SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •
Special Focus on Theory and Applications of Models of Computation

April 2022, Vol. 65 140601:1–140601:9
https://doi.org/10.1007/s11432-021-3411-7

An approximation algorithm for lower-bounded k-median with constant factor

Xiaoliang WU[1,2], Feng SHI[1,2*], Yutian GUO[1,2], Zhen ZHANG[3], Junyu HUANG[1,2] & Jianxin WANG[1,2]

[1]*School of Computer Science and Engineering, Central South University, Changsha* 410083*, China;*
[2]*Hunan Provincial Key Lab on Bioinformatics, Central South University, Changsha* 410083*, China;*
[3]*School of Frontier Crossover Studies, Hunan University of Technology and Business, Changsha* 410083*, China*

Received 12 March 2021/Revised 15 October 2021/Accepted 13 January 2022/Published online 14 March 2022

Abstract The lower-bounded k-median problem plays a key role in many applications related to privacy protection, which requires that the amount of assigned client to each facility should not be less than the requirement. Unfortunately, the lower-bounded clustering problem remains elusive under the widely studied k-median objective. Within this paper, we convert this problem to the capacitated facility location problem and successfully give a $(516+\epsilon)$-approximation for this problem.

Keywords approximation algorithm, k-median, lower-bounded k-median

Citation Wu X L, Shi F, Guo Y T, et al. An approximation algorithm for lower-bounded k-median with constant factor. Sci China Inf Sci, 2022, 65(4): 140601, https://doi.org/10.1007/s11432-021-3411-7

1 Introduction

Clustering is a fundamental problem [1–8] in the field of theoretical computer science. Given a set of points, its goal is to partition the point-set into several disjoint subsets (called clusters in the remaining text) such that the points of the same subsets are more similar to each other. In the paper, we center on the k-median problem, which is one of the most frequently encountered clustering problems. An instance of k-median consists of a facility set $\mathcal{F}$ and a client set $\mathcal{C}$ in a metric space with a distance function d and a non-negative integer k. The goal aims to identify a subset $F \subseteq \mathcal{F}$ with $|F| \leqslant k$, such that the cost $\sum_{j\in\mathcal{C}} d(j,F)$ is as small as possible, where $d(j,F) = \min_{f\in F} d(j,f)$. Obviously, the subset F partitions $\mathcal{C}$ with $|F|$ clusters by $d(j,F)$. The problem is known to be NP-hard [9] and appeals to lots of interests in developing its approximation algorithms [10–16]. On the basis of the primal-dual approach given in [15], Byrka et al. [10] obtained an algorithm with ratio $(2.675+\epsilon)$, which is the best-known approximation ratio.

In the setting of the classic clustering problem, once the set of the opened facility is determined, each client is assigned to its closest opened facility by default. Unfortunately, the clustering applications related to privacy protection in real-world often involve a conception of lower bound with the facilities, which requires that the amount of assigned clients to each facility has tobe greater or equal to the corresponding lower bound [17, 18]. For example, if the size of the cluster is not large enough, then the clients in the cluster can be identified by the attributes of the cluster, and thus increasing the size of the cluster can protect the privacy of inner clients. To handle such applications, Karger and Minkoff [19] and Guha et al. [20] proposed the lower-bounded facility location problem that associates a lower bound of the facilities (i.e., the facilities share the same lower bound) and presented several bi-criteria approximation algorithms for the problem. These algorithms yield a solution whose cost is at most a constant factor of the optimal solution but violate the lower bound constraint of the facilities by a constant factor. Then Svitkina [21] presented a $(448+\epsilon)$-approximation algorithm based on the techniques given in [19, 20]. Afterwards Ahmadian and Swamy [22] improved the ratio to $82.6+\epsilon$. For the setting of non-uniform

* Corresponding author (email: fengshi@csu.edu.cn)

lower bounds (i.e., the facilities have different lower bounds), Li [23] presented a $(3926+\epsilon)$-approximation algorithm.

Within this investigation, we give the lower-bounded k-median problem.

Definition 1 (Lower-bounded k-median problem (LBK))**.** The input consists of a client set $\mathcal{C}$ and a facility set $\mathcal{F}$ in a metric space with a distance function d, a positive integer bound k and a lower bound B. The goal is to find a facility subset $F \subseteq \mathcal{F}$ with $|F| \leqslant k$ and a mapping $\phi : \mathcal{C} \to F$ so as $|\{j \in \mathcal{C} \mid \phi(j) = i\}| \geqslant B$ for each facility $i \in F$, and the cost $\sum_{j\in\mathcal{C}} d(j, \phi(j))$ is as small as possible.

Considering an instance $\mathcal{I} = (\mathcal{C}, \mathcal{F}, k, d, B)$ of LBK, a feasible solution of $\mathcal{I}$ is a pair (F, ϕ), where F is a subset of $\mathcal{F}$ with $|F| \leqslant k$ and ϕ is a mapping with $|\{j \in \mathcal{C} \mid \phi(j) = i\}| \geqslant B$ for each $i \in F$ (to simplify the notation, let $\phi^{-1}(i) = \{j \in \mathcal{C} \mid \phi(j) = i\}$). In d-dimension Euclidean space, Ding and Xu [24] showed that LBK admits an algorithm with ratio $(1+\epsilon)$ with running time $O(n^2 d \cdot (\log n)^{k+2} \cdot 2^{\mathrm{poly}(k/\epsilon)})$. Later Bhattacharya et al. [25] presented a faster algorithm with running time $O(n^2 d \cdot (\log n)^2 \cdot (\frac{k}{\epsilon})^{O(k/\epsilon)})$ and the approximation factor $(1 + \epsilon)$. Note that neither of the above two algorithms has polynomial running time. Ahmadian and Swamy [26] showed that LBK has a polynomial-time $O(1)$-approximation algorithm, but the approximation ratio is very large. For the setting of non-uniform lower bounds, Feng et al. [27] gave a $(3 + \epsilon)$-approximation algorithm with running time $(k\epsilon)^{O(k)} n^{O(1)}$. As the opposite of LBK, the capacitated k-median problem was proposed by Demirci and Li [28] who required that the amount of clients assigned to each opened facility is at most a given bound. They presented an algorithm with ratio $O(1)$ for this problem, but the solution obtained by the algorithm violates the capacities by $(1 + \epsilon)$. Then Li [15] obtained the first algorithm with ratio $O(1)$ that allows $(1 + \epsilon)k$ facilities can be opened without violating the capacity.

Theorem 1. There is a $(516 + \epsilon)$-approximation algorithm for LBK with polynomial runtime.

Given an LBK instance $\mathcal{I} = (\mathcal{C}, \mathcal{F}, k, d, B)$, clearly, if $|\mathcal{C}| < B$, then there exists no feasible solution to the instance. Therefore, our default lower bound B is strictly less than the amount of the client set $\mathcal{C}$. The general idea of our algorithm consists of two steps. Firstly, give a bi-criteria approximation algorithm for the problem, which obtains a solution whose cost is bounded by a constant factor in approximation guarantee but violates the lower bound constraint mildly. Secondly, convert the obtained bi-criteria approximate solution of $\mathcal{I}$ to a feasible one by closing some facilities and reassigning some clients, to meet the lower bound constraint. Obviously, the conversion causes a loss in the cost. Thus we reduce the problem to the capacitated facility location problem and successfully show that the loss can be bounded by a constant factor in approximation guarantee.

Definition 2 (Capacitated facility location problem (CFL))**.** The input consists of a client set $\mathcal{C}$ and a facility set $\mathcal{F}$ in a metric space with a distance function d, a positive integer bound k and an upper bound U. The goal is to find a facility subset $F \subseteq \mathcal{F}$ and a mapping $\phi : \mathcal{C} \to F$ so as $|\{j \in \mathcal{C} \mid \phi(j) = i\}| \leqslant u_i$ for each facility $i \in F$, and the sum of the connection cost $\sum_{j\in\mathcal{C}} d(j, F)$ and the opening cost $\sum_{i\in F} f_i$ is as small as possible, where $d(j, F) = \min_{f\in F} d(j, f)$, and f_i and u_i are the opening cost and capacity of facility $i \in \mathcal{F}$, respectively.

2 A bi-criteria approximate solution

In this section, we first show the close relationship between LBK and the k-facility location problem, where the formulation of the k-facility location problem is given below. Then based on the relationship, a bi-criteria approximation algorithm for LBK can be given by calling a known approximation algorithm for the k-facility location problem.

Definition 3 (k-facility location problem (KFL))**.** The input consists of a client set $\mathcal{C}$ and a facility set $\mathcal{F}$ in a metric space with a distance function d, and a positive integer bound k. The goal is to look for a facility subset $F \subseteq \mathcal{F}$ with $|F| \leqslant k$, such that the sum of the connection cost $\sum_{j\in\mathcal{C}} d(j, F)$ and the opening cost $\sum_{i\in F} f_i$ is minimized, where $d(j, F) = \min_{f\in F} d(j, f)$ and f_i is the opening cost of facility $i \in \mathcal{F}$.

Consider an instance $\mathcal{I} = (\mathcal{C}, \mathcal{F}, k, d, B)$ of LBK and a solution (F, ϕ) of $\mathcal{I}$. Let $\Gamma_{\mathcal{I}}(F, \phi)$ denote the cost of the solution (F, ϕ) (i.e., $\Gamma_{\mathcal{I}}(F, \phi) = \sum_{j\in\mathcal{C}} d(j, \phi(j))$). If each facility i of F is assigned at least αB clients with respect to ϕ, where $\alpha \in (0, 1)$, then (F, ϕ) is called an α-covered solution. Let $T_i \subseteq \mathcal{C}$ denote the set containing the B clients nearest to i for each facility $i \in \mathcal{F}$.

An instance $\mathcal{I}' = (\mathcal{C}, \mathcal{F}, k, d, f)$ of KFL corresponding to $\mathcal{I}$ is constructed by assigning opening cost to

the facilities and omitting the lower bound B; more specifically, each facility $i \in \mathcal{F}$ is entitled to an opening cost $f_i = \frac{2\alpha}{1-\alpha}\sum_{j\in T_i} d(i,j)$. Then a solution F' of $\mathcal{I}'$ is obtained by calling the 3.25-approximation algorithm given in [29]. Let $\Gamma_{\mathcal{I}'}(F')$ denote the cost of F', i.e., $\Gamma_{\mathcal{I}'}(F') = \sum_{j\in\mathcal{C}} d(j,F') + \sum_{i\in F'} f_i$.

In the following, we show that any solution of $\mathcal{I}'$ can be transformed into an α-covered solution of $\mathcal{I}$.

Lemma 1. If there is a feasible solution (F,ϕ) of $\mathcal{I}$, then F is a solution of $\mathcal{I}'$ with

$$\sum_{i\in F} f_i \leqslant \frac{2\alpha}{1-\alpha}\Gamma_{\mathcal{I}}(F,\phi).$$

Proof. For any facility $i \in F$, $|\phi^{-1}(i)| \geqslant B = |T_i|$ due to the feasibility of (F,ϕ). Combining the inequality and the definition of T_i gives that $\sum_{j\in\phi^{-1}(i)} d(i,j) \geqslant \sum_{j\in T_i} d(i,j)$. Consequently, we know that

$$\begin{aligned}\sum_{i\in F} f_i &= \sum_{i\in F}\left(\frac{2\alpha}{1-\alpha}\sum_{j\in T_i} d(i,j)\right)\\ &\leqslant \frac{2\alpha}{1-\alpha}\sum_{i\in F}\sum_{j\in\phi^{-1}(i)} d(i,j)\\ &= \frac{2\alpha}{1-\alpha}\sum_{j\in\mathcal{C}} d(j,\phi(j))\\ &= \frac{2\alpha}{1-\alpha}\Gamma_{\mathcal{I}}(F,\phi),\end{aligned}$$

where the first inequality holds due to $|\phi^{-1}(i)| \geqslant |T_i|$ for any $i \in F$.

For any feasible solution (F,ϕ) of instance $\mathcal{I}$, by Lemma 1 and $\Gamma_{\mathcal{I}'}(F) = \sum_{i\in F} f_i + \Gamma_{\mathcal{I}}(F,\phi)$, we know that $\Gamma_{\mathcal{I}'}(F) \leqslant \frac{1+\alpha}{1-\alpha}\Gamma_{\mathcal{I}}(F,\phi)$.

Lemma 2. Based on a solution F' of $\mathcal{I}'$, an α-covered solution (F,ϕ) of $\mathcal{I}$ can be obtained in polynomial time with $\Gamma_{\mathcal{I}}(F,\phi) \leqslant \Gamma_{\mathcal{I}'}(F')$.

Proof. A k-median problem instance $\mathcal{I}^k$ is constructed by omitting the opening cost of facilities on the basis of instance $\mathcal{I}'$. Now we construct a solution (F,ϕ) of $\mathcal{I}$ as follows. Firstly, let $F'' = F'$. Secondly, removing the facility $i \in F''$ from F'' if $\Gamma_{\mathcal{I}'}(F''\backslash\{i\}) \leqslant \Gamma_{\mathcal{I}'}(F'')$, i.e., $F'' = F''\backslash\{i\}$. Thirdly, repeat the second step until it is not applicable. Finally, let $F = F''$ and ϕ be the mapping that assigns each client $j \in \mathcal{C}$ to its closest facility in F. The resulted solution (F,ϕ) is a minimal feasible solution of $\mathcal{I}^k$. Let $\Gamma_{\mathcal{I}^k}(F,\phi)$ be the cost of solution (F,ϕ). Obviously, $\Gamma_{\mathcal{I}^k}(F,\phi) \leqslant \Gamma_{\mathcal{I}'}(F)$ (as facilities of $\mathcal{I}^k$ have no opening cost), implying that $\Gamma_{\mathcal{I}^k}(F,\phi) \leqslant \Gamma_{\mathcal{I}'}(F) \leqslant \Gamma_{\mathcal{I}'}(F')$.

We now show that the solution (F,ϕ) is an α-covered solution of $\mathcal{I}$, i.e., at least αB clients are assigned to i for each $i \in F$. Assume that there is a facility $i \in F$ with $|\phi^{-1}(i)| < \alpha B$, which implies that

$$|T_i\backslash\phi^{-1}(i)| \geqslant (1-\alpha)B. \tag{1}$$

Thus there is a client $j' \in T_i\backslash\phi^{-1}(i)$, such that

$$d(i,j') \leqslant \frac{1}{|T_i\setminus\phi^{-1}(i)|}\sum_{j\in T_i\backslash\phi^{-1}(i)} d(i,j) \leqslant \frac{1}{(1-\alpha)B}\sum_{j\in T_i\backslash\phi^{-1}(i)} d(i,j) \leqslant \frac{1}{(1-\alpha)B}\sum_{j\in T_i} d(i,j), \tag{2}$$

where the second inequality follows from inequality (1). Note that the client j' is not assigned to facility i but another one $i' \in F$ with $i' = \phi(j)$ and the distance from j' to i' is no more than the distance from j' to i. Consequently, we know

$$\begin{aligned}\sum_{j\in\phi^{-1}(i)} d(j,i') &\leqslant \sum_{j\in\phi^{-1}(i)} (d(j,i) + d(i,j') + d(j',i')) \quad \triangleright\text{by the triangle inequality}\\ &\leqslant |\phi^{-1}(i)| \times 2d(i,j') + \sum_{j\in\phi^{-1}(i)} d(j,i) \quad \triangleright\text{by } d(j',i') \leqslant d(i,j')\\ &\leqslant \alpha B \times \frac{2}{(1-\alpha)B}\sum_{j\in T_i} d(i,j) + \sum_{j\in\phi^{-1}(i)} d(j,i) \quad \triangleright\text{by inequality (2)}\end{aligned}$$

$$= \frac{2\alpha}{(1-\alpha)} \sum_{j \in T_i} d(i,j) + \sum_{j \in \phi^{-1}(i)} d(j,i).$$

If the facility i is closed and all clients of $\phi^{-1}(i)$ are assigned to i', then the opening cost is decreased by f_i and the connection cost is increased by at most $\frac{2\alpha}{(1-\alpha)} \sum_{j \in T_i} d(i,j) = f_i$. Thus we know $\Gamma_{\mathcal{I}'}(F \backslash \{i\}) \leqslant \Gamma_{\mathcal{I}'}(F)$, contradicting that (F, ϕ) is a minimal feasible solution of $\mathcal{I}^k$.

Combining Lemmas 1 and 2, we have the following theorem.

Theorem 2. There is a $(3.25\frac{1+\alpha}{1-\alpha}, \alpha)$ bi-criteria approximation algorithm for LBK, where the cost of the returned solution is at most $3.25\frac{1+\alpha}{1-\alpha}$ times the optimal cost and the lower bound constraint is violated by the factor α.

Proof. Consider an instance $\mathcal{I}$ of LBK and the instance $\mathcal{I}'$ of KFL constructed by the way given before. By the 3.25-approximation algorithm for KFL [29], we have a solution F' of $\mathcal{I}'$. Let (F^*, ϕ^*) and $F^\dagger$ be the optimal solutions of $\mathcal{I}$ and $\mathcal{I}'$, respectively. Then we have that $\Gamma_{\mathcal{I}'}(F') \leqslant 3.25\Gamma_{\mathcal{I}'}(F^\dagger) \leqslant 3.25\Gamma_{\mathcal{I}'}(F^*)$. Thus,

$$\begin{aligned}
\Gamma_{\mathcal{I}'}(F') &\leqslant 3.25\Gamma_{\mathcal{I}'}(F^*) \\
&\leqslant 3.25 \left(\sum_{i \in F^*} f_i + \sum_{j \in \mathcal{C}} d(j, F^*) \right) \\
&\leqslant 3.25 \left(\frac{2\alpha}{1-\alpha} \Gamma_{\mathcal{I}}(F^*, \phi^*) + \sum_{j \in \mathcal{C}} d(j, F^*) \right) \\
&\leqslant 3.25 \left(\frac{2\alpha}{1-\alpha} \Gamma_{\mathcal{I}}(F^*, \phi^*) + \Gamma_{\mathcal{I}}(F^*, \phi^*) \right) \\
&= 3.25 \frac{1+\alpha}{1-\alpha} \Gamma_{\mathcal{I}}(F^*, \phi^*),
\end{aligned}$$

where the third inequality holds due to Lemma 1. Combining the above expression and Lemma 2 shows that an α-covered solution (F_0, ϕ_0) of $\mathcal{I}$ can be obtained in polynomial time, satisfying

$$\Gamma_{\mathcal{I}}(F_0, \phi_0) \leqslant \Gamma_{\mathcal{I}'}(F') \leqslant 3.25 \left(\frac{1+\alpha}{1-\alpha} \right) \Gamma_{\mathcal{I}}(F^*, \phi^*).$$

3 The approximation algorithm

The section proposes a way to convert the bi-criteria approximate solution (F_0, ϕ_0) of the instance $\mathcal{I} = (\mathcal{C}, \mathcal{F}, k, d, B)$ of LBK that is obtained by the way given in Theorem 2 to a feasible one, by reducing instance $\mathcal{I}$ to an instance of CFL, where the process consists of four phases. In Phase I, a different instance $\mathcal{I}_1 = (\mathcal{C}, \mathcal{F}, k, d_1, B)$ of LBK will be constructed by changing the metric d to d_1 based on the bi-criteria approximate solution (F_0, ϕ_0). In Phase II, we construct an instance $\mathcal{I}_2 = (\mathcal{C}, F_0, B, d_1, k)$ by trimming the facility set of $\mathcal{I}_1$. In Phase III, by adding the penalty cost for each closed facility in F_0, an instance $\mathcal{I}_3 = (\mathcal{C}, F_0, B, d_1, k, P_{\mathcal{I}_3})$ can be constructed. In Phase IV, based on the instance $\mathcal{I}_3$, we instead construct a CFL instance $\mathcal{I}_4$ since it can be solved by a known approximation algorithm.

3.1 Phase I: consolidating clients

Recall that (F_0, ϕ_0) is the bi-criteria approximate solution of the instance $\mathcal{I} = (\mathcal{C}, \mathcal{F}, k, d, B)$. A new instance $\mathcal{I}_1 = (\mathcal{C}, \mathcal{F}, k, d_1, B)$ of LBK is constructed based on $\mathcal{I} = (\mathcal{C}, \mathcal{F}, k, d, B)$, where the metric d_1 is defined as follows. For any two clients $j_1, j_2 \in \mathcal{C}$ and two facilities $i_1, i_2 \in \mathcal{F}$, $d_1(i_1, i_2) = d(i_1, i_2)$, $d_1(i_1, j_1) = d(i_1, \phi_0(j_1))$, and $d_1(j_1, j_2) = d(\phi_0(j_1), \phi_0(j_2))$. Note that the distance between any two facilities does not change in $\mathcal{I}_1$ compared to $\mathcal{I}$. Thus the d_1-metric space can be constructed by moving each client j to the facility $\phi_0(j)$ in the d-metric space; i.e., the clients locate at the same positions with the facilities of F_0.

Lemma 3. If there is a feasible solution (F, ϕ) of $\mathcal{I}$, $\Gamma_{\mathcal{I}_1}(F, \phi) \leqslant (1 + 3.25\frac{1+\alpha}{1-\alpha})\Gamma_{\mathcal{I}}(F, \phi)$.

Proof. It can be know that for any facility $i \in \mathcal{F}$ and client $j \in \mathcal{C}$ under the metric d_1,

$$\begin{aligned} d_1(i,j) &= d(i,\phi_0(j)) \quad \triangleright\text{by the definition of the metric } d_1 \\ &\leqslant d(i,j) + d(j,\phi_0(j)). \quad \triangleright\text{by the triangle inequality} \end{aligned} \tag{3}$$

Thus it can be known that

$$\begin{aligned} \Gamma_{\mathcal{I}_1}(F,\phi) &\leqslant \Gamma_{\mathcal{I}}(F,\phi) + \Gamma_{\mathcal{I}}(F_0,\phi_0) \\ &\leqslant \Gamma_{\mathcal{I}}(F,\phi) + 3.25\frac{1+\alpha}{1-\alpha}\Gamma_{\mathcal{I}}(F,\phi) \\ &= \left(1 + 3.25\frac{1+\alpha}{1-\alpha}\right)\Gamma_{\mathcal{I}}(F,\phi), \end{aligned} \tag{4}$$

where the first inequality holds due to inequality (3) and the second one follows from Theorem 2.

The reduction method given in [21] gives the following theorem.

Theorem 3 ([21])**.** If there is a β_1-approximate solution of $\mathcal{I}_1$, then a β-approximate solution of $\mathcal{I}$ can be obtained in polynomial time with

$$\beta = \left(1 + 3.25\frac{1+\alpha}{1-\alpha}\right)\beta_1 + 3.25\frac{1+\alpha}{1-\alpha}.$$

3.2 Phase II: trim on facility set

In this subsection, we center on instance $\mathcal{I}_1 = (\mathcal{C}, \mathcal{F}, B, d_1, k)$. To simplify the notation, let $\delta_i = \{j \in \mathcal{C} \mid \phi_0(j) = i\}$ for each facility $i \in F_0$. Recall that $|\delta_i| \geqslant \alpha B$ for each $i \in F_0$ as (F_0, ϕ_0) is an α-covered solution of $\mathcal{I}$. Now a new instance $\mathcal{I}_2 = (\mathcal{C}, F_0, B, d_1, k)$ of LBK is constructed by removing all facilities in $\mathcal{F}\backslash F_0$ from $\mathcal{I}_1$.

Lemma 4. If there is a feasible solution (F_1, ϕ_1) of $\mathcal{I}_1$, then a feasible solution (F_2, ϕ_2) of $\mathcal{I}_2$ can be obtained in polynomial time with $\Gamma_{\mathcal{I}_2}(F_2, \phi_2) \leqslant 2\Gamma_{\mathcal{I}_1}(F_1, \phi_1)$.

Proof. For each facility $i \in \mathcal{F}\backslash F_0$, let i' denote the facility closest to i in F_0. For $\mathcal{I}_2$, a solution (F_2, ϕ_2) can be constructed by opening the facilities of $F_1 \cap F_0$ and the ones i' with $i \in F_1\backslash F_0$ and reassigning the clients assigned to i with respect to ϕ_1 to i'. Note that for any client j reassigned, the definition of metric d_1 and triangle inequality imply that the increment of its cost is at most $d_1(i,i') = d(i,i') \leqslant d(i,\phi_0(j)) = d_1(i,j)$ (since i' is the closest facility in F_0 to i). Considering this above all $j \in \mathcal{C}$, the increment of the cost of (F_2, ϕ_2) compared with (F_1, ϕ_1) is bounded by $\Gamma_{\mathcal{I}_1}(F_1, \phi_1)$, hence $\Gamma_{\mathcal{I}_2}(F_2, \phi_2) \leqslant 2\Gamma_{\mathcal{I}_1}(F_1, \phi_1)$.

Observe that any feasible solution (F, ϕ) of instance $\mathcal{I}_2$ is also a feasible solution of instance $\mathcal{I}_1$, and $\Gamma_{\mathcal{I}_1}(F, \phi) = \Gamma_{\mathcal{I}_2}(F, \phi)$.

Theorem 4. If there is a β_2-approximate solution of $\mathcal{I}_2$, then a β_1-approximate solution of $\mathcal{I}_1$ can be obtained in polynomial time with $\beta_1 = 2\beta_2$.

Proof. Let (F_1^*, ϕ_1^*) be an optimal solution of $\mathcal{I}_1$, and (F, ϕ) be a β_2-approximate solution of $\mathcal{I}_2$. Lemma 4 shows that a solution (F_2, ϕ_2) of $\mathcal{I}_2$ can be obtained with $\Gamma_{\mathcal{I}_2}(F_2, \phi_2) \leqslant 2\Gamma_{\mathcal{I}_1}(F_1^*, \phi_1^*)$. Hence, for the β_2-approximate solution (F, ϕ) of $\mathcal{I}_2$, we have $\Gamma_{\mathcal{I}_2}(F, \phi) \leqslant 2\beta_2\Gamma_{\mathcal{I}_1}(F_1^*, \phi_1^*)$.

Observe that any feasible solution of instance $\mathcal{I}_2$ is also a feasible solution of instance $\mathcal{I}_1$. Therefore, (F, ϕ) is also a feasible solution of $\mathcal{I}_1$, and $\Gamma_{\mathcal{I}_1}(F, \phi) = \Gamma_{\mathcal{I}_2}(F, \phi) \leqslant 2\beta_2\Gamma_{\mathcal{I}_1}(F_1^*, \phi_1^*)$.

3.3 Phase III: entitling penalties to instance $\mathcal{I}_2$

Based on instance $\mathcal{I}_2 = (\mathcal{C}, F_0, B, d_1, k)$, a new instance $\mathcal{I}_3 = (\mathcal{C}, F_0, B, d_1, k, P_{\mathcal{I}_3})$ is constructed, which penalizes the close of the facility in F_0. A penalty cost $P_{\mathcal{I}_3}(i) = \frac{2\alpha-1}{\alpha}\ell_i|\delta_i|$ with $\ell_i = \min_{i' \in F_0\backslash\{i\}} d(i,i')$ is paid if facility i is closed in the solution of $\mathcal{I}_3$. For a solution (F, ϕ) of $\mathcal{I}_3$, let $\Gamma_{\mathcal{I}_3}(F, \phi)$ denote the cost of the solution (F, ϕ) of $\mathcal{I}_3$, which is composed of the connection cost $\omega_{\mathcal{I}_3}(F, \phi)$ and the penalty cost $P_{\mathcal{I}_3}(F, \phi) = \sum_{i \in F_0\backslash F} P_{\mathcal{I}_3}(i)$.

Lemma 5. For any feasible solution (F, ϕ) of $\mathcal{I}_2$ and $\mathcal{I}_3$,

$$\Gamma_{\mathcal{I}_2}(F, \phi) \leqslant \Gamma_{\mathcal{I}_3}(F, \phi) \leqslant \frac{3\alpha - 1}{\alpha}\Gamma_{\mathcal{I}_2}(F, \phi).$$

Proof. The cost of the solution (F,ϕ) of $\mathcal{I}_3$ is composed of the connection cost $\omega_{\mathcal{I}_3}(F,\phi)$ and the penalty cost $P_{\mathcal{I}_3}(F,\phi)$. It is obvious that the connection cost $\omega_{\mathcal{I}_3}(F,\phi)$ equals the cost of the solution (F,ϕ) of $\mathcal{I}_2$, i.e., $\omega_{\mathcal{I}_3}(F,\phi)=\Gamma_{\mathcal{I}_2}(F,\phi)$. For the penalty cost $P_{\mathcal{I}_3}(F,\phi)$, we have

$$\begin{aligned}P_{\mathcal{I}_3}(F,\phi)=\sum_{i\in F_0\setminus F}P_{\mathcal{I}_3}(i)&=\sum_{i\in F_0\setminus F}\frac{2\alpha-1}{\alpha}\ell_i|\delta_i|\\&\leqslant\frac{2\alpha-1}{\alpha}\sum_{i\in F_0\setminus F}\sum_{j\in\delta_i}d(i,\phi(j))\\&=\frac{2\alpha-1}{\alpha}\sum_{i\in F_0\setminus F}\sum_{j\in\delta_i}d_1(j,\phi(j))\\&\leqslant\frac{2\alpha-1}{\alpha}\sum_{j\in\mathcal{C}}d_1(j,\phi(j))\\&=\frac{2\alpha-1}{\alpha}\Gamma_{\mathcal{I}_2}(F,\phi),\end{aligned}$$

where the first inequality is obtained by replacing the symbol δ_i and scaling ℓ_i, and the third equality follows from $i=\phi_0(j)$. Summarizing the conclusions obtained above gives that

$$\Gamma_{\mathcal{I}_3}(F,\phi)=\omega_{\mathcal{I}_3}(F,\phi)+P_{\mathcal{I}_3}(F,\phi)=\Gamma_{\mathcal{I}_2}(F,\phi)+P_{\mathcal{I}_3}(F,\phi)\leqslant\frac{3\alpha-1}{\alpha}\Gamma_{\mathcal{I}_2}(F,\phi).$$

By Lemma 5, we have the following theorem.

Theorem 5. If there is a β_3-approximate solution of $\mathcal{I}_3$, then a β_2-approximate solution of $\mathcal{I}_2$ can be obtained in polynomial time with $\beta_2=\frac{3\alpha-1}{\alpha}\beta_3$.

Proof. Let (F_2^*,ϕ_2^*) denote an optimal solution of $\mathcal{I}_2$, and let (F,ϕ) be a β_3-approximate solution of $\mathcal{I}_3$. Lemma 5 gives that $\Gamma_{\mathcal{I}_2}(F,\phi)\leqslant\Gamma_{\mathcal{I}_3}(F,\phi)\leqslant\frac{3\alpha-1}{\alpha}\Gamma_{\mathcal{I}_2}(F,\phi)\leqslant\frac{3\alpha-1}{\alpha}\beta_3\Gamma_{\mathcal{I}_2}(F_2^*,\phi_2^*)$, as desired.

3.4 Phase IV: final reduction

In this subsection, we show that instance $\mathcal{I}_3=(\mathcal{C},F_0,B,d_1,k,P_{\mathcal{I}_3})$ can be reduced to an instance $\mathcal{I}_4=(\mathcal{C}_4,F_4,U,d_1,f)$ of CFL, where $\mathcal{C}_4$ and F_4 are the client-set and facility-set, respectively, and U and f are the capacity and opening cost defined on F_4, respectively. To avoid confusion, the facilities and clients considered in the CFL instances are called $\mathcal{C}$-facilities and $\mathcal{C}$-clients, respectively, in the remaining text.

Let P_i be the position of i for each facility $i\in F_0$. Recall that the clients of $\mathcal{C}$ locate at the same positions with the facilities of F_0 in the d_1-metric space, and that the facilities considered in LBK have no upper bound on their capacities, thus without loss of generality, the facilities are assumed to be at different positions, i.e., $P_i\neq P_j$ for any $i,j\in F_0$. In addition, for each facility $i\in F_0$, let $\tau_i^1=|\delta_i|$, $\tau_i^2=|\delta_i|-B$, and $\tau_i\in\{\tau_i^1,\tau_i^2\}$.

For any feasible solution (F,ϕ) of $\mathcal{I}_3$, if $i\in F$ (i.e., facility i is opened in $\mathcal{I}_3$), let $\tau_i=\tau_i^2$. If $\tau_i^2>0$, then τ_i^2 clients of δ_i can be reassigned maintaining the lower bound constraint of i with respect to ϕ_0; otherwise (i.e., $\tau_i^2\leqslant 0$), $|\tau_i^2|$ clients should be reassigned to i with respect to ϕ_0. If $i\in F_0\setminus F$ (i.e., facility i is not opened in $\mathcal{I}_3$), let $\tau_i=\tau_i^1$, and then the τ_i^1 clients of δ_i should be reassigned with respect to ϕ_0.

Now we construct the instance $\mathcal{I}_4=(\mathcal{C}_4,F_4,U,d_1,f)$ of CFL by the following two steps. Firstly, for each position P_i with $i\in F_0$, a $\mathcal{C}$-facility with opening cost $\frac{2\alpha-1}{\alpha}\ell_i|\delta_i|$ and capacity $\tau_i^1-\tau_i^2$ is constructed. Secondly, according to the value of τ_i^2, the operation is different at P_i. If $\tau_i^2\leqslant 0$, then a set of $|\tau_i^2|$ $\mathcal{C}$-clients is constructed; if $\tau_i^2>0$, then a $\mathcal{C}$-facility with opening cost 0 and capacity τ_i^2 is constructed. Note that there may be two $\mathcal{C}$-facilities at the same position in $\mathcal{I}_4$. For a solution (F,ϕ) of $\mathcal{I}_4$, let $\Gamma_{\mathcal{I}_4}(F,\phi)=f_{\mathcal{I}_4}(F,\phi)+\omega_{\mathcal{I}_4}(F,\phi)$ denote the cost of (F,ϕ), where $f_{\mathcal{I}_4}(F,\phi)$ is the opening cost of the $\mathcal{C}$-facilities in F and $\omega_{\mathcal{I}_4}(F,\phi)$ is the connection cost of assigning the $\mathcal{C}$-clients to the $\mathcal{C}$-facilities in F.

Lemma 6. If there is a feasible solution (F,ϕ) of $\mathcal{I}_3$, then a feasible solution (F_c,ϕ_c) of $\mathcal{I}_4$ of CFL can be obtained in polynomial time with $\Gamma_{\mathcal{I}_4}(F_c,\phi_c)\leqslant\Gamma_{\mathcal{I}_3}(F,\phi)$.

Proof. Note that the values of τ_i^1,τ_i^2,τ_i are known for each P_i with $i\in F_0$ by solution (F,ϕ). The set F_c of the opened $\mathcal{C}$-facilities of $\mathcal{I}_4$ is constructed with the following two steps. Firstly, for each P_i with $i\in F_0$, the $\mathcal{C}$-facility with opening cost 0 and capacity τ_i^2 (if any) is opened. Secondly, for each P_i with $i\in F_0\setminus F$, the $\mathcal{C}$-facility with opening cost $\frac{2\alpha-1}{\alpha}\ell_i|\delta_i|$ and capacity $\tau_i^1-\tau_i^2$ is opened.

Observe that given a subset $S \subseteq F_0$ and a facility $j \in F_0 \backslash S$ considered in $\mathcal{I}_3$, if there are η_i clients located at P_i assigned to j with respect to ϕ for any $i \in S$, then $\sum_{i\in S} \eta_i$ $\mathcal{C}$-clients located at P_j would be constructed in $\mathcal{I}_4$. Now based on F_c, we construct the assignment ϕ_c from the $\mathcal{C}$-clients to F_c by the following two steps. Firstly, for any position P_i with $i \in F_0 \backslash F$, if there are $|\tau_i^2|$ $\mathcal{C}$-clients located at P_i (note that $\tau_i^2 \leqslant 0$ under this case), then ϕ_c assigns the $|\tau_i^2|$ $\mathcal{C}$-clients to the $\mathcal{C}$-facility located at P_i. Secondly, for any P_i, P_j with $i, j \in F_0$ considered in $\mathcal{I}_3$, if there are $\eta > 0$ clients of δ_i that are assigned to the facility j located at P_j by ϕ, then ϕ_c reassigns the η $\mathcal{C}$-clients located at P_j to the $\mathcal{C}$-facilities located at P_i in $\mathcal{I}_4$ (recall the observation given at the beginning of this paragraph). By the assignment ϕ_c of solution (F_c, ϕ_c), we have that the connection cost of solution (F_c, ϕ_c) of $\mathcal{I}_4$ is no more than the connection cost of (F, ϕ) of $\mathcal{I}_3$.

Observe that given a subset $Q \subseteq F_0$ and a facility $i \in F_0 \backslash Q$ considered in $\mathcal{I}_3$, if there are η_j clients located at P_i assigned to j with respect to ϕ for any $j \in Q$, then the total capacity of the $\mathcal{C}$-facilities located at P_i is no less than $\sum_{j\in Q} \eta_j$ in $\mathcal{I}_4$ (note that there may be more than one $\mathcal{C}$-facility at P_i). Now we show that (F_c, ϕ_c) is a feasible solution of $\mathcal{I}_4$. For each position P_j with $j \in F_0 \backslash F$ and $\tau_j^2 \leqslant 0$, there are $|\tau_j^2|$ $\mathcal{C}$-clients at P_j, and they are assigned to the $\mathcal{C}$-facility at P_j with respect to ϕ_c, whose capacity is $\tau_j^1 - \tau_j^2$, implying that the $|\tau_j^2|$ $\mathcal{C}$-clients can be assigned to it and its capacity constraint is satisfied. Note that under this case, the unique $\mathcal{C}$-facility at P_j has room to be assigned τ_j^1 $\mathcal{C}$-clients. For each position P_j with $j \in F$ and $\tau_j^2 \leqslant 0$, we know that $\eta \leqslant |\delta_i| = \tau_i^1$ clients at P_i are assigned to the facility at P_j with respect to ϕ in $\mathcal{I}_3$, and that $\eta \leqslant |\delta_i| = \tau_i^1$ $\mathcal{C}$-clients at P_j are assigned to the $\mathcal{C}$-facilities at P_i with respect to ϕ_c in $\mathcal{I}_3$. By the observation given at the beginning of this paragraph, we have that the $\mathcal{C}$-facilities at P_i have the capacity to contain the $\mathcal{C}$-clients for all j satisfying the above conditions. Summarizing the above discussion gives that (F_c, ϕ_c) is a feasible solution of $\mathcal{I}_4$.

Now we consider the cost of (F_c, ϕ_c). If there is a facility $i \in F_0$ that is not opened with respect to the solution (F, ϕ) of $\mathcal{I}_3$, then we pay a penalty cost $\frac{2\alpha-1}{\alpha}\ell_i|\delta_i|$ that equals the opening cost of a $\mathcal{C}$-facility with capacity $\tau_i^1 - \tau_i^2$ in $\mathcal{I}_4$. Consequently, we have

$$\begin{aligned}\Gamma_{\mathcal{I}_4}(F_c, \phi_c) &= f_{\mathcal{I}_4}(F_c, \phi_c) + \omega_{\mathcal{I}_4}(F_c, \phi_c)\\ &\leqslant P_{\mathcal{I}_3}(F, \phi) + \omega_{\mathcal{I}_3}(F, \phi)\\ &= \Gamma_{\mathcal{I}_3}(F, \phi).\end{aligned}$$

Lemma 7. If there is a feasible solution (F_c, ϕ_c) of $\mathcal{I}_4$, then a feasible solution (F, ϕ) of $\mathcal{I}_3$ can be obtained in polynomial time with $\Gamma_{\mathcal{I}_3}(F, \phi) \leqslant \frac{2\alpha}{2\alpha-1}\Gamma_{\mathcal{I}_4}(F_c, \phi_c)$.

Proof. A subset $F' \subseteq F_0$ of $\mathcal{I}_3$ in which the facilities are opened is constructed as follows. Firstly, for each position P_i with $i \in F_0$, if the $\mathcal{C}$-facility with capacity $\tau_i^1 - \tau_i^2$ and opening cost $\frac{2\alpha-1}{\alpha}\ell_i|\delta_i|$ is opened with respect to (F_c, ϕ_c) of $\mathcal{I}_4$, then the facility at P_i is closed in $\mathcal{I}_3$; otherwise, it is opened.

Now based on F', we first construct an assignment ϕ' from the clients to F' by the following two steps. Step-1, for any position P_i, P_j with $i, j \in F_c$ considered in $\mathcal{I}_4$, if there are η $\mathcal{C}$-clients located at P_i that are assigned to the $\mathcal{C}$-facilities located at P_j by ϕ_c (i.e., the $\mathcal{C}$-facility located at P_i is not in F_c and the facility located at P_i is in F'), then ϕ' assigns η clients located at P_j to the facility located at P_i. Step-2, for any facility $i \in F'$, ϕ' assigns the unassigned clients of $\delta(i)$ in Step-1 to i. By the assignment ϕ' of (F', ϕ'), we have that the connection cost of solution (F_c, ϕ_c) of $\mathcal{I}_4$ equals that of (F', ϕ') of $\mathcal{I}_3$ (i.e., $\omega_{\mathcal{I}_4}(F_c, \phi_c) = \omega_{\mathcal{I}_3}(F', \phi')$), and the opening cost of solution (F_c, ϕ_c) of $\mathcal{I}_4$ equals the penalty cost of (F', ϕ') of $\mathcal{I}_3$ (i.e., $f_{\mathcal{I}_4}(F_c, \phi_c) = P_{\mathcal{I}_3}(F', \phi')$).

Unfortunately, there are some clients who are not unassigned with respect to ϕ'. Obviously, these unassigned clients are located at the position P_i with $i \in F_0 \backslash F'$ (i.e., the facility at P_i is not opened in $\mathcal{I}_3$).

Let $F = F'$ and $\phi = \phi'$. For each position P_i with $i \in F_0 \backslash F'$, let ζ_{P_i} denote the set of unassigned clients at P_i. Now we give the following operation to complete ϕ. If $\zeta_{P_i} \geqslant B$, then facility i ($F = F \cup \{i\}$) is opened, and ϕ assigns all clients of ζ_{P_i} to i without violating its lower bound. If $0 < \zeta_{P_i} < B$, then we consider the facility i' in F_0 that is the closest to i. If i' is opened, then ϕ assigns all clients of ζ_{P_i} to i', and the increment of the connection cost is bounded by $\sum_{j\in\zeta_{P_i}} d_1(i', j) = \sum_{j\in\zeta_{P_i}} d(i', i) \leqslant B\ell_i \leqslant \frac{|\delta_i|}{\alpha}\ell_i$. If i' is not opened, it involves the following two cases: (1) $|\zeta_{P_i}| + |\zeta_{P_{i'}}| \geqslant B$ and (2) $|\zeta_{P_i}| + |\zeta_{P_{i'}}| < B$. For case (1), facility i' is opened, let $F = F \cup \{i'\}$, and ϕ assigns all clients of ζ_{P_i} to i'. In this case, the lower bound of i' is satisfied. For case (2), all clients of ζ_{P_i} are assigned to facility i' and let $\zeta_{P_{i'}} = \zeta_{P_i} \cup \zeta_{P_{i'}}$. Repeat the operation until all clients are assigned without violating the lower bound constraint.

The critical issue is that the procedure of repeating the above operation may be caught in several facilities; more specifically, the sequence of the facilities on which the operation is performed comprises a directed cycle such that no feasible assignment can be found by repeating the operation. Let i and i' be two adjacent facilities in the directed cycle, where i is located previously to i'. Our idea is to assign these clients by ϕ to a facility $i'' \in F$ that has been opened such that the connection cost is minimized. Since i and i' are not opened, $|\zeta_{P_{i'}}| < B$ and $|\delta_{i'}| + |\delta_i| \geqslant 2\alpha B$. Thus, there are at least $|\delta_{i'}| + |\delta_i| - |\zeta_{P_{i'}}| \geqslant (2\alpha - 1)B$ clients that have been assigned by ϕ' and the connection cost is at most

$$\sum_{j \in \delta_{i'} \cup \delta_i \setminus \zeta_{P_{i'}}} d_1(\phi'(j), j) \geqslant (2\alpha - 1)B d_1(\phi'(j), j) \geqslant (2\alpha - 1)B \min\{d_1(i'', i), d_1(i'', i')\}. \tag{5}$$

Inequality (5) and triangle inequality imply that

$$\begin{aligned}\sum_{j \in \zeta_{P_{i'}}} d_1(i'', j) &\leqslant B d_1(i'', i) \leqslant B(\min\{d_1(i'', i), d_1(i'', i')\} + \ell_i) \\ &\leqslant \frac{1}{2\alpha - 1} \sum_{j \in \delta_{i'} \cup \delta_i \setminus \zeta_{P_{i'}}} d_1(\phi'(j), j) + \frac{\ell_i |\delta_i|}{\alpha}.\end{aligned}$$

Thus the total increment of the connection cost induced by the unassigned clients is at most

$$\frac{1}{2\alpha - 1} \omega_{\mathcal{I}_3}(F', \phi') + \frac{1}{2\alpha - 1} P_{\mathcal{I}_3}(F', \phi').$$

Summarizing the above discussion gives that (F, ϕ) is a feasible solution of $\mathcal{I}_3$. Now we consider the cost of (F, ϕ). Based on the conclusions above, we have that

$$\begin{aligned}\Gamma_{\mathcal{I}_3}(F, \phi) &\leqslant \omega_{\mathcal{I}_3}(F', \phi') + P_{\mathcal{I}_3}(F', \phi') + \frac{1}{2\alpha - 1} \omega_{\mathcal{I}_3}(F', \phi') + \frac{1}{2\alpha - 1} P_{\mathcal{I}_3}(F', \phi') \\ &= \frac{2\alpha}{2\alpha - 1} P_{\mathcal{I}_3}(F', \phi') + \frac{2\alpha}{2\alpha - 1} \omega_{\mathcal{I}_3}(F', \phi') \\ &= \frac{2\alpha}{2\alpha - 1} f_{\mathcal{I}_4}(F_c, \phi_c) + \frac{2\alpha}{2\alpha - 1} \omega_{\mathcal{I}_4}(F_c, \phi_c) \\ &= \frac{2\alpha}{2\alpha - 1} \Gamma_{\mathcal{I}_4}(F_c, \phi_c).\end{aligned}$$

Theorem 6. If there is a β_4-approximate solution of $\mathcal{I}_4$, then a β_3-approximate solution of instance $\mathcal{I}_3$ can be obtained in polynomial time with

$$\beta_3 = \frac{2\alpha}{2\alpha - 1} \beta_4.$$

Proof. Let (F_3^*, ϕ_3^*) denote an optimal solution of $\mathcal{I}_3$, and (F', ϕ') denote a β_4-approximate solution of $\mathcal{I}_4$. Lemma 6 shows that there is a feasible solution (F_c, ϕ_c) of $\mathcal{I}_4$ such that $\Gamma_{\mathcal{I}_4}(F_c, \phi_c) \leqslant \Gamma_{\mathcal{I}_3}(F_3^*, \phi_3^*)$. Hence, for the β_4-approximate solution (F', ϕ') of $\mathcal{I}_4$, we have that $\Gamma_{\mathcal{I}_4}(F', \phi') \leqslant \beta_4 \Gamma_{\mathcal{I}_3}(F_3^*, \phi_3^*)$. Furthermore, by Lemma 7, we can get a solution (F, ϕ) of $\mathcal{I}_3$ that satisfies $\Gamma_{\mathcal{I}_3}(F, \phi) \leqslant \frac{2\alpha}{2\alpha - 1} \Gamma_{\mathcal{I}_4}(F', \phi') \leqslant \frac{2\alpha}{2\alpha - 1} \beta_4 \Gamma_{\mathcal{I}_3}(F_3^*, \phi_3^*)$.

3.5 Summarizing everything

A $(1 + \sqrt{2})$-approximate solution of instance $\mathcal{I}_4$ can be obtained by applying the algorithm for CFL due to Ahmadian and Swamy [22]. Let $\alpha = \frac{2}{3}$. By Theorems 3–6, we have that

$$\begin{aligned}&\beta_3 = \frac{2\alpha}{2\alpha - 1}(1 + \sqrt{2}) = 4(1 + \sqrt{2}), \\ &\beta_2 = \frac{3\alpha - 1}{\alpha} \beta_3 = \frac{3\alpha - 1}{\alpha} \times 4(1 + \sqrt{2}) = 6(1 + \sqrt{2}), \\ &\beta_1 = 2\beta_2 = 2 \times 6(1 + \sqrt{2}) = 12(1 + \sqrt{2}), \\ &\beta = \beta_1 \left(1 + 3.25 \frac{1 + \alpha}{1 - \alpha}\right) + 3.25 \frac{1 + \alpha}{1 - \alpha} = 12(1 + \sqrt{2}) \times \left(1 + 3.25 \frac{1 + \alpha}{1 - \alpha}\right) + 3.25 \frac{1 + \alpha}{1 - \alpha} \approx 516.\end{aligned}$$

Therefore, Theorem 1 holds.

4 Conclusion

Within this paper, we present a reduction-based algorithm for LBK in polynomial time, which has the guarantee of yielding a 516-approximate solution. The main contribution of this paper is a reduction method that converts the instance of LBK to the instance of CFL. We think that this reduction method is of independent interest and can find applications in other clustering problems with lower bound constraints.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61872450, 62172446, 61802441).

References

1 Ravishankar K, Li S, Sandeep S. Constant approximation for k-median and k-means with outliers via iterative rounding. In: Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, 2018. 646–659
2 Hochbaum D S, Shmoys D B. A best possible heuristic for the k-center problem. Math Oper Res, 1985, 10: 180–184
3 Li S, Svensson O. Approximating k-median via pseudo-approximation. SIAM J Comput, 2016, 45: 530–547
4 Chen D Z, Huang Z Y, Liu Y W, et al. On clustering induced Voronoi diagrams. In: Proceedings of the 54th Annual Symposium on Foundations of Computer Science, 2013. 390–399
5 Hochbaum D S, Shmoys D B. A unified approach to approximation algorithms for bottleneck problems. J ACM, 1986, 33: 533–550
6 Zhang Z, Feng Q L, Huang J Y, et al. A local search algorithm for k-means with outliers. Neurocomputing, 2021, 450: 230–241
7 Feng Q L, Hu J X, Huang N, et al. Improved PTAS for the constrained k-means problem. J Comb Optim, 2019, 37: 1091–1110
8 Feng Q L, Zhang Z, Shi F, et al. An improved approximation algorithm for the k-means problem with penalties. In: Proceedings of the 13th International Workshop on Frontiers in Algorithmics. Berlin: Springer, 2019. 170–181
9 Lin J H, Vitter J S. Approximation algorithms for geometric median problems. Inf Process Lett, 1992, 44: 245–249
10 Byrka J, Pensyl T, Rybicki B, et al. An improved approximation for k-median and positive correlation in budgeted optimization. ACM Trans Algorithms, 2017, 13: 1–31
11 Zhang Z, Feng Q L, Xu J H, et al. An approximation algorithm for k-median with priorities. Sci China Inf Sci, 2021, 64: 150104
12 Cohen-Addad V, Klein P N, Mathieu C. Local search yields approximation schemes for k-means and k-median in euclidean and minor-free metrics. In: Proceedings of the 57th Annual Symposium on Foundations of Computer Science, 2016. 353–364
13 Jain K, Vazirani V V. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation. J ACM, 2001, 48: 274–296
14 Kumar A, Sabharwal Y, Sen S. Linear-time approximation schemes for clustering problems in any dimensions. J ACM, 2010, 57: 1–32
15 Li S. Approximating capacitated k-median with $(1 + \epsilon)k$ open facilities. In: Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms, 2016. 786–796
16 Feng Q L, Zhang Z, Huang Z Y, et al. Improved algorithms for clustering with outliers. In: Proceedings of the 30th International Symposium on Algorithms and Computation, 2019. 1–12
17 Qi L L, Zhang X Y, Dou W C, et al. A two-stage locality-sensitive hashing based approach for privacy-preserving mobile service recommendation in cross-platform edge environment. Future Generation Comput Syst, 2018, 88: 636–643
18 Qi L Y, Wang R L, Hu C H, et al. Time-aware distributed service recommendation with privacy-preservation. Inf Sci, 2019, 480: 354–364
19 Karger D R, Minkoff M. Building Steiner trees with incomplete global knowledge. In: Proceedings of the 41st Annual Symposium on Foundations of Computer Science, 2000. 613–623
20 Guha S, Meyerson A, Munagala K. Hierarchical placement and network design problems. In: Proceedings of the 41st Annual Symposium on Foundations of Computer Science, 2000. 603–612
21 Svitkina Z. Lower-bounded facility location. ACM Trans Algorithms, 2010, 6: 1–16
22 Ahmadian S, Swamy C. Improved approximation guarantees for lower-bounded facility location. In: Proceedings of the 10th International Approximation and Online Algorithms Workshop, 2012. 257–271
23 Li S. On facility location with general lower bounds. In: Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms, 2019. 2279–2290
24 Ding H, Xu J H. A unified framework for clustering constrained data without locality property. In: Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms, 2015. 1471–1490
25 Bhattacharya A, Jaiswal R, Kumar A. Faster algorithms for the constrained k-means problem. Theor Comput Syst, 2018, 62: 93–115
26 Ahmadian S, Swamy C. Approximation algorithms for clustering problems with lower bounds and outliers. In: Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming, 2016. 1–15
27 Feng Q L, Zhang Z, Huang Z Y, et al. A unified framework of FPT approximation algorithms for clustering problems. In: Proceedings of the 31st International Symposium on Algorithms and Computation, 2020. 1–17
28 Demirci G, Li S. Constant approximation for capacitated k-median with $(1+\epsilon)$-capacity violation. In: Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming, 2016. 1–14
29 Charikar M, Li S. A dependent LP-rounding approach for the k-median problem. In: Proceedings of the 39th International Colloquium on Automata, Languages, and Programming, 2012. 194–205