Incremental Medians via Online Bidding*

Marek Chrobak† Claire Kenyon‡ John Noga§ Neal E. Young¶

June 13, 2006

Abstract

In the k-median problem we are given sets of facilities and customers, and distances between them. For a given set F of facilities, the cost of serving a customer u is the minimum distance between u and a facility in F. The goal is to find a set F of k facilities that minimizes the sum, over all customers, of their service costs.

Following the work of Mettu and Plaxton, we study the *incremental* medians problem, where k is not known in advance. An incremental algorithm produces a nested sequence of facility sets $F_1 \subseteq F_2 \subseteq ... \subseteq F_n$, where $|F_k| = k$ for each k. Such an algorithm is called *c-cost-competitive* if the cost of each F_k is at most c times the optimum k-median cost. We give improved incremental algorithms for the metric version of this problem: an 8-cost-competitive deterministic algorithm, a $2e \approx 5.44$-cost-competitive randomized algorithm, $(24+\epsilon)$-cost-competitive, polynomial-time deterministic algorithm, and a $6e + \epsilon \approx 16.31$-cost-competitive, polynomial-time randomized algorithm.

We also consider the competitive ratio with respect to *size*. An algorithm is *s-size-competitive* if the cost of each F_k is at most the minimum cost of any set of k facilities, while the size of F_k is at most sk. We show that the optimal size-competitive ratios for this problem, in the deterministic and randomized cases, are 4 and e. For polynomial-time algorithms, we present the first polynomial-time $O(\log m)$-size-approximation algorithm for the offline problem, as well as a polynomial-time $O(\log m)$-size-competitive algorithm for the incremental problem.

Our upper bound proofs reduce the incremental medians problem to the following *online bidding* problem: faced with some unknown threshold $T \in \mathbb{R}^+$, an algorithm must submit "bids" $b \in \mathbb{R}^+$ until it submits a bid $b \geq T$, paying the sum of all its bids. We present folklore algorithms for online bidding and prove that they are optimally competitive.

We extend some of the above results for incremental medians to approximately metric distance functions and to incremental fractional medians. Finally, we present an optimally cost-competitive deterministic algorithm for a restricted version of the incremental medians problem where k is restricted to one of two given values.

*The conference version of this paper appeared in [8].

†Department of Computer Science, University of California, Riverside, CA 92521. Research supported by NSF Grant CCR-0208856.

‡Computer Science Department, Brown University, Providence, RI 02912.

§Department of Computer Science, California State University, Northridge, CA 91330.

¶Department of Computer Science, University of California, Riverside, CA 92521.

1 Introduction and Summary of Results

The k-median problem. An instance of the k-median problem is specified by a finite set $\mathcal{C}$ of customers, a finite set $\mathcal{F}$ of facilities, and, for each customer u and facility f, a distance $d_{uf} \geq 0$ from u to f representing the cost of serving u from f. The cost of a set of facilities $F \subseteq \mathcal{F}$ is $cost(F) = \sum_{u \in \mathcal{C}} d_{uF}$, where $d_{uF} = \min_{f \in F} d_{uf}$. For a given k, the *offline* k-median problem is to compute a *k-median*, that is, a set $F \subseteq \mathcal{F}$ of cardinality k for which $cost(F)$ is minimum (among all sets of cardinality k). This minimum cost is denoted opt_k. An instance of the k-median problem is called *metric* if the distance function is a metric (the shortest u-to-f path has length d_{uf} for each $u \in \mathcal{C}$ and $f \in \mathcal{F}$).

The k-median problem is a well-known NP-hard facility location problem. Substantial work has been done on efficient approximation algorithms that, given k, find a set F_k of k facilities of approximately minimum cost [2, 1, 6, 5, 14, 13, 25, 21]. In particular, for the metric version, Arya *et al.* [2] show that, for any $\epsilon > 0$, a set F_k of cost at most $(3 + \epsilon)opt_k$ can be found in polynomial time.

Incremental medians. The *incremental* medians problem is a version of the k-median problem where k is not specified in advance [23, 22]. Instead, authorizations for additional facilities arrive over time. Given an instance of the k-median problem, a (possibly randomized) algorithm produces a sequence $\bar{F} = (F_1, F_2, \ldots, F_n)$ of facility sets, where $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n \subseteq \mathcal{F}$, and $|F_k| \leq k$ for all k.

In general, in an incremental solution, the F_k's cannot all simultaneously have minimum cost. An algorithm is said to be *c-cost-competitive*, or to have *cost-competitive ratio* of c, if it produces a (possibly random) sequence $\bar{F}$ of sets which is c-cost-competitive, that is, such that for each k, the set F_k has size at most k and (expected) cost at most $c \cdot opt_k$.

Alternatively, an algorithm is *s-size*-competitive if it produces a (possibly random) sequence $\bar{F}$ of sets which is s-size-competitive, that is, such that each set F_k has cost at most opt_k and (expected) *size* at most sk.

For offline solutions we use the term “approximate” instead of “competitive”.

Online bidding. To analyze incremental medians, we reduce the various incremental medians problems to the following folklore *online bidding* problem. An algorithm repeatedly submits “bids” $b \in \mathbb{R}^+$, until it submits a bid b that is at least as large as some unknown threshold $T \in \mathbb{R}^+$. The algorithm's cost is the sum of the submitted bids. The algorithm is β-competitive if, for any $T \in \mathbb{R}^+$, its cost is at most βT (or, if the algorithm is randomized, its expected cost is at most βT). More generally, the algorithm may be given in advance a closed universe $\mathcal{U} \subseteq \mathbb{R}^+$, with a guarantee that the threshold T is in $\mathcal{U}$ and a requirement that all bids be in $\mathcal{U}$. (To handle the case when $\mathcal{U}$ includes arbitrarily small positive numbers, we allow the bidding sequence to be left-infinite – see Section 2 for a formal definition.)

In Section 2 we completely characterize optimal competitive ratios for deterministic and randomized algorithms for online bidding:

Theorem 1 [folklore] (a) *The online bidding problem has a deterministic 4-competitive algorithm. If $\mathcal{U}$ is finite, this algorithm runs in time polynomial in $|\mathcal{U}|$.* **(b)** *The online bidding problem has a randomized e-competitive algorithm. If $\mathcal{U}$ is finite, this algorithm runs in time polynomial in $|\mathcal{U}|$.*

Throughout, $[n]$ denotes $\{1, 2, \ldots, n\}$.

Theorem 2 **(a)** **[folklore]** *No deterministic algorithm for online bidding is less than* 4*-competitive, even when restricted to instances of the form* $\mathcal{U} = [n]$ *for some integer* n. **(b)** *No randomized algorithm for online bidding is less than* e*-competitive, even when restricted to instances of the form* $\mathcal{U} = [n]$ *for some integer* n.

Portions of Theorems 1 and 2 are folklore. In particular, Theorem 1(a) uses a doubling algorithm that has been used previously in several papers, first in [16, 24] and later in [12, 3, 4, 10]. Some of these papers also have the randomized upper bound. We include proofs of all bounds for completeness. Our main new contribution in this setting is Theorem 2(b), a randomized lower bound that matches the known upper bound.

Size-competitiveness. To our knowledge, size-competitive algorithms for incremental medians have not been studied, although other online problems have been analyzed in an analogous setting of *resource augmentation* (e.g. [15, 7, 18]). For unrestricted (possibly non-polynomial-time) algorithms, we completely characterize the optimal size-competitive ratios:

Theorem 3 **(a)** *The incremental medians problem has a* 4*-size-competitive deterministic algorithm.* **(b)** *No deterministic incremental algorithm has size-competitive ratio smaller than* 4.

Theorem 4 **(a)** *The incremental medians problem has an* e*-size-competitive randomized algorithm.* **(b)** *No randomized incremental algorithm has size-competitive ratio smaller than* e.

We stress that the upper and lower bounds in Theorems 3 and 4 are for unrestricted algorithms and hold for both the metric and non-metric problems.

Regarding polynomial-time algorithms, previously no polynomial-time size-approximation algorithms for k-medians were known, even for the offline problem. The best previous result known for the offline problem is a bicriteria-approximation algorithm, which finds a facility set of size $O(k \log(m + m/\epsilon))$ and cost at most $(1 + \epsilon)opt_k$ [25] (this improved [21]). We first improve this result to obtain an offline size-approximation algorithm:

Theorem 5 *The offline* k*-median problem has a polynomial-time* $O(\log m)$*-size-approximation algorithm, where* $m = |\mathcal{C}|$ *is the number of customers.*

Note that this algorithm finds a true (not bicriteria) approximate solution: a facility set of size $O(k \log m)$ and cost at most opt_k. We use this result and a reduction to give a polynomial-time size-approximation algorithm for the incremental problem:

Theorem 6 *The incremental medians problem has a polynomial-time* $O(\log m)$*-size-competitive algorithm.*

Note that this is optimal in the sense that no polynomial-time algorithm (incremental or offline) is $o(\log m)$-size-competitive unless P=NP, even for the metric case. This follows from known results on the hardness of approximating set cover.

Theorems 3, 4, and 6 (proven in Section 3) imply the size-competitive ratios shown in Figure 1.

Cost-competitive incremental medians. For incremental medians, Mettu and Plaxton [23, 22] give a c-cost-competitive linear-time algorithm with $c \approx 30$. We improve this result. The problem is difficult both because (i) the solution must be incremental, and (ii) even the offline problem is NP-hard. To study separately the effects of the two difficulties, we consider both polynomial-time and unrestricted algorithms, proving the following two theorems (see Section 4 for the proofs):

problem:	**cost-competitive, metric**		**size-competitive**		**bidding**
time:	polynomial	unrestricted	polynomial	unrestricted	polynomial
deterministic	$24+\epsilon$	8	$\boldsymbol{O(\log m)}$	**4**	**4**
randomized	$6e+\epsilon < 16.31$	$2e < 5.44$	$\boldsymbol{O(\log m)}$	$\boldsymbol{e} < 2.72$	$\boldsymbol{e} < 2.72$

Figure 1: Competitive ratios shown for incremental medians and online bidding. Ratios in bold are optimal.

Theorem 7 **(a)** *The metric incremental medians problem has an* 8*-cost-competitive deterministic algorithm.* **(b)** *Suppose that the offline metric k-median problem has a polynomial-time c-cost-approximation algorithm. Then the incremental medians problem has a polynomial-time $8c$-cost-competitive deterministic algorithm.*

Theorem 8 **(a)** *The metric incremental medians problem has a $2e$-cost-competitive randomized algorithm.* **(b)** *Suppose that the offline metric k-median problem has a polynomial-time c-cost-approximation algorithm. Then the incremental medians problem has a polynomial-time $2ec$-cost-competitive randomized algorithm.*

As it is known that there is a polynomial-time $(3+\epsilon)$-cost-approximation algorithm for the metric k-median problem [2], Theorems 7 and 8 imply the cost-competitive ratios shown in Figure 1.

Theorems 7 and 8 were recently and independently discovered by Lin, Nagarajan, Rajaraman and Williamson [19]. For polynomial-time algorithms, they improve the cost-competitive ratios further to 16 and $4e$, in the deterministic and randomized cases, respectively. They also generalize the approach to incremental versions of k-MST, k-vertex cover, k-set cover, facility location, and to hierarchical k-median.

λ-Relaxed metrics. Mettu and Plaxton show that their incremental medians algorithm also works in "λ-approximate" metric spaces, achieving cost-competitive-ratio $O(\lambda^4)$ [23, 22]. We get a similar result. Let $\lambda \geq 1$. We say that the cost function d is a *λ-relaxed metric* if $d_{fy} \leq \lambda(d_{fx} + d_{xg} + d_{gy})$ for any facilities f, g and customers x and y. (This condition is somewhat less restrictive than the one in [23, 22]. A related concept was studied in [11].) In Section 5, we prove that Theorems 7 and 8 generalize as follows:

Theorem 9 **(a)** *The incremental medians problem for λ-relaxed metrics has a deterministic algorithm with cost-competitive ratio $8\lambda^2$.* **(b)** *If the offline k-median problem for λ-relaxed metrics has a polynomial-time c-cost-approximation algorithm, then the incremental version has a deterministic polynomial-time algorithm with cost-competitive ratio $8\lambda^2 c$.*

Theorem 10 **(a)** *The incremental medians problem for λ-relaxed metrics has a randomized algorithm with cost-competitive ratio $2e\lambda^2$.* **(b)** *If the offline k-median problem for λ-relaxed metrics has a polynomial-time c-cost-approximation algorithm, then the incremental version has a randomized polynomial-time algorithm with cost-competitive ratio $2e\lambda^2 c$.*

Incremental fractional medians. A *fractional k-median* is a solution to the standard linear program relaxation for the k-median problem (see Section 6). The natural *incremental* version of this fractional problem asks for a sequence of fractional solutions, one for each k, such that, for each facility f, as k increases, the fractional extent to which f is opened does not decrease. Specifically, an incremental fractional solution specifies, for each $k \in [n]$, a fractional k-median $(x_{uf}^{(k)}), (y_f^{(k)})$ meeting the constraints of the linear program, as well as $y_f^{(k)} \leq y_f^{(k+1)}$ (for all $f \in \mathcal{F}$ and $k < n$).

Such an incremental solution is c-cost-competitive if, for each k, the cost of the fractional solution is at most c times the optimal (non-incremental) fractional solution for that k.

The proof of the theorem below (see Section 6) extends the proof of Theorem 8, along with the observation that the randomized algorithm for the fractional problem can be derandomized without increasing the competitive ratio.

Theorem 11 *The incremental fractional metric medians problem has a deterministic polynomial-time algorithm that is* $2e$*-cost-competitive.*

The kl-medians problem. A natural question to ask is whether better competitive ratios are possible if the number of facilities can take only some limited number of values. As shown in [23, 22], no algorithm can be better than 2-competitive even when there are only two possible numbers of facilities, either 1 or k, for some large k.

For any $1 \leq k < l \leq n$, we define the *kl-medians* problem as the restricted version of the incremental medians problem where the number of facilities is either k or l. In Section 7, we determine the exact competitive ratios of the kl-medians problem in the deterministic case:

Theorem 12 *Let* $1 \leq k < l \leq n$. (a) *If the distance function is metric, then there is a deterministic kl-median algorithm with cost-competitive ratio* $2 - 1/l$, *and* (b) *no better ratio is possible.*

Bicriteria approximations. We say that an algorithm for the k-median problem is a *bicriteria (c, s)-approximation* algorithm if it approximates the cost within the ratio of c and the size within the ratio of s. Such bicriteria approximation algorithms for k-medians have been studied by many authors, [20, 21, 17, 25]. We remark without proof that, analogously to Theorem 7 and Theorem 3, one can transform any offline polynomial-time bicriteria (c, s)-approximation algorithm into a polynomial-time *incremental* algorithm whose bicriteria performance guarantee is either $(c, 4s)$ or $(8c, s)$. For example, for metric spaces, using the approximation results from [2, 20, 17] one can obtain incremental polynomial-time algorithms with the following respective bicriteria ratios: $(3 + \epsilon, 4)$, $(2 + \epsilon, 4(1 + 2\epsilon^{-1}))$, and $(1 + \epsilon, 4(3 + 5\epsilon^{-1}))$, where $\epsilon > 0$.

Weighted medians. We remark without proof that many of the results in this paper extend to a version of k-medians where facilities and customers are given non-negative weights. Denote by w_u the weight of $u \in \mathcal{F} \cup \mathcal{C}$. In this generalization, for a facility set F, one constrains the total facility weight $\sum_{f \in F} w_f$ to be at most k, and one takes $cost(F) = \sum_{u \in \mathcal{C}} w_u d_{uF}$.

2 Online Bidding

In this section we provide a complete analysis of online bidding by proving Theorems 1 and 2.

Throughout the paper, $\mathbb{R}^+$ denotes the set of non-negative reals, $\mathbb{Z}$ the set of integers, and $\mathbb{N}^+$ the set of positive integers. For $n \in \mathbb{N}^+$, let $[n] = \{1, 2, \ldots, n\}$.

Definition 1 *Given a universe $\mathcal{U}$ which is a closed subset of $\mathbb{R}^+$, an* online bidding *algorithm outputs a* bid set $\mathcal{B} \subseteq \mathcal{U}$. *Against a particular threshold $T \in \mathcal{U}$, the algorithm's cost is*

$$\sum \{b \in \mathcal{B} : b < T\} + \min\{b \in \mathcal{B} : b \geq T\}.$$

*The bid set $\mathcal{B}$ is β-*competitive *if, for any $T \in \mathcal{U}$, this cost is at most βT.*

Proof of Theorem 1(a): We give a deterministic 4-competitive algorithm.

First, consider the case $\mathcal{U} = \mathbb{R}^+$. Define the algorithm to produce the set of bids $\mathcal{B} = \{0\} \cup \{2^i : i \in \mathbb{Z}\}$. If the threshold T is zero, the algorithm pays zero. For any other threshold $T > 0$, let p be such that $T \in (2^{p-1}, 2^p]$. The algorithm pays $\sum_{i \le p} 2^i = 2^{p+1} \le 4T$, and thus its competitive ratio is at most 4.

Next, we reduce the general case to the case $\mathcal{U} = \mathbb{R}^+$. Let $\mathcal{B}$ be a 4-competitive bid set for $\mathcal{U} = \mathbb{R}^+$. For an arbitrary closed universe $\mathcal{U}' \subseteq \mathbb{R}^+$, the algorithm produces the bid set

$$\mathcal{B}' \;=\; \big\{\max(\mathcal{U}' \cap [0, b]) \;:\; b \in \mathcal{U}\big\}.$$

In other words, we by replace each $b \in \mathcal{B}$ with the maximum element in $\mathcal{U}' \cap [0, b]$ (if any). The cost incurred against any threshold $T \in \mathcal{U}'$ is at most the cost incurred when using the bid set $\mathcal{B}$ against the same threshold T. Thus, the bid set $\mathcal{B}'$ is also 4-competitive. Note that for the bid set $\mathcal{B}$ described in the previous paragraph, the bid set $\mathcal{B}'$ can be computed in time polynomial in $|\mathcal{U}|$ if $\mathcal{U}$ is finite. □

Proof of Theorem 1(b): We give a randomized e-competitive algorithm.

First, we consider the case $\mathcal{U} = \mathbb{R}^+$. The algorithm picks a real number $\xi \in (0, 1]$ uniformly at random, then produces the set of bids $\mathcal{B} = \{0\} \cup \{e^\xi e^i : i \in \mathbb{Z}\}$.

For the analysis, let $T > 0$ be the threshold. Observe that we can equivalently describe the algorithm as producing the set of bids $\mathcal{B} = \{0\} \cup \{e^{\xi + \ln(T) + i} : i \in \mathbb{Z}\}$ where ξ is distributed uniformly in $[0, 1)$. Let b be the random variable equal to the largest bid paid by the algorithm, defined by $e^{\xi + \ln(T) + i - 1} < T \le e^{\xi + \ln(T) + i} = b$, or in other words, $e^{\xi + i - 1} < e^0 \le e^{\xi + i} = b/T$. Thus $i = 0$ and b/T is distributed like e^ξ with ξ uniform in $[0, 1)$. It follows that the expectation of b is $T \int_0^1 e^z \, dz = T(e - 1)$, the expected total payment incurred by the algorithm is $E[\sum_{i=0}^\infty be^{-i}] = E[b]e/(e - 1) = eT$, and the algorithm is e-competitive.

The general case, for an arbitrary closed universe, reduces to the case $\mathcal{U} = \mathbb{R}^+$ just as in the proof of Theorem 1(a) above. □

Proof of Theorem 2(a): We now show a deterministic lower bound matching the upper bound.

In the simple case where $\mathcal{U} = \mathbb{N}^+$, assume, for a contradiction, that there exists an online algorithm with competitive ratio $a < 4$. Let $\mathcal{B} = \{b_i\}_{i \ge 1}$ be the bid set produced by the algorithm, $s_j = \sum_{i=1}^j b_i$ and $y_j = s_{j+1}/s_j$, for $j \ge 1$. Against the threshold $T = b_j + 1$, the algorithm pays s_{j+1}. By our assumption, $s_{j+1} \le a(b_j + 1)$ for all j. Substituting $b_j = s_j - s_{j-1}$, rearranging, and using $s_j \ge j$, we get:

$$(y_{j+1} - y_j)y_j \;\le\; -y_j^2 + ay_j - a(1 - 1/s_j) \;\le\; -y_j^2 + ay_j - a(1 - 1/j).$$

For $a < 4$ and j large enough, the discriminant $a^2 - 4a(1 - 1/j)$ is negative, thus the rightmost expression is negative, and so the sequence (y_j) is ultimately decreasing. As it is bounded below by 1, it converges to some limit y which, by continuity, must satisfy $0 = (y - y)y \le -y^2 + ay - a$ – a contradiction, since $-y^2 + ay - a < 0$ for $0 < a < 4$ and any y.

In the case when $\mathcal{U} = [n]$, for arbitrarily large n, the proof is similar but a bit more explicit. Let $\mathcal{B} = \{b_i\}_{i=1}^m$. We start out the same. Note that $s_{j+1} \le a(b_j + 1)$ implies $s_{j+1} < 8s_j$, and so $y_j < 8$ for all j. Let $j_0 = j_0(a)$ be the smallest positive integer such that $a^2 - 4a(1 - 1/j) = -\delta < 0$. For $j \ge j_0$, we have:

$$(y_{j+1} - y_j)y_j \;\le\; \max_y(-y^2 + ay - a(1 - 1/j)) \;=\; -\delta/4.$$

Since $y_j \leq 8$, the above inequality implies that $y_{j+1} - y_j \leq -\delta/32$. Therefore, for $j = j_0, ..., m$, we have $1 \leq y_j \leq y_{j_0} - (j - j_0)\delta/32 \leq 8 - (j - j_0)\delta/32$, so, in particular, $m < j_0 + 224/\delta$. Since $b_m = n$ (otherwise, $\mathcal{B}$ would not be competitive at all) and $b_{j+1} \leq 8b_j$ for all $j = 1, ..., m-1$, we get $n < 8^{j_0 + 224/\delta}$, contradicting our assumption that n can be arbitrarily large. □

Proof of of Theorem 2(b): We now show a randomized lower bound matching the upper bound. The proof consists of the two lemmas below.

Lemma 1 *Let $n \in \mathbb{N}^+$. Given $\mu : [n] \to \mathbb{R}^+$ and $\pi : [n] \to \mathbb{R}^+$ which satisfy*

$$\sum_{T=t}^{n} \frac{1}{T}\pi(T) \;\geq\; \frac{1}{b}\sum_{T=t}^{b} \mu(T) \quad (\forall b, t : 1 \leq t \leq b \leq n), \tag{1}$$

any randomized β-competitive online bidding algorithm for $\mathcal{U} = [n]$ must have $\beta \geq \dfrac{\sum_{T=1}^{n} \mu(T)}{\sum_{T=1}^{n} \pi(T)}$.

Proof: Consider a β-competitive randomized algorithm for $\mathcal{U}$, and let $\mathcal{B} = \{b_1, b_2, \ldots, b_m\}$ be the ordered (random) sequence of bids produced by the algorithm. Without loss of generality, $b_m = n$.

For $t \leq b$, let $X(t, b)$ be the characteristic function of the event that $t - 1$ and b are consecutive elements of $\{0\} \cup \mathcal{B}$. The algorithm pays $\sum_{t \leq T} \sum_b bX(t, b)$ against threshold T. Since it is β-competitive, this is at most βT in expectation. Since the algorithm only stops when reaching a bid greater than or equal to T, we must always have $\sum_{t,b: t \leq T \leq b} X(t, b) \geq 1$. Hence, together with β, the expectation $x(t, b)$ of $X(t, b)$ must form a feasible solution to the following linear program (LP):

$$\text{minimize}_{\beta,x}\ \beta \ \text{ subject to } \begin{cases} \beta - \displaystyle\sum_{b=1}^{n} \frac{b}{T} \sum_{t=1}^{T} x(t, b) \;\geq\; 0 & (\forall T \in [n]) \\ \displaystyle\sum_{b=T}^{n} \sum_{t=1}^{T} x(t, b) \;\geq\; 1 & (\forall T \in [n]) \\ x(t, b) \;\geq\; 0 & (\forall t, b \in [n]). \end{cases}$$

Thus, the value of (LP) is a lower bound on the optimal competitive ratio of the randomized algorithm. To get a lower bound on the value of (LP), we use the dual (DLP) (where the dual variables $\mu(T)$ correspond to the first set of constraints and $\pi(T)$ to the second set of constraints):

$$\text{maximize}_{\mu,\pi} \sum_{T=1}^{n} \mu(T) \quad \text{subject to } \begin{cases} \displaystyle\sum_{T=1}^{n} \pi(T) \;\leq\; 1 & \\ \displaystyle\sum_{T=t}^{b} \mu(T) - \sum_{T=t}^{n} \frac{b}{T}\pi(T) \;\leq\; 0 & (\forall t, b \in [n]) \\ \mu(T), \pi(T) \;\geq\; 0 & (\forall T \in [n]). \end{cases}$$

Now, given any μ and π meeting the condition of the lemma, if we scale μ and π by dividing by $\sum_T \pi(T)$, we get a feasible dual solution whose value is $\sum_T \mu(T) \,/\, \sum_T \pi(T)$, hence the lemma. □

Lemma 2 *There exist $\mu : [n] \to \mathbb{R}^+$ and $\pi : [n] \to \mathbb{R}^+$ satisfying Condition (1) of Lemma 1 and such that $\sum_T \mu(T)/\sum_T \pi(T) \geq (1 - o(1))e$ as n goes to infinity.*

Proof: Fix U arbitrarily large and let $n = \lceil U^2 \log U \rceil$. Let $\alpha > 0$ be a constant to be determined later. We will choose α so that Condition (1) holds, and then show that the corresponding lower bound is $e(1 - o(1))$ as $U \to \infty$. Define

$$\mu(T) = \begin{cases} \alpha/T & \text{if } U \le T \le U^2 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \pi(T) = \begin{cases} 1/T & \text{if } U \le T \le U^2 \log U \\ 0 & \text{otherwise.} \end{cases}.$$

If $t \ge U^2$, then the right-hand side of Condition (1) has value 0, so the condition holds trivially. On the other hand, since $\pi(T)$ and $\mu(T)$ are zero for $T < U$, if the condition holds for $t = U$, then it also holds for $t < U$. So, we need only verify the condition for t in the range $U \le t \le U^2$. The expression on the left-hand side of (1) then has value

$$\sum_{T=t}^{U^2 \log U} \frac{1}{T^2} \;\ge\; \int_t^{1+U^2\log U} \frac{1}{T^2}\, dT \;=\; \frac{1}{t} - \frac{1}{1+U^2\log U} \;\ge\; \frac{1}{t}(1-o(1)).$$

In comparison, the expression on the right-hand side has value at most

$$\max_{b \ge t} \frac{1}{b} \sum_{T=t}^{b} \frac{\alpha}{T} \;\le\; \alpha \max_{b\ge t} \frac{1}{b} \int_{t-1}^{b} \frac{1}{T}\, dT \;=\; \alpha \max_{b \ge t} \frac{1}{b} \ln \frac{b}{t-1} \;=\; \frac{\alpha}{e\, t(1-o(1))}.$$

(The second equation follows by calculus, for the maximum occurs when $b = e(t-1)$.) Thus, Condition (1) is met for $\alpha = (1 - o(1))e$. Then, Lemma 1 gives a lower bound on the competitive ratio of

$$\frac{\sum_T \mu(T)}{\sum_T \pi(T)} \;=\; \frac{\sum_{T=U}^{U^2} \alpha/T}{\sum_{T=U}^{U^2 \log U} 1/T} \;=\; (1-o(1))\alpha \frac{\ln(U^2/U)}{\ln((U^2\log U)/U)} \;=\; (1-o(1))e.$$

□

Theorem 2 follows directly from Lemmas 1, and 2. □

3 Incremental Size-Competitive Medians

In this section we prove Theorems 3, 4, 5, and 6. The proofs are based on the reduction shown in the next lemma. We show that from a β-competitive algorithm for online bidding, and a c-size-approximation algorithm for the offline k-median problem, we can construct a $c\beta$-size-competitive algorithm for the incremental medians problem. This reduction works even for the non-metric case. The reduction takes polynomial time, so, if the offline size-approximation algorithm runs in polynomial time, the reduction yields a polynomial-time size-competitive algorithm. Likewise, if the online bidding algorithm is randomized then the size-competitive algorithm will be randomized.

Lemma 3 (a) *Let $\beta \ge 1$ and assume that there exists a (possibly randomized) polynomial-time β-competitive algorithm for online bidding. Suppose also that for each k we have a set of facilities F_k^* of size at most sk and cost at most opt_k. Then, in polynomial time, we can compute a (randomized)* incremental *solution $\bar{F} = (F_1, ..., F_n)$ where each F_k has size at most βsk and cost at most opt_k.*

Proof: We give the proof in the deterministic setting. (The proof in the randomized setting is an easy extension, and we omit it.) Let $\mathcal{U} = [n]$ and take $\mathcal{B}$ to be the set of bids used by the β-competitive online bidding algorithm for universe $\mathcal{U}$. Let $\mathcal{B}_k$ be the set of bids issued against threshold $T = k$. We define $F_k = \bigcup_{b \in \mathcal{B}_k} F_b^*$. Note that $\mathcal{B}_j \subseteq \mathcal{B}_{j+1}$ for all $j < n$, and so $\bar{F}$ is indeed an incremental solution. Further, F_k contains F_b^* for some $b \geq k$, so $cost(F_k) \leq cost(F_b^*) \leq opt_b \leq opt_k$. Finally, $|F_k| \leq \sum_{b \in \mathcal{B}_k} |F_b^*| \leq \sum_{b \in \mathcal{B}_k} sb \leq s\beta k$ since $\mathcal{B}$ is β-competitive. □

Proof of Theorems 3(a) and 4(a): By Theorems 1(a), and 2(a), there are deterministic and randomized algorithms for online bidding with competitive ratios of 4 and e, respectively. Using these online bidding algorithms, and taking each F_k^* to be the optimal k-median, the reduction in Lemma 3 gives a 4-size-competitive deterministic algorithm and an e-size-competitive randomized algorithm for incremental medians. □

Next, we turn our attention to the proof of Theorem 3(b), our lower bound on the size-competitiveness of unrestricted algorithms for incremental medians. We do this by showing the converse of the reduction in Lemma 3. That is, we show that competitive online bidding reduces to size-competitive incremental medians:

Lemma 4 *Let $s \geq 1$ and assume that, for incremental medians (metric or not), there is a (possibly randomized) s-size-competitive algorithm. Then, for any integer n, there is a (randomized) s-competitive algorithm for online bidding with $\mathcal{U} = [n]$.*

Proof: We give the proof in the deterministic setting. (The proof in the randomized setting is an easy extension, and we omit it.) For any arbitrarily large m, we construct a set $\mathcal{C}$ of customers, a set $\mathcal{F}$ of facilities, and a metric distance function d_{uf}, for $u \in \mathcal{C}$ and $f \in \mathcal{F}$. The facility set $\mathcal{F}$ will be partitioned into sets $M_1, M_2, \ldots, M_m$, where $|M_k| = k$ for each k, with the following properties:

(i) For all k, $cost(M_k) > cost(M_{k+1})$, and

(ii) For all k, and for every set F of facilities, if $cost(F) \leq cost(M_k)$ then there exists $\ell \geq k$ such that M_ℓ is contained in F.

Notice that the two conditions above imply that each M_k is the unique optimum k-median.

Assume for the moment that there exists such a metric space, and consider an s-size-competitive incremental median $\bar{F} = (F_1, ..., F_m)$ for it. Let $\mathcal{B} = \{k : M_k \subseteq F_k\}$. For $k = m$, (ii) implies that $M_m \subseteq F_m$, and thus $\mathcal{B} \neq \emptyset$.

We show that $\mathcal{B}$ is an s-competitive bid set for universe $\mathcal{U} = [m]$. Against any threshold $T \in [m]$, the total of the bids paid will be

$$C \;=\; \sum\{k : k < T, M_k \subseteq F_k\} + \min\{\ell : \ell \geq T, M_\ell \subseteq F_\ell\}.$$

Now, $\sum\{k : k < T, M_k \subseteq F_k\} \leq \sum\{k : k < T, M_k \subseteq F_T\}$ since $\bar{F}$ is a nested sequence. Similarly, we have

$$\min\{\ell : \ell \geq T, M_\ell \subseteq F_\ell\} \;\leq\; \min\{\ell : \ell \geq T, M_\ell \subseteq F_T\}.$$

(By (ii), $M_\ell \subseteq F_T$ for some $\ell \geq T$, so the minimum on the right is well-defined for $T \in [m]$.) Thus we can bound the cost of $\mathcal{B}$ as follows:

$$
\begin{aligned}
C &\leq \sum\{k : k < T, M_k \subseteq F_T\} + \min\{\ell : \ell \geq T, M_\ell \subseteq F_T\} \\
&= \sum\{|M_k| : k < T, M_k \subseteq F_T\} + \min\{|M_\ell| : \ell \geq T, M_\ell \subseteq F_T\} && \text{because } |M_k| = k \\
&\leq \sum\{|M_k| : M_k \subseteq F_T\} \\
&\leq |F_T| && \text{since the } M_k\text{'s are disjoint} \\
&\leq sT. && \text{since } \bar{F} \text{ is } s\text{-size-competitive.}
\end{aligned}
$$

Thus, the bid set $\mathcal{B}$ is s-competitive for universe $\mathcal{U} = [m]$.

We now present the construction of the metric space satisfying conditions (i) and (ii). Let $\mathcal{C} = [1] \times [2] \times ... \times [m]$, that is, $\mathcal{C}$ is the set of integer vectors $\bar{u} = (u_1, u_2, \ldots, u_m)$ where $u_\ell \in [\ell]$ for all $\ell = 1, 2, \ldots, m$. For each $\ell \in [m]$, introduce a set $M_\ell = \{\mu_{\ell,1}, \mu_{\ell,2}, \ldots, \mu_{\ell,\ell}\}$, and for each node $\bar{u}$ in $\mathcal{C}$, connect $\bar{u}$ to μ_{ℓ,u_ℓ} with an edge of length $\delta_\ell = 1 + (m!)^{-\ell}$. The set of facilities is $\mathcal{F} = \bigcup_\ell M_\ell$. All distances between points in $\mathcal{C} \cup \mathcal{F}$ other than those specified above are determined by shortest-path distances. Since $1 < \delta_\ell \leq 2$ for all ℓ, the resulting distance function satisfies the triangle inequality.

We have $cost(M_j) = m!\delta_j$ for each $j \in [m]$, so (i) holds. We prove (ii) by contradiction. Fix some index j and consider a set $F \subseteq \mathcal{F}$ that does not contain M_ℓ for any $\ell \geq j$: for each $\ell \geq j$ there is $i_\ell \leq \ell$ such that $\mu_{\ell,i_\ell} \notin F$. Define a customer $\bar{u}$ as follows: $u_i = 1$ for $\ell = 1, \ldots, j-1$ and $u_i = i_\ell$ for $\ell = j, \ldots, m$. Then the facility $\mu_{\ell,i} \in F$ serving this $\bar{u}$ must have $\ell < j$ or $i \neq i_\ell$. Either way, it is at distance at least δ_{j-1} from $\bar{u}$. Since each other customers pays strictly more than 1, we get $cost(F) > m! - 1 + \delta_{j-1} = m!\delta_j = cost(M_j)$ – a contradiction. □

Proof of Theorems 3(b) and Theorem 4(b): Lemma 4 and the lower bounds for online bidding in Theorem 2 immediately imply Theorems 3(b) and 4(b). □

Next we turn attention to polynomial-time algorithms. We start with the offline problem. We prove Theorem 5 — that there exists a polynomial-time $O(\log m)$-size-approximation algorithm for the offline problem. (Given k and a problem instance, in polynomial time, our algorithm finds a facility set of size $O(k \log m)$ and cost at most opt_k.) Here $m = |\mathcal{C}|$ is the number of customers.

Proof of Theorem 5: Without loss of generality, we assume that each customer has distance 0 to its closest facility. (If not, for each customer u we subtract $d_{u\mathcal{F}}$ from the distance d_{uf} to each facility. This decreases the cost of each facility set by a uniform amount, so any solution having optimal cost in the modified instance will also have optimal cost in the original instance.)

The algorithm first runs the bicriteria approximation algorithm from [25], which finds a facility set of size $O(k \log(m + m/\epsilon))$ and cost at most $(1+\epsilon)\, opt_k$. The algorithm takes $\epsilon = 1/m$. This gives a facility set F of size $O(k \log m)$ and cost at most $(1 + 1/m)\, opt_k$. The algorithm then greedily adds to F a single facility f that minimizes $cost(F \cup \{f\})$. The algorithm returns $F \cup \{f\}$.

To finish, we show that the facility set $F \cup \{f\}$ has size $O(k \log m)$ and cost at most opt_k. The size bound is immediate from the size bound on F. To show the cost bound, note that if we were to add to F, for each of the m customers, the closest facility to that customer, the cost would decrease to 0 (recall our assumption that $d_{u\mathcal{F}} = 0$ for each customer u). Thus, by the submodularity of the cost function, there is some facility $f \notin F$ whose addition to F decreases its cost by at least $1/m$ times the current cost, and we get $cost(F \cup \{f\}) \leq (1 - 1/m)\, cost(F) \leq (1 - 1/m)(1 + 1/m)\, opt_k \leq opt_k$.
□

By the reduction in Lemma 3, this gives a polynomial-time size-approximation algorithm for the incremental problem:

Proof of Theorem 6: By Theorem 1(a), there is a deterministic polynomial-time algorithm for online bidding with competitive ratio $O(1)$. Using this online bidding algorithm, and using Theorem 5 to compute an $O(\log m)$-size-approximate offline k-median F_k^* for each k, the reduction in Lemma 3 gives an $O(\log m)$-size-competitive deterministic polynomial-time algorithm for incremental medians. $\square$

4 Incremental Cost-Competitive Medians

In this section we prove Theorems 7 and 8. The argument is based on another reduction from online bidding to incremental medians. We show, in the lemma below, that from a β-competitive (possibly randomized) algorithm for online bidding, and a c-cost-approximation algorithm for the offline k-median problem, we can construct (in polynomial time) a $2\beta c$-cost competitive (possibly randomized) algorithm for the incremental medians problem. This implies the theorems.

Lemma 5 *Let $\beta \geq 1$ and assume that there exists a polynomial-time β-competitive algorithm for online bidding. Consider an instance of the metric medians problem. Suppose that for each k we have a set of facilities F_k^* of size at most k and cost at most $c \cdot opt_k$. Then in polynomial time we can compute an* incremental *solution $\bar{F} = (F_1, ..., F_n)$, where each F_k has size at most k and cost at most $2\beta c \cdot opt_k$.*

Proof: Assume, without loss of generality, that $cost(F_k^*) \geq cost(F_{k+1}^*)$ for all k. The algorithm constructs the incremental solution $\bar{F}$ from $F_1^*, ..., F_n^*$ in several steps. First, fix some index set $\mathcal{K} \subseteq [n]$, with $1 \in \mathcal{K}$, by a method to be described later, and let $\kappa(1) < \kappa(2) < \ldots < \kappa(m)$ denote the indices in $\mathcal{K}$ in increasing order.

Compute $F_{\kappa(i)}$ as follows. Start by defining $F_{\kappa(m)} = F_{\kappa(m)}^*$. Then, working backwards, for $i = m-1, m-2, ..., 1$, inductively define $F_{\kappa(i)}$ to contain the facilities within $F_{\kappa(i+1)}$ that are "closest" to $F_{\kappa(i)}^*$ in sense defined below.

Definition 2 *Introduce distances between facilities: given two $f, g \in \mathcal{F}$, let $d_{fg} = \min_{x \in \mathcal{C}}(d_{fx} + d_{xg})$. Given two sets of facilities A and B, let $\Gamma(A, B)$ be a set $C \subseteq B$ of cardinality at most $|A|$ such that $d_{aB} = d_{aC}$ for all $a \in A$. (Ties broken arbitrarily.) In other words, for each $a \in A$ we add to C an element of B closest to a.*

With this notation, we define $F_{\kappa(i)} = \Gamma(F_{\kappa(i)}^*, F_{\kappa(i+1)})$. Directly from the definition, $|F_{\kappa(i)}| \leq |F_{\kappa(i)}^*| \leq \kappa(i)$.

Finally, for indices $k \notin \mathcal{K}$, define $F_k = F_{\kappa(i)}$, where i is maximum such that $\kappa(i) \leq k$ (this i is well defined, since $1 \in \mathcal{K}$.) Obviously, $|F_k| \leq \kappa(i) \leq k$. To complete the construction, it remains to describe how to compute $\mathcal{K}$, which we momentarily defer.

We will use the following fact, whose proof can be found in [9] and is also implicit in [14]. (See also Sections 5 and 6 for generalizations with proofs.)

Fact 1 *Assume that the distance function is metric. Consider two sets of facilities A and B, and let $C = \Gamma(A, B)$. Then for every customer u we have $d_{uC} \leq 2d_{uA} + d_{uB}$.*

To analyze the cost, let i be maximum such that $\kappa(i) \leq k$. From the definition of $F_{\kappa(i)}$ and Fact 1 summed over all $u \in \mathcal{C}$, we have: $cost(F_{\kappa(i)}) \leq 2cost(F^*_{\kappa(i)}) + cost(F_{\kappa(i+1)})$. Applying this inequality repeatedly gives:

$$cost(F_k) \;=\; cost(F_{\kappa(i)}) \;\leq\; 2\sum_{j\geq i} cost(F^*_{\kappa(j)}).$$

By assumption $cost(F^*_k) \leq c \cdot opt_k$. To complete the proof of the lemma, it suffices to define $\mathcal{K}$ so that

$$\sum_{j\geq i} cost(F^*_{\kappa(j)}) \;\leq\; \beta\, cost(F^*_k). \tag{2}$$

Let $\mathcal{U} = \{cost(F^*_k) : 1 \leq k \leq n\}$ and take $\mathcal{B}$ to be the set of bids used by the β-competitive online bidding algorithm for universe $\mathcal{U}$. We define $\mathcal{K} = \{\kappa : cost(F^*_\kappa) \in \mathcal{B}\}$, with ties broken in favor of smaller indices (ties may happen if several sets of facilities have the same cost). Note that with this tie-breaking rule, we always have $1 \in \mathcal{K}$. (If the online bidding algorithm is randomized, then $\mathcal{B}$ and therefore $\mathcal{K}$ are random.) Then the left-hand side of (2) is exactly the sum of the bids paid by the online bidding algorithm for threshold $T = cost(F^*_k)$. Since $\mathcal{B}$ is β-competitive, this cost is at most $\beta\, cost(F^*_k)$, so (2) holds. This completes the proof. □

The algorithm of Theorem 7(a) is obtained by combining the reduction used in the proof of Lemma 5 with the deterministic 4-competitive online bidding algorithm given in the proof of Theorem 1(a). The resulting algorithm is detailed in the displayed Algorithm 1.

Algorithm 1 The 8-cost-competitive deterministic oblivious median algorithm.

```
for all k = 1, ..., n, do
    compute a k-median solution F*_k
end for
K ← {1}
for all k = 2, ..., n, do
    if cost(F*_k) = 0 < cost(F*_{k-1}) then add k to K
    if cost(F*_k) > 0 then if ⌈log cost(F*_k)⌉ < ⌈log cost(F*_{k-1})⌉ then add k to K
end for
ℓ ← max(K); F_ℓ ← F*_ℓ
for all k = ℓ + 1, ..., n do F_k ← F_ℓ
for each k ∈ K − {max(K)}, in decreasing order do
    F_k ← Γ(F*_k, F_ℓ)
    for all k' = k + 1, ..., ℓ − 1 do F_{k'} ← F_k
    ℓ ← k
end for
```

5 λ-Relaxed Metrics

In this section we prove Theorem 9, namely the upper bounds for cost-competitive ratios when the distance function is a λ-relaxed metric. By a λ-relaxed metric, we mean $d_{yf} \leq \lambda(d_{xf} + d_{xg} + d_{yg})$ for all $f, g \in \mathcal{F}$ and $x, y \in \mathcal{C}$.

We start with a generalization of Fact 1 from the previous section. Recall that $\Gamma(A, B)$ is the set of up to $|A|$ elements of B that are closest to the elements of A. (See Definition 2.)

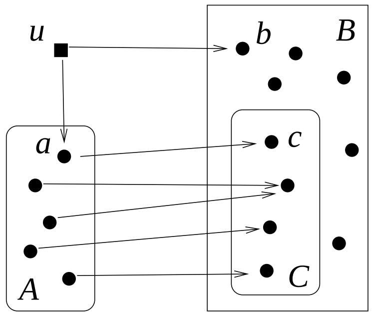

Figure 2: Illustration of Lemma 6

.

Lemma 6 *Assume that the distance function is a λ-relaxed metric. Given two sets of facilities A and B, let $C = \Gamma(A, B)$. Then for every customer $u \in \mathcal{C}$ we have $d_{uC} \le 2\lambda d_{uA} + \lambda d_{uB}$.*

Proof: For a given $u \in \mathcal{C}$, choose $a \in A$ such that $d_{uA} = d_{ua}$, $b \in B$ such that $d_{uB} = d_{ub}$, and $c \in C$ such that $d_{aB} = d_{ac}$, (See Figure 2.) By the definition of the λ-relaxed metric, for every customer $x \in \mathcal{C}$ we have $d_{uc} \le \lambda(d_{ua} + d_{xa} + d_{xc})$. There is $x \in \mathcal{C}$ for which $d_{ac} = d_{xa} + d_{xc}$, and choosing this x we get $d_{uc} \le \lambda(d_{ua} + d_{ac})$. Thus:

$$d_{uC} \le d_{uc} \le \lambda(d_{ua} + d_{ac}) \le \lambda(d_{ua} + d_{ab}) \le \lambda(d_{ua} + (d_{ua} + d_{ub})) = 2\lambda d_{uA} + \lambda d_{uB},$$

completing the proof. $\square$

Proof of Theorem 9: The algorithm used to prove Theorem 9 is the same as in the metric case (see the proof of Lemma 5), except for the definition of $\mathcal{B}$. Let $\mathcal{U} = \{cost(F_k^*) : 1 \le k \le n\}$.

$\mathcal{B}$ contains 0 iff $\mathcal{U}$ does, plus the following elements: in the deterministic case, for every $i \in \mathbb{Z}$, $\mathcal{B}$ contains the maximum element in the set $\mathcal{U} \cap [0, (2\lambda)^i]$ (if it is non-empty); in the randomized case, pick a real number $\xi \in (0, 1]$ uniformly at random, and for every $i \in \mathbb{Z}$, $\mathcal{B}$ contains the maximum element in the set $\mathcal{U} \cap [0, e^{\xi}(e\lambda)^i]$ (if it is non-empty).

Let $\mathcal{K} = \{\kappa(i) : 1 \le i \le |\mathcal{B}|\}$ be defined from $\mathcal{U}$ and $\mathcal{B}$ as in the proof of Theorem 7 (at the end of Section 4).

The analysis is similar to the proof of Theorem 1. Let i be maximum such that $\kappa(i) \le k$. From Lemma 6 summed over $u \in \mathcal{C}$, we have: $cost(F_{\kappa(i)}) \le 2\lambda cost(F^*_{\kappa(i)}) + \lambda cost(F_{\kappa(i+1)})$. Applying this inequality repeatedly gives

$$cost(F_k) = cost(F_{\kappa(i)}) \le 2\lambda \sum_{j \ge i} \lambda^{(j-i)} cost(F^*_{\kappa(j)}).$$

To continue the analysis in the deterministic case, let p be such that $(2\lambda)^{p-1} < cost(F^*_{\kappa(i)}) \le (2\lambda)^p$. By the definition of $\kappa(i)$, we have:

$$\begin{aligned} cost(F_k) &\le 2\lambda(2\lambda)^p \sum_{j \ge i} \lambda^{(j-i)} (2\lambda)^{-(j-i)} \\ &= 2(2\lambda)(2\lambda)^p \\ &\le 4\lambda(2\lambda) cost(F_k^*), \end{aligned}$$

since $k < \kappa(i+1)$ and so $cost(F_k^*) \geq (2\lambda)^{p-1}$. Hence our algorithm is $8\lambda^2$-cost-competitive when F_k^* is the optimal k-median solution and is $8\lambda^2 c$-cost-competitive when F_k^* is a c-approximation.

In the randomized case, let p such that $e^\xi(e\lambda)^{p-1} < cost(F^*_{\kappa(i)}) \leq e^\xi(e\lambda)^p$. By the definition of $\kappa(i)$, we have:

$$cost(F_k) \;\leq\; 2\lambda e^\xi (e\lambda)^p \sum_{j\geq i} \lambda^{(j-i)}(e\lambda)^{-(j-i)} \;=\; \frac{1}{1-1/e} 2\lambda e^\xi (e\lambda)^p.$$

As in the proof of Theorem 1(b), the expected value of $e^\xi(e\lambda)^p/cost(F_k^*)$ is distributed like $e^\xi\lambda$ where ξ is uniform in $(0,1]$, so

$$E[cost(F_k)] \;\leq\; \frac{1}{1-1/e}(2\lambda)\, cost(F_k^*)\lambda(e-1) \;=\; 2e\lambda^2\, cost(F_k^*),$$

and the theorem follows. □

6 Incremental Fractional Medians

In this section we prove Theorem 11.

The following linear program is the relaxation of the standard integer program for the k-median problem. Given a solution (x, y), we refer to x as the *assignment* and y as the *fractional k-median.*

$$\text{minimize}_{x,y} \sum_{uf} d_{uf}x_{uf} \quad \text{subject to} \begin{cases} \sum_f x_{uf} & \geq & 1 & (\forall u \text{ customer}) \\ x_{uf} & \leq & y_f & (\forall u \text{ customer}, \forall f \text{ facility}) \\ \sum_f y_f & \leq & k. \end{cases}$$

We let $\overline{opt}_k$ denote the value of an optimal solution to this linear program.

A c-cost-competitive fractional *incremental* solution gives, for every integer $k \in [n]$, a fractional k-median of cost at most $c \cdot \overline{opt}_k$ and satisfying $y_f^{(k)} \leq y_f^{(k+1)}$ for all $f \in \mathcal{F}$ and $k < n$. In this section we show that a $2e$-cost-competitive fractional oblivious solution can be computed in polynomial time.

As in the proof of Theorem 7, for convenience we define distances between facilities, $d_{fg} = \min_{u\in\mathcal{C}}(d_{uf} + d_{ug})$. Given a fractional h-median y and a fractional k-median z, define d_{yz} to be the minimum cost of "serving" y from z. That is, for each facility f, we order the facilities $g(1), g(2), \ldots$ in order of increasing distance from f, then iteratively define $x_{f,g(i)} = \min\{z_{g(i)}, y_f - \sum_{j<i} x_{f,g(j)}\}$. Then $d_{yz} = \sum_{fg} d_{fg}x_{fg}$. We say that x *realizes* d_{yz} and will use the notation $\mathsf{greedy}_{fg}(y, z)$ to denote x_{fg} — how much (of the z_g available) of facility g is used to serve the y_f fraction of facility f that needs serving. Now, define $w = \Gamma(y, z)$ to be the fractional median where $w_g = \max_f x_{fg}$.

Note that $w = \Gamma(y, z)$ has the following properties: $w \leq z$ (analogous to $\Gamma(A, B) \subseteq B$ in the integral setting); $d_{yw} = d_{yz}$ (analogous to $d_{a\Gamma(A,B)} = d_{aB}$ for every $a \in A$ in the integral setting); and Γ can be computed in polynomial time. Moreover, $\sum_g w_g \leq \sum_{fg} x_{fg} \leq \sum_f y_f \leq h$ and so w is an h-median (analogous to $|\Gamma(A, B)| \leq |A|$ in the integral setting).

Lemma 7 *Assume that the distance function is metric. Given two fractional medians y and z, let $w = \Gamma(y, z)$. Then for every customer u we have $d_{uw} \leq 2d_{uy} + d_{uz}$.*

Proof: Let d_{uy} be the minimum cost of serving u from y. That is, we order the facilities $f(1), f(2), \ldots$ in order of increasing distance from u, then iteratively define $x_{uf(i)} = \min\{y_{f(i)}, 1 -$

$\sum_{j<i} y_{f(j)}\}$, and finally set $d_{uy} = \sum_f d_{uf} x_{uf}$. Let $v \le y$ be the fractional 1-median defined by $v_f = x_{uf}$ for all f. By the choice of v, $d_{uv} = d_{uy}$.

We will prove the following sequence of inequalities:

$$\begin{aligned} d_{uw} &\le d_{uv} + d_{vw} && (3)\\ &\le d_{uv} + d_{vz} && (4)\\ &\le d_{uv} + d_{uv} + d_{uz} && (5)\\ &= 2d_{uy} + d_{uz}. && (6) \end{aligned}$$

Inequality (3): Let x' realize d_{vw}. Note, by definition of v, that $x_{uf} = v_f = \sum_g x'_{fg}$ for each f. Consider serving u by serving $\sum_f x'_{fg}$ units from each facility g. A total of 1 unit is served, and (by the choice of x'), at most w_g comes from each g. Since d_{uw} is the minimum cost to serve u while serving at most w_g from each facility g, we have

$$d_{uw} \le \sum_{fg} x'_{fg} d_{ug} \le \sum_{fg} x'_{fg}(d_{uf} + d_{fg}) = \sum_f x_{uf} d_{uf} + \sum_{fg} x'_{fg} d_{fg} = d_{uv} + d_{vw}.$$

Inequality (4): We will show that $d_{vw} \le d_{vz}$.

Let $w' = \Gamma(v,z)$. By definition of Γ, we have $d_{vw'} = d_{vz}$. So our goal reduces to showing $d_{vw} \le d_{vw'}$. To show this, by the definition of d, it is enough to show $w' \le w$.

Recall that $w = \Gamma(y,z)$ and $v \le y$. To see that this implies $w' \le w$, we need only consider that Γ is determined by the greedy assignment.

$$w'_g = \Gamma_g(v,z) = \max_f \mathsf{greedy}_{fg}(v,z) \le \max_f \mathsf{greedy}_{fg}(y,z) = \Gamma_g(y,z) = w_g.$$

Inequality (5): We use the triangle inequality to show $d_{vz} \le d_{uv} + d_{uz}$, similarly to the argument for Inequality (3). Namely, consider serving v from z via u. Let α be a flow that sends $\alpha(f,u) = v_f$ units from f to u for each $f \in \mathcal{F}$, and $\alpha(u,g) = z_g$ units from u to g for each $g \in \mathcal{F}$. Note the total flow into u is $\sum_f v_f = 1$, while the total flow out of u is also $\sum_g z_g = 1$. The cost of this flow is $\sum_f d_{uf}\alpha(f,u) + \sum_g d_{ug}\alpha(u,g) = d_{uv} + d_{uz}$. By a standard "short-cutting" of this flow α, we get a flow α' that that serves v from z and has cost at most the cost of α. (To short-cut α, repeatedly choose any f and g such that $\alpha(f,u), \alpha(u,g) > 0$, then reduce $\alpha(f,u)$ and $\alpha(u,g)$ by some ϵ, while increasing $\alpha(f,g)$ by ϵ. By the triangle inequality, $d_{fg} \le d_{uf} + d_{ug}$, so this does not increase the cost of the flow.)

Inequality (6): This holds because $d_{uv} = d_{uy}$. □

With the lemma in place, the proof of the theorem is essentially the same as the proof of Theorem 7, along with a minor observation about fractional strategies being closed under randomization.

Proof of of Theorem 11: We first show how to compute an 8-competitive fractional solution, using the deterministic 4-competitive online bidding algorithm. We then describe how to improve this to a $2e$-competitive fractional solution, using the randomized $2e$-competitive online bidding algorithm.

For each $k \in [n]$, compute the optimal fractional k-median; denote it $z^{(k)}$. (Note $\overline{opt}_k = cost(z^{(k)})$.) Let $\mathcal{U} = \{cost(z^{(k)}) : 1 \le k \le n\}$ and take $\mathcal{B}$ to be the set of bids used by the 4-competitive online bidding algorithm for universe $\mathcal{U}$. Let $\mathcal{K} = \{\kappa(i) : 1 \le i \le m\}$ be a minimal set (containing 1) such that $\mathcal{B} = \{cost(z^{(\kappa(i))}) : 1 \le i \le m\}$.

Let $\kappa(1) < \kappa(2) < \ldots < \kappa(m)$ denote the indices in $\mathcal{K}$ in increasing order. Compute $y^{(\kappa(i))}$ as follows. Start by defining $y^{(\kappa(m))} = z^{(\kappa(m))}$. Then, working backwards, inductively for $i = m-1, m-2, ..., 1$, define $y^{(\kappa(i))} = \Gamma(z^{(\kappa(i))}, y^{(\kappa(i+1))})$.

Finally, define $y^{(k)}$ for $k \notin \mathcal{K}$: let $k^- = \max\{i \in \mathcal{K} : i \leq k\}$ (note $1 \in \mathcal{K}$); then $y^{(k)} = y^{(k^-)}$.

To analyze the size, note that $\sum_f y_f^{(k)} \leq \sum_f y_f^{(k^-)} \leq \sum_f z_f^{(k^-)} \leq k^- \leq k$. To analyze the cost, we use Lemma 7 to show that

$$cost(y^{(k)}) \;\leq\; 2 \sum_{\ell \geq k^-, \ell \in \mathcal{K}} cost(z^{(\ell)}). \tag{7}$$

Indeed, for indices $k \in \mathcal{K}$, we have $k = k^-$, and (7) follows from Lemma 7 summed over all $u \in \mathcal{C}$ and from the construction of $y^{(k)}$ (for $k = \kappa(m), \ldots, \kappa(1)$). For $k \notin \mathcal{K}$, inequality (7) holds as well, simply because $y^{(k)} = y^{(k^-)}$, the bound holds for $k = k^-$, and $(k^-)^- = k^-$.

Since the bidding set $\mathcal{B}$ is 4-competitive,

$$\sum_{\ell \geq k^-, \ell \in \mathcal{K}} cost(z^{(\ell)}) \;\leq\; 4\, cost(z^{(k)}). \tag{8}$$

(To see why, the left-hand side of (8) is exactly the sum of the bids paid from the bid set for threshold $T = cost(z^{(k)})$. Since the bid set is 4-competitive, this is at most $4T$.)

This shows that we can compute an 8-cost-competitive fractional solution in polynomial time. Next we improve this ratio to $2e$ using the e-competitive randomized bidding algorithm instead of the 4-competitive deterministic one.

Consider carrying out the above construction using the e-competitive randomized online bidding algorithm. That algorithm generates a random bidding set $\mathcal{B}$ for our universe $\mathcal{U}$. Applying the construction above gives us a random incremental fractional solution $(y^{(k)})_k$.

For each k and f, take $\bar{y}_f^{(k)}$ to be the expectation of $y_f^{(k)}$ for this random solution. By a standard convexity argument, this gives an incremental fractional solution whose cost, for each k, is at most the expected cost of the random solution. Thus, $(\bar{y}^{(k)})_k$ is $2e$-competitive.

To finish, we need to show that $(\bar{y}^{(k)})_k$ can be computed in polynomial time. Consider our randomized bidding algorithm. We claim that, given a universe of size n, the algorithm produces one of at most n possible bidding sets $\mathcal{B}$. The algorithm works by choosing $\tau \in [0,1]$ uniformly at random, then taking the bid set $\mathcal{B}$ to contain each bid $b \in \mathcal{U}$ that is the largest in some set $\mathcal{U} \cap [0, e^{\tau+i}]$ for some integer i. Considering all values of τ as τ increases from 0 to 1, $\mathcal{B}$ (considered as a function of τ) will only change for τ such that $e^{\tau+i} = b$ for some integer i and $b \in \mathcal{U}$. Consideration shows that, for any b, this equation holds for only one value of τ in the range $(0,1]$. Thus, the bid set $\mathcal{B}$ changes at most $|\mathcal{U}| = n$ times. Furthermore, for any particular $\mathcal{B}$, the probability that $\mathcal{B}$ is produced can be computed in polynomial time. Thus, the incremental fractional solution $\bar{y}$ described above can be computed in polynomial time. □

7 Incremental Algorithms for kl-Medians

In this section we prove Theorem 12. Recall that in the kl-median problem, for given $1 \leq k < l \leq n$, we need to compute two facility sets $F_k \subseteq F_l$ with $|F_k| = k$ and $|F_l| = l$, minimizing the cost-competitive ratio

$$c \;=\; \max\left\{ \frac{cost(F_k)}{opt_k}, \frac{cost(F_l)}{opt_l} \right\}.$$

We now prove that the optimal ratio c for this problem is $2 - 1/l$.

Proof of Theorem 12(a): We start with the upper bound proof. Our method here is very different from the previous bounds in this paper and it does not rely on online bidding. Let F and G denote, respectively, the optimum k-median and the optimum l-median. Without loss of generality, we can assume that $F \cap G = \emptyset$, for otherwise we can duplicate the facilities in $F \cap G$. Our algorithm chooses the better of two options below (the one with the better ratio):

(i) $F_k = F$, and F_l is a set with minimum $cost(F_l)$, such that $|F_l| = l$ and $F \subseteq F_l \subseteq \mathcal{F}$, or

(ii) F_k is a set of with minimum $cost(F_k)$ such that $|F_k| = k$ and $F_k \subseteq G$, and $F_l = G$.

We now show that this algorithm's competitive ratio is at most $2 - 1/l$. It is sufficient to show that there exists a k-element set $X \subseteq G$ such that

$$cost(X) + cost(F \cup G - X) \quad \leq \quad (2 - 1/)[cost(F) + cost(G)]. \tag{9}$$

Indeed, Inequality (9) implies that at least one of the following two options must hold: either $cost(X) \leq (2 - 1/l)cost(F)$, or $cost(F \cup G - X) \leq (2 - 1/l)cost(G)$. In the first case, our algorithm can choose option (ii) with $F_k = X$, in the second case it can choose option (i) with $F_l = F \cup G - X$. In both cases, the algorithm's ratio is at most $2 - 1/l$.

It remains to show that there is a k-element set $X \subseteq G$ that satisfies (9). Our proof is based on a probabilistic argument. We start with some notation. For each $f \in F$ and $g \in G$, denote by C_{fg} the set of customers served by f at step k and by g at step l. Let w_{fg} be the cardinality of C_{fg}. By a_{fg} (resp. b_{fg}) we denote the average distance between a customer $x \in C_{fg}$ and f (resp. g.) Formally, $a_{fg} = \sum_{u \in C_{fg}} d_{uf}/|C_{fg}|$ and $b_{fg} = \sum_{u \in C_{fg}} d_{ug}/|C_{fg}|$. It is convenient to think of C_{fg} as a single point with weight w_{fg} whose distances to f and g are a_{fg} and b_{fg}, respectively. The costs of F and G can then be written as $cost(F) = \sum_{f \in F, g \in G} w_{fg} a_{fg}$, and $cost(G) = \sum_{f \in F, g \in G} w_{fg} b_{fg}$.

We now define a probability distribution on k-element subsets of G. Let $w_{fG} = \sum_{g \in G} w_{fg}$. For each $f \in F$ and $g \in G$, define $\bar{w}_{fg} = w_{fg}(w_{fG} - w_{fg})^{-1}$. (Assume for now that $w_{fg} > 0$ for all f, g. We will explain later how to extend the argument to the general case.) Choose a random mapping $\pi : F \to G$ as follows: for any $f \in F$ set $\pi(f) = g$ with probability $\bar{w}_{fg} / \sum_{h \in G} \bar{w}_{fh}$. Fix some ordering of G, and for any such mapping π let X_π be the k-element subset of G that consists of $\pi(F)$ and the first $k - |\pi(F)|$ elements of $G - \pi(F)$, if any. (Intuitively, we would like to take our random k-set $X \subseteq G$ to be $\pi(F)$, but then $F \cup G - \pi(F)$ may have cardinality larger than l and will not be a valid l-median. This is why we add these additional elements to X_π.)

If $h \in G - X_\pi$ and $g \in X_\pi$, then, using the triangle inequality, the cost of serving C_{fg} from g is at most $w_{fh} a_{fh} + w_{fh} \min_{u \in C_{fg}} (d_{uf} + d_{ug}) \leq w_{fh}(a_{fh} + a_{fg} + b_{fg})$. Therefore

$$\begin{aligned}
cost(X_\pi) + cost(F \cup G - X_\pi) \quad &\leq \quad \sum_{f \in F} \Big[\sum_{h \in G - X_\pi} w_{fh}(a_{fh} + a_{f\pi(f)} + b_{f\pi(f)} + b_{fh}) \\
&\qquad\qquad + \sum_{h \in X_\pi} w_{fh}(a_{fh} + b_{fh}) \Big] \\
&= \quad cost(F) + cost(G) + \sum_{f \in F} \sum_{h \in G - X_\pi} w_{fh}(a_{f\pi(f)} + b_{f\pi(f)}) \\
&\leq \quad cost(F) + cost(G) + \sum_{f \in F} (w_{fG} - w_{f\pi(f)})(a_{f\pi(f)} + b_{f\pi(f)}).
\end{aligned}$$

Now, by the linearity of expectation, we have

$$
\begin{aligned}
E[\,cost(X_\pi) + cost(F \cup G - X_\pi)\,] &\le cost(F) + cost(G) \\
&\quad + \sum_{f\in F}\sum_{g\in G} \frac{\bar{w}_{fg}}{\sum_{h\in G}\bar{w}_{fh}}(w_{fG} - w_{fg})(a_{fg} + b_{fg}) \\
&= cost(F) + cost(G) + \sum_{f\in F} \frac{1}{\sum_{h\in G}\bar{w}_{fh}} \sum_{g\in G} w_{fg}(a_{fg} + b_{fg}) \\
&\le (2 - 1/l)[cost(F) + cost(G)].
\end{aligned}
$$

The last inequality holds because, for each $f \in F$, $\sum_{h\in G}\bar{w}_{fh}$ is minimized when $w_{fh} = w_{fG}/l$ for all $h \in G$, and thus $(\sum_{h\in G}\bar{w}_{fh})^{-1} \le 1 - 1/l$. We conclude that there exists a set X_π that satisfies $(*)$, as claimed.

To complete the proof, we still need to extend the argument to the general case, when some weights w_{fg} are zero. Suppose first that we allow arbitrary positive weights. Choose an arbitrarily small $\epsilon > 0$ and set all zero weights w_{fg} to ϵ instead. The earlier argument implies that for each ϵ there is X that satisfies $(*)$. Since there are finitely many choices for X, there is X that satisfies $(*)$ for infinitely many values of ϵ. The continuity of $cost(X) + cost(F \cup G - X)$ with respect to ϵ implies that $(*)$ holds for $\epsilon = 0$ as well.

In our case the weights are integer, so we cannot use arbitrarily small weights. Instead, we create a large number of copies of each customer (same for each), and then add one customer to each empty set C_{fg}. Then the same asymptotic argument as above applies. □

Proof of Theorem 12(b): Our lower bound a slight refinement of the one in [23, 22]. Let $k = 1$ and $1 < l \le n$. Consider the metric space M with customers $v_1, ..., v_l$ and facilities $f_1, g_1, g_2, ..., g_l$. Each customer v_j is connected to facility g_j by an edge of length $\delta = 1/l$. All customers are also connected to facility f_1 with edges of length 1. (See Figure 3.) All other distances are measured along the above defined edges.

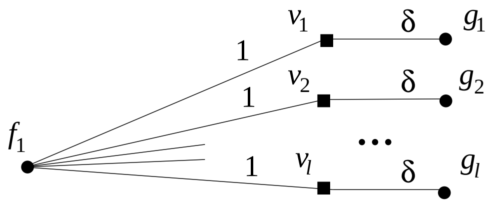

Figure 3: The metric space in the lower bound. Facilities are represented by squares and customers by circles.

Let $G = \{g_1, \ldots, g_l\}$. We have $cost(f_1) = l$ and $cost(G) = l\delta$. Further, for each i, we have $cost(g_i) = \delta + (l-1)(2+\delta)$ and $cost(G - \{g_i\} \cup \{f_1\}) = (l-1)\delta + 1$. Thus, substituting $\delta = 1/l$, we get

$$
\begin{aligned}
\frac{cost(g_i)}{cost(f_1)} &= 2 - 1/l, \\
\frac{cost(G - \{g_i\} \cup \{f_1\})}{cost(G)} &= 2 - 1/l,
\end{aligned}
$$

for all i. So, if an incremental algorithm chooses $F_1 = \{g_i\}$, for some i, the ratio is at least $2 - 1/l$. On the other hand, if it chooses $F_1 = \{f_1\}$, then $F_l = G - \{g_i\} \cup \{f_1\}$, for some i, and the ratio is again at least $2 - 1/l$.

This completes the lower bound proof for $k = 1$. To extend the argument to any $k < l$, add additional $k - 1$ facilities $f_2, ..., f_k$ at a very large distance from the above space and each other, and $k - 1$ customers $v_{l+1} = f_2, ..., v_{l+k-1} = f_k$. Then any k-median must include $f_2, ..., f_k$, and the argument above applies to this new space. $\square$

Acknowledgments

We are grateful to anonymous referees for suggestions to improve the presentation. We also wish to thank Yossi Azar for pointing out references to previous work on online bidding and simplifying the proof of Theorem 2(a).

References

[1] A. Archer, R. Rajagopalan, and D.B. Shmoys. Lagrangian relaxation for the k-median problem: new insights and continuity properties. In *Proc. 11th European Symp. on Algorithms (ESA)*, pages 31–42, 2003.

[2] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristic for k-median and facility location problems. In *Proc. 33rd Symp. Theory of Computing (STOC)*, pages 21–29. ACM, 2001.

[3] S. Chakrabarti, C.A. Phillips, A.S. Schulz, D.B. Shmoys, C. Stein, and J. Wein. Improved scheduling algorithms for minsum criteria. In *Automata, Languages and Programming*, pages 646–657, 1996.

[4] M. Charikar, C. Chekuri, T. Feder, and R. Motwani. Incremental clustering and dynamic information retrieval. In *Proc. 29th Symp. Theory of Computing (STOC)*, pages 626–635. ACM, 1997.

[5] M. Charikar and S. Guha. Improved combinatorial algorithms for the facility location and k-median problems. In *Proc. 40th Symp. Foundations of Computer Science (FOCS)*, pages 378–388. IEEE, 1999.

[6] M. Charikar, S. Guha, E. Tardos, and D.B. Shmoys. A constant-factor approximation algorithm for the k-median problem. In *Proc. 31st Symp. Theory of Computing (STOC)*, pages 1–10. ACM, 1999.

[7] C. Chekuri, A. Goel, S. Khanna, and A. Kumar. Multi-processor scheduling to minimize flow time with ϵ-resource augmentation. In *Proc. 36th Symp. Theory of Computing (STOC)*, pages 363–372. ACM, 2004.

[8] M. Chrobak, C. Kenyon, J. Noga, and N.E. Young. Oblivious medians via online bidding. In José R. Correa, Alejandro Hevia, and Marcos A. Kiwi, editors, *LATIN*, volume 3887 of *Lecture Notes in Computer Science*, pages 311–322. Springer, 2006.

[9] M. Chrobak, C. Kenyon, and N.E. Young. The reverse greedy algorithm for the k-median problem. *Information Processing Letters*, 97:68–72, 2006.

[10] S. Dasgupta and P.M. Long. Performance guarantees for hierarchical clustering. *Journal of Computer and System Sciences*, 70(4):555–569, 2005.

[11] R. Fagin and L. Stockmeyer. Relaxing the triangle inequality in pattern matching. *IJCV: International Journal of Computer Vision*, 30:219–231, 1998.

[12] M. Goemans and J. Kleinberg. An improved approximation ratio for the minimum latency problem. In *Proc. 7th Symp. on Discrete Algorithms (SODA)*, pages 152 – 158. ACM/SIAM, 1996.

[13] K. Jain, M. Mahdian, and A. Saberi. A new greedy approach for facility location problems. In *Proc. 34th Symp. Theory of Computing (STOC)*, pages 731–740. ACM, 2002.

[14] K. Jain and V.V. Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation. *Journal of ACM*, 48:274–296, 2001.

[15] B. Kalyanasundaram and K. Pruhs. Speed is as powerful as clairvoyance. *J. ACM*, 47:214–221, 2000.

[16] M. Kao, J.H. Reif, and S. Tate. Searching in an unknown environment: An optimal randomized algorithm for the cow-path problem. *Information and Computation*, 131(1):63–80, 1996. Preliminary version appeared in the Proceedings of the Symp. on Discrete Algorithms, Austin, TX, Jan 1993.

[17] M.R. Korupolu, C.G. Plaxton, and R. Rajaraman. Analysis of a local search heuristic for facility location problems. *Journal of Algorithms*, 37:146–188, 2000.

[18] E. Koutsoupias. Weak adversaries for the k-server problem. In *Proc. 40th Symp. Foundations of Computer Science (FOCS)*, pages 444–449. IEEE, 1999.

[19] G. Lin, C. Nagarajan, R. Rajamaran, and D.P. Williamson. A general approach for incremental approximation and hierarchical clustering. In *Proc. 17th Symp. on Discrete Algorithms (SODA)*. ACM/SIAM, 2006.

[20] J-H. Lin and J.S. Vitter. Approximation algorithms for geometric median problems. *Information Processing Letters*, 44:245–249, 1992.

[21] J-H. Lin and J.S. Vitter. ϵ-approximations with minimum packing constraint violation (extended abstract). In *Proc. 24th Symp. Theory of Computing (STOC)*, pages 771–782. ACM, 1992.

[22] R. Mettu and C.G. Plaxton. The online median problem. *SIAM Journal on Computing*, 32:816–832, 2003.

[23] Ramgopal R Mettu and C. Greg Plaxton. The online median problem. In *Proc. 41st Symp. Foundations of Computer Science (FOCS)*, pages 339–348. IEEE, 2000.

[24] R. Motwani, S. Phillips, and E. Torng. Nonclairvoyant scheduling. *Theoretical Computer Science*, 130(1):17–47, 1994.

[25] N.E. Young. K-medians, facility location, and the Chernoff-Wald bound. In *Proc. 11th Symp. on Discrete Algorithms (SODA)*, pages 86–95. ACM/SIAM, January 2000.