Online Medians via Online Bribery
(Extended Abstract)

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

Abstract

Following Mettu and Plaxton [19, 20], we study online algorithms for the k-medians problem. Such an algorithm produces an incremental sequence of facility sets. Mettu and Plaxton show that online metric medians has a (roughly) 30-competitive deterministic polynomial-time algorithm. We give improved algorithms, including a $(24+\epsilon)$-competitive deterministic polynomial-time algorithm and a $2e \approx 5.44$-competitive, randomized, non-polynomial-time algorithm. Our approach is similar to that of [16], which was derived independently.

We then consider the competitive ratio with respect to *size*. An algorithm is *s-size-competitive* if, for each k, the cost of 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 present optimally competitive algorithms for this problem.

Our proofs reduce online medians to the following *online bribery* problem: faced with some unknown threshold $T \in \mathbb{R}^+$, an algorithm must submit "bids" $b \in \mathbb{R}^+$ until it submits a bid as large as T. The algorithm pays the sum of its bids. We describe optimally competitive algorithms for online bribery.

Our results on cost-competitive online medians extend to approximately metric distance functions, online fractional medians, and online bicriteria approximation.

We also consider the case when the number of medians can be either k or l, for two given numbers $k < l$. For this case, using a probabilistic argument, we show that the tight cost-competitive ratio is $2 - 1/l$.

1 Introduction and summary of results

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 $X \subseteq \mathcal{F}$ is $cost(X) = \sum_{u \in \mathcal{C}} d_{uX}$, where $d_{uX} = \min_{f \in X} d_{uf}$. For a given k, the *k-median problem* is to compute a *k-median*, that is, a set $X \subseteq \mathcal{F}$ of cardinality k and for which $cost(X) = opt_k$ is minimum (among all sets of cardinality k). *Metric* k-median refers to the case where the distance function is metric (the shortest u-to-f path has length d_{uf} for each u and f).

The k-median problem is a well-known NP-hard facility location problem, and, for the metric variant, substantial work has been done on efficient approximation algorithms that, given k, find a

*Department of Computer Science, University of California, Riverside, CA 92521. Email: marek@cs.ucr.edu.
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. Email: jnoga@csun.edu.
§Department of Computer Science, University of California, Riverside, CA 92521. Email: neal@cs.ucr.edu.

set F_k of k medians of approximately minimum cost [2, 1, 6, 5, 12, 11]. In particular, Arya et al. show that for any $\epsilon > 0$, a set F_k of cost at most $(3+\epsilon)\,opt_k$ can be found in polynomial time [2].

Online medians is the online version of the k-median problem where k is not specified in advance [19, 20]. Instead, authorizations for additional facilities arrive over time. A (possibly randomized) *online algorithm* produces a sequence $\bar{F} = (F_1, F_2, \ldots, F_n)$ of facility sets which must satisfy the online constraint $F_1 \subseteq F_2 \subseteq \cdots \subseteq F_n \subseteq \mathcal{F}$. In general, in an online solution, the F_k's cannot all simultaneously have minimum cost. The 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 $c \cdot opt_k$. For offline solutions we use the term "approximate" instead of "competitive".

Mettu and Plaxton [19, 20] give a c-cost-competitive linear time online algorithm with $c \approx 30$. One can use Theorem 1 to improve this ratio. Note that the online medians problem is difficult for two reasons: (1) the algorithm must be online, and (2) even the offline problem is NP-hard. By considering both polynomial-time and non-polynomial-time algorithms we separate the two effects.

Theorem 1 **(a)** *Online metric medians has non-polynomial-time deterministic and randomized algorithms that are* 8*-cost-competitive and* $2e$*-cost-competitive, respectively.* **(b)** *If (offline) metric k-median has a polynomial-time c-cost-approximation algorithm, then the online problem has polynomial-time deterministic and randomized algorithms that are* $8c$*-cost-competitive and* $2ec$*-cost-competitive, respectively.*

As it is known that there is a polynomial-time $(3+\epsilon)$-cost-approximation algorithm for the offline metric medians [2], Theorem 1 implies the cost-competitive ratios shown in Fig. 1. Theorem 1 was just recently independently discovered by Lin, Nagarajan, Rajaraman and Williamson [16], who then used it together with a Lagrangian multiplier preserving approximation algorithm for facility lication so as to derive a 16-cost-competitive and a randomized $4e$-competitive polynomial-time algorithm for metric medians.

We also consider here online algorithms that are *s-size-competitive*: they are allowed to use extra medians, but must achieve the optimal cost for each k. A (possibly randomized) algorithm is *s-size-competitive* if it produces a (possibly random) sequence $\bar{F}$ 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.

To our knowledge, size-competitive algorithms for online medians have not been studied yet, although many other online problems have been analyzed in an analogous setting called *resource augmentation* (e.g. [13, 7, 15]). Our next contribution is to completely characterize the optimal size-competitive ratios for online medians:

Theorem 2 **(a)** *Online medians has non-polynomial-time deterministic and randomized online algorithms that are* 4*-size-competitive and* e*-size-competitive, respectively.* **(b)** *No deterministic or randomized online algorithm is less than* 4*-size-competitive or* e*-size-competitive, respectively.*

online problem:	**cost-competitive metric medians**		**size-competitive medians**		**bribery**
time:	polynomial	non-polynomial	polynomial	non-polynomial	polynomial
deterministic	$24+\epsilon$	8	$\mathbf{O(\log n)}$	**4**	**4**
randomized	$6e+\epsilon < 16.31$	$2e < 5.44$	$\mathbf{O(\log n)}$	$\mathbf{e} < 2.72$	$\mathbf{e} < 2.72$

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

(c) *If offline k-median has a polynomial-time c-size-competitive algorithm, then the online problem has polynomial-time deterministic and randomized algorithms that are* $4c$*-size-competitive and* ec*-size-competitive, respectively.*

The upper and lower bounds in Theorem 2 hold for both the metric and non-metric problems (with respect to size-competitiveness, the metric and non-metric problems are equivalent). Part (c) on polynomial-time algorithms is included for completeness, as is the following result for offline k-medians:

Theorem 3 *Offline k-medians has a polynomial-time* $O(\log(n))$*-size-approximation algorithm.*

(The proof is in the appendix. The best comparable previous result is a bicriteria approximation algorithm that finds a facility set of size $\ln(n + n/\epsilon)k$ and cost $(1 + \epsilon)opt_k$ [22].)

Theorems 2 and 3 imply the size-competitive ratios shown in Fig. 1. Note also that no polynomial-time algorithm (online or offline) is $o(\log n)$-size-competitive unless P=NP, even for the metric case.

To analyze online medians, we reduce the size- and cost-competitive online problems to the following folklore online problem, which we call "*online bribery.*": 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 $T \in \mathbb{R}^+$. Its cost is the total 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 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 $\mathcal{U}$. (This restriction only helps the algorithm.)

For $\mathcal{U} = \mathbb{R}^+$, it is known that an optimal deterministic strategy bids increasing powers of 2, and that there is a better randomized strategy which bids (randomly translated) powers of e. We complete the complete characterization of the best competitive ratios for online bribery by proving that the randomized strategy is optimal.

Theorem 4 **(a)** *Online bribery has deterministic and randomized algorithms that are* 4*-competitive and* e*-competitive, respectively. Furthermore, if* $\mathcal{U}$ *is finite, the algorithms run in time polynomial in* $|\mathcal{U}|$. **(b)** *No deterministic or randomized algorithm is less than* 4*-competitive or* e*-competitive, respectively, even when restricted to instances of the form* $\mathcal{U} = \{1, 2, \ldots, n\}$*, for some integer* n.

Weighted medians. The k-medians problem generalizes to allow the facilities and the customers to have non-negative weights w. In this case, for a facility set X, one constrains the total weight $\sum_{f \in X} w(f)$ (rather than the size) to be at most k, and one takes $cost(X) = \sum_{u \in \mathcal{C}} w(u) d_{uX}$. All of the results in this paper extend directly to this case.

Approximate triangle inequality. Mettu and Plaxton show that their online median algorithm also works in what they call "λ-approximate" metric spaces, achieving cost-competitive-ratio $O(\lambda^4)$ [19, 20]. We reduce this ratio to $O(\lambda^2)$.

Let $d : \mathcal{F} \times \mathcal{C} \to \mathbb{R}^+$ be a cost function. We say that d is a *λ-relaxed metric* if $d_{fy} \le \lambda(d_{fx} + d_{xg} + d_{gy})$ for any $f, g \in \mathcal{F}$ and $x, y \in \mathcal{C}$. (This is a less restrictive condition than assumed by Mettu and Plaxton, but it appears that their analysis would go through with this condition too. A related concept was studied in [9].) In this setting Theorem 1 generalizes as follows:

Theorem 5 **(a)** *Online λ-relaxed metric medians has non-polynomial-time deterministic and randomized algorithms that are* $8\lambda^2$*-cost-competitive and* $2e\lambda^2$*-cost-competitive, respectively.* **(b)** *If* *offline λ-relaxed metric k-median has a polynomial-time c-cost-approximation algorithm , then the online problem has deterministic and randomized polynomial-time algorithms that are $8\lambda^2 c$-cost-competitive and $2e\lambda^2 c$-cost-competitive, respectively.*

The proof is a generalization of the proof of Theorem 1. We sketch the proof in the appendix.

The kl-medians problem. As shown in [19, 20], no online algorithm for the online medians problem can be better than 2-competitive. The lower bound proof is an adversary strategy that uses only two possible numbers of medians, either 1 or k, for some large k. A natural question to ask is whether better competitive ratios are possible if the number of medians can take only some limited number of values. Here, we completely solve the kl-median case, where the number of medians is either k or $l > k$.

Theorem 6 *For any values of k, l, with $k < l$, there is a deterministic online algorithm for kl-medians with competitive ratio $2 - 1/l$, and no better ratio is possible.*

Online fractional medians. A *fractional k-median* is a solution to the following linear program, which is the relaxation of the standard integer program for the k-median problem:

$$\text{minimize}_{x,y} \sum_u \sum_f x_{uf} d_{uf} \;\; \text{subject to} \;\; \begin{cases} \sum_f y_f & \leq & k & \\ \sum_f x_{if} & \geq & 1 & (\forall i) \\ x_{if} & \leq & y_f & (\forall i, f) \\ x_{if}, y_f & \geq & 0 & (\forall i, f). \end{cases}$$

The natural *online* version of this fractional problem is to find a $c \geq 1$ and, for every integer $k \in [n]$ simultaneously, a pair $(x_{if}^{(k)}), (y_f^{(k)})$ meeting the constraints of the above linear program, as well as $y_f^{(k)} \leq y_f^{(k+1)}$ (for all f) and $\sum_u \sum_f x_{uf} d_{uf} \leq c \cdot opt_k$ (where opt_k is the minimum cost of any fractional k-median). The goal of the online problem is to minimize the competitive ratio c.

An possible approach (different than the one we use here) for finding a good online (non-fractional) solution would be to solve the online fractional problem above optimally using linear programming, and then to somehow randomly round that solution to an online integer solution. To analyze such an approach, it may be useful to have a bound on the competitive ratio c for the fractional problem. By adapting the analyses of our cost-competitive algorithms to the integer problem, we can show the following:

Theorem 7 *Online fractional metric medians has a deterministic polynomial-time algorithm that is $2e$-cost-competitive.*

The proof is an extension of the proof for the non-fractional case, along with the observation that the randomized algorithm for the fractional problem can be derandomized without increasing the competitive ratio. We omit the proof from this abstract.

Bicriteria approximations for online metric medians. Combining Theorem 2, Theorem 8, and offline bicriteria results from [2, 17, 18, 14], we can obtain online, polynomial-time algorithms with the following bicriteria (c, s)-competitiveness guarantees for online metric medians. The first quantity c is the competitive ratio with respect to cost and the second quantity s is the competitive ratio with respect to size: (a) $(3 + \epsilon, 4)$, for any $\epsilon > 0$, (b) $(2 + \epsilon, 4(1 + 2\epsilon^{-1}))$, for any $\epsilon > 0$, (c) $(1 + \epsilon, 4(3 + 5\epsilon^{-1}))$, for any $\epsilon > 0$.

Notation. Throughout we use the following terminology for online bribery. Given the universe $\mathcal{U}$, the algorithm outputs a *bid set* $\mathcal{B} \subseteq \mathcal{U}$. Against a particular threshold T, the algorithm pays for the bids $\{b \in \mathcal{B} : b \leq T^+\}$, where $T^+ = \min\{b \in \mathcal{B} : b \geq T\}$. The bid set $\mathcal{B}$ is *β-competitive* if, for any $T \in \mathcal{U}$, this payment is at most βT.

Also, $\mathbb{R}^+$ 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 $[n] = \{1, 2, \ldots, n\}$.

Plan of the paper. We prove our upper bounds on competitive algorithms for online medians (Theorem 1 for cost-competitive algorithms and Theorem 2(a) for size-competitive algorithms) by reducing online medians to online bribery (Theorem 8, below) and then proving the upper bounds for online bribery (Theorem 4).

We prove our lower bounds on size-competitive algorithms for online medians (Theorem 2(b)) by reducing online bribery to size-competitive medians (Theorem 9, below) and then proving the lower bounds for online bribery in Theorem 4.

We prove the reductions in Section 2 and analyze online bribery in Section 3. In Section 4 we prove Theorem 6.

2 Reductions between online medians and online bribery

We start with the reductions from online medians to online bribery. We show that (a) $2c\beta$-cost-competitive online metric medians reduces (in polynomial time) to β-competitive online bribery and c-cost-approximate offline medians, and (b) $s\beta$-size-competitive online medians reduces (in polynomial time) to β-competitive online bribery and s-size-approximate offline medians.

Note that part (b) holds even for non-metric medians. Also, if allowing non-polynomial time, one can take F_k^* to be the optimal k-median in Theorem 8, which is both 1-cost-approximate and 1-size-approximate; then the online solution $\bar{F}$ is (a) 2β-cost-competitive or (b) β-size-competitive.

Theorem 8 *Let $\beta \geq 1$ and suppose that there exists a (possibly randomized) polynomial-time β-competitive algorithm for online bribery. Fix an instance of k-median.*

(a) *If the problem is metric and for each $i \in [1, n]$, we have a set of facilities F_i^* with size at most i and cost at most $c \cdot opt_i$, then we can compute an* online *solution $(F_i)_i$ where F_i has size at most i and cost at most $2c\beta \cdot opt_i$ in (possibly randomized) polynomial time.*

(b) *If for each $i \in [1, n]$, we have a set of facilities F_i^* with size at most $s \cdot i$ and cost at most opt_i, then we can compute an* online *solution $(F_i)_i$ where F_i has size at most $s\beta \cdot i$ and cost at most opt_i in (possibly randomized) polynomial time.*

Proof: We first prove part (a) of Theorem 8 in the deterministic case. The proof in the randomized setting is similar and we omit it.

For convenience, we introduce distances between facilities: given two $f, g \in \mathcal{F}$, let $d'_{fg} = \min_{x \in \mathcal{C}}(d_{fx} + d_{xg})$. This extension satisfies the triangle inequality. By assumption, each F_k^* is c-cost-approximate: $|F_k^*| \leq k$ and $cost(F_k^*) \leq c \cdot opt_k$. Assume without loss of generality that $cost(F_k^*) \leq cost(F_{k+1}^*)$ for all k. (Otherwise, repeatedly pick the minimum k violating this condition, and replace F_{k+1}^* by F_k^*.)

The algorithm constructs the online solution $(F_i)_i$ from $(F_i^*)_i$ 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. Next, compute F_k just for $k \in \mathcal{K}$. Start by defining $F_{\kappa(m)} = F_{\kappa(m)}^*$. Then, working backwards, inductively define $F_{\kappa(i)}$ to contain the facilities within $F_{\kappa(i+1)}$ that are "closest" to $F_{\kappa(i)}^*$.

More precisely, given two subsets A, B of $\mathcal{F}$, let $\Gamma(A, B)$ denote a subset Γ of B, minimal with respect to inclusion, and such that $d'_{\mu\Gamma} = d'_{\mu B}$ for all $\mu \in A$ (breaking ties arbitrarily). Obviously, $|\Gamma(A, B)| \leq |A|$, and $\Gamma(A, B)$ can be computed in polynomial time given A and B. Then $F_{\kappa(i)} = \Gamma(F^*_{\kappa(i)}, F_{\kappa(i+1)})$.

Finally, define F_k for $k \notin \mathcal{K}$ as follows. Let $k^- = \max\{i \in \mathcal{K} : i \leq k\}$ (the largest index in $\mathcal{K}$ less than or equal to k, which always exists since $1 \in \mathcal{K}$) and define $F_k = F_{k^-}$. To complete the construction, it only remains to describe how to compute the initial index set $\mathcal{K}$, which we momentarily defer.

To analyze the size, note that $|F_k| \leq k$, because for $k \in \mathcal{K}$, by definition of Γ we have $|F_k| \leq |F^*_k| \leq k$, while for $k \notin \mathcal{K}$, we have $|F_k| = |F_{k^-}| \leq k^- < k$.

To analyze the cost, we will use the following Lemma from [8] whose proof requires the space to be metric.

Lemma 1 *[8] Consider any two sets $A, B \subseteq \mathcal{F}$ and let $\Gamma = \Gamma(A, B)$. Then for every $x \in X$ we have $c_{x\Gamma} \leq 2c_{xA} + c_{xB}$.*

We now claim that

$$cost(F_k) \;\leq\; 2 \sum_{\ell \geq k^-, \ell \in \mathcal{K}} cost(F^*_\ell). \tag{1}$$

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

Since $cost(F^*_k) \leq c\, opt_k$, to make F $2c\beta$-cost-competitive we will choose $\mathcal{K}$ so that, for all k,

$$\sum_{\ell \geq k^-, \ell \in \mathcal{K}} cost(F^*_\ell) \;\leq\; \beta\, cost(F^*_k). \tag{2}$$

To compute the set $\mathcal{K}$, let $\mathcal{U} = \{cost(F^*_n), cost(F^*_{n-1}), \ldots, cost(F^*_1)\}$ and take $\mathcal{B}$ to be any β-competitive bid set for universe $\mathcal{U}$. Define $\mathcal{K} = \{\kappa(m), \kappa(m-1), \ldots, \kappa(1)\}$ to be a minimal set (containing 1) such that the bid set is $\mathcal{B} = \{cost(F^*_{\kappa(m)}), cost(F^*_{\kappa(m-1)}), \ldots, cost(F^*_{\kappa(1)})\}$. Then the left-hand side of (2) is exactly the sum of the bids paid from the bid set for threshold $T = cost(F^*_k)$. Since the bid set is β-competitive, this is at most $\beta\, cost(F^*_k)$, so (2) holds.

This completes the proof of part (a).

We now prove part (b) of Theorem 8. By assumption each F^*_k is s-size-approximate, that is, $|F^*_k| \leq sk$ and $cost(F^*_k) \leq opt_k$.

Fix some β-competitive bid set $\mathcal{B}$. Let $\mathcal{B}_k$ be the set of bids in $\mathcal{B}$ paid against threshold $T = k$ with $\mathcal{U} = \{1, 2, \ldots, n\}$. Define $F_k = \bigcup_{b \in \mathcal{B}_k} F^*_b$. Then $\bar{F} = (F_1, F_2, ..., F_n)$ is an online solution because $\mathcal{B}_k \subseteq \mathcal{B}_\ell$ for $\ell \geq k$. Further, $cost(F_k) \leq opt_k$ because F_k contains F^*_b for some $b \geq k$, so $cost(F_k) \leq cost(F^*_b) \leq opt_b \leq opt_k$. Since $\mathcal{B}$ is β-competitive, we have $|F_k| \leq \sum_{b \in \mathcal{B}_k} |F^*_b| \leq \sum_{b \in \mathcal{B}_k} sb \leq s\beta k$. □

Our next reduction will prove that competitive online bribery reduces to size-competitive online medians. Note that, together with Theorem 8(b), this implies that online bribery and size-competitive online medians are equivalent.

Theorem 9 *Let $s \geq 1$ and suppose that, for online medians (either metric or not), there is a (possibly randomized) s-size-competitive algorithm. Then, for any integer n, for online bribery with $\mathcal{U} = \{1, 2, \ldots, n\}$, there is a (randomized) s-competitive algorithm.*

Proof: We give the proof in the deterministic setting. The proof in the randomized setting is similar and we omit it.

For any arbitrarily large m, we construct a metric space $\mathcal{C} \times \mathcal{F}$ and a sequence of disjoint facility sets $M_1, M_2, \ldots, M_m \subseteq \mathcal{F}$, where $|M_k| = k$ for each k, with the following properties:

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

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

Note that these conditions imply that each M_k is the unique optimum k-median.

Assuming for the moment that there exists such a metric space, we show that any s-size-competitive online median $\bar{F}$ for it corresponds to an s-competitive bid set $\mathcal{B}$ for online bribery with universe $\mathcal{U} = [m]$. The reason is that, by (ii), the only way for F_T to have cost at most opt_T is for F_T to contain M_k for some $k \geq T$. Since the M_k's are disjoint, and the F_T's have to be nested, this effectively turns the problem into online bribery.

Define $\mathcal{B} = \{k : M_k \subseteq F_k\}$. 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

$$\sum\{k : k < T, M_k \subseteq F_k\} \; + \; \min\{\ell : \ell \geq T, M_\ell \subseteq F_\ell\}. \tag{3}$$

Since $M_k \subseteq F_k$ implies $M_k \subseteq F_T$ for $k < T$, and $M_\ell \subseteq F_T$ implies $M_\ell \subseteq F_\ell$ for $\ell \geq T$, we obtain that Expression (3) is at most

$$\sum\{k : k < T, M_k \subseteq F_T\} \; + \; \min\{\ell : \ell \geq T, M_\ell \subseteq F_T\}. \tag{4}$$

(By (ii), $M_\ell \subseteq F_T$ for some $\ell \geq T$, so the minimum on the right is well-defined for $T \in [m]$.) Since $|M_k| = k$, and the M_k's are disjoint, Expression (4) is at most

$$\sum\{|M_k| : k < T, M_k \subseteq F_T\} + \sum\{|M_\ell| : \ell \geq T, M_\ell \subseteq F_T\} \; \leq \; \sum\{|M_k| : M_k \subseteq F_T\} \; \leq \; |F_T| \leq sT.$$

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

We now present the construction of the sets M_j satisfying conditions (i) and (ii). Let $\mathcal{C}$ be the set of vectors $\bar{u} = (u_1, u_2, \ldots, u_m)$ where $1 \leq u_\ell \leq \ell$ for all $\ell = 1, 2, \ldots, m$. Clearly, $|\mathcal{C}| = m!$. For each ℓ, let $\mathcal{C}_{\ell,i}$ be the set of those $\bar{u}$ with $u_\ell = i$. Then $\mathcal{C}_{\ell,1}, \mathcal{C}_{\ell,2}, \ldots, \mathcal{C}_{\ell,\ell}$ is a partition of $\mathcal{C}$. For each ℓ, introduce a set $M_\ell = \{\mu_{\ell,1}, \mu_{\ell,2}, \ldots, \mu_{\ell,\ell}\}$, and for each $i = 1, \ldots, \ell$, connect all nodes in $\mathcal{C}_{\ell,i}$ to $\mu_{\ell,i}$ with an edge of length $\delta_\ell = 1 + (m!)^{-\ell}$. The set of facilities is $\mathcal{F} = \bigcup_\ell M_\ell$. All distances in $\mathcal{C} \times \mathcal{F}$ other than those specified above are determined by shortest-path distances. The resulting distance function satisfies the triangle inequality.

We have $\mathit{cost}(M_j) = m!\delta_j$ for each j, so (i) holds. We prove (ii) by contradiction. Fix some index j and consider a set $F \subseteq \mathcal{F}$ that does not contain any M_ℓ for $\ell \geq j$. This means that 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 this $\bar{u}$ is served either by some $\mu_{\ell,i} \in F$ with $\ell < j$, or by some $\mu_{\ell,i} \in F$ with $\ell \geq j$ and $i \neq i_\ell$. So the closest facility to $\bar{u}$ is at distance at least δ_{j-1}. Therefore we get $\mathit{cost}(F) > m! - 1 + \delta_{j-1} = m!\delta_j = \mathit{cost}(M_j)$, which is a contradiction. □

3 Online bribery

In this section we prove Theorem 4. For completeness, we give proofs of the (folklore) deterministic and randomized upper bounds and deterministic lower bound. This type of doubling algorithms has been used in papers such as [21, 10, 3, 4] and it is in fact possible that some of the proofs therein can be rewritten via a reduction to online bribery. We then prove a randomized lower bound which matches the upper bound.

Lemma 2 (upper bound for deterministic online bribery) *For online bribery, there is a deterministic* 4*-competitive algorithm. If* $\mathcal{U}$ *is finite, the algorithm runs in time polynomial in* $|\mathcal{U}|$.

Proof: First consider the case $\mathcal{U} = \mathbb{R}^+$. Define the algorithm to produce the set of bids $\{0\} \cup \{2^j : j \in \mathbb{N}\}$. Let $i = \lceil \log_2 T \rceil$, where $T > 0$ is the threshold: the algorithm pays $\sum_{j \le i} 2^j = 2^{i+1} \le 4T$, hence is 4-competitive.

Next, we reduce the general case to the case $\mathcal{U} = \mathbb{R}^+$. Knowing that $T \in \mathcal{U}$, the algorithm, when it would have bid $b \notin \mathcal{U}$, will instead bid the next smaller bid in $\mathcal{U}$ (if there is one, and otherwise the bid is skipped). This only decreases the cost the algorithm pays against any threshold $T \in \mathcal{U}$. Note that the modified algorithm can be implemented in time polynomial in $|\mathcal{U}|$ if $\mathcal{U}$ is finite. $\square$

Lemma 3 (lower bound for deterministic online bribery) *For online bribery, no deterministic algorithm is less than* 4*-competitive, even for* $\mathcal{U} = \mathbb{N}^+$.

Proof: let x_n be the nth bid, $s_n = \sum_1^n x_i$ and $y_n = s_{n+1}/s_n$. Suppose, for a contradiction, that there exists $a < 4$ such that $s_{n+1}/x_n < a$ for all n. Rewriting, we get $y_{n+1} \le (1 - 1/y_n)a$. Since $1 - 1/z < z/4$, this implies $y_{n+1} < (y_n/4)a$; thus $y_n < (a/4)^n y_0$, and so eventually $s_{n+1} < s_n$, which is a contradiction. $\square$

Lemma 4 (upper bound for randomized online bribery) *For online bribery, there is a randomized* e*-competitive algorithm. If* $\mathcal{U}$ *is finite, then the algorithm runs in time polynomial in* $|\mathcal{U}|$.

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

For the analysis, let random variable b be the largest bid paid by the algorithm against threshold $T > 0$. The total paid by the algorithm is less than $\sum_{i=0}^{\infty} be^{-i} = be/(e-1)$. Since b/T is distributed like e^ξ where ξ is distributed uniformly in $[0, 1)$, the expectation of b is $T \int_0^1 e^z \, dz = T(e-1)$. Thus, the expected total payment is eT, and the algorithm is e-competitive.

The general case reduces to the case $\mathcal{U} = \mathbb{R}^+$ just as in the proof of Lemma 2. $\square$

Lemma 5 (lower bound for randomized online bribery, part 1) *Fix any* $n \in \mathbb{N}^+$*. Suppose* $\mu : [n] \to \mathbb{R}^+$ *and* $\pi : [n] \to \mathbb{R}^+$ *satisfy*

$$\sum_{T=t}^{n} \frac{1}{T}\pi(T) \;\ge\; \frac{1}{b}\sum_{T=t}^{b} \mu(T) \quad (\forall b, t : 1 \le t \le b \le n). \tag{5}$$

For online bribery with $\mathcal{U} = [n]$*, there is no randomized algorithm with competitive ratio better than* $\sum_{T=1}^{n} \mu(T) \;/\; \sum_{T=1}^{n} \pi(T)$.

Proof: Consider a random set $\mathcal{B}$ of bids generated by any β-competitive randomized algorithm when $\mathcal{U} = [n]$. Without loss of generality, the maximum bid in $\mathcal{B}$ is n.

Let $\mathcal{B} = \{b_1, b_2, \ldots, b_m = n\}$ be the ordered sequence of bids in $\mathcal{B}$. Consider the sequence of intervals $([1, b_1], [b_1+1, b_2], [b_2+1, b_3], \ldots, [b_{m-1}+1, b_m])$, which exactly covers the points $1, 2, \ldots, n$. Let $x(t, b)$ denote the probability (over all random $\mathcal{B}$) that $[t, b]$ is one of these intervals. The algorithm pays bid b against threshold T if and only if, for some integer $t \le T$, $[t, b]$ is one of these intervals. Thus, for any threshold T and bid b, $\sum_{t=1}^{T} x(t, b)$ is the probability that bid b is made against threshold T. (We will use this below.)

We claim that β and x form a feasible solution to the following linear program (LP):

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

1. The first constraint is met because, for any threshold T, $\sum_{t \le T; b} b\, x(t, b)$ is the expected sum of the bids made by the algorithm if T is the threshold. This is at most βT because the algorithm has competitive ratio β.

2. The second constraint is met because for any threshold T, the algorithm must have at least one bid above the threshold, hence at least one $[t, b]$ with $t \le T \le b$.

Thus, the value of this linear program (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):

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

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 $\sum_T \mu(T) \,/ \sum_T \pi(T)$. Since the value of any feasible dual solution is a lower bound on the value of any feasible solution to the primal, it follows that the competitive ratio β of the randomized algorithm is at least $\sum_T \mu(T) \,/ \sum_T \pi(T)$. $\square$

Lemma 6 (lower bound for randomized online bribery, part 2) *There exists $\mu : [n] \to \mathbb{R}^+$ and $\pi : [n] \to \mathbb{R}^+$ satisfying Condition (5) of Lemma 5 and such that $\sum_T \mu(T)/\sum_T \pi(T) \ge (1 - o(1))e$.*

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 (5) holds, and then show that the corresponding lower bound is $e(1-o(1))$ as $U \to \infty$. Define

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

If $T \ge U^2$, then the right-hand side of Condition (5) 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 (5) then has value

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

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

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

(The second equation follows by calculus, for the maximum occurs when $b = e(t-1)$.)

Thus, Condition (5) is met for $\alpha = (1-o(1))e$. Then, Lemma 5 gives a lower bound on the competitive ratio of

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

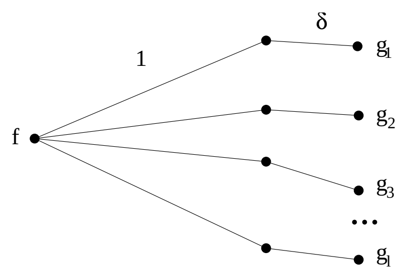
□

Theorem 4 follows directly from Lemmas 2, 3, 4, 5, and 6.

Remark that the reader may be tempted to use a different linear programming formulation, with variables $x(T,b)$ indicating whether bid b is used against threshold T, but that linear program would only yield a lower bound of 1.

4 Online Algorithms for kl-Medians

The lower bound. The lower bound is a slight refinement of the one in [19, 20]. Consider the following metric space. It consists of l customers, where customers j is connected to facility g_j by an edge of length $\delta = 1/l$. All customers are also connected to a facility f with edges of length 1.



Let $G = \{g_1, \ldots, g_l\}$. Then G is the optimal l-median. We have $cost(f) = l$, $cost(G) = l\delta$, $cost(g_i) = \delta + (l-1)(2+\delta)$, and $cost(G - g_i + f) = (l-1)\delta + 1$. So for $\delta = 1/l$, we get:

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

The upper bound. Our upper bound proof is very different from the previous bounds in this paper and it does not rely on online bribery.

Formally, in the kl-median problem we need to compute two sets $F_k \subseteq F_l$ with $|F_k| = k$ and $|F_l| = l$, minimizing the competitive ratio $R = \max\{cost(F_k)/opt_k, cost(F_l)/opt_l\}$.

The algorithm. Let F and G denote, respectively, the optimum k-median and the optimum l-median. Choose the better of two options: either

(a) $F_k = F$ and $F_l = F \cup G - X$, where $X \subseteq G$ is a set of cardinality k that minimizes $cost(F \cup G - X)$, or

(b) $F_k = Y$, where $Y \subseteq G$ is a set of cardinality k that minimizes $cost(Y)$, and $F_l = G$.

For each $f \in F$ and $g \in G$, denote by C_{fg} the cluster 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.) 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 total optimal costs are $cost(F) = \sum_{f\in F, g\in G} w_{fg}a_{fg}$, and $cost(G) = \sum_{f\in F, g\in G} w_{fg}b_{fg}$.

Analysis. The proof is based on a probabilistic argument. For each $f \in F$ let

$$\Lambda_f = \sum_{g\in G} w_{fg}(w_{fG} - w_{fg})^{-1},$$

where $w_{fG} = \sum_{g\in G} w_{fg}$. Define a probability distribution on mappings $\pi : F \to G$ as follows: for any $f \in F$ set $\pi(f) = g$ with probability $w_{fg}(w_{fG} - w_{fg})^{-1}/\Lambda_f$. Fix some ordering of G, and for any such mapping π let X_π be a k-element subset of G that consists of $\pi(F)$ and the first $k - |\pi(F)|$ elements of $G - \pi(F)$, if any.

Using the triangle inequality to reroute some service paths we get

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

Now, by linearity of expectation, we have

$$\begin{aligned}
\text{Exp}[cost(X_\pi) + cost(F \cup G - X_\pi)] &\leq cost(F) + cost(G) \\
&\quad + \sum_{f \in F} \sum_{g \in G} \frac{w_{fg}}{(w_{fG} - w_{fg})\Lambda_f} (w_{fG} - w_{fg})(a_{fg} + b_{fg}) \\
&= cost(F) + cost(G) + \sum_{f \in F} \Lambda_f^{-1} \sum_{g \in G} w_{fg}(a_{fg} + b_{fg}) \\
&\leq (2 - 1/l)[cost(F) + cost(G)].
\end{aligned}$$

The last inequality holds because each Λ_f is minimized when $w_{fg} = w_{fG}/l$ for each $g \in G$.

The above inequality implies that at least one of the following two options must hold: (a) $cost(F_k) = cost(F)$ is optimal and $cost(F_l) = cost(F \cup G - X) \leq \text{Exp}[cost(F \cup G - X_\pi)] \leq (2 - 1/l)cost(G)$, or (b) $cost(F_k) = cost(Y) \leq \text{Exp}[cost(X_\pi)] \leq (2 - 1/l)cost(F)$ and $cost(F_l) = cost(G)$ is optimal. The $2 - 1/l$ upper bound on the competitive ratio follows.

Acknowledgments. The authors wish to thank Yossi Azar for pointing out references to previous results on online bribery and providing a simpler proof of the deterministic lower bound.

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] Vijay Arya, Naveen Garg, Rohit Khandekar, Kamesh Munagala, and Vinayaka 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] Soumen Chakrabarti, Cynthia A. Phillips, Andreas S. Schulz, David B. Shmoys, Clifford Stein, and Joel 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 *29th Annual ACM Symposium on Theory of Computing (STOC)*. ACM, 1997.

[5] Moses Charikar and Sudipto 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] Moses Charikar, Sudipto Guha, Eva Tardos, and David 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] Chandra Chekuri, Ashish Goel, Sanjeev Khanna, and Amit 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] Marek Chrobak, Claire Kenyon, and Neal E. Young. The reverse greedy algorithm for the k-median problem. *IPL: Information Processing Letters.* to appear. A preliminary version appears in the 11th International Computing and Combinatorics Conference (COCOON'05).

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

[10] M. Goemans and J. Kleinberg. An improved approximation ratio for the minimum latency problem. In *Symposium on Discrete Algorithms (SODA).* ACM-SIAM, 1996.

[11] 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.

[12] Kamal Jain and Vijay 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.

[13] Bala Kalyanasundaram and Kirk Pruhs. Speed is as powerful as clairvoyance. *J. ACM*, 47:214–221, 2000.

[14] 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.

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

[16] Guolong Lin, Chandrashekhar Nagarajan, Rajmohan Rajamaran, and David P. Williamson. A general approach for incremental approximation and hierarchical clustering. In *Symposium on Discrete Algorithms (SODA).* ACM-SIAM, 2006.

[17] Jyh-Han Lin and Jeffrey Scott Vitter. Approximation algorithms for geometric median problems. *Information Processing Letters*, 44:245–249, 1992.

[18] Jyh-Han Lin and Jeffrey Scott Vitter. ϵ-approximations with minimum packing constraint violation (extended abstract). In *Proc. 24th Symp. Theory of Computing (STOC)*, pages 771–782. ACM, 1992.

[19] 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.

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

[21] Rajeev Motwani, Steven Phillips, and Eric Torng. Nonclairvoyant scheduling. *Theoretical Computer Science*, 130(1):17–47, 1994.

[22] Neal 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.

Appendix

Proof of Theorem 5. (Sketch.) Fix an instance of the λ-relaxed k-median problem, and let $\bar{F}^* = (F_1^*, \ldots, F_n^*)$ denote a (not necessarily online) sequence of facility sets. Assume that $\bar{F}^*$ is c-cost-approximate (note that c can be 1 for non-polynomial-time algorithms). Extend the λ-relaxed metric to $\mathcal{F} \times \mathcal{F}$ by defining $d'_{fg} = \min_{x \in \mathcal{C}}(d_{xf} + d_{xg})$. The proof hinges on the following straightforward generalization of Fact 1.

Fact 1 *For any two sets $A, B \subseteq \mathcal{F}$, we have $cost(\Gamma(A, B)) \leq \lambda(2cost(A) + cost(B))$.*

Using notation from the proof of Fact 1, Fact 1 follows from $d_{xg} \leq \lambda(d_{x\mu} + d'_{\mu g}) \leq \lambda(d_{x\mu} + d'_{\mu f}) \leq \lambda(2d_{x\mu} + d_{xf})$.

Continue following the proof of Theorem 8, replacing Inequality (1) by

$$cost(F_k) \ \leq \ 2\lambda \sum_{\kappa(j) \geq k^-} \lambda^{j-i} cost(F^*_{\kappa(j)})$$

where $k^- = \kappa(i)$. Instead of a reduction, as in the remark we complete the proof directly by defining $\mathcal{K}$ inductively, setting $\kappa(1) = 1$ and defining $\kappa(i+1)$ to be minimum such that $cost(F^*_{\kappa(i+1)}) \leq cost(F^*_{\kappa(i)})/(2\lambda)$. It is then easy to conclude the proof of the deterministic version of Theorem 5. We omit the proof in the randomized setting.

Proof of Theorem 3. Fix an instance of k-medians. Without loss of generality, assume that $opt_n = 0$, otherwise modify the distance function by subtracting, for each customer u, $d_{u\mathcal{F}}$ from each d_uf. Any c-approximate solution to the modified problem will also be a c-approximate solution to the original problem.

The polynomial-time algorithm from [22] finds a solution of size at most $\ln(n + n/\epsilon)k$ and cost at most $(1+\epsilon)opt_k$. Our algorithm computes such a solution F with $\epsilon = 1/(n-k)$, then adds a single additional facility to minimize the resulting cost. This gives at most $\ln(n + n(n-k))k + 1$ medians.

To finish, we prove that adding the extra facility reduces the cost by at least $opt_k/(n-k)$. To see why, note that if we were to add to F the remaining $n-k$ facilities not in F, one at a time, the cost would be reduced to $opt_n = 0$. Thus, one of those $n-k$ facilities, at the time it would be added, would reduce the cost by at least $opt_k/(n-k)$. Since the cost function is sub-modular, adding that facility to F also reduces the cost by at least $opt_k/(n-k)$.