname{rank} M_f)$

Lovett: $D(f) = O(\sqrt{\operatorname{rank} M_f} \cdot \log \operatorname{rank} M_f)$

Thm: Any single-pass det algorithm for deciding if a graph on n vertices has a perfect matching needs $\Omega(n^2)$ bits of memory.

Pf. stream



$D(IDX_N(X, i)) = \Omega(N)$ $\leftarrow$

$R(IDX_N(x, i)) = \Omega(N)$

Given

$$X = \begin{array}{cc} & j \\ i & \begin{bmatrix} 0 & 0 & 1 \\ & 0 & \\ \cdots & 1 & \cdots \\ & & \end{bmatrix} n \\ & n \end{array} \qquad M_A$$

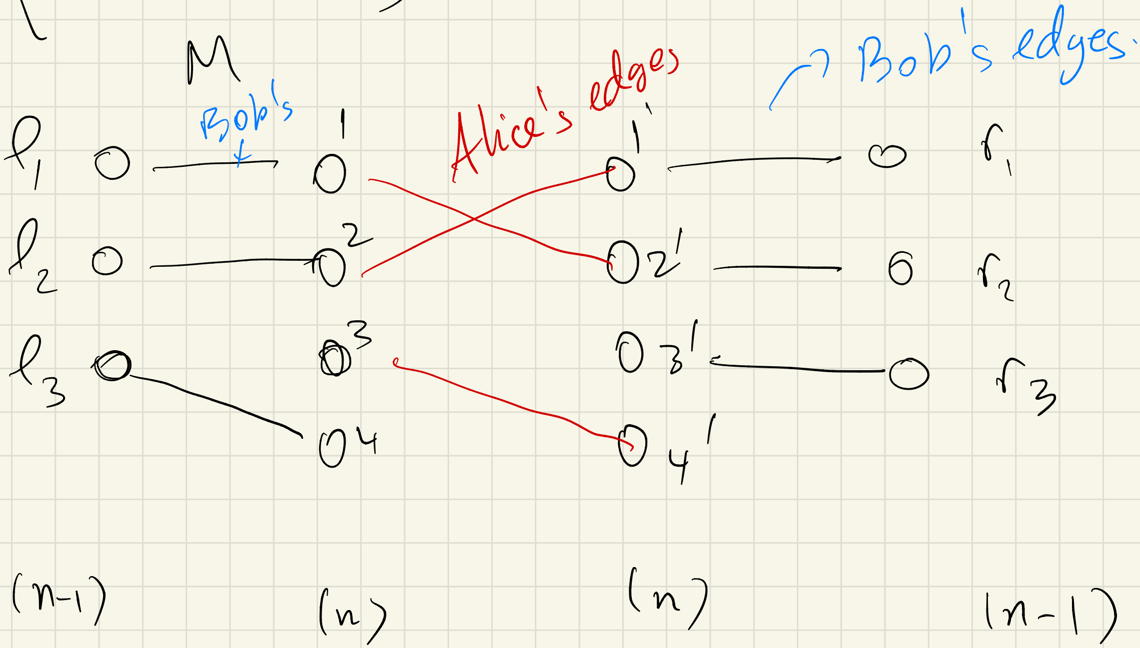
Alice gets A, Bob gets (i, j)

$\Omega(IDX_n(x, i)) = \Omega(n^2)$

Eg.

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad (i,j) = (3,4)$$

M



G

<u>Claim:</u> $M(i,j) = 1$ iff G has a perfect matching. ✓

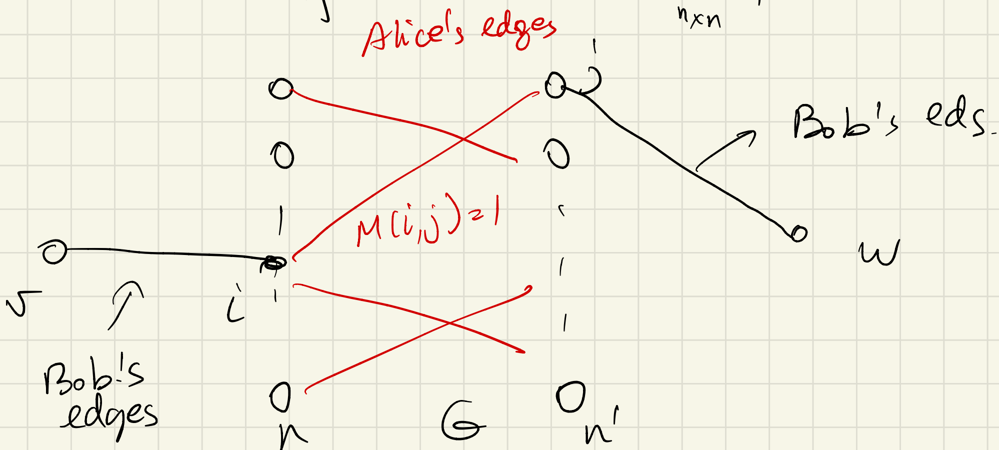
So, stream Alice's before Bob's edges.

Can come up with protocol in which Alice sends memory content to Bob as soon as her stream of edges was processed.

So, if can solve streaming problem in $o(n^2)$ space then can solve IND_{n^2} problem in $o(n^2)$ bits communicated, a contradiction.

Thm: Any streaming deterministic one-pass approx alg for ShortestPath(v,w) with approx factor $\leq 5/3$ uses $\Omega(n^2)$ space.

Pf: Reduce from $IDX(M_{n\times n}, i, j)$.

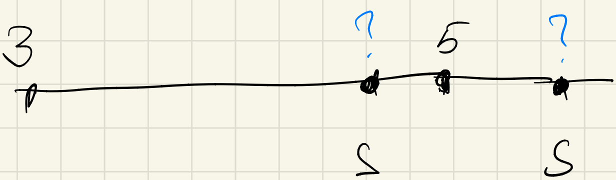


Use the streaming alg with $o(n^2)$ space to compute $5/3$ approx for $d(v,w)$, & output S. Stream Alice's edges before Bob's. In the CC protocol Alice sends her memory content to Bob.

If $M(i,j) = 1$ then $d(v,w) = 3$

ow $d(v,w) \geq 5$

(could be ∞)



Use output S to decide whether edge (i,j) exists.

If $S < 5$ output $IDX(M,i,j) = 1$

ow output $IDX(M,i,j) = 0$

(Indeed, if $OPT = 3 \Rightarrow S \leq 5$

ow. $S > OPT = 5$)

So, can solve IND_{n^2} in $O(n^2)$. Contradiction.