1-2008

A Plant Location Guide for the Unsure

Barbara M. Anthony
Carnegie Mellon University

Vineet Goyal
Carnegie Mellon University

Anupam Gupta
Carnegie Mellon University, anupamg@cs.cmu.edu

Viswanath Nagarajan
Carnegie Mellon University

Published In
Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms , 1164- 1173.

A Plant Location Guide for the Unsure

Barbara M. Anthony* Vineet Goyal† Anupam Gupta‡ Viswanath Nagarajan†

Abstract

This paper studies an extension of the k-median problem where we are given a metric space (V, d) and not just one but m client sets $\{S_i \subseteq V\}_{i=1}^m$, and the goal is to open k facilities F to minimize:

$$\max_{i\in[m]} \left\{ \sum_{j\in S_i} d(j, F) \right\},$$

i.e., the worst-case cost over all the client sets. This is a "min-max" or "robust" version of the k-median problem; however, note that in contrast to previous papers on robust/stochastic problems, we have only one stage of decision-making—*where should we place the facilities?* We present an $O(\log n + \log m)$ approximation for robust k-median: The algorithm is combinatorial and very simple, and is based on reweighting/Lagrangean-relaxation ideas. In fact, we give a general framework for (minimization) facility location problems where there is a bound on the number of open facilities. For *robust* and *stochastic* versions of such location problems, we show that if the problem satisfies a certain "projection" property, essentially the same algorithm gives a logarithmic approximation ratio in both versions. We use our framework to give the first approximation algorithms for robust/stochastic versions of k-tree, capacitated k-median, and fault-tolerant k-median.

1 Introduction

Consider the following class of *k-facility location* problems: given a metric space (V, d), and a subset of clients $S \subseteq V$ who want service, we want to locate k facilities $F \subseteq V$ to minimize some objective function $\Phi(F \mid S)$. For instance, $\Phi(F \mid S) = \sum_{x\in S} d(x, F)$ gives us the *k-median* objective function, $\Phi(F \mid S) = \max_{x\in S} d(x, F)$ is *k-center*, $\Phi(F \mid S) =$ *total distance traveled by salesmen, one at each $f \in F$, to visit all clients in S* is *k-person TSP*, etc. Many such problems are known to be NP-hard, and have been extensively studied in both the computer science and operations research communities.

In this paper, we consider the stochastic and robust versions of these problems. These are cases where we are given not just one, but several sets $S_1, S_2, \ldots, S_m$ of clients. Again, we want to locate k facilities, and the goal is to be *simultaneously good for all these client sets*—more precisely, we want to minimize

$$\text{Robust-}\Phi = \max_i \Phi(F \mid S_i)$$

in the robust or max-min version, and

$$\text{Stochastic-}\Phi = \sum p_i \cdot \Phi(F \mid S_i)$$

in the stochastic version (for given probability values p_i for each scenario).

Robust and stochastic versions of problems naturally model cases with uncertain or dynamic systems. For instance, we might want to locate our facilities knowing that one of several scenarios is 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 [2, 20, 28, 18, 32].

1.1 Our Results and Techniques We consider the k-median problem as our running example, since it helps to illustrate the basic ideas we use. Our first result in Section 2 is obtained using the classical reweighting/Lagrangean relaxation techniques (see, e.g., [37, 3]) in conjunction with a natural reverse-greedy algorithm [7].

Theorem 1.1. (Robust k-Median Result) *There is a combinatorial $O(\log m + \log k)$-approximation algorithm for the robust k-median problem.*

In passing, let us mention that a natural approach to solve the problem by embedding the metric space into random trees does not seem to give us an advantage here: we do not currently know how to solve the problem even on the uniform metric to better than a logarithmic guarantee.

*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213. Email: banthony@andrew.cmu.edu

†Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213, USA. Supported in part by NSF ITR grant CCR-0122581 (The ALADDIN project) and NSF grant CCF-0430751. Email: {vineet, viswa}@cmu.edu

‡Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213. Research was partly supported by the NSF CAREER award CCF-0448095 and an Alfred P. Sloan Fellowship.

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: (The formal definition is given in (3.7).)

> *Given any instance of a problem Φ with client set S, and an infeasible solution F with $K > k$ facilities, there are $K-k$ facilities $F' \subseteq F$ such that shutting down any facility in F' increases the cost by at most $\beta\mathsf{Opt}$, and shutting down a random facility in F' (chosen uniformly) causes the cost to rise by at most $\frac{\beta\,\mathsf{Opt}}{K-k}$ in expectation.*

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 $F^* = F \setminus F'$. Clearly, the cost increase in shutting down any facility in F' is at most Opt. Further, the average cost of shutting down a facility in F' will be at most $\frac{\mathsf{Opt}}{|F'|} = \frac{\mathsf{Opt}}{K-k}$. 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 essentially says that even if $F^* \not\subseteq F$, we can "project" the F^* onto some k nodes 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.

THEOREM 1.2. (GENERAL FRAMEWORK) *Given any k-facility location problem Φ that satisfies the above β-projection property, and where we can efficiently compute the increase in objective function value on dropping a single facility, there is*

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

THEOREM 1.3. (APPLICATIONS) *The following problems have $\beta = O(1)$: (a) hard-capacitated k-median with uniform capacities, (b) fault-tolerant k-median with non-uniform requirements, and (c) k-tree. Hence, all these problems admit $O(\log m + \log n)$-approximation guarantees for their robust version, and $O(\log n)$-approximations for their stochastic versions.*

We note in passing that the results for hard-capacitated k-median and non-uniform fault-tolerant k-median seem to be the first logarithmic approximation guarantees known for *even the one-scenario versions* of these problems. Moreover, our algorithms give *incremental solutions* for these problems in the sense of [26, 24]: the output is a permutation of the vertices such that for every i, the solution consisting of the first i vertices in this permutation is a good robust/stochastic solution for the i-location problem on that instance; we elaborate on this in the full version of the paper.

Finally, we show that not all k-facility location problems give good results using this framework, since they do not satisfy the projection property:

THEOREM 1.4. (STOCHASTIC k-CENTER) *The Stochastic k-center problem is as hard to approximate as the (minimization) Dense-k-subgraph problem.*

Roadmap. In Section 2 we present results for the Robust k-median problem: first for the uniform metric, which gives many of the ideas, and then for general metrics. We then abstract out the general framework in Section 3. Due to lack of space, we can present only one of the β-projection proofs in this version (that for k-Tree in Section 4.1): the rest appear in the final version of the paper, as do many of the other proofs throughout the paper. Finally, we give the hardness of the Stochastic k-center problem in Section 5.

1.2 Other Related Work

Location problems under uncertainty have long been studied in the operations research literature due to their vast applicability in real world scenarios. Sheppard [31] used a scenario based approach to model uncertainty in demand and minimize the expected cost, while Cooper [9] was among the first to consider the robust objective on location problems. Following this, similar models for location problems such as k-median and UFL were studied [27, 36, 29]. See Louveaux [25] and Daskin and Owen [10] 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 [33]. The papers by van Hentenryck et al. [35] 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 [6, 5, 21, 1, 26, 7] as well as its capacititated [8] and fault-tolerant versions [34], as well as for the k-tree [12] and k-center problems [30, 19]. Out of these, the one most relevant to our work is the reverse-greedy algorithm of Chrobak et al [7] 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., [20, 28, 18, 32], and robust optimization (see, e.g., [11, 17, 14]), it is not clear how to use their techniques to solve the problems we consider here. As an aside, our paper gives results in the explicit scenarios model: it would be interesting to extend results to cases where exponentially many scenarios can be handed, e.g., like in [18, 32, 4, 14].

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. [23].

2 The Robust k-Median Problem

The *robust k-median* is the following: given an n-vertex metric space (V, d), m subsets $S_1, \cdots, S_m \subseteq V$ of vertices, and a number k, find a subset $T \subseteq V$ of size $|T| = k$ that minimizes the objective $\max_{i=1}^{m} \sum_{v \in S_i} d(v, T)$, where given a vertex $x \in V$ and $S \subseteq V$, $d(x, S) = \min_{y \in S} d(x, y)$. In this section, we will give a logarithmic approximation algorithm for the problem on general metric spaces.

2.1 A Warm Up: the Uniform Metric When (V, d) is the uniform metric, the analysis gives some intuition for subsequent algorithms. Note that the problem can be recast as the following:

> Given a ground set V, and a family of m sets $\{S_i \subseteq V\}_{i=1}^{m}$, find a set $T \subseteq V$ of size $|T| = k$ such that the *maximum "exposure"* $\max_i |S_i \setminus T|$ is minimized.

By a reduction from vertex cover, this is NP-hard to approximate to better than a factor of 2 (details are deferred to a full version of the paper). To solve this problem, we consider the following algorithm which starts with all the elements, and repeatedly drops the element that exposes the fewest sets. This would just minimize the exposure on average, so to get a result for the maximum exposure, we "penalize" the newly exposed sets by increasing their weights, so that exposing them further costs us more. Formally, the algorithm is given as Algorithm I.

Algorithm I: Uniform Metric Robust k-Median

1. Set $w_i^1 = 1$ $(1 \le i \le m)$ and open facilities $F^1 = V$.
2. For $t = 1, \cdots, n-k$ do:
   (a) For $v \in F^t$, $W^t(v)$ = weight of sets containing v.
   (b) Let v^t be the element that minimizes $W^t(v)$
   (c) Drop this element to get $F^{t+1} \leftarrow F^t \setminus v^t$.
   (d) Set: $w_i^{t+1} = 2\, w_i^t$ if $S_i \ni v^t$, and $w_i^{t+1} = w_i^t$ if $S_i \not\ni v^t$.
3. Output F^{n-k+1} of size k.

Lemma 2.1. *If some set S_i is exposed ℓ times by the end of the algorithm, its weight is 2^ℓ.*

Lemma 2.2. *Let $W^t = \sum_i w_i^t$ be the total weight at the beginning of any round t, and let the optimal solution expose at most ℓ^* elements in each set. Then $W^{t+1} \le W^t(1 + \frac{\ell^*}{n-k-t+1})$.*

Proof. Let T^* be the k elements picked in the optimal solution: they expose at most ℓ^* in each of the sets. There are $n-k$ elements in $V \setminus T^*$; in round t, $t-1$ of these elements might have been discarded in the previous rounds, leaving at least $n-k-t+1$ candidates in $(V \setminus T^*) \cap F^t$. Since each set contains at most ℓ^* of these elements, an averaging argument shows that there must be one element that is contained in sets whose weight is at most $W^t \times \frac{\ell^*}{n-k-t+1}$, and thus $W^t(v^t)$ is at most this quantity. But the weight adjustment step gives us that $W^{t+1} = W^t + W^t(v^t)$, which proves the lemma.

Theorem 2.1. *If the optimal solution exposes ℓ^* sets, then the number of sets exposed by the above algorithm is at most $O(\log n)\ell^* + O(\log m)$.*

Proof. By Lemma 2.2, the total weight of the system at the end of the algorithm is at most

$$W^0 \prod_{t=1}^{n-k} \left(1 + \frac{\ell^*}{n-k-t+1}\right) \le m \exp\left\{\sum_{t=1}^{n-k} \frac{\ell^*}{n-k-t+1}\right\} = m \exp\{O(\ell^* \log(n-k))\}. \quad (2.1)$$

Now if some set is exposed ℓ times, its weight (and hence the total weight of the set system) is at least 2^ℓ by Lemma 2.1; hence 2^ℓ is at most the quantity in (2.1). Taking logs, we get that $\ell \le \log m + O(\ell^* \log(n-k))$, which proves the theorem.

We note that an LP-rounding algorithm can give an improved $2 \cdot l^* + O(\log m)$ guarantee for this problem (details in Appendix A). However, that algorithm does not seem to extend to the general case we consider next.

2.2 Robust k-Median on General Metrics The algorithm on uniform metrics can be generalized to obtain an $O(\log m + \log k)$-approximation algorithm for this problem on arbitrary metric spaces: this is based on a reverse greedy algorithm for k-median due to Chrobak et al. [7] combined with a weight-update scheme similar to the one above. In Algorithm II, we first guess 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 instance. Note that a polynomial number of guesses suffice, and we can assume that $B \geq 1$ (by scaling distances).

Algorithm II: Robust k-Median for general metrics

1. Set $w_i^1 = 1$ for all $1 \leq i \leq m$ and $F^1 = V$.
2. For $t = 1, \cdots, n-k$ do:
   (a) For each $v \in F^t$ and each $1 \leq i \leq m$, 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 $F^t \setminus v$.
   (b) Set $\widehat{F^t} = \{v \in F^t \mid \delta_i^t(v) \leq \frac{B}{2} \quad \forall 1 \leq i \leq m\}$.
   (c) Let $v^t = \text{argmin}\{\sum_{i=1}^m w_i^t \cdot \delta_i^t(v) : v \in \widehat{F^t}\}$. Drop this vertex and set $F^{t+1} \leftarrow F^t \setminus v^t$.
   (d) Update $w_i^{t+1} = w_i^t \cdot (1 + \frac{1}{B})^{\delta_i^t(v^t)}$ for all $1 \leq i \leq m$.
3. Output F^{n-k+1} with k facilities.

Let us first prove a technical claim about the algorithm:

LEMMA 2.3. *At any time instant t, there exists a set $Q^t \subseteq F^t$ of size at most k such that for any scenario S_i*

$$\textstyle\sum_{v \in F^t \setminus Q^t} \delta_i^t(v) \leq 2\,\mathsf{Opt}.$$

Proof. Our arguments here follow those by Chrobak et al. [7]. Let T^* denote the optimal solution to the robust k-median problem, and $Q^t \subseteq F^t$ be the "projection" of T^* onto F^t—that is, for each vertex in T^* choose one closest vertex in F^t, and let Q^t be the resulting set of chosen vertices. Note that the size of the projection is $|Q^t| \leq |T^*| = k$.

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

$$\begin{aligned}
\textstyle\sum_{v \in R} \delta_i(v) &= \textstyle\sum_{v \in R} (\sum_{x \in S_i} d(x, F \setminus v) - \sum_{x \in S_i} d(x, F)) \\
&= \textstyle\sum_{x \in S_i} \sum_{v \in R} (d(x, F \setminus v) - d(x, F)) \\
&\leq \textstyle\sum_{x \in S_i} (d(x, Q) - d(x, F)) \qquad (2.2)\\
&\leq \textstyle\sum_{x \in S_i} (2 \cdot d(x, T^*) + d(x, F) - d(x, F)) \qquad (2.3)\\
&= 2 \textstyle\sum_{x \in S_i} d(x, T^*) \qquad (2.4)
\end{aligned}$$

To derive inequality (2.2), consider some demand $x \in S_i$, and $f_x \in F$ that serves x in solution F. If $f_x \in Q$, then for any $v \in R$, $d(x, F \setminus v) = d(x, F)$, and hence $\sum_{v \in R} (d(x, F \setminus v) - d(x, F)) = 0$; if $f_x \in R$, then $\sum_{v \in R} (d(x, F \setminus v) - d(x, F)) = d(x, F \setminus f_x) - d(x, F) \leq d(x, Q) - d(x, F)$. The inequality (2.3) follows from the triangle inequality:

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

where $f_x^* \in T^*$ is the facility that serves x in solution T^*. Finally, the last expression (2.4) is bounded above by $2 \cdot \max_{i=1}^m \sum_{x \in S_i} d(x, T^*) = 2 \cdot \mathsf{Opt}$.

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

LEMMA 2.4. *If $|F^t| > k$, the set $\widehat{F^t} \subseteq F^t$ in step 2b is non-empty.*

Proof. By Lemma 2.3, there is a set Q^t such that $\sum_{v \in F^t \setminus Q^t} \delta_i^t(v) \leq 2\mathsf{Opt} \leq B/2$ for all scenarios i. Moreover, the δ's are non-negative: hence each individual cost increase $\delta_i^t(v) \leq B/2$ for all $v \in R$, implying that $\widehat{F^t} \supseteq F^t \setminus Q^t$. Since Q^t has size at most k, and $|F^t| > k$, this must be non-empty.

LEMMA 2.5. *In iteration t, we have*

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

Proof. Let us fix Q^t as in Lemma 2.3, and sum the weighted $\delta_i^t(v)$ values, both over vertices in $F^t \setminus Q^t$ and over scenarios:

$$\begin{aligned}
\textstyle\sum_{v \in F^t \setminus Q^t} \sum_{i=1}^m w_i \cdot \delta_i^t(v) &= \textstyle\sum_{i=1}^m w_i \cdot \sum_{v \in F^t \setminus Q^t} \delta_i^t(v) \\
&\leq \textstyle\sum_{i=1}^m w_i \cdot 2\mathsf{Opt}
\end{aligned}$$

The inequality above follows from Lemma 2.3. Finally, a simple averaging shows that

$$\begin{aligned}
&\min_{v \in \widehat{F^t}} \{\textstyle\sum_{i=1}^m w_i^t \cdot \delta_i^t(v)\} \leq \min_{v \in F^t \setminus Q^t} \{\sum_{i=1}^m w_i^t \cdot \delta_i^t(v)\} \\
&\leq \frac{1}{|F^t \setminus Q^t|} \textstyle\sum_{v \in F^T \setminus Q^t} \sum_{i=1}^m w_i \cdot \delta_i^t(v) \leq \frac{2 \cdot \mathsf{Opt}}{n-k-t+1} \sum_{i=1}^m w_i^t,
\end{aligned}$$

which completes the proof of the lemma.

LEMMA 2.6. *Let $W^t = \sum_{i=1}^m w_i^t$ denote the weight in iteration t. Then $W^{t+1} \leq W^t \cdot e^{O(1)/(n-k-t+1)}$.*

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

$$w_i^{t+1} = w_i^t (1 + \tfrac{1}{B})^{\delta_i^t(v^t)} \leq w_i^t \cdot e^{\delta_i^t(v^t)/B},$$

where $v^t \in \widehat{F}$ is the facility that is dropped in iteration t. From the definition of the set $\widehat{F}$, $\delta_i^t(v^t)/B \in (0, \frac{1}{2}]$; moreover, for $y \in [0, \frac{1}{2}]$, we have $e^y \leq 1 + \sqrt{e} \cdot y$. This implies $w_i^{t+1} \leq w_i^t \cdot (1 + \sqrt{e}\,\frac{\delta_i^t(v^t)}{B})$, and hence $W^{t+1} \leq \sum_{i=1}^m w_i^t \cdot (1 + \sqrt{e}\,\frac{\delta_i^t(v^t)}{B}) \leq W^t + \frac{\sqrt{e}}{B} \sum_{i=1}^m w_i^t \cdot \delta_i^t(v^t)$. Using Lemma 2.5 and the fact that $B = \Theta(\mathsf{Opt})$, it follows that

$$W^{t+1} \leq W^t + \frac{\sqrt{e}}{B} \cdot \frac{2 \cdot \mathsf{Opt}}{n-k-t+1} \cdot W^t \leq (1 + \frac{\sqrt{e}/2}{n-k-t+1}) W^t.$$

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

THEOREM 2.2. *There is an $O(\log m + \log k)$-approximation algorithm for robust k-median.*

Proof. Let $\mathsf{Alg} = \max_i \sum_t \delta_i^t(v^t)$ denote the value of the solution F at the end of the algorithm, and let i_0 be the scenario achieving the maximum in the above expression. Hence, the total weight $W^{n-k+1} \geq w_{i_0}^{n-k+1} = (1+\frac{1}{B})^{\mathsf{Alg}}$. Furthermore, repeated applications of Lemma 2.6 implies that $W^{n-k+1} \leq W^1 \cdot e^{O(\log(n-k))} = m \cdot e^{\log(n-k)}$. Taking logarithms, we get $\mathsf{Alg} \leq (\log m + O(\log n))/\log(1+\frac{1}{B})$. Using $B = \Theta(\mathsf{Opt})$ and the fact that $\log(1+y) \geq y$ for $y \in [0,1]$, we get $\mathsf{Alg} \leq O(\log m + \log n) \cdot \mathsf{Opt}$. We can perform a filtering step to ensure that each scenario consists of at most k demands, which implies $n \leq mk$. Hence we have the theorem.

3 A General Framework for Robust Problems

Consider a location problem Π defined by an objective function Φ: given a metric space (V,d) and a set of demand points $S \subseteq V$, the cost of serving S from a set of facilities $T \subseteq V$ is given by $\Phi(T \mid S)$. We assume that Φ is a monotone non-increasing function in the set of facilities: i.e., $\Phi(T \cup \{x\} \mid S) \leq \Phi(T \mid S)$ for all facility sets T; in other words, opening more facilities does not cause the cost to increase. Given this monotonicity property, the natural solution would be to open all of V as facilities: however, we are given a parameter k, and want to choose a set of *at most k facilities* $T \subseteq V$ that minimizes the resulting cost $\Phi(T \mid S)$. (For instance, $\Phi(T \mid S) = \sum_{v \in S} d(x,T)$ defines the k-median objective function.)

In the robust version $\mathsf{Robust}(\Pi)$ of the location problem Π, given m different scenarios $S_1, S_2, \ldots, S_m \subseteq V$, the goal is to open k facilities T that minimize $\max_i \Phi(T \mid S_i)$. Let the optimal solution to this robust problem be denoted by T^*, the cost in scenario i be denoted by $\mathsf{Opt}_i = \Phi(T^* \mid S_i)$, and the global cost be denoted by $\mathsf{Opt} = \max_i \mathsf{Opt}_i$.

We show that a simple greedy-like procedure gives good approximations for this problem, given the following properties hold.

P1. (**Incremental Cost Computation**) For each facility set $T \subseteq V$ and each $x \in T$, we can efficiently compute the incremental cost of dropping x, to each scenario i:

$$\delta_i(T,x) \doteq \Phi(T \setminus x \mid S_i) - \Phi(T \mid S_i). \quad (3.5)$$

Note that the monotonicity property implies that this value is non-negative.

P2. (k-**Projection**) Given any set $F^* \subseteq V$ of size k and a set $F \subseteq V$ of size greater than k, we can prove the existence of a "small" set $Q \subseteq F$ of size $|Q| \leq k$ such that for all scenarios i,

$$\sum_{x \in F \setminus Q} \delta_i(F,x) \leq \beta \cdot \Phi(F^* \mid S_i). \quad (3.6)$$

In applications, it also suffices to prove (**P**1) and (**P**2) with any 'good' lower bound Φ' in place of Φ.[1] This modification is useful in cases where the incremental cost-functions δ_i's are approximable under Φ', but not under Φ. The first property (**P**1) naturally arises in a reverse-greedy-style algorithm for location problems. The second property (**P**2) is only required to prove the performance guarantee: it seems somewhat mysterious at first, and is useful in the same way as Lemma 2.3 was for the Robust k-Median problem. In fact, for applications to robust location problems it suffices to drop the quantification on F^* and just prove (**P**2) with $\beta \cdot \mathsf{Opt}$ in the right hand side of (3.7).

3.1 The General Algorithm Recall that we start off with a metric (V,d), demand sets $\{S_i\}_{i=1}^m$, and the cost incurred for demand set S_i and facility set T is given by $\Phi(T \mid S_i)$. Moreover, we assume that we have guessed a value $B \in [2\beta\,\mathsf{Opt}, 4\beta\,\mathsf{Opt}]$. The algorithm is the natural extension of that for Robust k-Median.

General Algorithm for Robust k-Location

1. Initialize weights $w_i^1 = 1$ for all $1 \leq i \leq m$ and set of facilities $F^1 = V$.
2. For $t = 1, \cdots, n-k$ do:
   (a) For each $v \in F^t$, let $\delta_i^t(v) = \Phi(F^t \setminus \{v\} \mid S_i) - \Phi(F^t \mid S_i)$.
   (b) Let $\widehat{F^t} = \{v \in F^t \mid \delta_i^t(v) \leq \frac{B}{2} \ \ \forall 1 \leq i \leq m\}$.
   (c) Let $v^t = \mathrm{argmin}\{\sum_{i=1}^m w_i^t \cdot \delta_i^t(v) : v \in \widehat{F^t}\}$ be a vertex with the least weighted increase.
   (d) Drop this vertex v^t and set $F^{t+1} \leftarrow F^t \setminus \{v^t\}$.
   (e) Update weights by $w_i^{t+1} = w_i^t \cdot (1+\frac{1}{B})^{\delta_i^t(v^t)}$ for all $1 \leq i \leq m$.
3. Output F^{n-k+1} with k facilities.

THEOREM 3.1. (GENERAL THM: ROBUST VERSION) *Given an instance of a robust location problem* $\mathsf{Robust}(\Pi)$ *satisfying the properties (**P1**) and (**P2**), the above algorithm is an $O(\beta(\log n + \log m))$ approximation, where m is the number of scenarios, and $n = |V|$.*

[1] Lower bound Φ' is said to be a γ-factor lower bound if it satisfies $\Phi'(T \mid S) \leq \Phi(T \mid S) \leq \gamma \cdot \Phi'(T \mid S)$ for all $T, S \subseteq V$. If we use Φ' in place of Φ, an additional factor γ appears in the approximation guarantees of Theorems 3.1 & 3.2.

The proof of this theorem is very similar to that for Robust k-median, and is deferred to the full version of the paper.

3.2 Framework for Stochastic Problems We can extend our framework to stochastic problems as well: given a location problem Π as in the previous section, scenarios $\{S_i\}$ that now come with probabilties $p_i \geq 0$ with $\sum_{i=1}^m p_i = 1$, the stochastic problem $\mathsf{Stoc}(\Pi)$ seeks to find a set T of size k that minimizes $\sum_{i=1}^m p_i\ \Phi(T \mid S_i)$. Once again, we denote the optimal set by T^*, each scenario's cost by $\mathsf{Opt}_i = \Phi(T^* \mid S_i)$, and $\mathsf{StocOpt} = \sum_i p_i \mathsf{Opt}_i$. We present the following algorithm for Stochastic location problems. We assume that we have guessed a value $B \in [2\beta\,\mathsf{Opt}, 4\beta\,\mathsf{Opt}]$. This algorithm is similar to that for the robust version; however it does not use the weight updates.

General Algorithm for Stochastic k-Location

1. Initialize set of facilities $F^1 = V$.
2. For $t = 1, \cdots, n-k$ do:
   (a) For each $v \in F^t$, let $\delta_i^t(v) = \Phi(F^t \setminus \{v\} \mid S_i) - \Phi(F^t \mid S_i)$.
   (b) Let $v^t = \text{argmin}\{\sum_{i=1}^m p_i\ \delta_i^t(v) : v \in F^t\}$ be a vertex with the least weighted increase (by probability).
   (c) Drop this vertex v^t and set $F^{t+1} \leftarrow F^t \setminus \{v^t\}$.
3. Output F^{n-k+1} with k facilities.

Theorem 3.2. (General Thm: Stoch. Version) *Given an instance of a stochastic location problem* $\mathsf{Stoc}(\Pi)$ *satisfying properties (**P**1) and (**P**2), the above algorithm is an $O(\beta \log n)$ approximation, where n is the number of vertices in V.*

Again, the proof of this theorem appears in the full version of the paper. We note that for this theorem it suffices to prove (**P**2) with F^* being the optimal stochastic solution T^*, in which case the right hand side in (3.7) becomes $\beta \cdot \mathsf{Opt}_i$. For all problems that we consider in this paper, we manage to prove (**P**2) itself, which immediately implies the corresponding results for both robust and stochastic versions.

4 Applications to Location Problems

In this section, we apply our general framework to the following location problems: k-tree, capacitated k-median, and fault-tolerant k-median. The main step in each of these applications is proving the k-projection property (**P**2). In the following subsection, we present the proof for the k-tree problem, which is representative of techniques used to prove the projection property. The proofs for capacitated k-median and fault-tolerant k-median appear in the full version of the paper.

4.1 The k-Tree Problem The *k-tree* problem is the following: given a metric space (V, d) and a set S of demand points, we want to open k facilities $T \subseteq V$ and build a forest of minimum cost in the induced metric $(T \cup S, d)$ so that each demand in S lies in a tree containing some facility in T. In particular, we want to minimize $d(\mathsf{forest}(T, S))$, where $\mathsf{forest}(T, S)$ denotes the minimum-cost forest in the metric induced on the set $T \cup S$ that connects each vertex in S to some vertex in T, so the objective function is:

$$\Phi(T \mid S) = d(\mathsf{forest}(T, S)) = \textstyle\sum_{e \in \mathsf{forest}(T,S)} d(e). \quad (4.7)$$

It is worth noting that once we choose the set T of facilities for a given demand set S, $\mathsf{forest}(T, S)$ corresponds to a minimum spanning tree in the metric obtained from $(T \cup S, d)$ by shrinking all the nodes in T to a single "root" vertex; hence the real effort is in choosing the set of facilities T.

Note that by taking an Euler tour of the forests constructed, we also get an approximate solution for the *k-person TSP* problem where we have to locate k salespersons, such that each demand point is visited by at least one salesperson, and the *total distance* traveled by these salespeople is minimized. The robust version of this problem can be thought of as the case when we locate k salespeople not knowing which of the m demand sets $S_1, \ldots, S_m$ will materialize: the objective function is then the total distance traversed by these salespeople on the worst of these m sets.

To apply the general framework of Section 3 to k-tree, it suffices to prove that this problem satisfies the two conditions. Property (**P**1) is easy, since we can compute the objective function given S and T, and hence can calculate the exact difference between the cost of any two solutions in polynomial time. The following lemma shows that the k-projection property (**P**2) is satisfied with $\beta = 4$.

Lemma 4.1. (Property **P**2 for k-Tree) *For every $F^* \subseteq V$ ($|F^*| = k$) and $F \subseteq V$ with $|F| > k$, there exists a subset $Q \subseteq F$ of size at most k such that $\forall\ 1 \leq i \leq m$:*

$$\textstyle\sum_{r \in F \setminus Q} [d(\mathsf{forest}(F \setminus r, S_i)) - d(\mathsf{forest}(F, S_i))] \leq 4 \cdot d(\mathsf{forest}(F^*, S_i))$$

Proof. We prove the lemma for a generic scenario $i \in [m]$. For each $f^* \in F^*$ choose a facility $\eta(f^*) \in F$ closest to f^*. Define $Q = \{\eta(f^*) \mid f^* \in F^*\}$ to be the chosen facilities; clearly $|Q| \leq |F^*| = k$. Let $\delta_i(r) = d(\mathsf{forest}(F \setminus r, S_i)) - d(\mathsf{forest}(F, S_i))$ denote the increase

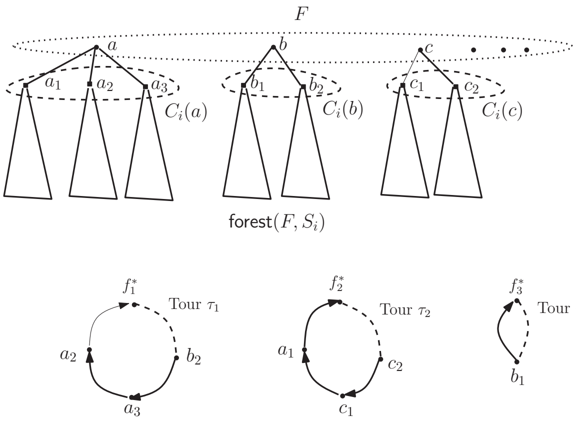

Figure 1: The solution $\mathsf{forest}(F, S_i)$ and one possible mapping $\sigma : C_i \to C_i \cup F^*$.

in the cost for S_i upon dropping facility $r \in F \setminus Q$. For each vertex $r \in F$, define:

$$C_i(r) = \{s \in S_i : (r, s) \text{ is an edge in } \mathsf{forest}(F, S_i)\}$$

Since each tree in $\mathsf{forest}(F, S_i)$ contains at most one vertex from F, we get that the sets $\{C_i(r)\}_{r \in F}$ are disjoint subsets of S_i; define $C_i = \sqcup_{r \in F} C_i(r)$. The following claim will be useful in the sequel.

Claim 1. *There exists a one-to-one map* $\sigma : C_i \to C_i \sqcup F^*$ *such that* $\sum_{v \in C_i} d(v, \sigma(v)) \leq 2 \cdot d(\mathsf{forest}(F^*, S_i))$.

Proof of Claim 1: Consider the optimal solution $\mathsf{forest}(F^*, S_i)$ induced on the vertices $F^* \cup S_i$, and double each tree in it to obtain $|F^*| = k$ vertex-disjoint tours $\tau_1, \cdots, \tau_k$ (some of which may be empty) on $F^* \cup S_i$ having total length at most $2 \cdot d(\mathsf{forest}(F^*, S_i))$. Note that each tour τ_j contains a distinct vertex from F^*, and together these tours contain all of $S_i \supseteq C_i$. We define a mapping $\sigma : C_i \to C_i \sqcup F^*$ as follows. Arbitrarily fix an orientation in each tour $\tau_1, \cdots, \tau_k$ and restrict each tour to vertices in $C_i \sqcup F^*$ (by short-cutting). Now each vertex $v \in C_i$ has a *unique* successor vertex $v' \in C_i \sqcup F^*$ in one of the (now oriented) tours: set $\sigma(v) = v'$. Observe that this mapping σ is one-to-one, and $\sum_{v \in C_i} d(v, \sigma(v)) \leq 2 \cdot d(\mathsf{forest}(F^*, S_i))$. ■

We now show how to modify the forest $\mathcal{F}' = \mathsf{forest}(F, S_i)$ to obtain a feasible forest $\mathcal{F}(r)$ that connects S_i to $F \setminus r$, for each $r \in F \setminus Q$. Furthermore, we will show that the length of each forest $\mathcal{F}(r)$ is not much more than $d(\mathcal{F}')$, which would bound $\delta_i(r)$. For any $r \in F \setminus Q$, the forest $\mathcal{F}(r)$ is constructed as follows: starting with the forest $\mathcal{F}'$, delete the edges $\{(r, v) \mid v \in C_i(r)\}$ adjacent to r in this forest, and add the following two edge-sets **(i)** $\{(v, \sigma(v)) \mid v \in C_i(r)\}$ where σ is the map defined in Claim 1; and **(ii)** $\{(f^*, \eta(f^*)) \mid f^* \in \sigma(C_i(r)) \cap F^*\}$ (recall that $\eta(f^*) \in Q$ is the closest facility to f^* in F).

We first show that $\mathcal{F}(r)$ is a feasible forest connecting S_i to $F \setminus r$: it suffices to argue that $C_i(r)$ is connected to $F \setminus r$. We claim that the edges $E_i(r) = \{(v, \sigma(v)) \mid v \in C_i(r)\}$ (edge-set **(i)** above) define a collection of paths such that each path contains a vertex in $(F^* \sqcup C_i) \setminus C_i(r)$. Clearly, the edges $E_i(r) \subseteq \sqcup_{1 \leq j \leq k} \tau_j$ correspond to a collection of paths or cycles. They are acyclic since each tour τ_j contains an F^*-vertex and $E_i(r)$ contains no σ-edge out of F^*. Now suppose some path in this collection contained only $C_i(r)$-vertices: the σ-edge out of one of its end points (which are in $C_i(r)$) would have to be missing from $E_i(r)$, which is a contradiction! So each path in $E_i(r)$ contains a vertex in $(F^* \sqcup C_i) \setminus C_i(r)$. The paths in $E_i(r)$ that contain a vertex from $C_i \setminus C_i(r)$ are clearly connected to $F \setminus r$, since the vertices in $C_i \setminus C_i(r)$ remain connected to $F \setminus r$. Each of the remaining paths contains a vertex from F^*, and the direct edges from $\sigma(C_i(r)) \cap F^*$ to $Q \subseteq (F \setminus r)$ (edge-set **(ii)** above) connect the remaining $C_i(r)$-vertices to $F \setminus r$.

Next, note that we can upper bound the increase in cost $\delta_i(r) \leq d(\mathcal{F}(r)) - d(\mathcal{F}')$ by $-\sum_{v \in C_i(r)} d(r, v) + \sum_{v \in C_i(r)} d(v, \sigma(v)) + \sum_{f^* \in \sigma(C_i(r)) \cap F^*} d(f^*, Q)$. The last term of this expression can be bounded thus:

$$\begin{aligned}
&\textstyle\sum_{f^* \in \sigma(C_i(r)) \cap F^*} d(f^*, Q) \\
&= \textstyle\sum_{f^* \in \sigma(C_i(r)) \cap F^*} d(f^*, F) \\
&\leq \textstyle\sum_{f^* \in \sigma(C_i(r)) \cap F^*} [d(f^*, \sigma^{-1}(f^*)) + d(\sigma^{-1}(f^*), F)] \\
&\leq \textstyle\sum_{v \in C_i(r)} [d(\sigma(v), v) + d(v, F)]
\end{aligned}$$

Above, the first equality is by the choice of $Q \subseteq F$, and subsequent inequality follows from triangle inequality and the observation that σ is one-to-one. Plugging the final expression into the original bound, the cost increase $\delta_i(r)$ is bounded by

$$\begin{aligned}
&2 \cdot \textstyle\sum_{v \in C_i(r)} d(v, \sigma(v)) + \sum_{v \in C_i(r)} [d(v, F) - d(v, r)] \\
&\leq 2 \cdot \textstyle\sum_{v \in C_i(r)} d(v, \sigma(v))
\end{aligned}$$

where the final inequality uses that $r \in F$. Now summing over all $r \in F \setminus Q$, we get $\sum_{r \in F \setminus Q} \delta_i(r) \leq 2 \cdot \sum_{r \in F \setminus Q} \sum_{v \in C_i(r)} d(v, \sigma(v)) \leq 2 \cdot \sum_{v \in C_i} d(v, \sigma(v))$. Finally, by Claim 1, this last expression is at most $4 \cdot d(\mathsf{forest}(F^*, S_i))$, which completes the proof.

For the k-person TSP problem (mentioned earlier), we can work with the k-tree lower bound in place of Φ in properties (**P**1) & (**P**2); as noted in Section 3, this suffices to obtain the robust/stochastic results for k-person TSP as well.

4.2 Capacitated k-Median Problem In the capacitated k-Median problem, we are given a metric (V, d), a subset $S \subset V$ of demand points, a number k and a capacity μ such that $|S| \leq k \cdot \mu$. We are required to open at most k facilities F and give an assignment $\rho : S \to F$ of demand points to open facilities such that at most μ demand points are assigned to any open facility (i.e., $|\rho^{-1}(f)| \leq \mu$ for all $f \in F$). Note that given the map ρ, the facility set $F = \rho(S)$ is implicitly specified; moreover, given a facility set F one can find the map ρ by solving a min-cost b-matching problem. The cost of a solution F is $\Phi(F \mid S) = \sum_{v \in S} d(v, \rho(v))$, for ρ as defined above. In the full version, we show that this problem satisfies the conditions for our general framework, and hence we obtain the corresponding results for robust and stochastic capacitated k-Median.

4.3 Fault-tolerant k-Median In this problem, we are given a demand set $S \subseteq V$, and also a *requirement* $r_v \in \{1, 2, \ldots, k\}$ for each client $v \in S$: the goal is to open some 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 a facility set F, each client $v \in S$ is connected to the set C_v of r_v *distinct* facilities in F closest to v. So the cost of the solution is $\Phi(F \mid S) = \sum_{v \in S} \sum_{f \in C_v} d(v, f)$. In the full version, we prove the k-projection property for this problem, and hence obtain a logarithmic approximation for robust fault-tolerant k-Median.

5 The Stochastic k-Center Problem

In the previous sections, we gave efficient approximation algorithms for some robust and stochastic location problems. In this section, we study another natural location problem, that turns out to be fairly difficult to approximate. An instance of the *Stochastic k-Center* problem consists of subsets $\{S_i\}_{i=1}^m$ of vertices in a metric space (V, d), and the goal is to open a set T of k facilities to minimize $\sum_{i=1}^m \max_{x \in S_i} d(x, T)$. (Under our generic definition of stochastic location problems, this is really only a special case where all scenarios have equal probabilities.) Note that if there is only one set S_i, this is the classical k-center problem, for which several 2-approximations are known, and this is the best we can do unless $P = NP$.

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 which maximize the number of edges in the induced subgraph. The *minimization version* of Dense k-Subgraph will also be useful, in which the goal is to pick k edges to minimize the number of vertices incident to these edges. The best result known is that of Feige et al. [15] who give an $O(n^\delta)$-approximation algorithm for some $\delta < 1/3$. The problem is believed to be fairly hard, and [13, 22] 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 Stochastic k-Center problem on the uniform metric, and hence can formulate it as a set-covering-type problem:

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

Theorem 5.1. (Stochastic k-Center Hardness) *An α-approximation to the Stochastic k-Center problem on the uniform metric implies an α-approximation for the minimization version of the* Dense k-Subgraph *problem.*

Proof. Given an instance G of the Dense k-Subgraph problem, we construct an auxiliary bipartite graph $H = (P, Q, E')$ with $P = V(G)$, $Q = E(G)$, and E' connects $p \in P$ and $q \in Q$ if the edge corresponding to q is incident to the vertex corresponding to p. Note that solving the minimization version of Dense k-Subgraph on G is the same as finding a subset $A \subseteq Q$ with size k such that the number of "vertices" in P that are adjacent (using the edges in H) to "edges" in $A \subseteq Q$ is minimized.[2]

We can now turn this problem on $H = (P, Q, E')$ into an instance of Stochastic k-Center. The metric we consider is the uniform metric on points corresponding to Q. We have one scenario for each $p \in P$, where the scenario S_p contains all the demand points $q \in Q$ such that $(p, q) \in E'$. Now consider this instance of the Stochastic k-Center problem with parameter $n - k$: it seeks to find a subset $A' \subseteq Q$ with $|A'| \leq n - k$ to minimize the number of sets S_p not contained within A'. In other words, it finds a set $A \subseteq Q$ of size at least k to minimize the the number of vertices in its neighborhood $\Gamma(A) = \{p \in P \mid \exists q \in A, (p, q) \in E'\}$. Hence, if we do not violate the cardinality constraints, any algorithm for the Stochastic k-Center problem on the uniform metric with approximation ratio α gives an identical approximation for the Dense k-Subgraph minimization problem.

[2]In other words, the Dense k-Subgraph problem reduces to a bipartite graph expansion problem, where we want to find a set of vertices of size at least k on the right hand side that expands the least.

References

[1] Arya, V., Garg, N., Khandekar, R., Meyerson, A., Munagala, K., and Pandit, V. Local search heuristic for k-median and facility location problems. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing* (2001), ACM Press, pp. 21–29.

[2] Birge, J. R., and Louveaux, F. *Introduction to stochastic programming*. Springer Series in Operations Research. Springer-Verlag, New York, 1997.

[3] Cesa-Bianchi, N., and Lugosi, G. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[4] Charikar, M., Chekuri, C., and Pál, M. Sampling bounds for stochastic optimization. In *Approximation, randomization and combinatorial optimization*, vol. 3624 of *Lecture Notes in Comput. Sci.* Springer, Berlin, 2005, pp. 257–269.

[5] Charikar, M., and Guha, S. Improved combinatorial algorithms for facility location problems. *SIAM J. Comput. 34*, 4 (2005), 803–824.

[6] Charikar, M., Guha, S., Tardos, É., and Shmoys, D. B. A constant-factor approximation algorithm for the k-median problem. *J. Comput. System Sci. 65*, 1 (2002), 129–149. Special issue on STOC, 1999 (Atlanta, GA).

[7] Chrobak, M., Kenyon, C., and Young, N. The reverse greedy algorithm for the metric k-median problem. *Inform. Process. Lett. 97*, 2 (2006), 68–72.

[8] Chuzhoy, J., and Rabani, Y. Approximating k-median with non-uniform capacities. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (New York, 2005), ACM, pp. 952–958 (electronic).

[9] Cooper, L. Bounds on the Weber problem solution under conditions of uncertainty. *Journal of Regional Science 18*, 1 (1978), 87–93.

[10] Daskin, M., and Owen, S. Location models in transportation. *Handbook of Transportation Science* (1999), 311–360.

[11] Dhamdhere, K., Goyal, V., Ravi, R., and Singh, M. How to pay, come what may: Approximation algorithms for demand-robust covering problems. In *46th FOCS* (2005), pp. 367–378.

[12] Even, G., Garg, N., Könemann, J., Ravi, R., and Sinha, A. Min-max tree covers of graphs. *Oper. Res. Lett. 32*, 4 (2004), 309–315.

[13] Feige, U. Relations between average case complexity and approximation complexity. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing* (New York, 2002), ACM, pp. 534–543.

[14] Feige, U., Jain, K., Mahdian, M., and Mirrokni, V. Robust combinatorial optimization with exponential scenarios. In *12th IPCO* (2007), pp. 439–453.

[15] Feige, U., Kortsarz, G., and Peleg, D. The dense k-subgraph problem. *Algorithmica 29*, 3 (2001), 410–421.

[16] Gandhi, R., Khuller, S., Parthasarathy, S., and Srinivasan, A. Dependent rounding and its applications to approximation algorithms. *J. ACM 53*, 3 (2006).

[17] Golovin, D., Goyal, V., and Ravi, R. Pay today for a rainy day: Improved approximation algorithms for demand-robust min-cut and shortest path problems. In *Proceedings of the 23rd Symposium on Theoretical Aspects of Computer Science, STACS 2006* (2006), B. Durand and W. Thomas, Eds., vol. 3884 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 206–217.

[18] Gupta, A., Pál, M., Ravi, R., and Sinha, A. Boosted sampling: Approximation algorithms for stochastic optimization problems. In *36th STOC* (2004), pp. 417–426.

[19] Hochbaum, D., and Shmoys, D. A best possible heuristic for the k-center problem. *Math. Oper. Res. 10*, 2 (1985), 180–184.

[20] Immorlica, N., Karger, D., Minkoff, M., and Mirrokni, V. On the costs and benefits of procrastination: Approximation algorithms for stochastic combinatorial optimization problems. In *15th SODA* (2004), pp. 684–693.

[21] Jain, K., and Vazirani, V. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangean relaxation. *J. ACM 48*, 2 (2001), 274–296. (Preliminary version in *40th FOCS*, pages 2–13, 1999).

[22] Khot, S. Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J. Comput. 36*, 4 (2006), 1025–1071.

[23] Krause, A., McMahan, B., Guestrin, C., and Gupta, A. Selecting observations under multiple objectives. submitted, 2007.

[24] Lin, G., Nagarajan, C., Rajaraman, R., and Williamson, D. P. A general approach for incremental approximation and hierarchical clustering. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm* (New York, NY, USA, 2006), ACM Press, pp. 1147–1156.

[25] Louveaux, F. Stochastic location analysis. *Location Science 1* (1993), 127–154.

[26] Mettu, R. R., and Plaxton, C. G. The online median problem. In *41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000)*. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000, pp. 339–348.

[27] Mirchandani, P., and Odoni, A. Locations of Medians on Stochastic Networks. *Transportation Science 13* (1979), 85–97.

[28] Ravi, R., and Sinha, A. Hedging uncertainty: Approximation algorithms for stochastic optimization problems. In *10th IPCO* (2004), pp. 101–115.

[29] Rosenblatt, M. J., and Lee, H. L. A robustness approach to facilities design. *International Journal of Production Research 25* (1987), 479–486.

[30] Sahni, S., and Gonzalez, T. P-complete approxi-

mation problems. *J. ACM 23* (1976), 555–565.
[31] Sheppard, E. A conceptual framework for dynamic location allocation analysis. *Environment and Planning A 6(5)* (1974), 547–564.
[32] Shmoys, D. B., and Swamy, C. An approximation scheme for stochastic linear programming and its application to stochastic integer programs. *J. ACM 53*, 6 (2006), 978–1012 (electronic).
[33] Snyder, L. Facility location under uncertainty: a review. *IIE Transactions 38*, 7 (2006), 547–564.
[34] Swamy, C., and Shmoys, D. B. Fault-tolerant facility location. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (Baltimore, MD, 2003)* (New York, 2003), ACM, pp. 735–736.
[35] van Hentenryck, P., Bent, R., and Upfal, E. Online stochastic optimization under time constraints. *Annals of Operations Research* (2007). To appear.
[36] Weaver, J., and Church, R. Computational procedure for location problems on stochastic networks. *Transportation Science 17*, 2 (1983), 168–180.
[37] Welzl, E. Suchen und Konstruieren durch Verdoppeln. In *Highlights der Informatik*, I. Wegener, Ed. Springer, Berlin, Germany, 1996, pp. 221–228.

A Improved guarantee for Robust k-Median on uniform metrics

We consider the following natural linear relaxation for the Robust k-Median problem on a uniform metric. Recall that there are n elements V, and m scenarios $S_1, \cdots, S_m \subseteq V$; the goal is to pick k elements so as to minimize the number of uncovered elements in any scenario.

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

Above 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. [16], which implies the following in our context.

Theorem A.1. (Gandhi et al. [16]) *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* $S \subseteq V$, *if* $\mu_S = E[\sum_{e \in S} X_e]$:

$$Pr[\sum_{e \in S} X_e > (1+\delta)\mu_S] \leq e^{\frac{-\delta^2 \mu_S}{2+\delta}} \qquad \forall \delta \geq 0$$

Using this rounding scheme, it is clear that we always pick exactly k elements. For any scenario S_i, we have $\mu_i = E[\sum_{e \in S_i} X_e] = \sum_{e \in S_i} x_e \leq z$. Using property (3) in the above theorem with $\delta_i = \frac{\mu_i + \alpha}{\mu_i}$ for each S_i ($\alpha > 0$ will be fixed later), we have for each $1 \leq i \leq m$:

$$Pr[\sum_{e \in S_i} X_e > 2\mu_i + \alpha] \leq e^{-\frac{(\mu_i+\alpha)^2}{3\mu_i+\alpha}} \leq e^{-\alpha/3}$$

Setting $\alpha = O(\log m)$, and using $\mu_i \leq z$ for all scenarios S_i, we get $Pr[\sum_{e \in S_i} X_e > 2 \cdot z + \alpha] \leq \frac{1}{m^2}$ for each $1 \leq i \leq m$. Now, by a union bound over all scenarios we obtain that with probability at least $1 - \frac{1}{m}$, the maximum number of uncovered elements in any scenario is at most $2 \cdot z + O(\log m)$.