Computing Inconsistency Measures Under Differential Privacy

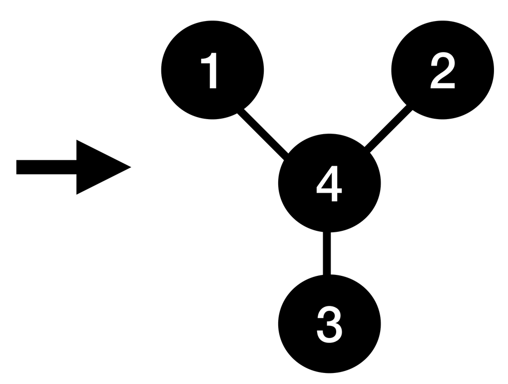
Shubhankar Mohapatra[1], Amir Gilad[2], Xi He[1], Benny Kimelfeld[3]

University of Waterloo[1], Hebrew University[2], Technion[3]

Background

Denial Constraints and Conflict Graphs

ID	Capital	Country
1	Ottawa	Canada
2	Ottawa	Canada
3	Ottawa	Canada
4	Ottawa	Kanada



σ : Capital —> Country

"country of two tuples must be the same if their capital is the same"

Inconsistency Measures [1]

Given, dataset D and constraint set Σ, inconsistency measures are of the form I(D, Σ) → ℝ:

1. Drastic Measure I_D (G) = existence of an edge ✗
2. Minimal inconsistency measure I_{MI} (G) = number of edges ✓
3. Problematic measure I_P (G) = number of vertices with positive degree ✓
4. Maximal consistency measure I_{MC} (G) = number of maximal independent sets ✗
5. Optimal repair measure I_R (G) = minimum vertex cover size ✓

✓ = our work

Differential Privacy [2]

A randomized algorithm $A: \mathscr{G} \to \mathscr{R}$ satisfies ε-differential privacy (DP) if for any two adjacent graphs $G, G' \in \mathscr{G}$ that differ in a node and for any subset of outputs $o \subseteq \mathscr{R}$ it holds that :

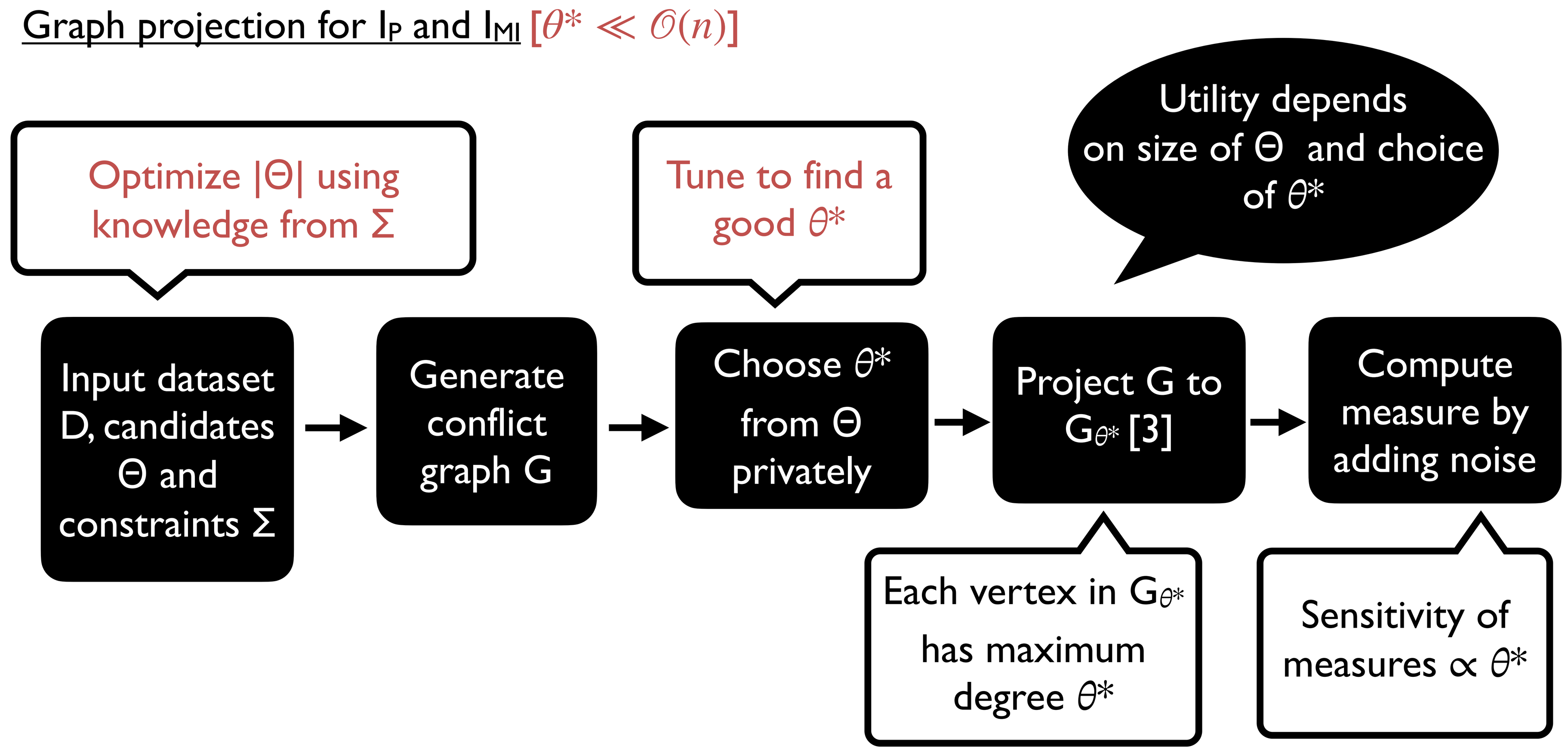
$$\Pr[A(G) \in o] \leq e^{\varepsilon} \Pr[A(G') \in o]$$

Differentially Private Inconsistency Measures

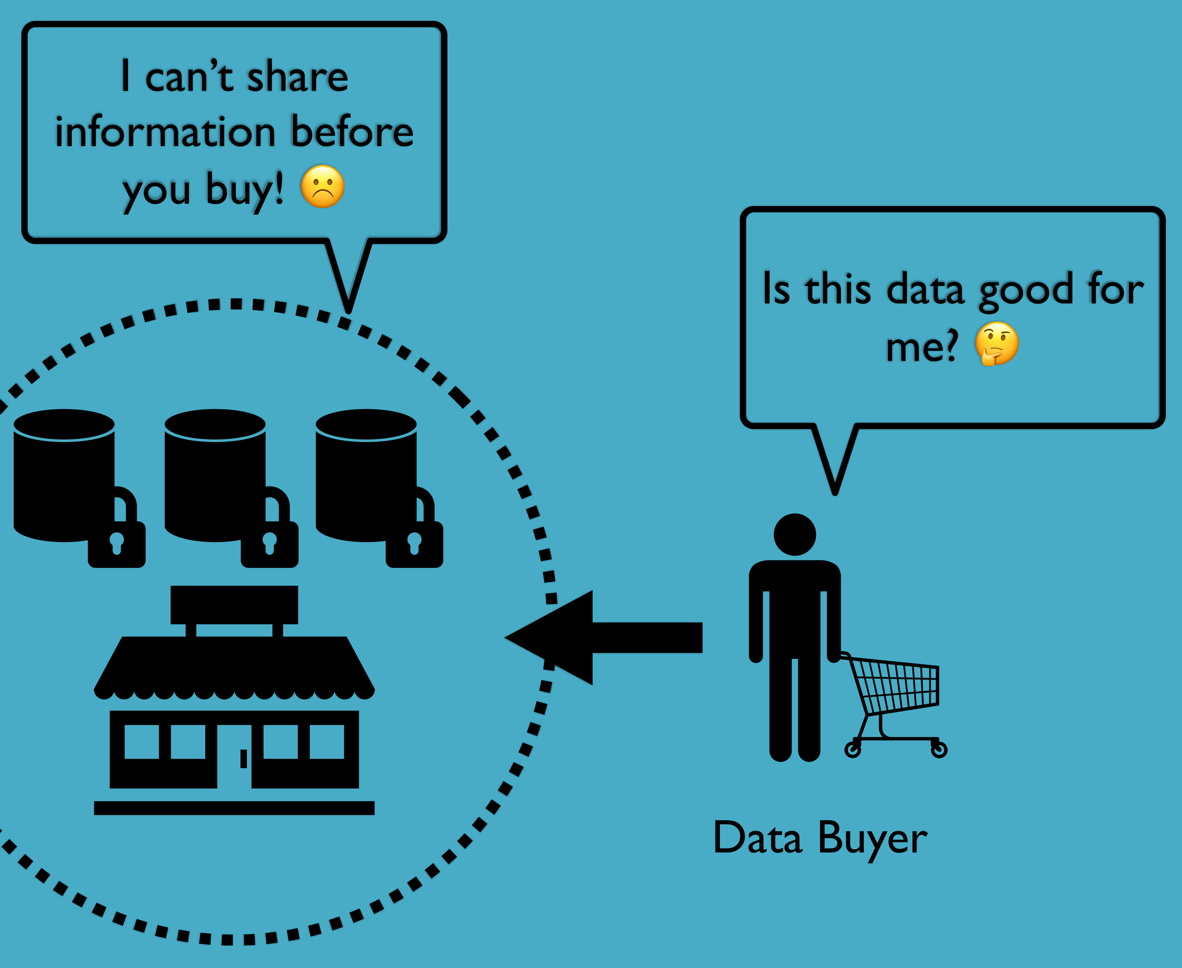
Challenges

1. Computational hardness: Minimum vertex cover (I_R) and #maximal independent sets (I_{MC}) are NP hard problems
2. High sensitivity: Maximum change in output when G is replaced by G' is $\mathcal{O}(n)$, where n is the number of nodes

Graph projection for I_P and I_{MI} [$\theta^* \ll \mathcal{O}(n)$]



Setup

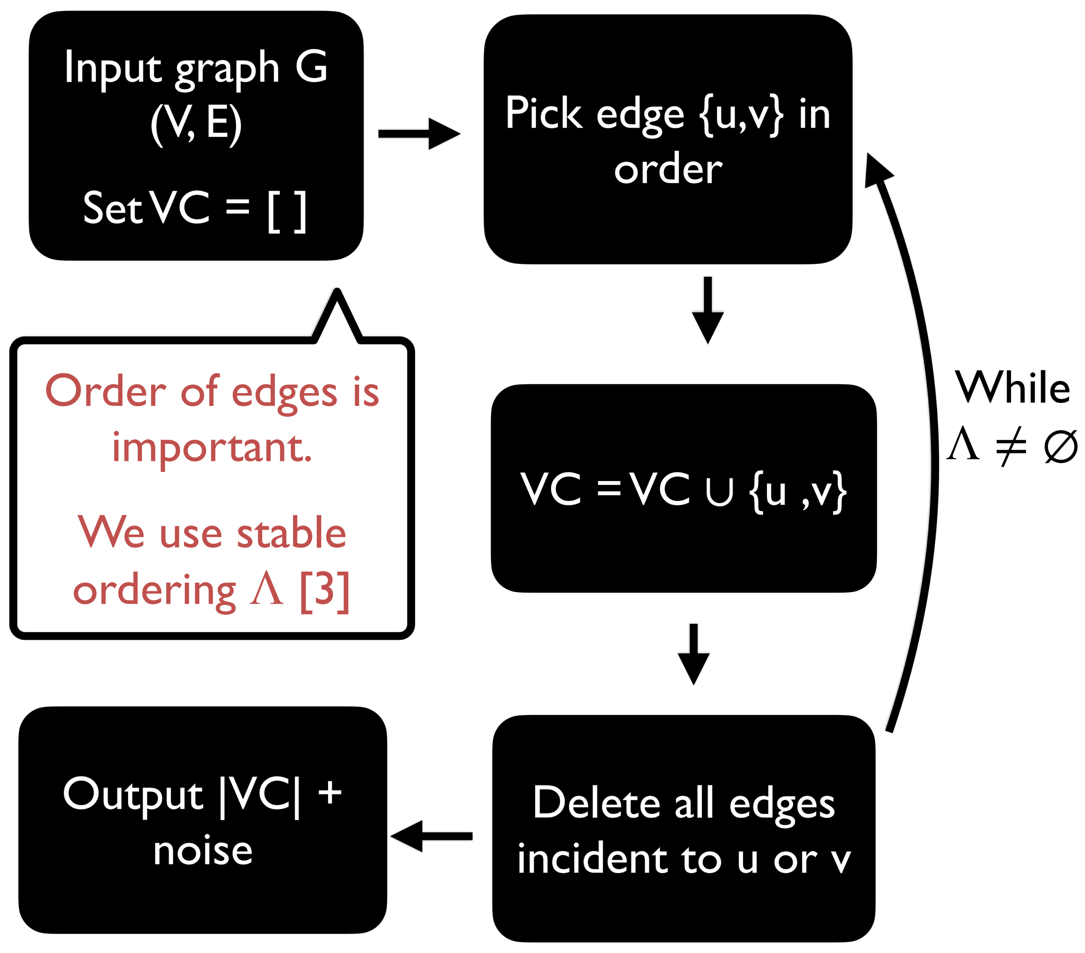


Research Questions:

- Are there errors in a private dataset?
- How much effort is required to repair?

DP vertex cover size for I_R [$2 \ll \mathcal{O}(n)$]

We analyse the 2-approximate vertex cover size algorithm and show that it has sensitivity of 2.



Experiments

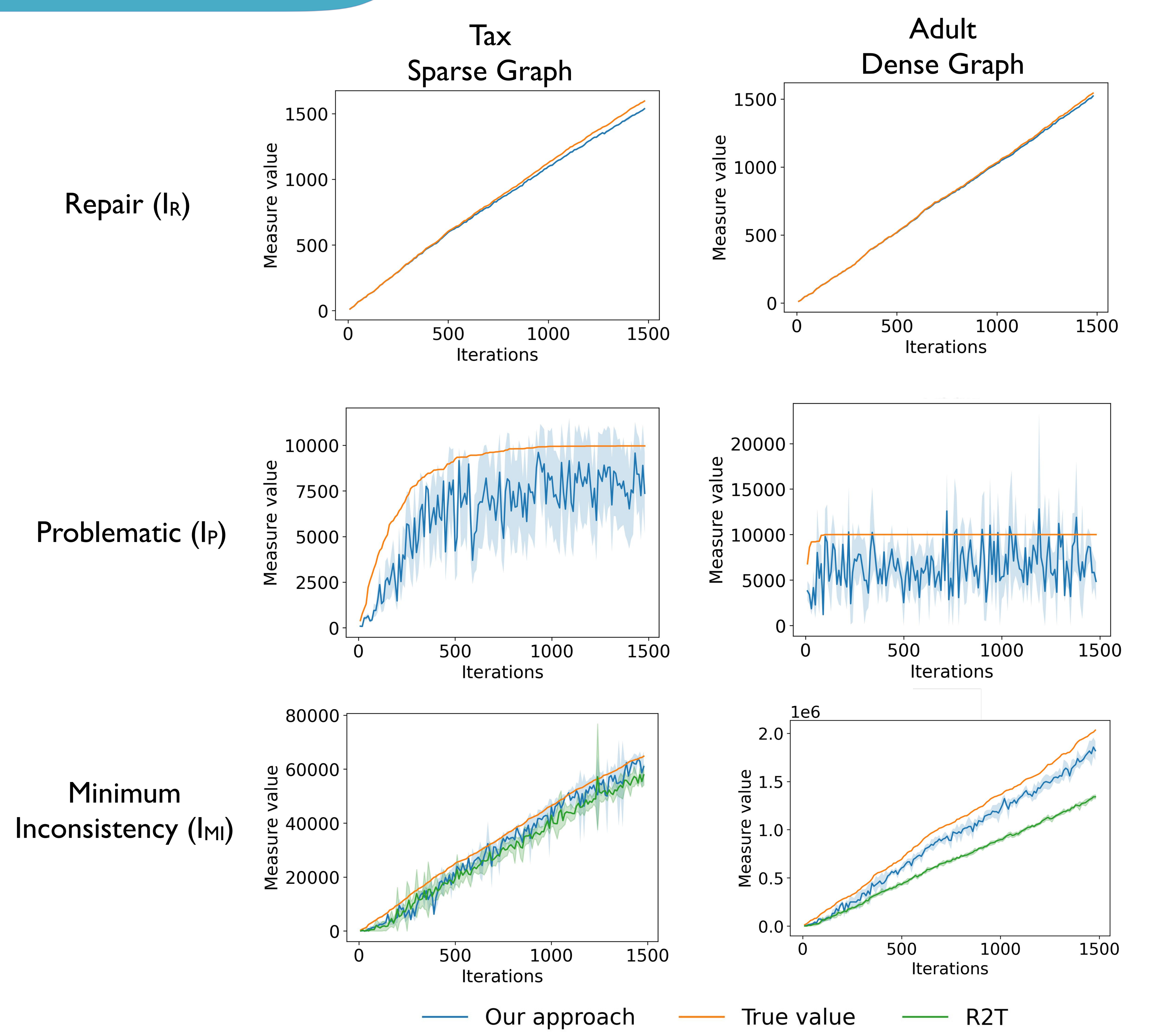
Setup

- Inconsistency: Random typo to 1% rows
- Privacy: ε = 1
- Measure: True vs private by adding one typo at a time

Datasets

- Five real-world datasets with varying conflict graph densities
- We experiment on subset of 10k and repeat for 10 times and average

Dataset	#Tuples	#Attributes	#Constraints	Graph density
Adult	32k	15	3	9635
Flight	500k	20	13	1520
Hospital	114k	15	7	793
Stock	122k	7	1	1
Tax	1M	15	9	373



References:

1. Livshits, Ester, et al. "Properties of inconsistency measures for databases." *International Conference on Management of Data*. 2021.
2. Dwork, Cynthia, et al. "Calibrating noise to sensitivity in private data analysis." *TCC 2006*
3. Day, Wei-Yen, et al. "Publishing graph degree distribution with node differential privacy." *International Conference on Management of Data* 2016.
4. Dong, Wei, et al. "R2T: Instance-optimal Truncation for Differentially Private Query Evaluation with Foreign Keys." *ACM SIGMOD* 2023