Metric 1-median selection: Query complexity vs. approximation ratio

Ching-Lueh Chang *
†

September 21, 2015

Abstract
Consider the problem of finding a point in a metric space \( \{1, 2, \ldots, n\}, d \) with the minimum average distance to other points. We show that this problem has no deterministic \( o(n^{1+1/(h-1)}) \)-query \( (2h - \Omega(1)) \)-approximation algorithms for any constant \( h \in \mathbb{Z}^+ \setminus \{1\} \).

1 Introduction

The metric 1-median problem asks for a point in an \( n \)-point metric space with the minimum average distance to other points. It has a Monte-Carlo \( O(n/\epsilon^2) \)-time \( (1 + \epsilon) \)-approximation algorithm for all \( \epsilon > 0 \) [6, 7]. In \( \mathbb{R}^D \), Kumar et al. [8] give a Monte-Carlo \( O(2^{\text{poly}(1/\epsilon)}D) \)-time \( (1 + \epsilon) \)-approximation algorithm for 1-median selection and another algorithm for \( k \)-median selection, where \( D \geq 1 \) and \( \epsilon > 0 \). Guha et al. [5] give streaming approximation algorithms for \( k \)-median selection in metric spaces.

Chang [3], Wu [11] and Chang [1] show that metric 1-median has a deterministic nonadaptive \( O(n^{1+1/h}) \)-time \( (2h) \)-approximation algorithm for all constants \( h \in \mathbb{Z}^+ \setminus \{1\} \). Furthermore, Chang [4] shows the nonexistence of deterministic \( o(n^2) \)-time \( (4 - \Omega(1)) \)-approximation algorithms for metric 1-median. This paper generalizes his result to show that metric 1-median has no deterministic \( o(n^{1+1/(h-1)}) \)-query \( (2h - \Omega(1)) \)-approximation algorithms for any constant \( h \in \mathbb{Z}^+ \setminus \{1\} \). Combining our result with an existing upper bound [1, 11],

\[
\min \{ c \geq 1 \mid \text{metric 1-median has a deterministic } O(n^{1+\epsilon}) \text{-query } c \text{-approx. alg.} \} = \min \{ c \geq 1 \mid \text{metric 1-median has a deterministic } O(n^{1+\epsilon}) \text{-time } c \text{-approx. alg.} \} = 2 \left\lceil \frac{1}{\epsilon} \right\rceil
\]

*Department of Computer Science and Engineering, Yuan Ze University, Taoyuan, Taiwan. Email: clchang@saturn.yzu.edu.tw
†Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taoyuan, Taiwan.
for all constants \(\epsilon \in (0, 1)\). That is, we determine the best approximation ratio of deterministic \(O(n^{1+\epsilon})\)-query (resp., \(O(n^{1+\epsilon})\)-time) algorithms for all \(\epsilon \in (0, 1)\).

As in the previous lower bounds for deterministic algorithms \([2, 4]\), we use an adversarial method. Roughly speaking, our proof proceeds as follows:

(i) Design an adversary \(\text{Adv}\) for answering the distance queries of any deterministic algorithm \(A\) with query complexity \(q(n) = o(n^{1+1/(h-1)})\).

(ii) Show that \(A\)'s output has a large average distance to other points, according to \(\text{Adv}\)'s answers to \(A\).

(iii) Construct a distance function with respect to which a certain point \(\hat{\alpha}\) has a small average distance to other points.

(iv) Construct the final distance function \(d(\cdot, \cdot)\) similar to that in item (iii).

(v) Show that \(d\) is a metric.

(vi) Show the consistency of \(d(\cdot, \cdot)\) with \(\text{Adv}\)'s answers.

(vii) Compare \(\hat{\alpha}\) in item (iii) with \(A\)'s output to establish our lower bound on \(A\)'s approximation ratio.

Central to our constructions are two graph sequences, \(\{H^{(i)}\}_{i=0}^{q(n)}\) and \(\{G^{(i)}\}_{i=0}^{q(n)}\) in Sec. 3, that are unseen in previous lower bounds \([2, 4, 9]\). Like in \([4]\), we need a small set \(S\) of points whose distances to other points are answered as large values during \(A\)'s execution, and yet we assign a small value to the distances from a certain point \(\hat{\alpha} \in S\) to many other points in item (iii).

This paper is organized as follows. Sec. 2 introduces the terminologies. Sec. 3 proves our main theorem that \textsc{metric 1-median} has no deterministic \(o(n^{1+1/(h-1)})\)-query \((2h-\Omega(1))\)-approximation algorithms for any constant \(h \in \mathbb{Z}^+ \setminus \{1\}\). In particular, Secs. 3.1, 3.2, 3.3 and 3.4 correspond to items (ii), (iii), (iv)–(vi) and (vii) above, respectively.

2 Definitions

A finite metric space \((M, d)\) is a finite set \(M\) endowed with a function \(d: M^2 \to [0, \infty)\) such that

- \(d(x, x) = 0\),
- \(d(x, y) > 0\) if \(x \neq y\),
- \(d(x, y) = d(y, x)\), and
- \(d(x, y) + d(y, z) \geq d(x, z)\)
for all \(x, y, z \in M\) \cite{10}. For all \(c \geq 1\), a point \(z \in M\) is said to be a \(c\)-approximate 1-median of \((M, d)\) if

\[
\sum_{x \in M} d(z, x) \leq c \cdot \sum_{x \in M} d(y, x)
\]

for all \(y \in M\). For convenience, \([n] \overset{\text{def.}}{=} \{1, 2, \ldots, n\}\).

For deterministic algorithms \(A\) and \(O\): \(\{1, 2, \ldots, n\}^2 \rightarrow \mathbb{R}\), denote by \(A^O(1^n)\) the execution of \(A\) with oracle access to \(O\) and with input \(1^n\), where \(n \in \mathbb{N}\). As the input to \(A\) will be \(1^n\) throughout this paper, abbreviate \(A^O(1^n)\) as \(A^O\). If \(A^d\) outputs a \(c\)-approximate 1-median of \(([n], d)\) for each finite metric space \(([n], d)\), then \(A\) is said to be \(c\)-approximate for METRIC 1-MEDIAN, where \(c \geq 1\).

**Fact 1** ([1, 3, 11]). For each constant \(h \in \mathbb{Z}^+ \setminus \{1\}\), METRIC 1-MEDIAN has a deterministic nonadaptive \(O(n^{1+1/h})\)-time \((2h)\)-approximation algorithm.

A weighted undirected graph \(G = (V, E, w)\) has a finite vertex set \(V\), an edge set \(E\) and a weight function \(w: E \rightarrow (0, \infty)\), where each edge is an unordered pair of distinct vertices in \(V\). If \(w: Y \rightarrow (0, \infty)\) for a superset \(Y\) of \(E\), interpret \((V, E, w)\) simply as \((V, E, w|_E)\), where \(w|_E\) denotes the restriction of \(w\) on \(E\). For all \(v \in V\), let

\[
N_G(v) \overset{\text{def.}}{=} \{ u \in V \mid (u, v) \in E \}
\]

and \(\text{deg}_G(v) \overset{\text{def.}}{=} |N_G(v)|\). For all \(S \subseteq V\), \(N_G(S) \overset{\text{def.}}{=} \bigcup_{v \in S} N_G(v)\). For all \(s, t \in V\), an \(s-t\) path \(P\) in \(G\) is a sequence \(\{v_i \in V\}_{i=0}^{k}\) satisfying \(k \in \mathbb{N}\), \(v_0 = s\), \(v_k = t\) and \((v_i, v_{i+1}) \in E\) for all \(i \in \{0, 1, \ldots, k-1\}\). Its weight (or length) is \(w(P) \overset{\text{def.}}{=} \sum_{i=0}^{k-1} w(v_i, v_{i+1})\).\(^1\) The shortest \(s-t\) distance in \(G\) is

\[
d_G(s, t) = \inf \{ w(P) \mid P \text{ is an } s-t \text{ path in } G \},
\]

where \(s, t \in V\). So \(d_G(s, t) = \infty\) if \(G\) has no \(s-t\) paths. Note that we allow only positive weights, i.e., \(\text{Im}(w) \subseteq (0, \infty)\). So a shortest \(s-t\) path must be simple, i.e., it does not repeat vertices. If \(w \equiv 1\), abbreviate \((V, E, w)\) as \((V, E)\) and call it an unweighted graph.

The following fact is well-known.

**Fact 2.** For each undirected graph \(G = (V, E)\),

\[
\sum_{v \in V} \text{deg}_G(v) = 2 \cdot |E|.
\]

For a predicate \(P\), let \(\chi[P] = 1\) if \(P\) is true and \(\chi[P] = 0\) otherwise. The following fact about geometric series is not hard to see.

**Fact 3.** For all \(r \geq 2\) and \(m \in \mathbb{N}\),

\[
\sum_{k=0}^{m} r^k \leq 2r^m.
\]

\(^1\)\(w(P)\) is a common and convenient abuse of notation.
3 Query complexity vs. approximation ratio

Throughout this section,

- $n \in \mathbb{Z}^+$,
- $\delta \in (0, 1)$ and $h \in \mathbb{Z}^+ \setminus \{1\}$ are constants (i.e., they are independent of $n$),
- $A$ is a deterministic $o(n^{1+1/(h-1)})$-query algorithm for METRIC 1-MEDIAN, and
- $S = [\lfloor \delta n \rfloor] \subseteq [n]$. \\

All pairs in $[n]^2$ are assumed to be unordered in this section. So, e.g., $(1, 2) \in \{2\} \times [n]$. By padding at most $n - 1$ dummy queries, assume without loss of generality that $A$ will have queried for the distances between its output and all other points when halting. Denote $A$’s query complexity by

$$q(n) = o\left(n^{1+1/(h-1)}\right).$$

Without loss of generality, forbid making the same query twice or querying for the distance from a point to itself, where the queries for $d(x, y)$ and $d(y, x)$ are considered to be the same for $x, y \in [n]$. Furthermore, let $n$ be sufficiently large to satisfy

1. $q(n) \leq \delta n^{1+1/(h-1)}$, 
2. $\frac{2q(n)}{|S| - 1} \leq \delta n^{1/(h-1)}$.

Define two unweighted undirected graphs $G^{(0)}$ and $H^{(0)}$ by

$$E^{(0)}_G \overset{\text{def.}}{=} \{(u, v) \mid (u, v \in [n] \setminus S) \wedge (u \neq v)\},$$

$$G^{(0)} \overset{\text{def.}}{=} ([n], E^{(0)}_G),$$

$$E^{(0)}_H \overset{\text{def.}}{=} \emptyset,$$

$$H^{(0)} \overset{\text{def.}}{=} ([n], E^{(0)}_H).$$

Algorithm $\text{Adv}$ in Fig. 1 answers $A$’s queries. In particular, for all $i \in [q(n)]$, the $i$th iteration of the loop of $\text{Adv}$ answers the $i$th query of $A$, denoted $(a_i, b_i) \in [n]^2$. It constructs three unweighted undirected graphs, $G^{(i)} = ([n], E^{(i)}_G)$, $H^{(i)} = ([n], E^{(i)}_H)$ and $Q^{(i)}$. As $G^{(i-1)}$ is unweighted for all $i \in [q(n)]$, $P_i$ in line 5 of $\text{Adv}$ is an $a_i$-$b_i$ path in $G^{(i-1)}$ with the minimum number of edges. By line 16 of $\text{Adv}$, the edges of $Q^{(i)}$ are precisely the first $i$ queries of $A$. 


Let $E^{(0)}_G, G^{(0)}, E^{(0)}_H$ and $H^{(0)}$ be as in equations (4)–(7);

1. for $i = 1, 2, \ldots, q(n)$ do
2. Receive the $i$th query of $A$, denoted $(a_i, b_i)$;
3. if $d_{G^{(i-1)}}(a_i, b_i) \leq h$ then
4. Find a shortest $a_i$-$b_i$ path $P_i$ in $G^{(i-1)}$;
5. $E^{(i)}_H \leftarrow E^{(i-1)}_H \cup \{e \mid e$ is an edge on $P_i\}$;
6. $H^{(i)} \leftarrow ([n], E^{(i)}_H)$;
7. $E^{(i)}_G \leftarrow E^{(i-1)}_G \setminus \{(u, v) \in E^{(i-1)}_G \setminus E^{(i)}_H \mid (\deg_{H^{(i)}}(u) \geq \delta n^{1/(h-1)} - 2) \lor (\deg_{H^{(i)}}(v) \geq \delta n^{1/(h-1)} - 2)\}$;
8. $G^{(i)} \leftarrow ([n], E^{(i)}_G)$;
9. else
10. $E^{(i)}_H \leftarrow E^{(i-1)}_H$;
11. $H^{(i)} \leftarrow ([n], E^{(i)}_H)$;
12. $E^{(i)}_G \leftarrow E^{(i-1)}_G$;
13. $G^{(i)} \leftarrow ([n], E^{(i)}_G)$;
14. end if
15. $Q^{(i)} \leftarrow ([n], \{(a_j, b_j) \mid j \in \{i\}\})$;
16. Output min$\{d_{H^{(i)}}(a_i, b_i), h - (1/2) \cdot \chi[\exists v \in \{a_i, b_i\}, (v \in S) \land (\deg_{Q^{(i)}}(v) \leq \delta n^{1/(h-1)})]\}$ as the answer to the $i$th query of $A$;
17. end for

Figure 1: Algorithm Adv for answering $A$’s queries
Lemma 4.

\[ E_H^{(0)} \subseteq E_H^{(1)} \subseteq \ldots \subseteq E_H^{(q(n))} \subseteq E_G^{(q(n))} \subseteq E_G^{(q(n)-1)} \subseteq \ldots \subseteq E_G^{(0)} . \]

**Proof.** By lines 6 and 11 of Adv in Fig. 1, \( E_H^{(i-1)} \subseteq E_H^{(i)} \) for all \( i \in [q(n)] \). By lines 8 and 13, \( E_G^{(i)} \subseteq E_G^{(i-1)} \) for all \( i \in [q(n)] \).

To show that \( E_H^{(q(n))} \subseteq E_G^{(q(n))} \), we shall prove the stronger statement that \( E_H^{(i)} \subseteq E_G^{(i)} \) for all \( i \in \{0, 1, \ldots, q(n)\} \) by mathematical induction. By equation (6), \( E_H^{(0)} \subseteq E_G^{(0)} \). Assume as the induction hypothesis that \( E_H^{(i-1)} \subseteq E_G^{(i-1)} \).

The following shows that \( E_H^{(i)} \subseteq E_G^{(i-1)} \) by examining each \( e \in E_H^{(i)} \):

Case 1: \( e \in E_H^{(i-1)} \). By the induction hypothesis, \( e \in E_G^{(i-1)} \).

Case 2: \( e \notin E_H^{(i-1)} \). As \( e \in E_H^{(i)} \setminus E_H^{(i-1)} \), lines 6 and 11 show that \( e \) is on \( P_i \) (and that the \( i \)th iteration of the loop of Adv runs line 6 rather than line 11). By line 5, each edge on \( P_i \) is in \( E_G^{(i-1)} \). In particular, \( e \in E_G^{(i-1)} \).

Having shown that \( E_H^{(i)} \subseteq E_G^{(i-1)} \), lines 8 and 13 will both result in \( E_H^{(i)} \subseteq E_G^{(i)} \), completing the induction step. □

**Lemma 5.** For all \( i \in [q(n)] \) with \( d_G^{(i-1)}(a_i, b_i) \leq h \),

\[ d_H^{(i)}(a_i, b_i) = d_H^{(q(n))}(a_i, b_i) = d_G^{(q(n))}(a_i, b_i) = d_G^{(i-1)}(a_i, b_i) . \]

**Proof.** By line 4 of Adv, the \( i \)th iteration of the loop runs lines 5–9. Lines 5–7 put (the edges of) a shortest \( a_i \)-\( b_i \) path in \( G^{(i-1)} \) into \( H^{(i)} \); hence

\[ d_H^{(i)}(a_i, b_i) \leq d_G^{(i-1)}(a_i, b_i) . \]

This and Lemma 4 complete the proof. □

Below is an easy consequence of Lemma 4.

**Lemma 6.** For all \( i \in [q(n)] \) with \( d_G^{(i-1)}(a_i, b_i) > h \),

\[ d_G^{(q(n))}(a_i, b_i) > h . \]

3.1 The average distance from \( A \)'s output to other points

This subsection shows that the output of \( A^{\text{Adv}} \) has a large average distance to other points, according to the answers of Adv.

**Lemma 7.** For all \( i \in [q(n)] \) and \( v \in [n] \),

\[ \deg_{H^{(i)}}(v) \leq \deg_{H^{(i-1)}}(v) + 2 . \]
Proof. If the $i$th iteration of the loop of $\text{Adv}$ runs lines 11–14 but not 5–9, then $H^{(i)} = H^{(i-1)}$, proving the lemma. So assume otherwise. Being shortest, $P_i$ in line 5 does not repeat vertices. Therefore, $v$ is incident to at most two edges on $P_i$, which together with lines 6–7 complete the proof. \hfill \qed

**Lemma 8.** For all $v \in [n]$, \[ \deg_{H^{(q(n))}}(v) < \delta n^{1/(h-1)}. \]

**Proof.** Assume \[ \deg_{H^{(q(n))}}(v) \geq \delta n^{1/(h-1)} - 2 \] for, otherwise, there is nothing to prove. Clearly, \[ \deg_{H^{(0)}}(v) \overset{(6)-(7)}{=} 0 \overset{(2)}{<} \delta n^{1/(h-1)} - 2. \] By inequalities (8)–(9), there exists $i \in [q(n)]$ satisfying \[ \deg_{H^{(i-1)}}(v) < \delta n^{1/(h-1)} - 2, \] \[ \deg_{H^{(i)}}(v) \geq \delta n^{1/(h-1)} - 2. \] Clearly, \[ \text{NG}(i)(v) = \left\{ u \in [n] \mid (u,v) \in E_G^{(i)} \right\}. \] As $H^{(i-1)} \neq H^{(i)}$ by inequalities (10)–(11), the $i$th iteration of the loop of $\text{Adv}$ runs lines 5–9 but not 11–14. By inequality (11) and line 8 of $\text{Adv}$, \[ \left\{ u \in [n] \mid (u,v) \in E_G^{(i)} \right\} = \left\{ u \in [n] \mid (u,v) \in E_G^{(i-1)} \setminus \left( E_G^{(i-1)} \setminus E_H^{(i)} \right) \right\}. \] Equations (12)–(13) and Lemma 4 give \[ \text{NG}(i)(v) = \left\{ u \in [n] \mid (u,v) \in E_H^{(i)} \right\}. \] By inequality (10) and Lemma 7, \[ \deg_{H^{(i)}}(v) < \delta n^{1/(h-1)}. \] This and equation (14) imply $\deg_{G(i)}(v) < \delta n^{1/(h-1)}$, which together with Lemma 4 completes the proof. \hfill \qed

**Lemma 9.** For all $v \in [n]$, \[ \left| \{ u \in [n] \mid d_{H^{(q(n))}}(v,u) < h \} \right| \leq 2\delta^{h-1}n. \]
Proof. By Lemma 8,

\[
\left| \{ u \in [n] \mid \exists \text{ } v-u \text{ path in } H^{(q(n))} \text{ with exactly } k \text{ edges} \} \right| \leq \left( \delta n^{1/(h-1)} \right)^k
\]

for all \( k \in \mathbb{N} \). Consequently,

\[
\left| \{ u \in [n] \mid \exists \text{ } v-u \text{ path in } H^{(q(n))} \text{ with at most } h-1 \text{ edges} \} \right| \leq \left( \delta n^{1/(h-1)} \right)^k \sum_{k=0}^{h-1} \left( \delta n^{1/(h-1)} \right)^k
\]

(2) and Fact 3 \leq 2\delta^{h-1}n.

Finally, recall that \( H^{(q(n))} \) is unweighted.

Denote the output of \( A^{\text{Adv}} \) by \( z \). Furthermore,

\[
I \overset{\text{def.}}{=} \{ j \in [q(n)] \mid z \in \{ a_j, b_j \} \}.
\]

(15)

The following lemma analyzes the sum of the distances, as answered by line 17 of \( A^{\text{Adv}} \), from \( z \) to other points.

**Lemma 10.**

\[
\sum_{i \in I} \min \left\{ d_{H^{(q(n))}}(a_i, b_i), h - \frac{1}{2} \cdot \chi \left[ \exists v \in \{ a_i, b_i \}, (v \in S) \land \left( \deg_{Q(i)}(v) \leq \delta n^{1/(h-1)} \right) \right] \right\}
\]

\[
\geq n \cdot \left( h - 2h\delta^{h-1} - o(1) - \delta \right).
\]

Proof. By Lemma 4,

\[
\sum_{i \in I} \min \left\{ d_{H^{(q(n))}}(a_i, b_i), h - \frac{1}{2} \cdot \chi \left[ \exists v \in \{ a_i, b_i \}, (v \in S) \land \left( \deg_{Q(i)}(v) \leq \delta n^{1/(h-1)} \right) \right] \right\}
\]

(16)

\[
\geq \sum_{i \in I} \min \left\{ d_{H^{(q(n))}}(a_i, b_i), h \right\}
\]

\[
- \sum_{i \in I} \frac{1}{2} \cdot \chi \left[ \exists v \in \{ a_i, b_i \}, (v \in S) \land \left( \deg_{Q(i)}(v) \leq \delta n^{1/(h-1)} \right) \right].
\]

For all \( i \in I \), there exists \( c_i \in [n] \) with \( \{ z, c_i \} = \{ a_i, b_i \} \) by equation (15). Therefore,

\[
\sum_{i \in I} \min \left\{ d_{H^{(q(n))}}(a_i, b_i), h \right\} = \sum_{i \in I} \min \left\{ d_{H^{(q(n))}}(z, c_i), h \right\}.
\]

As we forbid repeated queries, \( \{ c_i \}_{i \in I} \) is a sequence of distinct points. So by Lemma 9,

\[
\sum_{i \in I} \min \left\{ d_{H^{(q(n))}}(z, c_i), h \right\} = h \cdot (|I| - 2\delta^{h-1}n).
\]
Recall that $A^{Adv}$ will have queried for the distances between its output (which is $z$) and all other points when halting. So

$$|I| \geq n - 1$$

by equation (15).^2

Clearly,

$$\sum_{i \in I} \chi \left[ \exists v \in \{a_i, b_i\}, (v \in S) \land (\deg_{Q(i)}(v) \leq \delta n^{1/(h-1)}) \right]$$

$$= \sum_{i \in I} \chi \left[ \exists v \in \{z, c_i\}, (v \in S) \land (\deg_{Q(i)}(v) \leq \delta n^{1/(h-1)}) \right]$$

$$\leq \sum_{i \in I} \chi \left[ (z \in S) \land (\deg_{Q(i)}(z) \leq \delta n^{1/(h-1)}) \right]$$

$$+ \sum_{i \in I} \chi \left[ (c_i \in S) \land (\deg_{Q(i)}(c_i) \leq \delta n^{1/(h-1)}) \right].$$

By line 16 of $Adv$ and equation (15),

$$\deg_{Q(i)}(z) = |\{j \in I \mid j \leq i\}|.$$ 

Therefore,

$$\sum_{i \in I} \chi \left[ (z \in S) \land (\deg_{Q(i)}(z) \leq \delta n^{1/(h-1)}) \right] \leq \sum_{i \in I} \chi \left[ |\{j \in I \mid j \leq i\}| \leq \delta n^{1/(h-1)} \right]$$

$$\leq \delta n^{1/(h-1)},$$

where the last inequality follows because $|\{j \in I \mid j \leq i\}| = k$ when $i$ is the $k$th smallest element of $I$, for all $k \in [|I|]$. Recall the distinctness of the points in $\{c_i\}_{i \in I}$. Therefore,

$$\sum_{i \in I} \chi \left[ (c_i \in S) \land (\deg_{Q(i)}(c_i) \leq \delta n^{1/(h-1)}) \right] \leq \sum_{i \in I} \chi [c_i \in S] \leq |S| = \lfloor \delta n \rfloor. \quad (17)$$

Inequalities (16)–(17) complete the proof.

3.2 Planting a point with a small average distance to other points

This subsection constructs a distance function with respect to which a certain point has an average distance of approximately $1/2$ to other points.

^2Because we forbid repeated queries and queries for the distance from a point to itself, we also have $|I| \leq n - 1$. 
Lemma 11.

\[ |E_{H}^{(q(n))}| \leq h \cdot q(n). \]

**Proof.** Consider the \( i \)-th iteration of the loop of \( \text{Adv} \), where \( i \in [q(n)] \).

- Running lines 4–5 results in \( P \) having at most \( h \) edges. Consequently,
  \[ |E_{H}^{(i)}| \leq |E_{H}^{(i-1)}| + h \]  
  by line 6.

- Running line 11 yields \( |E_{H}^{(i)}| = |E_{H}^{(i-1)}| \), implying inequality (18) as well.

Now,

\[ |E_{H}^{(q(n))}| - |E_{H}^{(0)}| = \sum_{i=1}^{q(n)} \left( |E_{H}^{(i)}| - |E_{H}^{(i-1)}| \right) \overset{(18)}{\leq} h \cdot q(n). \]

Finally, \( |E_{H}^{(0)}| = 0 \) by equation (6).

\[ \square \]

Lemma 12.

\[ \left| \{ u \in [n] \mid \deg_{H^{(q(n))}}(u) \geq \delta n^{1/(h-1)} - 2 \} \right| = \frac{h}{\delta} \cdot o(n). \]

**Proof.** By Fact 2, the average degree in \( H^{(q(n))} \) is

\[ \frac{1}{n} \cdot 2 \cdot \left| E_{H}^{(q(n))} \right|. \]

So by the averaging argument (that any finite nonempty sequence of nonnegative numbers with average \( \bar{a} \) has at most an \( \bar{a}/t \) fraction of numbers that are greater than or equal to \( t > 0 \)),

\[ \frac{1}{n} \cdot \left| \{ u \in [n] \mid \deg_{H^{(q(n))}}(u) \geq \delta n^{1/(h-1)} - 2 \} \right| \leq \frac{1}{n} \cdot 2 \cdot |E_{H}^{(q(n))}| \cdot \frac{1}{\delta n^{1/(h-1)} - 2}, \]

where the rightmost denominator is positive and is \( \Theta(\delta n^{1/(h-1)}) \) by equation (2).

This and Lemma 11 complete the proof.

By inequality (2), \( S \setminus \{ z \} \neq \emptyset \). Let

\[ \hat{\alpha} \overset{\text{def.}}{=} \arg\min_{\alpha \in S \setminus \{ z \}} \deg_{Q^{(q(n))}}(\alpha), \]

breaking ties arbitrarily.

---

\(^3\)We explicitly write down the constants \( h \) and \( \delta \) on the right-hand side for clarity, although they can be absorbed within \( o(\cdot) \).
Lemma 13. For all $i \in [q(n)]$,
\[
\deg_{Q(i)}(\hat{\alpha}) \leq \delta n^{1/(h-1)}.
\]

Proof. By line 16 of \textit{Adv},
\[
\deg_{Q(i)}(\hat{\alpha}) \leq \deg_{Q(q(n))}(\hat{\alpha}).
\tag{20}
\]

By equation (19) and the averaging argument,
\[
\deg_{Q(q(n))}(\hat{\alpha}) \leq \frac{1}{|S \setminus \{z\}|} \cdot \sum_{\alpha \in S \setminus \{z\}} \deg_{Q(q(n))}(\alpha).
\]

Furthermore,
\[
\sum_{\alpha \in S \setminus \{z\}} \deg_{Q(q(n))}(\alpha) \leq \sum_{\alpha \in [n]} \deg_{Q(q(n))}(\alpha) = 2q(n),
\tag{21}
\]

where the equality follows from Fact 2, line 16 of \textit{Adv} and the non-repeating of queries. Finally,
\[
\deg_{Q(i)}(\hat{\alpha}) \stackrel{(20)-(21)}{\leq} 2q(n) \frac{1}{|S| - 1} \stackrel{(3)}{\leq} \delta n^{1/(h-1)}.
\]

Inductively, let
\[
V_0 \overset{\text{def.}}{=} \{\hat{\alpha}\},
\tag{22}
V_1 \overset{\text{def.}}{=} N_{Q(q(n))}(\hat{\alpha}) \setminus V_0,
\tag{23}
V_{j+1} \overset{\text{def.}}{=} N_{H(q(n))}(V_j) \setminus \left( \bigcup_{i=0}^{j} V_i \right)
\tag{24}
\]

for all $j \in [h-2]$. Furthermore,
\[
V_h \overset{\text{def.}}{=} [n] \setminus \left( \bigcup_{i=0}^{h-1} V_i \right).
\tag{25}
\]

The following lemma is not hard to see from equations (22)–(25).

Lemma 14. $(V_0, V_1, \ldots, V_h)$ is a partition of $[n]$, i.e., $\bigcup_{k=0}^{h} V_k = [n]$ and $V_i \cap V_j = \emptyset$ for all distinct $i, j \in \{0, 1, \ldots, h\}$.
Let
\[ B = \{ u \in [n] \mid \deg_{H^{(q(n))}(u)} \geq \delta n^{1/(h-1)} - 2 \} , \]  
\[ \mathcal{E} \overset{\text{def.}}{=} \left[ E_{G^{(q(n))}} \left( \bigcup_{i, j \in \{0, 1, \ldots, h\}, |i-j| \geq 2} V_i \times V_j \right) \right] \cup \left( \{ \hat{\alpha} \} \times (V_h \setminus (B \cup S)) \right) . \]  
By equation (19), \( \hat{\alpha} \notin V_h \setminus (B \cup S) \), which together with equation (4) and Lemma 4 forbids any edge in \( \mathcal{E} \) from being a self-loop. For all distinct \( u, v \in [n] \),
\[ w(u, v) \overset{\text{def.}}{=} \begin{cases} 1/2, & \text{if one of } u \text{ and } v \text{ is } \hat{\alpha} \text{ and the other is in } V_h \setminus (B \cup S), \\ 1, & \text{otherwise}. \end{cases} \]  
Furthermore, let
\[ \mathcal{G} \overset{\text{def.}}{=} ([n], \mathcal{E}, w) \]  
be a weighted undirected graph.

**Lemma 15.**
\[ \sum_{j=1}^{h-1} |V_j| \leq 2\delta^{h-1}n. \]

**Proof.** By Lemma 8 and equation (24),
\[ |V_{j+1}| \leq |V_j| \cdot \delta n^{1/(h-1)} \]
for all \( j \in [h-2] \). Therefore, \( \sum_{j=1}^{h-1} |V_j| \) is bounded from above by the \( (h-1) \)-term geometric series with the common ratio of \( \delta n^{1/(h-1)} \) and the initial value of \( |V_1| \). Consequently,
\[ \sum_{j=1}^{h-1} |V_j|^{(2) \text{ and Fact } 3} \leq 2 \cdot |V_1| \cdot \delta^{h-2}n^{(h-2)/(h-1)}. \]  
By Lemma 13, \( |N_{Q^{(q(n))}}(\hat{\alpha})| \leq \delta n^{1/(h-1)} \). So by equation (23), we have \( |V_1| \leq \delta n^{1/(h-1)} \), which together with inequality (30) completes the proof.

**Lemma 16.**
\[ |V_h \setminus (B \cup S)| \geq n \left( 1 - 2\delta^{h-1} - \frac{h}{\delta} \cdot o(1) - \delta \right) . \]

**Proof.** By Lemma 12 and equation (26), \( |B| = (h/\delta) \cdot o(n) \). By construction, \( |S| = [\delta n] \). Finally,
\[ |V_h| \overset{\text{Lemmas } 14-15}{\geq} n - 2\delta^{h-1}n - |V_0| \overset{(22)}{=} n - 2\delta^{h-1}n - 1. \]  
\[ \square \]
The following lemma says that $\hat{\alpha}$ has an average distance of approximately $1/2$ to other points w.r.t. the distance function $\min\{d_G(\cdot, \cdot), h\}$.

Lemma 17.

$$\sum_{v \in [n]} \min\{d_G(\hat{\alpha}, v), h\} \leq n \cdot \left(\frac{1}{2} + 2h\delta^{-1} + \frac{h^2}{\delta} \cdot o(1) + h\delta\right).$$

Proof. By equations (27)–(29), $d_G(\hat{\alpha}, v) \leq 1/2$ for all $v \in V_h \setminus (B \cup S)$. This and Lemma 16 complete the proof.

3.3 A metric consistent with Adv’s answers

This subsection constructs a metric $d: [n]^2 \to [0, \infty)$ consistent with Adv’s answers in line 17. So Lemma 10 will require $z$, which is the output of $A_{Adv}$, to have an average distance (w.r.t. $d$) of at least approximately $h$ to other points. Although $d(\cdot, \cdot)$ will not be exactly $\min\{d_G(\cdot, \cdot), h\}$, Lemma 17 will forbid $\sum_{v \in [n]} d(\hat{\alpha}, v)/n$ from exceeding $1/2$ by too much. Details follow.

Recall that $H_i$ and $G_i$ are unweighted for all $i \in \{0, 1, \ldots, q(n)\}$. They can be treated as having the weight function $w$ while preserving $d_{H(i)}(\cdot, \cdot)$ and $d_{G(i)}(\cdot, \cdot)$, as shown by the lemma below.

Lemma 18. For all $i \in \{0, 1, \ldots, q(n)\}$, each path $P$ in $H(i)$ or $G(i)$ has exactly $w(P)$ edges.

Proof. As $\hat{\alpha} \in S$ by equation (19), equation (28) implies $w(u, v) = 1$ for all distinct $u, v \in [n] \setminus S$. This and equation (4) imply that all edges in $E_G(0)$ have weight 1 w.r.t. $w$. So by Lemma 4, the edges in $E_H(0) \cup E_G(0)$ have weight 1 w.r.t. $w$. Finally, recall that $H(i) = ([n], E_H(i))$ and $G(i) = ([n], E_G(i))$.

We now show that $H(q(n))$ has an edge in $V_i \times V_j$ only if $|i - j| \leq 1$.

Lemma 19.

$$E_H^{(q(n))} \cap \left(\bigcup_{i,j \in \{0,1,\ldots,h\}, |i-j| \geq 2} V_i \times V_j\right) = \emptyset.$$ 

Proof. Suppose for contradiction that there exists $e \in E_H^{(q(n))}$ with an endpoint in $V_k$ and the other in $V_{\ell}$, where $k, \ell \in \{0, 1, \ldots, h\}$ and $\ell \geq k + 2$. Then $N_{H^{(q(n))}}(V_k) \cap V_{\ell} \neq \emptyset$, which together with Lemma 14 and $\ell \geq k + 2$ implies

$$N_{H^{(q(n))}}(V_k) \nsubseteq \bigcup_{j=0}^{k+1} V_j.$$ 

As $\ell \geq k + 2$ and $k, \ell \in \{0, 1, \ldots, h\}$, we have $0 \leq k \leq h - 2$. 

13
Case 1: \( k = 0 \). By equations (19) and (22), \( V_0 \subseteq S \). So \( N_{G(0)}(V_0) = \emptyset \) by equations (4)–(5). Consequently, \( N_{H(0)(\alpha)}(V_0) = \emptyset \) by Lemma 4, contradicting relation (31).

Case 2: \( k \in [h - 2] \). Relation (31) contradicts equation (24) (with \( j \leftarrow k \)).

A contradiction occurs in either case. \( \square \)

Lemma 20. \( E^{(q(n))}_{H} \subseteq E \).

Proof. By Lemma 19 and equation (27), \( E^{(q(n))}_{G} \cap E^{(q(n))}_{H} \subseteq E \). This and Lemma 4 complete the proof. \( \square \)

Lemma 21. Let \( P \) be a path in \( G \) that visits no edges in \( \{ \tilde{\alpha} \} \times (V_h \setminus (B \cup S)) \). If the first and the last vertices of \( P \) are in \( V_h \) and \( V_1 \), respectively, then \( w(P) \geq h - 1 \).

Proof. By Lemma 14, \( \bigcup_{k=0}^{h} V_k = \{ n \} \), \( V_{i+1} \cap V_i = \emptyset \) and \( (V_{i+1} \times V_i) \cap (V_{j+1} \times V_j) = \emptyset \) for all distinct \( i, j \in [h] \). Because \( P \) is a path in \( G \), visiting no edges in \( \{ \tilde{\alpha} \} \times (V_h \setminus (B \cup S)) \), no edges on \( P \) are in \( V_i \times V_j \) for any \( i, j \in \{ 0, 1, \ldots, h \} \) with \( |i - j| \geq 2 \) by equations (27) and (29). This forces \( P \), which is a \( V_h \)-\( V_1 \) path, to visit at least one edge in \( V_{i+1} \times V_i \) for each \( i \in [h-1] \) (for a total of at least \( h - 1 \) edges). As \( \tilde{\alpha} \notin \bigcup_{i=1}^{h} V_i \) by equations (22)–(25), equation (28) gives \( w(u, v) = 1 \) for all \( (u, v) \in \bigcup_{i=1}^{n} V_{i+1} \times V_i \). We have shown that \( P \) has at least \( h - 1 \) edges of weight (w.r.t. \( w \)) 1. \( \square \)

We proceed to analyze shortest \( a_i \)-\( b_i \) paths in \( G \), where \( i \in [q(n)] \). Clearly, such paths must be simple.

Lemma 22. Let \( P \) be a shortest \( a_i \)-\( b_i \) path in \( G \), where \( i \in [q(n)] \). If \( P \) visits exactly one edge in \( \{ \tilde{\alpha} \} \times (V_h \setminus (B \cup S)) \) and \( \tilde{\alpha} = \{ a_i, b_i \} \), then \( w(P) \geq h - 1/2 \).

Proof. Being shortest, \( P \) must be simple. Assume \( \hat{\alpha} = a_i \) for now. Because \( P \) is a simple \( \hat{\alpha} \)-\( b_i \) path in \( G \) visiting exactly one edge in \( \{ \hat{\alpha} \} \times (V_h \setminus (B \cup S)) \), it can be decomposed into an edge \( (\hat{\alpha}, v) \), where \( v \in V_h \setminus (B \cup S) \), and a \( v \)-\( b_i \) path \( \hat{P} \) in \( G \) that visits no edges in \( \{ \hat{\alpha} \} \times (V_h \setminus (B \cup S)) \).\(^4\) As \( \hat{\alpha} = a_i \), we have \( b_i \in N_{Q(\alpha)}(\hat{\alpha}) \) by line 16 of Adv. So by equations (22)–(23), \( b_i \in V_1 \cup \{ \hat{\alpha} \} \), implying \( b_i \in V_1 \) because querying for the distance from a point to itself is forbidden and \( \hat{\alpha} = a_i \). In summary, \( \hat{P} \) is a path in \( G \), from \( v \in V_h \setminus (B \cup S) \) to \( b_i \in V_1 \), that visits no edges in \( \{ \hat{\alpha} \} \times (V_h \setminus (B \cup S)) \). So by Lemma 21 (with \( P \leftarrow \hat{P} \)),

\[
    w(\hat{P}) \geq h - 1. \tag{32}
\]

\(^4\)If the first edge on \( P \) is not in \( \{ \hat{\alpha} \} \times (V_h \setminus (B \cup S)) \), then \( P \)'s later visit of an edge in \( \{ \hat{\alpha} \} \times (V_h \setminus (B \cup S)) \) must make \( P \) non-simple, a contradiction.
As $v \in V_h$, we have $\hat{\alpha} \neq v$ by equations (22) and (25). By the construction of $\hat{P}$,

$$w(P) = w(\hat{\alpha}, v) + w(\hat{P}) \geq \frac{1}{2} + w(\hat{P}).$$

Inequalities (32)–(33) show that $w(P) \geq h - 1/2$. The case of $\hat{\alpha} = b_i$ is symmetric: Reverse $P$ and exchange all the above occurrences of “$a_i$” with “$b_i$.” □

**Lemma 23.** For all $i \in [q(n)]$ with $\hat{\alpha} \in \{a_i, b_i\}$,

$$\chi \left[ \exists v \in \{a_i, b_i\}, (v \in S) \wedge (\deg_{Q(i)}(v) \leq \delta n^{1/(h-1)}) \right] = 1.$$

**Proof.** By equation (19), $\hat{\alpha} \in S$. This and Lemma 13 complete the proof. □

**Lemma 24.** For all distinct $u, v \in [n] \setminus (B \cup S)$, we have $(u, v) \in E_G^{(q(n))})$.

**Proof.** As $u, v \in [n] \setminus B$, equation (26) implies

$$\deg_{H(i)}(u) < \delta n^{1/(h-1)} - 2,$$

$$\deg_{H(i)}(v) < \delta n^{1/(h-1)} - 2$$

when $i = q(n)$. So by Lemma 4, inequalities (34)–(35) hold for all $i \in [q(n)]$.

As $u, v \in [n] \setminus S$ and $u \neq v$, we have $(u, v) \in E_G^{(0)}$ by equation (4). By lines 8 and 13 of Adv,

$$E_G^{(i-1)} \setminus \left\{ (x, y) \in [n]^2 \mid \left( \deg_{H(i)}(x) \geq \delta n^{1/(h-1)} - 2 \right) \vee \left( \deg_{H(i)}(y) \geq \delta n^{1/(h-1)} - 2 \right) \right\} \subseteq E_G^{(i)}$$

for all $i \in [q(n)]$. By inequalities (34)–(35) and relation (36), $(u, v) \in E_G^{(i)}$ if $(u, v) \in E_G^{(i-1)}$, for all $i \in [q(n)]$. The proof is complete by mathematical induction. □

**Lemma 25.** Let $P$ be a shortest $a_i$-$b_i$ path in $G$, where $i \in [q(n)]$. If $P$ visits exactly two edges in $\{\hat{\alpha}\} \times (V_h \setminus (B \cup S))$, then $G^{(q(n))}$ has an $a_i$-$b_i$ path with exactly $w(P)$ edges.

**Proof.** Being shortest, $P$ must be simple. Therefore, the two edges of $P$ in $\{\hat{\alpha}\} \times (V_h \setminus (B \cup S))$, denoted $(u, \hat{\alpha})$ and $(\hat{\alpha}, v)$, are consecutive on $\hat{P}$. Clearly, $u \neq v$. Replace the subpath $(u, \hat{\alpha}, v)$ of $P$ by the edge $(u, v)$ to yield an $a_i$-$b_i$ path $\hat{P}$. Except for the two edges of $P$ in $\{\hat{\alpha}\} \times (V_h \setminus (B \cup S))$ (which are $(u, \hat{\alpha})$ and $(\hat{\alpha}, v)$), all edges of $P$ are in $E_G^{(q(n))}$ by equation (27) and $P$’s being a path in $G = ([n], E, w)$. As $u, v \in V_h \setminus (B \cup S)$ and $u \neq v$, $(u, v) \in E_G^{(q(n))}$ by Lemma 24. In summary, all the edges of $\hat{P}$ (including $(u, v)$ and the edges of $P$ not in $\{\hat{\alpha}\} \times (V_h \setminus (B \cup S))$) are in $E_G^{(q(n))}$. Consequently, $\hat{P}$ is an $a_i$-$b_i$ path in $G^{(q(n))} = ([n], E_G^{(q(n))})$. So we are left only to prove that $\hat{P}$ has exactly $w(P)$ edges.
edges, which, by Lemma 18 (with \( P \leftarrow \tilde{P} \) and \( i \leftarrow q(n) \)), is equivalent to proving \( w(\tilde{P}) = w(P) \).

Note that \( \tilde{\alpha} \notin V_h \setminus (B \cup S) \) by equation (19). By the construction of \( \tilde{P} \) and recalling that \( u, v \in V_h \setminus (B \cup S) \) and \( u \neq v \),

\[
w(\tilde{P}) = w(P) - w(u, \tilde{\alpha}) - w(\tilde{\alpha}, v) + w(u, v) \overset{(28)}{=} w(P) - \frac{1}{2} - \frac{1}{2} + 1 = w(P).
\]

\[\square\]

**Lemma 26.** Every simple path in \( G \) visiting exactly one edge in \( \{\tilde{\alpha}\} \times (V_h \setminus (B \cup S)) \) either starts or ends at \( \tilde{\alpha} \).

**Proof.** By equation (19), \( \tilde{\alpha} \in S \). So by equation (4) and Lemma 4, \( \tilde{\alpha} \) is incident to no edges in \( E(q(n)) \) of \( G \). Consequently, the set of all edges of \( G \) incident to \( \tilde{\alpha} \) is \( \{\tilde{\alpha}\} \times (V_h \setminus (B \cup S)) \) by equation (27). The lemma is now easy to see. \[\square\]

**Lemma 27.** For all \( i \in [q(n)] \),

\[
\min \left\{ d_{H(i)}(a_i, b_i), h - \frac{1}{2} \cdot \left[ \exists v \in \{a_i, b_i\}, (v \in S) \land \left( \deg_{Q(i)}(v) \leq \delta n^{1/(h-1)} \right) \right] \right\}
\leq \min \left\{ d_G(a_i, b_i), h - \frac{1}{2} \cdot \left[ \exists v \in \{a_i, b_i\}, (v \in S) \land \left( \deg_{Q(i)}(v) \leq \delta n^{1/(h-1)} \right) \right] \right\}.
\]

\[
(37)
\]

**Proof.** Assume the existence of an \( a_i-b_i \) path in \( G \) for, otherwise, \( d_G(a_i, b_i) = \infty \) and inequality (37) trivially holds. Pick any shortest \( a_i-b_i \) path \( P \) in \( G([n], \mathcal{E}, w) \). Clearly,

\[
w(P) = d_G(a_i, b_i) .
\]

(38)

Being shortest, \( P \) must be simple.

We establish inequality (37) in the following exhaustive cases:

**Case 1:** \( P \) visits no edges in \( \{\tilde{\alpha}\} \times (V_h \setminus (B \cup S)) \). By equation (27), all edges of \( P \) are in \( E_G(q(n)) \), i.e., \( P \) is a path in \( G(q(n)) \). So by Lemma 18 (with \( i \leftarrow q(n) \)), \( w(\tilde{P}) \) equals the length of \( P \) in the unweighted graph \( G(q(n)) \). Therefore,

\[
d_{G(q(n))}(a_i, b_i) \leq w(P).
\]

(39)

If \( d_{G(q(n))}(a_i, b_i) \leq h \), then

\[
d_{H(i)}(a_i, b_i) = d_{G(q(n))}(a_i, b_i)
\]

by Lemma 5. Otherwise, \( d_{G(q(n))}(a_i, b_i) > h \) by Lemma 6. In either case, equations (38)–(39) imply inequality (37).
Case 2: $P$ visits exactly one edge in $\{\hat{a}\} \times (V_h \setminus (B \cup S))$ and $\hat{a} \in \{a_i, b_i\}$. By Lemma 22 and equation (38), $d_G(a_i, b_i) \geq h - 1/2$. This and Lemma 23 force the right-hand side of inequality (37) to equal $h - 1/2$. By Lemma 23, the left-hand side of inequality (37) is less than or equal to $h - 1/2$. We have verified inequality (37).

Case 3: $P$ visits exactly one edge in $\{\hat{a}\} \times (V_h \setminus (B \cup S))$ and $\hat{a} \notin \{a_i, b_i\}$. A contradiction to Lemma 26 occurs.

Case 4: $P$ visits exactly two edges in $\{\hat{a}\} \times (V_h \setminus (B \cup S))$. Lemma 25 and that $G(q(n))$ is unweighted imply inequality (39). Proceeding as in Case 1, equations (38)–(39) and Lemmas 5–6 imply inequality (37) no matter $d_{G(i-1)}(a_i, b_i) \leq h$ or otherwise.

Case 5: $P$ visits at least three edges in $\{\hat{a}\} \times (V_h \setminus (B \cup S))$. Clearly, $P$ is non-simple, a contradiction.

Define $d: [n]^2 \to [0, \infty)$ by

$$d(a_i, b_i) = d(b_i, a_i)$$

$$d(u, v) = \min \left\{ d_G(a_i, b_i), h - \frac{1}{2} \cdot \chi \left( \exists v \in \{a_i, b_i\}, (v \in S) \land (\deg_G(v) \leq \delta n^{1/(h-1)}) \right) \right\}, \quad (40)$$

$$d(u, v) = \min \{ d_G(u, v), h \} \quad (41)$$

for all $i \in [q(n)]$ and $(u, v) \in [n]^2 \setminus \{(a_j, b_j) \mid j \in [q(n)]\}$. Because all pairs in $[n]^2$ are unordered in this section, $(b_i, a_i) \notin [n]^2 \setminus \{(a_j, b_j) \mid j \in [q(n)]\}$ for all $i \in [q(n)]$. Consequently, equation (41) does not redefine $d(b_i, a_i)$. Because $G$ is undirected, the right-hand side of equation (41) remains intact with $u$ and $v$ interchanged. As $A$ does not repeat queries, equation (40) defines $d(a_i, b_i)$ and $d(b_i, a_i)$ only once for each $i \in [q(n)]$ (note that forbidding repeated queries implies the nonexistence of distinct $i, j \in [q(n)]$ satisfying (1) $a_i = a_j$ and $b_i = b_j$ or (2) $a_i = b_j$ and $b_i = a_j$). It is now clear that $d(\cdot, \cdot)$ is a well-defined function on $[n]^2$, a set of unordered pairs.\footnote{Even if we considered each pair in $[n]^2$ to be ordered, our arguments would still have shown that $d(\cdot, \cdot)$ is well-defined and symmetric.}

So we have the following lemma.

**Lemma 28.** For all $x, y \in [n]$, $d(x, y) = d(y, x)$.

**Lemma 29.** For all distinct $x, y \in [n]$, $d(x, x) = 0$ and $d(x, y) \geq 1/2$. 
Proof. Recall that \( G = ([n], E, w) \). As \( \text{Im}(w) \subseteq [1/2, \infty) \) by equation (28), we have \( d_G(x, y), d_G(y, x) \geq 1/2 \). So by equations (40)–(41) and \( h \in \mathbb{Z}^+ \setminus \{1\} \), \( d(x, y) \geq 1/2 \). Because we forbid queries for the distance from a point to itself, \( d(x, x) \) is not defined by equations (40). By equation (41), \( d(x, x) = 0 \).

Lemma 30. \(([n], d)\) is a metric space.

Proof. By Lemmas 28–29, we only need to show that
\[
d(x, y) + d(y, z) \geq d(x, z)
\]
for all \( x, y, z \in [n] \). It is well-known that a positively-weighted undirected graph induces a distance function obeying the triangle inequality; hence
\[
d_G(x, y) + d_G(y, z) \geq d_G(x, z).
\]
Because \( G \) is undirected, \( d_G(\cdot, \cdot) \) is symmetric. So by equations (40)–(41),
\[
d(x, y) \in \left\{ \min \left\{ d_G(x, y), h \right\}, \min \left\{ d_G(x, y), h - \frac{1}{2} \right\} \right\}
\]
for all \( x, y \in [n] \). Now verify inequality (42) in the following exhaustive (but not mutually exclusive) cases:

Case 1: \( x = y, y = z \) or \( x = z \). Lemma 29 implies inequality (42).

Case 2: \( d_G(x, y) \geq h - 1/2 \) and \( y \neq z \). By relation (44), \( d(x, y) \geq h - 1/2 \). As \( y \neq z \), \( d(y, z) \geq 1/2 \) by Lemma 29. By relation (44), \( d(x, z) \leq h \). Summarizing the above proves inequality (42).

Case 3: \( d_G(y, z) \geq h - 1/2 \) and \( x \neq y \). Replace “(x, y),” “(y, z)” and “y \neq z” in the analysis of Case 2 by “(y, z),” “(x, y)” and “x \neq y,” respectively.

Case 4: \( d_G(x, y) < h - 1/2 \) and \( d_G(y, z) < h - 1/2 \). By relation (44), \( d(x, y) = d_G(x, y) \) and \( d(y, z) = d_G(y, z) \). So inequalities (42)–(43) share a common left-hand side. To deduce inequality (42) from inequality (43), therefore, it suffices to show that \( d_G(x, z) \geq d(x, z) \), which follows from relation (44).

Lemma 31. For all \( i \in [q(n)] \),
\[
d_H(i) (a_i, b_i) \geq d_G(a_i, b_i).
\]
Proof. Assume the existence of an \(a_i\)-\(b_i\) path in \(H^{(i)}\) for, otherwise, \(d_{H^{(i)}}(a_i, b_i) = \infty\) and there is nothing to prove. Take a shortest \(a_i\)-\(b_i\) path \(P\) in the unweighted graph \(H^{(i)} = ([n], E^{(i)}_H)\). So \(d_{H^{(i)}}(a_i, b_i)\) is the number of \(P\)'s edges. By Lemma 18, \(P\)'s number of edges equals \(w(P)\). By Lemma 4, \(P\)'s edges are in \(E^{(q(n))}_H\). So by Lemma 20, \(P\) is a path in \(G = ([n], E, w)\), implying \(d_G(a_i, b_i) \leq w(P)\). Summarizing the above proves the lemma.

The following lemma says that line 17 of \(\text{Adv}\) answers queries consistently with \(d(\cdot, \cdot)\).

**Lemma 32.** For all \(i \in [q(n)]\),

\[
\min \left\{ d_{H^{(i)}}(a_i, b_i), h - \frac{1}{2} \cdot X \left[ \exists v \in \{a_i, b_i\}, (v \in S) \land (\deg_{Q^{(i)}}(v) \leq \delta n^{1/(h-1)}) \right] \right\} = d(a_i, b_i). 
\]

(45)

**Proof.** Lemma 27 and equation (40) prove the “\(\leq\)" part of equation (45). On the other hand, Lemma 31 and equation (40) imply the “\(\geq\)" part of equation (45).

### 3.4 Putting things together

We now arrive at our main result.

**Theorem 33.** Metric 1-median has no deterministic \(o(n^{1+1/(h-1)})\)-query \((2h - \epsilon)\)-approximation algorithms for any constants \(h \in \mathbb{Z}^+ \setminus \{1\}\) and \(\epsilon > 0\).

**Proof.** By Lemma 32 and line 17 of \(\text{Adv}\), \(\text{Adv}\) answers \(A\)'s queries consistently with \(d(\cdot, \cdot)\). This implies that \(A^{\text{Adv}}\) and \(A^d\) have the same output.\(^6\) That is, \(A^d\) outputs \(z\). By Lemma 30, \(([n], d)\) is a metric space.

By relation (44), \(d(x, y) \leq \min\{d_G(x, y), h\}\) for all \(x, y \in [n]\). Therefore,

\[
\sum_{v \in [n]} d(\hat{x}, v) \leq n \cdot \left( \frac{1}{2} + 2h\delta^{h-1} + \frac{h^2}{\delta} \cdot o(1) + h\delta \right) \tag{46} 
\]

by Lemma 17.

Recall that \(A\) does not repeat queries. So by equation (15) and Lemmas 28–29,

\[
\sum_{v \in [n]} d(z, v) \geq \sum_{i \in I} d(a_i, b_i). 
\]

\(^7\) By Lemmas 10 and 32,

\[
\sum_{i \in I} d(a_i, b_i) \geq n \cdot (h - 2h\delta^{h-1} - o(1) - \delta). \tag{47} 
\]

\(^6\)See, e.g., [2, Lemma 8].

\(^7\)In fact, this is an equality because \(A^{\text{Adv}}\) will have queried for the distances between its output and all other points when halting.
By inequalities (46)–(47),
\[
\frac{\sum_{v \in [n]} d(z, v)}{\sum_{v \in [n]} d(\hat{\alpha}, v)} \geq \frac{h - 2h\delta^{h-1} - o(1) - \delta}{1/2 + 2h\delta^{h-1} + (h^2/\delta) \cdot o(1) + h\delta}.
\] (48)

Note that all the derivations so far have been valid for all constants \(h \in \mathbb{Z}^+ \setminus \{1\}\) and \(\delta \in (0, 1)\). Take \(\delta = \delta(h, \epsilon) > 0\) to be sufficiently small and \(n\) to be sufficiently large so that the right-hand side of inequality (48) is greater than \(2h - \epsilon\).\(^8\) Then inequality (48) forbids \(z\), which is the common output of \(A^{\text{adv}}\) and \(A^d\), from being a \((2h - \epsilon)\)-approximate 1-median of \([n], d\). Note that \(A\) can be any deterministic \(o(n^{1+1/(h-1)})\)-query algorithm from the beginning of this section.

Next, we use Theorem 33 and Fact 1 to determine the minimum value of \(c \geq 1\) such that \textsc{metric} 1-MEDIAN has a deterministic \(O(n^{1+c})\)-query (resp., \(O(n^{1+c})\)-time) \(c\)-approximation algorithm, for each constant \(\epsilon \in (0, 1)\).

**Theorem 34.** For each constant \(\epsilon \in (0, 1)\),
\[
\min \{c \geq 1 \mid \textsc{metric} 1\text{-median} \text{ has a deterministic } O(n^{1+c})\text{-query } c\text{-approx. alg.} \} = \min \{c \geq 1 \mid \textsc{metric} 1\text{-median} \text{ has a deterministic } O(n^{1+c})\text{-time } c\text{-approx. alg.} \} = 2 \left\lceil \frac{1}{\epsilon} \right\rceil.
\]

**Proof.** Take \(h = \lceil 1/\epsilon \rceil\); hence \(h \in \mathbb{Z}^+ \setminus \{1\}\). It is easy to verify that \(n^{1+c} = o(n^{1+1/(h-1)})\). So by Theorem 33, \textsc{metric} 1-MEDIAN does not have a deterministic \(O(n^{1+c})\)-query \((2\lceil 1/\epsilon \rceil - \epsilon')\)-approximation algorithm for any constant \(\epsilon' > 0\). Clearly, \(n^{1+1/h} = O(n^{1+c})\). So by Fact 1, \textsc{metric} 1-MEDIAN has a deterministic \(O(n^{1+c})\)-time \((2\lceil 1/\epsilon \rceil)\)-approximation algorithm.

The above analyses remain valid with “query” and “time” exchanged because every \(O(n^{1+c})\)-time algorithm makes \(O(n^{1+c})\) queries. Consequently, deterministic \(O(n^{1+c})\)-query (resp., \(O(n^{1+c})\)-time) algorithms can be \((2\lceil 1/\epsilon \rceil)\)-approximate but not \((2\lceil 1/\epsilon \rceil - \epsilon')\)-approximate for any constant \(\epsilon' > 0\).\(\square\)

The brute-force exact algorithm for \textsc{metric} 1-MEDIAN is well-known to run in \(O(n^2)\) time. Therefore, there is no need to extend Theorem 34 to the case of \(\epsilon = 1\). On the other hand, the following corollary deals with the case of \(\epsilon = 0\).

**Corollary 35.** \textsc{metric} 1-MEDIAN does not have a deterministic \(O(n^{1+o(1)})\)-query (resp., \(O(n^{1+o(1)})\)-time) \(O(1)\)-approximation algorithm.

**Proof.** Take \(h \to \infty\) in Theorem 33.\(\square\)

\(^8\)Alternatively, we may take
\[
\delta = \delta(n) = \left(\frac{\max\{q(n), n\}}{n^{1+1/(h-1)}}\right)^{1/3}
\]
from the beginning of this section. Then, as \(q(n) = o(n^{1+1/(h-1)})\), the right-hand side of inequality (48) is \(2h - o(1)\), and inequalities (1)–(3) remain true for all sufficiently large \(n\).
Acknowledgments

The author is supported in part by the Ministry of Science and Technology of Taiwan under grant 103-2221-E-155-026-MY2.

A Optimizing the hidden factors in Theorem 33

This appendix discusses how the bound of $o(n^{1+1/(h-1)})$ in Theorem 33 hides factors dependent on $h$. For all $i \in [q(n)]$,

$$B_{i-1} \overset{\text{def}}{=} \{ v \in [n] \mid \deg_{H(i-1)}(v) \geq \delta n^{1/(h-1)} - 2 \}. \quad (49)$$

**Lemma 36.** For all $i \in [q(n)]$ and distinct $u, v \in [n] \setminus (B_{i-1} \cup S)$, we have $(u, v) \in E^{(i-1)}_G$.

**Proof.** As $u, v \in [n] \setminus B_{i-1}$,

$$\deg_{H(0)}(u) < \delta n^{1/(h-1)} - 2,$$
$$\deg_{H(0)}(v) < \delta n^{1/(h-1)} - 2$$

for all $j \in \{0, 1, \ldots, i - 1\}$ by equation (49) and Lemma 4. So by lines 8 and 13 of Adv, $(u, v) \in E^{(j)}_G$ if $(u, v) \in E^{(j-1)}_G$, for all $j \in [i - 1]$. By equation (4), $(u, v) \in E^{(0)}_G$. The proof is complete by mathematical induction. \qed

**Lemma 37.** For each $i \in [q(n)]$ such that the $i$th iteration of the loop of Adv runs lines 5–9, $P_i$ in line 5 does not have two non-consecutive vertices in $[n] \setminus (B_{i-1} \cup S)$.

**Proof.** By line 5 of Adv, two non-consecutive vertices on $P_i$ are not connected by an edge in $E^{(i-1)}_G$. This and Lemma 36 complete the proof. \qed

**Lemma 38.** For all $i \in [q(n)]$ and $v \in B_{i-1}$,

$$N_{G^{(i-1)}}(v) \subseteq N_{H^{(i-1)}}(v).$$

**Proof.** By equation (49),

$$\deg_{H^{(i-1)}}(v) \geq \delta n^{1/(h-1)} - 2.$$  

Clearly,

$$\deg_{H^{(0)}}(v) \overset{(6)}{=} 0 < \delta n^{1/(h-1)} - 2.$$  

So there exists $j \in [i - 1]$ satisfying

$$\deg_{H^{(j-1)}}(v) < \delta n^{1/(h-1)} - 2,$$
$$\deg_{H^{(j)}}(v) \geq \delta n^{1/(h-1)} - 2.$$  

21
Clearly,
\[
N_{G^{(j)}}(v) = \{ u \in [n] \mid (u, v) \in E_G^{(j)} \}.
\] (52)

As \( H^{(j-1)} \neq H^{(j)} \) by inequalities (50)–(51), the \( j \)th iteration of the loop of \text{Adv} runs lines 5–9 but not 11–14. By inequality (51) and line 8 of \text{Adv},
\[
\{ u \in [n] \mid (u, v) \in E_G^{(j)} \} = \{ u \in [n] \mid (u, v) \in E_G^{(j-1)} \setminus \left( E_H^{(j-1)} \setminus E_H^{(j)} \right) \}.
\] (53)

Equations (52)–(53) and Lemma 4 give
\[
N_{G^{(j)}}(v) = N_{H^{(j)}}(v).
\] This and Lemma 4 complete the proof.

**Lemma 39.** For all \( i \in \lfloor q(n) \rfloor \),
\[
|E_H^{(i)}| \leq |E_H^{(i-1)}| + 1.
\]

*Proof.* Clearly, we may assume that the \( i \)th iteration of the loop of \text{Adv} runs lines 5–9 but not 11–14. By line 6, we only need to show that
\[
\left| \left\{ e \mid (e \text{ is an edge on } P_i) \land (e \notin E_H^{(i-1)}) \right\} \right| \leq 1.
\] (54)

By Lemma 37, \( P_i \) in line 5 has at most one edge in \( ([n] \setminus (B_{i-1} \cup S))^2 \). So, to prove inequality (54), it suffices to show that each edge \( (u, v) \) on \( P_i \) with \( (u, v) \notin ([n] \setminus (B_{i-1} \cup S))^2 \) satisfies \( (u, v) \in E_H^{(i-1)} \), as done below:

Case 1: \( \{u, v\} \cap S \neq \emptyset \). By equation (4) and Lemma 4, \( (u, v) \notin E_G^{(i-1)} \). Consequently, \( P_i \) has an edge not in \( E_G^{(i-1)} \), contradicting line 5.

Case 2: \( \{u, v\} \cap B_{i-1} \neq \emptyset \). By symmetry, assume \( v \in B_{i-1} \). So by Lemma 38, \( N_{G^{(i-1)}}(v) \subseteq N_{H^{(i-1)}}(v) \). Because \( P_i \) is a path in \( G^{(i-1)} \) by line 5 and \( (u, v) \) is on \( P_i \), \( u \in N_{G^{(i-1)}}(v) \). In summary, \( u \in N_{H^{(i-1)}}(v) \). I.e., \( (u, v) \in E_H^{(i-1)} \).

The following improvement over Lemma 11 is immediate from equation (6) and Lemma 39.

**Lemma 40.**
\[
|E_H^{(q(n))}| \leq q(n).
\]

Assuming \( 100 \leq h = o(n^{1/(h-1)}) \), the following modifications to this paper show that the bound of \( o(n^{1+1/(h-1)}) \) in Theorem 33 depends on \( h \) as \( o(n^{1+1/(h-1)}/h) \):
(1) Take

\[ q(n) = o\left(\frac{n^{1+1/(h-1)}}{h}\right), \]

\[ \delta = h \cdot \max\{q(n), n\} \cdot \frac{n^{1+1/(h-1)}}{h^{1+1/(h-1)}}, \]

\[ \lambda = \delta^{b/8}, \]

\[ S = \lfloor \lambda n \rfloor. \]

(2) Replace “\( \delta \)” by “\( \sqrt{\delta} \)” in inequality (2).

(3) Replace “\( \delta \)” by \( 1/\delta^{h/4} \) in inequality (3).

(4) Replace the two occurrences of “\( \delta \)” by “\( \sqrt{\delta} \)” in line 8 of Adv.

(5) Replace “\( \delta \)” by “\( 1/\delta^{h/4} \)” in line 17 of Adv.

(6) Replace all occurrences of “\( \delta \)” by “\( \sqrt{\delta} \)” in Lemma 8 and its proof.

(7) Replace all occurrences of “\( \delta \)” by “\( \sqrt{\delta} \)” in Lemma 9 and its proof.

(8) Replace the two occurrences of “\( \delta \)” by “\( \sqrt{\delta} \)” in line 8 of Adv.

(9) Replace all occurrences of “\( \delta \)” by “\( 1/\delta^{h/4} \)” in line 17 of Adv.

(10) That \( \hat{\alpha} \) is well-defined in equation (19) follows from \( |S| \geq 2 \), which holds for all sufficiently large \( n \) by item (1) and \( h \geq 100 \).

(11) Replace all occurrences of “\( \delta \)” by “\( 1/\delta^{h/4} \)” in Lemma 13 and its proof.

(12) Replace all occurrences of “\( \delta \)” by “\( 1/\delta^{h/4} \)” in Lemma 16.

(13) Replace “\( \delta \)” by “\( \sqrt{\delta} \)” in equation (26).

(14) Replace “\( \delta^{h-1} \)” by “\( \delta^{h-1} \)” in the statement of Lemma 15.

(15) Replace all occurrences of “\( \delta \)” by “\( \sqrt{\delta} \)” and “\( 1/\delta^{h/4} \)” respectively, in the first and the second paragraphs of the proof of Lemma 15.

(16) Replace “\( 1-2\delta^{h-1}-(h/\delta)\cdot o(1)-\delta \)” by “\( 1-2\delta^{h-1}-(1/\sqrt{\delta})\cdot O(q(n)/n^{1+1/(h-1)})-\lambda \)” in the statement of Lemma 16.
(17) Replace all occurrences of “$h/\delta \cdot o(n)$,” “$[\delta n]$” and “$\delta^{h-1}$” by “$(1/\sqrt{\delta}) \cdot O(q(n)/n^{1/(h-1)}),$” “[n]” and “$\delta^{h/4-1},$” respectively, in the proof of Lemma 16.

(18) Replace “$\delta^{h-1},$” “$h^2/\delta \cdot o(1)$” and “$h\delta$” by “$\delta^{h/4-1},$” “$(h/\sqrt{\delta}) \cdot O(q(n)/n^{1+1/(h-1)}$)” and “$h\lambda,$” respectively, in the statement of Lemma 17.

(19) Replace “$\delta$” by “$1/\delta^{h/4}$” in the statement of Lemma 23.

(20) Replace all occurrences of “$\delta$” by “$\sqrt{\delta}$” in the proof of Lemma 24.

(21) Replace the two occurrences of “$\delta$” by “$1/\delta^{h/4}$” in the statement of Lemma 27.

(22) Replace “$\delta$” by “$1/\delta^{h/4}$” in equation (40).

(23) Replace “$\delta$” by “$1/\delta^{h/4}$” in the statement of Lemma 32.

(24) Replace “$\delta^{h-1},$” “$h^2/\delta \cdot o(1)$” and “$h\delta$” by “$\delta^{h/4-1},$” “$(h/\sqrt{\delta}) \cdot O(q(n)/n^{1+1/(h-1)}$)” and “$h\lambda,$” respectively, in inequality (46).

(25) Replace “$h-2h\delta^{h-1}-o(1)-\delta$” by “$h-2h\sqrt{\delta^{h-1}}-o(1)-\lambda/2-1/(2\delta^{h/4} n^{1-1/(h-1)})$” in the right-hand side of inequality (47).

(26) Replace the numerator and the denominator on the right-hand side of inequality (48) by “$h-2h\sqrt{\delta^{h-1}}-o(1)-\lambda/2-1/(2\delta^{h/4} n^{1-1/(h-1)})$” and “$1/2+2h\delta^{h/4-1}+(h/\sqrt{\delta}) \cdot O(q(n)/n^{1+1/(h-1)})+h\lambda,$” respectively.

(27) Verify that the right-hand side of inequality (48) is $2h-o(1)$. To see this, use item (1) and $100 \leq h = o(n^{1/(h-1)})$ to verify that $\delta = o(1),$ max$_{x \geq 1} x \cdot \delta^{x/8} = O(\delta) = o(1)$ (which requires elementary calculus and reveals that $h\sqrt{\delta^{h-1}} = o(1), h \delta^{h/4-1} = o(1)$ and $h \lambda = h \delta^{h/8} = o(1)).$ $\lambda = o(1), \delta^{h/4} \geq 1/n^{h/(4(h-1))},$ $\delta^{h/4} n^{1/(h-1)} = n^{\Omega(1)}, \sqrt{\delta} \geq \sqrt{h \cdot q(n)/n^{1+1/(h-1)}}$ and $\sqrt{h \cdot q(n)/n^{1+1/(h-1)}} = o(1).

(28) Replace all occurrences of “$\delta$” by “$\sqrt{\delta}$” in equation (49) as well as in the proofs of Lemmas 36 and 38.

References

[1] C.-L. Chang. A deterministic sublinear-time nonadaptive algorithm for metric 1-median selection. To appear in Theoretical Computer Science.

[2] C.-L. Chang. Some results on approximate 1-median selection in metric spaces. Theoretical Computer Science, 426:1–12, 2012.

[3] C.-L. Chang. Deterministic sublinear-time approximations for metric 1-median selection. Information Processing Letters, 113(8):288–292, 2013.
[4] C.-L. Chang. A lower bound for metric 1-median selection. Technical Report arXiv: 1401.2195, 2014.

[5] S. Guha, A. Meyerson, N. Mishra, R. Motwani, and L. O’Callaghan. Clustering data streams: Theory and practice. IEEE Transactions on Knowledge and Data Engineering, 15(3):515–528, 2003.

[6] P. Indyk. Sublinear time algorithms for metric space problems. In Proceedings of the 31st Annual ACM Symposium on Theory of Computing, pages 428–434, 1999.

[7] P. Indyk. High-Dimensional Computational Geometry. PhD thesis, Stanford University, 2000.

[8] A. Kumar, Y. Sabharwal, and S. Sen. Linear-time approximation schemes for clustering problems in any dimensions. Journal of the ACM, 57(2):5, 2010.

[9] R. R. Mettu and C. G. Plaxton. Optimal time bounds for approximate clustering. Machine Learning, 56(1–3):35–60, 2004.

[10] W. Rudin. Principles of Mathematical Analysis. McGraw-Hill, 3rd edition, 1976.

[11] B.-Y. Wu. On approximating metric 1-median in sublinear time. Information Processing Letters, 114(4):163–166, 2014.