

# Distributed Statistical Estimation of Matrix Products with Applications

David Woodruff

CMU

Qin Zhang

IUB

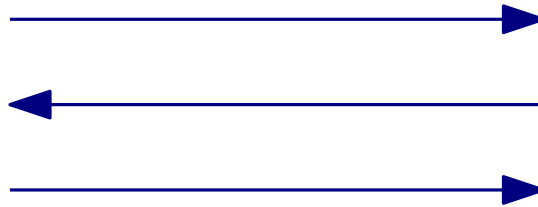
PODS 2018

June, 2018

# The Distributed Computation Model



$p$ -norms,  
heavy-hitters, ...



$$A \in \{0, 1\}^{m \times n}$$

$$B \in \{0, 1\}^{n \times m}$$

Alice and Bob want to compute some function on  $C = A \times B$  with the minimum amount of communication and number of rounds

Communication: sum of message lengths  
(maximized over all choices of  $A$ ,  $B$ , and randomness)

The protocol can fail with prob. 0.01 (over its randomness)

# Statistics of Matrix Products: $p$ -Norms

- Alice holds  $A \in \{0, 1\}^{m \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times m}$
- Let  $C = A \cdot B$ . Alice and Bob want to approximate

$$\|C\|_p = \left( \sum_{i,j \in [n]} |C_{i,j}|^p \right)^{1/p}$$

# Statistics of Matrix Products: $p$ -Norms

- Alice holds  $A \in \{0, 1\}^{m \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times m}$

- Let  $C = A \cdot B$ . Alice and Bob want to approximate

$$\|C\|_p = \left( \sum_{i,j \in [n]} |C_{i,j}|^p \right)^{1/p}$$

- $p = 0$ : number of non-zero entries of  $C$

$\Rightarrow$  size of **set-intersection join**

$i$ -th row of  $A$  as set  $A_i$ ,  $j$ -th column of  $B$  as set  $B_j$ ,  
compute  $\#(i, j)$  s.t.  $A_i \cap B_j \neq \emptyset$

# Statistics of Matrix Products: $p$ -Norms

- Alice holds  $A \in \{0, 1\}^{m \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times m}$

- Let  $C = A \cdot B$ . Alice and Bob want to approximate

$$\|C\|_p = \left( \sum_{i,j \in [n]} |C_{i,j}|^p \right)^{1/p}$$

- $p = 0$ : number of non-zero entries of  $C$

⇒ size of **set-intersection join**

$i$ -th row of  $A$  as set  $A_i$ ,  $j$ -th column of  $B$  as set  $B_j$ ,  
compute  $\#(i, j)$  s.t.  $A_i \cap B_j \neq \emptyset$

- $p = 1$ : sum of entries of  $C$

⇒ size of corresponding **natural join**

compute  $\#(i, k, j)$  s.t.  $k \in A_i \cap B_j$

# Statistics of Matrix Products: $p$ -Norms

- Alice holds  $A \in \{0, 1\}^{m \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times m}$

- Let  $C = A \cdot B$ . Alice and Bob want to approximate

$$\|C\|_p = \left( \sum_{i,j \in [n]} |C_{i,j}|^p \right)^{1/p}$$

- $p = 0$ : number of non-zero entries of  $C$

⇒ size of **set-intersection join**

$i$ -th row of  $A$  as set  $A_i$ ,  $j$ -th column of  $B$  as set  $B_j$ ,  
compute  $\#(i, j)$  s.t.  $A_i \cap B_j \neq \emptyset$

- $p = 1$ : sum of entries of  $C$

⇒ size of corresponding **natural join**

compute  $\#(i, k, j)$  s.t.  $k \in A_i \cap B_j$

- $p = \infty$ : maximum entry of  $C$

⇒ most “similar”  $(A_i, B_j)$  pair

# Application of set-intersection join

Applicant	Skills	Skills	Opening
$A_1$	$S_1, S_4, S_9, S_{13}$	$S_2, S_3, S_4$	$B_1$
$A_2$	$S_2, S_9, S_{10}$	$S_3, S_4, S_9, S_{11}$	$B_2$
$A_m$	$S_6, S_7, S_8, S_{15}$	$S_4, S_8$	$B_m$

Find all candidate (Applicant, Opening) pairs

# $\|C\|_1$ corresponds to natural join

$U$	$V$
1	2
1	4
2	1
2	2
2	3
...	



$V$	$W$
1	1
2	1
4	1
1	2
3	2
...	

$V$

$U$	$V$	$W$	$X$
0	1	0	1
1	1	1	0
...			



$W$

$V$	$W$	
1	1	
1	0	
0	1	
1	0	
...		

$A$

$B$



# $\|C\|_1$ corresponds to natural join

$U$	$V$
1	2
1	4
2	1
2	2
2	3
...	



$V$	$W$
1	1
2	1
4	1
1	2
3	2
...	

$\Rightarrow (2, 1, 2)$

$V$

$U$	$V$	$W$	$X$
0	1	0	1
1	1	1	0
...			



$W$

1	1	
1	0	...
0	1	
1	0	



	+1	

$A$

$B$

$C$

# Statistics of Matrix Products: Heavy Hitters

- Alice holds  $A \in \{0, 1\}^{m \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times m}$
- Let  $C = A \cdot B$ , and let

$$\text{HH}_{\phi}^p(C) = \{(i, j) \mid C_{i,j} \geq \phi \|C\|_p\}$$

# Statistics of Matrix Products: Heavy Hitters

- Alice holds  $A \in \{0, 1\}^{m \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times m}$
- Let  $C = A \cdot B$ , and let

$$\text{HH}_{\phi}^p(C) = \{(i, j) \mid C_{i,j} \geq \phi \|C\|_p\}$$

- $\ell_p$ - $(\phi, \epsilon)$ -heavy-hitter ( $0 < \epsilon \leq \phi \leq 1$ ): output a set  $S \subseteq \{(i, j) \mid i, j \in [m]\}$  such that

$$\text{HH}_{\phi}^p(C) \subseteq S \subseteq \text{HH}_{\phi-\epsilon}^p(C)$$

Pairs  $(A_i, B_j)$  that are similar; for **similarity join**

# Statistics of Matrix Products: Heavy Hitters

- Alice holds  $A \in \{0, 1\}^{m \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times m}$
- Let  $C = A \cdot B$ , and let

$$\text{HH}_{\phi}^p(C) = \{(i, j) \mid C_{i,j} \geq \phi \|C\|_p\}$$

- $\ell_p$ - $(\phi, \epsilon)$ -heavy-hitter ( $0 < \epsilon \leq \phi \leq 1$ ): output a set  $S \subseteq \{(i, j) \mid i, j \in [m]\}$  such that

$$\text{HH}_{\phi}^p(C) \subseteq S \subseteq \text{HH}_{\phi-\epsilon}^p(C)$$

Pairs  $(A_i, B_j)$  that are similar; for **similarity join**

Both the  $p$ -norm estimation and heavy-hitters are well-studied in the data stream literature, but not as much in the distributed model (model next slide)

# Our Main Results – $\ell_p$ ( $p \in [0, 2]$ )

For simplicity, assume  $m = n$

- For any  $p \in [0, 2]$ , a 2-round  $\tilde{O}(n/\epsilon)$ -bit algorithm that approximates  $\|AB\|_p$  within a  $(1 + \epsilon)$  factor

# Our Main Results – $\ell_p$ ( $p \in [0, 2]$ )

For simplicity, assume  $m = n$

- For any  $p \in [0, 2]$ , a 2-round  $\tilde{O}(n/\epsilon)$ -bit algorithm that approximates  $\|AB\|_p$  within a  $(1 + \epsilon)$  factor
  - For  $p = 0$ , this improves the previous result  $\tilde{O}(n/\epsilon^2)$  (Van Gucht et al., PODS'15)
  - Same paper shows a lower bound of  $\Omega(n/\epsilon^{2/3})$ .

# Our Main Results – $\ell_p$ ( $p \in [0, 2]$ )

For simplicity, assume  $m = n$

- For any  $p \in [0, 2]$ , a 2-round  $\tilde{O}(n/\epsilon)$ -bit algorithm that approximates  $\|AB\|_p$  within a  $(1 + \epsilon)$  factor
    - For  $p = 0$ , this improves the previous result  $\tilde{O}(n/\epsilon^2)$  (Van Gucht et al., PODS'15)
    - Same paper shows a lower bound of  $\Omega(n/\epsilon^{2/3})$ .
- If we restrict the communication to be one-way, then we have a lower bound  $\Omega(n/\epsilon^2)$ .

# Our Main Results – $\ell_\infty$

- $O(1)$ -round algorithms that approximate  $\|AB\|_\infty$ 
  - within a factor of  $(2 + \epsilon)$  use  $\tilde{O}(n^{1.5}/\epsilon)$  bits
  - within a factor of  $\kappa$  use  $\tilde{O}(n^{1.5}/\kappa)$  bits



# Our Main Results – $\ell_\infty$

- $O(1)$ -round algorithms that approximate  $\|AB\|_\infty$ 
  - within a factor of  $(2 + \epsilon)$  use  $\tilde{O}(n^{1.5}/\epsilon)$  bits
  - within a factor of  $\kappa$  use  $\tilde{O}(n^{1.5}/\kappa)$  bits
- Any algorithm (regardless of the #rounds used) that approximates  $\|AB\|_\infty$  within a factor of
  - within a factor of 2 needs  $\Omega(n^2)$  bits
  - within a factor of  $\kappa \geq 4$  needs  $\Omega(n^{1.5}/\kappa)$  bits

# Our Main Results – $\ell_\infty$

- $O(1)$ -round algorithms that approximate  $\|AB\|_\infty$ 
  - within a factor of  $(2 + \epsilon)$  use  $\tilde{O}(n^{1.5}/\epsilon)$  bits
  - within a factor of  $\kappa$  use  $\tilde{O}(n^{1.5}/\kappa)$  bits
- Any algorithm (regardless of the #rounds used) that approximates  $\|AB\|_\infty$  within a factor of
  - within a factor of 2 needs  $\Omega(n^2)$  bits
  - within a factor of  $\kappa \geq 4$  needs  $\Omega(n^{1.5}/\kappa)$  bits
- The above results hold for **binary** matrices  $A$  and  $B$ . For **general** matrices  $A, B \in \Sigma^{n \times m}$ , the bound is  $\tilde{\Theta}(n^2/\kappa^2)$  bits ( $O(1)$ -round for UB, any round for LB)

# Our Main Results – Heavy Hitters

- For **binary** matrices  $A$  and  $B$ , for any  $p \in (0, 2]$ , an  $O(1)$ -round  $\tilde{O}(n + \frac{\phi}{\epsilon^2})$ -bit algorithm that computes  $\ell_p$ - $(\phi, \epsilon)$ -heavy-hitters

# Our Main Results – Heavy Hitters

- For **binary** matrices  $A$  and  $B$ , for any  $p \in (0, 2]$ , an  $O(1)$ -round  $\tilde{O}(n + \frac{\phi}{\epsilon^2})$ -bit algorithm that computes  $\ell_p$ - $(\phi, \epsilon)$ -heavy-hitters
- For **general** matrices  $A$  and  $B$ , for any  $p \in (0, 2]$ , we obtain  $O(1)$ -round  $\tilde{O}(\frac{\sqrt{\phi}}{\epsilon} \cdot n)$  bits algorithms

# Our Main Results – Heavy Hitters

- For **binary** matrices  $A$  and  $B$ , for any  $p \in (0, 2]$ , an  $O(1)$ -round  $\tilde{O}(n + \frac{\phi}{\epsilon^2})$ -bit algorithm that computes  $\ell_p$ - $(\phi, \epsilon)$ -heavy-hitters
- For **general** matrices  $A$  and  $B$ , for any  $p \in (0, 2]$ , we obtain  $O(1)$ -round  $\tilde{O}(\frac{\sqrt{\phi}}{\epsilon} \cdot n)$  bits algorithms

All of our results above can be easily extended to rectangular matrices where  $A \in \Sigma^{m \times n}$  and  $B \in \Sigma^{n \times m}$

# Previous Results

- Most relevant (Van Gucht et al., PODS'15). It studies *set-intersection/disjointness/equality/at-least- $T$  joins* in the 2-party communication model

The only overlap between Van Gucht et al. and this paper is the estimation of  $\|AB\|_0$  mentioned before

# Previous Results

- Most relevant (Van Gucht et al., PODS'15). It studies *set-intersection/disjointness/equality/at-least- $T$  joins* in the 2-party communication model

The only overlap between Van Gucht et al. and this paper is the estimation of  $\|AB\|_0$  mentioned before

- A number of recent works look at distributed linear algebra problems (Balcan et al. KDD'16; Boutsidis et al. STOC'16; Woodruff&Zhong, ICDE'16; etc.)

These works concern statistics estimation on  $C = A + B$ , compared with  $C = A \cdot B$  studied in this paper

# Previous Results

- Most relevant (Van Gucht et al., PODS'15). It studies *set-intersection/disjointness/equality/at-least- $T$  joins* in the 2-party communication model

The only overlap between Van Gucht et al. and this paper is the estimation of  $\|AB\|_0$  mentioned before

- A number of recent works look at distributed linear algebra problems (Balcan et al. KDD'16; Boutsidis et al. STOC'16; Woodruff&Zhong, ICDE'16; etc.)

These works concern statistics estimation on  $C = A + B$ , compared with  $C = A \cdot B$  studied in this paper

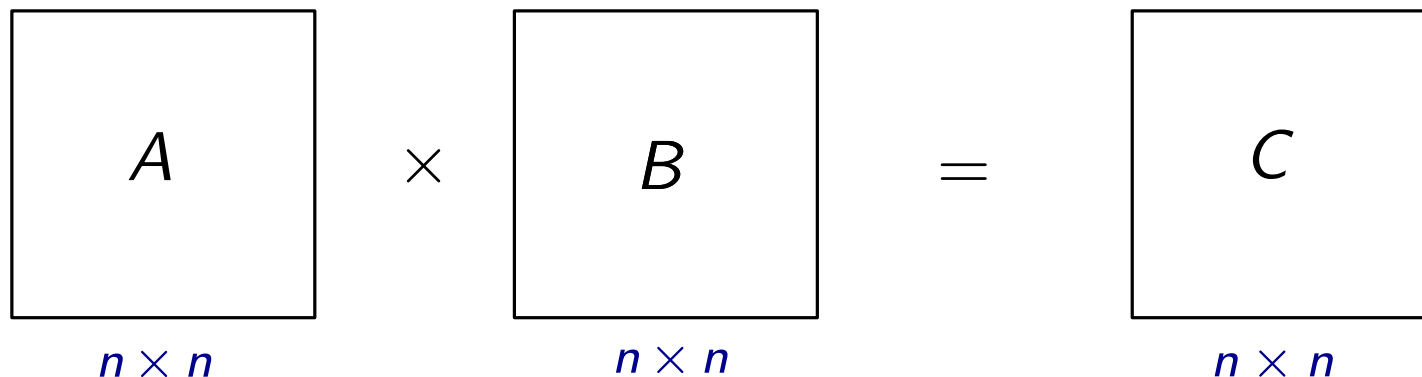
- Similar problems have been studied in the RAM model (Cohen&Lewis, J. Algorithms, '99; Pagh TOCT'13; etc.)



$(1 + \epsilon)$ -approximate  $\ell_0$

# $(1 + \epsilon)$ -approximate $\ell_0$

- Alice holds  $A \in \{0, 1\}^{n \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times n}$
- Let  $C = A \cdot B$ . Goal:  $(1 + \epsilon)$ -approximate  $\|C\|_0$



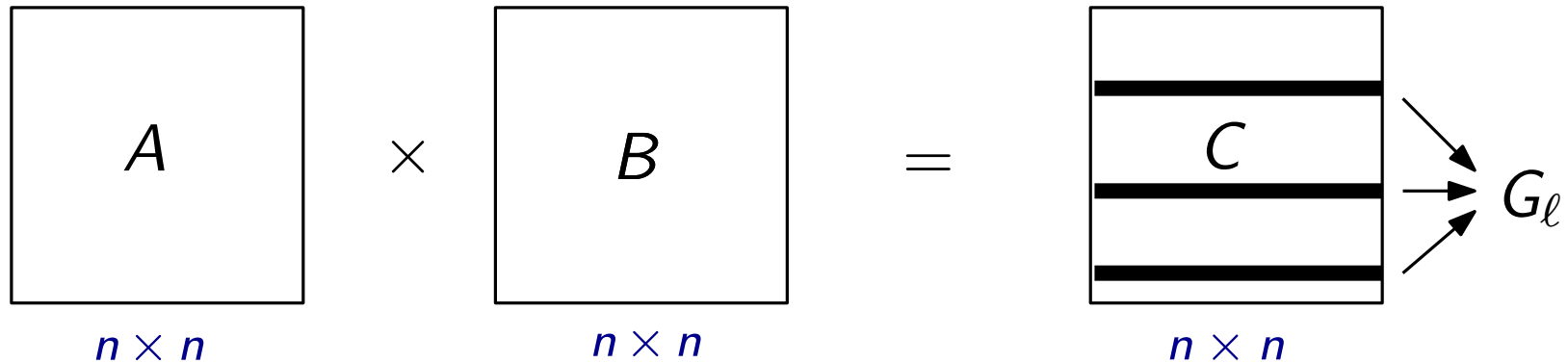
A diagram illustrating the matrix multiplication  $A \cdot B = C$ . Three square boxes are arranged horizontally. The first box contains the letter 'A' and is labeled  $n \times n$  below it. To its right is a multiplication symbol  $\times$ . The second box contains the letter 'B' and is labeled  $n \times n$  below it. To its right is an equals sign  $=$ . The third box contains the letter 'C' and is labeled  $n \times n$  below it.

## High level idea:

1. First perform a **rough estimation** of the number of non-zero entries in the rows of  $C$
2. Use the rough estimation to partition the rows of  $C$  to groups s.t. **rows in the same group have similar #non-zero entries**
3. Sample rows in each group of  $C$  with a probability proportional to the (estimated) average #non-zero entries of the group
4. Use sampled rows to estimate #non-zero entries of  $C$

# $(1 + \epsilon)$ -approximate $\ell_0$

- Alice holds  $A \in \{0, 1\}^{n \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times n}$
- Let  $C = A \cdot B$ . Goal:  $(1 + \epsilon)$ -approximate  $\|C\|_0$



## High level idea:

1. First perform a **rough estimation** of the number of non-zero entries in the rows of  $C$
2. Use the rough estimation to partition the rows of  $C$  to groups s.t. **rows in the same group have similar #non-zero entries**
3. Sample rows in each group of  $C$  with a probability proportional to the (estimated) average #non-zero entries of the group
4. Use sampled rows to estimate #non-zero entries of  $C$

# $(1 + \epsilon)$ -approximate $\ell_0$ (cont.)

- Alice holds  $A \in \{0, 1\}^{n \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times n}$
- Let  $C = A \cdot B$ . Goal:  $(1 + \epsilon)$ -approximate  $\|C\|_0$

**Algorithm.** Set  $\beta = \sqrt{\epsilon}$ ,  $\rho = \Theta(1/\epsilon)$

1. Use the algorithm in VWVZ'15 to get a  $(1 + \beta)$ -approx of  $C$  (denoted by  $\tilde{C}$ )
2. Partition the  $n$  rows of  $\tilde{C}$  to  $L = O(\log n/\beta)$  groups  $G_1, \dots, G_L$ , s.t.  $G_\ell$  contains all row  $i \in [n]$  with  $(1 + \beta)^\ell \leq \|\tilde{C}_{\ell,*}\|_0 \leq (1 + \beta)^{\ell+1}$
3. For each group  $\ell$ , Alice samples each row  $i \in G_\ell$  w.pr.  
$$p_\ell = \frac{\rho}{\|\tilde{C}\|_0} \cdot \frac{\sum_{i \in G_\ell} \|\tilde{C}_{i,*}\|_0}{|G_\ell|}$$
 $A'$  : matrix containing sampled rows of  $A$ .  
Alice sends  $A'$  to Bob
4. Bob computes  $C' \leftarrow A'B$ , outputs  $\sum_{\ell \in [L]} \sum_{i \in G_\ell} \frac{1}{p_\ell} \|C'_{i,*}\|_0$

# $(1 + \epsilon)$ -approximate $\ell_0$ (cont.)

- Alice holds  $A \in \{0, 1\}^{n \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times n}$
- Let  $C = A \cdot B$ . Goal:  $(1 + \epsilon)$ -approximate  $\|C\|_0$

Correctness: expectation  
+ variance estimation

**Algorithm.** Set  $\beta = \sqrt{\epsilon}, \rho = \Theta(1/\epsilon)$

1. Use the algorithm in VWWZ'15 to get a  $(1 + \beta)$ -approx of  $C$  (denoted by  $\tilde{C}$ )
2. Partition the  $n$  rows of  $\tilde{C}$  to  $L = O(\log n/\beta)$  groups  $G_1, \dots, G_L$ , s.t.  $G_\ell$  contains all row  $i \in [n]$  with  $(1 + \beta)^\ell \leq \|\tilde{C}_{\ell,*}\|_0 \leq (1 + \beta)^{\ell+1}$
3. For each group  $\ell$ , Alice samples each row  $i \in G_\ell$  w.pr.  
$$p_\ell = \frac{\rho}{\|\tilde{C}\|_0} \cdot \frac{\sum_{i \in G_\ell} \|\tilde{C}_{i,*}\|_0}{|G_\ell|}$$
 $A'$  : matrix containing sampled rows of  $A$ .  
Alice sends  $A'$  to Bob
4. Bob computes  $C' \leftarrow A'B$ , outputs  $\sum_{\ell \in [L]} \sum_{i \in G_\ell} \frac{1}{p_\ell} \|C'_{i,*}\|_0$

# $(1 + \epsilon)$ -approximate $\ell_0$ (cont.)

- Alice holds  $A \in \{0, 1\}^{n \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times n}$
- Let  $C = A \cdot B$ . Goal:  $(1 + \epsilon)$ -approximate  $\|C\|_0$

Correctness: expectation  
+ variance estimation

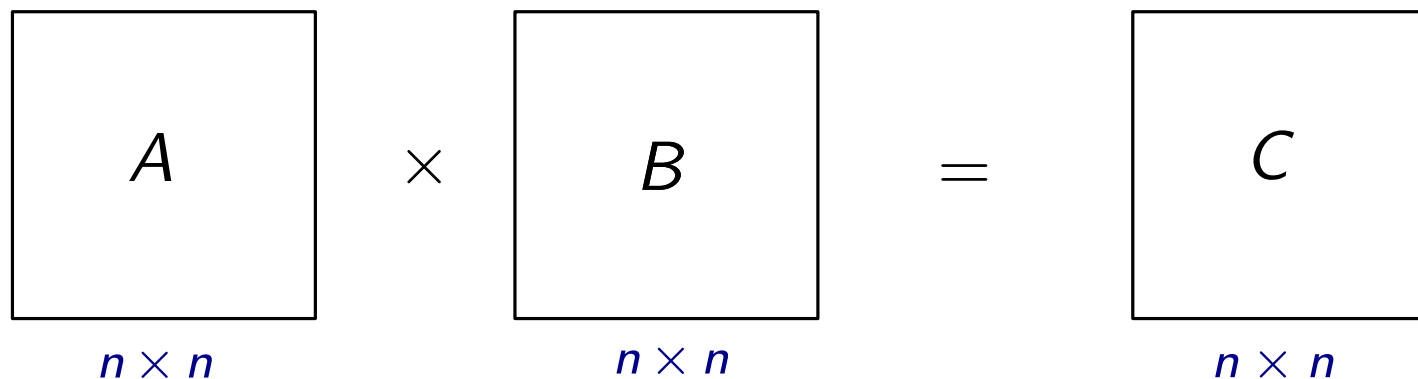
**Algorithm.** Set  $\beta = \sqrt{\epsilon}, \rho = \Theta(1/\epsilon)$

1. Use the algorithm in VWVZ'15 to get a  $(1 + \beta)$ -approx of  $C$  (denoted by  $\tilde{C}$ )  
 $\tilde{O}(n/\beta^2) = \tilde{O}(n/\epsilon)$  bits
2. Partition the  $n$  rows of  $\tilde{C}$  to  $L = O(\log n/\beta)$  groups  $G_1, \dots, G_L$ , s.t.  $G_\ell$  contains all row  $i \in [n]$  with  $(1 + \beta)^\ell \leq \|\tilde{C}_{i,*}\|_0 \leq (1 + \beta)^{\ell+1}$
3. For each group  $\ell$ , Alice samples each row  $i \in G_\ell$  w.pr.  
 $p_\ell = \frac{\rho}{\|\tilde{C}\|_0} \cdot \frac{\sum_{i \in G_\ell} \|\tilde{C}_{i,*}\|_0}{|G_\ell|}$ .  $A'$  : matrix containing sampled rows of  $A$ .  
Alice sends  $A'$  to Bob  
 $\tilde{O}(n \cdot \rho) = \tilde{O}(n/\epsilon)$  bits
4. Bob computes  $C' \leftarrow A'B$ , outputs  $\sum_{\ell \in [L]} \sum_{i \in G_\ell} \frac{1}{p_\ell} \|C'_{i,*}\|_0$

$(2 + \epsilon)$ -approximate  $l_\infty$

# $(2 + \epsilon)$ -approximate $l_\infty$

- Alice holds  $A \in \{0, 1\}^{n \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times n}$
- Let  $C = A \cdot B$ . Goal:  $(1 + \epsilon)$ -approximate  $\|C\|_0$



$$A^0 = A$$

$A^1$  = subsample each  
entry of  $A$  w.pr.  $\frac{1}{1+\epsilon}$

$A^2$  = subsample each  
entry of  $A$  w.pr.  $\frac{1}{(1+\epsilon)^2}$

...

$$C^0 = A^0 \times B$$

$$C^1 = A^1 \times B$$

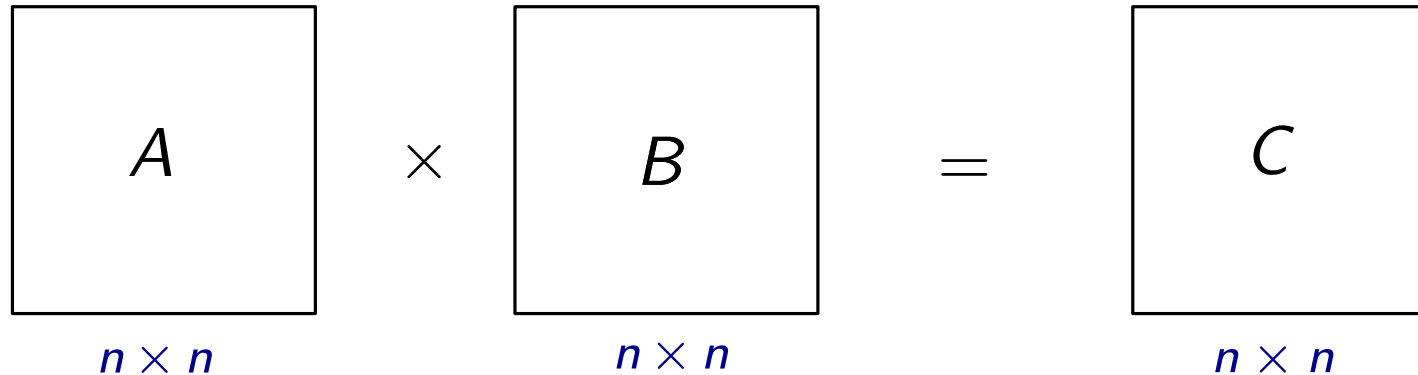
$$C^2 = A^2 \times B$$

...



# $(2 + \epsilon)$ -approximate $l_\infty$

- Alice holds  $A \in \{0, 1\}^{n \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times n}$
- Let  $C = A \cdot B$ . Goal:  $(1 + \epsilon)$ -approximate  $\|C\|_0$



$$A^0 = A$$

$A^1 =$  subsample each entry of  $A$  w.pr.  $\frac{1}{1+\epsilon}$

$A^2 =$  subsample each entry of  $A$  w.pr.  $\frac{1}{(1+\epsilon)^2}$

...

$$C^0 = A^0 \times B$$

$$C^1 = A^1 \times B$$

$$C^2 = A^2 \times B$$

...

**The idea:** subsample  $C$  (via subsampling  $A$ ) to a level  $\ell$  s.t.

(1)  $\ell$  is as large as possible, or,  $l_1(C^\ell)$  is as small as possible

(2)  $l_\infty(C^\ell) \cdot (1 + \epsilon)^\ell$  still approximates  $l_\infty(C)$  well.

# $(2 + \epsilon)$ -approximate $\ell_\infty$ (cont.)

- Alice holds  $A \in \{0, 1\}^{n \times n}$ , Bob holds  $B \in \{0, 1\}^{n \times n}$
- Let  $C = A \cdot B$ . Goal:  $(1 + \epsilon)$ -approximate  $\|C\|_0$

**Algorithm** Set  $L = O(\log n/\epsilon)$ ,  $\gamma = \Theta(\log n/\epsilon^2)$

1. For  $\ell = 0, 1, \dots, L$ , let  $C^\ell \leftarrow A^\ell B$   
 $A^\ell \leftarrow$  sample each '1' in  $A$  w.pr.  $p_\ell = \frac{1}{(1+\epsilon)^\ell}$ .
2. Let  $\ell^*$  be the **smallest**  $\ell$  for which  $\|C^\ell\|_1 \leq \gamma n^2$ .
3. For each  $j \in [n]$ 
  - (a)  $u_j$ : #'1's in  $j$ -th column of  $A^{\ell^*}$ ;  
 $v_j$ : #'1's in  $j$ -th row of  $B$
  - (b) If  $u_j \leq v_j$ , then Alice sends  $j$ -th column of  $A^{\ell^*}$  to Bob;  
otherwise Bob sends  $j$ -th row of  $B$  to Alice
4. Alice and Bob use received information to compute matrices  $C_A$  and  $C_B$  respectively, s.t.  $C_A + C_B = C^{\ell^*}$
5. Output  $\max \left\{ \frac{\|C_A\|_\infty}{p_{\ell^*}}, \frac{\|C_B\|_\infty}{p_{\ell^*}} \right\}$

# $(2 + \epsilon)$ -approximate $l_\infty$ (cont.)

- Correctness

**Lemma:** With probability  $1 - \frac{1}{n^2}$ ,

$\frac{\|C^{\ell^*}\|_\infty}{p_{\ell^*}}$  approximates  $\|C\|_\infty$  within a factor of  $1 + \epsilon$ .

**Simple Fact:** If  $C_A + C_B = C^{\ell^*}$ , then

$$\frac{\|C^{\ell^*}\|_\infty}{2} \leq \max\{\|C_A\|_\infty, \|C_B\|_\infty\} \leq \|C^{\ell^*}\|_\infty$$

Put together,  $\max\left\{\frac{\|C_A\|_\infty}{p_{\ell^*}}, \frac{\|C_B\|_\infty}{p_{\ell^*}}\right\}$  approximates  $\|C\|_\infty$  within a factor of  $2 + \epsilon$ .

# $(2 + \epsilon)$ -approximate $l_\infty$ (cont.)

- Communication cost

**Bottleneck:**

For each  $j \in [n]$

- (a)  $u_j$ : #'1's in  $j$ -th column of  $A^{\ell^*}$  ;  
 $v_j$ : #'1's in  $j$ -th row of  $B$
- (b) If  $u_j \leq v_j$ , then Alice sends  $j$ -th column of  $A^{\ell^*}$  to Bob;  
otherwise Bob sends  $j$ -th row of  $B$  to Alice

# $(2 + \epsilon)$ -approximate $\ell_\infty$ (cont.)

- Communication cost

**Bottleneck:**

For each  $j \in [n]$

- (a)  $u_j$ : # '1's in  $j$ -th column of  $A^{\ell^*}$ ;  
 $v_j$ : # '1's in  $j$ -th row of  $B$
- (b) If  $u_j \leq v_j$ , then Alice sends  $j$ -th column of  $A^{\ell^*}$  to Bob;  
otherwise Bob sends  $j$ -th row of  $B$  to Alice

For each  $j \in [n]$ , we analyze two cases:

- If  $u_j, v_j > \sqrt{n}/\epsilon$ , # such  $j$  is bounded by  $\|C^{\ell^*}\|_1 / (\sqrt{n}/\epsilon)^2$

The comm. cost can be bounded by  $\tilde{O}\left(\frac{n^{1.5}}{\epsilon}\right)$ .

- If  $\min\{u_j, v_j\} \leq \sqrt{n}/\epsilon$ , the communication is bounded by

$$\sum_{j: \min\{u_j, v_j\} \leq \frac{\sqrt{n}}{\epsilon}} \min\{u_j, v_j\} \leq n \times \frac{\sqrt{n}}{\epsilon} \leq \frac{n^{1.5}}{\epsilon}.$$

# Concluding Remarks

## Main results:

- The communication cost for  $(1 + \epsilon)$ -approximating  $\ell_p$  ( $p \in [0, 2]$ ) with  $\Sigma = \mathbb{Z}$  is  $\tilde{O}(n/\epsilon)$ .
- The communication cost for  $(2 + \epsilon)$ -approximating  $\ell_\infty$  with  $\Sigma = \{0, 1\}$  is  $\tilde{\Theta}(n^{1.5}/\epsilon)$ ;  
that for  $\kappa$ -approximating  $\ell_\infty$  with  $\Sigma = \{0, 1\}$  is  $\tilde{\Theta}(n^{1.5}/\kappa)$
- The communication cost for  $\ell_p$ - $(\phi, \epsilon)$ -heavy-hitters with  $\Sigma = \mathbb{Z}$  is  $\tilde{O}(\frac{\sqrt{\phi}}{\epsilon} n)$ ;  
that for  $\ell_p$ - $(\phi, \epsilon)$ -heavy-hitters with  $\Sigma = \{0, 1\}$  is  $\tilde{O}(n + \frac{\phi}{\epsilon^2})$ .

# Concluding Remarks

## Main results:

- The communication cost for  $(1 + \epsilon)$ -approximating  $\ell_p$  ( $p \in [0, 2]$ ) with  $\Sigma = \mathbb{Z}$  is  $\tilde{O}(n/\epsilon)$ .  
Open question: close the gap between this UB and the  $\Omega(n/\epsilon^{2/3})$  LB.
- The communication cost for  $(2 + \epsilon)$ -approximating  $\ell_\infty$  with  $\Sigma = \{0, 1\}$  is  $\tilde{\Theta}(n^{1.5}/\epsilon)$ ;  
that for  $\kappa$ -approximating  $\ell_\infty$  with  $\Sigma = \{0, 1\}$  is  $\tilde{\Theta}(n^{1.5}/\kappa)$
- The communication cost for  $\ell_p$ - $(\phi, \epsilon)$ -heavy-hitters with  $\Sigma = \mathbb{Z}$  is  $\tilde{O}(\frac{\sqrt{\phi}}{\epsilon} n)$ ;  
that for  $\ell_p$ - $(\phi, \epsilon)$ -heavy-hitters with  $\Sigma = \{0, 1\}$  is  $\tilde{O}(n + \frac{\phi}{\epsilon^2})$ .  
Open question: tight LBs?

Thank you!  
Questions?



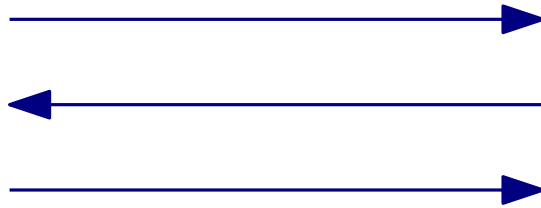
A lower bound for  $l_\infty$

# A Lower Bound for $l_\infty$

## Set-disjointness



$\exists i, x[i] = y[i] = 1?$



$$x \in \{0, 1\}^{n^2/4}$$

$$y \in \{0, 1\}^{n^2/4}$$

**Theorem** (Bar-Yossef et al. JCSS'04): Any randomized algorithm that solves the above set-disjointness problem w.pr. 0.51 needs  $\Omega(n^2)$  comm.

# A Lower Bound for $\ell_\infty$ (cont.)

## A reduction from set-disjointness to $\ell_\infty(AB)$

1. Alice partitions  $x \in \{0, 1\}^{n^2/4}$  to  $n/2$  chunks of size  $n/2$  each, and uses them as rows to construct  $A' \in \{0, 1\}^{\frac{n}{2} \times \frac{n}{2}}$ . Further, let

$$A = \begin{bmatrix} A' & I \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

2. Similarly, Bob uses  $y \in \{0, 1\}^{n^2/4}$  to construct  $B' \in \{0, 1\}^{\frac{n}{2} \times \frac{n}{2}}$ , and further let

$$B = \begin{bmatrix} I & \mathbf{0} \\ B' & \mathbf{0} \end{bmatrix}.$$

3. We have  $\|A \cdot B\|_\infty = \|A' + B'\|_\infty$ , which is 2 if  $x \cap y \neq \emptyset$ , and 1 otherwise.