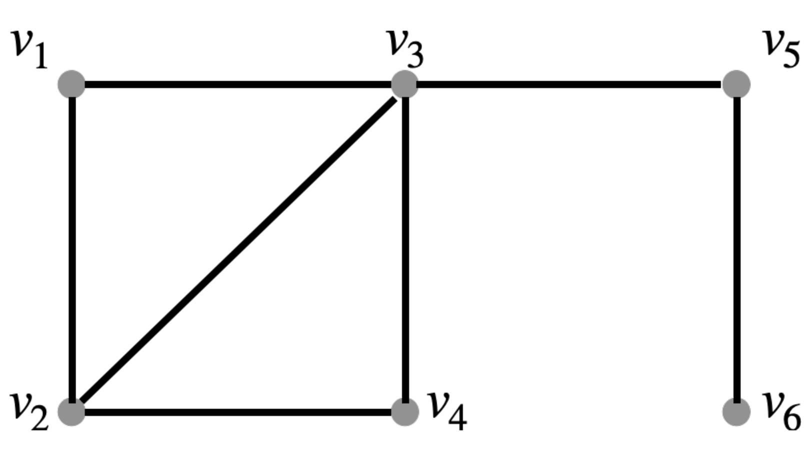
On the Price of Differential Privacy for Hierarchical Clustering

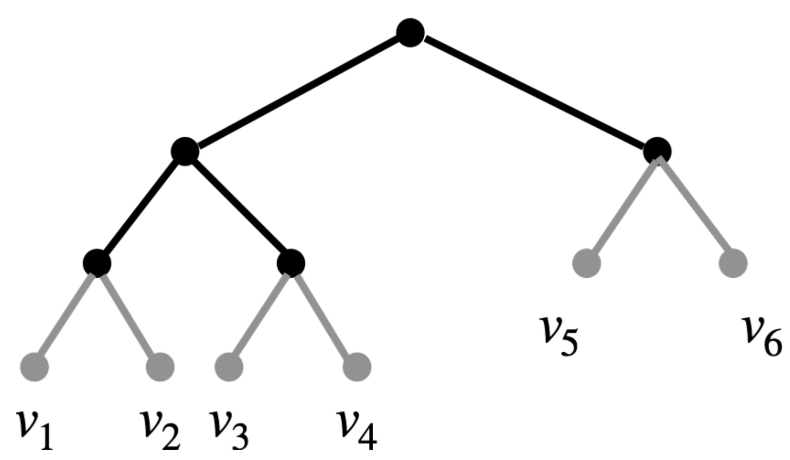
Chengyuan Deng, Jie Gao, Jalaj Upadhyay, Chen Wang, Samson Zhou

Rutgers University, Rice University, Texas A&M University

Dasgupta's Cost for HC

➢ **Hierarchical Clustering:** organize data with a hierarchy of clusters





- Vertex: data point
- Edge weight: pair-wise similarity

The hierarchical clustering output can be represented by a tree

➢ **Dasgupta's Cost:** an objective to capture the "quality" of the clustering.

$$\mathrm{cost}_G(T) = \sum_{(i,j)\in E} w_{ij} \cdot |\mathrm{leaves}(T[i \vee j])|$$

$T[i \vee j]$ is the subtree rooted on the lowest common ancestor of i, j

➢ Cost of the above example:

$$\mathrm{cost}_G(T) = 1{\times}6 + 3{\times}4 + 3{\times}2$$

➢ [Das'15, CC'16]: Finding the clustering with minimum cost is NP-hard. However, it can be approximated using sparsest cut in polynomial time.

Background

Approximate Dasgupta's cost with balanced sparsest cut:

- Sparsity of a graph: $\phi_G = \min_{S\subseteq V} \frac{w(S, V\setminus S)}{|S|}$.
- Balanced sparsest cut: the cut induced by $(S^*, V\setminus S^*)$. And $|S^*| = O(n)$.
- $O(\alpha)$-approximation of balanced sparsest cut w.r.t. the sparsity gives $O(\alpha)$-approximation of the optimal Dasgupta's cost.
- $\alpha = O(\sqrt{\log n})$ for balanced sparsest cut has a poly-time algorithm (A-1).

Differential privacy models on graphs:

- Node-DP: Neighboring graphs differ by a node.
- Edge-DP: Neighboring graphs differ by an edge.
- Weight-DP: Neighboring graphs have weights with l_1 difference ≤ 1.

Prior work on DP Hierarchical Clustering [Imola et al. 23]:

- Upper Bound: $\tilde{O}(n^2/\varepsilon)$ additive error. But requires exponential time.
- Lower Bound: $\Omega(n^2/\varepsilon)$ additive error even when the optimal cost is $O(n)$.
- Achieving ε-DP in the Edge-DP model.

What is the "correct" notion of privacy for hierarchical clustering?

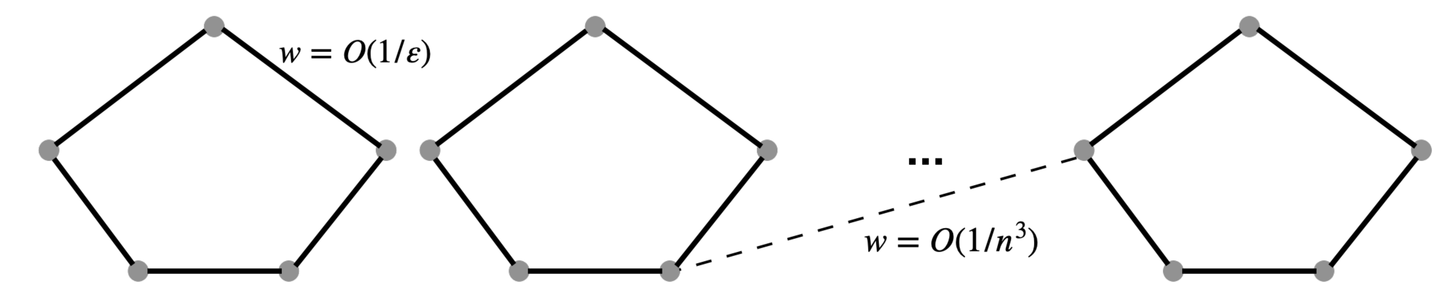
Main Results

1. The $\Omega(n^2/\varepsilon)$ lower bound even holds for weight-DP model.
2. However, with one assumption that all weights are at least 1, we can achieve $O(\log^{1.5} n)$-approximation in poly-time.
3. First algorithm with reasonable implementations for general graphs.

Lower Bound for Weight-DP

The additive error of $\Omega(n^2/\varepsilon)$ is necessary in the weight-DP model:

- Inspired by the lower bound instance for edge-DP
- Main idea: Embed the disjoint 5-cycle graphs into a complete graph



Main argument:

- To minimize Dasgupta's cost, the algorithm should always partition disconnected components first.
- Any private algorithm will have to cut cycle edges early.

Private Algorithm

Unit weight assumption:

❑ We can't afford "significant changes" on important edges.
❑ Perspective from DP: neighboring graphs differ at an atomic level.
❑ Not a trivialization: input perturbation or output perturbation ✕

Add $\mathrm{Lap}(1/\varepsilon)$ to each edge, weight, then compute the sparsest cut

Compute all cuts, then add $\mathrm{Lap}(\phi_G/\varepsilon)$ to each and output the sparsest cut

❑ Adding noise may change the sparsity a lot!

Private algorithm for balanced sparsest cut:

Amplify the gap between sparse and non-sparse cuts

- Add $O(\log n/\varepsilon)$ to all edge weights.
- Add independent $\mathrm{Lap}(1/\varepsilon)$ to all edge weights (input perturbation).
- Run algorithm A-1 on the perturbed graph and return the cut.

Recursively call the above algorithm, we get private algorithm for HC.

Analysis:

✓ Privacy: Achieved by Input perturbation with Laplace mechanism.

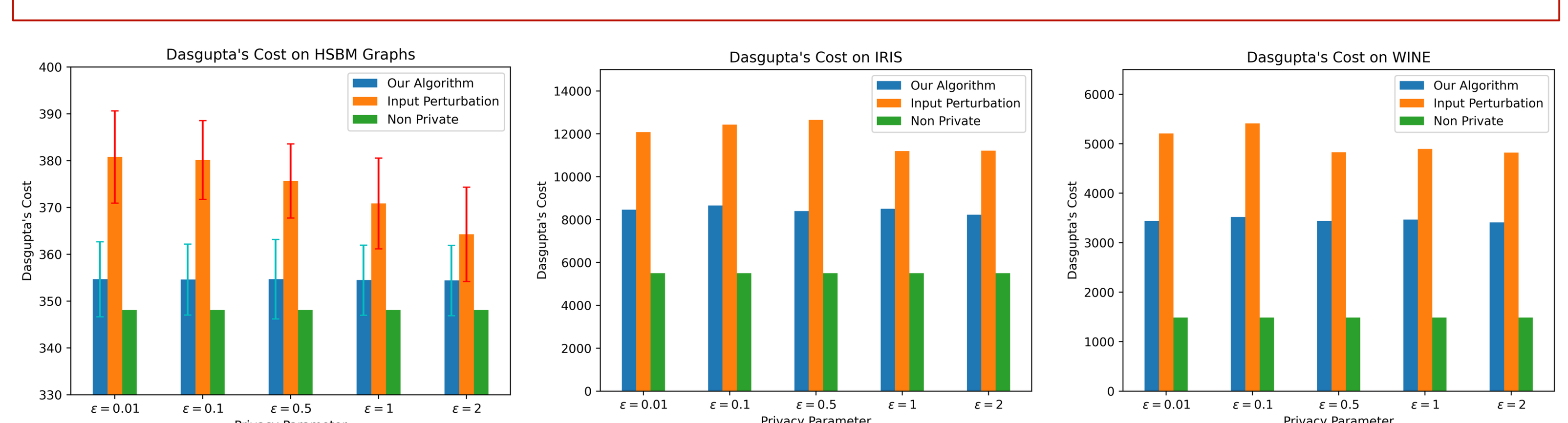
✓ Utility: $O(\log^{1.5} n)$ multiplicative error for HC has two sources:

a) $O(\sqrt{\log n})$-approximation from algorithm A1 for sparsest cut.
b) $O(\log n)$-approximation due to the Laplace noise.

Experiments

Datasets and Baseline:

- Synthetic: SBM and HSBM graphs
- Real-world: datasets from sk-learn (standard in the literature)
- Baseline: Input perturbation



Comparing the Dasgupta's Cost:

- For well-clustered graphs, our algorithm performs close to non-private.
- For all graphs, our algorithm performs better than input perturbation.

Our algorithm scales well to large graphs.

Open Problem

1. Any approximate-DP algorithm for this model?
2. More applications of the DP sparsest cut algorithm?
3. Long shot: Hierarchical Agglomerative Clustering?

References

Dasgupta, Sanjoy. "A cost function for similarity-based hierarchical clustering." *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing.* 2016.

Charikar, Moses, and Vaggos Chatziafratis. "Approximate hierarchical clustering via sparsest cut and spreading metrics." *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms.* Society for Industrial and Applied Mathematics, 2017.

Imola, Jacob, et al. "Differentially private hierarchical clustering with provable approximation guarantees." *International Conference on Machine Learning.* PMLR, 2023.