Addressing the Hardness of k-Facility Relocation Problem: A Pair of Approximate Solutions

Hu Wang
Xidian University
Xi'an, China
wh2mail@163.com

Hui Li
Xidian University
Xi'an, China
hli@xidian.edu.cn

Meng Wang
Xi'an Polytechnic University
Xi'an, China
wangmeng@xpu.edu.cn

Jiangtao Cui
Xidian University
Xi'an, China
cuijt@xidian.edu.cn

ABSTRACT

Facility Relocation (FR), which is an effort to reallocate the placement of facilities to adapt to the changes of urban planning and population distribution, has remarkable impact on many application areas. Existing solutions to the FR problem either focus on relocating one facility (*i.e.,* 1-FR) or fail to guarantee the result quality on relocating $k > 1$ facilities (*i.e.,* k-FR). As k-FR problem is NP-hard and is not submodular or non-decreasing, traditional hill-climb approximate algorithm cannot be directly applied. In light of that, we propose to transform k-FR into another facility placement problem, which is submodular and non-decreasing. We theoretically prove that the optimal solution of both problems are equivalent. Accordingly, we are able to present the first approximate solution towards the k-FR, namely FR2FP. Our extensive comparison over both FR2FP and the state-of-the-art heuristic solution shows that FR2FP, although provides approximation guarantee, cannot necessarily given superior results to the heuristic solution. The comparison motivates and, more importantly, directs us to present an advanced approximate solution, namely FR2FP-ex. Extensive experimental study over both real-world and synthetic datasets have verified that, FR2FP-ex demonstrates the best result quality. In addition, we also exactly unveil the scenarios when the state-of-the-art heuristic would fail to provide satisfied results in practice.

CCS CONCEPTS

• **Theory of computation → Facility location and clustering**; *Approximation algorithms analysis*; • **Information systems →** Wrappers (data mining).

KEYWORDS

Facility Relocation, Submodular, Approximate Algorithm

ACM ISBN 978-1-4503-8446-9/21/11…$15.00
https://doi.org/10.1145/3459637.3482411

ACM Reference Format:
Hu Wang, Hui Li, Meng Wang, and Jiangtao Cui. 2021. Addressing the Hardness of k-Facility Relocation Problem: A Pair of Approximate Solutions. In *Proceedings of the 30th ACM Int'l Conf. on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3459637.3482411

1 INTRODUCTION

The facility relocation (FR) problem aims to reallocate facilities in light of changes in users' locations. This can improve service quality and is useful in many applications. For example, when a new subway line is launched, many people may resettle to different locations. As a result, facilities (*e.g.,* chain store, firehouse, *etc.*) may need to be reallocated.

Several studies [7, 11, 18, 19, 23] have been undertaken to solve the FR problem under the Min-dist criteria with a variety of constraints. Specifically, given a set of users U , existing facility locations F, a set of new locations C, assume that each user is associated with a nearest facility in F, it is rational to expect that the distance between each user and her nearest facility is minimized. Driven by that, FR [18] aims to relocate k arbitrary facilities $f \in F$ with new locations $c \in C$ in order that the average distance between all users and their nearest facilities is minimized.

Most of the existing solutions to the FR problem only consider the situation where we only relocate one facility [18, 21], *i.e.,* k = 1. However, in practice, the k should not be limited in order to minimize the service distance. For instance, consider population changes in 254 counties in Texas. During the period from 2010 to 2021[1], there are 38 counties with a population change rate greater than 20% (including both increase and decrease). Among them, 19 counties have a change rate over 30%. Suppose there is a chain store (*e.g.,* McDonald's) which deployed facilities according to the population distribution in 2010. By 2021, due to the change in population distribution, it will definitely be unable to provide the best service and thus fail to obtain the maximum benefits especially when their competitors have improved the service by relocating facilities. Therefore, there exist enough motivation for the store to relocate a series of the existing facilities to improve the service network (*i.e.,* to minimize the service distance).

Notably, given the existing solutions towards FR problem with $k = 1$, referred to as 1-FR, extensively addressing the problem with

[1]https://worldpopulationreview.com/us-counties/states/tx

$k > 1$, referred to as k-FR, is not a trivial task due to the following challenges. Firstly, the solution space changes from $|F| \times |C|$ to $\sum_{i=1}^{k} \binom{i}{|F|} \times \binom{i}{|C|}$, which is not polynomial. Note that the solution space does not necessarily become $\binom{k}{|F|} \times \binom{k}{|C|}$, but is eventually much larger. The reason is, although we hope to relocate k facilities, the average distance of users may be already minimized when $k'(k' < k)$ facilities are replaced. Obviously, the k-FR problem can degenerate to the well-known k-median problem [6] when we set $F = \emptyset$. Therefore, the k-FR problem is at least as hard as the k-median problem, which is NP-hard [8]. Secondly, as we shall show in Section 5, k-FR is not submodular or non-decreasing, such that greedy approximate strategy cannot be directly applied. Although there exist heuristic solution [21] towards the k-FR problem. They can not provide approximate ratio over the results, such that the efficacy of the solutions are not reliable.

In addition to the hardness of the k-FR problem, we advocate that the users considered in FR problem should not be necessarily static, but allowed to be dynamic. That is, when considering the service distance from a facility towards a user, the user cannot be simply modeled as a static position, but a series of positions along her movement track. Similar problem statement can also be found in a series of related works [3, 21].

In order to address the limitations above, in this paper, we propose to transform k-FR into a facility placement problem [4, 9, 20], which is submodular and non-decreasing, and prove that the transformed problem is equivalent to the k-FR. By converting the k-FR problem into the facility placement problem, we proposed the first approximate solution towards the k-FR problem, the results of which can provide an approximation guarantee to the optimal ones. By comparing with the state-of-the-art heuristic solution [21], we observe that the practical efficacy of the basic approximate algorithm is not superior to the heuristic one. In fact, as k is small, the result quality of the basic solution is poor than that of [21]. We explained the reason for this phenomenon and accordingly, proposed an advanced approximate solution, called FR2FP-ex, which can obtain the best performance over the competitors while ensuring the same approximation ratio.

In summary, the major contributions of this paper are as follows:

- As the k-FR problem does not satisfy submodularity or non-decreasing, it cannot be directly approximately addressed by hill-climb solutions. In light of that, we propose to equivalently transform the k-FR in to a facility placement problem, where we propose an approximate solution. To the best of our knowledge, it is the first approximate solution towards the k-FR problem.
- We observe that, although the approximate solution provides result guarantee, it does not necessarily provide better results practically, sometimes worse. Our thorough investigation, both theoretical and empirical, unveils the secret behind the unexpected performance of the existing heuristic solution.
- Given our insight in the performance comparison between the basic approximate solution and state-of-the-art heuristic one, we propose an advanced approximate solution, which demonstrate superior performance compared to the state-of-the-art heuristic and ensures the approximation ratio in the quality of the result.

The rest of this paper is organized as follows. We introduce the related work In Section 2. In Section 3, we discuss how to model the movements of users. Section 4 formally gives the definition of the k-FR problem. In Section 5, two methods are proposed to solve k-FR problem with approximation rate. Section 6 presents our experiments and results. In Section 7, we extensively discuss the scenarios where the state-of-the-art heuristic will fail in practice. In Section 8 we conclude this article.

2 RELATED WORK

Min-dist in Euclidean. [14] studied the k-medios problem based on the minimum distance. The goal is to select k center points from all points so that the summed distance between all other points and the nearest k centers is the smallest. [29] studied another problem, by assuming that there are already some facilities and users, find a location in a given area to establish a new facility to minimize the average distance from the user to the nearest facility. The author proposed a progressive algorithm that can gradually obtain the optimal solution. [17] studied a discrete form of the above problem. The possible location facilities are no longer the entire continuous space, but a set of discrete locations specified in advance. Their goal is to find a location from these locations to build a new facility which minimizes the average distance between users and the nearest facility. The above researches are all based on Euclidean space, which is different from the issue we consider on the road network.

Min-dist on the road. [25] studied the optimal location problem on the road network. They proposed a framework based on the divide-and-conquer strategy to place multiple facilities at the same time. The problem studied by [2] is the same as that of [25]. Based on the idea of nearest local network [5], they proposed an efficient algorithm to solve the problem. As discussed in [18], in real applications, we are always allowed to choose from some candidate locations. Hence the answers generated by these approaches may not eventually be valid in practice. [16] finds a facility among existing ones that has the minimum average distance to all the users, *w.r.t.* Euclidean distance, while [28] solved the problem on road network utilizing network connectivity information and spatial locality. In [32], Voronoi diagram based look-up tables were designed to avoid network traversal. [26] presented a two-phase convex-hull-based pruning technique for both exact and approximate solutions. As these efforts only consider the facility placement but not reallocation, thus are orthogonal to our problem setting and cannot be applied to address k-FR.

Facility relocation. For the first time, [23] considered the removal of old facilities in the facility location problem, which is the prototype of the FR problem, and proposed three approximate algorithms to solve the problem. But in this study, the distance between users and facilities is assumed to be known, which is not guaranteed in our problem. [11] studied the re-planning of ambulances, which is essentially a special FR problem, which means that all facilities are relocated. They proposed a PAM-based method to solve this problem. Since it may not be necessary to relocate all facilities in k-FR, this method cannot solve the k-FR problem. [7] studied the mobile facility location problem, in which users and facilities need to be relocated at the same time. However, only facilities can be relocated in k-FR, so this method is not applicable. Therefore, none of the above methods can be used to solve the k-FR problem.

[18] considered a FR problem very similar to our setting. Based on Replacement Influence Distance, which is used to restrict the search space, they solved this problem in Euclidean space. [21] studied the FR problem on road network and they proposed two methods, depending on whether an index can be prepared beforehand, to solve the problem. However, they only consider the 1-FR problem, but fails to provide robust solution towards the k-FR. In fact, although [21] focuses on the 1-FR, the author extensively discussed how their solution to 1-FR can be extended to a heuristic solution in order to address the k-FR. Besides, they failed to provide any theoretical or empirical study for their efficacy or efficiency in k-FR setting. In this work, we give a thorough study over their proposed strategy and unveil that 1) although failing to guarantee the result approximation ratio, the heuristic solution can be as good as, or even superior to, our basic approximation scheme; 2) the scenarios where the heuristic solution fails to provide acceptable results are also identified by our exhaustive study.

3 PRELIMINARIES

According to our discussion in Section 1, we should not restrict the users to be static, thus hereby we shall introduce a group of preliminaries that are used to model the movement of users.

3.1 Mobile Objects

In real life, moving objects are ubiquitous. To represent the movement of an object, there are two ways: discrete (*e.g.,* check-in data [27]) and continuous (*e.g.,* trajectory [31]). No matter which method is used, a moving object is modeled as a set of positions [22] (*e.g.,* sample points of a trajectory or check-ins at POIs). We must not consider all of them, because it not only leads to costly computation, but also is inappropriate due to three kinds of valueless points: noisy, passing-by and outlier. Noise [30] is caused by data or GPS errors. Passing-by points are the points that the user only passes through without performing any meaningful activities. The outlier points are those visited occasionally. These points have little value for us to find the correct result of FR.

On the other hand, as remarked in [12], moving users are intimately associated with two major behaviors: frequently appearing at some place and staying for a duration. Obvisouly, the points in these locations are more important. Therefore, it makes sense to identify these points (we called reference locations) from raw movement history data, which enables us to pave the way to effectively capture daily activity places for handling the k-FR problem.

3.2 Capturing Reference Locations

We adopt the same strategy as [21] to identify the reference locations, employing the kernel method. Kernel method has been widely used in a variety of domains, including the detection of frequent activity places for humans [24]. We use the standard bivariate normal density kernel, which is

$$f(\langle x,y\rangle) = \frac{1}{nh^2}\sum_{i=1}^{n}\frac{1}{2\pi}exp(-\frac{d_\varepsilon(\langle x,y\rangle)(\langle x_i,y_i\rangle)}{2h^2})$$

and set $h : h = \frac{1}{2}(\zeta_x^2, \zeta_y^2)^{\frac{1}{2}} n^{-\frac{1}{6}}$ following [24]. For a more detailed introduction to the kernel method, please refer to [24].

In line with [12], we capture reference locations of each user as follows. We discretize the continuous space into grids and evaluate the density for each of them. The top-5% grids with the highest density are selected. Inspired by the notion of reference spot [12], we aggregate the adjacent grids among the selected ones together to form a series of grid groups. In each group, the peak grid with the highest density is intuitively regarded as a reference location. Moreover, the density accumulation of each group is normalized and viewed as the probability that a user appears nearby the corresponding reference location.

4 PROBLEM DEFINITION

In order to formally define the k-FR problem, we introduce some necessary terminologies. A location l in this paper is a planar position on an edge in a given directed road network $G(V,E)$, with a geographical coordinate (*i.e.,* latitude and longitude). Each directed edge in E between a pair of vertices in V is associated with a positive cost, *i.e.,* travel distance or time *etc.* Given any two locations l_1 and l_2, the directed network distance from l_1 to l_2 is denoted by $d(l_1,l_2)$, which may not be equal to $d(l_2,l_1)$. Since the locations of existing facilities (*resp.,* candidates to deploy a substitute) can usually be obtained precisely, we denote facilities (*resp.,* candidates) as a set of locations $F = \{f_1,\dots,f_{|F|}\}$(resp.,$C = \{c_1,\dots,c_{|C|}\}$), where $|F|$ (*resp.,* $|C|$) is the cardinality of F (*resp.,* C). For a user at location l the network nearest facilities with respect to F are denoted as $nn(F,l)$, and the network distances from l to it is defined as $dnn(F,l) = d(l, nn(F,l))$. A facility relocation (FR) pair that consists of k obsolete facilities $F_k \subseteq F$ and k candidates $C_k \subseteq C$ for substitution is defined as $\langle F_k, C_k\rangle$. For a user located at l if $\langle F_k, C_k\rangle$ is carried out, the distance to his nearest facility will become $dnn(F \setminus F_k \cup C_k, l)$.

To minimize the average distance between users and their respective nearest facilities, we have to identify their locations first. However, as discussed in Section 3, a mobile object u may be present at a set of n_u reference locations $L(u) = \{r_1,\dots,r_{n_u}\}$. Let $l(\cdot)$ denote a reference location(s) where "$\cdot$" is(are) present at, then $l(u) \in L(u)$ and $\sum_{i=1}^{n_u} Pr[l(u) = r_i] = 1$. Observe that this differs from the classical Min-dist criteria, where each object has only a single location. How can we select the facility relocation pairs when considering the users movement?

Similar to [3], we employ the notion of **possible world** [1] to model the movements of users. Given a set of m users $U = u_1,\dots,u_m$, each user u_i is associated with reference locations $L(u_i)$, then a **possible world** $w = (r_1^w,\dots,r_m^w)$ is a list of location instances, one for each user, where $r_i^w \in L(u_i)$. Assume that the reference locations of users are independent from each other, then $Pr\left[l(U) = w\right] = \prod_{i=1}^{m} Pr\left[l(u_i) = r_i^w\right]$. Let W be all possible worlds, then $|W| = \binom{|L(u_1)|}{1}\dots\binom{|L(u_m)|}{1} = \prod_{i=1}^{m}|L(u_i)|$. It is obvious that $\sum_{w\in W} Pr[l(U) = w] = 1$.

Note that for a particular possible world w, each user is associated with only a single reference location. This is in line with the setting of the Min-dist FR problem [13, 19].

Definition 4.1. Given a facility set F, a possible world w, and a FR pair $\langle F_k, C_k\rangle$, **the change of total distance** between all users

U and their respective nearest facilities is defined as

$$\Delta_w(\langle F_k, C_k \rangle) = \sum_{i=1}^{m} \left(d_{nn}\left(F, r_i^w\right) - d_{nn}\left(F \setminus F_k \cup C_k, r_i^w\right) \right)$$

$\Delta_w(\langle F_k, C_k \rangle)$ can be an alternative to evaluate the final total distance. When $\Delta_w(\langle F_k, C_k \rangle)$ is maximum, the final distance is minimum. Hence, the FR pair with the maximal $\Delta_w(\langle F_k, C_k \rangle)$ is the optimum in w with respect to the Min-dist criterion. As the distribution of Δ_w for any FR pair is a random variable, it is reasonable to evaluate the expected value over all possible worlds.

Definition 4.2. Given a set of facilities F and a set of moving users U for all possible worlds W, **the expected change of total distance (ED)** with respect to a FR pair $\langle F_k, C_k \rangle$ is defined as

$$\Delta(\langle F_k, C_k \rangle) = \sum_{w \in W} \left(\Delta_w(\langle F_k, C_k \rangle) \times Pr[l(U) = w] \right).$$

We are now ready to define the k-FR problem addressed in this paper that incorporates the concepts of reference location and the expected Min-dist criterion following possible world semantics.

Definition 4.3. Given a directed road network G and a set of users U, each of whose movement can be modeled as a set of reference locations, the k **facility relocation (k-FR) problem** aims to find an FR pair $\langle F_k, C_k \rangle$ among a set of existing facilities F and a set of candidate locations C such that

$$\langle F_k, C_k \rangle_{OPT} = argmax_{\langle F_k, C_k \rangle} \Delta(\langle F_k, C_k \rangle).$$

To find the optimal FR pair with the maximum **ED**, we need to calculate $\Delta_w(\langle F_k, C_k \rangle)$ for every $w \in W$, which requires enumerating all possible worlds. Unfortunately, since $|W|$ increases exponentially with $|U|$, it is impractical to directly compute **ED** using Definition 4.2. To address this challenge, we will show how the **ED** computation can be transformed from the aspect of reference locations and can be completed in polynomial time.

Definition 4.4. Let r^w be the reference location of a user u in a possible world w. Then the **expected nearest facility distance** of u with respect to F in all possible worlds W is defined as

$$E\left[d_{nn}(F, L(u))\right] = \sum_{w \in W} \left(d_{nn}\left(F, r^w\right) \times Pr\left[l(U) = w\right] \right).$$

Lemma 4.5. *[3]* $E\left[d_{nn}(F, L(u))\right] = \sum_{r \in L(u)} \left(d_{nn}(F, r) \times p(r)\right)$.

Theorem 4.6. *[3] Given a set of facilities F, a set of users U, and a FR pair $\langle F_k, C_k \rangle$ the **expected change of total distance** in Definition 4.2 can be computed as*

$$\Delta(\langle F_k, C_k \rangle) = \sum_{i=1}^{m} \left(E\left[d_{nn}(F, L(u_i))\right] - E\left[d_{nn}(F \setminus F_k \cup C_k, L(u_i))\right] \right).$$

Given the assumption that reference locations of users are independent with each other, Theorem 4.6 can be derived, which means that we no longer need to enumerate all possible worlds and the complexity of **ED** significantly drops to linear.

5 SOLUTIONS TO k-FR

Intuitively, a brute-force approach to address the problem in Definition 4.3 can be described as follows. Find all possible k-FR pairs $\langle F_p, C_p \rangle$ $(1 \le p \le k)$, and return the best one *i.e.,* with the maximum $\Delta(\cdot)$. Unfortunately, the solution space is up to $\sum_{i=1}^{k} \binom{i}{|F|} \times \binom{i}{|C|}$. Obviously, the well-known k-median problem [6] is a special case of k-FR when $F = \emptyset$, and k-FR is at least as hard as the k-median. Since the k-median problem is NP-Hard [8], it follows that problem in Definition 4.3 is NP-Hard. Therefore, one practical way to address the problem is to find a polynomial solution, either heuristic or approximated.

To this end, [21] present a heuristic solution, referred to as kLNB, which greedily selects the best 1-FR pair[2] one-after-another, in each iteration the best 1-FR pair can be acquired by their proposed solution to the 1-FR problem. However, as the k-FR problem is neither submodular nor non-decreasing, this greedy strategy cannot provide any approximation ratio [15]. We shall prove that in the follows.

Definition 5.1. Consider an arbitrary function $\sigma(\cdot)$ that maps subsets of a finite set P to non-negative real numbers. We say that σ is **submodular** if $\sigma(A \cup \{v\}) - \sigma(A) \ge \sigma(B \cup \{v\}) - \sigma(B)$, for all elements v and all pairs of sets $A \subseteq B \subseteq P$.

Lemma 5.2. *[15] For a non-decreasing, submodular function σ, let S be a set of size k obtained by iteratively selecting the element with the largest marginal increase in the function. Then $\sigma(S) \ge \left(1 - \frac{1}{e}\right) \cdot \sigma(S^*)$, where S^* is the optimal solution; in other words, S provides an $\left(1 - \frac{1}{e}\right)$-approximation ratio.*

For the k-FR problem, where the function to be maximized is $\Delta(\cdot)$, which maps subsets of finite set $F \times C$ to real numbers. But this function does not satisfy non-decreasing or submodular (Example 1 and Figure 2 justify that). Therefore, according to Lemma 5.2, the kLNB algorithm fails to provide the approximation ratio.

Example 1. *In Figure 1 there are four references $\{r_{11}, r_{12}, r_{21}, r_{22}\}$, two current facilities $\{f_1, f_2\}$ and two candidate facilities $\{c_1, c_2\}$. If we want to relocate 2 facilities, kLNB algorithm would find the 1-FR pair $\langle f, c \rangle$ that brings the greatest distance reduction. In the first iteration, all pairs and corresponding distance reductions are shown in Figure 2(a). Obviously, the kLNB will select $\langle f_2, c_2 \rangle$, whose distance reduction is 2. Then for the second iteration, pairs and reductions are shown in Figure 2(b). $\langle f_1, c_1 \rangle$ will be selected. Obviously, the marginal gain in the second iteration (*i.e., *4) is larger than the first (*i.e., *2).*

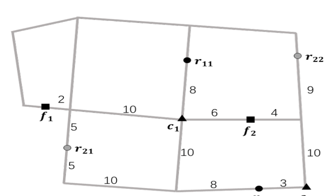


Figure 1: An example of relocation of two facilities

5.1 A Basic Approximate Solution

5.1.1 ***Facility placement problem.*** Before presenting our first approximate solution, we introduce a facility placement problem.

[2]a special case of FR pair $\langle F_k, C_k \rangle$, where $k = 1$

pair	distance reduction
$< f_1, c_1 >$	-2
$< f_1, c_2 >$	0
$< f_2, c_1 >$	-1
$< f_2, c_2 >$	**2**

(a) 1-st iteration

pair	distance reduction
$< f_1, f_2 >$	-2
$< f_1, c_1 >$	**4**
$< c_2, c_1 >$	-3
$< c_2, f_2 >$	-2

(b) 2-nd iteration

Figure 2: Marginal gain in different iterations

Algorithm 1: FR2FP Algorithm

Input: Road network $G(V, E)$, V and E are vertice and edges, respectively; current facility locations F; candidate facility locations C; reference locations R; budget k
Output: Facility relocation (FR) pair $\langle F_k, C_k \rangle$
Let $L = F \cup C$;
The result set $S = \emptyset$;
The number of locations in S that belong to C: $num = 0$;
for i *in 1 to* $|F|$ **do**
    Calculate the total distance $D(l)$ for each location in L if it is selected ;
    select the location l_i whose $D(l)$ is minimum ;
    $S = S \cup l_i$;
    if $l_i \in C$ **then**
        $num = num + 1$;
        if $num == k$ **then**
            $L = L \setminus C$;
        end
    else
        $L = L \setminus \{l_i\}$;
    end
end
return $\langle F \setminus S, C \cap S \rangle$;

Definition 5.3. Given a directed road network $G = (V, E)$, a set of existing facility locations F, a set of candidate facility locations C, and a set of users U, suppose $D_U(F)$ denote the aggregate distance from each user in U to the nearest facility in F, then the goal of **the facility placement problem** is to find $C' \subseteq C$ ($|C'| = k$), so that the total distance reduction, denoted as $\omega_{U,F}(F \cup C') = D_U(F) - D_U(F \cup C')$, is maximized[3].

Theorem 5.4. *$\omega(\cdot)$ in Definition 5.3 is a non-decreasing submodular function.*

Proof. Firstly, $\forall A \subseteq C, \omega(A) \geq 0$. This is obvious, because when a new facility is added, the distance from the user to the nearest facility may only decrease or remain the same.

Secondly, $\forall A \subseteq C, \forall l \in C - A, \omega(A \cup \{l\}) \geq \omega(A)$. The reason is the same as the above. Thus, the function $\omega(\cdot)$ is non-decreasing.

Lastly, suppose $A \subseteq B \subseteq C, \omega(A) \geq 0, \omega(B) \geq 0$
$\forall l, l \in C - B, \omega(A \cup \{l\}) - \omega(A) \geq \omega(B \cup \{l\}) - \omega(B)$
Thus the function is submodular. □

5.1.2 ***Algorithm design.*** As discussed above, k-FR is not submodular or non-decreasing, thus presenting an approximate solution is extremely challenging.

In fact, according to our study in Section 5.1.1, facility placement problem is submodular and non-decreasing. Although the k-FR itself does not satisfy either property, is it possible for us to transform it into a facility placement problem such that both properties can be realized? Motivated by that, we propose to consider the k-FR problem in the perspective of facility placement as follows.

Definition 5.5. Given a directed road network $G = (V, E)$, current facility locations set F, candidate facility locations set C and a set of users U, each of whose movement can be modeled as a set of reference locations, and $\omega(\cdot)$ is the total distance reduction in Definition 5.3, then the k-**FR problem from facility placement perspective** aims to find a set of locations H from $F \cup C$ so that:

$$Maximize \quad \omega(H) \text{ s.t. } |H| = |F|, |H \cap C| \leq k \tag{1}$$

Theorem 5.6. *Let H_{OPT} be the optimal solution to problem[4] 5.5, then it can be classified into two subsets, namely $H^F_{OPT} = H_{OPT} \cap F$ and $H^C_{OPT} = H_{OPT} \cap C$ where $|H^C_{OPT}| \leq k$. Then $\langle F \setminus H^F_{OPT}, H^C_{OPT} \rangle$ is an optimal solution to problem 4.3. That is, the optimal solution of problem 4.3 can be directly acquired from the optimal solution of problem 5.5.*

Proof. We shall prove, through contradiction, that

$$\forall \langle F_k, C_k \rangle, \Delta(\langle F_k, C_k \rangle) \leq \Delta(\langle F \setminus H^F_{OPT}, H^C_{OPT} \rangle).$$

Without loss of generality, suppose $D(\Phi)$ denote the aggregate distance for each user to her nearest facilities in Φ, then

$$\omega(H) = D(F) - D(H), \text{ and } \Delta(\langle F_k, C_k \rangle) = D(F) - D(F \setminus F_k \cup C_k). \tag{2}$$

As H_{OPT} is the optimal solution to problem 5.5, then there is no other H such that $D(H) < D(H_{OPT})$.

Suppose $\exists F_k, C_k$ subject to $F_k \neq F \setminus H^F_{OPT}$ or $C_k \neq H^C_{OPT}$ such that $\Delta(\langle F_k, C_k \rangle) > \Delta(\langle F \setminus H^F_{OPT}, H^C_{OPT} \rangle)$, then.

$$D(F \setminus F_k \cup C_k) < D(F \setminus (F \setminus H^F_{OPT}) \cup H^C_{OPT}) = D(H^F_{OPT} \cup H^C_{OPT}) \tag{3}$$

Based on F_k, C_k, we can construct a candidate solution $H = F \setminus F_k \cup C_k$ to problem 5.5. Taking into account Equation 3, we hereby constructed a solution H to problem 5.5 such that $D(H) < D(H_{OPT})$, which contradict to the fact that there is no other H satisfying $D(H) < D(H_{OPT})$.

Hence, $\langle F \setminus H^F_{OPT}, H^C_{OPT} \rangle$ is an optimal solution to problem 4.3. □

Corollary 1. *Suppose H_{app} is a solution of problem 5.5 with approximation ratio p, i.e., $\omega(H_{app}) \geq p\omega(H_{OPT})$, then $\langle F \setminus H^F_{app}, H^C_{app} \rangle$ is a p-approximate solution to problem 4.3.*

Proof. According to Equation 2, $\omega(H_{app}) = D(F) - D(H^F_{app} \cup H^C_{app})$. Similarly, $\omega(H_{OPT}) = D(F) - D(H^F_{OPT} \cup H^C_{OPT})$. Therefore, $D(F) - D(H^F_{app} \cup H^C_{app}) \geq p(D(F) - D(H^F_{OPT} \cup H^C_{OPT}))$.

Additionally, according to Equation 2,

$$\begin{aligned} \Delta(\langle F \setminus H^F_{app}, H^C_{app} \rangle) &= D(F) - D(F \setminus (F \setminus H^F_{app}) \cup H^C_{app}) \\ &\geq p(D(F) - D(H^F_{OPT} \cup H^C_{OPT})) \\ &= p\Delta(\langle F \setminus H^F_{OPT}, H^C_{OPT} \rangle) \end{aligned}$$

[3]As U, F are fixed in the problem, we shall use $\omega(\cdot), D(\cdot)$ for short in the sequel

[4]We shall use problem x, for short, to refer to the problem defined in Definition x.

facility	total distance
f_1	88
f_2	65
c_1	60
c_2	76

(a) 1-st iteration

facility	total distance
f_1	52
f_2	53
c_2	45

(b) 2-nd iteration

Figure 3: Total distances in iterations according to FR2FP

As $\langle F \backslash H_{OPT}^F, H_{OPT}^C \rangle$ is an optimal solution to problem 4.3 according to Theorem 5.6, then $\langle F \setminus H_{app}^F, H_{app}^C \rangle$ is a p-approximate solution to problem 4.3. □

As problem 5.5 is a special case of facility placement problem. It is both submodular and non-decreasing, such that a hill-climb solution can guarantee the approximation ratio in the results. Accordingly, we are now ready to propose a basic approximation strategy towards both problems, namely FR2FP (Facility Relocation to Facility Placement), showed in Algorithm 1. The algorithm begins by computing the $D(l)$ for each location $l \in F \cup C$ (Line 5), which represents the total distance between all users and his nearest facility after joining the location l. Every time it will select the location l with the minimum $D(l)$ until $|F|$ locations (Line 6). Each time a location $c \in C$ is selected, it will increase the count of locations selected from C. Whenever the number is up to k, it removes all locations in C from L (Lines 8-12).

Example 2. *In Figure 1, FR2FP assumes that there is no current facility, and then gradually selects the facility that can bring the minimum value in $D()$. In the first iteration, the total distance after adding each facility is shown in Figure 3(a). Because the distance after adding c_1 is the smallest, it will be selected as the first facility. Since c_1 belongs to C, we need to judge whether the number of locations selected from C (1 at the current timestamp) has reached k (i.e., 2 in the example). If not, we carry on to select the next facility. According to Figure 3(b), c_2 is the best choice with the minimum total distance 45 in the second iteration. The total number of locations selected now reaches $k = 2$. Thus, FR2FP will stop with the final facility set $\{c_1, c_2\}$. The k-FR pairs selected by FR2FP is $\langle F_2, C_2 \rangle$ ($F_2 = \{f_1, f_2\}, C_2 = \{c_1, c_2\}$).*

5.1.3 ***Theoretical study.*** In this part, we shall theoretically prove that FR2FP guarantees the results quality of problem 4.3. Firstly, we prove the approximation ratio of Algorithm 2 to problem 5.7, as well as the correlation between Algorithm 1 and Algorithm 2. Finally, we obtain the approximation ratio of Algorithm 1 to problem 5.5.

To prove that, we shall firstly remove the constraint of $|H \cap C| \le k$ in Equation 1, which results in a new problem as follows.

Definition 5.7. Given a directed road network $G = (V, E)$, the set of selected locations $\emptyset$, current facility locations set F, candidate facility locations set C and a set of users U, each of whose movement can be modeled as a set of reference locations, the **new problem** aims to find a set of locations H from $F \cup C$ so that:

$$Maximum \quad \omega(H) \text{ s.t. } |H| = |F|$$

As this problem is in fact a facility placement problem defined in Definition 5.3, according to Theorem 5.4, it is also non-decrease and submodular. Suppose the optimal solutions of the problem 5.7

Algorithm 2: Relaxed FR2FP Algorithm

Input: Road network $G(V, E)$, V and E are vertice and edges, respectively; current facility locations F; candidate facility locations C; reference locations R;
Output: Final facility location set $S(|S| = |F|)$

```
Let L = F ∪ C ;
The result set S = ∅ ;
The number of locations in S that belong to C num = 0 ;
for i in 1 to |F| do
    Calculate the total distance D(l) for each location in L if it is
      selected ;
    select the location l_i whose D(l) is minimum ;
    S = S ∪ l_i ;
    L = L \ {l_i} ;
end
return S ;
```

and 5.5 are respective S^* and S, it's clear that

$$\omega\left(S^*\right) \geq \omega\left(S\right) \tag{4}$$

For the new problem 5.7, we can reuse FR2FP generally, except for removing the restriction $|H \cap C| \le k$, such that a relaxed version is proposed in Algorithm 2. It is almost the same as Algorithm 1, except that the returned format is different and the number of selected locations in C is not checked. Therefore, every time it can select the location that brings the most profit (maximum distance reduction or minimum total distance).

According to Lemma 5.2 and our discussion of Definition 5.7, Algorithm 2 can achieve $\left(1 - \frac{1}{e}\right)$ approximation ratio for problem 5.7.

Lemma 5.8. *Let $p_l(H) = \omega(H \cup \{l\}) - \omega(H)$. In i-th iteration, suppose H_i and l_i denote the locations that have been selected beforehand and in the current iteration (*i.e., *i) by Algorithm 2, respectively, then*

$$p_{l_i}(H_i) \geq p_{l_{i+1}}(H_{i+1})$$

Proof. Algorithm 2 selects the location which brings the most profit every time, so $p_{l_i}(H_i) \geq p_{l_{i+1}}(H_i)$.

As the function $\omega(\cdot)$ is submodular, then $p_{l_{i+1}}(H_i) \geq p_{l_{i+1}}(H_{i+1})$.

Finally, we have $p_{l_i}(H_i) \geq p_{l_{i+1}}(H_{i+1})$. □

Lemma 5.9. *For problem 5.7, suppose the facility sets selected by Algorithm 1 and 2 are S_b, S_c, respectively. Then $\omega(S_b) \geq \frac{g}{|F|} \cdot \omega(S_c)$, where g ($g \geq k$) denotes the number of locations eventually selected in $F \cup C$.*

Proof. Suppose the locations selected by Algorithm 2 is $l_1, \ldots, l_{|F|}$ in order. According to Lemma 5.8, $p_{l_1}(H_1) \geq \ldots \geq p_{l_{|F|}}(H_{|F|})$. Result of Algorithm 1 is exactly the same as that obtained by Algorithm 2 before finding k facilities in C. Assuming that the k-th facility in C is found at the g-th iteration ($g \geq k$, because it contains facilities in F), then Algorithm 1 obtains the top g maximums in the result of Algorithm 2. Let the result obtained by the Algorithm 1 be divided into two parts: S_{b1} and S_{b2}. S_{b1} represents the first g locations. Then

$$\omega(S_{b1}) \geq \frac{g}{|F|} \cdot \omega(S_c).$$

As $\omega(S_{b1} \cup S_{b2}) \geq \omega(S_{b1})$, then $\omega(S_b) \geq \frac{g}{|F|} \cdot \omega(S_c)$. □

Algorithm 3: FR2FP-ex Algorithm

Input: Road network $G(V, E)$, V and E are vertice and edges, respectively; current facility locations F; candidate facility locations C; reference locations R; budget k;
Output: Facility relocation (FR) pair $\langle F_k, C_k \rangle$

1 FR pair $\langle F_k, C_k \rangle$ = FR2FP(G, F, C, R);
2 Let $S = F \cup C_k \setminus F_k$;
`// final facility locations selected by FR2FP`
3 Let $L = C \cup F_k \setminus C_k$;
`// all locations not selected by FR2FP`
4 **while** *true* **do**
5 Select the interchange pair $< l, s >$ ($l \in L, s \in S'$) which is the same type and brings the maximum distance reduction ;
6 **if** *the reduction* ≤ 0 **then**
7 break ;
8 **else**
9 $S' = S' \cup \{l\} \setminus \{s\}$;
10 $L = L \cup \{s\} \setminus \{l\}$;
11 **end**
12 **end**
13 return $\langle F \setminus S', S' \cap C \rangle$

Theorem 5.10. *For the original facility relocation problem in Definition 5.5 the FR2FP algorithm showed in Algorithm 1 can achieve $\frac{g}{|F|}\left(1 - \frac{1}{e}\right)$ approximation ratio.*

Proof. According to Lemma 5.2 and 5.9

$$\omega(S_b) \geq \frac{g}{|F|}\left(1 - \frac{1}{e}\right) \cdot \omega(S^*)$$

In addition, as $\omega(S^*) \geq \omega(S)$, then

$$\omega(S_b) \geq \frac{g}{|F|}\left(1 - \frac{1}{e}\right) \cdot \omega(S)$$

□

According to Theorem 5.10 as well as Corollary 1, FR2FP can achieve $\frac{g}{|F|}\left(1 - \frac{1}{e}\right)$ approximation ratio for problem 4.3.

Moreover, if we use the greedy algorithm to find the solution of a non-decreasing submodular function, there is an another theorem.

Theorem 5.11. *[15] For the non-decreasing submodule set function Z defined on the set N and the objective function $\max_{S \subseteq N}\{Z(S) : |S| \leq k\}$, if the greedy algorithm is used to solve the problem and it stops at step $k' \leq k$, then greedy algorithm get the optimal solution.*

According to Theorem 5.11, if Algorithm 2 has reached the maximum distance reduction after finding $n(n < |F|)$ facilities, then the solution is the optimal solution. Then the approximate ratio of Algorithm 1 will correspondingly become $\frac{g}{|F|}$.

5.2 An Advanced Approximate Solution

The approximation rate of the FR2FP algorithm is $\frac{g}{|F|}\left(1 - \frac{1}{e}\right)$, where g is lower bounded by k. Only when k is big enough, the approximation ratio of the FR2FP is high (empirically justified in 6). To avoid this problem, we propose an advanced algorithm, namely FR2FP-ex (FR2FP with exchange). The idea is to add an exchange after getting the results from FR2FP. The pseudo code is shown in Algorithm 3. Assuming that all facilities are classified into two groups according to whether they originally belonged to F or C, then the facilities in S also belong to both groups accordingly. Let L be all the facilities except S, which also belong to either F or C. The exchange is performed iteratively. Each time it selects the exchange pair $\langle l, s \rangle$ that brings the greatest distance reduction (Line 5). Both facilities in the pair must belong to the same group (both in F or C), which ensures the number of facilities in C will not exceed k. Whenever the current exchange pair does not cause the total distance to decrease, the algorithm stops (Lines 6-7).

The idea of the exchange is not complicated, but when k is small, it shows significant improvement comparing the FR2FP. Moreover the FR2FP-ex is executed based on the results of FR2FP, the results it obtains is at least as good as that of the FR2FP. Therefore, FR2FP-ex still guarantees the approximate ratio.

Table 1: Real facilities of CA and BJ

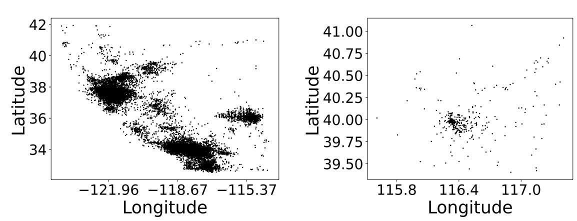
CA	$\|F\|$	BJ	$\|F\|$
Hospital	1000	Bank	1000
Park	1000	Cafe	1000
Post Office	1000	Logistic	1000
School	1000	Gas Station	1000

6 EXPERIMENTS

In this section, we empirically evaluate the performance of the proposed approximate solutions.

6.1 Experimental Setup

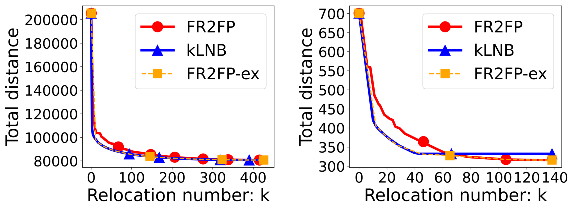
Datasets. Two real-world datasets, California (CA)[10] and Beijing (BJ), are adopted in our experiment. Both of them are get from the author of [21]. There are 21,693 bidirectional edges and 21,047 vertices in CA. BJ consists of 433,391 unidirectional edges and 171,504 vertices. The type and number of facilities in each dataset are shown in Table 1. Notably, there are many type of facilities in each dataset, we randomly select 4 types from each dataset in our empirical study. The users in BJ dataset is available from[31]. Each user has 136,686 sample points on average. For CA, we adopt a real-world check-in data[5]. The real data are all clustered and distributed in a certain area as shown in Figure 4. Hence, we constructed user data evenly distributed on the road network and the probability of reference locations of each user is generated randomly. Candidate facilities are constructed in two ways[6]: 1) uniformly randomly generated geographically; 2) constructed according to the distribution of users. The former is used by default.



(a) CA (b) BJ

Figure 4: User distribution in the two datasets

[5] Obtained from http://snap.stanford.edu/data/.
[6] For the objects that are not exactly located on roads, we shift them to the closest point of on the road network.



(a) CA:1000 park (b) BJ:1000 bank

Figure 5: Effectiveness comparison

Compared solutions. We report the performance of the following three algorithms, which are implemented in C++ and tested on a 3.2GHz quad-core machine with 16G RAM.

- kLNB: The heuristic solution of [21], which iteratively select the optimal 1-FR pair for k times.
- FR2FP: The approximate solution proposed in Algorithm 1.
- FR2FP-ex: The advanced solution proposed in Algorithm 3.

6.2 Experimental Results

Comparison of the solutions. Firstly, we vary the size of F from 100 to 1000 to test the performance of the three approaches. The phenomenon found in all the tests are similar, so for each dataset we only select one type of facility, shown in Figure 5. The x-axis and y-axis respectively represent k and the aggregated distance from the user to the nearest facility. From the figure, we can see that when k is small, the performance of FR2FP is the worst, even inferior to kLNB. This is mainly caused by the following two reasons:

- The approximate ratio of FR2FP is $\frac{g}{|F|}\left(1-\frac{1}{e}\right)$. This approximate rate is closely related to the value of g, which is lower bounded by k. When k is small, g has small lower bound, leading to a low approximation ratio.
- Although the kLNB does not provide any guarantee, if only one facility is relocated, this algorithm can guarantee the optimal solution.

As k continues to increase, the performance of FR2FP gets better and better. According to the empirical results of FR2FP algorithm, when k is large enough, the value of g approaches $|F|$, which means that the approximate rate of FR2FP is nearly $\left(1-\frac{1}{e}\right)$.

There is another interesting phenomenon: for the BJ dataset, with the increase of k, FR2FP is ultimately better than kLNB. In comparison, in CA both solutions finally reached the same effectiveness. We shall discuss our insight of this difference in the follows.

As the number of reference locations in the BJ dataset is only 400, there exist the following facts during the execution of FR2FP. The FR2FP algorithm originally needs to select $|F|$ facilities. But when $n(n \leq |F|)$ facilities are selected, even if more facilities are added, the total distance won't decrease anymore. That is, the marginal benefit becomes 0. According to Theorem 5.11, when the marginal benefit is 0, the approximate ratio of FR2FP is $\frac{g}{|F|}$. With the increase of k, g is close to $|F|$. In the CA dataset, the number of reference points is up to 22227, and the above scenario will not occur. Thus, the approximate ratio of the FR2FP is still $\frac{g}{|F|}(1-\frac{1}{e})$. Therefore, in the BJ dataset, the performance of FR2FP is better than kLNB when k is large enough.

We can also find that FR2FP-ex consistently performs the best among the three. Notably, as response time is not a key factor in this mining task, and there is no significant difference in the running time for the solutions empirically, we select not to report them in detail due to limit of space.

Effect of facility distribution. In order to investigate whether the distribution of facilities F will affect the performance of the algorithms, we randomly select 2 types of facilities in both datasets, *i.e.,* the cafe and station of BJ and the park and school of CA. The distributions of the facilities are shown in Figure 6. The performance for the three solutions accordingly are shown in Figure 7. Compared with Figure 5, we can see that the distribution of facilities has negligible effect on the performance of the algorithms.

Effect of users distribution. Besides, we also test whether the distribution of users will affect the performance of the algorithms. To this end, we randomly generate a series of uniformly distributed user points while keeping all the facilities, candidates and road networks the same as the above experiments. Since the experimental results are similar, we select to show two of them, in Figure 8, due to the limit of space. We can found the result is similar to Figure 5, *i.e.,* FR2FP-ex performs the best, FR2FP and kLNB dominate each other depending on k.

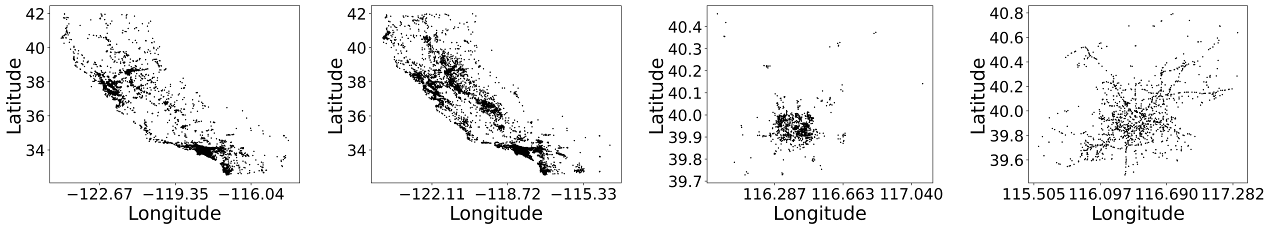
Effect of candidate distribution. Finally, we also propose to test the effect of candidate distribution. To this end, we vary the candidate facility distribution according to 1) uniform distribution; 2) clustered distribution according to the density of users, respectively. By comparing between the results *w.r.t.* both distributions, there is no oblivious difference. That is, the candidate distribution has negligible effect on the performance of the three solutions.

7 EXTENSIVE STUDY OF kLNB IN [21]

Through the above experiments, we found that although the kLNB does not have an approximate ratio, its performance is surprisingly superior to the FR2FP solution when k is small. However, as kLNB fails to guarantee the results quality while FR2FP does, there should exist cases that kLNB fails to work. Driven by that, we conduct extensive study, and finally find the cases that kLNB fail to perform well: *when multiple facilities need to be relocated at the same time, and the relocation of anyone of the facilities alone cannot bring the distance reduction.*

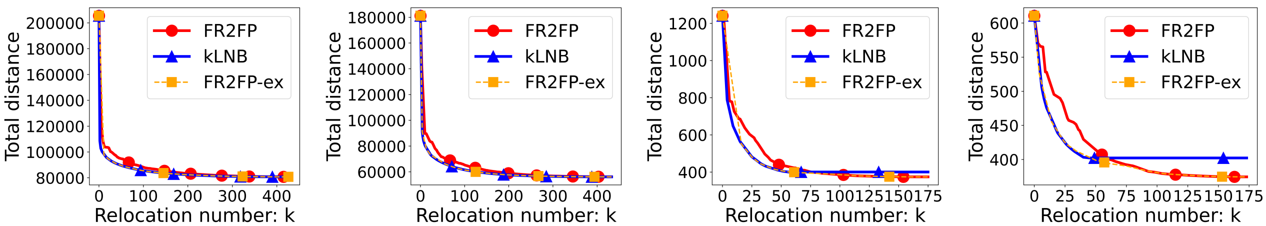
Figure 10 shows an example of this scenario. For ease of discussion, we assume that Euclidean distance, which does not affect our conclusion in this example. According to the distance (the blue number) marked on the figure, we can calculate the distance between each user and the facilities, shown in Table 2. Suppose we are relocating two facilities, according to Table 3, no matter which pair is selected, it cannot bring a positive distance reduction. Therefore, kLNB will output nothing. Nevertheless, if we relocate f_1 and f_2 at the same time, the distance will decrease eventually. Suppose FR2FP is adopted in this scenario, it will selects c_1 first, and then c_2, which is obviously better than kLNB.

Although the kLNB algorithm does not have an approximate rate guarantee, it performs good in many cases, even better than FR2FP when k is small. However, it may encounter some awkward situations as shown in Figure 10. Due to that, we recommend to use FR2FP-ex algorithm to solve the k-FR problem, as it not only



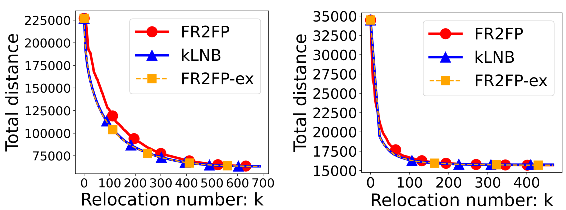
(a) CA park (b) CA school (c) BJ cafe (d) BJ station

Figure 6: Distribution of different facilities



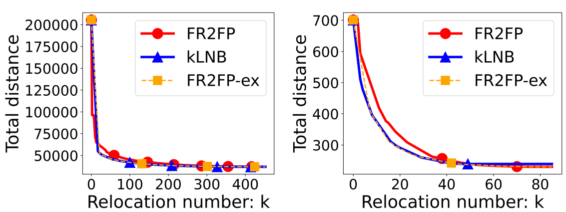
(a) CA park (b) CA school (c) BJ cafe (d) BJ station

Figure 7: Effect of facilities distribution



(a) CA:1000 park (b) BJ:1000 bank

Figure 8: Results of evenly distributed users



(a) CA:1000 park (b) BJ:1000 bank

Figure 9: Effect of candidates distribution

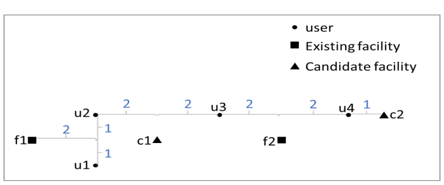


Figure 10: A failure scenario of kLNB

has the same result quality guarantee as FR2FP, but also exhibits the best practical performance among the solutions.

8 CONCLUSION

Although k-FR problem has a wide range of applications in real life, the NP-hardness makes it difficult to find an algorithm reliable in real application. In this paper, we proposed a pair of approximate solution to k-FR by transforming the problem into an equivalent facility placement one. To the best of our knowledge, they are the first approximate solutions. We also found that the state-of-the-art heuristic solution, kLNB proposed by [21], is effective in most cases, although it does not provide an approximation ratio. Results of exhaustive experiments show that the FR2FP-ex algorithm is currently the best method for solving k-FR problem. We also extensively show the specific scenarios where kLNB fail to find valid solutions.

Table 2: Distances between users and facilities of Figure 10

user \ facility	f_1	f_2	c_1	c_2
u_1	$\sqrt{5}$	$\sqrt{37}$	$\sqrt{5}$	$\sqrt{85}$
u_2	$\sqrt{5}$	$\sqrt{37}$	$\sqrt{5}$	9
u_3	$\sqrt{37}$	$\sqrt{5}$	$\sqrt{5}$	5
u_4	$\sqrt{101}$	$\sqrt{5}$	$\sqrt{37}$	1

Table 3: Distance reductions for all 1-FR pairs in Figure 10

pairs	reductions
$\langle f_1, c_1\rangle$	0
$\langle f_1, c_2\rangle$	$3\sqrt{5} - 1 - 2\sqrt{37} < 0$
$\langle f_2, c_1\rangle$	$\sqrt{5} - \sqrt{37} < 0$
$\langle f_2, c_2\rangle$	$2\sqrt{5} - 6 < 0$

ACKNOWLEDGMENTS

This work is supported by National Natural Science Foundation of China (No. 61972309), CCF-Huawei Database System Innovation Research Plan (No. 2020010B), Key Scientific Research Program of Shaanxi Provincial Department of Education (No. 20JY014), National Natural Science Foundation of China (No. 61976168) and Natural Science Basic Research Program of Shaanxi (No. 2020JM-575).

REFERENCES

[1] Muhammad Aamir Cheema, Xuemin Lin, Wei Wang, Wenjie Zhang, and Jian Pei. 2009. Probabilistic reverse nearest neighbor queries on uncertain data. *IEEE Transactions on Knowledge and Data Engineering* 22, 4 (2009), 550–564.

[2] Zitong Chen, Yubao Liu, Raymond Chi-Wing Wong, Jiamin Xiong, Ganglin Mai, and Cheng Long. 2014. Efficient algorithms for optimal location queries in road networks. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. 123–134.

[3] Jiangtao Cui, Meng Wang, Hui Li, and Yang Cai. 2018. Place Your Next Branch with MILE-RUN: Min-dist Location Selection over User Movement. *Information Sciences* 463-464 (2018), 1–20.

[4] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Mauro Sozio. 2015. Scalable Facility Location for Massive Graphs on Pregel-like Systems. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*. ACM, 273–282.

[5] Parisa Ghaemi, Kaveh Shahabi, John P Wilson, and Farnoush Banaei-Kashani. 2010. Optimal network location queries. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. 478–481.

[6] S Louis Hakimi. 1965. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations research* 13, 3 (1965), 462–475.

[7] Russell Halper, S Raghavan, and Mustafa Sahin. 2015. Local search heuristics for the mobile facility location problem. *Computers & Operations Research* 62 (2015), 210–223.

[8] O. Kariv and S. Hakimi. 1979. An Algorithmic Approach to Network Location Problems. II: The p-Medians. *Siam Journal on Applied Mathematics* 37 (1979), 539–560.

[9] Thomas Lambert, David Guyon, and Shadi Ibrahim. 2020. Rethinking Operators Placement of Stream Data Application in the Edge. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*. ACM, 2101–2104.

[10] Feifei Li, Dihan Cheng, Marios Hadjieleftheriou, George Kollios, and Shang-Hua Teng. 2005. On trip planning queries in spatial databases. In *International symposium on spatial and temporal databases*. Springer, 273–290.

[11] Yuhong Li, Yu Zheng, Shenggong Ji, Wenjun Wang, Leong Hou U, and Zhiguo Gong. 2015. Location selection for ambulance stations: a data-driven approach. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 1–4.

[12] Zhenhui Li, Bolin Ding, Jiawei Han, Roland Kays, and Peter Nye. 2010. Mining periodic behaviors for moving objects. (2010), 1099–1108.

[13] Gino J Lim and Ayse Durukan Sonmez. 2013. γ-Robust facility relocation problem. *European Journal of Operational Research* 229, 1 (2013), 67–74.

[14] Kyriakos Mouratidis, Dimitris Papadias, and Spiros Papadimitriou. 2005. Medoid queries in large spatial databases. In *International Symposium on Spatial and Temporal Databases*. Springer, 55–72.

[15] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. 1978. An analysis of approximations for maximizing submodular set functions—I. *Mathematical programming* 14, 1 (1978), 265–294.

[16] Dimitris Papadias, Yufei Tao, Kyriakos Mouratidis, and Chun Kit Hui. 2005. Aggregate nearest neighbor queries in spatial databases. *ACM Transactions on Database Systems (TODS)* 30, 2 (2005), 529–576.

[17] Jianzhong Qi, Rui Zhang, Lars Kulik, Dan Lin, and Yuan Xue. 2012. The min-dist location selection query. In *2012 IEEE 28th International Conference on Data Engineering*. IEEE, 366–377.

[18] Jianzhong Qi, Rui Zhang, Yanqiu Wang, Andy Yuan Xue, Ge Yu, and Lars Kulik. 2014. The min-dist location selection and facility replacement queries. *World Wide Web* 17, 6 (2014), 1261–1293.

[19] Ayse Durukan Sonmez and Gino J Lim. 2012. A decomposition approach for facility location and relocation problem with uncertain number of future facilities. *European Journal of Operational Research* 218, 2 (2012), 327–338.

[20] Feng Wang, Li Chen, and Weike Pan. 2016. Where to Place Your Next Restaurant?: Optimal Restaurant Placement via Leveraging User-Generated Reviews. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*. ACM, 2371–2376.

[21] Meng Wang, Hui Li, Jiangtao Cui, Sourav S. Bhowmick, and Ping Liu. 2020. FROST: Movement History–Conscious Facility Relocation. *ACM Trans. Intell. Syst. Technol.* 11, 1, Article 9 (Jan. 2020), 26 pages.

[22] Meng Wang, Hui Li, Jiangtao Cui, Ke Deng, Sourav S Bhowmick, and Zhenhua Dong. 2016. Pinocchio: Probabilistic influence-based location selection over moving objects. *IEEE Transactions on Knowledge and Data Engineering* 28, 11 (2016), 3068–3082.

[23] Qian Wang, Rajan Batta, Joyendu Bhadury, and Christopher M Rump. 2003. Budget constrained location problem with opening and closing of facilities. *Computers & Operations Research* 30, 13 (2003), 2047–2069.

[24] Brian J Worton. 1989. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* 70, 1 (1989), 164–168.

[25] Xiaokui Xiao, Bin Yao, and Feifei Li. 2011. Optimal location queries in road network databases. In *2011 IEEE 27th international conference on data engineering*. IEEE, 804–815.

[26] Da Yan, Zhou Zhao, and Wilfred Ng. 2011. Efficient algorithms for finding optimal meeting point on road networks. *Proceedings of the VLDB Endowment* 4, 11 (2011), 968–979.

[27] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. 2011. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. 325–334.

[28] Man Lung Yiu, Nikos Mamoulis, and Dimitris Papadias. 2005. Aggregate nearest neighbor queries in road networks. *IEEE transactions on knowledge and data engineering* 17, 6 (2005), 820–833.

[29] Donghui Zhang, Yang Du, Tian Xia, and Yufei Tao. 2006. Progressive computation of the min-dist optimal-location query. In *Proceedings of the 32nd international conference on Very large data bases*. 643–654.

[30] Yu Zhang, Wei Wei, Binxuan Huang, Kathleen M. Carley, and Yan Zhang. 2017. RATE: Overcoming Noise and Sparsity of Textual Features in Real-Time Location Estimation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*. ACM, 2423–2426.

[31] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. 2009. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web*. 791–800.

[32] Liang Zhu, Yinan Jing, Weiwei Sun, Dingding Mao, and Peng Liu. 2010. Voronoi-based aggregate nearest neighbor query processing in road networks. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 518–521.