Improved Approximation Guarantees for Lower-Bounded Facility Location*

Sara Ahmadian **1 and Chaitanya Swamy **1

Combinatorics and Optimization, Univ. Waterloo, Waterloo, ON N2L 3G1.
{sahmadian,cswamy}@math.uwaterloo.ca

Abstract. We consider the *lower-bounded facility location* (LBFL) problem, which is a generalization of *uncapacitated facility location* (UFL), where each open facility is required to serve a certain *minimum* amount of demand. The current best approximation ratio for LBFL is 448 [17]. We substantially advance the state-of-the-art for LBFL by improving its approximation ratio from 448 [17] to 82.6.
Our improvement comes from a variety of ideas in algorithm design and analysis, which also yield new insights into LBFL. Our chief algorithmic novelty is to present an improved method for solving a more-structured LBFL instance obtained from $\mathcal{I}$ via a bicriteria approximation algorithm for LBFL, wherein all clients are aggregated at a subset $\mathcal{F}'$ of facilities, each having at least αM co-located clients (for some $\alpha \in [0,1]$). The algorithm in [17] proceeds by reducing $\mathcal{I}_2(\alpha)$ to CFL. One of our key insights is that one can reduce the resulting LBFL instance, denoted $\mathcal{I}_2(\alpha)$, to a problem we introduce, called *capacity-discounted UFL* (CDUFL), which is a structured special case of capacitated facility location (CFL). We give a simple local-search algorithm for CDUFL based on add, delete, and swap moves that achieves a significantly-better approximation ratio than the current-best approximation ratio for CFL, which is one of the reasons behind our algorithm's improved approximation ratio.

1 Introduction

Facility location problems have been widely studied in the Operations Research community (see, e.g., [13]). In its simplest version, *uncapacitated facility location* (UFL), we are given a set of facilities with opening costs, and a set of clients, and we want to open some facilities and assign each client to an open facility so as to minimize the sum of the facility-opening and client-assignment costs. This problem has a wide range of applications. For example, a company might want to open its warehouses at some locations so that its total cost of opening warehouses and servicing customers is minimized.

We consider the *lower-bounded facility location* (LBFL) problem, which is a generalization of UFL where each open facility is required to serve a certain *minimum* amount of demand. More formally, an LBFL instance $\mathcal{I}$ is specified by a set $\mathcal{F}$ of facilities, a set $\mathcal{D}$ of clients, and an integer M. Opening facility

* A full version [1] is available on the CS arXiv

** Research supported in part by NSERC grant 327620-09 and the second author's Ontario Early Researcher Award.

i incurs a *facility-opening cost* f_i, and assigning a client j to a facility i incurs a *connection cost* c_{ij}. A feasible solution specifies a subset $F \subseteq \mathcal{F}$ of facilities, and assigns each client j to an open facility $i(j) \in F$ so that *each open facility serves at least M clients*. The cost of such a solution is the sum of the facility-opening and connection costs, that is, $\sum_{i \in F} f_i + \sum_j c_{i(j)j}$, and the goal is to find a feasible solution of minimum cost. As is standard in the study of facility location problems, we assume throughout that c_{ij}s form a metric. We use the terms connection cost and assignment cost interchangeably in the sequel.

LBFL can be motivated from various perspectives. This problem was introduced independently by Karger and Minkoff [8], and Guha et al. [5] (who called the problem *load-balanced facility location*, both of whom arrived at LBFL as a means of solving their respective buy-at-bulk style network design problems. LBFL arises as a natural subroutine in solving buy-at-bulk problems because obtaining a near-optimal solution often entails aggregating a certain minimum demand at certain hub locations, and then connecting the hubs via links of lower per-unit-demand cost (and higher fixed cost). LBFL also finds direct applications in supply-chain logistics problems, where the lower-bound constraint can be used to model the fact that it is not profitable or feasible to use services unless they satisfy a certain minimum demand. For example (as noted in [17]), Lim et al. [11], use LBFL to abstract a transportation problem faced by a company that has to determine the allocation of cargo from customers to carriers, who then ship their cargo overseas. Here the lower bound arises because each carrier, if used, is required (by regulation) to deliver a minimum amount of cargo.

Clearly, LBFL with $M = 1$ is simply UFL, and hence, is *NP*-hard; consequently, we are interested in designing approximation algorithms for LBFL. The first constant-factor approximation algorithm for LBFL was devised by Svitkina [17], whose approximation ratio is 448. Prior to this, the only known approximation guarantees were *bicriteria guarantees*. [8] and [5] independently devised (ρ, α)-approximation algorithms via a reduction to UFL: these algorithms return a solution of cost at most ρ times the optimum where each open facility serves at least αM clients ($\alpha < 1$, ρ is a function of α).

Our results and techniques. We devise an approximation algorithm for LBFL that achieves a substantially-improved approximation guarantee of 82.6 (Theorem 1), thus significantly advancing the state-of-the-art for LBFL. Our improvement comes from a combination of ideas in algorithm design and analysis, and yields new insights about the approximability of LBFL. In order to describe the ideas underlying our improvement, we first briefly sketch Svitkina's algorithm.

Svitkina's algorithm begins by using the reduction in [8, 5] to obtain a bicriteria solution for $\mathcal{I}$, which is then used to convert $\mathcal{I}$ into an LBFL instance $\mathcal{I}_2$ with facility-set $\mathcal{F}' \subseteq \mathcal{F}$ having the following structure: (i) all clients are aggregated at $\mathcal{F}'$ with each facility $i \in \mathcal{F}'$ having $n_i \geq \alpha M$ co-located clients; (ii) all facilities in $\mathcal{F}'$ have zero opening costs; and (iii) near-optimal solutions to $\mathcal{I}_2$ translate to near-optimal solutions to $\mathcal{I}$ (and vice versa). The goal now is to identify a subset of $\mathcal{F}'$ to close, such that transferring the clients aggregated at these closed facilities to the remaining (open) facilities in $\mathcal{F}'$ ensures that each remaining facility serves at least M demand (and the cost incurred is "small"). [17] shows that one can achieve this by solving a suitable CFL instance. Essentially the idea is that a facility i that remains open corresponds to a *demand point* in the CFL instance that requires $M - n_i$ units of demand, and a facility i that is closed maps to a *supply point* in the CFL instance having n_i units that can be supplied to demand points (i.e., open facilities). Of course, one does not know beforehand which facilities will be closed and which will remain open; so to encode this correspondence in the CFL instance, we create at every location $i \in \mathcal{F}'$, a supply point with (suitable opening cost and) capacity M, and a demand point with demand $M - n_i$ if $n_i \leq M$ (so the supply point at i has n_i residual capacity after satisfying this demand). (Assume $n_i \leq M$ for simplicity; facilities with $n_i > M$ are treated differently.) [17] argues that a CFL-solution can be mapped to an $\mathcal{I}_2$-solution without increasing the cost incurred by much; since CFL admits an $O(1)$-approximation algorithm, one obtains an $O(1)$-approximate solution to $\mathcal{I}_2$, and hence to the original LBFL instance $\mathcal{I}$.

Our algorithm also proceeds by (a) obtaining an LBFL instance $\mathcal{I}_2$ satisfying properties (i)–(iii) mentioned above, (b) solving $\mathcal{I}_2$, and (c) mapping the $\mathcal{I}_2$-solution to a solution to $\mathcal{I}$, but our implementation of steps (a) and (b) differs from that in Svitkina's algorithm. These modified implementations, which are independent of each other and yield significant improvements in the overall approximation ratio even when considered in isolation, result in our much-improved approximation ratio. We detail how we perform step (a) later, and focus first on describing how we solve $\mathcal{I}_2$, which is our chief algorithmic contribution.

Our key insight is that one can solve the LBFL instance $\mathcal{I}_2$ by reducing it to a new problem we introduce that we call *capacity-discounted* UFL (CDUFL), which closely resembles UFL and admits an algorithm (that we devise) with a much better approximation ratio than CFL. A CDUFL-instance has the property that every facility is either uncapacitated (i.e., has infinite capacity), or has finite capacity and *zero* facility cost. The CDUFL instance we construct consists of the same supply and demand points as in the reduction of $\mathcal{I}_2$ to CFL in [17], except that all supply points with non-zero opening cost are now uncapacitated. (Interestingly, if all facilities in $\mathcal{I}_2$ have $n_i \leq M$, the CDUFL instance is in fact a UFL-instance!)

We prove two crucial algorithmic results. The "standard" integrality-gap example for the natural LP-relaxation of CFL can be cast as a CDUFL instance, thus showing that the natural LP-relaxation for CDUFL has a large integrality gap, and we are not aware of any LP-relaxation with constant integrality gap. Circumventing this difficulty, we devise a local-search algorithm for CDUFL based on add, swap, and delete moves that achieves the *same performance guarantees* as the corresponding local-search algorithm for UFL [2] (Section 4.2). The local-search algorithm yields significant dividends in the overall approximation ratio because not only is its approximation ratio for CDUFL better than the state-of-the-art for CFL, but also because it yields separate (asymmetric) guarantees on the facility-opening and assignment costs, which allows one to perform a tighter analysis. Second, we show that any near-optimal CDUFL-solution can be mapped to a near-optimal solution to $\mathcal{I}_2$ (Section 4.1). As in [17], in the CDUFL-solution, an open supply point i (which corresponds to closing facility i) may send less than n_i supply to other demand points, so that closing down i entails transferring its residual clients to open facilities. But since some supply points are now uncapacitated, it could also be that i sends *more* than n_i supply to other demand points. We argue that this artifact can also be handled without increasing the solution cost by much, by opening the facilities in a carefully-chosen subset of $\{i\} \cup \{\text{demand points satisfied by } i\}$ and closing down the remaining facilities. For *every* α (recall that the LBFL instance $\mathcal{I}_2$ is specified in terms of a parameter α), the resulting approximation factor for $\mathcal{I}_2$ (Theorem 5) is better than the guarantee obtained for $\mathcal{I}_2$ in Svitkina's algorithm; this in turn translates (by choosing α suitably) to an improved solution to the original instance.

We now discuss how we implement step (a), that is, how we obtain instance $\mathcal{I}_2$. As in [17], we arrive at $\mathcal{I}_2$ by computing a bicriteria solution to LBFL, but we obtain this bicriteria solution in a different fashion (see Section 3). The reduction from LBFL to UFL in [8, 5] proceeds by setting the opening cost of facility i to $f_i + \frac{2\alpha}{1-\alpha} \cdot \sum_{j \in \mathcal{D}(i)} c_{ij}$, where $\mathcal{D}(i)$ is the set of M clients closest to i, solving the resulting UFL instance, and postprocessing using (single-facility) delete moves if such a move improves the solution cost. We modify this reduction subtly by creating a UFL instance, where facility i's opening cost is instead set to $f_i + 2\alpha M R_i(\alpha)$, where $R_i(\alpha)$ is the distance between i and the αM-closest client to it. As in the case of the earlier reduction, we argue that each open facility i in the resulting solution (obtained by solving UFL and postprocessing) serves at least αM clients. The overall bound we obtain on the total cost now includes various $R_i(\alpha)$ terms. Instead of plugging in the (weak) bound $MR_i(\alpha) \leq \frac{\sum_{j \in \mathcal{D}(i)} c_{ij}}{1-\alpha}$ (which would yield the same guarantee as that obtained via the earlier reduction), we are able to perform a tighter analysis by choosing α from a suitable distribution and leveraging the fact that $M \int_0^1 R_i(\alpha) d\alpha = \sum_{j \in \mathcal{D}(i)} c_{ij}$. (This can easily be derandomized, since there are only M combinatorially distinct choices for α.) These simple modifications yield a surprising amount of improvement in the approximation factor, which is reminiscent of the mileage provided by (random) α-points for various scheduling problems and UFL [15, 16]. Also, we observe that one can obtain further improvements by using the local-search algorithm of [3, 2] to solve the above UFL instance: this is because the resulting solution is then already postprocessed, which allows us to exploit the asymmetric bounds on the facility-opening and assignment costs provided by the local-search algorithm via scaling, and improve the approximation ratio.

Finally, we remark that the study of CDUFL may provide useful and interesting insights about CFL. CDUFL is a special case of CFL that despite its special structure inherits the intractability of CFL with respect to LP-based approximation guarantees. If one seeks to develop LP-based techniques and algorithms for CFL (which has been a long-standing and intriguing open question), then one needs to understand how one can leverage LP-based techniques for CDUFL, and it is plausible that LP-based insights developed for CDUFL may yield similar insights for CFL (and potentially LP-based approximation guarantees for CFL).

Related work. LBFL was independently introduced by [8] and [5], who used it as a subroutine to solve the *maybecast* and *access network design* problems respectively. Their ideas lead to bicriteria guarantees for LBFL and play a preprocessing role both in Svitkina's algorithm [17] and (slightly indirectly) in our algorithm.

There is a large body of literature that deals with approximation algorithms for (metric) UFL, CFL and its variants; see [14] for a survey on UFL. The first constant approximation guarantee for UFL was obtained by Shmoys et al. [15] via an LP-rounding algorithm, and the current state-of-the-art is a 1.488-approximation algorithm due to Li [10]. Local-search techniques have also been utilized to obtain $O(1)$-approximation guarantees for UFL [9, 3, 2]. We apply some of the ideas of [3, 2] in our algorithm. Starting with the work of Korupolu et al. [9], various local-search algorithms with constant approximation ratios have been devised for CFL, with the current-best approximation ratio being $5.83 + \epsilon$ [18]. Local-search approaches are however not known to work for LBFL; in the full version [1], we show that local search based on add, delete, and swap moves yields poor approximation guarantees. A related problem is *universal facility location* (UniFL), a generalization of UFL where the facility cost depends on the number of clients served by the facility. UniFL with non-decreasing functions was introduced by [6, 12], and [12] obtained a constant approximation algorithm. We are not aware of any work on UniFL with arbitrary non-increasing functions (which generalizes LBFL). [4] give a constant approximation for the case where the cost-functions do not decrease too steeply (the constant depends on the steepness); notice that LBFL does not fall into this class so their results do not apply here.

2 Problem definition and notation

Recall that an LBFL instance $\mathcal{I}$ consists of a set $\mathcal{F}$ of facilities with facility-opening costs $\{f_i\}$, a set $\mathcal{D}$ of clients, metric connection (or assignment) costs $\{c_{ij}\}$ specifying the cost of assigning client j to facility i, and a (integer) parameter M. Our objective is to open a subset F of facilities and assign each client j to an open facility $i(j) \in F$, so that at least M clients are assigned to each open facility, and the total cost incurred, $\sum_{i \in F} f_i + \sum_j c_{i(j)j}$, is minimized.

Let F^* and C^* denote respectively the facility-opening and assignment cost of an optimal solution to $\mathcal{I}$; we will often refer to this solution as "the optimal solution" in the sequel. We sometimes abuse notation and also use F^* to denote the set of open facilities in this optimal solution. Let $OPT = F^* + C^*$ denote the total optimal cost. For a facility $i \in \mathcal{F}$, let $\mathcal{D}(i)$ be the set of M clients closest to i, and $R_i(\alpha)$ denote the distance between i and the $\lceil \alpha M \rceil$-closest client to i; that is, if $\mathcal{D}(i) = \{j_1, \ldots, j_M\}$, where $c_{ij_1} \le \ldots \le c_{ij_M}$, then $R_i(\alpha) = c_{ij_{\lceil \alpha M \rceil}}$ (for $0 < \alpha \le 1$). Let $R^*(\alpha) = \sum_{i \in F^*} R_i(\alpha)$. Observe that each $R_i(\alpha)$ is an increasing function of α, $M \int_0^1 R_i(\alpha) d\alpha = \sum_{j \in \mathcal{D}(i)} c_{ij}$, and $R_i(\alpha) \le \big(\sum_{j \in \mathcal{D}(i)} c_{ij}\big)/(M - \lceil \alpha M \rceil + 1) \le \frac{\sum_{j \in \mathcal{D}(i)} c_{ij}}{M(1-\alpha)}$. Hence, $R^*(\alpha)$ is an increasing function of α, $M \int_0^1 R^*(\alpha) d\alpha \le C^*$, and $R^*(\alpha) \le \frac{C^*}{M(1-\alpha)}$.

3 Our algorithm and the main theorem

We now give a high-level description of our algorithm using certain building blocks that are supplied in the subsequent sections.

(1) **Obtaining a bicriteria solution.** Construct a UFL instance with the same set of facilities and clients, and the same assignment costs as $\mathcal{I}$, where the opening cost of facility i is set to $f_i + 2\alpha M R_i(\alpha)$. Use the local-search algorithm for UFL in [3] or [2] with scaling parameter $\gamma > 0$ to solve the resulting UFL instance. (We set α, γ suitably to get the desired approximation; see Theorem 1.) Let $\mathcal{F}' \subseteq \mathcal{F}$ be the set of facilities opened in the UFL-solution. Claim 2 and Lemma 3 show that each $i \in F'$ serves at least αM clients.

(2) **Transforming to a structured LBFL instance.** We use the bicriteria solution obtained above to transform $\mathcal{I}$ into another structured LBFL instance $\mathcal{I}_2$ as in [17]. In the instance $\mathcal{I}_2$, we set the opening cost of each $i \in \mathcal{F}'$ to zero, and we "move" to i all the $n_i \geq \alpha M$ clients assigned to it, that is, all these clients are now co-located at i. So $\mathcal{I}_2$ consists of only the points in $\mathcal{F}'$ (which forms both the facility-set and client-set). We sometimes denote this instance by $\mathcal{I}_2(\alpha)$ to indicate explicitly that its specification depends on α.

(3) Solve $\mathcal{I}_2$ using the method described in Section 4. Obtain a solution to $\mathcal{I}$ by opening the same facilities and making the same client assignments as in the solution to $\mathcal{I}_2$.

Analysis. Our main theorem is as follows.

Theorem 1. *For any* $\alpha \in (0.5, 1]$ *and* $\gamma > 0$*, the above algorithm returns a solution to* $\mathcal{I}$ *of cost at most*

$$F^*\big(1 + \gamma h(\alpha)\big) + C^*\Big(2h(\alpha) - 1 + \tfrac{2}{\gamma}\Big) + 2\gamma\alpha M R^*(\alpha)h(\alpha) + 2\alpha M R^*(\alpha)$$

where $h(\alpha) = 1 + \frac{4}{\alpha} + \frac{4\alpha}{2\alpha-1} + 4\sqrt{\frac{6}{2\alpha-1}}$. *Thus, we can compute efficiently a solution to* $\mathcal{I}$ *of cost at most: (i)* $92.84 \cdot OPT$, *by setting* $\alpha = 0.75, \gamma = 3/h(\alpha)$*; (ii)* $82.6 \cdot OPT$, *by letting* γ *be a suitable function of* α, *and choosing* α *randomly from the interval* $[0.67, 1]$ *according to the density function* $p(x) = \frac{1}{\ln(1/0.67)x}$.

The roadmap for proving Theorem 1 is as follows. We first bound the cost of the bicriteria solution obtained in step (1) in terms of OPT (Lemma 3). This will allow us to bound the cost of an optimal solution to $\mathcal{I}_2$, and argue that mapping an $\mathcal{I}_2$-solution to a solution to $\mathcal{I}$ does not increase the cost by much (Lemma 4). The only missing ingredient is a guarantee on the cost of the solution to $\mathcal{I}_2$ found in step (3), which we supply in Theorem 5, whose proof appears in Section 4.

The following claim follows from essentially the same arguments as in [8, 5].

Claim 2. Let S' *be a* delete-optimal *solution to the above* UFL *instance; that is, the total* UFL*-cost does not decrease by deleting any open facility of* S'. *Then, each facility of* S' *serves at least* αM *clients.*

The local-search algorithms for UFL in [3, 2] have the same performance guarantees and both include a delete-move as a local-search operation, so upon termination, we obtain a delete-optimal solution.[1] Opening the same facilities and

[1] A subtle point is that typically local-search algorithms terminate only with an "approximate" local optimum. However, one can then execute all delete moves that improve the solution cost, and thereby obtain a delete-optimal solution.

making the same client assignments as in the optimal solution to $\mathcal{I}$ yields a solution S to the UFL instance constructed in step (1) of the algorithm with facility cost $F^S \leq F^* + 2\alpha M R^*(\alpha)$ and assignment cost $C^S \leq C^*$. Combined with the analysis in [3, 2], this yields the following. (For simplicity, we assume that local search terminates with a local optimum; standard arguments show that dropping this assumption increases the approximation by at most a $(1+\epsilon)$ factor.)

Lemma 3. *For a given parameter $\gamma > 0$, executing the local-search algorithm in [3, 2] on the above UFL instance returns a solution with facility cost F_b and assignment cost C_b satisfying $F_b \leq F^* + 2\alpha M R^*(\alpha) + 2C^*/\gamma,\ C_b \leq \gamma\big(F^* + 2\alpha M R^*(\alpha)\big) + C^*$, where each open facility serves at least αM clients.*

Lemma 4 ([17]). *(i) The cost $C^*_{\mathcal{I}_2}$ of an optimal solution to $\mathcal{I}_2$ is at most $2(C_b + C^*)$. (ii) Any solution to $\mathcal{I}_2$ of cost C yields a solution to $\mathcal{I}$ of cost at most $F_b + C_b + C$.*

Theorem 5. *For any $\alpha > 0.5$, there is a $g(\alpha)$-approximation algorithm for $\mathcal{I}_2(\alpha)$, where $g(\alpha) = \frac{2}{\alpha} + \frac{2\alpha}{2\alpha-1} + 2\sqrt{\frac{2}{\alpha^2} + \frac{4}{2\alpha-1}}$.*

Remark 6. Our $g(\alpha)$-approximation ratio for $\mathcal{I}_2(\alpha)$ improves upon the approximation obtained in [17] by a factor of roughly 2 *for all* α. Thus, plugging in our algorithm for solving $\mathcal{I}_2$ in the LBFL-algorithm in [17] (and choosing a suitable α), already yields an improved approximation factor of 218 for LBFL.

Proof of Theorem 1. Recall that $h(\alpha) = 1 + \frac{4}{\alpha} + \frac{4\alpha}{2\alpha-1} + 4\sqrt{\frac{6}{2\alpha-1}}$. Note that $2g(\alpha)+1 \leq h(\alpha)$ for all $\alpha \in [0,1]$; we use this upper bound throughout below. Combining Theorem 5 and the bounds in Lemmas 3 and 4, we obtain a solution to $\mathcal{I}$ of cost at most $F_b + \big(2g(\alpha)+1\big)C_b + 2g(\alpha)C^* \leq F_b + h(\alpha)C_b + \big(h(\alpha)-1\big)C^*$

$$\leq F^*\big(1+\gamma h(\alpha)\big) + C^*\Big(2h(\alpha) - 1 + \frac{2}{\gamma}\Big) + 2\gamma\alpha M R^*(\alpha)h(\alpha) + 2\alpha M R^*(\alpha).$$

Part (i) follows by plugging in the values of α and γ, and using the bound $R^*(\alpha) \leq \frac{C^*}{M(1-\alpha)}$. Let $\beta = 0.67$. For part (ii), we set $\gamma = \frac{K}{\sqrt{h(\alpha)}}$, where $K = \left(\ln^2(1/\beta) \cdot \mathrm{E}_\alpha\left[h(\alpha)\right] / \big(\frac{\int_\beta^1 h(x)dx}{1-\beta}\big)\right)^{\frac{1}{4}}$. Hence, the cost incurred is at most

$$F^*\big(1+K\sqrt{h(\alpha)}\big)+C^*\Big(2h(\alpha)-1+\frac{2}{K}\sqrt{h(\alpha)}\Big)+2K\alpha M R^*(\alpha)\sqrt{h(\alpha)}+2\alpha M R^*(\alpha).$$

We now bound the expected cost incurred when one chooses α randomly according to the stated density function. This will also yield an explicit expression for K (as a function of β), thus showing that K (and hence, γ) can be computed efficiently. We note that $\mathrm{E}\left[\sqrt{X}\right] \leq \sqrt{\mathrm{E}\left[X\right]}$ and utilize Chebyshev's Integral inequality (see [7]): if f and g are non-increasing and non-decreasing functions respectively from $[a,b]$ to $\mathbb{R}_+$, then $\int_a^b f(x)g(x)dx \leq \frac{(\int_a^b f(x)dx)(\int_a^b g(x)dx)}{b-a}$. Observe that $h(\alpha)$ decreases with α. Recall that $\beta = 0.67$. We have the following.

$$\mathrm{E}_\alpha\left[h(\alpha)\right] = c_2(\beta) := \tfrac{4}{\beta\ln(1/\beta)} - \tfrac{4}{\ln(1/\beta)} + \tfrac{8\sqrt{6}\left(\pi/4-\tan^{-1}(\sqrt{2\beta-1})\right)}{\ln(1/\beta)} + \tfrac{2\ln(1/(2\beta-1))}{\ln(1/\beta)} + 1$$

$$\mathrm{E}_\alpha\left[\alpha M R^*(\alpha)\right] \;=\; M\Big(\int_\beta^1 R^*(x)dx\Big)/\ln(1/\beta) \le C^*/\ln(1/\beta).$$

$$\mathrm{E}_\alpha\left[\alpha M R^*(\alpha)\sqrt{h(\alpha)}\right] \le \left[M\Big(\int_\beta^1 R^*(x)dx\Big)\tfrac{\int_\beta^1 dx\sqrt{h(x)}}{1-\beta}\right]/\ln(1/\beta) \le \frac{C^*\sqrt{c_3(\beta)}}{\ln(1/\beta)}, \quad \text{where}$$

$$c_3(\beta) := \frac{\int_\beta^1 h(x)dx}{1-\beta} = \left[4\ln\Big(\frac{1}{\beta}\Big) + 4\sqrt{6}\big(1-\sqrt{2\beta-1}\big) + 3(1-\beta) + \ln\Big(\frac{1}{2\beta-1}\Big)\right]/(1-\beta).$$

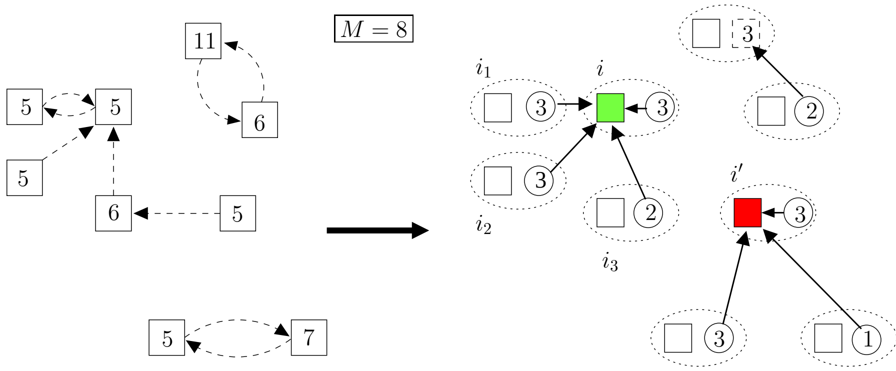
The second inequality follows since $\left(\int_\beta^1 dx\sqrt{h(x)}\right)/(1-\beta) = \mathrm{E}_{\alpha\sim\text{uniform in }[\beta,1]}\left[\sqrt{h(\alpha)}\right]$. These bounds yield $K = \left(\ln^2(1/\beta)c_2(\beta)/c_3(\beta)\right)^{0.25}$, and the total cost is at most

$$F^*\left(1+\big(\tfrac{\ln^2(1/\beta)(c_2(\beta))^3}{c_3(\beta)}\big)^{\frac{1}{4}}\right)+C^*\left(2c_2(\beta)-1+4\big(\tfrac{c_2(\beta)c_3(\beta)}{\ln^2(1/\beta)}\big)^{\frac{1}{4}}+\tfrac{2}{\ln(1/\beta)}\right) < 82.59(F^*+C^*). \quad \square$$

4 Solving instance $\mathcal{I}_2(\alpha)$

We now describe our algorithm for solving instance $\mathcal{I}_2(\alpha)$ and analyze its performance guarantee, thereby proving Theorem 5. As mentioned earlier, one of the key differences between our algorithm and the one in [17] is that instead of reducing $\mathcal{I}_2$ to capacitated facility location (CFL), we solve $\mathcal{I}_2$ by reducing it to a new problem that we call *capacity-discounted UFL* (CDUFL). CDUFL is a special case of CFL where all facilities with non-zero opening cost are uncapacitated (i.e., have infinite capacity). Perhaps surprisingly, despite this special structure, CDUFL inherits the intractability of CFL with respect to LP-based approximation guarantees: the natural LP-relaxation for CDUFL has bad integrality gap, and there is no known LP-relaxation with constant integrality gap. However, we show in Section 4.2 that a simple local-search algorithm for CDUFL yields a better approximation ratio than the current-best approximation for CFL.

Recall that $\mathcal{I}_2$ has only the points in $\mathcal{F}' \subseteq \mathcal{F}$, and there are $n_i \ge \alpha M$ co-located clients at each $i \in \mathcal{F}'$. Let $l(i) = \min_{i'\in\mathcal{F}', i'\neq i} c_{ii'}$. To avoid confusion, we refer to the facilities and clients in the CDUFL instance as supply points and demand points respectively. The CDUFL instance created to solve $\mathcal{I}_2$ resembles the CFL instance created in [17]; the difference is that supply points with non-zero opening costs are now uncapacitated. At each $i \in \mathcal{F}'$, we create an uncapacitated supply point with opening cost $\delta\min\{n_i, M\}l(i)$, where δ will be fixed later. If $n_i > M$ we create a second supply point at i with capacity $n_i - M$ and zero opening cost. If $n_i < M$, we create a demand point at i with demand $M - n_i$. Let $\mathcal{I}'$ denote this CDUFL instance (see Fig. 1). Let $\mathcal{F}^u$, $\mathcal{F}^c$ denote respectively the set of uncapacitated and capacitated supply points of $\mathcal{I}'$. Roughly speaking, satisfying a demand point i by non-co-located supply points translates to leaving facility i open in the $\mathcal{I}_2$ solution; hence, its demand is set to $M - n_i$, which is the number of additional clients it needs. Conversely, opening the uncapacitated supply point at i and supplying demand points from i translates to closing i in the $\mathcal{I}_2$ solution and transferring its co-located clients to other open facilities.



(a) $\mathcal{I}_2$ (b) $\mathcal{I}'$, and a solution S for $\mathcal{I}'$

Fig. 1. (a) An $\mathcal{I}_2$ instance. Each box denotes a facility and the number inside it is the number of co-located clients; $i \dashrightarrow i'$ indicates that i' is the closest facility to i.
(b) The corresponding $\mathcal{I}'$ instance. The boxes and circles represent supply points and demand points respectively, and points inside a dotted oval are co-located. A solid box denotes an uncapacitated supply point, and a dashed box denotes a capacitated facility whose capacity is shown inside the box. The number inside a circle is the demand of that demand point. The arrows indicate a solution S to $\mathcal{I}'$, where i and i' are the two open uncapacitated supply points.

Lemma 7 ([17]). *There exists a solution to $\mathcal{I}'$ with facility cost $F \leq \delta C^*_{\mathcal{I}_2}$ and assignment cost $C \leq C^*_{\mathcal{I}_2}$.*

Theorem 8. *(i) Given any CDUFL instance, one can efficiently compute a solution with facility-opening cost $\widehat{F} \leq F^{\text{sol}} + 2C^{\text{sol}}$ and assignment cost $\widehat{C} \leq F^{\text{sol}} + C^{\text{sol}}$, where F^{sol} and C^{sol} are the facility and assignment costs of an arbitrary solution to the CDUFL instance.*
(ii) Thus, Lemma 7 implies that one can compute a solution to $\mathcal{I}'$ with facility cost $F_{\mathcal{I}'}$ and assignment cost $C_{\mathcal{I}'}$ satisfying $F_{\mathcal{I}'} \leq (2+\delta)C^*_{\mathcal{I}_2}$, $C_{\mathcal{I}'} \leq (1+\delta)C^*_{\mathcal{I}_2}$.

4.1 Mapping an $\mathcal{I}'$-solution to an $\mathcal{I}_2$-solution

An $\mathcal{I}'$-solution need not directly translate to an $\mathcal{I}_2$ solution because an open supply point i may not supply (and hence, transfer) exactly n_i units of demand (e.g., i and i' in Fig. 1(b)). Since we have uncapacitated supply points, we have to consider both the cases where i supplies more than n_i demand (which is not encountered in [17]), and less than n_i demand. Suppose that we are given a solution S to $\mathcal{I}'$ with facility cost F^S and assignment cost C^S (see Fig. 1(b)). Again, we abuse notation and use F^S to also denote the set of supply points that are opened in S. Let N_i initialized to n_i keep track of the number of clients at location $i \in \mathcal{F}'$. Our goal is to reassign clients (using S as a template) so that at the end we have $N_i = 0$ or $N_i \geq M$ for each $i \in \mathcal{F}'$. We may assume that: (i) $\mathcal{F}^c \subseteq F^S$; (ii) if S opens an uncapacitated supply point located at some $i \in \mathcal{F}'$ with $n_i > M$, then the demand assigned to the capacitated supply point at i equals its capacity $n_i - M$; (iii) for each $i \in \mathcal{F}'$ with $n_i \leq M$, if the supply point at i is open then it serves the entire demand of the co-located demand point; and

(iv) at most one *uncapacitated* supply point serves, maybe partially, the demand of any demand point; we say that this uncapacitated supply point satisfies the demand point. We reassign clients in three phases.

A1. **Removing capacitated supply points.** Consider any $i \in \mathcal{F}'$ with $n_i > M$. Let i^1 and i^2 denote respectively the capacitated and uncapacitated supply points located at i. If i^1 supplies x units to the demand point at location i', we transfer x clients from location i to i'. Now if i^1 has $y > 0$ leftover units of capacity in S, then we "move" y clients to i^2 (which is not open in S). We update the N_is accordingly. This reassignment effectively gets rid of all capacitated supply points. Thus, there is now exactly one uncapacitated supply point and at most one demand point at each location $i \in \mathcal{F}'$; we refer to these simply as supply point i and demand point i below.

Let X_i be the total demand from other locations assigned to supply point i. Let $\mathcal{F}^G = \{i \in \mathcal{F}' : N_i < X_i\}$, $\mathcal{F}^R = \{i \in \mathcal{F}' : N_i \geq X_i > 0\}$, and $\mathcal{F}^B = \{i \in \mathcal{F}' : X_i = 0\}$ ($\mathcal{F}^B$ is the set of supply points not opened in S). Note that $N_i \geq \min\{n_i, M\} \geq \alpha M$ for all $i \in \mathcal{F}'$, and $N_i = \min\{n_i, M\}$ for all $i \in \mathcal{F}^R \cup \mathcal{F}^G$.

A2. **Taking care of $\mathcal{F}^R$ and demand points satisfied by $\mathcal{F}^R$.** For each $i \in \mathcal{F}^R$, if i supplies x units to demand point i', we move x clients from i to i', and update $N_i, N_{i'}$. We now have $N_i = \min\{n_i, M\} - X_i$ residual clients at each $i \in \mathcal{F}^R$, which we must reduce to 0, or increase to at least M.

We follow the same procedure as in [17]. For each $i \in \mathcal{F}^R$, we include an edge (i, i') where $i' \in \mathcal{F}'$ is the facility nearest to i (recall that $c_{ii'} = l(i)$). We use an arbitrary but fixed tie-breaking rule here, so each component of the resulting digraph is a directed tree rooted at either (i) a node $r \in \mathcal{F}' \setminus \mathcal{F}^R$, or (ii) a 2-cycle $(r, r'), (r', r)$, where $r, r' \in \mathcal{F}^R$. We break up each component Γ into a collection of smaller components. Essentially, we move the residual clients of supply points in Γ bottom-up from the leaves up to the root, cut off Γ at the first node u that accumulates at least M clients, and recurse on the portion of Γ not containing u. More precisely, let Γ_u denote the subtree of Γ rooted at node $u \in \Gamma$ (if u belongs to a 2-cycle then we do not include the other node of this 2-cycle in Γ_u). If $\sum_{i \in \Gamma} N_i < M$, or if Γ is of type (i) and all children u of the root satisfy $\sum_{i \in \Gamma_u} N_i < M$, we leave Γ unchanged. Otherwise, let u be a deepest (i.e., furthest from root) node in Γ such that $\sum_{i \in \Gamma_u} N_i \geq M$. We delete the arc leaving u. If this disconnects u from $\Gamma \setminus \Gamma_u$, then we recurse on $\Gamma \setminus \Gamma_u$. Otherwise u must belong to the root 2-cycle of Γ. Let r' be the other node of this 2-cycle. If $\sum_{i \in \Gamma_{r'}} N_i \geq M$, we delete r''s outgoing arc (thus splitting Γ into Γ_u and $\Gamma_{r'}$).

After applying the above procedure (to all components), if we are left with a component of type (ii) with $\sum_{i \in \text{ component}} N_i \geq M$, we convert it to type (i) by arbitrarily deleting one of the arcs of the 2-cycle. Let T be a component at the end of this process. If T rooted at a node r, we move the N_i residual clients of each non-root node $i \in T$ to r. Otherwise, T is of type (ii) with root $\{r, r'\}$, and we have $\sum_{i \in T} N_i < M$. Let $i' \in \mathcal{F}^B$ be the location nearest to $\{r, r'\}$; we move the N_i residual clients of each $i \in T$ to i'. Update the N_is to reflect the above reassignment. Observe that we now have $N_i = 0$ or

$N_i \geq M$ for each $i \in \mathcal{F}^R$, and each $i \in \mathcal{F}^B$ has $n_i \geq M$, or is a demand point satisfied by a supply point in $\mathcal{F}^G$.

For example, executing step (A1 and) A2 on the solution shown in Fig. 1(b) results in $i' \in \mathcal{F}^R$ having one client left after moving its co-located clients to the bottom two facilities; this residual client is then transferred to i_3.

A3. **Taking care of $\mathcal{F}^G$ and demand points satisfied by $\mathcal{F}^G$.** For $i \in \mathcal{F}^G$, let $D(i)$ be the set of demand points $j \in \mathcal{F}'$, $j \neq i$ satisfied by i, and let $D'(i) = \{j \in D(i) : N_j < M\}$. Note that $D(i) \subseteq \mathcal{F}^B$. Phase A2 may only increase N_j for all j in $\mathcal{F}^B \cup \mathcal{F}^G$, so $N_j \geq \alpha M$ for all $j \in \mathcal{F}^G \cup \big(\bigcup_{i \in \mathcal{F}^G} D(i)\big)$. Fix $i \in \mathcal{F}^G$. We reassign clients so that $N_j = 0$ or $N_j \geq M$ for all $j \in \{i\} \cup D'(i)$, without decreasing N_j for $j \in D(i) \setminus D'(i)$. Doing this for all supply points in $\mathcal{F}^G$ will complete our task. Define $Y_j = M - N_j$ for $j \in D'(i)$. (1) If $\sum_{j \in D'(i)} Y_j \leq N_i$, for each $j \in D'(i)$, if i supplies x units to j, we transfer x clients from i to j. If i is now left with less than M residual clients, we move these residual clients to the location in $D(i)$ nearest to i. (2) If $\sum_{j \in D'(i)} Y_j > N_i$, set $i_0 = i$, and $D'(i) = \{i_1, \ldots, i_t\}$, where $c_{i_1 i} \leq \ldots \leq c_{i_t i}$. Let $\ell = t - \left\lfloor \frac{\sum_{r=0}^{t} N_{i_r}}{M} \right\rfloor = \left\lceil \frac{\sum_{r=1}^{t} Y_{i_r} - N_{i_0}}{M} \right\rceil$ (so $1 \leq \ell < t$ since $N_{i_0} + N_{i_1} \geq M$), which is the unique index such that $\sum_{r=\ell+1}^{t} Y_{i_r} \leq \sum_{r=0}^{\ell} N_{i_r} < \sum_{r=\ell+1}^{t} Y_{i_r} + M$. This enables us to transfer Y_{i_q} clients to each i_q, $q = \ell+1, \ldots, t$ from the locations $i_\ell, \ldots, i_0$—we do this by transferring all clients of i_r (where $1 \leq r \leq \ell$) before considering i_{r-1}—and be left with at most M residual clients in $\{i_0, \ldots, i_\ell\}$. We argue that these residual clients are all concentrated at i_0 and i_1, with i_1 having at most $(1-\alpha)M$ residual clients. We transfer these residual clients to $i_{\ell+1}$.

In the solution shown in Fig. 1(b), we have $Y_{i_1} = 3 = Y_{i_2},\ Y_{i_1} = 1,\ N_i = 5$, so case 2 applies; we transfer 1 client to i_3 and 9 clients to i_2 from $\{i, i_1\}$.

Theorem 9. *The above algorithm returns an $\mathcal{I}_2$-solution of cost at most $\frac{F^S}{\delta\alpha} + C^S\big(\frac{1}{\alpha} + \frac{2\alpha}{2\alpha-1}\big)$. Thus, taking S to be the solution mentioned in part (ii) of Theorem 8, and $\delta = \sqrt{\frac{2/\alpha}{1/\alpha + (2\alpha)/(2\alpha-1)}}$, we obtain a solution to $\mathcal{I}_2(\alpha)$ satisfying the approximation bound stated in Theorem 5.*

Proof. Let S_2 denote the solution computed for $\mathcal{I}_2$. For a supply point i opened in S, let C_i^S denote the cost incurred in supplying demand from i to the demand points satisfied by i. At various steps, we transfer clients between locations according to the assignment in the CDUFL solution S, and the cost incurred in doing so can be charged to the C_i^Ss of the appropriate supply points. So the cost of phase A1 is $\sum_{i \in \mathcal{F}^c} C_i^S$, and the cost of the first step of phase A2 is $\sum_{i \in \mathcal{F}^R} C_i^S$. As in [17], we can bound the remaining cost of phase A2, incurred in transferring clients according to the tree edges, by $F^S/\delta\alpha + \big(\sum_{i \in \mathcal{F}^R} C_i^S\big)/(2\alpha - 1)$.

Finally, consider phase A3 and some $i \in \mathcal{F}^G$. If $\sum_{j \in D'(i)} Y_j \leq N_i$, then the cost incurred is at most $C_i^S + M \cdot \frac{C_i^S}{X_i} \leq C_i^S\big(1 + \frac{1}{\alpha}\big)$ (as $X_i > N_i \geq \alpha M$). Now consider the case $\sum_{j \in D'(i)} Y_j > N_i$. For any $i_q \in \{i_{\ell+1}, \ldots, i_t\}$ and any $i_r \in \{i_0, \ldots, i_\ell\}$, we have $c_{i_r i_q} \leq 2c_{i i_q}$, so the cost of transferring $Y_{i_q} \leq M - n_{i_q}$ clients to each i_q, $q = \ell+1, \dots, t$ is at most $2C_i^S$. Observe that $(t-\ell+1)M > \sum_{r=0}^{t} N_{i_r}$, i.e., $M + \sum_{q=\ell+1}^{t} Y_{i_r} > \sum_{r=0}^{\ell} N_{i_r}$, so after this reassignment, there are less than M residual clients in $i_0, \dots, i_\ell$. By our order of transferring clients, all these residual clients are at i_0, i_1 (otherwise we would have at least $N_{i_0} + N_{i_1} \geq M$ residual clients) with at most $M - N_{i_0} \leq (1-\alpha)M$ of them located at i_1. The cost of reassigning these residual clients is at most $(1-\alpha)Mc_{ii_1} + Mc_{ii_{\ell+1}} \leq (1-\alpha)M \cdot \frac{C_i^S}{\sum_{r=1}^{t} Y_{i_r}} + M \cdot \frac{C_i^S}{\sum_{r=\ell+1}^{t} Y_{i_r}}$, since C_i^S is the total cost of supplying at least Y_{i_r} demand to each i_r, $r = 1, \dots, t$. The latter expression is at most $C_i^S\big(\frac{1-\alpha}{\alpha} + \frac{1}{2\alpha-1}\big)$, since $\sum_{r=1}^{t} Y_{i_r} > N_{i_0} \geq \alpha M$, $\sum_{r=\ell+1}^{t} Y_{i_r} > \sum_{r=0}^{\ell} N_{i_r} - M \geq (2\alpha-1)M$.) Thus, the cost of S_2 is at most $\frac{F^S}{\delta\alpha} + \sum_{i \in \mathcal{F}^c} C_i^S + \sum_{i \in \mathcal{F}^R} C_i^S \cdot \big(1 + \frac{1}{2\alpha-1}\big) + \sum_{i \in \mathcal{F}^G} C_i^S \cdot \max\big\{1 + \frac{1}{\alpha}, 2 + \frac{1-\alpha}{\alpha} + \frac{1}{2\alpha-1}\big\} \leq \frac{F^S}{\delta\alpha} + C^S\big(\frac{1}{\alpha} + \frac{2\alpha}{2\alpha-1}\big)$. So if S is the solution given by part (ii) of Theorem 8, the cost of S_2 is at most $\big(\frac{2}{\delta\alpha} + \frac{1}{\alpha} + (1+\delta)(\frac{1}{\alpha} + \frac{2\alpha}{2\alpha-1})\big)C^*_{\mathcal{I}_2}$, and plugging in the value of δ yields the $g(\alpha) = \frac{2}{\alpha} + \frac{2\alpha}{2\alpha-1} + 2\sqrt{\frac{2}{\alpha^2} + \frac{4}{2\alpha-1}}$ approximation stated in Theorem 5. □

4.2 A local-search based approximation algorithm for CDUFL

We now describe our local-search algorithm for CDUFL, which leads to the proof of Theorem 8. Let $\widehat{\mathcal{F}} = \widehat{\mathcal{F}}^u \cup \widehat{\mathcal{F}}^c$ be the facility-set of the CDUFL instance, where $\widehat{\mathcal{F}}^u \cap \widehat{\mathcal{F}}^c = \emptyset$. Here, $\widehat{\mathcal{F}}^u$ are the uncapacitated facilities with opening costs $\{\widehat{f}_i\}$, and facilities in $\widehat{\mathcal{F}}^c$ have (finite) capacities $\{u_i\}$ and zero opening costs. Let $\widehat{\mathcal{D}}$ be the set of clients and $\widehat{c}_{ij}$ be the cost of assigning client j to facility i. The goal is to open facilities and assign clients to open facilities (respecting the capacities) so as to minimize the sum of the facility-opening and client-assignment costs. We can find the best assignment of clients to open facilities by solving a network flow problem, so we focus on determining the set of facilities to open.

The local-search algorithm consists of add(i'), delete(i), swap(i, i') moves, which respectively, add a facility i' not currently open, delete a facility i that is currently open, and swap facility i that is open with facility i' that is not open. We note that *all* previous (local-search) algorithms for CFL with non-uniform capacities use moves that are more complicated than the moves above. The algorithm repeatedly executes the best cost-improving move until no such move exists. We may assume without loss of generality that each client has unit demand.

Analysis. Let $\widehat{S}$ denote a local-optimum returned by the algorithm, with facility-opening cost $\widehat{F}$ and assignment cost $\widehat{C}$. Let sol be an arbitrary CDUFL solution, with facility-cost F^{sol} and assignment cost C^{sol}. We also use $\widehat{F}$ and F^{sol} to denote the set of open facilities in $\widehat{S}$ and sol respectively. We may assume that $\widehat{\mathcal{F}}^c \subseteq \widehat{F} \cap F^{\text{sol}}$. For a facility i, we use $\widehat{\mathcal{D}}_{\widehat{S}}(i)$ and $\widehat{\mathcal{D}}_{\text{sol}}(i)$ to denote respectively the (possibly empty) set of clients served by i in $\widehat{S}$ and sol. For a client j, let $\widehat{C}_j$ and C_j^{sol} be the assignment cost of j in $\widehat{S}$ and sol respectively.

We borrow ideas from the analysis of the corresponding local-search algorithm for UFL in [2], but to handle capacities we need to reassign clients more carefully to analyze the change in assignment cost due to a local-search move. In particular, unlike the analysis in [2], where upon deletion of a facility $s \in \widehat{F}$ we reassign only the clients currently assigned to s, in our case (as in the analysis of local-search algorithms for CFL), we need to perform a more "global" reassignment (i.e., even clients not assigned to s may get reassigned) along certain paths in a suitable graph. This also means that we need to construct a suitable mapping between paths instead of the client-mapping considered in [2].

Consider a directed graph G with node-set $\widehat{\mathcal{D}} \cup \widehat{\mathcal{F}}$, and arcs from i to all clients in $\widehat{\mathcal{D}}_{\widehat{S}}(i)$ and arcs from all clients in $\widehat{\mathcal{D}}_{\text{sol}}(i)$ to i, for every facility i. Via standard flow-decomposition, we can decompose G into a collection of (simple) paths $\mathcal{P}$, and cycles $\mathcal{R}$, so that (i) each facility i appears as the starting point of $\max\{0, |\widehat{\mathcal{D}}_{\widehat{S}}(i)| - |\widehat{\mathcal{D}}_{\text{sol}}(i)|\}$ paths, and the ending point of $\max\{0, |\widehat{\mathcal{D}}_{\text{sol}}(i)| - |\widehat{\mathcal{D}}_{\widehat{S}}(i)|\}$ paths, and (ii) each client j appears on a unique path P_j or on a cycle. Let $\mathcal{P}^{\mathsf{st}}(s) \subseteq \mathcal{P}$ and $\mathcal{P}^{\mathsf{end}}(o) \subseteq \mathcal{P}$ denote respectively the collection of paths starting at s and ending at o, and $\mathcal{P}(s,o) = \mathcal{P}^{\mathsf{st}}(s) \cap \mathcal{P}^{\mathsf{end}}(o)$. For a path $P = \{i_0, j_0, i_1, j_1, \ldots, i_k, j_k, i_{k+1} := o\} \in \mathcal{P}$, define $\widehat{\mathcal{D}}(P) = \{j_0, \ldots, j_k\}$, $\mathit{head}(P) = j_0$, and $\mathit{tail}(P) = j_k$. A *shift* along P means that we reassign client j_r to i_{r+1} for each $r = 0, \ldots, k$ (opening o if necessary). Note that this is feasible, since if $o \in \widehat{\mathcal{F}}^c$, we know that $|\widehat{\mathcal{D}}_{\widehat{S}}(o)| \leq |\widehat{\mathcal{D}}_{\text{sol}}(o)| - 1 \leq u_o - 1$. Let $\mathit{shift}(P) := \sum_{j \in \widehat{\mathcal{D}}(P)} \big(C_j^{\text{sol}} - \widehat{C}_j\big)$ be the increase in assignment cost due to this reassignment, which is an upper bound on the actual increase in assignment cost if o is added to $\widehat{F}$. Let $\mathit{cost}(P) := \sum_{j \in \widehat{\mathcal{D}}(P)} \big(C_j^{\text{sol}} + \widehat{C}_j\big)$. We define a shift along a cycle $R \in \mathcal{R}$ similarly, letting $\mathit{shift}(R) := \sum_{j \in \widehat{\mathcal{D}} \cap R} \big(C_j^{\text{sol}} - \widehat{C}_j\big)$. By considering a shift operation for every path and cycle in $\mathcal{P} \cup \mathcal{R}$ (i.e., suitable add moves) and adding the resulting inequalities, we get the following result.

Lemma 10. $\widehat{C} \leq F^{\text{sol}} + C^{\text{sol}}$.

To bound $\widehat{F}$, we only need paths starting at facilities in $\widehat{F} \setminus F^{\text{sol}}$. Note that facilities in $(\widehat{F} \setminus F^{\text{sol}}) \cup (F^{\text{sol}} \setminus \widehat{F})$ are *uncapacitated*. To avoid excessive notation, for a facility $o \in F^{\text{sol}} \setminus \widehat{F}$, we now use $\mathcal{P}^{\mathsf{end}}(o)$ to refer to the collection of paths ending in o that start in $\widehat{F} \setminus F^{\text{sol}}$. (As before, $\mathcal{P}(s,o)$ is the set of paths that start at s and end at o.) Let $\text{capt}_s \subseteq F^{\text{sol}} \setminus \widehat{F}$ be the facilities captured by s. For any $o \in F^{\text{sol}} \setminus \widehat{F}$, we can obtain a 1-1 mapping $\pi : \mathcal{P}^{\mathsf{end}}(o) \mapsto \mathcal{P}^{\mathsf{end}}(o)$ such that if $P \in \mathcal{P}(s,o)$, $\pi(P) = P' \in \mathcal{P}(s',o)$ then (i) if $o \notin \text{capt}_s$, we have $s \neq s'$; (ii) if $s = s'$, then $P = P'$; and (iii) $\pi(P') = P$. Say that $o \in F^{\text{sol}} \setminus \widehat{F}$ is *captured* by s if $|\mathcal{P}(s,o)| > \frac{|\mathcal{P}^{\mathsf{end}}(o)|}{2}$. Call a facility in $\widehat{F} \setminus F^{\text{sol}}$ *good* if $\text{capt}_s = \emptyset$, and *bad* otherwise. For a bad facility s, let $o_s \in \text{capt}_s$ be the facility nearest to s.

Lemma 11. *Let s be a facility in $\widehat{F} \setminus F^{\text{sol}}$.*

If s is good,
$$\widehat{f}_s \leq \sum_{P \in \mathcal{P}^{\mathsf{st}}(s)} \mathit{shift}(P) + \sum_{o \notin \widehat{F}, P \in \mathcal{P}(s,o)} \mathit{cost}\big(\pi(P)\big). \tag{1}$$

If s is bad,
$$\widehat{f}_s \leq \sum_{o \in \text{capt}_s} \widehat{f}_o + \sum_{P \in \mathcal{P}^{\mathsf{st}}(s)} \mathit{shift}(P) + \sum_{\substack{o \notin \widehat{F}, P \in \mathcal{P}(s,o): \\ \pi(P) \neq P}} \mathit{cost}\big(\pi(P)\big) + \sum_{\substack{o \in \text{capt}_s \setminus \{o_s\} \\ P \in \mathcal{P}(s,o): \pi(P) = P}} \mathit{cost}(P). \tag{2}$$

Proof Sketch of Theorem 8. We focus on part (i); part (ii) follows directly from part (i) and Lemma 7. Lemma 10 bounds $\widehat{C}$. Consider adding (1) for all good facilities and (2) for all bad facilities, and the vacuous equality $\widehat{f}_i = \widehat{f}_i$ for all $i \in \widehat{F} \cap F^{\mathrm{sol}}$. The LHS of the resulting inequality is precisely $\widehat{F}$. The $\widehat{f}_i$s on the RHS add up to give at most F^{sol}. One can argue that each path $P \in \bigcup_{s \in \widehat{F} \setminus F^{\mathrm{sol}}} \mathcal{P}^{\mathsf{st}}(s)$ contributes at most $\mathit{shift}(P) + \mathit{cost}(P) = 2\sum_{j \in \widehat{\mathcal{D}}(P)} C_j^{\mathrm{sol}}$ to the RHS. Thus the RHS is at most $F^{\mathrm{sol}} + 2C^{\mathrm{sol}}$, and we obtain that $\widehat{F} \leq F^{\mathrm{sol}} + 2C^{\mathrm{sol}}$. □

References

1. S. Ahmadian and C. Swamy. Improved approximation guarantees for lower-bounded facility location. *CS arXiV*, September 2012.
2. V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristics for k-median and facility location problems. *SIAM Journal on Computing*, 33(3):544–562, 2004.
3. M. Charikar and S. Guha. Improved combinatorial algorithms for facility location problems. *SIAM Journal on Computing*, 34(4):803–824, 2005.
4. S. Guha, A. Meyerson, and K. Munagala. Facility location with demand dependent costs and generalized clustering. *Manuscript*, 2000.
5. S. Guha, A. Meyerson, and K. Munagala. Hierarchical placement and network design problems. In *Proceedings of the 41st FOCS*, pages 603–612, 2000.
6. M. Hajiaghayi, M. Mahdian, and V. Mirrokni. The facility location problem with general cost functions. *Networks*, 42:42–47, 2003.
7. G. Hardy, J. Littlewood, and G. Pólya. *Inequalities*. Cambridge Univ. Press, 1952.
8. D. R. Karger and M. Minkoff. Building Steiner trees with incomplete global knowledge. In *Proceedings of the 41st FOCS*, pages 613–623, 2000.
9. M. R. Korupolu, C. G. Plaxton, and R. Rajaraman. Analysis of a local search heuristic for facility location problems. *Journal of Algorithms*, 37(1):146–188, 2000.
10. S. Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. In *Proceedings of the 38th ICALP*, pages 77–88, 2011.
11. A. Lim, F. Wang, and Z. Xu. A transportation problem with minimum quantity commitment. *Transportation Science*, 40(1):117–129, 2006.
12. M. Mahdian and M. Pál. Universal facility location. In *Proceedings of the 11th ESA*, pages 409–421, 2003.
13. P. Mirchandani and R. Francis, editors. *Discrete Location Theory*. John Wiley and Sons, Inc., New York, 1990.
14. D. B. Shmoys. The design and analysis of approximation algorithms: facility location as a case study. In S. Hosten, J. Lee, and R. Thomas, eds. *Trends in Optimization, AMS Proc. of Symposia in Applied Math. 61*, pages 85–97, 2004.
15. D. B. Shmoys, É. Tardos, and K. I. Aardal. Approximation algorithms for facility location problems. In *Proceedings of the 29th STOC*, pages 265–274, 1997.
16. M. Sviridenko. An improved approximation algorithm for the metric uncapacitated facility location problem. In *Proceedings of 9th IPCO*, pages 240–257, 2002.
17. Z. Svitkina. Lower-bounded facility location. *Trans. on Algorithms*, 6(4), 2010.
18. J. Zhang, B. Chen, and Y. Ye. A multi-exchange local search algorithm for the capacitated facility location problem. *Math. of Oper. Research*, 30:389–403, 2005.