MATHEMATICS OF OPERATIONS RESEARCH
Vol. 35, No. 1, February 2010, pp. 79–101
ISSN 0364-765X | EISSN 1526-5471 | 10 | 3501 | 0079

DOI 10.1287/moor.1090.0428

A Plant Location Guide for the Unsure: Approximation Algorithms for Min-Max Location Problems

Barbara Anthony
Department of Mathematics and Computer Science, Southwestern University, Georgetown, Texas 78626,
anthonyb@southwestern.edu

Vineet Goyal
Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139,
goyalv@mit.edu

Anupam Gupta
Computer Science Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213,
anupamg@cs.cmu.edu

Viswanath Nagarajan
IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, viswanath@us.ibm.com

This paper studies an extension of the k-median problem under uncertain demand. We are given an n-vertex metric space (V, d) and m client sets $\{S_i \subseteq V\}_{i=1}^m$. The goal is to open a set of k facilities F such that the worst-case connection cost over all the client sets is minimized, i.e.,

$$\min_{F\subseteq V,\,|F|=k} \max_{i\in[m]} \left\{ \sum_{j\in S_i} d(j, F) \right\},$$

where for any $F \subseteq V$, $d(j, F) = \min_{f\in F} d(j, f)$. This is a "min-max" or "robust" version of the k-median problem. Note that in contrast to the recent papers on robust and stochastic problems, we have only one stage of decision-making where we select a set of k facilities to open. Once a set of open facilities is fixed, each client in the uncertain client-set connects to the closest open facility. We present a simple, combinatorial $O(\log n + \log m)$-approximation algorithm for the robust k-median problem that is based on reweighting/Lagrangean-relaxation ideas. In fact, we give a general framework for (minimization) k-facility location problems where there is a bound on the number of open facilities. We show that if the location problem satisfies a certain "projection" property, then both the *robust* and *stochastic* versions of the location problem admit approximation algorithms with logarithmic ratios. We use our framework to give the first approximation algorithms for robust and stochastic versions of several location problems such as k-tree, capacitated k-median, and fault-tolerant k-median.

Key words: approximation algorithms; robust optimization; stochastic optimization; facility location
MSC2000 subject classification: Primary: 90C27, 90C47, 68W25
OR/MS subject classification: Primary: stochastic/robust facility location; approximation algorithms
History: Received July 22, 2008; revised May 30, 2009. Published online in *Articles in Advance* December 8, 2009.

1. Introduction. Consider the following class of *facility location* problems: given a metric space (V, d) with $|V| = n$ locations, and a subset of locations $S \subseteq V$ containing clients that want service, we want to locate a set of k facilities $F \subseteq V$ to minimize the cost of servicing clients S from the facilities F, denoted by $\Phi(F \mid S)$. Because the number of facilities to be opened is constrained by the number k, we refer to such problems as *k-facility location* problems.

Below we list several examples of problems that fall into this category.

(i) k-median: For the k-median problem,

$$\Phi(F \mid S) = \sum_{x\in S} d(x, F),$$

where we define $d(x, F) = \min_{f\in F} d(x, f)$.

(ii) k-center: For the k-center problem,

$$\Phi(F \mid S) = \max_{x\in S} d(x, F).$$

(iii) k-person TSP: For the k-person TSP,

$$\Phi(F \mid S) = \text{minimum total distance traveled by salesmen, one at each } f \in F, \text{ so as to visit all clients in } S.$$

Many k-facility location problems are known to be NP-hard, and have been extensively studied in both the computer science and operations research literature.

In this paper, we study several k-facility location problems under uncertainty in demands, i.e., when the client-set is not fixed in advance. Specifically, we consider the following *stochastic* and *robust* versions of these problems. We are given several sets $S_1, S_2, \ldots, S_m$ of clients, which are called *scenarios*. The goal is to locate k facilities that are *simultaneously good for all scenarios*—more precisely, we want to minimize the objective function

$$\text{Robust-}\Phi = \max_{i=1}^{m} \Phi(F \mid S_i),$$

in the robust (or min-max) version, and

$$\text{Stochastic-}\Phi = \sum_{i=1}^{m} p_i \cdot \Phi(F \mid S_i),$$

in the stochastic version (for given probability values p_i for each scenario S_i). Recall that $\Phi(F \mid S_i)$ denotes the cost of servicing client set S_i using the set of facilities F.

The robust and the stochastic versions of these location problems naturally model cases with uncertain or dynamic systems. For instance, we might want to locate our facilities knowing that one of several scenarios are likely to happen but we do not know which. Or, we might know consumer demand patterns on each day of the week (and maybe on special holidays) and might want to locate facilities to be simultaneously good given these scenarios. Note that these problems only have a *single stage of decision-making*, in contrast to much work that has been done on two-stage stochastic optimization (Birge and Louveaux [2], Immorlica et al. [22], Ravi and Sinha [32], Gupta et al. [19], Shmoys and Swamy [36]).

1.1. Our results and techniques. We use the k-median problem as an example to illustrate the basic ideas of our algorithm. We present an $O(\log m + \log n)$-approximation algorithm for the robust k-median problem in §3 where m is the number of different client sets and n is the number of vertices in the given metric. The algorithm uses ideas from the classical reweighting/Lagrangean-relaxation techniques (see, e.g., Welzl [42], Cesa-Bianchi and Lugosi [4]) in conjunction with a reverse-greedy algorithm (Chrobak et al. [8]). We note that the natural approach to solving the problem by embedding the metric space into a tree metric does not seem to give us an advantage here as we do not know how to obtain a better than logarithmic approximation for the problem even on a *uniform metric*. (The uniform metric is one where all points are at equal distance from each other, and it is a tree metric, because it can be represented as the shortest-path metric on the leaves of the unweighted star graph $K_{1,n}$.)

We then show that, in fact, a similar algorithm works for any k-facility location problem that satisfies the following "β-projection" property for the single-scenario version (this is formalized in (6)).

> Given any instance of a k-location problem with objective function Φ, client set S, and an infeasible solution F with $K > k$ facilities, there are $K - k$ facilities $F' \subseteq F$ such that shutting down a random facility in F' (chosen uniformly) causes the cost to rise in expectation by at most $\beta/(K-k)$ times the optimum.

To give some intuition for this property, consider the k-median problem and the special case when the set F *contains the optimal solution* F^*: in this case we can set $F' = F \setminus F^*$ and when we close a facility $f \in F'$, we assign all the clients originally assigned to f to the facilities these clients were assigned to in $F^* = F \setminus F'$. The sum over all $f \in F'$ of the cost increase in shutting down facility f is at most OPT, where OPT denotes the optimal objective value. Hence the average cost increase of shutting down a facility in F' is at most $\mathsf{OPT}/|F'| = \mathsf{OPT}/(K-k)$. Note that we looked only at a special case, and one has to consider other cases when $F^* \not\subseteq F$, but loosely, the projection property says that even if $F^* \not\subseteq F$, we can "project" the F^* onto some k vertices in F, such that closing a random facility from the other $K - k$ facilities $F \setminus F^*$ behaves more-or-less in the above-mentioned fashion.

In §4, we show that for any k-location problem Φ with the above β-projection property and where the objective function Φ is computable in polynomial time, there is

(i) an $O(\beta \cdot (\log n + \log m))$-approximation algorithm for the robust version of Φ, and
(ii) an $O(\beta \cdot \log n)$-approximation algorithm for the stochastic version of Φ.

Additionally, the algorithm for stochastic k-location problems is *incremental* (Mettu and Plaxton [28], Lin et al. [26]) in the following sense. We obtain a permutation π of all locations V such that for any bound $1 \le t \le n$ on the number of facilities, $\{\pi_1, \ldots, \pi_t\}$ is an approximately optimal solution to the stochastic t-location problem.

We show that the projection property holds for the following problems with $\beta = O(1)$:

(i) Hard-capacitated k-median with uniform capacities (the nonuniform soft-capacitated version was studied in Chuzhoy and Rabani [9]),

(ii) Fault-tolerant k-median with nonuniform requirements (the uniform version was studied in Swamy and Shmoys [38]), and

(iii) k-tree.

Hence, the robust and the stochastic versions of all these problems admit logarithmic approximation guarantees. We also note that the results for hard-capacitated k-median and nonuniform fault-tolerant k-median seem to be the first logarithmic approximation guarantees known for even the deterministic versions of these problems (where there is only one scenario or client set).

Finally, we show that not all natural k-facility location problems give good results using this framework, because they do not satisfy the projection property. In particular, we show that the stochastic k-center problem is as hard to approximate as the (minimization) *dense-k-subgraph* problem. Dense-k-subgraph is a well-studied problem for which the best approximation guarantee is $O(n^{\delta})$ (for some constant $\delta < 1/3$) (Feige et al. [16]), and improving on this is a long-standing open question.

We would like to point out that in all the k-location problems we consider, we do not have costs associated with opening facilities at specific locations.

1.2. Related work. Location problems under uncertainty have long been studied in the operations research literature because of their vast applicability in real-world scenarios. Sheppard [35] used a scenario-based approach to model uncertainty in demand and minimize the expected cost, while Cooper [11] was among the first to consider the robust objective on location problems. Following this, similar models for location problems such as k-median and uncapacitated facility location were studied (Mirchandani and Odoni [29], Weaver and Church [41], Rosenblatt and Lee [33]). See Louveaux [27] and Daskin and Owen [12] for more thorough surveys of location problems under uncertainty with robust and stochastic objectives; a good summary can be found in the recent survey by Snyder [37]. The papers by Van Hentenryck et al. [39] have also proposed online stochastic algorithms for some stochastic location problems. However, to the best of our knowledge, no algorithms with provable guarantees have been given for robust k-median and the other stochastic/robust location problems we consider in our work.

In the single-scenario case, many results are known for the k-median problem (Charikar et al. [7], Charikar and Guha [6], Jain and Vazirani [23], Arya et al. [1], Mettu and Plaxton [28], Chrobak et al. [8]) as well as its capacitated (Chuzhoy and Rabani [9]) and fault-tolerant versions (Swamy and Shmoys [38]), and k-center problems (Sahni and Gonzalez [34], Hochbaum and Shmoys [20]). Out of these, the one most relevant to our work is the reverse-greedy algorithm of Chrobak et al. [8] whose work we adapt and extend: our proofs of the projection property give reverse-greedy $O(\log n)$-approximation algorithms for all the problems we consider.

While facility location problems have been considered in the context of stochastic optimization (see, e.g., Immorlica et al. [22], Ravi and Sinha [32], Gupta et al. [19], Shmoys and Swamy [36]), and robust optimization (see, e.g., Dhamdhere et al. [13], Golovin et al. [18], Feige et al. [15]), it is not clear how to use the techniques in these previous papers to solve the problems we consider where we have a strict bound on the number of open facilities.

Bicriteria results for robust versions of profit maximization k-location problems (e.g., locating k depots such that one salesman can start at each of these depots and travel for at most some time budget B, so as to maximize the number of clients visited) can be obtained by recent work on robust submodular function maximization by Krause et al. [25].

Outline. In §3, we present our results for the robust k-median problem. We first consider the case of uniform metrics, which gives many of the ideas, and then extend the ideas to general metrics. We then abstract out the general framework in §4. In the following sections (§§5, 6, and 7), we show that the k-tree problem, capacitated k-median problem, and the fault-tolerant k-median problem satisfy our general framework and thus admit $O(\log m + \log n)$-approximation guarantees for their robust version and $O(\log n)$-approximations for their stochastic versions. Finally, we give evidence of the hardness of approximating stochastic k-center in §8.

2. Notation and preliminaries. In the following discussion, we consider finite metric spaces (V, d) with $|V| = n$ points. The function $d\colon V \times V \to \mathbb{R}_+$ satisfies the following two conditions:

(i) $d(u, v) + d(v, w) \geq d(u, w)$ for all $u, v, w \in V$ (triangle inequality), and

(ii) $d(u, v) = d(v, u)$ for any $u, v \in V$ (symmetry).

A metric (V, d) is uniform if $d(x, y) = 1$ for all $x, y \in V$, $x \neq y$. For a set $S \subseteq V$ and $j \in V$, we define $d(j, S) = \min_{j' \in S} d(j, j')$. We let $\mathsf{diam}(V, d)$ denote the *diameter* of the metric; i.e.,

$$\mathsf{diam}(V, d) = \max_{i, j \in V} d(i, j).$$

For any integer $t \geq 1$, $[t]$ denotes the set $\{1, 2, \ldots, t\}$. All logarithms in the paper are base-2 logarithms, unless otherwise specified. The tth harmonic number is $H_t = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{t}$. We will often use the standard approximation that $H_t = O(\log t)$. We also use the notation $\oplus$ for symmetric difference; i.e., for sets A and B, $A \oplus B = (A \backslash B) \cup (B \backslash A)$.

Approximation algorithm. Given a minimization problem Π and a parameter $\alpha \geq 1$, an *α-approximation algorithm* for the problem Π is an algorithm that, on every input instance $\mathcal{I}$, outputs a feasible solution whose cost is at most a factor α times the cost of an optimal solution for the instance $\mathcal{I}$ (Vazirani [40]) in time polynomial in the length of the input.

3. The robust k-median problem.

In the *robust k-median* problem, we are given an n-vertex metric space (V, d), m subsets $S_1, \ldots, S_m \subseteq V$ of clients, and a bound k on the number of facilities. We want to find a set of k facilities $F \subseteq V$ that minimizes the objective,

$$\max_{i=1}^{m} \sum_{v \in S_i} d(v, F).$$

In this section, we prove the following.

THEOREM 3.1 (ROBUST k-MEDIAN RESULT). *There is an $O(\log m + \log n)$-approximation algorithm for the robust k-median problem where m is the number of client sets and n is the number of vertices in the given metric.*

3.1. A warm up: The uniform metric.

We first study the special case when (V, d) is a uniform metric. The analysis here illustrates the basic ideas for the subsequent algorithms. In the uniform metric case, the problem can be recast as follows:

> Given a ground set V and a family of m sets $S_1, \ldots, S_m \subseteq V$, find a set $F \subseteq V$ where $|F| = k$ such that the *maximum "exposure"* $\max_{i=1}^{m} |S_i \backslash F|$ is minimized.

The set F corresponds to open facilities and the *exposure* of any set S_i is the number of elements in S_i that are left uncovered by the open facilities; i.e., $|S_i \backslash F|$. We first observe that it is NP-hard to approximate robust k-median on uniform metrics to better than a factor of two (see Figure 3 for an illustration).

THEOREM 3.2. *The robust k-median problem on uniform metrics is NP-hard.*

PROOF. We reduce from the decision version of the minimum vertex cover problem: given a graph $G = (V, E)$ and a parameter k, the goal is to decide if there is a subset $V' \subset V$ with $|V'| \leq k$ such that for each edge $(u, v) \in E$, at least one of u and v is in V'.

Given an instance $\langle G = (V, E), k \rangle$ of vertex cover, we construct a robust k-median instance as follows. We consider a uniform metric on the vertex set V, and corresponding to each edge $e = (u, v) \in E$ there is a scenario $S_e = \{u, v\}$. The goal is to open k facilities F so as to minimize the maximum exposure, $\max_{e=(u,v) \in E} |\{u, v\} \backslash F|$.

If G has a vertex cover V' of size k, then setting $F = V'$ would cover at least one vertex from each set $\{S_e \mid e \in E\}$, and hence the optimal value of robust k-median is at most 1. On the other hand, if G has no vertex cover of size at most k, then any choice of $F \subseteq V$ (of size k) would miss both vertices of some edge in E. Hence in this case, the optimal value of the robust k-median instance would be 2. Because the vertex cover problem is NP-hard, this proves the theorem.

The same reduction implies that robust k-median on general metrics is $(2 - \varepsilon)$-hard to approximate for any $\varepsilon > 0$. For this, we modify the above uniform metric by introducing L (some large number) copies of each vertex in V (all copies of a vertex are at zero distance from each other), and for each edge $e = (u, v) \in E$ scenario S_e consists of all copies of vertices u and v. In this case, if G has a vertex cover of size k, then the optimal value is L; otherwise the optimal value is $2L$. Thus it is NP-hard to approximate robust k-median on general metrics to better than factor of two. □

Our algorithm for the uniform metric robust k-median problem is based on maintaining "weights" for each scenario and reweighting them appropriately. This technique is similar to the *Experts algorithm* in learning theory (see the survey by Blum [3]), and the fast combinatorial algorithms for solving *fractional covering/packing* linear programs (Plotkin et al. [31]).

We first observe that two natural greedy algorithms do not work well for this problem. One simple approach is to start with all elements and repeatedly drop the element that increases the exposure of the fewest sets until the number of elements is k. Another greedy approach would be to repeatedly drop any element that keeps the maximum exposure minimized until the number of elements is k. Appendix B gives bad examples for both these algorithms.

To get a result for minimizing the *maximum* exposure, we "penalize" the newly exposed sets by increasing their weights, so that exposing them further costs us even more. Formally, the algorithm is as stated in Algorithm 1.

Algorithm 1 (Uniform metric robust k-median)

- Set $w_i^1 \leftarrow 1$ $(1 \le i \le m)$ and open facilities $F^1 = V$.
- For $t = 1, \ldots, n-k$ do:
  - (i) For each $v \in F^t$, $W^t(v) := \sum_{i|S_i \ni v} w_i^t$; i.e., total weight of sets containing v.
  - (ii) Let v^t be the element $v \in F^t$ that minimizes $W^t(v)$.
  - (iii) Drop this element to get $F^{t+1} \leftarrow F^t \backslash \{v^t\}$.
  - (iv) Set $w_i^{t+1} \leftarrow 2 \cdot w_i^t$ if $S_i \ni v^t$, and $w_i^{t+1} \leftarrow w_i^t$ if $S_i \not\ni v^t$.
- Output F^{n-k+1} of size k.

The next claim follows immediately from the statement of the algorithm.

Claim 3.1. *If the exposure of some set S_i at the end of the algorithm is l, then its weight is 2^l.*

Claim 3.2. *Let $W^t = \sum_{i=1}^m w_i^t$ be the total weight at the beginning of round $1 \le t \le n-k$, and let the maximum exposure of the optimal solution be l^*. Then*

$$W^{t+1} \le W^t \left(1 + \frac{l^*}{n-k-t+1}\right).$$

Proof. Let F^* be the k elements picked in the optimal solution: they expose at most l^* in each of the sets. Note that there are $n-k$ elements in $V \backslash F^*$, and by round t at most $t-1$ of these elements might have been discarded, leaving at least $n-k-t+1$ elements in $F^t \backslash F^*$. Because each set contains at most l^* of these elements, an averaging argument shows that there must be an element such that the total weight of sets containing it is at most $W^t \times l^*/(n-k-t+1)$. Thus $W^t(v^t)$ is at most this quantity. But the weight adjustment step implies that $W^{t+1} = W^t + W^t(v^t)$, which proves the lemma. □

We are now ready to prove the performance guarantee of Algorithm 1.

Theorem 3.3. *If the maximum exposure of the optimal solution is l^*, then the maximum exposure in the solution found by Algorithm 1 is $O(\log n) \cdot l^* + O(\log m)$. Hence there is an $O(\log m + \log n)$-approximation algorithm for robust k-median on uniform metrics.*

Proof. By Claim 3.2, the total weight of the m sets at the end of Algorithm 1 is at most:

$$W^1 \cdot \prod_{t=1}^{n-k} \left(1 + \frac{l^*}{n-k-t+1}\right) \le m \cdot \exp\left\{\sum_{t=1}^{n-k} \frac{l^*}{n-k-t+1}\right\}$$
$$= m \cdot \exp\{l^* H_{n-k}\},$$

where H_t is the tth harmonic number. Now, if some set is exposed l times, its weight (and hence the total weight of all sets) is at least 2^l by Claim 3.1. Therefore,

$$2^l \le m \cdot \exp\{l^* H_{n-k}\}. \tag{1}$$

Taking logarithms on both sides of (1), we get that

$$l \le \log m + O(l^* \cdot H_{n-k}).$$

Finally using $H_t = O(\log t)$ proves the theorem. □

We note that an algorithm based on solving a suitable linear programming relaxation followed by randomized rounding gives an improved $(1+\epsilon)l^* + O(\log m/\epsilon)$ guarantee (with any constant $0 < \epsilon < 1$) for robust k-median on uniform metrics, where l^* is the optimal value; the details appear in Appendix A. However, that algorithm does not extend to the case of general metrics considered in the next section.

3.2. Robust k-median on general metrics. In this section, we generalize the algorithm on uniform metrics to obtain an $O(\log m + \log n)$-approximation algorithm for the robust k-median problem on general metrics. This algorithm is based on a reverse greedy algorithm for k-median because of Chrobak et al. [8] combined with a weight-update scheme similar to the one described above. We assume (by scaling) that distances in the metric are at least one and we let $\Delta := \mathsf{diam}(V, d)$. Then observe that the optimal value of any robust k-median instance lies in the interval $[1, n\Delta]$; recall that $n = |V|$.

In Algorithm 2 (described below), we also assume that we know a value B such that $4 \cdot \mathsf{OPT} \le B \le 8 \cdot \mathsf{OPT}$, where OPT is the optimal value of the given robust k-median instance. This assumption can be discharged by running the algorithm several times, trying all values of B that are powers of two in the interval $[1, 8n\Delta]$ and finally taking the minimum cost solution. We need to try $O(\log(n\Delta))$ different values for B, which is polynomial in the size of the input.

Algorithm 2 (Robust k-median for general metrics).

- Set $w_i^1 \leftarrow 1$ for all $1 \le i \le m$ and $F^1 \leftarrow V$.
- For $t = 1, \ldots, n-k$ do:

(i) For each $v \in F^t$ and each $1 \le i \le m$, let $\delta_i^t(v)$ be the increase in the k-median objective for S_i when the set of facilities changes from F^t to $F^t \backslash \{v\}$; i.e.,

$$\delta_i^t(v) := \sum_{x \in S_i} (d(x, F^t \backslash v) - d(x, F^t)).$$

(ii) Set $\hat{F}^t \leftarrow \{v \in F^t \mid \delta_i^t(v) \le B/2 \ \forall\, 1 \le i \le m\}$.

(iii) Set $v^t \leftarrow \arg\min\{\sum_{i=1}^m w_i^t \cdot \delta_i^t(v) \colon v \in \hat{F}^t\}$. Drop this vertex and set $F^{t+1} \leftarrow F^t \backslash \{v^t\}$.

(iv) For all $1 \le i \le m$, update

$$w_i^{t+1} \leftarrow w_i^t \cdot \left(1 + \frac{1}{B}\right)^{\delta_i^t(v^t)}.$$

- Output F^{n-k+1} with k facilities.

We first prove the following lemma.

LEMMA 3.1. *In any iteration $1 \le t \le n-k$, there exists a set $Q^t \subseteq F^t$ of size at most k such that for each scenario $\{S_i\}_{i=1}^m$,*

$$\sum_{v \in F^t \backslash Q^t} \delta_i^t(v) \le 2 \sum_{x \in S_i} d(x, F^*) \le 2\,\mathsf{OPT},$$

where F^* is an optimal solution to the robust k-median instance, and $\mathsf{OPT} = \max_{i=1}^m \sum_{x \in S_i} d(x, F^*)$.

PROOF. Our arguments here follow those by Chrobak et al. [8]. For each $f^* \in F^*$, let $\eta(f^*) := \arg\min_{g \in F^t} d(f^*, g)$; i.e., the vertex in F^t closest to f^*. Let $Q^t \subseteq F^t$ be the "projection" of F^* onto F^t; i.e., the vertices in F^t closest to F^*. Formally, $Q^t := \{\eta(f^*) \mid f^* \in F^*\}$. Note that the size of the projection is $|Q^t| \le |F^*| = k$.

In the following discussion, fix any $i \in \{1, \ldots, m\}$; the superscripts t are dropped for brevity. Summing the changes $\delta_i(v)$ over all the vertices in $R = F \backslash Q$, we get

$$\begin{aligned}
\sum_{v \in R} \delta_i(v) &= \sum_{v \in R} \left(\sum_{x \in S_i} d(x, F \backslash v) - \sum_{x \in S_i} d(x, F) \right) \\
&= \sum_{x \in S_i} \sum_{v \in R} (d(x, F \backslash v) - d(x, F)) \\
&\le \sum_{x \in S_i} (d(x, Q) - d(x, F)) \qquad (2) \\
&\le \sum_{x \in S_i} (2 \cdot d(x, F^*) + d(x, F) - d(x, F)) \qquad (3) \\
&= 2 \sum_{x \in S_i} d(x, F^*). \qquad (4)
\end{aligned}$$

To derive inequality (2), consider some client $x \in S_i$ and $f_x \in F$ that serves x in solution F (see also Figure 1). If $f_x \in Q$, then for any $v \in R$,

$$d(x, F \backslash v) = d(x, F) \ \Rightarrow \ \sum_{v \in R} (d(x, F \backslash v) - d(x, F)) = 0.$$

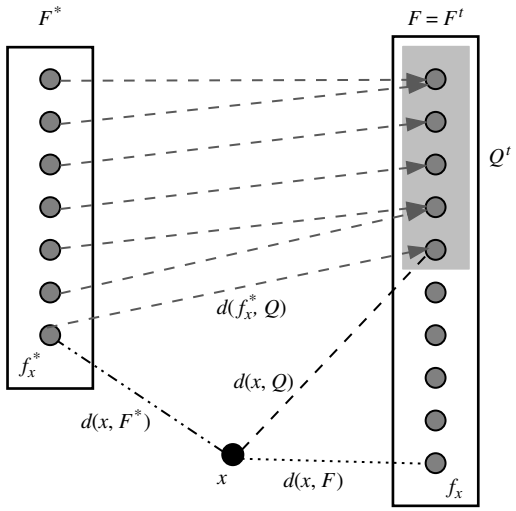

FIGURE 1. Argument for robust k-median. The shaded portion on the right is $Q = Q^t$, and the grey arrows indicate the "projection" from F^* to F.

If $f_x \in R$, then

$$\sum_{v \in R} (d(x, F \backslash v) - d(x, F)) = d(x, F \backslash f_x) - d(x, F) \leq d(x, Q) - d(x, F).$$

To obtain inequality (3), let $f_x^* \in F^*$ be the facility that serves x in solution F^* (see Figure 1). Using the triangle inequality, we have that

$$\begin{aligned} d(x, Q) &\leq d(x, f_x^*) + d(f_x^*, Q) \\ &= d(x, f_x^*) + d(f_x^*, F) \\ &\leq d(x, f_x^*) + d(f_x^*, x) + d(x, F) \\ &= 2 \cdot d(x, F^*) + d(x, F). \end{aligned}$$

Finally, for any $i \in [m]$, the last expression (4) is bounded from above by $2 \cdot \mathsf{OPT}$. □

The following claim ensures that the algorithm is well defined and always terminates with a feasible solution to robust k-median.

CLAIM 3.3. *Assuming $B \geq 4 \cdot \mathsf{OPT}$, in any iteration $1 \leq t \leq n - k$ and in step* (ii) *of Algorithm 2, we have $F^t \backslash Q^t \subseteq \hat{F}^t$; hence $\hat{F}^t$ is nonempty.*

PROOF. By Lemma 3.1, there is a set Q^t (of size at most k) such that $\sum_{v \in F^t \backslash Q^t} \delta_i^t(v) \leq 2 \cdot \mathsf{OPT} \leq B/2$ for all scenarios $i \in [m]$. Moreover, the δ's are nonnegative: hence, each cost increase $\delta_i^t(v)$ is at most $B/2$ for all $v \in F^t \backslash Q^t$ and for all $i \in [m]$, implying that $\hat{F}^t \supseteq F^t \backslash Q^t$. Since Q^t has size at most k, and $|F^t| > k$, the set $\hat{F}^t$ is nonempty. □

CLAIM 3.4. *Assuming $B \geq 4 \cdot \mathsf{OPT}$, in any iteration $1 \leq t \leq n - k$, we have*

$$\min_{v \in \hat{F}^t} \left\{ \sum_{i=1}^m w_i^t \cdot \delta_i^t(v) \right\} \leq \frac{2 \cdot \mathsf{OPT}}{n - k - t + 1} \sum_{i=1}^m w_i^t.$$

PROOF. Let us fix Q^t as in Lemma 3.1. Let us sum the weighted $\delta_i^t(v)$ values, this time both over vertices in $F^t \backslash Q^t$ and over scenarios $i \in [m]$:

$$\sum_{v \in F^t \backslash Q^t} \sum_{i=1}^m w_i^t \cdot \delta_i^t(v) = \sum_{i=1}^m w_i^t \cdot \sum_{v \in F^t \backslash Q^t} \delta_i^t(v) \leq \sum_{i=1}^m w_i^t \cdot 2 \cdot \mathsf{OPT}.$$

The last inequality follows from Lemma 3.1. Finally, simple averaging using the fact that $F^t \setminus Q^t \subseteq \hat{F}^t$ (from Claim 3.3) shows that:

$$\min_{v \in \hat{F}^t} \left\{ \sum_{i=1}^{m} w_i^t \cdot \delta_i^t(v) \right\} \le \min_{v \in F^t \setminus Q^t} \left\{ \sum_{i=1}^{m} w_i^t \cdot \delta_i^t(v) \right\}$$
$$\le \frac{1}{|F^t \setminus Q^t|} \sum_{v \in F^t \setminus Q^t} \sum_{i=1}^{m} w_i^t \cdot \delta_i^t(v)$$
$$\le \frac{2 \cdot \mathsf{OPT}}{n-k-t+1} \sum_{i=1}^{m} w_i^t.$$

The last inequality uses $|F^t \setminus Q^t| \ge n-t-k+1$ since $|F^t| = n-t+1$ and $|Q^t| \le k$. □

Lemma 3.2. *Assume that $B \ge 4 \cdot \mathsf{OPT}$. Let $W^t = \sum_{i=1}^{m} w_i^t$ denote the total weight of all scenarios at the start of iteration $1 \le t \le n-k$. Then the total weight at the start of the next iteration,*

$$W^{t+1} \le W^t \cdot \exp\left(\frac{1}{n-k-t+1}\right).$$

Proof. For any iteration t and scenario i, the weight update step ensures that

$$w_i^{t+1} = w_i^t \left(1 + \frac{1}{B}\right)^{\delta_i^t(v^t)} \le w_i^t \cdot \exp(\delta_i^t(v^t)/B),$$

where $v^t \in \hat{F}^t$ is the facility that is dropped in iteration t. From the definition of the set $\hat{F}^t$, we have $0 \le \delta_i^t(v^t)/B \le 1/2$. Moreover, for $y \in [0, 1/2]$, we have $e^y \le 1 + \sqrt{e} \cdot y$. This implies

$$w_i^{t+1} \le w_i^t \cdot \left(1 + \sqrt{e}\frac{\delta_i^t(v^t)}{B}\right),$$

and hence

$$W^{t+1} \le \sum_{i=1}^{m} w_i^t \cdot \left(1 + \sqrt{e}\frac{\delta_i^t(v^t)}{B}\right) = W^t + \frac{\sqrt{e}}{B} \cdot \sum_{i=1}^{m} w_i^t \cdot \delta_i^t(v^t).$$

Using Claim 3.4 and the facts that $B \ge \mathsf{OPT}/4$ and $\sqrt{e} < 2$, it follows that

$$W^{t+1} \le W^t + \frac{\sqrt{e}}{B} \cdot \frac{2 \cdot \mathsf{OPT}}{n-k-t+1} \cdot W^t \le \left(1 + \frac{1}{n-k-t+1}\right) W^t \le W^t \cdot \exp(1/(n-k-t+1)),$$

where the last inequality follows as $1 + x \le e^x$. □

We now prove the main result of this section.

Proof of Theorem 3.1. Let $\mathsf{Alg} = \max_{i=1}^{m} \sum_{t=1}^{n-k} \delta_i^t(v^t)$ denote the value of the solution F^{n-k+1} at the end of the algorithm, and let i_0 be the value of i achieving the maximum in the above expression. Hence, the total weight

$$W^{n-k+1} \ge w_{i_0}^{n-k+1} = \left(1 + \frac{1}{B}\right)^{\mathsf{Alg}}.$$

Furthermore, repeated applications of Lemma 3.2 imply that

$$W^{n-k+1} \le W^1 \cdot e^{H_{n-k}} = m \cdot e^{H_{n-k}}.$$

Taking logarithms and approximating the harmonic number by a logarithm, we get

$$\mathsf{Alg} \le O(\log m + \log n) / \log\left(1 + \frac{1}{B}\right).$$

Using $B = \Theta(\mathsf{OPT})$, $B \ge 1$, and the fact that $\log_2(1+y) \ge y$ for $y \in [0, 1]$, we get

$$\mathsf{Alg} \le O(\log m + \log n) \cdot \mathsf{OPT}. \quad \square$$

In the next section, we show how a similar algorithm works for robust and stochastic location problems satisfying certain properties, and give a general framework for solving such problems.

We note that the *stochastic k-median* problem can be easily reduced to the usual k-median problem with weights on clients. Because this latter problem admits a constant factor approximation algorithm (Jain and Vazirani [23], Arya et al. [1]), the same holds for stochastic k-median as well.

4. A general framework for robust and stochastic location. Consider a location problem Π on a metric space (V, d) where the cost of serving a set of clients $S \subseteq V$ from a set of facilities $F \subseteq V$ is given by $\Phi(F \mid S)$. We assume that Φ is a monotone nonincreasing function in the set of facilities; i.e., $\Phi(F \cup \{x\} \mid S) \leq \Phi(F \mid S)$ for all $F, S \subseteq V$ and $x \in V$. In other words, opening more facilities does not cause the cost to increase. We are also given a parameter $k \leq n = |V|$ and want to choose a set of *at most k facilities* $F \subseteq V$ that minimizes the resulting cost $\Phi(F \mid S)$. For instance, $\Phi(F \mid S) = \sum_{v \in S} d(v, F)$ defines the k-median objective function.

Robust version, Robust(Π). In the robust version Robust(Π) of the location problem Π, we are given m different scenarios $S_1, S_2, \ldots, S_m \subseteq V$ and the goal is to open a set of k facilities F that minimizes

$$\max_{i=1}^{m} \Phi(F \mid S_i).$$

Stochastic version, Stoc(Π). In the stochastic version Stoc(Π) of the location problem Π, we are given scenarios $\{S_i\}_{i=1}^m$ each occurring with probabilities $\{p_i\}_{i=1}^m$ (with $\sum_{i=1}^m p_i = 1$) and the goal is to find a set F of size k that minimizes

$$\sum_{i=1}^{m} p_i \Phi(F \mid S_i).$$

We show that simple greedy-like procedures give good approximations to both these versions of the location problem Π, given that the following properties hold:

P1 (Cost Computation). For any facility set $F \subseteq V$ and client set $S \subseteq V$, the objective value $\Phi(F \mid S)$ is computable in polynomial time. This implies that for any client set $S \subseteq V$, facility set $F \subseteq V$, and $x \in F$, we can compute, in polynomial time, the incremental cost of dropping x:

$$\delta(F, x \mid S) := \Phi(F \backslash x \mid S) - \Phi(F \mid S). \tag{5}$$

Note that the monotonicity property implies that this value is always nonnegative.

P2 (β-Projection). There is a $\beta \geq 1$ such that, for any set $F^* \subseteq V$ of size k and a set $F \subseteq V$ of size greater than k, there exists a "small" set $Q \subseteq F$ of size $|Q| \leq k$ such that for all client-sets $S \subseteq V$,

$$\sum_{x \in F \backslash Q} \delta(F, x \mid S) \leq \beta \cdot \Phi(F^* \mid S). \tag{6}$$

In applications, it also suffices to prove (P1) and (P2) with any lower bound Φ' in place of Φ, that satisfies the following properties:

- Φ' is a γ-*factor* lower bound for Φ; i.e., it satisfies $\Phi'(F \mid S) \leq \Phi(F \mid S) \leq \gamma \cdot \Phi'(F \mid S)$ for all $F, S \subseteq V$. In addition, there is a polynomial-time algorithm that given any $F, S \subseteq V$, outputs a solution satisfying clients S from facilities F, having cost at most $\gamma \cdot \Phi'(F \mid S)$.
- Φ' is monotone; i.e., $\Phi'(F \cup \{x\} \mid S) \leq \Phi'(F \mid S)$ for all $F, S \subseteq V$ and $x \in V$.

If we use Φ' in place of Φ, an additional factor γ appears in the approximation guarantees of Theorems 4.1 and 4.2. This modification is useful in cases where the lower bound Φ' is polynomial-time-computable, but the objective function Φ itself is not; e.g., the k-person TSP (§5).

The first property (P1) naturally arises in a reverse-greedy-style algorithm for location problems. The second property (P2) is only required to prove the performance guarantee: it seems somewhat mysterious at first, and is useful in the same way as Lemma 3.1 was for the robust k-median problem. Proving this property is very problem specific; see §1.1 for some intuition for property (P2) applied to the k-median problem.

4.1. Algorithm for robust location. Recall that the input consists of a metric (V, d), client sets $\{S_i\}_{i=1}^m$, and objective function Φ satisfying properties (P1) and (P2). We assume (by scaling) that distances in the metric are at least one. We also assume that there is a polynomial-time computable upper bound U such that (i) $\Phi(F \mid S) \leq U$ for every $F \subseteq V$ (with $|F| = k$) and $S \subseteq V$, and (ii) $\log U$ is polynomial in the input size. This is a mild assumption, and (to the best of our knowledge) is satisfied by all previously studied location problems. For example, in the k-median problem $U = n \cdot \mathsf{diam}(V, d)$.

The general algorithm (described below) is a natural extension of the algorithm for the robust k-median problem. We assume that the algorithm knows a value $B \in [2\beta\,\mathsf{OPT}, 4\beta\,\mathsf{OPT}]$, where OPT denotes the optimal value of the robust k-location instance. As in the k-median case, this can be achieved by trying all values of B that are powers of two in the interval $[1, 4\beta U]$ and finally taking the minimum cost solution.

General algorithm for robust k-location

1. Initialize weights $w_i^1 \leftarrow 1$ for all $1 \le i \le m$ and the set of facilities $F^1 \leftarrow V$.
2. For $t = 1, \ldots, n-k$ do:
   (a) For each $v \in F^t$ and $i \in [m]$, let $\delta_i^t(v) := \delta(F^t, v \mid S_i) = \Phi(F^t \backslash \{v\} \mid S_i) - \Phi(F^t \mid S_i)$.
   (b) Set $\hat{F}^t \leftarrow \{v \in F^t \mid \delta_i^t(v) \le B/2 \ \forall\, 1 \le i \le m\}$.
   (c) Let $v^t = \arg\min\{\sum_{i=1}^m w_i^t \cdot \delta_i^t(v) \colon v \in \hat{F}^t\}$ be a vertex with the least weighted increase.
   (d) Drop this vertex v^t and set $F^{t+1} \leftarrow F^t \backslash \{v^t\}$.
   (e) Update weights by $w_i^{t+1} \leftarrow w_i^t \cdot (1 + 1/B)^{\delta_i^t(v^t)}$ for all $1 \le i \le m$.
3. Output F^{n-k+1} with k facilities.

THEOREM 4.1 (FRAMEWORK: ROBUST VERSION). *Given a robust location problem* Robust(Π), *where* Π *satisfies properties* (P1) *and* (P2), *there is an* $O(\beta \cdot \log(n+m))$*-approximation algorithm for* Robust(Π), *where* m *is the number of scenarios and* $n = |V|$.

The proof is almost identical to that for robust k-median, and is given here for completeness.

LEMMA 4.1. *Assuming that* $B \ge 2\beta \cdot \mathsf{OPT}$, *in any iteration* $1 \le t \le n-k$ *and in step* 2(*b*), *we have* $F^t \backslash Q^t \subseteq \hat{F}^t$, *and hence* $\hat{F}^t$ *is nonempty.*

PROOF. By the projection property (P2), there is a set Q^t (of size at most k) such that $\sum_{v \in F^t \backslash Q^t} \delta_i^t(v) \le \beta \cdot \mathsf{OPT} \le B/2$ for all scenarios i. Because the δs are nonnegative, each cost increase $\delta_i^t(v)$ is at most $B/2$ for all $v \in F^t \backslash Q^t$ and $i \in [m]$. This implies that $\hat{F}^t \supseteq F^t \backslash Q^t$. Since Q^t has size at most k, and $|F^t| > k$, the set $\hat{F}^t$ must be nonempty. □

LEMMA 4.2. *Assuming that* $B \ge 2\beta \cdot \mathsf{OPT}$, *in any iteration* $1 \le t \le n-k$, *we have*

$$\min_{v \in \hat{F}^t} \left\{ \sum_{i=1}^m w_i^t \cdot \delta_i^t(v) \right\} \le \frac{\beta \cdot \mathsf{OPT}}{n-k-t+1} \sum_{i=1}^m w_i^t.$$

PROOF. Let us fix Q^t as promised by the projection property (P2), and sum the $\delta_i^t(v)$ values both over vertices $v \in F^t \backslash Q^t$ and over scenarios $i \in [m]$:

$$\sum_{v \in F^t \backslash Q^t} \sum_{i=1}^m w_i^t \cdot \delta_i^t(v) = \sum_{i=1}^m w_i^t \cdot \sum_{v \in F^t \backslash Q^t} \delta_i^t(v) = \sum_{i=1}^m w_i^t \cdot \sum_{v \in F^t \backslash Q^t} \delta(F^t, v \mid S_i) \le \sum_{i=1}^m w_i^t \cdot \beta \cdot \mathsf{OPT}.$$

The last inequality follows from Property (P2). Since $F^t \backslash Q^t \subseteq \hat{F}^t$ by Lemma 4.1, we have:

$$\begin{aligned} \min_{v \in \hat{F}^t} \left\{ \sum_{i=1}^m w_i^t \cdot \delta_i^t(v) \right\} &\le \min_{v \in F^t \backslash Q^t} \left\{ \sum_{i=1}^m w_i^t \cdot \delta_i^t(v) \right\} \\ &\le \frac{1}{|F^t \backslash Q^t|} \sum_{v \in F^t \backslash Q^t} \sum_{i=1}^m w_i^t \cdot \delta_i^t(v) \\ &\le \frac{\beta \cdot \mathsf{OPT}}{n-k-t+1} \sum_{i=1}^m w_i^t. \end{aligned}$$

The last inequality uses $|F^t \backslash Q^t| \ge n-t-k+1$ since $|F^t| = n-t+1$ and $|Q^t| \le k$. □

LEMMA 4.3. *Assume that* $B \ge 2\beta \cdot \mathsf{OPT}$. *Let* $W^t = \sum_{i=1}^m w_i^t$ *denote the total weight of all scenarios at the start of iteration* $1 \le t \le n-k$. *Then the total weight at the start of the next iteration,*

$$W^{t+1} \le W^t \cdot e^{1/(n-k-t+1)}.$$

PROOF. For any iteration t and scenario i, the weight update step ensures that

$$w_i^{t+1} = w_i^t \left(1 + \frac{1}{B}\right)^{\delta_i^t(v^t)} \le w_i^t \cdot \exp(\delta_i^t(v^t)/B),$$

where $v^t \in \hat{F}^t$ is the facility that is dropped in iteration t. From the definition of the set $\hat{F}^t$, we have $0 \le \delta_i^t(v^t)/B \le 1/2$. Moreover, for $y \in [0, 1/2]$, we have $e^y \le 1 + \sqrt{e} \cdot y$. This implies

$$w_i^{t+1} \le w_i^t \cdot \left(1 + \sqrt{e} \frac{\delta_i^t(v^t)}{B}\right),$$

and hence

$$W^{t+1} \le \sum_{i=1}^{m} w_i^t \cdot \left(1 + \sqrt{e}\frac{\delta_i^t(v^t)}{B}\right) = W^t + \frac{\sqrt{e}}{B} \cdot \sum_{i=1}^{m} w_i^t \cdot \delta_i^t(v^t).$$

Using Lemma 4.2 and the facts that $B \ge 2\beta \cdot \mathsf{OPT}$ and $\sqrt{e} < 2$, it follows that

$$W^{t+1} \le W^t + \frac{\sqrt{e}}{B} \cdot \frac{\beta \cdot \mathsf{OPT}}{n-k-t+1} \cdot W^t \le \left(1 + \frac{1}{n-k-t+1}\right) W^t.$$

Finally, the inequality $1 + x \le e^x$ implies the lemma. □

Proof of Theorem 4.1. Let $\mathsf{Alg} = \max_{i=1}^{m} \sum_{t=1}^{n-k} \delta_i^t(v^t)$ denote the value of the solution F^{n-k+1} at the end of the algorithm, and let i_0 be the value of i achieving the maximum in the above expression. Hence the total weight

$$W^{n-k+1} \ge w_{i_0}^{n-k+1} = \left(1 + \frac{1}{B}\right)^{\mathsf{Alg}}.$$

A repeated application of Lemma 4.3 implies that

$$W^{n-k+1} \le W^1 \cdot e^{H_{n-k}} = m \cdot e^{H_{n-k}}.$$

Taking logarithms and approximating the harmonic number by a logarithm, we get

$$\mathsf{Alg} \le O(\log m + \log n) / \log\left(1 + \frac{1}{B}\right).$$

Using the fact that $\log_2(1+y) \ge y$ for $y \in [0, 1]$, we get

$$\mathsf{Alg} \le O(\log m + \log n) \cdot \beta \cdot \mathsf{OPT}. \quad \square$$

4.2. Algorithm for stochastic location. We can extend our framework to stochastic problems as well: given a location problem Π as in the previous section and scenarios $\{S_i\}_{i=1}^m$ that now come with probabilities $\{p_i\}_{i=1}^m$ with $\sum_{i=1}^m p_i = 1$, the stochastic problem $\mathsf{Stoc}(\Pi)$ seeks to find a set F of size k that minimizes $\sum_{i=1}^m p_i \Phi(F \mid S_i)$. We denote the optimal set by F^*, each scenario's cost by $\mathsf{OPT}_i = \Phi(F^* \mid S_i)$, and $\mathsf{StocOpt} = \sum_{i=1}^m p_i \mathsf{OPT}_i$. The algorithm we present for stochastic location problems is similar to that for the robust version, but is even simpler because it does not use the weight updates.

General algorithm for stochastic k-location

1. Initialize the set of facilities $F^1 \leftarrow V$.
2. For $t = 1, \ldots, n-k$ do:
   (a) For each $v \in F^t$ and $i \in [m]$, let $\delta_i^t(v) := \delta(F^t, v \mid S_i) = \Phi(F^t \backslash \{v\} \mid S_i) - \Phi(F^t \mid S_i)$.
   (b) Let $v^t = \arg\min\{\sum_{i=1}^m p_i \delta_i^t(v) \colon v \in F^t\}$ be a vertex with the least expected increase.
   (c) Drop this vertex v^t and set $F^{t+1} \leftarrow F^t \backslash \{v^t\}$.
3. Output F^{n-k+1} with k facilities.

Theorem 4.2 (Framework: Stochastic Version). *Given a stochastic location problem* $\mathsf{Stoc}(\Pi)$, *where* Π *satisfies properties* (P1) *and* (P2), *there is an* $O(\beta \cdot \log n)$*-approximation algorithm for* $\mathsf{Stoc}(\Pi)$. *Here* $n = |V|$ *is the number of vertices.*

Lemma 4.4. *In any iteration* $1 \le t \le n-k$, *we have*

$$\min_{v \in F^t} \left\{ \sum_{i=1}^{m} p_i \delta_i^t(v) \right\} \le \frac{\beta \cdot \mathsf{StocOpt}}{n-k-t+1}.$$

Proof. Let us fix Q^t as promised by the projection property (P2). Then

$$\sum_{v \in F^t \backslash Q^t} \sum_{i=1}^{m} p_i \cdot \delta_i^t(v) = \sum_{i=1}^{m} p_i \sum_{v \in F^t \backslash Q^t} \delta_i^t(v) \le \sum_{i=1}^{m} p_i \cdot \beta \cdot \mathsf{OPT}_i = \beta\, \mathsf{StocOpt}.$$

The second inequality follows from the projection property (P2). Observe that $|F^t| = n - t + 1 \geq k + 1$ and $|Q^t| \leq k$; hence $|F^t \setminus Q^t| \geq n - k - t + 1 \geq 1$. Now,

$$\begin{aligned}\min_{v \in F^t} \left\{ \sum_{i=1}^{m} p_i \cdot \delta_i^t(v) \right\} &\leq \min_{v \in F^t \setminus Q^t} \left\{ \sum_{i=1}^{m} p_i \cdot \delta_i^t(v) \right\} \\ &\leq \frac{1}{|F^t \setminus Q^t|} \sum_{v \in F^t \setminus Q^t} \sum_{i=1}^{m} p_i \cdot \delta_i^t(v) \\ &\leq \frac{\beta \cdot \mathsf{StocOpt}}{n - k - t + 1},\end{aligned}$$

which completes the proof of the lemma. □

Proof of Theorem 4.2. The cost of the solution found by the algorithm is eventually:

$$\sum_{i=1}^{m} p_i \cdot \sum_{t=1}^{n-k} \delta_i^t(v^t) = \sum_{t=1}^{n-k} \sum_{i=1}^{m} p_i \cdot \delta_i^t(v^t) \leq \sum_{t=1}^{n-k} \frac{\beta \cdot \mathsf{StocOpt}}{n - k - t + 1},$$

using Lemma 4.4. But this is just $O(\beta \cdot \log(n-k)) \cdot \mathsf{StocOpt}$, proving the theorem. □

Remark 1. Our algorithm for stochastic location problems is also *incremental* in the sense of Mettu and Plaxton [28] and Lin et al. [26]: given metric (V, d) and scenarios $\{S_i, p_i\}_{i=1}^m$, the output is a *single permutation* of the vertices such that for every $1 \leq k \leq |V|$, the solution consisting of the first k vertices in this permutation is an approximate solution to the stochastic k-location instance.

Remark 2. Our framework for stochastic location problems also extends to the model where the demand distribution is not given explicitly, instead by means of a sampleable black-box. This model is well studied in the context of two-stage stochastic optimization problems, e.g., Gupta et al. [19], Shmoys and Swamy [36], and Charikar et al. [5]. Let $\mathscr{D}$ denote the demand distribution; i.e., the actual client-set $S \subseteq V$ is drawn according to $\mathscr{D}$. We now describe the modifications required in the above algorithm for stochastic location. In any iteration $1 \leq t \leq n - k$, define $\delta^t(v) := E_{S \leftarrow \mathscr{D}}[\delta(F^t, v \mid S)]$ for each $v \in F^t$. For each $v \in F^t$, let $\tilde{\delta}^t(v)$ denote an *estimate* of $\delta^t(v)$ obtained by taking the average of a large (polynomial) number of independent samples from $\mathscr{D}$. The algorithm for general demand distributions replaces steps 2(a) and 2(b) by the following:

(2a′) For each $v \in F^t$, compute $\tilde{\delta}^t(v)$ by sampling from $\mathscr{D}$.

(2b′) Let $v^t \leftarrow \arg\min\{\tilde{\delta}^t(v) \mid v \in F^t\}$.

Using Chernoff bounds (Motwani and Raghavan [30]), it can be shown that with high probability (w.h.p.), all the estimates $\tilde{\delta}^t(v)$ obtained in the algorithm are within a factor of two of the respective true values $\delta^t(v)$. Then the same analysis as above implies that w.h.p. the solution F^{n-k+1} has objective value $O(\beta \cdot \log(n-k)) \cdot \mathsf{StocOpt}$.

5. The k-tree and k-person TSP problems. In the *k-tree* problem, we are given a metric space (V, d) and a set S of clients, and we want to open a set of k facilities $F \subseteq V$ and build a forest of minimum cost in the induced metric $(F \cup S, d)$ so that for each client $v \in S$, there is some facility $f \in F$ such that this forest contains a path from v to f (and we say that the forest connects v to f). In particular, we want to minimize $d(\kappa(F, S))$, where $\kappa(F, S)$ denotes the minimum-cost forest in the metric induced on the set $F \cup S$ that connects each vertex in S to some vertex in F. Thus the objective function is:

$$\Phi(F \mid S) = d(\kappa(F, S)) = \sum_{e \in \kappa(F,S)} d(e). \tag{7}$$

It is worth noting that once we choose the set F of facilities, $\kappa(F, S)$ (for a given client-set S) is a minimum spanning tree in the distance function obtained from $(F \cup S, d)$ by shrinking all the nodes in F to a single "root" vertex; hence the real effort is in choosing the set of facilities F. This also implies that property (P1) holds for the k-tree problem.

***k*-person TSP.** In this problem, given a metric space (V, d) and a set S of clients, the goal is to open a set of k facilities $F \subseteq V$; but now the goal is to build k tours, so that for each $i \in [k]$, the ith tour contains the ith facility in F, each client in S is visited by some tour, and the sum of the tour lengths is minimized. Given an instance $\mathscr{I}$ of the k-person TSP problem, the cost of the optimal k-tree for $\mathscr{I}$ is a lower bound on the cost of the optimal k-person TSP for $\mathscr{I}$; moreover, given a forest which is a k-tree solution to the instance $\mathscr{I}$, taking Euler tours for each of the trees in the forest gives a solution for k-person TSP with cost at most twice as much.

Hence, an α-approximation algorithm for the robust version of the k-tree problem gives a 2α-approximation algorithm for the robust version of the k-person TSP problem. In this section we focus on the k-tree problem.

To apply the general framework of §4 to k-tree, we show that this problem satisfies the two required conditions. Property (P1) has been established above. The subsequent lemma shows that the β-projection property (P2) is satisfied with $\beta = 4$.

LEMMA 5.1 (PROPERTY (P2) FOR k-TREE). *For every $F^* \subseteq V$ with $|F^*| = k$ and $F \subseteq V$ with $|F| > k$, there exists a subset $Q \subseteq F$ of size at most k such that for all $S \subseteq V$,*

$$\sum_{r \in F \setminus Q} [d(\kappa(F \setminus r, S)) - d(\kappa(F, S))] \leq 4 \cdot d(\kappa(F^*, S)).$$

PROOF. For each $f^* \in F^*$ choose a facility $\eta(f^*) \in F$ closest to f^*; i.e., $\eta(f^*) := \arg\min_{g \in F} d(f^*, g)$. Define $Q = \{\eta(f^*) \mid f^* \in F^*\}$; clearly $|Q| \leq |F^*| = k$. Fix an arbitrary client set $S \subseteq V$. Recall that $\delta(F, r \mid S) = d(\kappa(F \setminus r, S)) - d(\kappa(F, S))$ denotes the increase in the cost for S upon dropping facility $r \in F$. For each vertex $r \in F$, define:

$$C(r) = \{s \in S\colon (r, s) \text{ is an edge in } \kappa(F, S)\} \qquad \text{and} \qquad D(r) = C(r) \setminus F^*.$$

Because each tree in $\kappa(F, S)$ contains exactly one vertex from F, the sets $\{D(r)\}_{r \in F}$ are disjoint subsets of $S \setminus F^*$; define $D := \bigcup_{r \in F} D(r)$. We define a useful map $\sigma\colon D \to D \cup F^*$ as follows (see also Figure 2).

1. Obtain an Euler tour on each tree in forest $\kappa(F^*, S)$. This corresponds to $|F^*| = k$ vertex-disjoint tours $\tau_1, \ldots, \tau_k$ on $F^* \cup S$.
2. For each $j \in [k]$, orient tour τ_j clockwise and restrict the tour to vertices in $D \cup F^*$, by short-cutting over vertices $S \setminus (D \cup F^*)$.
3. For each vertex $v \in D$, let τ_j (some $j \in [k]$) be the tour that contains v; and set $\sigma(v) \leftarrow v'$ where $v' \in D \cup F^*$ is the *unique* successor vertex of v, given by the clockwise orientation in τ_j.

Note that this map is indeed well defined: since $D \subseteq S$, each vertex in D appears in some tour $\{\tau_j\}_{j=1}^k$ and hence has a unique successor as required in the last step above. Because each vertex of $D \cup F^*$ has in-degree one in the orientation of tours, it follows that map σ is one-to-one. Finally, the total length of the tours $\{\tau_j\}_{j=1}^k$ is at most $2 \cdot d(\kappa(F^*, s))$, which implies:

$$\sum_{v \in D} d(v, \sigma(v)) \leq 2 \cdot d(\kappa(F^*, S)). \tag{8}$$

Now fix any $r \in F \setminus Q$; we will upper bound $\delta(F, r \mid S)$. To show this, we modify $\mathcal{F}' = \kappa(F, S)$ to obtain a feasible forest $\mathcal{F}(r)$ such that each S-vertex is connected to some vertex in $F \setminus r$. We will show that the length of forest $\mathcal{F}(r)$ is not much more than $d(\mathcal{F}')$, which would bound $\delta(F, r \mid S)$. The forest $\mathcal{F}(r)$ is constructed

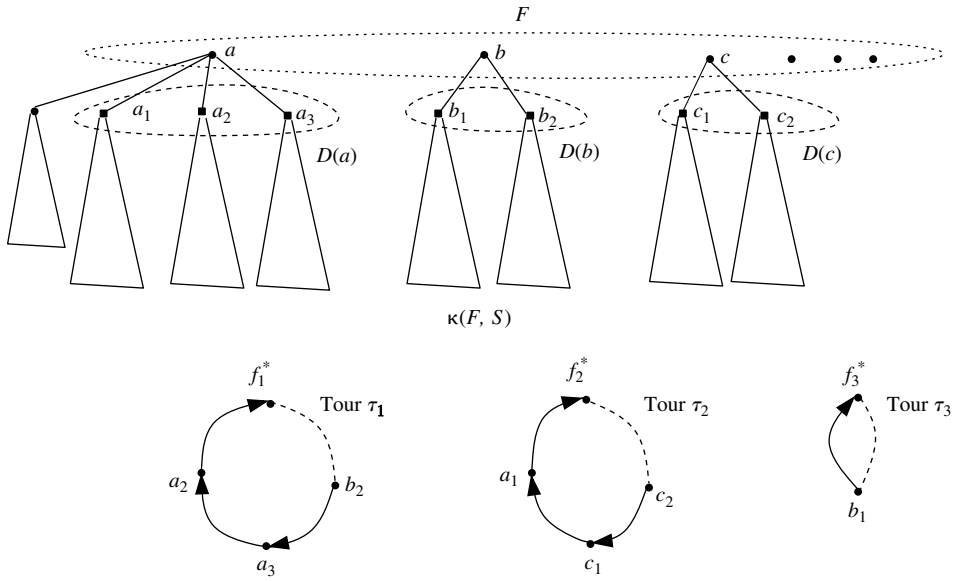

FIGURE 2. The solution $\kappa(F, S)$ and one possible mapping $\sigma\colon D \to D \cup F^*$.

as follows: starting with the forest $\mathcal{F}'$, delete the edges $\{(r, v) \mid v \in C(r)\}$ adjacent to r in this forest, and add the following edge-sets: (i) $\{(v, \sigma(v)) \mid v \in D(r)\}$ where σ is the map defined earlier; (ii) $\{(f^*, \eta(f^*)) \mid f^* \in \sigma(D(r)) \cap F^*\}$; and (iii) $\{(g, \eta(g)) \mid g \in C(r) \cap F^*\}$. Recall that for any $f \in F^*$, $\eta(f) \in Q$ is the closest facility to f in F.

Feasibility of $\mathcal{F}(r)$. We first show that $\mathcal{F}(r)$ is a feasible forest connecting each S-vertex to $F \setminus r$: it suffices to argue that $C(r)$ is connected to $F \setminus r$. Note that vertices $C(r) \cap F^*$ are directly connected to $Q \subseteq F \setminus r$ by edge-set (iii) above. We will now show that edge-sets (i) and (ii) suffice to connect $D(r) = C(r) \setminus F^*$ to $F \setminus r$ as well. Let $E(r) := \{(v, \sigma(v)) \mid v \in D(r)\}$; i.e., edge-set (i) above.

Claim 5.1. *The edges $E(r)$ connect each $D(r)$-vertex to some vertex in $(D \setminus D(r)) \cup (F^* \cap \sigma(D(r)))$.*

Proof of Claim 5.1. Fix any $v \in D(r)$, and let T_v denote the set of vertices that are connected to v using edges $E(r)$. Note that by definition of the map σ, we have $T_v \subseteq D \cup (F^* \cap \sigma(D(r)))$. Hence if $T_v \setminus D(r) \neq \varnothing$, it follows that v is connected to some vertex in $(D \setminus D(r)) \cup (F^* \cap \sigma(D(r)))$. Suppose (for a contradiction) that $T_v \setminus D(r) = \varnothing$; i.e., $T_v \subseteq D(r)$. Since $F^* \cap D(r) = \varnothing$, it must be (again by construction of σ) that there is some vertex $u \in T_v$ with $\sigma(u) \notin T_v$. But as $(u, \sigma(u)) \in E(r)$, this contradicts the definition of T_v. Thus it must be that T_v contains some vertex from $(D \setminus D(r)) \cup (F^* \cap \sigma(D(r)))$. □

Note that edge-set (ii) connects each vertex in $F^* \cap \sigma(D(r))$ to Q. Combined with Claim 5.1, each $D(r)$-vertex is connected to some vertex in $D \setminus D(r)$ or Q. Finally observe that each vertex in $D \setminus D(r)$ remains connected to $F \setminus r$ in forest $\mathcal{F}(r)$; and since $Q \subseteq F \setminus r$, we obtain that $\mathcal{F}(r)$ connects each $D(r)$-vertex to some vertex in $F \setminus r$.

Bounding cost of $\mathcal{F}(r)$. Next, we upper bound the increase in cost $\delta(F, r \mid S) \leq d(\mathcal{F}(r)) - d(\mathcal{F}')$ by:

$$\delta(F, r \mid S) \leq -\sum_{v \in C(r)} d(r, v) + \sum_{v \in D(r)} d(v, \sigma(v)) + \sum_{g \in C(r) \cap F^*} d(g, F) + \sum_{f^* \in \sigma(D(r)) \cap F^*} d(f^*, F). \tag{9}$$

The last term of this expression can be bounded by

$$\begin{aligned} \sum_{f^* \in \sigma(D(r)) \cap F^*} d(f^*, F) &\leq \sum_{v \in D(r)} d(\sigma(v), F) \\ &\leq \sum_{v \in D(r)} [d(\sigma(v), v) + d(v, F)], \end{aligned}$$

where the last inequality follows from triangle inequality. Plugging the final expression above into (9),

$$\delta(F, r \mid S) \leq 2 \cdot \sum_{v \in D(r)} d(v, \sigma(v)) + \sum_{v \in C(r)} [d(v, F) - d(v, r)] \leq 2 \cdot \sum_{v \in D(r)} d(v, \sigma(v)),$$

where the final inequality uses that $r \in F$.

Now summing over all $r \in F \setminus Q$, we get:

$$\sum_{r \in F \setminus Q} \delta(F, r \mid S) \leq 2 \cdot \sum_{r \in F \setminus Q} \sum_{v \in D(r)} d(v, \sigma(v)) \leq 2 \cdot \sum_{v \in D} d(v, \sigma(v)).$$

Finally, by (8), this last expression is at most $4 \cdot d(\kappa(F^*, S))$, which completes the proof. □

Using this lemma within our general framework, we obtain:

Corollary 5.2 (Robust/Stochastic k-Tree Result). *There is an $O(\log(m + n))$-approximation algorithm for the robust k-tree problem, and an $O(\log n)$-approximation algorithm for the stochastic k-tree problem.*

Note that we could also consider robust/stochastic versions of the *k-Steiner-tree* problem, where given clients S and facilities F, the goal is to construct a forest that is *not necessarily induced* on $F \cup S$, connecting each S-vertex to some F-vertex; i.e., the solution may use vertices outside $F \cup S$ as Steiner points. In the k-tree problem considered above, $\kappa(F, S)$ was required to be induced on $F \cup S$. However, these two objectives are within a factor two of each other, and we obtain the same approximation results for robust/stochastic k-Steiner-tree. As mentioned earlier, we also obtain identical guarantees for the robust/stochastic versions of the k-person TSP.

6. Capacitated k-median problem. In this problem, we are given a metric (V, d), a client-set $S \subset V$, a number k, and a capacity μ such that $|S| \le k \cdot \mu$; the goal is to open a set of k facilities $F \subseteq V$ and construct an assignment $\rho\colon S \to F$ of clients to open facilities such that at most μ clients are assigned to any open facility (i.e., $|\rho^{-1}(f)| \le \mu$ for all $f \in F$), and the objective $\sum_{v \in S} d(v, \rho(v))$ is minimized. A map $\rho\colon S \to F$ is said to be *feasible* iff $|\rho^{-1}(f)| \le \mu$ for all $f \in F$; additionally we define the *cost* of mapping ρ as $D(\rho) := \sum_{v \in S} d(v, \rho(v))$. Thus, given client-set $S \subseteq V$ and facility-set $F \subseteq V$, the objective in capacitated k-median is:

$$\Phi(F \mid S) = \min\{D(\rho) \mid \text{map } \rho\colon S \to F \text{ is feasible}\}.$$

Note that $\Phi(F \mid S)$ and the map $\rho\colon S \to F$ achieving the minimum can be found in polynomial time by solving a minimum cost b-matching problem (Cook et al. [10]). In this section we show that the capacitated k-median problem satisfies the conditions for our general framework, and hence we obtain logarithmic approximations for its robust and stochastic versions.

To the best of our knowledge, our algorithm for the robust version gives the first nontrivial approximation guarantee for even the deterministic version of the problem, with *hard* capacity constraints. Chuzhoy and Rabani [9] obtain a constant factor approximation for the deterministic version with nonuniform *soft* capacities where the algorithm violates capacities by a constant factor.

To apply our general framework for robust and stochastic location problems, we establish the two properties (P1) and (P2). Property (P1) holds trivially: as noted above, given facilities $F \subset V$ and clients $S \subseteq V$, $\Phi(F \mid S_i)$ can be computed in polynomial time via b-matching (Cook et al. [10]).

We will prove the β-projection property (P2) with $\beta = 2$. Recall that we are given any set $F^* \subseteq V$ of k facilities, and another set $F \subseteq V$ of more than k facilities. Define $\sigma\colon F^* \to F$ to be a minimum cost matching between F^* and F that assigns each vertex of F^* to a distinct vertex in F. We set $Q := \sigma(F^*)$ to be those facilities in F that are matched to some facility in F^*. Note that $|Q| = |F^*| = k$, as required. In the rest of this section we show that for any $S \subseteq V$,

$$\sum_{r \in F \setminus Q} \delta(F, r \mid S) \le 2 \cdot \Phi(F^* \mid S). \tag{10}$$

This would establish property (P2) with $\beta = 2$. Let $\rho^*\colon S \to F^*$ (resp. $\rho\colon S \to F$) denote the minimum-cost feasible mapping from S to F^* (resp. S to F). To establish (10), we construct for each $r \in F \setminus Q$, a feasible mapping $\rho^{(r)}\colon S \to F \setminus \{r\}$ such that:

$$\sum_{r \in F \setminus Q} (D(\rho^{(r)}) - D(\rho)) \le 2 \cdot \Phi(F^* \mid S); \tag{11}$$

this suffices since $\delta(F, r \mid S) \le D(\rho^{(r)}) - D(\rho)$ for any $r \in F \setminus Q$. In the next subsection we describe how these new mappings are constructed, and in the following subsection we bound the cost increases.

6.1. Constructing new assignments. A useful assignment is $\rho' := \sigma \circ \rho^*$, which maps S to Q. This is a candidate choice for $\rho^{(r)}$ for every $r \in F \setminus Q$; however this may result in a large increase in cost. Another natural choice for $\rho^{(r)}$ is to map (i) all $v \in S$ with $\rho(v) = r$ to $\rho'(v) \in Q$, and (ii) all other $u \in S$ to $\rho(u)$. However this might violate capacity at some facilities. Hence, defining the new mappings requires several clients to be reassigned, as described below.

It will be convenient to view any map $\theta\colon S \to F$ as a bipartite graph on disjoint vertex-sets S and F with edge-set $E(\theta) := \{(v, \theta(v)) \mid v \in S\}$. Note that for any feasible map θ, in the resulting bipartite graph, vertices in S have degree one, and those in F have degree at most μ. Recall that both ρ and $\rho' = \sigma \circ \rho$ map S to F. Define a bipartite multigraph H with disjoint vertex-sets S and F, and edge-set $E_H := E(\rho) \sqcup E(\rho')$ (i.e., if an edge appears in both $E(\rho)$ and $E(\rho')$, then graph H contains two distinct copies of it). Note that in graph H, each S-vertex has degree exactly two, vertices in $F \setminus Q$ have degree at most μ, and vertices in Q have degree at most 2μ.

A path $P \subseteq E_H$ is called an *alternating path* if it starts at a vertex in $F \setminus Q$, ends at a vertex in Q, and uses edges alternately from $E(\rho)$ and $E(\rho')$. For every vertex $v \in \rho^{-1}(F \setminus Q)$, we will show the existence of an alternating path P_v starting at vertex $\rho(v) \in F \setminus S$ and with edge $(\rho(v), v) \in E(\rho)$. Define $\mathcal{P} := \{P_v \mid v \in \rho^{-1}(F \setminus Q)\}$ and $E(\mathcal{P}) := \bigcup P_v$. We will also ensure the following two conditions for this collection $\mathcal{P}$ of paths.

Condition 1. The paths in $\mathcal{P}$ are edge disjoint; i.e., $P_v \cap P_u = \varnothing$ for all distinct $u, v \in \rho^{-1}(F \setminus Q)$.

Condition 2. The bipartite graph on vertex-sets S and F with edges $E(\mathcal{P}) \oplus E(\rho)$ has each S-vertex of degree one and each F-vertex of degree at most μ.

To establish the existence of this collection $\mathcal{P}$, we first prove the existence of a circulation in a suitably defined network flow problem. Then we show how this circulation gives rise to the desired collection $\mathcal{P}$.

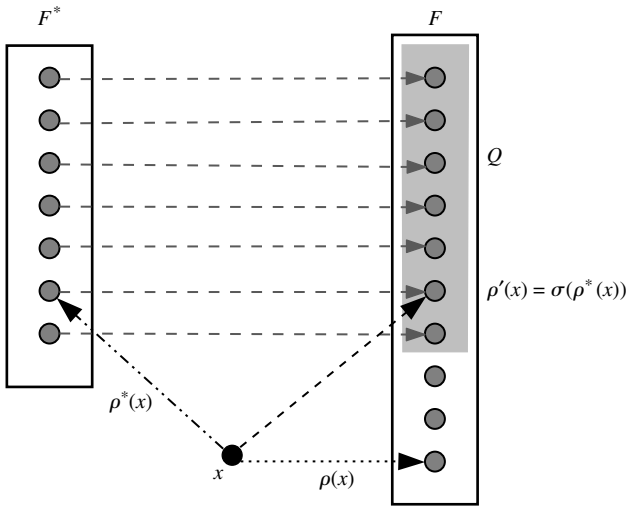

FIGURE 3. Figure showing maps $\rho\colon S \to F$, optimal map $\rho^*\colon S \to F^*$ applied to point $x \in S$, and also the matching $\sigma\colon F^* \to F$.

An auxiliary flow problem. Consider a directed multigraph T on vertex set $V(T) := F \cup \{s\}$ where s is a new vertex. The arcs in T are given by the *multiset* $A(T) := \{(\rho(v), \rho'(v)) \mid v \in S\} \cup \{(s, f) \mid f \in F \setminus Q\} \cup \{(q, s) \mid q \in Q\}$. For any vertex $u \in F$, denote by $\delta^+(u)$ (resp. $\delta^-(u)$) the number of arcs in the multiset $\{(\rho(v), \rho'(v)) \mid v \in S\}$ leaving (resp. entering) vertex u. By the properties of mappings ρ and ρ', it follows that $\delta^+(u), \delta^-(u) \le \mu$ for all $u \in F$; furthermore, $\delta^-(u) = 0$ for all $u \in F \setminus Q$. Consider an instance of the *circulation problem* on T given by integer lower-bounds $\{\kappa_a \mid a \in A(T)\}$ and upper-bounds $\{\eta_a \mid a \in A(T)\}$ on arcs. An assignment of integer values $x\colon A(T) \to \mathbb{Z}_+$ to the arcs is called a circulation iff:

$$\sum_{(u,v)\in A(T)} x(u,v) - \sum_{(v,u)\in A(T)} x(v,u) = 0, \quad \forall\, u \in V(T) \qquad \text{and} \qquad \kappa_a \le x(a) \le \eta_a, \quad \forall\, a \in A(T).$$

Hoffman's circulation theorem (Hoffman [21]) states that there exists a circulation if and only if:

$$\sum_{(u,v)\in A(T),\, u\in X,\, v\notin X} \eta_{u,v} \ge \sum_{(v,u)\in A(T),\, u\in X,\, v\notin X} \kappa_{v,u}, \quad \text{for all } X \subseteq V(T). \tag{12}$$

Set the upper and lower bounds on arcs of $A(T)$ as follows:

$$\eta_{u,v} := \begin{cases} \mu - \delta^+(u) & \text{if } v = s \text{ and } u \in Q, \\ 1 & \text{if } u, v \in F, \\ \infty & \text{if } u = s \text{ and } v \in F \setminus Q, \end{cases} \qquad \kappa_{u,v} := \begin{cases} 0 & \text{if } v = s \text{ and } u \in Q, \\ 0 & \text{if } u, v \in F, \\ \delta^+(v) & \text{if } u = s \text{ and } v \in F \setminus Q. \end{cases}$$

We claim that this circulation instance satisfies (12). Consider any $X \subseteq V(T)$. We denote the left-hand side in (12) by $\eta(X)$, and the right-hand side by $\kappa(X)$. If $s \in X$, then it follows that $\kappa(X) = 0$ and $\eta(X) \ge \kappa(X)$ trivially. Now suppose $s \notin X$, and let $X_1 = X \cap (F \setminus Q)$ and $X_2 = X \cap Q$; so $X = X_1 \cup X_2$. It is clear that $\kappa(X) = \sum_{u\in X_1} \delta^+(u)$. For any $Y \subseteq F$, define $\Delta^+(Y)$ to be the number of arcs $(u, v) \in \{(\rho(w), \rho'(w))\}_{w\in S}$ with $u \in Y$ and $v \notin Y$. Observe that $\eta(X) = \mu \cdot |X_2| - \sum_{u\in X_2} \delta^+(u) + \Delta^+(X)$. To establish $\eta(X) \ge \kappa(X)$, it suffices to show that $\Delta^+(X) + \mu \cdot |X_2| \ge \sum_{u\in X} \delta^+(u)$.

Consider the arcs $A_X \subseteq \{(\rho(w), \rho'(w))\}_{w\in S}$ having their tail[1] in X: there are $|A_X| = \sum_{u\in X} \delta^+(u)$ such arcs. Since $\delta^-(u) = 0$ for all $u \in F \setminus Q$, we can partition A_X into $A'_X \subseteq A_X$ having a head in $X_2 = X \cap Q$, and $A''_X \subseteq A_X$ having a head in $Q \setminus X$. Since $\delta^-(u) \le \mu$ for all $u \in Q$, we have $|A'_X| \le \sum_{u\in X_2} \delta^-(u) \le \mu \cdot |X_2|$. Observe that each arc in A''_X has a tail in X and a head in $Q \setminus X$: thus $|A''_X| \le \Delta^+(X)$. Finally,

$$\sum_{u\in X} \delta^+(u) = |A_X| = |A'_X| + |A''_X| \le \mu \cdot |X_2| + \Delta^+(X),$$

which gives $\eta(X) \ge \kappa(X)$ that, in turn, implies (12).

[1] We use the standard terminology for directed graphs, the *tail* of an arc (u, v) is u, and its *head* is v.

Constructing alternating paths $\mathcal{P}$. We now show how this integral circulation x in T can be used to construct the alternating paths $\mathcal{P}$. Note that by the definition of the circulation instance, x restricted to arcs $\{(\rho(w), \rho'(w))\}_{w\in S}$ can be decomposed into an arc-disjoint collection of paths $\{\hat{P}_v \mid v \in \rho^{-1}(F\backslash Q)\}$ where each $\hat{P}_v$ originates at $\rho(v)$ using arc $(\rho(v), \rho'(v))$ and ends at some Q-vertex. Moreover, the total number of paths ending at any $q \in Q$ is at most $\mu - \delta^+(q)$ because of the upper-bounds η_as. Note that there is a one-to-one correspondence between arcs $\{(\rho(w), \rho'(w))\}_{w\in S}$ and set S. Using this, for each $v \in \rho^{-1}(F\backslash S)$, directed path $\hat{P}_v$ corresponds to path P_v in the (undirected) bipartite graph H, defined as follows: for every arc $(\rho(w), \rho'(w))$ (where $w \in S$) in $\hat{P}_v$, path P_v contains two edges $(\rho(w), w)$ and $(w, \rho'(w))$. Thus it follows that each P_v is an *alternating path* in H, that starts at vertex $\rho(v)$ using edge $(\rho(v), v)$ and ends at some Q-vertex.

Proving Condition 1. Since the collection $\{\hat{P}_v \mid v \in \rho^{-1}(F\backslash Q)\}$ is arc-disjoint in T, paths $\mathcal{P} = \{P_v \mid v \in \rho^{-1}(F\backslash S)\}$ are edge-disjoint in H.

Proving Condition 2. Let $\{\hat{P}_v \mid v \in \rho^{-1}(F\backslash Q)\}$ consist of arcs $\{(\rho(w), \rho'(w)) \mid w \in W\}$, where $W \subseteq S$. Then $E(\mathcal{P}) = \{(\rho(w), w)\}_{w\in W} \cup \{(\rho'(w), w)\}_{w\in W}$. So $E'_H := E(\rho) \oplus E(\mathcal{P}) = \{(\rho(y), y)\}_{y\in S\setminus W} \cup \{(\rho'(w), w)\}_{w\in W}$. Clearly each S-vertex has degree exactly one in E'_H. Vertices in $F\backslash Q$ have zero degree in E'_H: since $W \supseteq \rho^{-1}(F\backslash Q)$ by construction and ρ' maps S to Q. We now upper bound the degree of any vertex $q \in Q$ in E'_H. Recall that $\delta^+(q)$ is the degree of vertex q in $E(\rho)$. Let $e(q)$ denote the number of directed paths in $\{\hat{P}_v \mid v \in \rho^{-1}(F\backslash Q)\}$ that end at vertex q; by construction of the circulation instance, $e(q) \le \mu - \delta^+(q)$. Additionally the number of paths in $\mathcal{P}$ ending at q also equals $e(q)$. Since $\mathcal{P}$ consists of edge-disjoint alternating paths in graph H, the degree of vertex q in $E'_H = E(\rho) \oplus E(\mathcal{P})$ equals $\delta^+(q) + e(q) \le \mu$. This completes the proof of Condition 2.

Defining mappings $\rho^{(r)}$. For each $r \in F\backslash Q$, we define $\rho^{(r)}: S \to F\backslash\{r\}$ as follows.

1. Consider bipartite graph G_r on disjoint vertex-sets S and F, and edge set:

$$E_r := E(\rho) \oplus \Big(\bigcup_{v\in\rho^{-1}(r)} P_v \Big).$$

2. For each $v \in S$, set $\rho^{(r)}(v) \leftarrow u$, where $(u, v) \in E_r$ is the unique such edge.

LEMMA 6.1. *For each $r \in F\backslash Q$, the map $\rho^{(r)}$ is well defined, feasible, and $\rho^{(r)}(S) \subseteq F\backslash\{r\}$.*

PROOF. Fix any $r \in F\backslash Q$ for this proof. Note that each S-vertex has degree one in $E(\rho)$, and degree zero or two in every alternating path of $\mathcal{P}$. So the degree of each S-vertex in E_r is odd. However, $E_r \subseteq E_H$ and each S-vertex has degree two in E_H; thus the degree of each S-vertex in E_r is exactly one. This implies that $\rho^{(r)}$ is indeed well defined.

To show that $\rho^{(r)}$ is feasible, we will prove that each F-vertex has degree at most μ in E_r. Let $W := \rho^{-1}(r) \subseteq S$ and $X := \rho^{-1}(F\backslash Q)$. Number the vertices in X from 1 to $|X|$ such that $W = \{1, 2, \ldots, |W|\}$. Consider the following iterative way of modifying $E(\rho)$. Starting with $J_0 \leftarrow E(\rho)$, define for each $i \in \{1, \ldots, |X|\}$, $J_i \leftarrow J_{i-1} \oplus P_i$. Since the paths $\mathcal{P} = \{P_i\}_{i=1}^{|X|}$ are edge-disjoint, it follows that $J_{|X|} = E(\rho) \oplus E(\mathcal{P})$ and $J_{|W|} = E_r$. For each $1 \le i \le |X|$, the following hold:

1. For any vertex in $F\backslash Q$, its degree in J_i is at most its degree in J_{i-1}. Note that P_i contains exactly one edge incident to a vertex in $F\backslash Q$, namely edge $(i, \rho(i))$. Thus the degree of vertices $F\backslash Q\backslash\{\rho(i)\}$ is unchanged going from J_{i-1} to J_i. Additionally, $(i, \rho(i)) \in J_{i-1}$ since it is in $E(\rho)$ and in *none* of $P_1, \ldots, P_{i-1}$. Hence the degree of vertex $\rho(i)$ in J_i is one less than in J_{i-1}.
2. For any $u \in Q$, its degree in J_i is at least its degree in J_{i-1}. Clearly the degree of Q-vertices not in P_i remain the same in J_i and J_{i-1}. For any vertex $u \in Q$ visited in P_i, there is an edge $(u, \rho'^{-1}(u))$ that lies in path P_i but not J_{i-1}; so the degree of u in J_i is at least that in J_{i-1}.

From the above, we obtain that (i) the degree in $J_{|W|}$ of any $(F\backslash Q)$-vertex is at most its degree in $J_0 = E(\rho)$; i.e., μ; and (ii) the degree in $J_{|W|}$ of any Q-vertex is at most its degree in $J_{|X|} = E(\rho) \oplus E(\mathcal{P})$, which is at most μ by Condition 2. Recall that $J_{|W|} = E_r$, so we obtain that E_r is a feasible map.

We now show that vertex r has degree zero in E_r, which would imply that $\rho^{(r)}(S) \subseteq F\backslash\{r\}$. By definition of the alternating paths $\mathcal{P}$, we have $\{(v, \rho(v)) \mid v \in W\} \subseteq \bigcup_{v\in W} P_v$. Clearly $\{(v, \rho(v)) \mid v \in W\} \subseteq E(\rho)$. Thus $\{(v, \rho(v)) \mid v \in W\} \cap E_r = \varnothing$. Finally, since $r \in F\backslash Q$, the only edges in H that are incident to r are $\{(v, \rho(v)) \mid v \in W\}$: recall that ρ' maps S to $Q \subseteq F$. Hence it follows that r has degree zero in E_r. □

6.2. Bounding the cost increase. Given the alternating paths $\mathcal{P}$ and the mappings in Lemma 6.1, we now complete the proof of (11) by bounding the cost increases; i.e., $D(\rho^{(r)}) - D(\rho)$ for $r \in F\backslash Q$.

Increase of single alternating path. Consider any $v \in \rho^{-1}(F\backslash Q)$ and let $S \cap P_v = \{u_1, u_2, \ldots, u_l\}$ (in that order) be the S-vertices in alternating path P_v; note that $u_1 = v$. By definition of an alternating path, we have $\rho'(u_j) = \rho(u_{j+1})$ for all $1 \le j \le l-1$. Define

$$\varepsilon(P_v) := \sum_{j=1}^{l} (d(u_j, \rho'(u_j)) - d(u_j, \rho(u_j))).$$

For any $1 \le j \le l$, using triangle inequality we have

$$\begin{aligned} d(u_j, \rho'(u_j)) - d(u_j, \rho(u_j)) &\le d(u_j, \rho^*(u_j)) + d(\rho^*(u_j), \rho'(u_j)) - d(u_j, \rho(u_j)) \\ &= 2 \cdot d(u_j, \rho^*(u_j)) + d(\rho^*(u_j), \rho'(u_j)) - (d(u_j, \rho(u_j)) + d(u_j, \rho^*(u_j))) \\ &\le 2 \cdot d(u_j, \rho^*(u_j)) + d(\rho^*(u_j), \rho'(u_j)) - d(\rho^*(u_j), \rho(u_j)). \end{aligned}$$

Using the above, we can bound

$$\varepsilon(P_v) \le 2 \sum_{j=1}^{l} d(u_j, \rho^*(u_j)) + \sum_{j=1}^{l} [d(\rho^*(u_j), \rho'(u_j)) - d(\rho^*(u_j), \rho(u_j))]. \tag{13}$$

Define a weighted bipartite graph M on disjoint vertex-sets F^* and F, with each edge (f, w) (for any $f \in F^*$ and $w \in F$) having cost $d(f, w)$. A subset of edges E' in M is said to be an F^*-matching if E' is a matching that contains some edge incident to each vertex in F^*. Recall that $\sigma\colon F^* \to F$ is the minimum cost F^*-matching in M. Let $E(\sigma) := \{(f, \sigma(f)) \mid f \in F^*\}$ denote the edges in this matching.

Consider the path in graph M defined by edges $\tilde{P}_v := \{(\rho(u_j), \rho^*(u_j))\}_{j=1}^{l} \cup \{(\rho^*(u_j), \rho'(u_j))\}_{j=1}^{l}$. Note that this indeed describes a path since $\rho'(u_j) = \rho(u_{j+1})$ for all $1 \le j \le l-1$. We claim that $E(\sigma) \oplus \tilde{P}_v$ is also an F^*-matching in M. This is because for each $j = 1, \ldots, l$, edge $(\rho^*(u_j), \rho'(u_j)) \in E(\sigma)$ and edge $(\rho(u_j), \rho^*(u_j)) \notin E(\sigma)$, and since $\rho(u_1) \in F\backslash Q$, it has zero degree in $E(\sigma)$. Since $E(\sigma)$ is the minimum cost F^*-matching, we have

$$\sum_{e \in E(\sigma)} d(e) - \sum_{e' \in E(\sigma) \oplus \tilde{P}_v} d(e') = \sum_{j=1}^{l} [d(\rho^*(u_j), \rho'(u_j)) - d(\rho^*(u_j), \rho(u_j))] \le 0.$$

Plugging this into (13), we get

$$\varepsilon(P_v) \le 2 \sum_{j=1}^{l} d(u_j, \rho^*(u_j)) = 2 \sum_{u \in S \cap P_v} d(u, \rho^*(u)). \tag{14}$$

Bounding cost of $\rho^{(r)}$. Fix any $r \in F\backslash Q$, and let $W(r) := \rho^{-1}(r)$. Now,

$$\begin{aligned} D(\rho^{(r)}) - D(\rho) &= \sum_{e \in E(\rho) \oplus (\cup_{v \in W(r)} P_v)} d(e) - \sum_{e' \in E(\rho)} d(e') \\ &= \sum_{v \in W(r)} \varepsilon(P_v) \\ &\le 2 \sum_{v \in W(r)} \sum_{u \in S \cap P_v} d(u, \rho^*(u)), \end{aligned} \tag{15}$$

where the first equality is by definition of $\rho^{(r)}$ (Lemma 6.1), the second equality uses the fact that $\{P_v \mid v \in W(r)\}$ are edge-disjoint, and the last inequality is by (14).

By Condition 1 the alternating paths in $\mathcal{P}$ are edge-disjoint in graph H. Furthermore, each S-vertex has degree two in H, and degree zero or two in each path of $\mathcal{P}$. Thus each S-vertex appears in at most one alternating path from $\mathcal{P}$. Thus we have

$$\sum_{r \in F\backslash Q} (D(\rho^{(r)}) - D(\rho)) \le 2 \sum_{r \in F\backslash Q} \sum_{v \in W(r)} \sum_{u \in S \cap P_v} d(u, \rho^*(u)) \le 2 \sum_{u' \in S} d(u', \rho^*(u')) \;=\; 2 \cdot \Phi(F^* \mid S).$$

Above, the first inequality is by (15), and the second inequality uses (i) $\{W(r) \mid r \in F\backslash Q\}$ are disjoint subsets of $\rho^{-1}(F\backslash Q)$, and (ii) $\{S \cap P_v \mid v \in \rho^{-1}(F\backslash Q)\}$ are disjoint (as argued above). This completes the proof of property (P2) for capacitated k-median with $\beta = 2$.

Corollary 6.1 (Robust/Stochastic Capacitated k-Median Result). *There is an $O(\log m + \log n)$-approximation algorithm for robust capacitated k-median and an $O(\log n)$-approximation algorithm for stochastic capacitated k-median.*

7. Fault-tolerant k-median. In this problem, we are given a client set $S \subseteq V$, and a *requirement* $r_v \in \{1, 2, \ldots, k\}$ for each client $v \in S$. The goal is to open a set of k facilities $F \subseteq V$ and connect each client v to r_v distinct facilities in F such that the total connection cost is minimized. Given the facility-set $F \subseteq V$, each client $v \in S$ should clearly be connected to be the r_v facilities in F that are closest to v, to minimize the objective. For any $F \subseteq V$, $v \in S$, and integer $0 \le h \le |F|$, let $\Gamma(v, F, h)$ denote the set of h *distinct* facilities in F that are closest to v; and let $\gamma(v, F, h)$ denote the cost of connecting v to the h distinct facilities in $\Gamma(v, F, h)$; i.e.,

$$\gamma(v, F, h) := \sum_{f \in \Gamma(v, F, h)} d(v, f), \quad \forall F \subseteq V, \quad v \in S, \quad h \in \{0, 1, \ldots, |F|\}.$$

Thus the objective in fault-tolerant k-median is:

$$\Phi(F \mid S) := \sum_{v \in S} \gamma(v, F, r_v).$$

To use our framework to solve the robust and stochastic fault-tolerant k-median problems, the next lemma proves the β-projection property (P2) with $\beta = 2$. Again property (P1) is trivial because it is easy to calculate the exact cost of any solution.

For the problem where all the requirements r_v are uniform, the best known approximation guarantee for the single scenario version is four (Swamy and Shmoys [38]). To the best of our knowledge, our result is the first nontrivial algorithm for nonuniform requirements, even for a single scenario.

LEMMA 7.1 ((P2) FOR FAULT-TOLERANT k-MEDIAN). *For every $F^* \subseteq V$ ($|F^*| = k$) and $F \subseteq V$ with $|F| > k$, there exists $Q \subseteq F$ with $|Q| = k$ such that*

$$\sum_{f \in F \setminus Q} \sum_{v \in S} [\gamma(v, F \setminus f, r_v) - \gamma(v, F, r_v)] \le 2 \cdot \Phi(F^* \mid S) \quad \forall S \subseteq V. \tag{16}$$

PROOF. Define the subset Q as a one-to-one mapping $\pi\colon F^* \to F$ as follows. Arbitrarily order the elements of F^* and initialize $Q \leftarrow \varnothing$. For each $g \in F^*$ in this order, set $\pi(g) := \arg\min\{d(g, f) \mid f \in F \setminus Q\}$ and $Q \leftarrow Q \cup \{\pi(g)\}$. Clearly $|Q| = |F^*| = k$ and mapping π is one-to-one.

Fix any client-set $S \subseteq V$. For each $v \in S$, we define a mapping $\tau_v\colon \Gamma(v, F, r_v) \to Q$ thus:

1. Define $G_v := \Gamma(v, F, r_v) \cap Q$. Set $\tau_v(g) \leftarrow g$ for all $g \in G_v$.
2. For each $g \in \Gamma(v, F, r_v) \setminus Q$, set $\tau_v(g)$ to be a distinct vertex from $\pi(\Gamma(v, F^*, r_v)) \setminus G_v$.

Note that the second step is well defined since $|\Gamma(v, F, r_v) \setminus Q| = r_v - |G_v| = |\pi(\Gamma(v, F^*, r_v))| - |G_v| \le |\pi(\Gamma(v, F^*, r_v)) \setminus G_v|$. Additionally τ_v indeed maps $\Gamma(v, F, r_v)$ to Q since $\pi(\Gamma(v, F^*, r_v)) \subseteq Q$. Finally observe that τ_v is also one-to-one. Next we bound the total cost of this map τ_v.

CLAIM 7.1. *For any $v \in S$,*

$$\sum_{g \in \Gamma(v, F, r_v)} d(v, \tau_v(g)) \le \gamma(v, F, r_v) + 2 \cdot \gamma(v, F^*, r_v).$$

PROOF. Note that by the definition of τ_v in step 1 above, $\sum_{g \in G_v} d(v, \tau_v(g)) = \sum_{g \in G_v} d(v, g)$ (recall that $G_v = \Gamma(v, F, r_v) \cap Q$). For each $g \in \Gamma(v, F, r_v) \setminus G_v$, define $f_g := \pi^{-1}(\tau_v(g))$; note that this is well defined since $\tau_v(g) \in Q$ and furthermore, $f_g \in \Gamma(v, F^*, r_v)$ by the definition of τ_v in step 2 above. Observe that the f_g's for each $g \in \Gamma(v, F, r_v) \setminus G_v$ are distinct.

Fix any $g \in \Gamma(v, F, r_v) \setminus G_v$. Now recall the greedy construction of Q, and consider the time when $f_g \in F^*$ was mapped under π to its nearest vertex in F. At that point, $g \in \Gamma(v, F, r_v) \setminus G_v$ was not in Q (because it is not in Q at the end of that process). Hence, the distance $d(f_g, \pi(f_g)) \le d(f_g, g) \le d(g, v) + d(v, f_g)$. Thus,

$$d(v, \tau_v(g)) = d(v, \pi(f_g)) \le d(v, f_g) + d(f_g, \pi(f_g)) \le d(g, v) + 2 \cdot d(v, f_g).$$

Summing this expression over all $g \in \Gamma(v, F, r_v) \setminus G_v$ and using the fact that $\{f_g \mid g \in \Gamma(v, F, r_v) \setminus G_v\}$ are distinct vertices in $\Gamma(v, F^*, r_v)$, we get:

$$\sum_{g \in \Gamma(v, F, r_v) \setminus G_v} d(v, \tau_v(g)) \le 2 \cdot \sum_{f \in \Gamma(v, F^*, r_v)} d(v, f) + \sum_{g \in \Gamma(v, F, r_v) \setminus G_v} d(v, g)$$

$$\Rightarrow \sum_{g \in \Gamma(v, F, r_v)} d(v, \tau_v(g)) \le 2 \cdot \gamma(v, F^*, r_v) + \sum_{g \in \Gamma(v, F, r_v)} d(v, g).$$

Rewriting the right-hand side as $2 \cdot \gamma(v, F^*, r_v) + \gamma(v, F, r_v)$ completes the proof of Claim 7.1. □

To complete the proof of Lemma 7.1, we show that for any $f \in F \setminus Q$, we can obtain a feasible assignment of each vertex $v \in S$ to r_v facilities $C_v(f) \subseteq F \setminus \{f\}$ as follows. If $f \notin \Gamma(v, F, r_v)$, then $C_v(f) := \Gamma(v, F, r_v)$. Otherwise, if $f \in \Gamma(v, F, r_v)$ (note also that $f \in F \setminus Q$), then $C_v := (\Gamma(v, F, r_v) \setminus \{f\}) \cup \{\tau_v(f)\}$. By the definition of map τ_v, we have $\tau_v(f) \in Q \setminus \Gamma(v, F, r_v)$; thus in either case, $C_v(f) \subseteq F \setminus \{f\}$ and $|C_v(f)| = r_v$. Observe that the increase in v's connection cost upon dropping f from F is $\gamma(v, F \setminus \{f\}, r_v) - \gamma(v, F, r_v) \le d(v, \tau_v(f)) - d(v, f)$. Now we have,

$$\begin{aligned}
\sum_{f \in F \setminus Q} \sum_{v \in S} (\gamma(v, F \setminus \{f\}, r_v) - \gamma(v, F, r_v)) &\le \sum_{f \in F \setminus Q} \sum_{v:\, f \in \Gamma(v, F, r_v)} (d(v, \tau_v(f)) - d(v, f)) \\
&= \sum_{v \in S} \sum_{f \in \Gamma(v, F, r_v) \setminus Q} (d(v, \tau_v(f)) - d(v, f)) \\
&= \sum_{v \in S} \sum_{f \in \Gamma(v, F, r_v)} (d(v, \tau_v(f)) - d(v, f)) \\
&= \sum_{v \in S} \left(\left(\sum_{f \in \Gamma(v, F, r_v)} d(v, \tau_v(f)) \right) - \gamma(v, F, r_v) \right) \\
&\le \sum_{v \in S} ((2 \cdot \gamma(v, F^*, r_v) + \gamma(v, F, r_v)) - \gamma(v, F, r_v)) \\
&= 2 \sum_{v \in S} \gamma(v, F^*, r_v) = 2 \cdot \Phi(F^* \mid S),
\end{aligned}$$

where the third to last inequality follows from Claim 7.1. This finishes the proof of the lemma. □

Corollary 7.2 (Robust/Stochastic Fault-Tolerant k-Median Result). *There is an $O(\log m + \log n)$-approximation algorithm for robust fault-tolerant k-median, and an $O(\log n)$-approximation algorithm for stochastic fault-tolerant k-median.*

8. The stochastic k-center problem. In the previous sections, we gave approximation algorithms for some robust and stochastic location problems. In this section, we study another natural stochastic location problem, *stochastic k-center*, and provide some evidence that it is hard to approximate well in polynomial time. We consider the uniform-probability stochastic k-center problem: given a metric space (V, d), subsets $S_1, \ldots, S_m \subseteq V$ and a bound k, the goal is to open a set F of k facilities to minimize

$$\sum_{i=1}^{m} \max_{x \in S_i} d(x, F).$$

Note that the deterministic version of this problem (i.e., $m = 1$) is the k-center problem, for which several 2-approximations are known, and this is the best one can do unless $P = NP$ (Vazirani [40]).

In this section we show that the stochastic k-center problem is closely related to the dense k-subgraph problem. Recall that in the standard (maximization) version of the dense k-subgraph problem, we are given a graph G with n vertices and a value k, and the goal is to pick k vertices that maximize the number of edges in the induced subgraph. One can also define the minimization version of dense k-subgraph, where the goal is now to pick k edges to minimize the number of vertices incident to these edges. The best result known for either version is that of Feige et al. [16] who gave an $O(n^{\delta})$-approximation algorithm for some $\delta < 1/3$. The problem is believed to be hard, and Feige [14] and Khot [24] showed that the dense k-subgraph problem is hard to approximate within some constant $\rho > 1$ under two different complexity-theoretic assumptions.

We study the (uniform-probability) stochastic k-center problem on the uniform metric, and hence can formulate it as a set-covering-type problem:

> Given m sets $\{S_i\}_{i=1}^m$ that are subsets of a ground set V, the goal is to pick a set $F \subseteq V$ of k elements to minimize the number of sets not contained within F; i.e., the objective is to minimize $|\{i \in [m] \mid S_i \not\subseteq F\}|$.

Theorem 8.1 (Stochastic k-Center Hardness). *Suppose there exists an α-approximation algorithm for the stochastic k-center problem on the uniform metric. Then there is an α-approximation algorithm for the minimization version of dense k-subgraph.*

Proof. Consider an instance of the minimization dense k-subgraph problem: given graph $G = (V(G), E(G))$ and parameter k_G, we want to pick at least k_G edges to minimize the number of vertices incident to these

edges. We construct an instance of stochastic k-center on the ground set $V := E(G)$. For each vertex $v \in V(G)$, we define a set $S_v := \{e \in E(G) \mid e \text{ is incident to } v\}$. Now consider the instance of stochastic k-center with V, $\{S_v \mid v \in V(G)\}$, and the parameter $k = |E(G)| - k_G$. Given any solution $F \subseteq V$ for this problem, consider the set $F' = V \setminus F$ of size k_G. Choosing F to minimize the number of sets S_v that are not contained within F is the same as choosing F' to minimize the number of sets that intersect F'—but because a set S_v intersects F' precisely when some edge in F' is incident to $v \in V(G)$, this is precisely the same as solving the minimization dense k-subgraph instance. In particular, if the bound k is not violated, any algorithm for stochastic k-center on uniform metrics with approximation ratio α gives an identical approximation ratio for the minimization dense k-subgraph problem. □

9. Closing remarks. In this paper we presented the first approximation algorithms for some natural classes of min-max (robust) and stochastic location problems. Our results propose a general framework for obtaining approximation guarantees of $O(\log m + \log n)$ for such problems, where m is the number of possible "scenarios" and n is the size of the metric space. For some of these problems, one can improve this to $O(\log m + \log k)$ by first preprocessing the instance to define a new weighted instance on a metric space of size $O(k)$, and then extend the current algorithms to work for weighted instances as well. As mentioned in §1, the algorithms in this paper only work in the case where there are no costs involved with opening facilities at particular locations. Can we give algorithms with similar performance guarantees for the situation with facility costs?

Appendix A. Improved guarantee for robust k-median on uniform metrics. We now consider a natural linear relaxation for the robust k-median problem on a uniform metric. Recall that there are n elements V, and m scenarios $S_1, \ldots, S_m \subseteq V$; the goal is to pick k elements so as to minimize the maximum number of uncovered elements in any scenario.

$$\begin{aligned} \min \quad & z \\ \text{s.t.} \quad & z \geq \sum_{e \in S_i} x_e \quad \forall\, 1 \leq i \leq m, \\ & \sum_{e \in V} x_e = n - k, \\ & 0 \leq x \leq 1, \\ & z \geq 0. \end{aligned}$$

In the above linear program, the variable x_e is 1 if element e is *not* picked, and 0 otherwise. Let us fix any solution (x, z) to this linear program. To round this solution, we use the dependent rounding scheme of Gandhi et al. [17], which implies the following in our context:

THEOREM A.1 (GANDHI ET AL. [17]). *There is a polynomial time randomized algorithm that generates $X_e \in \{0, 1\}$ for all $e \in V$ such that*:

1. $\Pr[X_e = 1] = x_e$ *for all* $e \in V$.
2. $\Pr[\sum_{e \in V} X_e = n - k] = 1$.
3. $\{X_e \mid e \in V\}$ *are negatively correlated. This implies that for any* $S \subseteq V$ *and* $\delta \geq 0$, *we have*:

$$\Pr\left[\sum_{e \in S} X_e > (1+\delta)\mu_S\right] \leq \min\left\{e^{(-\delta^2\mu_S)/(2+\delta)}, \left(\frac{e}{\delta+1}\right)^{\mu_S(1+\delta)}\right\}.$$

Here $\mu_S = \mathrm{E}[\sum_{e \in S} X_e]$.

Using this rounding scheme, it is clear that we always pick exactly k elements. For any scenario S_i, we have $\mu_i = \mathrm{E}[\sum_{e \in S_i} X_e] = \sum_{e \in S_i} x_e \leq z$. Fix any constant $\epsilon \in (0, 1)$, and set $\alpha := (8/\epsilon) \ln m$. Using the first expression in property 3 of Theorem 0 with $\delta_i = (\epsilon \cdot \mu_i + \alpha)/\mu_i$ for each S_i, we have for each $1 \leq i \leq m$:

$$\Pr\left[\sum_{e \in S_i} X_e > (1+\epsilon) \cdot \mu_i + \alpha\right] \leq \exp\left(-\frac{(\epsilon\mu_i + \alpha)^2}{2\mu_i + \epsilon\mu_i + \alpha}\right) \leq \frac{1}{m^2},$$

where the last inequality uses the following calculation:

$$\frac{(\epsilon\mu_i + \alpha)^2}{2\mu_i + \epsilon\mu_i + \alpha} \geq \alpha \cdot \left(\frac{\epsilon\mu_i + \alpha}{2\mu_i + \epsilon\mu_i + \alpha}\right) = \alpha \cdot \left(1 + \frac{2\mu_i}{\epsilon\mu_i + \alpha}\right)^{-1} \geq \alpha \cdot \left(1 + \frac{2}{\epsilon}\right)^{-1} \geq \frac{\epsilon}{4} \cdot \alpha = 2 \ln m.$$

Using $\mu_i \le z$ for all scenarios S_i, we get $\Pr[\sum_{e\in S_i} X_e > (1+\epsilon)\cdot z + \alpha] \le 1/m^2$ for each $1 \le i \le m$. Now, by a union bound over all scenarios we obtain that with probability at least $1 - 1/m$, the maximum number of uncovered elements in any scenario is at most $(1+\epsilon)\cdot z + (8/\epsilon)\cdot \ln m$. Thus we have:

THEOREM 0. *For any constant $0 < \epsilon < 1$, there is a (randomized) approximation algorithm for robust k-median on uniform metrics that, given any instance, returns a solution of value at most $(1+\epsilon)\cdot l^* + (8/\epsilon)\cdot \ln m$, where l^* is the optimal value of the given instance.*

This randomized rounding algorithm can also be shown to achieve a better multiplicative approximation guarantee. Set $\beta = 4\ln m/(\ln\ln m)$. For each scenario S_i, choose δ_i so that $\mu_i(1+\delta_i) = \beta\lceil z\rceil$; recall that z is the LP objective and $\mu_i = \mathrm{E}[\sum_{e\in S_i} X_e] = \sum_{e\in S_i} x_e \le z$. We assume that $z > 0$; otherwise the robust k-median instance is trivial. Since $\mu_i \le z$, we have $1 + \delta_i \ge \beta$ for all $i \in [m]$. Now using the second expression in property 3 of Theorem 0, for any $i \in [m]$,

$$\Pr\left[\sum_{e\in S_i} X_e > \beta\lceil z\rceil\right] \le \left(\frac{e}{1+\delta_i}\right)^{\beta\lceil z\rceil} \le (e/\beta)^{\beta} \le \exp\left(-\beta\cdot\frac{1}{2}\ln\ln m\right) \le \frac{1}{m^2}.$$

Now, again by a union bound, with probability at least $1 - 1/m$, the maximum number of uncovered elements in any scenario is at most $\beta\lceil z\rceil$, which implies:

THEOREM 0. *There is a randomized $O(\ln m/(\ln\ln m))$-approximation algorithm for the robust k-median problem on uniform metrics.*

Appendix B. Bad examples from §3. Here we give bad examples for the two greedy algorithms for robust k-median on uniform metrics, that were mentioned in §3.

Consider first the greedy algorithm that drops the element increasing the exposure of fewest sets. This algorithm performs badly on the following instance. The universe $V = \{a_1, a_2, \ldots, a_{3t}\} \cup \{b_1, b_2, \ldots, b_t\}$, and the sets/scenarios are as follows:

- There is a set $S_0 = \{a_1, a_2, \ldots, a_{3t}\}$.
- For each $j \in [t]$, there are three sets $S_{j1} = \{b_j, a_j\}$, $S_{j2} = \{b_j, a_{j+t}\}$, and $S_{j2} = \{b_j, a_{j+2t}\}$.

Note that each element b_j lies in three sets, whereas each element a_j lies in two sets. The total number of elements is $n = 4t$, and suppose we want to choose a set $F \subseteq V$ with $k = 3t$ to minimize the maximum exposure. An optimal solution is to choose $F = S_0$, which results in a maximum exposure of one. However, if we keep greedily dropping elements which increase the exposure of the fewest sets, we will drop some t of the elements in S_0, which will give us an exposure of $t = |V|/4$.

Next, consider the greedy algorithm that drops any element that keeps the maximum exposure minimized. The bad instance consists of universe $V = \{c_1, \ldots, c_t\} \cup \{d_1, \ldots, d_t\}$, and let $C := \{c_1, \ldots, c_t\}$. For each $j \in [t]$, there is a set $C \cup \{d_j\}$. The bound $k = t$. Clearly the optimal solution picks elements C, resulting in a maximum exposure of one. However, one possible run of this greedy algorithm is to repeatedly drop each element in C; this results in a solution having maximum exposure t.

Acknowledgements. The authors thank the referees for their incisive comments that helped improve the presentation of this paper. Part of this research was done while the first author was at the Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213. The second and the fourth authors were supported by NSF Grant CCF-0430751, and part of the work was done while both were at Tepper School of Business at Carnegie Mellon University in Pittsburgh, PA. The third author was partly supported by NSF grants CCF-0448095 and CCF-0729022, as well as an Alfred P. Sloan Fellowship.

References

[1] Arya, V., N. Garg, R. Khandekar, A. Meyerson, K. Munagala, V. Pandit. 2004. Local search heuristics for k-median and facility location problems. *SIAM J. Comput.* **33**(3) 544–562.
[2] Birge, J. R., F. Louveaux. 1997. *Introduction to Stochastic Programming. Springer Series in Operations Research*. Springer-Verlag, New York.
[3] Blum, A. 1998. On-line algorithms in machine learning. A. Fiat, G. Woeginger, eds. *Online Algorithms: The State of the Art*. Springer-Verlag, London, 306–325.
[4] Cesa-Bianchi, N., G. Lugosi. 2006. *Prediction, Learning, and Games*. Cambridge University Press, New York.
[5] Charikar, M., C. Chekuri, M. Pál. 2005. Sampling bounds for stochastic optimization. *Approximation, Randomization and Combinatorial Optimization*, Vol. 3624, *Lecture Notes in Computer Science*. Springer, Berlin, 257–269.
[6] Charikar, M., S. Guha. 2005. Improved combinatorial algorithms for facility location problems. *SIAM J. Comput.* **34**(4) 803–824.

[7] Charikar, M., S. Guha, É. Tardos, D. B. Shmoys. 2002. A constant-factor approximation algorithm for the k-median problem. *J. Comput. System Sci.* **65**(1) 129–149.
[8] Chrobak, M., C. Kenyon, N. Young. 2006. The reverse greedy algorithm for the metric k-median problem. *Inform. Process. Lett.* **97**(2) 68–72.
[9] Chuzhoy, J., Y. Rabani. 2005. Approximating k-median with non-uniform capacities. *Proc. Sixteenth Annual ACM-SIAM Sympos. Discrete Algorithms*. ACM, New York, 952–958.
[10] Cook, W. J., W. H. Cunningham, W. R. Pulleyblank, A. Schrijver. 1998. *Combinatorial Optimization*. John Wiley and Sons, New York.
[11] Cooper, L. 1978. Bounds on the Weber problem solution under conditions of uncertainty. *J. Regional Sci.* **18**(1) 87–93.
[12] Daskin, M. S., S. H. Owen. 1999. Location models in transportation. *Handbook of Transportation Science*. Kluwer Academic, Norwell, MA, 311–360.
[13] Dhamdhere, K., V. Goyal, R. Ravi, M. Singh. 2005. How to pay, come what may: Approximation algorithms for demand-robust covering problems. *Proc. 46th Sympos. Foundations Comput. Sci.* (*FOCS*), IEEE Computer Society, Washington, DC, 367–378.
[14] Feige, U. 2002. Relations between average case complexity and approximation complexity. *Proc. Thirty-Fourth Annual ACM Sympos. Theory Comput.* ACM, New York, 534–543.
[15] Feige, U., K. Jain, M. Mahdian, V. Mirrokni. 2007. Robust combinatorial optimization with exponential scenarios. *Proc. 12th Internat. Conf. Integer Programming Combin. Optim.*, Springer-Verlag, Berlin, 439–453.
[16] Feige, U., G. Kortsarz, D. Peleg. 2001. The dense k-subgraph problem. *Algorithmica* **29**(3) 410–421.
[17] Gandhi, R., S. Khuller, S. Parthasarathy, A. Srinivasan. 2006. Dependent rounding and its applications to approximation algorithms. *J. ACM* **53**(3) 324–360.
[18] Golovin, D., V. Goyal, R. Ravi. 2006. Pay today for a rainy day: Improved approximation algorithms for demand-robust min-cut and shortest path problems. *Proc. 23rd Sympos. Theoret. Aspects Comput. Sci.*, Springer-Verlag, Berlin, 206–217.
[19] Gupta, A., M. Pal, R. Ravi, A. Sinha. 2004. Boosted sampling: Approximation algorithms for stochastic optimization. *Proc. 36th Annual ACM Sympos. Theory Comput.*, ACM, New York, 417–426.
[20] Hochbaum, D., D. B. Shmoys. 1985. A best possible heuristic for the k-center problem. *Math. Oper. Res.* **10**(2) 180–184.
[21] Hoffman, A. J. 1960. Some recent applications of the theory of linear inequalities to extremal combinatorial analysis. *Sympos. Appl. Math.* **10** 113–127.
[22] Immorlica, N., D. Karger, M. Minkoff, V. S. Mirrokni. 2004. On the costs and benefits of procrastination: Approximation algorithms for stochastic combinatorial optimization problems. *Proc. Fifteenth Annual ACM-SIAM Sympos. Discrete Algorithms*, SIAM, Philadelphia, 691–700.
[23] Jain, K., V. V. Vazirani. 2001. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation. *J. ACM* **48**(2) 274–296.
[24] Khot, S. 2006. Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J. Comput.* **36**(4) 1025–1071.
[25] Krause, A., B. McMahan, C. Guestrin, A. Gupta. 2007. Selecting observations under multiple objectives. *Adv. Neural Inform. Processing Systems* **20**.
[26] Lin, G., C. Nagarajan, R. Rajaraman, D. P. Williamson. 2006. A general approach for incremental approximation and hierarchical clustering. *Proc. 17th Annual ACM-SIAM Sympos. Discrete Algorithms*, ACM, New York, 1147–1156.
[27] F. Louveaux. 1993. Stochastic location analysis. *Location Sci.* **1** 127–154.
[28] Mettu, R. R., C. G. Plaxton. 2003. The online median problem. *SIAM J. Comput.* **32**(3) 816–832.
[29] Mirchandani, P. B., A. R. Odoni. 1979. Locations of medians on stochastic networks. *Transportation Sci.* **13** 85–97.
[30] Motwani, R., P. Raghavan. 1995. *Randomized Algorithms*. Cambridge University Press, New York.
[31] Plotkin, S. A., D. B. Shmoys, E. Tardos. 1995. Fast approximation algorithms for fractional packing and covering problems. *Math. Oper. Res.* **20**(2) 257–301.
[32] Ravi, R., A. Sinha. 2006. Hedging uncertainty: Approximation algorithms for stochastic optimization problems. *Math. Programming* **108**(1) 97–114.
[33] Rosenblatt, M. J., H. L. Lee. 1987. A robustness approach to facilities design. *Internat. J. Production Res.* **25** 479–486.
[34] Sahni, S., T. Gonzalez. 1976. P-complete approximation problems. *J. ACM* **23** 555–565.
[35] Sheppard, E. S. 1974. A conceptual framework for dynamic location allocation analysis. *Environment Planning A* **6**(5) 547–564.
[36] Shmoys, D. B., C. Swamy. 2006. An approximation scheme for stochastic linear programming and its application to stochastic integer programs. *J. ACM* **53**(6) 978–1012.
[37] Snyder, L. V. 2006. Facility location under uncertainty: A review. *IIE Trans.* **38**(7) 547–564.
[38] Swamy, C., D. B. Shmoys. 2003. Fault-tolerant facility location. *Proc. 14th Annual ACM-SIAM Sympos. Discrete Algorithms*, Society for Industrial and Applied Mathematics, Philadelphia, 735–736.
[39] Van Hentenryck, P., R. Bent, E. Upfal. Online stochastic optimization under time constraints. *Ann. Oper. Res.* 1–33. http://springerlink.com/content/e53ux62618717244/.
[40] Vazirani, V. V. 2001. *Approximation Algorithms*. Springer-Verlag, Berlin.
[41] Weaver, J. R., R. L. Church. 1983. Computational procedure for location problems on stochastic networks. *Transportation Sci.* **17**(2) 168–180.
[42] Welzl, E. 1996. Suchen und konstruieren durch verdoppeln. I. Wegener, ed. *Highlights der Informatik*. Springer, Berlin, 221–228.