Covering Clients with Types and Budgets

Dimitris Fotakis
Yahoo Research-New York, USA & National Technical University of Athens, Greece
dfotakis@oath.com

Laurent Gourvès
Université Paris-Dauphine, PSL University, CNRS, LAMSADE, 75016 Paris, France
laurent.gourves@dauphine.fr

Claire Mathieu
CNRS, France
clairemmathieu@gmail.com

Abhinav Srivastav
ENS Paris & Université Paris-Dauphine, France
abhinavsriva@gmail.com

Abstract

In this paper, we consider a variant of the facility location problem. Imagine the scenario where facilities are categorized into multiple types such as schools, hospitals, post offices, etc. and the cost of connecting a client to a facility is realized by the distance between them. Each client has a total budget on the distance she/he is willing to travel. The goal is to open the minimum number of facilities such that the aggregate distance of each client to multiple types is within her/his budget. This problem closely resembles to the SET COVER and r-DOMINATION problems. Here, we study this problem in different settings. Specifically, we present some positive and negative results in the general setting, where no assumption is made on the distance values. Then we show that better results can be achieved when clients and facilities lie in a metric space.

2012 ACM Subject Classification Theory of computation → Packing and covering problems, Theory of computation → Facility location and clustering

Keywords and phrases Facility Location, Geometric Set Cover, Local Search

Digital Object Identifier 10.4230/LIPIcs.ISAAC.2018.73

Funding PSL Project MULTIFAC

1 Introduction

Consider the problem of opening a set of facilities, such as public service centres, in a city such that all clients (people living in the city) are within a pre-specified distance of some facility. The objective here is to open the minimum number of facilities. This problem closely resembles the r-DOMINATING SET problem where given a metric space (V, d) and a distance threshold r, the goal is to find a minimum-size set M of points such that every point in V is within distance r to some point in M. This is a special case of the classical set cover problem and can be approximated within a factor of $1 + \ln |V|$. In the Euclidean plane, however, a polynomial time approximation scheme follows from the results on geometric covering problems of Hochbaum and Maass [17].

In this paper, we study the generalization of the r-DOMINATING SET problem with different types of covering points. Consider the setting where facilities can be categorized into multiple types, such as schools, hospitals, post offices, etc., and the cost of connecting a client to a facility can be realized by the distance between them. Each client has a total budget on the distance he/she is willing travel. As in SET COVER and r-DOMINATING SET problems, the goal is to open the minimum number of facilities so that the aggregate distance of each client to the nearest facilities of all types is within his/her budget. Intuitively, each client is willing to accept tradeoffs among his/her distance to different facility types. Facility location with multiple types has been previously studied in [15, 3]. Hajiaghayi et al. [15] considered a variant of k-MEDIAN problem with two facility sets (red and blue), where we can open at most k_r red and k_b blue facilities. As opposed to our problem, each client is assigned to a single nearest facility that can be either red or blue. The goal is to minimize the total distance of the clients to their facility.

Problem Definition. We are given a set $\mathcal{F}$ of m facilities that are partitioned into L types $F_1, F_2, \ldots, F_L$, and a set $\mathcal{C}$ of n clients, each with a budget B_j for $j \in \{1, \ldots, n\}$. We assume that L is a constant and that $m = \mathcal{O}(n^c)$, for some fixed constant c. Moreover, we are given a distance matrix $\mathcal{D}$ of size $|\mathcal{F}| \times |\mathcal{C}|$, where each element d_{ij} represents the distance between facility i and client j. We say that a client j is *served* or *covered* by a type-ℓ facility i, if i is the nearest open facility of type-ℓ to j. Furthermore, we say that j has a *service cost* or *covering cost* of d_{ij} for facilities of type-ℓ. The *total service (or covering) cost* of j is the sum of j's service costs over all types. Our goal is to compute a set S of facilities of minimum cardinality such that each client j is served by one open facility of each type in S and the total service cost of j is at most B_j. We refer to this problem as FLT that is, *Facility Location with Types*.

In this paper, we present bi-criteria approximations of FLT problem in different settings. Let OPT denote the number of facilities opened by a fixed optimal solution. We say that a solution $S' \subseteq \mathcal{F}$ is (α, β)-approximate iff the number of facilities opened in S' is at most αOPT and the total service cost of each client j with respect to S' is at most βB_j. As usual, an algorithm $\mathcal{A}$ is (α, β)-approximate if it outputs an (α, β)-approximate solution for every instance.

Related Work. The FLT problem with $L = 1$ corresponds to the SET COVER problem, where given a set of elements $\mathcal{U}$ and a collection $\mathcal{S}$ of subsets of $\mathcal{U}$, the aim is to choose a minimum number of sets in $\mathcal{S}$ such that every element in $\mathcal{U}$ is covered. The analogy with FLT is straightforward: $\mathcal{U}$ and $\mathcal{S}$ correspond to the set of clients $\mathcal{C}$ and the set of the facilities $\mathcal{F}$, respectively such that a client j is contained in the set corresponding to facility i if $d_{ij} \leq B_j$. It is known that the SET COVER problem admits $\mathbb{H}_n$ and f approximation algorithms where $\mathbb{H}_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n} \leq (1 + \ln n)$ and f is the maximum frequency of any element in $\mathcal{U}$.

Dinur and Steurer [9] proved that it is NP-hard to approximate the SET COVER problem within a ratio of $(1 - \epsilon) \ln n$, for any $\epsilon > 0$. Another problem equivalent to the SET COVER problem is the HITTING-SET problem, where given a set of element $\mathcal{U}$ and a collection $\mathcal{S}$ of subsets of $\mathcal{U}$, the aim is to choose the minimal set of elements P in $\mathcal{U}$ such that $P \cap S \neq \emptyset$, for all $S \in \mathcal{S}$. In a general setting, all results for the SET COVER problem extend to the HITTING-SET problem.

Surprisingly, the HITTING-SET problem admits a better approximation ratio in $\mathbb{R}^2$ (also called GEOMETRIC HITTING-SET or GHS). Mustafa and Ray [20] showed that a simple local search algorithm is a PTAS for the problem where elements in $\mathcal{U}$ and subsets in $\mathcal{S}$ correspond to points and pseudo-disks, respectively, in $\mathbb{R}^2$. As such, there are no fully polynomial approximation scheme for this problem unless $\mathcal{NP} = \mathcal{P}$ [14]. The FLT problem in $\mathbb{R}^2$ is closely related to the problem of covering a set of points with ellipses. For a special case of this problem where the ellipses are axis-parallel, Efrat et al. [11] presented an $\mathcal{O}(n^* \log n^*)$ approximation, where n^* is the size of an optimal cover.

Another problem related to FLT problem is the RED-BLUE SET COVER problem [8]. Here, elements in $\mathcal{U}$ are partitioned into two sets: RED set $R = \{r_1, r_2, \ldots, r_k\}$ and BLUE set $B = \{b_1, b_2, \ldots, b_l\}$. The objective is to find a collection of subsets in $\mathcal{S}$ such that all BLUE elements are covered and the number of RED elements covered is minimized. Carr et al. [8] showed that RED-BLUE SET COVER problem cannot be approximated within a factor of $\mathcal{O}(2^{\log^{1-\epsilon} n'})$ for any $\epsilon > 0$, where $n' = |\mathcal{S}|^4$ (also see [12] for a similar inapproximability result). Further, Carr et al. [8] showed that RED-BLUE SET COVER admits an $\mathcal{O}((c\rho)^{1-1/c} \log \rho)$-approximation, where $\rho = |\mathcal{S}|$ and $c \geq |S \cap R|$ for all $S \in \mathcal{S}$.

FLT requires that every client is covered by L facilities, which is reminiscent of the SET MULTI-COVER problem [4] and the *fault-tolerant facility location* problem (FT-FL in short) [18, 21]. Every element (resp. client) has a demand which is a lower bound on the number of sets where the element must appear (resp. a lower bound on the number of facilities to which the client is assigned). However, unlike FLT, the sets (resp. facilities) are not categorized and a coverage with one set (resp. facility) of each category is not imposed. FLT also bears some remote resemblance to *multilevel facility location* problems, where facilities are partitioned into k levels and each client must travel to a facility at level k through a path that goes through one facility at each level $1, \ldots, k$ (see e.g., [1, 6] and the references therein). Unlike FLT, in multilevel facility location, the clients move from lower to higher levels and there is no budget on the total length of the path.

The literature contains various aggregate functions for capturing the distance between a client and its L covering facilities: maximum distance, sum of the distances, or more generally with the use of an *ordered weighted average* [22]. In this article we consider the sum, like for FT-FL. However, the clients' total covering costs are part of the objective function in FT-FL, whereas they are treated as constraints in FLT, i.e. client j's total service cost should not exceed a prescribed budget B_j.

1.1 Our Contribution

To the best of our knowledge, the approximability of covering problems with multiple types and a constraint on the combined "quality" of each client's covering has not been studied before. In this work, we give an almost complete picture of the approximability of FLT for both *general* and *metric* instances. For *general instances*, no specific assumption is made on the distance values in $\mathcal{D}$. For *metric instances*, we assume that the values in $\mathcal{D}$ satisfy the triangle inequality. We obtain stronger results for *Euclidean instances*, where the clients and the facilities lie in either $\mathbb{R}$ or $\mathbb{R}^2$ and the Euclidean distance is used. Many of our results (especially those for general instances) can be extended to non-uniform facility opening costs.

General Instances. For general instances, we almost match the approximability of SET COVER, by slightly violating the budget constraint. If we insist on satisfying the budget constraint, FLT becomes difficult to approximate even for $L = 2$. More specifically, in Section 2, we obtain the following results:

1. A greedy algorithm achieves an approximation ratio of $n(\sqrt[L]{\mathbb{H}_n/n})$. This matches the classical result for SET COVER when $L = 1$. We also present an example showing that our analysis is almost tight.
2. By extending the greedy algorithm for SET COVER, we obtain bi-criteria approximation algorithms with approximation guarantees of $(\mathbb{H}_n, L)$ and $(2\mathbb{H}_n, L - 1 + 1/L)$ for FLT.
3. By generalizing the randomized rounding algorithm for the SET COVER problem, we obtain a bi-criteria approximation of $(\mathcal{O}(\log n/\epsilon), 1 + \epsilon)$. So, we can achieve an asymptotically best possible logarithmic approximation, if we violate the budget constraint by a small constant factor. This result holds for non-uniform facility costs as well.

4. We propose a nontrivial generalization of the frequency parameter used for SET COVER. Formally, for $L = 2$, we introduce a parameter ψ, which is always bounded from above by the maximum number of facility pairs that can serve a client. Then, we obtain an LP-based ψ-approximation algorithm for $L = 2$ which satisfies the budget constraint.
5. If we insist on satisfying the budget constraint, one should not expect much better approximation guarantees. Using a transformation from SYMMETRIC LABEL COVER [10], we show that FLT cannot be approximated within a ratio of $\mathcal{O}(2^{\log^{1/2-\epsilon}\tau})$, for any $\epsilon > 0$, unless $\mathcal{NP}$ is in quasipolynomial time. Here, τ is no less than the maximum number of facility pairs that can cover any client.

Metric Instances. FLT becomes significantly easier to approximate in metric instances. This is especially true for Euclidean instances. More formally, in Section 3, we obtain the following results:

1. We show that a natural greedy algorithm achieves a bi-criteria guarantee of $(1, 3L)$, if the distance matrix $\mathcal{D}$ satisfies the triangle inequality.
2. By extending the dynamic programming algorithm for k-median on the real line, we show that FLT can be solved optimally in polynomial time for linear instances. This result can be extended to non-uniform facility costs (with a slightly different recursion though).
3. By extending the techniques of Mustafa and Ray [20], we obtain bi-criteria approximation algorithms with guarantees of $(1 + \epsilon, L)$ and $(2 + \epsilon, L - 1 + 1/L)$ for instances on $\mathbb{R}^2$.
4. Our main result is that FLT on the Euclidean plane admits a bi-criteria polynomial-time approximation scheme, with an approximation guarantee of $(1 + \epsilon, 1 + \epsilon)$, if all clients have a uniform budget B.

2 General Instances

Recall that for *general instances* of the FLT problem no specific assumptions are made on the distances in $\mathcal{D}$. The following lemma presented in [8] can be adapted to the FLT problem.

▶ **Lemma 1.** RED-BLUE SET COVER *has a $\mathcal{O}((c\rho)^{1-1/c}\log\rho)$-approximation algorithm when $\forall S \in \mathcal{S}, |S \cap R| \le c$ where $\rho = |\mathcal{S}|$ [8].*

If we restrict the type of facilities to 2 that is, $L = 2$ the Lemma 1 implies that there exists an $\mathcal{O}((\sqrt{2n})\log n)$-approximation algorithm. Below, we present a simple greedy algorithm that achieves an approximation ratio of $\sqrt{n\mathbb{H}_n}$ for 2 types of facilities.

2.1 Deterministic Algorithm

In Algorithm 1, we say that a client c is *covered* by a tuple $(i_1, \ldots, i_L) \in F_1 \times \cdots \times F_L$ if $\sum_{\ell=1}^{L} d_{c\,i_\ell} \le B_c$. Note that when considering $(i_1, \ldots, i_L)$, the algorithm does not take into account the clients of U that are covered by a tuple consisting of some facilities in $(i_1, \ldots, i_L)$ and some facilities that are already present in S and had been selected in previous rounds.

▶ **Theorem 2.** *Algorithm 1 is an $(n\left(\frac{\mathbb{H}_n}{n}\right)^{1/L})$-approximation algorithm for the* FLT *problem.*

Proof. Fix an instance and its optimal solution Y. Suppose $|Y \cap F_\ell| = a_\ell$, $\forall \ell \in [L]$. Then, $|Y| = \sum_{\ell=1}^{L} a_\ell$. For each tuple of L facilities $(i_1, \cdots, i_L) \in (Y \cap F_1) \times \cdots \times (Y \cap F_L)$, create a *bag* $B(i_1, \cdots, i_L)$. Each client c is put in exactly one bag $B(i_1, \cdots, i_L)$ such that $\sum_{\ell=1}^{L} d_{ci_\ell} \le B_c$. Break ties arbitrarily for the clients who can be placed in several bags. The $\prod_{\ell=1}^{L} a_\ell$ bags form a partition of $\mathcal{C}$.

Algorithm 1:

Initialize $S \leftarrow \emptyset$ and $U \leftarrow \mathcal{C}$
while $U \neq \emptyset$ **do**
  Choose $(i_1, \dots, i_L) \in F_1 \times \cdots \times F_L$ such that the number of clients in U covered by $(i_1, \dots, i_L)$ is maximized
  Add $\{i_1, \dots, i_L\}$ to S and remove from U the clients covered by $(i_1, \dots, i_L)$
return S

We repeatedly use the arithmetic-geometric means inequality: $\sum_{\ell=1}^{L} a_\ell \geq L(\prod_{\ell=1}^{L} a_\ell)^{1/L}$. Let X denote the solution output by Algorithm 1. We claim that

$$|X| \leq L(\prod_{\ell=1}^{L} a_\ell)\mathsf{H}_n. \tag{1}$$

To see this, observe the choices made by Algorithm 1 on the bags defined above. The first greedy choice is at least as good as covering the largest bag (i.e. selecting its corresponding facilities). Afterwards, update the bags by removing the clients currently covered by the partial greedy solution. The next choice is again, at least as good as covering the largest bag, and so on. Because there are $\prod_{\ell=1}^{L} a_\ell$ bags, the optimal solution uses at most $\prod_{\ell=1}^{L} a_\ell$ sets to cover the n clients. As for SET COVER, Algorithm 1 needs at most $\prod_{\ell=1}^{L} a_\ell \mathsf{H}_n$ rounds to cover all the clients, each round requiring at most L new facilities.

Suppose $\prod_{\ell=1}^{L} a_\ell \leq \frac{n}{\mathsf{H}_n}$. It follows from $|Y| = \sum_{\ell=1}^{L} a_\ell$ and (1) that the approximation ratio is at most $\frac{L(\prod_{\ell=1}^{L} a_\ell)\mathsf{H}_n}{\sum_{\ell=1}^{L} a_\ell}$. Combining this with the arithmetic-geometric means inequality, we obtain that $\frac{L(\prod_{\ell=1}^{L} a_\ell)\mathsf{H}_n}{\sum_{\ell=1}^{L} a_\ell} \leq (\prod_{\ell=1}^{L} a_\ell)^{1-1/L}\mathsf{H}_n$. Using the fact that $\prod_{\ell=1}^{L} a_\ell \leq \frac{n}{\mathsf{H}_n}$, we get that $(\prod_{\ell=1}^{L} a_\ell)^{1-1/L}\mathsf{H}_n \leq (\frac{n}{\mathsf{H}_n})^{1-1/L}\mathsf{H}_n = n^{1-1/L}(\mathsf{H}_n)^{1/L}$.

Now suppose $\prod_{\ell=1}^{L} a_\ell > \frac{n}{\mathsf{H}_n}$. We have $|X| \leq Ln$ because in the worst case, each client requires its own tuple of L facilities. The approximation ratio is at most $\frac{Ln}{\sum_{\ell=1}^{L} a_\ell}$. Using the arithmetic-geometric means inequality, we get that $\frac{Ln}{\sum_{\ell=1}^{L} a_\ell} \leq \frac{n}{(\prod_{\ell=1}^{L} a_\ell)^{1/L}} = \frac{n^{1-1/L} n^{1/L}}{(\prod_{\ell=1}^{L} a_\ell)^{1/L}}$ and $\prod_{\ell=1}^{L} a_\ell > \frac{n}{\mathsf{H}_n}$ raised to the power of $1/L$ to get that $\frac{n^{1-1/L} n^{1/L}}{(\prod_{\ell=1}^{L} a_\ell)^{1/L}} \leq n^{1-1/L}(\mathsf{H}_n)^{1/L}$. ◀

An almost tight instance: Take a positive integer t and create a set of $n = t^L$ clients $\{1, \cdots, t\}^L$. Each client is associated with a vector $\vec{c} \in \{1, \cdots, t\}^L$. The client with vector $\vec{c}$ can be covered by two separate sets of facilities: $(f_{\vec{c}_1}, \dots, f_{\vec{c}_L})$ and $(g_1^{\vec{c}}, \dots g_L^{\vec{c}})$. The optimum takes the "f" facilities (there are Lt such facilities) whereas the greedy algorithm can pick the "g" facilities (there are Lt^L such facilities). For the described family of instances, Algorithm 1 returns a $t^{L-1} = n^{1-1/L}$-approximate solution.

2.2 Bi-criteria Approximations

Theorem 2 gives a bi-criteria $\left(n\left(\frac{\mathbb{H}_n}{n}\right)^{1/L}, 1\right)$-approximation result for FLT problem. The simple strategy of solving L separate instances of SET COVER provides a bi-criteria $(\mathbb{H}_n, L)$-approximation algorithm. The exact proposition and proof is omitted due to space constraints. Next, we present another incomparable bi-criteria approximation algorithm.

▶ **Proposition 3.** FLT *admits a* $(2\mathbb{H}_n, L - 1 + 1/L)$*-approximate algorithm.*

Proof. From the instance of FLT, create an instance $\mathcal{I}_0$ of SET COVER as follows. Each facility i corresponds to a set that covers client j iff $d_{ij} \leq B_j/L$. The facilities' types are ignored. A $\mathbb{H}_n$-approximate solution S_0 is computed for $\mathcal{I}_0$ (greedy algorithm).

Let $\mathcal{C}_\ell$ be the clients assigned to a facility of type-ℓ in S_0. For every $\ell \in [L]$, create an instance $\mathcal{I}_\ell$ of SET COVER as follows. Each facility i of type ℓ corresponds to a set that covers client j iff $d_{ij} \leq B_j$ and $j \in \mathcal{C} \setminus \mathcal{C}_\ell$. A $\mathbb{H}_n$-approximate solution S_ℓ is computed for $\mathcal{I}_\ell$. Let T be an optimal solution to FLT. Let T_0 be a subset of T satisfying: $\forall j \in \mathcal{C}$, T_0 contains at least one facility $i \in T$ such that $d_{ij} \leq B_j/L$. Note that i must exist. As T_0 is a feasible solution to $\mathcal{I}_0$, we get that

$$|S_0| \leq \mathbb{H}_n |T_0| \leq \mathbb{H}_n |T| \tag{2}$$

For every $\ell \in [L]$, let T_ℓ be the restriction of T to its facilities of type-ℓ. Since T_ℓ is a feasible solution to $\mathcal{I}_\ell$, we get for every $\ell \in [L]$ that $|S_\ell| \leq \mathbb{H}_n |T_\ell|$. It follows that

$$|\bigcup_{\ell=1}^{L} S_\ell| = \sum_{\ell=1}^{L} |S_\ell| \leq \mathbb{H}_n \sum_{\ell=1}^{L} |T_\ell| = \mathbb{H}_n |T|. \tag{3}$$

Combine (2) and (3) to get that $|\bigcup_{\ell=0}^{L} S_\ell| \leq |S_0| + |\bigcup_{\ell=1}^{L} S_\ell| \leq 2\mathbb{H}_n|T|$. Since every client $j \in \mathcal{C}_\ell$ is at distance at most B_j/L from its assigned facility in S_0, and at distance at most B_j from its assigned facility in the $L-1$ instances of $\{\mathcal{I}_t : t \in [L] \setminus \{\ell\}\}$, client j is at total distance at most $(L - 1 + 1/L)B_j$ from its assigned facilities in $\bigcup_{\ell=0}^{L} S_\ell$. Thus, $\bigcup_{\ell=0}^{L} S_\ell$ is $(2\mathbb{H}_n, L - 1 + 1/L)$-approximate. ◀

2.3 LP-Based Approximations

Consider the FLT problem where $L = 2$. For each facility $i \in \mathcal{F}$, define a variable y_i such that $y_i = 1$, if the facility i is open and otherwise 0. For each client $j \in \mathcal{C}$, define $\mathcal{T}_j$ as the subset of $F_1 \times F_2$ such that $(i, i') \in \mathcal{T}_j$ if and only if $d_{ij} + d_{i'j} \leq B_j$. Let $\mathcal{T} = \bigcup_{j \in \mathcal{C}} \mathcal{T}_j$. An LP formulation of our problem is as follows:

$$\text{(LP-A) minimize } \sum_{i \in \mathcal{F}} y_i$$

$$\text{subject to: } y_i \geq x_{ii'},\ \forall (i, i') \in \mathcal{T} \tag{4}$$

$$\sum_{(i,i') \in \mathcal{T}_j} x_{ii'} \geq 1,\ \forall j \in \mathcal{C} \tag{5}$$

$$x_{ii'}, y_i \in \{0, 1\} \tag{6}$$

where $x_{ii'} = 1$ means that the pair of facilities (i, i') is opened.

Let $\phi_j := |\mathcal{T}_j|$ and $\phi := \max_{j \in \mathcal{C}} \phi_j$. Since ϕ is the maximum number of facility pairs which can serve a client, it is an adapted notion of frequency. If one solves the relaxation of LP-A and open every facility i such that $y_i \geq \phi^{-1}$, then the solution is feasible and ϕ-approximate. We are going to define a new parameter ψ such that $\psi \leq \phi$ and present an approximation algorithm with performance guarantee ψ.

Fix a client j and consider the bipartite graph $\mathcal{G}_j$ with vertex set $V_j \subseteq \mathcal{F}$ and edge set E_j. There is an edge $(i, i') \in E_j$ if and only if $i \in F_1$, $i' \in F_2$, and $d_{ij} + d_{i'j} \leq B_j$. Equivalently, $(i, i') \in E_j$ if and only if $(i, i') \in \mathcal{T}_j$. Furthermore, we impose that every vertex of F_j must have a positive degree.

▶ **Lemma 4.** *S is a feasible solution to* FLT *where $L = 2$ if, $\forall j \in \mathcal{C}$, $\forall$ vertex cover Q of $\mathcal{G}_j$, $S \cap Q \neq \emptyset$.*

Proof. Let S be a feasible solution. Fix a client $j \in \mathcal{C}$ for which S contains two facilities $i_1 \in F_1$ and $i_2 \in F_2$ such that $d_{ij} + d_{i'j} \leq B_j$. In other words, $\mathcal{G}_j$ has an edge (i_1, i_2). Since every vertex cover Q of $\mathcal{G}_j$ must contain either i_1 or i_2, we have that $S \cap Q \neq \emptyset$.

Now, let S' be a subset of $\mathcal{F}$ which intersects every vertex cover Q of every graph $\mathcal{G}_j$. Suppose by contradiction that S' is not a feasible solution. At least one client, say j', is not covered. Thus $S_{j'} := S' \cap V_{j'}$ is an independent set of $\mathcal{G}_j$. A contradiction is reached because $V_{j'} \setminus S'$ is a vertex cover of $\mathcal{G}_j$ that S' does not intersect. ◀

Lemma 4 provides a new formulation of FLT problem inspired from [8]. Let $\mathcal{Q}_j$ denote the set of all vertex covers of $\mathcal{G}_j$, and $\mathcal{Q} := \bigcup_{j \in \mathcal{C}} \mathcal{Q}_j$.

$$\text{(LP-B) minimize} \sum_{i \in \mathcal{F}} y_i \tag{7}$$

$$\text{subject to:} \sum_{i \in Q} y_i \geq 1, \ \forall Q \in \mathcal{Q} \tag{8}$$

$$y_i \in \{0, 1\}, \ \forall i \in \mathcal{F}$$

The relaxation of LP-B can be solved in polynomial time (the proof is omitted due to space constraints).

Let $\tilde{\mathcal{Q}}_j$ denote the set of all vertex covers of $\mathcal{G}_j$ that are exclusion-wise minimal, and $\tilde{\mathcal{Q}} := \bigcup_{j \in \mathcal{C}} \tilde{\mathcal{Q}}_j$. That is, $\tilde{\mathcal{Q}}$ is obtained from $\mathcal{Q}$ by discarding every member Q such that another member $\tilde{Q}$ satisfies $\tilde{Q} \subsetneq Q$. Note that a solution to LP-B satisfies $\forall \tilde{Q} \in \tilde{\mathcal{Q}}$, $\sum_{i \in \tilde{Q}} y_i \geq 1$, which is (8) where $\mathcal{Q}$ is substituted for $\tilde{\mathcal{Q}}$. Let ψ denote the size of the largest member of $\tilde{\mathcal{Q}}$. Interestingly, $\psi \leq \phi$ always holds (the proof is omitted due to space constraints).

▶ **Theorem 5.** FLT *admits a polynomial time ψ-approximation algorithm when $L = 2$.*

Proof. Solve the relaxation of LP-B and denote by y the solution. Guess ψ (with binary search) and open every facility i such that $y_i \geq \psi^{-1}$. The solution is feasible (Lemma 4) and ψ-approximate. For every $\tilde{Q} \in \tilde{\mathcal{Q}}$, at least one facility $i \in \tilde{Q}$ satisfies $y_i \geq |\tilde{Q}|^{-1} \geq \psi^{-1}$. Thus, at least one facility of every $\tilde{Q} \in \tilde{\mathcal{Q}}$ is open, and the same goes for $\mathcal{Q}$. ◀

2.4 Inapproximability

The SYMMETRIC LABEL COVER problem (SLC) is a variant of LABEL COVER introduced in [10] and defined as follows. We are given a complete bipartite graph where V_1 and V_2 are the two parts of the bipartition, and $|V_1| = |V_2| = q$. Two sets of labels $\mathcal{L}_1$ and $\mathcal{L}_2$ are given. For each $(v_1, v_2) \in V_1 \times V_2$, a relation $R(v_1, v_2) \subseteq \mathcal{L}_1 \times \mathcal{L}_2$ is given. A feasible solution is a pair of mappings $\mu_1 : V_1 \to 2^{\mathcal{L}_1}$ and $\mu_2 : V_2 \to 2^{\mathcal{L}_2}$ such that each edge (v_1, v_2) is *consistent*, that is there exists a pair $(\ell_1, \ell_2) \in \mu_1(v_1) \times \mu_2(v_2)$ such that $(\ell_1, \ell_2) \in R(v_1, v_2)$. The objective is to minimize $\sum_{v \in V_1} |\mu_1(v)| + \sum_{v \in V_2} |\mu_2(v)|$. An instance of SLC has size $\Theta(\sigma)$ where $\sigma := \sum_{(v_1,v_2) \in V_1 \times V_2} |R(v_1, v_2)|$ [7]. Unless $\mathcal{NP} \subseteq \mathcal{QP}$ (quasi-polynomial time), SLC cannot be approximated within a factor $\mathcal{O}(2^{\log^{1/2-\epsilon} \sigma})$ for any $\epsilon > 0$ [7].

Following the notation of Section 2.3, $\mathcal{T}_j$ is the set of pairs of facilities that cover client j and let $\tau := |\bigcup_{j \in \mathcal{C}} \mathcal{T}_j|$. Thus, the size of an instance of FLT with $L = 2$ is $\Theta(\tau)$.

▶ **Theorem 6.** *Unless $\mathcal{NP} \subseteq \mathcal{QP}$,* FLT *with $L = 2$ cannot be approximated within a factor $\mathcal{O}(2^{\log^{1/2-\epsilon} \tau})$ for any $\epsilon > 0$.*

Proof. Take an instance of SLC and build an instance of FLT with $L = 2$ as follows. Each edge $(x, y) \in V_1 \times V_2$ corresponds to a client j_{xy}. For each pair $(\ell_1, \ell_2) \in R(x, y)$, for some edge (x, y), create facilities (x, ℓ_1) and (y, ℓ_2) of types 1 and 2, respectively. We have $(\ell_1, \ell_2) \in R(x, y)$ iff the facilities $(x, \ell_1), (y, \ell_2)$ cover j_{xy}, i.e. $((x, \ell_1), (y, \ell_2)) \in \mathcal{T}_{j_{xy}}$.

From a feasible solution to SLC, build a solution with no greater cost: for each edge (x, y), take $\ell_1 \in \mu_1(x)$ and $\ell_2 \in \mu_2(y)$ such that $(\ell_1, \ell_2) \in R(x, y)$, and open facilities (x, ℓ_1) and (y, ℓ_2). Note that such a pair (ℓ_1, ℓ_2) exists since (μ_1, μ_2) form a feasible solution to the SLC instance. From a feasible solution to FLT, build a solution to SLC having the same cost. At the beginning, $\mu_i(v)$ is empty for every $v \in V_i$ and $i \in \{1, 2\}$. Then, for each open facility $(v, \ell) \in V_i \times \mathcal{L}_i$, add ℓ to $\mu_i(v)$. In this reduction, σ is equal to τ. ◀

2.5 Randomized Algorithm

Consider a natural LP formulation of the FLT problem. For each facility $i \in \mathcal{F}$, we define a variable y_i. For each pair of a client j and a facility i, we define a variable x_{ij} such that $x_{ij} = 1$ if i serves j. Then the FLT problem can be formulated as:

$$\text{(LP1)} \quad \min \sum_{i \in \mathcal{F}} y_i$$

$$\begin{aligned}
\text{subject to} \quad & y_i \geq x_{ij}, && \forall i, j \in \mathcal{F} \times \mathcal{C} && (9)\\
& \sum_{i \in F_\ell} x_{ij} \geq 1, && \forall \ell, j \in [L] \times \mathcal{C} && (10)\\
& \sum_{i \in \mathcal{F}} x_{ij} d_{ij} \leq B_j, && \forall j \in \mathcal{C} && (11)\\
& x_{ij} \in \{0, 1\}, && \forall i, j \in \mathcal{F} \times \mathcal{C} && (12)\\
& y_i \in \{0, 1\} && \forall i \in \mathcal{F} && (13)
\end{aligned}$$

The relaxation of LP1 is obtained by letting the variables x_{ij} and y_i attain fractional values between 0 and 1. Note that the objective value of an optimal solution to the relaxed LP is a lower bound on the objective value of an optimal integral solution. A simple and natural idea for rounding an optimal fractional solution is to consider the fractions as probabilities. Below, we show that this idea leads to a $\mathcal{O}(\log n/\epsilon)$-approximation algorithm where the service constraint (11) is relaxed by a factor of at most $(1 + \epsilon)$. The proof of Theorem 7 is omitted due to space constraints.

▶ **Theorem 7.** *There exists a randomized algorithm with the performance guarantee of $(\mathcal{O}(\log(n)/\epsilon), (1 + \epsilon))$ where $\epsilon \in (0, 1)$ for the* FLT *problem.*

3 Metric Instances

In this section, we assume that facilities and clients are placed in a metric space where the distances satisfy the triangle inequality. We next show the following:

▶ **Theorem 8.** *For the* FLT *problem, there exists a $(1, 3L)$-approximation algorithm when values in the distance matrix $\mathcal{D}$ follow triangle inequality.*

Algorithm 2: Greedy algorithm in metric space.

Data: $\mathcal{C}$, $\ell \in [L]$, F_ℓ, $(B_1, \dots, B_n)$

Initialize $S_\ell \leftarrow \emptyset$, $U \leftarrow \mathcal{C}$ and $q \leftarrow 1$

while $U \neq \emptyset$ **do**

    Find $j^q \in U$ with smallest budget B_{j^q}

    Let p_ℓ^q be a facility of F_ℓ such that $d_{j^q p_\ell^q} \leq B_{j^q}$

    $S_\ell \leftarrow S_\ell \cup \{p_\ell^q\}$

    Let $\mathcal{C}_\ell^q = \{j \in U \mid d_{j p_\ell^q} \leq 3B_j\}$

    The clients of $\mathcal{C}_\ell^q$ are *assigned* to p_ℓ^q, and $j^q \in \mathcal{C}_\ell^q$ is the *representative* of p_ℓ^q

    $U \leftarrow U \setminus \mathcal{C}_\ell^q$

    $q \leftarrow q + 1$

return S_ℓ

Proof. Consider the algorithm 2. The algorithm 2 run for different values of $\ell \in [L]$. For a fix $\ell \in L$, it identifies a set of facilities S_ℓ and $|S_\ell|$ representatives. By construction, the representatives j, j' of two different facilities in S_ℓ must be at distance strictly larger than $2\max(B_j, B_{j'})$ from one another: take the representative j^q of p_ℓ^q; we have $d_{j^q p_\ell^q} \leq B_{j^q}$. Take a representative j^g of p_ℓ^g such that $g > q$. Thus $B_{j^g} \geq B_{j^q}$ and $3B_{j^g} < d_{j^g p_\ell^q}$. By the triangle inequality $d_{j^g p_\ell^q} \leq d_{j^g j^q} + d_{j^q p_\ell^q}$. We get that $d_{j^g j^q} > 3B_{j^g} - B_{j^q} \geq 2B_{j^g} = 2\max(B_{j^g}, B_{j^q})$.

Because $d_{jj'} > 2\max(B_j, B_{j'})$, two representatives j, j' cannot share an ℓ-facility in the optimum. Therefore there are at least $|S_\ell|$ facilities of type ℓ in an optimal solution. It follows that $\bigcup_{\ell=1}^{L} S_\ell$ is a 1-approximation of the optimum concerning the number of open facilities. Since every client j is at distance at most $3B_j$ from its assigned facility of type ℓ, for every $\ell \in [L]$, the approximation ratio on the distance is $3L$. ◀

3.1 Euclidean Instances

In this section, we consider instances where the clients and the facilities lie in either $\mathbb{R}$ or $\mathbb{R}^2$ and the Euclidean distance is used. We show that:

▶ **Proposition 9.** *There is an $\mathcal{O}(nm^L(n+m^L))$-time optimal dynamic programming algorithm for linear instances of* FLT*, where all clients and facilities lie in $\mathbb{R}$.*

The proof is omitted due to space constraints. Next, we present the result of Mustafa and Ray [20] for the geometric HITTING SET problem. The algorithm presented in [20] is a simple local search which starts with any feasible solution (for example open all facilities) and iteratively reduces the size of this set as long as there does not exist a set of k facilities which can be replaced by $k-1$ facilities, where k is some integer given as an input. This algorithm is known as a k-level local search algorithm. Their main result is the following:

▶ **Lemma 10.** *There exists a constant c such that $(c/\epsilon)^2$-level local search algorithm returns a hitting set of size at most $(1+\epsilon)$ times the size of an optimal hitting set where $\epsilon \in (0,1)$ [20].*

If the same reasoning as for propositions 3 (replace the greedy algorithm with the PTAS of [20]), then in $\mathbb{R}^2$, FLT admits approximation algorithms with guarantees $(1+\epsilon, L)$ and $(2+\epsilon, L-1+1/L)$.

3.1.1 A Local Search Algorithm in $\mathbb{R}^2$

Recall that L is a constant. We say that S is *ϵ-feasible* if each client j is served by a type-ℓ facility and the total service cost for j is at most $(1+\mathcal{O}(\epsilon))B_j$ for $0<\epsilon<1$. Let S and S' denote two ϵ-feasible solutions. Then $S \oplus S'$ denotes the symmetric difference between S and S', that is $S \oplus S' := (S'-S) \cup (S-S')$.

Local Search Algorithm. Start with any ϵ-feasible solution S. While possible, replace S with an *ϵ-feasible* set of facilities S' such that $|S'|<|S|$ and $|S \oplus S'| \le \mathcal{O}(1/\epsilon^4)$.

Observe that the local search algorithm is similar to the k-level local search algorithm mentioned in [20]. The only difference is in the definition of feasibility. That is, a solution in the k-level local search algorithm is considered feasible if the budget for each client j is at most B_j (the budget corresponds to the radius of disks), whereas our local search algorithm relaxes the budget for each client by a factor of $1+\mathcal{O}(\epsilon)$.

▶ **Lemma 11.** *The running time of the local search algorithm is polynomial in the size of the input.*

Proof. An initial *ϵ-feasible* solution is to open, for each client j, the closest facility of each type $\ell \in [L]$. Hence the initial solution opens at most nL facilities. Since in each iteration the local search algorithm reduces the number of facilities by at least one, the total number of iterations is at most nL. In each iteration, the number of possible different combinations to check is at most $\binom{m}{\mathcal{O}(1/\epsilon^4)}$. Hence the total running time of the algorithm is $nLm^{\mathcal{O}(\frac{1}{\epsilon^4})}$. The lemma follows since L is a constant. ◀

▶ **Theorem 12.** *Assume that clients have uniform upper bound on the service cost, that is, $\forall j \in \mathcal{C}, B_j = B$. Then, the local search algorithm achieves a $(1+\mathcal{O}(\epsilon), 1+\mathcal{O}(\epsilon))$-approximation ratio for the* FLT *problem in $\mathbb{R}^2$ where $\epsilon \in (0,1)$.*

Proof. We assume w.l.o.g. that $\frac{1}{\epsilon}$ is an integer, and that the set of clients $\mathcal{C}$ and the set of facilities $\mathcal{F}$ are enclosed in an area $\mathcal{A}$. Let R, R^{2B} and R^{4B} denote squares centered at a given point p of width $\frac{2B}{\epsilon}$, $\left(\frac{2B}{\epsilon}+2B\right)$, and $\left(\frac{2B}{\epsilon}+4B\right)$, respectively. Let ALG and OPT denote the set of facilities opened by the local search algorithm and in some fixed optimal solution, respectively. Further, let $ALG(R')$ and $OPT(R')$ represent the restrictions of ALG and OPT to the square R', respectively.

Next, we grid the entire region $\mathcal{A}$ such that the internode distance is ϵB. Let K denote the set of small squares of width ϵB. Let OPT' be a solution such that for each *tiny* square $k \in K$ and each type $\ell \in [L]$, one type-ℓ facility is opened if and only if OPT has at least one open type-ℓ facility in k. Thus, we have $|OPT'| \le |OPT|$. Let $OPT'(R')$ represent the set of facilities open inside the region R' in OPT'. Also, we have $|OPT'(R^{4B})| \le |OPT(R^{4B})|$.

Consider the intermediate solution M formed by removing all the facilities opened by the local search algorithm in the square R from ALG and adding all the facilities opened in OPT' inside the square R^{4B} that is,

$$M = (ALG \setminus ALG(R)) \cup OPT'(R^{4B})$$

▶ **Claim 13.** *M forms an* ϵ-feasible *solution to the* FLT *problem.*

Proof. Observe that the facilities opened inside the square R can serve clients in the square R^{2B}. Therefore, closing the facilities inside R can lead to infeasible solution for clients in the square R^{2B}. Let j be some client enclosed in square R^{2B}. Let F_j^o denote the set of facilities in OPT that serve j. Observe that j can only be served by facilities in the square R^{4B}. Hence $F_j^o \subseteq OPT(R^{4B})$. Consider the grid formed with the internode distances ϵB in R^{4B} . Recall that OPT', for each *tiny* square with width ϵB and each type ℓ, opens a facility of type-ℓ if and only if OPT has at least an open facility type-ℓ in the same square. Thus for each type ℓ facility in F_j^o, there exists a facility a type ℓ facility in $OPT'(R^{4B})$ such that the service cost of j to a type-ℓ facility is at most the service cost of j to type-ℓ in OPT plus $\sqrt{2}\epsilon B$. Summing over all types, the claim follows. ◀

Observe that $|M \oplus ALG|$ can be much larger than $\mathcal{O}(\frac{1}{\epsilon^4})$ if $ALG(R)$ is huge. However, M can be realized after few steps of the local search algorithm wherein each iteration, the local search algorithm closes $\mathcal{O}(\frac{1}{\epsilon^4})$ facilities from $ALG(R) \setminus OPT'(R^{4B})$. Thus, the local exchange argument states that

$$|ALG| \le |M| = |(ALG \setminus ALG(R)) \cup OPT'(R^{4B})|$$
$$|ALG \setminus ALG(R)| + |ALG(R)| \le |(ALG \setminus ALG(R))| + |OPT'(R^{4B})|$$
$$|ALG(R)| \le |OPT'(R)| + |OPT'(R^{4B} - R)|$$

Let $\mathcal{R}^P$ denote the set of all regions in $\mathcal{A}$ according to some partitioning scheme P. For each region $R \in \mathcal{R}^p$, the above local exchange argument holds. Summing over all regions, we have

$$\begin{aligned}\sum_{R\in\mathcal{R}^p} |ALG(R)| &\le \sum_{R\in\mathcal{R}^p} (|OPT'(R)| + |OPT'(R^{4B} - R)|)\\ &\le |OPT'| + \sum_{R\in\mathcal{R}^p} |OPT'(R^{4B} - R)|\end{aligned}$$

▶ **Claim 14.** *There exists a partition Q such that* $\sum_{R\in\mathcal{R}^Q} |OPT'(R^{4B} - R)| = O(\epsilon)|OPT'|$.

Proof. In this proof, we use the idea of the "grid shifting strategy" mentioned in [16]. Due to space constraints, the proof is omitted. ◀

From above claim it follows that $|ALG| \le (1 + \mathcal{O}(\epsilon))|OPT'| \le (1 + \mathcal{O}(\epsilon))|OPT|$. ◀

4 Conclusion and Directions for Further Research

We introduce and study the approximability of covering problems with multiple types and a hard constraint on the combined "quality" of each client's covering. Our work leaves many promising directions for future research. A natural question is whether we could obtain strong approximation guarantees for metric instances of constant doubling dimension.

Given that it is very difficult, if even possible, to achieve good approximation guarantees without violating the budget constraint (for both general and metric instances), it would be very interesting to investigate the approximability of FLT with penalties (see e.g., [19, 5] for the approximability of other covering problems with penalties). In FLT with penalties, there is a *covering penalty*, which is a non-decreasing function of each client's total covering cost. In the simplest case, the covering penalty is 0, if the budget constraint is satisfied, and some $p_j > 0$, if the budget constraint is not satisfied for client j. The cost of the solution is the sum of the facility opening costs and the covering penalties for all clients.

Another natural research direction is to determine the competitive ratio of online FLT, where the clients arrive one-by-one and must be covered by a facility of each type upon arrival. A promising starting point is the ideas and techniques applied to ONLINE SET COVER [2] for general instances and to ONLINE FACILITY LOCATION problems (see e.g., [13] and the references therein) for metric instances.

References

1 A. Ageev, Y. Ye, and J. Zhang. Improved Combinatorial Approximation Algorithms for the k-Level Facility Location Problem. *SIAM J. Discrete Math.*, 18(1):207–2017, 2004.

2 N. Alon, B. Awerbuch, Y. Azar, N. Buchbinder, and J. Naor. The Online Set Cover Problem. *SIAM J. Computing*, 39(2):361–370, 2009.

3 S. Arora, N. Gupta, S. Khuller, Y. Sabharwal, and S. Singhal. Facility Location with Red-blue Demands. *Operation Research Letter*, 42(6):462–465, September 2014.

4 N. Bansal and K. Pruhs. Weighted geometric set multi-cover via quasi-uniform sampling. *J. Comp. Geometry*, 7(1):221–236, 2016.

5 R. Bar-Yehuda, G. Flysher, J. Mestre, and D. Rawitz. Approximation of Partial Capacitated Vertex Cover. *SIAM J. Discrete Math.*, 24(4):1441–1469, 2010.

6 J. Byrka, S. Li, and B. Rybicki. Improved Approximation Algorithm for k-level Uncapacitated Facility Location Problem (with Penalties). *Theory of Computing Systems*, 58(1):19–44, 2016.

7 A. Caprara, G. F. Italiano, G. Mohan, A. Panconesi, and A. Srinivasan. Wavelength rerouting in optical networks, or the Venetian Routing problem. *J. Algorithms*, 45(2):93–125, 2002.

8 R.D. Carr, S. Doddi, G. Konjevod, and M.V. Marathe. On the red-blue set cover problem. In *Proc. of SODA*, pages 345–353, 2000.

9 I. Dinur and D. Steurer. Analytical approach to parallel repetition. In *Proc. of STOC*, pages 624–633, 2014.

10 Y. Dodis and S. Khanna. Designing Networks with Bounded Pairwise Distance. In *Proc. of STOC*, pages 750–759, 1999.

11 A. Efrat, F. Hoffmann, C. Knauer, K. Kriegel, G. Rote, and C. Wenk. Covering Shapes by Ellipses. In *Proc. of SODA*, pages 453–454, 2002.

12 M. Elkin and D. Peleg. The Hardness of Approximating Spanner Problems. *Theory Comp. Systems*, 41(4):691–729, 2007.

13 D. Fotakis. Online and Incremental Algorithms for Facility Location. *SIGACT News*, 42(1):97–131, 2011.

14 M.R. Garey and D.S. Johnson. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., 1990.

15 M. Hajiaghayi, R. Khandekar, and G. Kortsarz. Budgeted Red-blue Median and Its Generalizations. In *Proc. of ESA*, pages 314–325, 2010.

16 D.S. Hochbaum and W. Maass. Approximation Schemes for Covering and Packing Problems in Image Processing and VLSI. *J. ACM*, 32(1):130–136, 1985.

17 D.S. Hochbaum and W. Maass. Fast approximation algorithms for a nonconvex covering problem. *J. Algorithms*, 8(3):305–323, 1987.

18 K. Jain and V.V. Vazirani. An approximation algorithm for the fault tolerant metric facility location problem. In *Proc. of APPROX*, pages 177–183, 2000.

19 J. Könemann, O. Parekh, and D. Segev. A Unified Approach to Approximating Partial Covering Problems. *Algorithmica*, 59(4):489–509, 2011.

20 N.H. Mustafa and S. Ray. Improved Results on Geometric Hitting Set Problems. *Discrete & Computational Geometry*, 44(4):883–895, 2010.

21 C. Swamy and D.B. Shmoys. Fault-tolerant facility location. *ACM Trans. Algorithms*, 4(4):51:1–51:27, 2008.

22 R.R. Yager. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans. on Systems, Man, and Cybernetics*, 18(1):183–190, 1988.