Facility Location Problem in Differential Privacy Model Revisited

Yunus Esencayi SUNY at Buffalo **Marco Gaboardi** Boston University **Shi Li** SUNY at Buffalo **Di Wang** SUNY at Buffalo

*Alphabetically sorted

Abstract

In this paper we study the uncapacitated facility location problem in the model of differential privacy (DP) with uniform facility cost. Specifically, we first show that, under the *hierarchically well-separated tree (HST) metrics* and the *super-set output setting* that was introduced in Gupta et. al., there is an ϵ-DP algorithm that achieves an $O(\frac{1}{\epsilon})$ (expected multiplicative) approximation ratio; this implies an $O(\frac{\log n}{\epsilon})$ approximation ratio for the general metric case, where n is the size of the input metric. These bounds improve the best-known results given by Gupta et. al. In particular, our approximation ratio for HST-metrics is independent of n, and the ratio for general metrics is independent of the aspect ratio of the input metric.

On the negative side, we show that the approximation ratio of any ϵ-DP algorithm is lower bounded by $\Omega(\frac{1}{\sqrt{\epsilon}})$, even for instances on HST metrics with uniform facility cost, under the super-set output setting. The lower bound shows that the dependence of the approximation ratio for HST metrics on ϵ can not be removed or greatly improved. Our novel methods and techniques for both the upper and lower bound may find additional applications.

Facility Location Problem

The input to the Uniform Facility Location (Uniform-FL) problem is a tuple $(V, d, f, \vec{N})$, where (V, d) is a n-point discrete metric, $f \in_{\geq 0}$ is the facility cost, and $\vec{N} = (N_v)_{v\in V} \in^V_{\geq 0}$ gives the number of clients in each location $v \in V$. The goal of the problem is to find a set of facility locations $S \subseteq V$ which minimize the following, where $d(v, S) = \min_{s\in S} d(v, s)$,

$$\min_{S\subseteq V} d(S; \vec{N}) := |S| \cdot f + \sum_{v\in V} N_v d(v, S). \qquad (1)$$

The first term of (1) is called the *facility cost* and the second term is called the *connection cost*.

Differential Privacy

A randomized algorithm $\mathcal{A}$ is ϵ-differentially private (DP) if for any client $i \in [N]$, any two possible data entries $v_i, v_i' \in V$, any dataset $D_{-i} \in V^{N-1}$ and for all events $\mathcal{T} \subseteq \mathcal{S}$ in the output space of $\mathcal{A}$, we have $\Pr[\mathcal{A}(v_i, D_{-i}) \in \mathcal{T}] \leq e^\epsilon \Pr[\mathcal{A}(v_i', D_{-i}) \in \mathcal{T}]$

Main Theorem

Given any UFL tuple $(V, d, f, \vec{N})$ where $|V| = n$ and $\epsilon > 0$, there is an efficient ϵ-DP algorithm $\mathcal{A}$ in the super-set output setting achieving an approximation ratio of $O(\frac{\log n}{\epsilon})$.

Previous Work

Gupta et. al. showed that, under the following setting, an $O(\frac{\log^2 n \log^2 \Delta}{\epsilon})$ approximation ratio under the ϵ-DP model is possible, where $\Delta = \max_{u,v\in V} d(u, v)$ is the diameter of the input metric.

They also showed that any 1-DP algorithm for UFL under general metric that outputs the set of open facilities must have a (multiplicative) approximation ratio of $\Omega(\sqrt{n})$ which negatively shows that UFL in DP model is useless. Thus one needs to consider some relaxed settings in order to address the issue.

Our Results

We show that under the so called *Hierarchical-Well-Separated-Tree (HST)* metrics, there is an algorithm that achieves $O(\frac{1}{\epsilon})$ approximation ratio. Using the classic FRT tree embedding technique of [?], we can achieve $O(\frac{\log n}{\epsilon})$ approximation ratio for any metrics, under the ϵ-DP model and the super-set output setting. These factors respectively improve upon a factor of $O(\log n \log^2 \Delta)$ in Gupta et. al. for HST and general metrics. Thus, for HST-metrics, our approximation only depends on ϵ. For general metrics, our result removed the poly-logarithmic dependence on Δ.

On the negative side, we show that the approximation ratio under ϵ-DP model is lower bounded by $\Omega(\frac{1}{\sqrt{\epsilon}})$ even if the metric is a star (which is a special case of a HST). This shows that the dependence on ϵ is unavoidable and can not been improved greatly.