A Note on Metric 1-median Selection

Ching-Lueh Chang(✉)

Department of Computer Science and Engineering, Yuan Ze University, Taoyuan, Taiwan
clchang@saturn.yzu.edu.tw

Abstract. METRIC 1-MEDIAN asks for $\text{argmin}_{p=1}^{n} \sum_{q=1}^{n} d(p,q)$, breaking ties arbitrarily, given a metric space $(\{1,2,\ldots,n\},d)$. Let A be any deterministic algorithm for METRIC 1-MEDIAN making each point in $\{1,2,\ldots,n\}$ involve in only $O(1)$ queries to d. We show A to not be $o(\log n)$-approximate.

Keywords: 1-median · Closeness centrality · Metric space

1 Introduction

For each positive integer n, $[n] \equiv \{1,2,\ldots,n\}$. Given a metric space $([n],d)$, METRIC 1-MEDIAN asks for

$$\underset{p\in[n]}{\text{argmin}} \sum_{q\in[n]} d\left(p,q\right),$$

breaking ties arbitrarily. An algorithm for METRIC 1-MEDIAN may query for $d(p,q)$ for any p, $q \in [n]$. Indyk [3,4] designs a Monte Carlo $O(n/\epsilon^2)$-time $(1+\epsilon)$-approximation algorithm for METRIC 1-MEDIAN, where $\epsilon > 0$.

Chang [1, Corollary 10] gives a deterministic, $O(\exp(O(1/\epsilon)) \cdot n)$-query, $(\epsilon \log n)$-approximation and nonadaptive algorithm for METRIC 1-MEDIAN, where $\epsilon > 0$ is any constant. For infinitely many n, his algorithm makes $O(1)$ queries concerning each point in $[n]$. We show that such a property forbids his algorithm to be $o(\log n)$-appproximate. As in previous lower bounds for METRIC 1-MEDIAN, our proof uses the adversarial method (see [2] and the references therein).

Chang [2] shows that METRIC 1-MEDIAN has no deterministic $o(n^{1+1/(h-1)})$-query $(2h \cdot (1-\epsilon))$-approximation algorithms for any constant $\epsilon > 0$ and any integer-valued $h = h(n) \geq 2$ satisfying $h = o(n^{1/(h-1)})$. His result does not imply ours in any obvious way.

Supported in part by the Ministry of Science and Technology of Taiwan under grant 107-2221-E-155-006-MY2.

C.-Y. Chang et al. (Eds.): ICS 2018, CCIS 1013, pp. 457–459, 2019.
https://doi.org/10.1007/978-981-13-9190-3_49

2 Main Result

For all metric spaces $([n], d)$ and algorithms A for METRIC 1-MEDIAN, define

$$Q_A^d \equiv \left\{(p,q) \in [n]^2 \mid A^d \text{ queries for } d\left(p,q\right)\right\}$$

to be the set of queries of A with oracle access to d, where (p, q) is interpreted as an unordered pair. So the maximum number of queries concerning a point is

$$\deg_A^d \equiv \max_{p \in [n]} \left|\left\{q \in [n] \mid (p,q) \in Q_A^d\right\}\right|.$$

Theorem 1. *Let A be a deterministic algorithm for* METRIC 1-MEDIAN *satisfying $deg_A^d = O(1)$ for all metric spaces $([n], d)$. Then A is not $o(\log n)$-approximate.*

Proof. Assume without loss of generality that A does not query for $d(p, p)$ for any $p \in [n]$. We will construct d as A queries. In particular, d is fully determined after A outputs.

Answer each query of A by 1. So $d(p, q) = 1$ if A ever queries for $d(p, q)$, where $(p, q) \in [n]^2$. Consider the undirected graph $G = ([n], Q)$, where Q denotes the set of all queries (as unordered pairs in $[n]^2$) of A. By padding dummy queries, assume $\mathrm{diam}(G) = O(\log n)$ without loss of generality (e.g., pick an $O(1)$-regular expander $G' = ([n], E')$ and assume $E' \subseteq Q$ by padding). All queries of A, having been answered by 1, are clearly consistent with d_G. Denote the output of A by p^*. As $\deg_A^d = O(1)$ for all metric spaces $([n], d)$, G has a maximum degree of $O(1)$. So there exists a small constant $\epsilon > 0$ such that p^* has distance in G greater than $\epsilon \log n$ to at least $n - \sqrt{n}$ points. That is,

$$U \equiv \{q \in [n] \mid d_G\left(p^*, q\right) > \epsilon \log n\}$$

satisfies $|U| \geq n - \sqrt{n}$. Define an undirected graph $H = ([n], Q \cup (U \times U))$ by adding to G an edge between each pair of points in U. Because $d_G(p^*, q) > \epsilon \log n$ for all $q \in U$ and $|U| \geq n - \sqrt{n}$,

$$\sum_{q \in U} d_H\left(p^*, q\right) \geq |U| \cdot \epsilon \log n = \Omega\left(n \log n\right). \tag{1}$$

For each $u \in U$,

$$\begin{aligned}
\sum_{q \in [n]} d_H\left(u, q\right) &= \sum_{q \in U} d_H\left(u, q\right) + \sum_{q \in [n] \setminus U} d_H\left(u, q\right) \\
&\leq |U| + \sum_{q \in [n] \setminus U} d_H\left(u, q\right) \\
&\leq |U| + O(\log n) \cdot (n - |U|) \\
&\leq n + o\left(n\right),
\end{aligned} \tag{2}$$

where the first, second and third inequalities follow from $H = ([n], Q \cup (U \times U))$, $\mathrm{diam}(H) \leq \mathrm{diam}(G) = O(\log n)$ and $|U| \geq n - \sqrt{n}$, respectively.

As $H = ([n], Q \cup (U \times U))$, $d_H(p,q) = 1$ for all $(p,q) \in Q$. Consequently, A^{d_H} outputs p^* (recall that A outputs p^* if every query is answered by 1). So Eqs. (1)–(2) forbid the output of A^{d_H} to be $o(\log n)$-approximate. In particular, A^{d_H} outputs a point with average d_H-distance to other points $\Omega(\log n)$ times the minimum possible.

Theorem 1 forbids Chang's [1, Corollary 10] deterministic $O(n)$-query algorithm to be $o(\log n)$-approximate. But in general, it is open whether METRIC 1-MEDIAN has a deterministic $O(n)$-query $o(\log n)$-approximation algorithm.

References

1. Chang, C.-L.: Metric 1-median selection with fewer queries. Technical Report arXiv: 1612.08654 (2016)
2. Chang, C.-L.: Metric 1-median selection: query complexity vs. approximation ratio. ACM Trans. Comput. Theory **9**(4), Article 20 (2018)
3. Indyk, P.: Sublinear time algorithms for metric space problems. In: Vitter, J., Larmore, L, Leighton, F. (eds.) The 31st Annual ACM Symposium on Theory of Computing 1999, pp. 428–434. ACM, New York (1999). https://doi.org/10.1145/301250.301366
4. Indyk, P.: High-dimensional computational geometry. Ph.D. thesis, Stanford University (2000)