Information Theory in Computer Science

Jaikumar Radhakrishnan

School of Technology and Computer Science
Tata Institute of Fundamental Research
MUMBAI

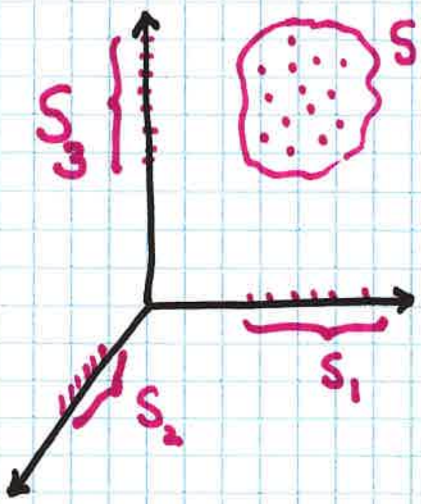
CAALM 2019, 21 Jan 2019

CMI, Chennai

PLAN

- A combinatorial example via information theory
- Circuit depth and communication complexity
- A communication lower bound via information theory

n points in $\mathbb{R}^3$



- $|S| = n$
- $|S_1| = n_1,\ |S_2| = n_2,\ |S_3| = n_3$

Then,

$$S \subseteq S_1 \times S_2 \times S_3$$

$$\Downarrow$$

$$|S| \leq |S_1| \cdot |S_2| \cdot |S_3|$$

i.e., $\quad n \leq n_1 n_2 n_3$

It is information that counts

- Pick a point $P \in S$ at random

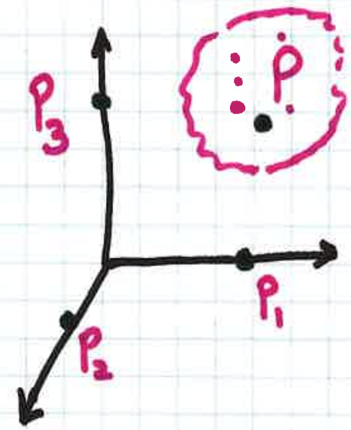
  Let P_1, P_2, P_3 be its projections on the axes.

- P has $\log n$ bits of information

  P_1 has $\log n_1$ bits of information

  P_2 has $\log n_2$ bits of information

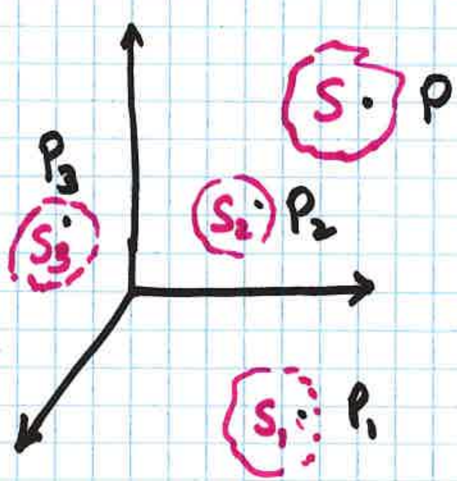
  P_3 has $\log n_3$ bits of information



- P can be recovered from (P_1, P_2, P_3).

  So,
  $$\log n_1 + \log n_2 + \log n_3 \geq \log n$$
  $$n_1 n_2 n_3 \geq n$$

The Loomis Whitney Inequality



$$|S| = n$$

$$|S_1| = n_1, \quad |S_2| = n_2, \quad |S_3| = n_3$$

Then,

$$n_1 n_2 n_3 \geqslant n^2$$

IDEA: Every piece of information about P is available from two sources. So,

$$\log n_1 + \log n_2 + \log n_3 \underset{?}{\geqslant} 2 \log n$$

Information

X: a random variable $\equiv \begin{pmatrix} x_1 & x_2 & \cdots & x_k \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}$

Entropy of X: measures the uncertainty in X

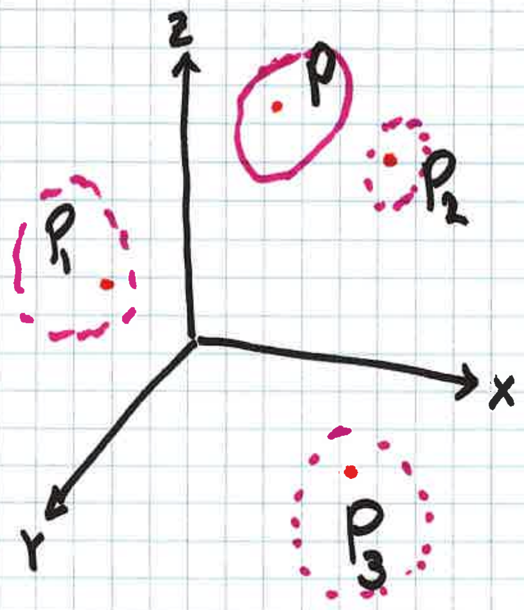
$$H[X] = -\sum_{i=1}^{k} p_i \log_2 p_i$$

Has many useful properties.

Properties of Entropy

- $H[X] = \log k$ if X is uniformly distributed on a set of size k.
- $H[(X,Y)] = H[X] + H[Y|X]$
- $H[Y|X] \leqslant H[Y]$
- $H[X] \leqslant \log k$ if X ranges over a set of size k

An Entropy Proof of Loomis-Whitney



- Pick P uniformly at random from S

$$P \equiv (X, Y, Z)$$

- $$\log n = H[P] = H[X] + H[Y|X] + H[Z|XY]$$
$$\log n_1 \geq H[P_1] = H[Y] + H[Z|Y]$$
$$\log n_2 \geq H[P_2] = H[X] + H[Z|X]$$
$$\log n_3 \geq H[P_3] = H[X] + H[Y|X]$$

$$\Rightarrow \quad 2\log n \leq \log n_1 + \log n_2 + \log n_3$$

Mutual Information

X, Y : Random variables

$$I[X:Y] = H[X] + H[Y] - H[XY]$$
$$= H[X] - H[X|Y]$$
$$= H[Y] - H[Y|X]$$

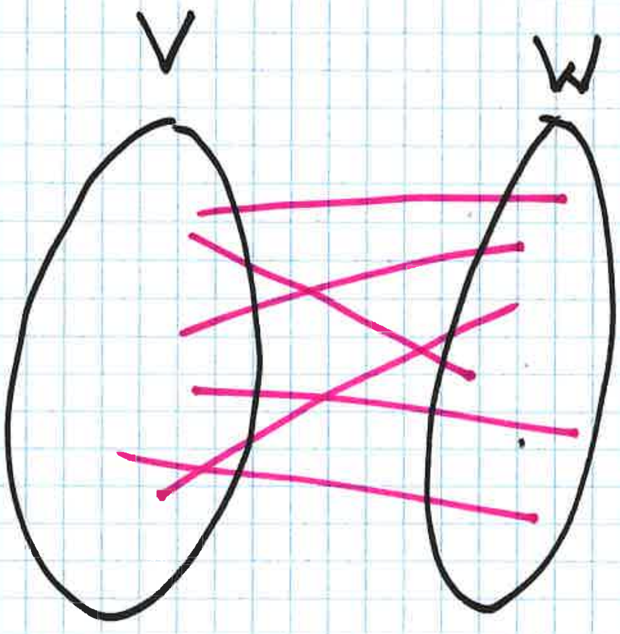
Key property

$X_1, X_2, \ldots, X_n$: independent

$$\Downarrow$$

$$H[T] \geq I[X_1 X_2 \ldots X_n : T] \geq \sum_{i=1}^{n} I[X_i : T]$$

BIPARTITE MATCHING

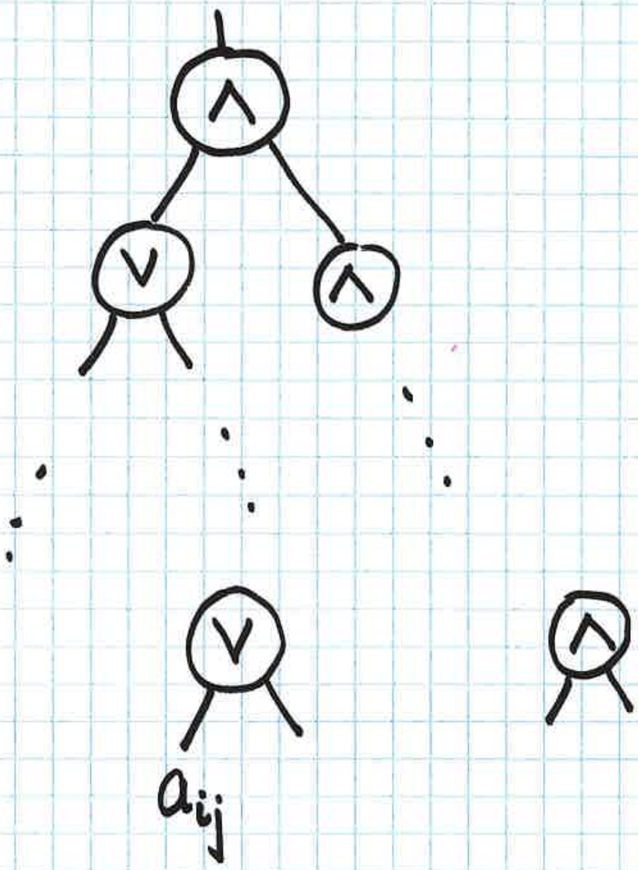


$|V| = |W| = n$

Input: ($n \times n$ adjacency matrix)

Task: Determine if the graph has a perfect matching

MONOTONE BOOLEAN FORMULAS



Theorem (Raz and Wigderson)

$$\text{depth} = \Omega(n).$$

IDEA: Formulas yield protocols.

Formulas Yield Protocols

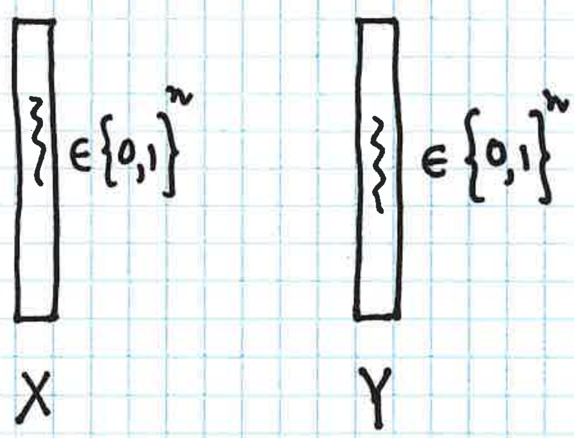
Alice	Bob
A bipartite graph with k disjoint edges	A set S of $k-1$ vertices

Task: Determine an edge in Alice's graph that is not incident on S.

(Formula of depth d) $\Rightarrow$ (Protocol with communication d.)

The Set Intersection Problem

Alice **Bob**



GOAL: $\bigvee_{i=1}^{n} (x_i \wedge y_i)$

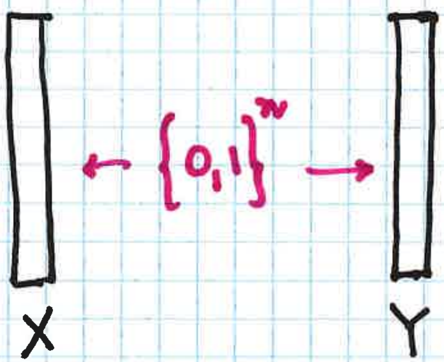
Determine if X and Y have a common 1.

- Alice and Bob exchange messages. $\rightsquigarrow$ protocol π
- They can toss coins.
- We allow errors.
- $\forall\, X, Y$

  $\Pr[\, \pi(X,Y) \text{ is correct}\,] \geq \frac{3}{4}$.
- Alice and Bob wish to minimize the number of bits exchanged.

The zero error case

Alice Bob



Goal: $\bigvee_{i=1}^{n} (X_i \wedge Y_i)$

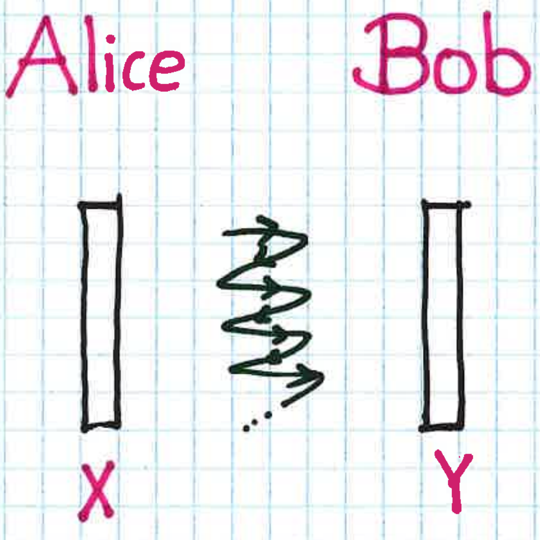
- Special Inputs

$\{(X, \neg X) : X \in \{0,1\}^n\}$

- Let T_X be the transcript when the protocol is run on input $(X, \neg X)$
- Different X's give different transcripts.

$$n = I[X : T_X] \leq \cancel{H[X]}\ H[T_X] \leq E[|T_X|]$$

The Set Intersection Problem
(Error $\leq \frac{1}{4}$)



Goal: $\bigvee_{i=1}^{n} (X_i \wedge Y_i)$

THEOREM

Alice and Bob must exchange $\Omega(n)$ bits in the worst case.

- Kalyanasundaram and Schnitger (1987)
- Razborov (1990)
- Bar-Yossef, Jayram, Kumar, and Sivakumar (2004)

PROOF IDEA

Assume communication $m \ll n$.

Feed random inputs to the protocol and observe the transcript.

Find a coordinate that both Alice and Bob neglect.

Argue that then the protocol makes errors with probability $\approx 1/2$.

An n-bit problem to a one-bit problem

RANDOM INPUTS

- Pick a pattern $T \in \{A, B\}^n$.
  There are 2^n such patterns.

- With each pattern T associate a distribution on inputs: $\mathcal{D}_T$ ← 2^n such distributions on $\{0,1\}^n \times \{0,1\}^n$

- $T_i = A \Rightarrow$

$$X_i = \begin{cases} 0 & \text{with prob. } 1/2 \\ 1 & \text{with prob } 1/2 \end{cases}$$

$$Y_i = 0$$

$T_i = B \Rightarrow$

$$X_i = 0$$

$$Y_i = \begin{cases} 0 & \text{with prob. } 1/2 \\ 1 & \text{with prob. } 1/2 \end{cases}$$

(independently for each coordinate)

The neglected coordinate

- Fix τ and consider $(X,Y) \sim D_\tau$
- $$I_\tau[X:T] \leqslant H[T] \leqslant m$$
$$\Downarrow$$
$$\sum_i I_\tau[X_i:T] \leqslant m$$
- Similarly, $\sum_i I_\tau[Y_i:T] \leqslant m$

$$\sum_i \underbrace{\left(I_\tau[X_i:T] + I_\tau[Y_i:T] \right)}_{\text{Attention paid to coordinate } i} \leqslant 2m$$

The neglected coordinate

$$\sum_i \left(I_\tau[X_i : T] + I_\tau[Y_i : T] \right) \leqslant 2m$$

Average over all $\tau \in \{A,B\}^n$

$$\sum_i \mathop{E}_\tau\left[\left(I_\tau[X_i : T] + I_\tau[Y_i : T] \right)\right] \leqslant 2m$$

$$\Downarrow$$

$\exists$ coordinate i^* such that

$$\mathop{E}_\tau\left[I_\tau[X_{i^*} : T] + I_\tau[Y_{i^*} : T] \right] \leqslant \frac{2m}{n}$$

FIX SUCH A NEGLECTED COORDINATE i^*

The one-bit set intersection problem
aka
AND

- Fix the pattern outside coordinate i^*.
- Outside i^* either Alice's input bit is random or Bob's input bit is random.
- Neither Alice nor Bob reveal much about their random input when the other party has 0. $\left(\frac{4m}{n}\right.$
- Yet the protocol computes the AND of their inputs with high probability.

THIS IS IMPOSSIBLE! (Why?)

Alice and Bob,
Went at it,
Dishing it out,
Bit by bit,
Neither of them,
Would take a hit,
So, $\Omega(n)$ it was,
By the time they quit!

CONCLUSION

THEOREM:

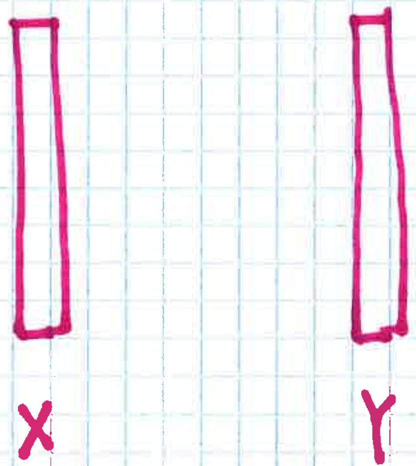
Alice and Bob must exchange $\Omega(n)$ bits in any protocol that is correct on all inputs with probability at least $3/4$.

(The lower bound holds even when Alice and Bob share a long sequence of random bits that is generated independent of the input (x, y).)

THE REDUCTION

SET INTERSECTION $\rightsquigarrow$ MATCHING VS COVER

Randomized



n edges, $n-1$ vertices

Successful ideas in science are those that are pervasive and invasive, are invitingly elegant and methodical, are open to extensions and variants, and answer an objective necessity, and capture a widespread but diffuse sense of dissatisfaction in a scientific community.

Christos Papadimitriou (1995)
in connection with the P vs. NP problem

David Galvin: Three tutorial lectures on entropy and counting

Ehud Friedgut: Hypergraphs, Entropy and Inequalities

Arkadev Chattopadhyay and Toni Pitassi: The story of set disjointness.

Anup Rao: Lecture notes