Parameterized Approximation Schemes for Fair Clustering in Doubling Metrics

Xiaoliang Wu[1] , **Ting Liang**[2] , **Zhize Li**[3] and **Qilong Feng**[2*]
[1]School of Computer Science and Technology, Changsha University of Science and Technology
[2]School of Computer Science and Engineering, Central South University
[3]School of Computing and Information Systems, Singapore Management University
xiaoliangwu@csust.edu.cn, tingliang@csu.edu.cn, zhizeli@smu.edu.sg, csufeng@mail.csu.edu.cn

Abstract

Fair clustering has garnered considerable attention, with various fairness notions proposed to ensure equitable representation across demographic groups. In this paper, we focus on the k-center problem in bounded doubling metrics under two popular fairness requirements: group fairness and data summarization fairness, referred to Group Fair k-Center (GF-k-CEN) and Data Summarization Fair k-Center (DSF-k-CEN), respectively. Both fairness notions extend classical clustering formulation by associating each data point with a demographic label. Motivated by recent advances in parameterized approximation results for fair clustering, we investigate whether these problems admit Fixed-Parameter Tractable (FPT) approximation schemes in bounded doubling metrics. The previous algorithms typically neglect the local structural properties induced by fairness constraints itself, which limits their approximation quality. By further leveraging the geometric properties of doubling metrics together with local fairness information, we develop a candidate-based structural method that yields $(1 + \varepsilon)$-approximation algorithms with FPT running times for both problems, parameterized by the number of selected centers. To the best of our knowledge, these results constitute the first parameterized approximation schemes for the GF-k-CEN and DSF-k-CEN problems in bounded doubling metrics.

1 Introduction

Clustering is a fundamental problem in machine learning with a wide range of applications in data mining, image analysis, etc. Given a set of points, the goal of clustering is to partition the point set into several disjoint clusters such that the points in the same cluster are close to each other, while points in different clusters are well separated. Several classic clustering formulations have been extensively studied, including k-center, k-median, and k-means. In this paper, we focus on the k-center problem, in which we are given a set C of points in a metric space $(\mathcal{X}, d)$, and a positive integer k. The goal is to select k centers from C such that the maximum distance from any point in C to its nearest center is minimized. The k-center problem is known to be NP-hard [Garey and Johnson, 1979], and admits a 2-approximation algorithm [Gonzalez, 1985; Hochbaum and Shmoys, 1985]. Moreover, finding an optimal solution for the k-center problem is W[2]-hard if parameterized by k [Demaine *et al.*, 2005]. Over the past decades, numerous variants of the k-center problem have been explored [Charikar *et al.*, 2001; Aggarwal *et al.*, 2010; Cygan *et al.*, 2012; Chierichetti *et al.*, 2017; Kleindessner *et al.*, 2019; Chen *et al.*, 2019; Mahabadi and Vakilian, 2020; Liang *et al.*, 2025].

Recently, fair clustering has become a major focus with many notions proposed, such as group fairness [Ahmadian *et al.*, 2019; Bercea *et al.*, 2019; Harb and Lam, 2020], data summarization fairness [Kleindessner *et al.*, 2019; Jones *et al.*, 2020], proportional fairness [Chen *et al.*, 2019; Micha and Shah, 2020], individual fairness [Mahabadi and Vakilian, 2020; Negahbani and Chakrabarty, 2021], etc. In this paper, we focus on the k-center problem under group fairness and data summarization fairness constraints in bounded doubling metrics, denoted as Group Fair k-Center (GF-k-CEN) and Data Summarization Fair k-Center (DSF-k-CEN), respectively. Motivated by recent progress on parameterized approximation algorithms for fair clustering [Bandyapadhyay *et al.*, 2021; Thejaswi *et al.*, 2022; Zhang *et al.*, 2024; Wu *et al.*, 2025], we investigate the existence of Fixed-Parameter Tractable (FPT) approximation schemes for the GF-k-CEN and DSF-k-CEN problems in bounded doubling metrics (an approximation scheme is a $(1+\varepsilon)$-approximation algorithm, where ε is an arbitrary small constant). In both fairness settings, each point is assigned a color indicating its demographic group. The two notions differ in how fairness is enforced: group fairness imposes the composition of each cluster such that the number of points with each color respects pre-specified proportions, whereas data summarization fairness requires the selected centers such that the number of centers with each color matches a pre-specified value. Indeed, both fairness notions extend the classical clustering formulation with demographic labels, where the former requires each cluster to proportionally reflect the population-level distribution of demographic groups, and the latter ensures the chosen centers maintaining the population-level demographic repre-

*Corresponding author

sentation of each group. We briefly illustrate the motivation for these notions. As noted in [Chierichetti *et al.*, 2017], if minority groups are ignored during the clustering process, the resulting clusters may substantially distort their representation, leading to systematically unfair outcomes. Data summarization similarly aims to select a small subset that represents the whole dataset. However, without fairness constraints, such summaries can be biased with respect to sensitive attributes. For instance, the Google Images search results for the query "CEO" have been shown to over-represent men relative to their real-world prevalence, reinforcing harmful stereotypes [Kay *et al.*, 2015].

[Chierichetti *et al.*, 2017] introduced the definition of fairness with only two colors, requiring that the proportion of two colors has approximately equal representation in each cluster. [Bercea *et al.*, 2019] formalized group fairness more generally, and presented a 3-approximation with an additive 1 violation of the group fairness constraints for the GF-k-CEN problem, using linear programming and min-cost flow network. The violation value represents the extent to which the fairness constraints are violated. [Ahmadian *et al.*, 2019] studied the GF-k-CEN problem with only an upper bound constraint, obtaining a 3-approximation with an additive 2 violation (under a notion of violation different from [Bercea *et al.*, 2019]), again via linear programming and min-cost flow network. When colors are allowed to overlap, a 4-approximation algorithm with $(4\Delta + 3)$ violation [Bera *et al.*, 2019] and a 3-approximation algorithm with $(4\Delta + 3)$ violation [Harb and Lam, 2020] were given for the GF-k-CEN problem, respectively, where Δ is the maximum number of colors to which any point may belong. Regarding parameterized result, [Goyal and Jaiswal, 2023] proposed a 2-approximation for the GF-k-CEN problem in FPT(k, m)-time, where m is the number of colors ($n^{O(1)} f(k)$ time for an input size of n and a positive function f, denoted by FPT(k)-time for brevity). [Bandyapadhyay *et al.*, 2021] employed the fair coreset technique to obtain $(3 + \varepsilon)$-approximation and $(9 + \varepsilon)$-approximation for the k-median and k-means objectives under group fairness constraints, respectively, both in FPT(k, m)-time.

[Kleindessner *et al.*, 2019] presented a $(3 \cdot 2^{m-1} - 1)$-approximation with running time $O(nkm^2 + km^4)$ based on a swap technique for the DSF-k-CEN problem. [Jones *et al.*, 2020] improved the time complexity to $O(nk)$, and maintained the approximation factor 3 by maximum matching method. More recently, [Angelidakis *et al.*, 2022] considered the lower-bounded version of the DSF-k-CEN problem, and gave a 15-approximation algorithm in time $O(nk^2 + k^5)$. [Thejaswi *et al.*, 2022] extended data summarization fairness by introducing an additional lower-bound constraint on the number of selected centers, and provided $(1 + 2/e + \varepsilon)$-approximation and $(1 + 8/e + \varepsilon)$-approximation for the k-median and k-means objectives, respectively, both running in FPT(k, m)-time. Moreover, the approximation factor can be further improved to $(1 + \varepsilon)$ in Euclidean space [Zhang *et al.*, 2024]. [Wu *et al.*, 2025] considered the doubly constrained fairness setting that combines group fairness and data summarization fairness constraints under the k-median objective, and gave a $(4 + \varepsilon)$-approximation algorithm of fairness violation by one in FPT(k)-time.

For the GF-k-CEN and DSF-k-CEN problems, several fundamental obstacles hinder the development of improved approximation algorithms. The FPT approximation algorithm in [Goyal and Jaiswal, 2023] for the GF-k-CEN problem starts with a bi-criteria approximate solution with $O(k \log n)$ centers. However, the gap between the bi-criteria solution and the optimal solution remains large, making it difficult to refine the solution to achieve better approximation guarantee. Moreover, for the k-center problem, it is known that achieving an approximation factor better than 2 is NP-hard unless P=NP [Hsu and Nemhauser, 1979]. Therefore, for the GF-k-CEN and DSF-k-CEN problems, the gaps between the ratios of the previous results and the lower bound ratio of k-center are still large. A common paradigm underlying existing algorithms for both problems is to first compute a set of k initial clustering centers and then convert it into a feasible solution. Unfortunately, even when the set of k initial clustering centers is of high-quality, converting it into a solution that satisfies the fairness constraints while incurring only a small loss in approximation guarantee remains troublesome, as this process often discards effective local structural information induced by the fairness constraints itself. Indeed, for the GF-k-CEN problem, even assigning data points to a fixed set of centers under fairness constraints is NP-hard [Esmaeili *et al.*, 2021]. Consequently, how to effectively convert a set of initial clustering centers to a desired feasible solution (the gap between the approximate solution obtained by the set of initial clustering centers and the optimal solution is small) is still a challenge for the GF-k-CEN and DSF-k-CEN problems. The hardness result of [Goyal and Jaiswal, 2023] shows that no FPT(k)-time algorithm can achieve an approximation ratio better than 2 for the GF-k-CEN problem in general metric spaces. Nevertheless, this negative result does not rule out the possibility of better approximation guarantees in more structured settings. In particular, their analysis applies only to general metrics, and leaves open the potential for improved approximation algorithms in restricted metric spaces such as doubling metrics. This naturally motivates the question of whether effective FPT time approximation schemes can be developed for the GF-k-CEN and DSF-k-CEN problems in such structured settings.

1.1 Our Contributions

In this paper, we take the first step toward exploiting the structural properties of doubling metrics for the GF-k-CEN and DSF-k-CEN problems, and propose a candidate-based structural method to overcome the aforementioned obstacles. Our main contributions provide a positive answer to the stated question, summarized as follows.

- By leveraging the geometric properties of doubling metrics and incorporating local fairness information, we develop a candidate-based structural method that effectively reduces the gap between the initial clustering centers and the final feasible solution. Based on the proposed method, we obtain $(1 + \varepsilon)$-approximation algorithms for the GF-k-CEN and DSF-k-CEN problems in bounded doubling metrics, with FPT running time.

Problem	Approximation	Time	Space	Reference
GF-k-CEN	3	Polynomial	Metric space	[Bercea *et al.*, 2019; Harb and Lam, 2020]
	2	FPT(k, m)	Metric space	[Goyal and Jaiswal, 2023]
	$1 + \varepsilon$	FPT(k)	Doubling metrics	Theorem 1
DSF-k-CEN	3	Polynomial	Metric space	[Jones *et al.*, 2020]
	$1 + \varepsilon$	FPT(k)	Doubling metrics	Theorem 2

Table 1: Approximation results for the GF-k-CEN and DSF-k-CEN problems.

- The proposed candidate-based structural method induces a crucial distance-bounded property that states theoretically the existence of $(1 + \varepsilon)$-approximate solutions for the GF-k-CEN and DSF-k-CEN instances.

We summarize the comparison between our results and existing work in Table 1. Formally, our main theoretical guarantees are stated as follows.

Theorem 1. *For a real number $\varepsilon > 0$, there exists a $(1 + \varepsilon)$-approximation algorithm of fairness violation by one for the* GF-k-CEN *problem in bounded doubling metrics with running time $f(k, \varepsilon)n^{O(1)}$, where $f(k, \varepsilon) = (k\varepsilon^{-1})^{O(k)}$.*

Theorem 2. *For a real number $\varepsilon > 0$, there exists a $(1+\varepsilon)$-approximation algorithm for the* DSF-k-CEN *problem in bounded doubling metrics with running time $f(k, \varepsilon)n^{O(1)}$, where $f(k, \varepsilon) = (k\varepsilon^{-1})^{O(k)}$.*

1.2 Other Related Work

The doubling dimension (see Section 2 with a formal definition) captures the intrinsic dimensionality of a general metric space and can be viewed as a natural extension of the notion of dimensionality in Euclidean space. As noted in [Pellizzoni *et al.*, 2025], in many real-world instances, the points of interest either lie in low-dimensional space or are supported on low-dimensional manifolds embedded in higher-dimensional metric space. Moreover, the structural properties of spaces with low doubling dimension have been exploited to design efficient clustering algorithms that achieve improved trade-offs between computational resources and approximation quality compared to dimension-insensitive strategies [Ceccarello *et al.*, 2017; Goranci *et al.*, 2021].

For the k-center problem in doubling metrics, existing research has primarily focused on the construction of coreset [Aghamolaei and Ghodsi, 2018; Ding *et al.*, 2019; Ceccarello *et al.*, 2019], and the design of approximation algorithm on fully dynamic model [Goranci *et al.*, 2021; Pellizzoni *et al.*, 2025]. More broadly, a wide range of algorithmic strategies have been proposed for the k-median and k-means problems in bounded doubling metrics [Huang *et al.*, 2018; Friggstad *et al.*, 2019; Cohen-Addad, 2020; Cohen-Addad *et al.*, 2021]. In contrast, the study of fair clustering in doubling metrics remains largely unexplored. With the exception of a few approximation algorithms developed in distributed settings [Wu *et al.*, 2024; Ceccarello *et al.*, 2024], there are no known sequential results for fair clustering in doubling metrics. This gap further motivates our investigation into fair clustering under geometric constraints.

2 Preliminaries

For any integer $m \geq 1$, let $[m] = \{1, \ldots, m\}$. Let $(\mathcal{X}, d)$ be a metric space, and let $C \subseteq \mathcal{X}$ denote the set of points considered in this paper. For any point $v \in C$ and an integer $r \geq 0$, the ball centered at v with radius r is defined as $B(v, r) = \{c \in C \mid d(c, v) \leq r\}$. For any point $v \in C$ and a subset $S \subseteq C$, let $d(v, S) = \min_{s \in S} d(v, s)$ denote the distance from v to its nearest center in S. In this paper, we mainly consider the metric space with bounded doubling dimension [Assouad, 1983; Gupta *et al.*, 2003], which is widely considered in clustering problems [Huang *et al.*, 2018; Ding *et al.*, 2019; Cohen-Addad, 2020; Cohen-Addad *et al.*, 2021; Ceccarello *et al.*, 2024]. The doubling dimension of $(\mathcal{X}, d)$ is the smallest number, denoted as D, such that for any point $v \in C$ and radius r, all points in the ball $B(v, r)$ are always covered by the union of at most 2^D balls with radius $r/2$.

Definition 1 (the k-center problem)**.** *An instance of the k-center problem is denoted by $((\mathcal{X}, d), C, k)$, where $(\mathcal{X}, d)$ is a metric space, $C \subseteq \mathcal{X}$ is a set of points, and k is a positive integer, respectively. The goal is to find a subset $S \subseteq C$ of k centers such that the cost $\max_{v \in C} d(v, S)$ is minimized.*

We now introduce the notions of fairness considered in this paper. Assume that the given data point C is divided m disjoint groups $\mathcal{P} = \{C_1, \ldots, C_m\}$ such that $\cup_{h=1}^{m} C_h = C$, where each point in C_h ($h \in [m]$) is said to have color h. We call such $\mathcal{P}$ a partition of C.

Definition 2 (the GF-k-CEN problem)**.** *An instance of the* GF-k-CEN *problem is denoted by $((\mathcal{X}, d), C, \mathcal{P}, \vec{\alpha}, \vec{\beta}, m, k)$, where $(\mathcal{X}, d)$ is a metric space, $C \subseteq \mathcal{X}$ is a set of points with a partition $\mathcal{P} = \{C_1, \ldots, C_m\}$, $\vec{\alpha} = (\alpha_1, \ldots, \alpha_m), \vec{\beta} = (\beta_1, \ldots, \beta_m)$ are two fairness vectors, and m, k are two positive integers, respectively. The goal is to find a subset $S \subseteq C$ of k centers, and a mapping $\phi : C \to S$ such that the cost $\max_{v \in C} d(v, \phi(v))$ is minimized, and ϕ satisfies the group fairness constraints, i.e., for any $s \in S, h \in [m]$, $\alpha_h \leq \frac{|\{v \in C_h | \phi(v) = s\}|}{|\{v \in C | \phi(v) = s\}|} \leq \beta_h$.*

Given an instance $((\mathcal{X}, d), C, \mathcal{P}, \vec{\alpha}, \vec{\beta}, m, k)$ of the GF-k-CEN problem, a pair (S, ϕ) is called a feasible solution of this instance if $S \subseteq C$ is a set with size k, and $\phi : C \to S$ is a mapping satisfying group fairness constraints.

Definition 3 (the DSF-k-CEN problem)**.** *An instance of the* DSF-k-CEN *problem is denoted by $((\mathcal{X}, d), C, \mathcal{P}, \vec{\theta}, m, k)$, where $(\mathcal{X}, d)$ is a metric space, $C \subseteq \mathcal{X}$ is a set of points with a partition $\mathcal{P} = \{C_1, \ldots, C_m\}$, $\vec{\theta} = (k_1, \ldots, k_m)$ is a* *fairness vector satisfying* $\sum_{h=1}^{m} k_h = k$*, and* m, k *are two positive integers, respectively. The goal is to find a subset* $S \subseteq C$ *of* k *centers such that the cost* $\max_{v \in C} d(v, S)$ *is minimized, and* S *satisfies the data summarization fairness constraints, i.e., for any* $h \in [m]$, $|S \cap C_h| = k_h$.

Given an instance $((\mathcal{X}, d), C, \mathcal{P}, \vec{\theta}, m, k)$ of the DSF-k-CEN problem, S is called a feasible solution of this instance if S is a subset of C with size k satisfying data summarization fairness constraints. Throughout this paper, let τ and τ_f denote the costs of optimal solutions of the unconstrained k-center instance and the corresponding GF-k-CEN (or DSF-k-CEN) instance, respectively. Since the feasible solution of the GF-k-CEN (or DSF-k-CEN) instance is also feasible to the corresponding k-center instance, it follows that $\tau \leq \tau_f$. Note that for the GF-k-CEN (or DSF-k-CEN) problem, the cost of optimal solution is always the distance between two points in C. Hence, the value of τ_f can be determined via a binary search over all pairwise distances, incurring an additional factor of $O(n^2)$ in the running time. In the subsequent analysis, we assume that τ_f is known for the given instance, which is a standard and widely used technique for the k-center problem and its variants [Khuller and Sussmann, 2000]. Let (S^*, ϕ^*) denote an optimal solution to the GF-k-CEN (or DSF-k-CEN) instance with cost τ_f, where $S^* = \{s_1^*, \ldots, s_k^*\}$ is the set of k optimal centers, and ϕ^* is the associated assignment function. Let $O^* = \{O_1^*, \ldots, O_k^*\}$ denote the set of k optimal clusters induced by ϕ^*. Note that for the DSF-k-CEN problem, the assignment function ϕ^* can be omitted.

3 An Overview of Our Algorithms

In this paper, we propose a candidate-based structural method for the GF-k-CEN and DSF-k-CEN problems in doubling metrics. Since the quality of a set of k initial clustering centers critically affects the finding of approximate solutions, our method begins by constructing a candidate set that contains a high-quality set of k initial clustering centers, leveraging the geometric properties of doubling metrics. Intuitively, the candidate construction process aims to build a representative subset of the given dataset such that it contains at least one point that approximately captures the location of optimal center. Compared with the bi-criteria solution given in [Goyal and Jaiswal, 2023], our method can effectively reduce the gap between the set of k initial clustering centers and the final feasible solution by working with the carefully designed candidate set. To further convert the set of initial clustering centers into a feasible solution without loss in approximation guarantee, our method explicitly exploits the local structural properties induced by fairness constraints itself. In particular, it establishes a distance-bounded property that enables the conversion while maintaining approximation quality. Specifically, this property theoretically states that there must exist a feasible solution with cost at most a factor $(1 + \varepsilon)$ of the optimal solution cost, where ε is a parameter used to control the approximation ratio. Equivalently, we can achieve a small gap between the approximate solution obtained by the set of initial clustering centers and the optimal solution. In contrast, the method of [Goyal and Jaiswal, 2023] directly relies on linear programming and min-cost flow network to convert a bi-criteria solution into a feasible one, without explicitly considering the local structural properties. As a result, their method does not effectively control the approximation gap, and is difficult to extend to the DSF-k-CEN problem. Our method can overcome these limitations to obtain desired approximate solutions in FPT time by combining with problem-specific selection algorithms.

We now provide an intuitive overview of our parameterized algorithms for the GF-k-CEN and DSF-k-CEN problems in doubling metrics. Instead of directly selecting the set of k initial clustering centers, our algorithm first utilizes the structure of doubling metrics to reduce the input to a carefully chosen candidate set without imposing fairness constraints. In the resulting candidate set, we then identify the high-quality set of clustering centers by enumerating all possible subsets of size k. Finally, depending on the specific fairness requirement, we apply appropriate algorithmic techniques to obtain desired feasible solutions with minimum clustering cost. Specifically, we employ linear programming and min-cost flow network for the GF-k-CEN problem, while the maximum matching method is used for the DSF-k-CEN problem. By the above process, we develop FPT time $(1 + \varepsilon)$-approximation algorithms for the GF-k-CEN and DSF-k-CEN problems, respectively, where $\varepsilon > 0$ is a given parameter.

4 Parameterized Approximation Algorithms

We now present our parameterized approximation algorithms for the GF-k-CEN and DSF-k-CEN problems. For brevity, we use the GF-k-CEN problem as an example for the analysis, which can be easily adapted to the DSF-k-CEN problem. Consider an instance $((\mathcal{X}, d), C, \mathcal{P}, \vec{\alpha}, \vec{\beta}, m, k)$ of the GF-k-CEN problem. By utilizing the geometric property of doubling metrics, we reduce the given data points to a candidate set of centers. Specifically, we consider the method in [Ceccarello *et al.*, 2019] with the following result.

Theorem 3. *Given an instance* $((\mathcal{X}, d), C, k)$ *of the* k*-center problem with the bounded doubling dimension* D *and a real number* $\varepsilon > 0$*, we can obtain a subset* $E \subseteq C$ *with size* $k \cdot (4/\varepsilon)^D$ *in polynomial time such that for any* $v \in C$, $d(v, E) \leq \varepsilon\tau$*, where* τ *is the optimal cost of this instance.*

Note that Theorem 3 still holds for the GF-k-CEN problem instance due to $\tau \leq \tau_f$. We now show that the candidate set E constains a high-quality set of initial clustering centers.

Lemma 1. *There must exist a subset* $S \subseteq E$ *with size* k *such that for any* $v \in C$, $d(v, S) \leq (1 + \varepsilon)\tau_f$.

Proof. For any $i \in [k]$, let $\pi(s_i^*) = \arg\min_{v \in E} d(v, s_i^*)$ be the closest point in E to s_i^*. Then, for any $v \in O_i^*$ ($i \in [k]$), we have $d(v, \pi(s_i^*)) \leq d(v, s_i^*) + d(s_i^*, \pi(s_i^*)) \leq \tau_f + \varepsilon\tau \leq (1 + \varepsilon)\tau_f$, where the first inequality follows from the triangle inequality, and the last inequality holds due to $\tau \leq \tau_f$ and Theorem 3. Hence, there must exist a subset $S = \{\pi(s_1^*), \ldots, \pi(s_k^*)\} \subseteq E$ with size k such that for any $v \in C$, $d(v, S) \leq (1 + \varepsilon)\tau_f$. □

Lemma 1 implies that there must exist a subset S of E that induces a $(1 + \varepsilon)$-approximation without satisfying fairness constraints by assigning each point of C to its closest center in S. Notably, S may be a multi-set, since it is possible that some optimal centers have the same closest point in E. Formally, we call such S a set of ε-optimal cluster centers (i.e., the high-quality set of initial clustering centers). Indeed, there are $|E|^k$ subsets with size k of E, and one of them induces a $(1+\varepsilon)$-approximation without satisfying fairness constraints. To determine which one is the set of ε-optimal cluster centers, a problem-specific selection algorithm (see [Ding and Xu, 2020] with more details) is needed to take each subset with size k of E as input, and output the subset with the minimum clustering cost. For simplicity, assume that we have found the set S of ε-optimal cluster centers for given instance. To convert the set of ε-optimal cluster centers obtained to one satisfying fairness constraints without approximation loss, in the following, we prove that for the GF-k-CEN problem, there must exist a $(1+\varepsilon)$-approximate solution (note that the solution satisfies fairness constraints).

Lemma 2. *Given an instance* $((\mathcal{X}, d), C, \mathcal{P}, \vec{\alpha}, \vec{\beta}, m, k)$ *of the* GF-k-CEN *problem and a real number* $\varepsilon > 0$, *let* $S = \{s_1, \ldots, s_k\}$ *be the set of* ε*-optimal cluster centers of this instance. Then, there must exist a* $(1+\varepsilon)$*-approximate solution of this instance based on* S.

Proof. For any $i \in [k]$ and $h \in [m]$, let $O_i^*(h)$ denote the set of points in O_i^* with color h. Obviously, for any $i \in [k]$, we have $O_i^* = \cup_{h\in[m]} O_i^*(h)$. For any $i \in [k]$, let $\pi(s_i^*) = \arg\min_{s\in S} d(s, s_i^*)$ denote the closest center in S to s_i^*. By the definition of the set S of ε-optimal cluster centers, it follows that $d(s_i^*, \pi(s_i^*)) \le \varepsilon\tau_f$. For any point $v \in C$, let $\phi(v) = \pi(\phi^*(v))$. We now claim that (S, ϕ) forms a $(1+\varepsilon)$-approximate solution. For any point $v \in O_i^*$ ($i \in [k]$), by the triangle inequality, we have $d(v, \phi(v)) = d(v, \pi(\phi^*(v))) \le d(v, \phi^*(v)) + d(\phi^*(v), \pi(\phi^*(v))) \le (1+\varepsilon)\tau_f$. Hence, the cost of (S, ϕ) is at most $(1+\varepsilon)\tau_f$. We next verify the group fairness constraints. Since (S^*, ϕ^*) is a feasible solution, for any $i \in [k]$ and $h \in [m]$, $\alpha_h \le |O_i^*(h)|/|O_i^*| \le \beta_h$. For any $s \in S$, let $N(s) = \{s_i^* \in S^* \mid \pi(s_i^*) = s\}$ denote all centers in S^* such that s is the closest center. Observe that $\{v \in C \mid \phi(v) = s\} = \cup_{s_i^* \in N(s)} O_i^*$, and similarly, for any $h \in [m]$, $\{v \in C_h \mid \phi(v) = s\} = \cup_{s_i^* \in N(s)} O_i^*(h)$. Consequently, for any $h \in [m]$ and $s \in S$, we have $\frac{|\{v\in C_h | \phi(v)=s\}|}{|\{v\in C|\phi(v)=s\}|} = \frac{\sum_{s_i^*\in N(s)} |O_i^*(h)|}{\sum_{s_i^*\in N(s)} |O_i^*|}$. Using a scaling technique, we have $\min_{s_i^*\in N(s)} \frac{|O_i^*(h)|}{|O_i^*|} \le \frac{\sum_{s_i^*\in N(s)} |O_i^*(h)|}{\sum_{s_i^*\in N(s)} |O_i^*|} \le \max_{s_i^*\in N(s)} \frac{|O_i^*(h)|}{|O_i^*|}$. Combining this with the feasibility of (S^*, ϕ^*), we conclude that $\alpha_h \le \frac{\sum_{s_i^*\in N(s)} |O_i^*(h)|}{\sum_{s_i^*\in N(s)} |O_i^*|} \le \beta_h$, which shows that ϕ satisfies group fairness constraints. □

Lemma 2 implies a distance-bounded property: there must exist a mapping ϕ satisfying the group fairness constraints that assigns each point in C to a center in S within distance $(1+\varepsilon)\tau_f$. The remaining task is to find such a desired approximate solution by the distance-bounded property. In the following, we will show how to obtain the desired FPT approximation algorithms by combining the proposed method with problem-specific selection algorithms for each problem.

Algorithm 1: An algorithm for GF-k-CEN

Input: An instance $((\mathcal{X}, d), C, \mathcal{P}, \vec{\alpha}, \vec{\beta}, m, k)$ of the GF-k-CEN problem and parameters ε, τ_f
Output: A pair (S, ϕ)

Let E be the candidate set constructed by Theorem 3;
for *each multi-set* $S \subseteq E$ **do**
    $\mathbf{x} \leftarrow$ solve the linear programming relaxation;
    Construct a min-cost flow network G based on $\mathbf{x}$;
    $\overline{\mathbf{x}} \leftarrow$ find a min-cost flow on G;
    $\forall i \in S, j \in C : \phi(j) \leftarrow i$ if $\overline{x}_{ij} > 0$;
Output (S, ϕ) such that the cost of (S, ϕ) is minimized among all considered subsets;
return (S, ϕ).

4.1 The GF-k-CEN Problem

In this section, we consider the GF-k-CEN problem. Based on S, our algorithm starts with linear programming to get a fractional solution with cost at most $(1+\varepsilon)\tau_f$. In order to modify the fractional solution obtained, it constructs a min-cost flow network to get an integral solution, which has the cost bounded by $(1+\varepsilon)\tau_f$, and violates the fairness constraints by one (see Algorithm 1). Formally, we start with the following linear programming.

$$\min \quad 1$$

$$\sum_{i\in S} x_{ij} = 1, \forall j \in C \tag{1}$$

$$\alpha_h \sum_{j\in C} x_{ij} \le \sum_{j\in C_h} x_{ij} \le \beta_h \sum_{j\in C} x_{ij}, \forall i \in S, h \in [m] \tag{2}$$

$$x_{ij} \in \{0,1\}, \forall i \in S, j \in C \tag{3}$$

$$x_{ij} = 0, \forall i \in S, j \in C, d(i,j) \le (1+\varepsilon)\tau_f \tag{4}$$

Here the binary variable x_{ij} indicates whether point $j \in C$ is assigned to center $i \in S$. Note that the value of x_{ij} represents the assignment ϕ. Constraint (1) ensures that each point $j \in C$ is assigned to a center. Constraint (2) captures group fairness constraints. Indeed, constraint (4) is based on the distance-bounded property that ensures the cost of the fractional solution obtained is at most $(1+\varepsilon)\tau_f$. Since the intractableness of integral linear programming, we relax constraint (3) by allowing $x_{ij} \in [0,1]$ for all $i \in S, j \in C$, yielding a linear programming relaxation. Let $\mathbf{x}$ denote an optimal solution to the linear programming relaxations, which has the property that the assignment satisfies the group fairness constraints, but the points may be fractionally assigned to several centers in S. Let $\mathrm{cost}(\mathbf{x})$ denote the cost of the solution $\mathbf{x}$. The remaining task is to obtain an integral solution $\overline{\mathbf{x}}$ with $\mathrm{cost}(\overline{\mathbf{x}}) \le (1+\varepsilon)\tau_f$. By constructing a min-cost flow network, we can obtain such an integral solution.

Lemma 3. *Given an instance* $((\mathcal{X}, d), C, \mathcal{P}, \vec{\alpha}, \vec{\beta}, m, k)$ *of the* GF-k-CEN *problem and a real number* $\varepsilon > 0$, *let* $\mathbf{x}$ *be the solution returned by the linear programming relaxations. Then, we can obtain a* $(1+\varepsilon)$*-approximate solution* $\overline{\mathbf{x}}$ *with an additive* 1 *violation for group fairness constraints in FPT time based on a min-cost flow network.*

Proof. For any $i \in S, h \in [m]$, let $T_i^h = \sum_{j \in C_h} x_{ij}, T_i = \sum_{j \in C} x_{ij}, B_i = \lfloor T_i \rfloor - \sum_{h \in [m]} \lfloor T_i^h \rfloor, B = |C| - \sum_{i \in S} \lfloor T_i \rfloor$. We construct a min-cost flow network $G = (V, A)$ with unit capacities and costs on the edges as well as demands on the nodes as follows: (1) $V = V_1 \cup V_2 \cup V_3 \cup V_4 \cup V_5$, where $V_1 = \{s\}$ with demand $|C|$, $V_2 = \{v_j \mid j \in C\}$ with demand 0, $V_3 = \{v_i^h \mid i \in S, h \in H\}$ with demand $-\lfloor T_i^h \rfloor$, $V_4 = \{v_i \mid i \in S\}$ with demand $-B_i$, and $V_5 = \{t\}$ with demand $-B$. (2) $A = A_1 \cup A_2 \cup A_3 \cup A_4$, where $A_1 = \{(s, v_j) \mid j \in C\}$ with cost 0, $A_2 = \{(v_j, v_i^h) \mid j \in C_h, i \in S, h \in [m], x_{ij} > 0)\}$ with cost $d(i,j)$, $A_3 = \{(v_i^h, v_i) \mid i \in S, h \in [m], T_i^h - \lfloor T_i^h \rfloor > 0\}$ with cost 0, and $A_4 = \{(v_i, t) \mid i \in S, T_i - \lfloor T_i \rfloor > 0\}$ with cost 0.

Observe that the sum of the demands over all nodes in V is equal to zero. Hence, the flow conservation holds. Moreover, all capacities, costs, and demands are integral. Consequently, the min-cost flow network $G = (V, A)$ admits an integral optimal solution. Let $\boldsymbol{f}$ be an optimal flow. It naturally induces a solution $\overline{\mathbf{x}}$ by setting $\overline{x}_{ij}$ equal to the flow value on edge $(v_j, v_i^h) \in A_2$ for each $j \in C$ and $i \in S$. Since the flow $\boldsymbol{f}$ is integral, this implies that $\overline{\mathbf{x}}$ is integral. Furthermore, since $\overline{x}_{ij} > 0$ only when $d(i,j) \leq (1+\varepsilon)\tau_f$, we have cost$(\overline{\mathbf{x}}) \leq (1+\varepsilon)\tau_f$.

We now show that the integral solution $\overline{\mathbf{x}}$ incurs an additive 1 violation of the group fairness constraints. For each $i \in S$ and $h \in [m]$, the demands of nodes v_i^h and v_i imply that at least $\lfloor T_i^h \rfloor$ points with color h and at least $\lfloor T_i \rfloor$ points in total are assigned to center i. On the other hand, since the arcs (v_i^h, v_i) and (v_i, t) have unit capacity, and each node v_i^h and v_i has at most one outgoing arc, at most $\lceil T_i^h \rceil$ points with color h and at most $\lceil T_i \rceil$ points in total can be assigned to i. Therefore, we have $\lfloor T_i^h \rfloor \leq \sum_{j \in C_h} \overline{x}_{ij} \leq \lceil T_i^h \rceil$ and $\lfloor T_i \rfloor \leq \sum_{j \in C} \overline{x}_{ij} \leq \lceil T_i \rceil$. By the definition of violation in [Bercea *et al.*, 2019], this implies that $\overline{\mathbf{x}}$ satisfies the group fairness constraints with an additive 1 violation.

Finally, we analyze the running time. The time required to compute the candidate set E is polynomial by Theorem 3. We then enumerate all subsets of E of size k, compute the corresponding clustering, and output the subset with minimum cost. There are $|E|^k$ such subsets. Since the metric space has bounded doubling dimension $D = O(1)$, we have $|E|^k = k^k \cdot (4/\varepsilon)^{Dk} = (k\varepsilon^{-1})^{O(k)}$. For each subset of size k, solving the associated linear programming and min-cost flow network is polynomial. Thus, the overall running time is bounded by $(k\varepsilon^{-1})^{O(k)} n^{O(1)}$, which is FPT($k$)-time. □

Lemma 3 shows that the min-cost flow network yields a $(1+\varepsilon)$-approximate solution with an additive 1 violation of the fairness constraints. Indeed, the violation can be removed using the mixed-integer linear programming of [Bandyapadhyay *et al.*, 2021] that introduces additional variables to capture the number of points with each color assigned to the centers in S. In particular, one can obtain a $(1+\varepsilon)$-approximate solution that satisfies the fairness constraints exactly in FPT(k, m)-time [Bandyapadhyay *et al.*, 2021], incurring an additional dependence on the number of colors m in running time. Compared with the method of [Bandyapadhyay *et al.*, 2021], our use of a min-cost flow network has polynomial running time, and is more efficient in practice. Combining these results completes the proof of Theorem 1.

Algorithm 2: An algorithm for DSF-k-CEN

Input: An instance $((\mathcal{X}, d), C, \mathcal{P}, \vec{\theta}, m, k)$ of the DSF-k-CEN problem and parameters ε, τ_f

Output: A solution $\hat{S}$

Let E be the candidate set constructed by Theorem 3;
for *each multi-set* $S \subseteq E$ **do**
  Construct a bipartite graph G based on S;
  Obtain $\hat{S}$ by maximum matching method on G;
Output $\hat{S}$ such that the cost of $\hat{S}$ is minimized among all considered subsets;
return $\hat{S}$.

4.2 The DSF-k-CEN Problem

In this section, we study the DSF-k-CEN problem, and employ the matching-based method of [Jones *et al.*, 2020] to solve it. Specifically, by combining our candidate-based structural method with the matching algorithm, we obtain a $(1+\varepsilon)$-approximate solution without fairness violation for the DSF-k-CEN problem with running time $(k\varepsilon^{-1})^{O(k)} n^{O(1)}$. Given the set S of ε-optimal cluster centers, the algorithm computes a maximum matching based on the bipartite graph, which induces a set $\hat{S}$ of k centers satisfying data summarization fairness constraints (see Algorithm 2). Theorem 2 then follows from the result stated below.

Lemma 4. *Given an instance* $((\mathcal{X}, d), C, \mathcal{P}, \vec{\theta}, m, k)$ *of the* DSF-k-CEN *problem and a real number* $\varepsilon > 0$*, let* S *be the set of* ε*-optimal cluster centers. Then, we can obtain a* $(1+\varepsilon')$*-approximate solution without fairness violation of this instance in FPT time, where* $\varepsilon' = 2\varepsilon$.

Proof. We first prove the existence of a $(1+\varepsilon')$-approximate solution for the given instance. Recall that the set S of ε-optimal cluster centers satisfies the following property: for each optimal center $s_i^* \in S^*$ ($i \in [k]$), there exists a center in S at distance at most $\varepsilon\tau_f$ from s_i^*. Consequently, for each $s_i \in S$ ($i \in [k]$), there must exist a point, denoted by $f(s_i)$, with the same color as s_i^* (possibly s_i^* itself) such that $d(s_i, f(s_i)) \leq \varepsilon\tau_f$. Let $F = \{f(s_1), \ldots, f(s_k)\}$. Then, for any point $v \in O_i^*$ ($i \in [k]$), by the triangle inequality, we have that $d(v, f(s_i)) \leq d(v, s_i) + d(s_i, f(s_i)) \leq (1+\varepsilon)\tau_f + \varepsilon\tau_f \leq (1+2\varepsilon)\tau_f$. Thus, for each point $v \in C$, we have $d(v, F) \leq (1+2\varepsilon)\tau_f$. Assigning each point in C to its closest center in F yields a feasible solution with cost at most $(1+2\varepsilon)\tau_f$. Let $\varepsilon' = 2\varepsilon$. Hence, there exists a $(1+\varepsilon')$-approximate solution for the given instance.

We now show how the matching algorithm constructs such a desired solution. The algorithm starts with a bipartite graph $G = (V_1 \cup V_2, A)$. The left vertex set V_1 contains k vertices in total, one vertex u_i for each center $s_i \in S$ ($i \in [k]$). The right vertex set is $V_2 = \cup_{h=1}^m V_h$, where V_h contains exactly k_h identical vertices corresponding to color h ($h \in [m]$). For

Problem	Dataset	n	d	m
GF-k-CEN	Reuters	2,500	10	50
	Victorian	4,500	10	45
	4area	35,385	8	4

Table 2: Datasets.

each $i \in [k]$ and $h \in [m]$, if there is a point $v \in C_h$ with $d(v, s_i) \leq \varepsilon\tau_f$, then we connect u_i to all vertices in V_h. Let M be the maximum matching returned by the Ford-Fulkerson algorithm based on G that runs in polynomial time. Then, M induces immediately a solution $\hat{S}$ as follows. For each matched edge $(u_i, b) \in M$, assume that vertex b is in V_h. We select a point $p \in C_h$ with $d(p, s_i) \leq \varepsilon\tau_f$, and include p in $\hat{S}$. By construction, for each center in S, there exists at least one such point matching the color of the corresponding optimal center in S^*. Moreover, these balls centered at s_i ($s_i \in S$) with radius $\varepsilon\tau_f$ are pairwise disjoint. Thus, the number of edges in maximum matching M is equal to k. Since $|V_h| = k_h$, and M is a maximum matching, $\hat{S}$ contains exactly k_h points with color h. Therefore, $\hat{S}$ satisfies data summarization fairness constraints. Then, by assigning each point $v \in C$ to its closest center in $\hat{S}$, we obtain a feasible solution $\hat{S}$ with cost $(1+\varepsilon)\tau_f$. The running time analysis follows the same argument as in Lemma 3. Since the matching algorithm runs in polynomial time, the total running time is $(k\varepsilon^{-1})^{O(k)}n^{O(1)}$, which is FPT($k$)-time. □

5 Experiments

While the primary focus of this paper is on theoretical aspect, we use the GF-k-CEN problem problem as an illustrative example, and try to compare our proposed parameterized approximation algorithm with the state-of-the-art algorithms in this section. All linear programs are solved using CPLEX. The experiments are conducted on a machine equipped with 72 Intel Xeon Gold 6230 CPUs and 500GB of memory.

Datasets. We conduct experiments on 3 real datasets (i.e., Reuters, Victorian and 4area) frequently used in fair clustering [Harb and Lam, 2020]. Reuters dataset contains 50 English language texts from each of 50 authors, where author is considered as the sensitive attribute to generate 50 groups. Victorian dataset contains the texts from 45 English language authors, in which each text consists of 1000-word sequences obtained from a book of the author. We use author as sensitive attribute to get 45 groups. 4area dataset contains researchers from four areas of computer science, where the main area of research is considered as the sensitive attribute to generate 4 groups. We normalize all used datasets to have zero mean and standard deviation of one. The datasets used in our experiments are summarized in Table 2.

Baseline Algorithms. We use two baseline algorithms in our experiments. The first one is the classical greedy algorithm that is a 2-approximation for the k-center problem, denoted as GREEDY-kC. We compare the cost returned by GREEDY-kC that is the maximum distance from a point to its closest center, and is also a lower bound of the optimal cost for the GF-k-CEN instance. Note that GREEDY-kC is not strictly a baseline algorithm, and it is included to provide a reference lower bound for evaluating the performance of fair clustering algorithm. The second one is a 3-approximation with additive violation 7 by [Harb and Lam, 2020], denoted as HARB. Indeed, we do not consider the algorithm in [Bercea *et al.*, 2019] that achieves a 3-approximation with additive violation 1 (note that the definition of violation is different from one in [Harb and Lam, 2020]), since the algorithm in [Harb and Lam, 2020] has better performance.

Dataset	$\vec{\alpha}$	$\vec{\beta}$	GREEDY-kC	HARB	Our
Reuters	0	0.05	2.146	2.681	**2.268**
	0	0.20	2.146	2.680	**2.268**
	0	0.40	2.146	2.680	**2.268**
Victorian	0	0.10	5.162	5.971	**5.369**
	0	0.30	5.162	5.966	**5.366**
	0	0.50	5.162	5.962	**5.365**
4area	0	0.45	12.651	13.226	**12.834**
	0	0.60	12.651	13.224	**12.834**
	0	0.80	12.651	13.219	**12.834**

Table 3: Comparison of average costs returned by the approximation algorithms for the GF-k-CEN problem on real datasets.

Parameters and Results. Our parameterized algorithm begins by computing a candidate set of size $k \cdot (4/\varepsilon)^D$. However, the doubling dimension D is hard to compute in practice, and is thus not treated as an explicit parameter in experiments. Following the settings in [Ceccarello *et al.*, 2019; Ceccarello *et al.*, 2024; Wu *et al.*, 2024], we instead vary the size of the candidate set directly. In our experiments, we fix the candidate set size to $3k$. Tables 3 compares the average clustering costs achieved by different algorithms for the GF-k-CEN problem. In particular, for the GF-k-CEN problem, we choose feasible values of $\vec{\alpha}, \vec{\beta}$ as in [Harb and Lam, 2020], and fix $k = 3$. As shown in Table 3, our parameterized algorithm consistently achieves lower clustering cost across a variety of datasets and parameter settings. Note that we do not include a comparison of running times in experiments. As expected, the running time of our parameterized algorithm is significantly higher than that of HARB, since our algorithm enumerate all possible subsets of candidate centers in order to obtain solutions with smaller clustering cost.

6 Conclusions

In this paper, we study the k-center problem under group fairness and data summarization fairness constraints in doubling metrics. Our proposed algorithms can be easily extended to the k-supplier variant, in which the set of selected centers is different from the set of data points. It is also interesting to extend our method to the setting where data points may belong to multiple demographic groups. More broadly, given the growing interest in theoretical foundations of fair clustering, we believe that the method developed in this paper will be applicable to a wide range of fairness-aware variants of the k-center problem.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62432016, 62502545) and the Central South University Research Program of Advanced Interdisciplinary Studies (No. 2023QYJC023).

References

[Aggarwal *et al.*, 2010] Gagan Aggarwal, Rina Panigrahy, Tomás Feder, Dilys Thomas, Krishnaram Kenthapadi, Samir Khuller, and An Zhu. Achieving anonymity via clustering. *ACM Transactions on Algorithms*, 6(3):49:1–49:19, 2010.

[Aghamolaei and Ghodsi, 2018] Sepideh Aghamolaei and Mohammad Ghodsi. A composable coreset for k-center in doubling metrics. In *Proceedings of the 30th Canadian Conference on Computational Geometry*, pages 165–171, 2018.

[Ahmadian *et al.*, 2019] Sara Ahmadian, Alessandro Epasto, Ravi Kumar, and Mohammad Mahdian. Clustering without over-representation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 267–275, 2019.

[Angelidakis *et al.*, 2022] Haris Angelidakis, Adam Kurpisz, Leon Sering, and Rico Zenklusen. Fair and fast k-center clustering for data summarization. In *Proceedings of the 39th International Conference on Machine Learning*, pages 669–702, 2022.

[Assouad, 1983] Patrice Assouad. Plongements lipschitziens dans R^n. *Bulletin de la Société Mathématique de France*, 111:429–448, 1983.

[Bandyapadhyay *et al.*, 2021] Sayan Bandyapadhyay, Fedor V Fomin, and Kirill Simonov. On coresets for fair clustering in metric and euclidean spaces and their applications. In *Proceedings of the 48th International Colloquium on Automata, Languages, and Programming*, pages 23:1–23:15, 2021.

[Bera *et al.*, 2019] Suman Kalyan Bera, Deeparnab Chakrabarty, Nicolas Flores, and Maryam Negahbani. Fair algorithms for clustering. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 4955–4966, 2019.

[Bercea *et al.*, 2019] Ioana Oriana Bercea, Martin Groß, Samir Khuller, Aounon Kumar, Clemens Rösner, Daniel Schmidt, and Melanie Schmidt. On the cost of essentially fair clusterings. In *Proceedings of the 22nd International Conference on Approximation Algorithms for Combinatorial Optimization Problems and 23rd International Conference on Randomization and Computation*, pages 18:1–18:22, 2019.

[Ceccarello *et al.*, 2017] Matteo Ceccarello, Andrea Pietracaprina, Geppino Pucci, and Eli Upfal. Mapreduce and streaming algorithms for diversity maximization in metric spaces of bounded doubling dimension. *Proceedings of the VLDB Endowment*, 10(5):469–480, 2017.

[Ceccarello *et al.*, 2019] Matteo Ceccarello, Andrea Pietracaprina, and Geppino Pucci. Solving k-center clustering (with outliers) in mapreduce and streaming, almost as accurately as sequentially. *Proceedings of the VLDB Endowment*, 12(7):766–778, 2019.

[Ceccarello *et al.*, 2024] Matteo Ceccarello, Andrea Pietracaprina, and Geppino Pucci. Fast and accurate fair k-center clustering in doubling metrics. In *Proceedings of the 33rd ACM Web Conference*, pages 756–767, 2024.

[Charikar *et al.*, 2001] Moses Charikar, Samir Khuller, David Mount, and Giri Narasimhan. Algorithms for facility location problems with outliers. In *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 642–651, 2001.

[Chen *et al.*, 2019] Xingyu Chen, Brandon Fain, Liang Lyu, and Kamesh Munagala. Proportionally fair clustering. In *Proceedings of the 36th International Conference on Machine Learning*, pages 1032–1041, 2019.

[Chierichetti *et al.*, 2017] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 5029–5037, 2017.

[Cohen-Addad *et al.*, 2021] Vincent Cohen-Addad, Andreas Emil Feldmann, and David Saulpic. Near-linear time approximation schemes for clustering in doubling metrics. *Journal of the ACM*, 68(6):1–34, 2021.

[Cohen-Addad, 2020] Vincent Cohen-Addad. Approximation schemes for capacitated clustering in doubling metrics. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2241–2259, 2020.

[Cygan *et al.*, 2012] Marek Cygan, MohammadTaghi Hajiaghayi, and Samir Khuller. LP rounding for k-centers with non-uniform hard capacities. In *Proceedings of the 53rd Annual Symposium on Foundations of Computer Science*, pages 273–282, 2012.

[Demaine *et al.*, 2005] Erik Demaine, Fedor Fomin, MohammadTaghi Hajiaghayi, and Dimitrios Thilikos. Fixed-parameter algorithms for (k, r)-center in planar graphs and map graphs. *ACM Transactions on Algorithms*, 1(1):33–47, 2005.

[Ding and Xu, 2020] Hu Ding and Jinhui Xu. A unified framework for clustering constrained data without locality property. *Algorithmica*, 82(4):808–852, 2020.

[Ding *et al.*, 2019] Hu Ding, Haikuo Yu, and Zixiu Wang. Greedy strategy works for k-center clustering with outliers and coreset construction. In *Proceedings of the 27th Annual European Symposium on Algorithms*, pages 40–1, 2019.

[Esmaeili *et al.*, 2021] Seyed Esmaeili, Brian Brubach, Aravind Srinivasan, and John Dickerson. Fair clustering under a bounded cost. In *Proceedings of the 34th International Conference on Neural Information Processing Systemss*, pages 14345–14357, 2021.

[Friggstad *et al.*, 2019] Zachary Friggstad, Mohsen Rezapour, and Mohammad Salavatipour. Local search yields a PTAS for k-means in doubling metrics. *SIAM Journal on Computing*, 48(2):452–480, 2019.

[Garey and Johnson, 1979] Michael Randolph Garey and David Stifler Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. WH Freeman, 1979.

[Gonzalez, 1985] Teofilo Francisco Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

[Goranci *et al.*, 2021] Gramoz Goranci, Monika Henzinger, Dariusz Leniowski, Christian Schulz, and Alexander Svozil. Fully dynamic k-center clustering in low dimensional metrics. In *Proceedings of the 23rd Symposium on Algorithm Engineering and Experiments*, pages 143–153, 2021.

[Goyal and Jaiswal, 2023] Dishant Goyal and Ragesh Jaiswal. Tight FPT approximation for constrained k-center and k-supplier. *Theoretical Computer Science*, 940:190–208, 2023.

[Gupta *et al.*, 2003] Anupam Gupta, Robert Krauthgamer, and James Lee. Bounded geometries, fractals, and low-distortion embeddings. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 534–543, 2003.

[Harb and Lam, 2020] Elfarouk Harb and Ho Shan Lam. KFC: A scalable approximation algorithm for k-center fair clustering. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, pages 14509–14519, 2020.

[Hochbaum and Shmoys, 1985] Dorit Hochbaum and David Bernard Shmoys. A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, 10(2):180–184, 1985.

[Hsu and Nemhauser, 1979] Wen-Lian Hsu and George Lann Nemhauser. Easy and hard bottleneck location problems. *Discrete Applied Mathematics*, 1(3):209–215, 1979.

[Huang *et al.*, 2018] Lingxiao Huang, Shaofeng Jiang, Jian Li, and Xuan Wu. Epsilon-coresets for clustering (with outliers) in doubling metrics. In *Proceedings of the 59th Annual Symposium on Foundations of Computer Science*, pages 814–825, 2018.

[Jones *et al.*, 2020] Matthew Jones, Huy Lê Nguyên, and Thy Nguyen. Fair k-centers via maximum matching. In *Proceedings of the 37th International Conference on Machine Learning*, pages 4940–4949, 2020.

[Kay *et al.*, 2015] Matthew Kay, Cynthia Matuszek, and Sean Munson. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3819–3828, 2015.

[Khuller and Sussmann, 2000] Samir Khuller and Yoram Jacob Sussmann. The capacitated k-center problem. *SIAM Journal on Discrete Mathematics*, 13(3):403–418, 2000.

[Kleindessner *et al.*, 2019] Matthäus Kleindessner, Pranjal Awasthi, and Jamie Morgenstern. Fair k-center clustering for data summarization. In *Proceeding of the 36th International Conference on Machine Learning*, pages 3448–3457, 2019.

[Liang *et al.*, 2025] Ting Liang, Xiaoliang Wu, Jinhui Xu, and Qilong Feng. The distributed algorithms for the lower-bounded k-center clustering in metric space. *Theoretical Computer Science*, 1027:114975, 2025.

[Mahabadi and Vakilian, 2020] Sepideh Mahabadi and Ali Vakilian. Individual fairness for k-clustering. In *Proceedings of the 37th International Conference on Machine Learning*, pages 6586–6596, 2020.

[Micha and Shah, 2020] Evi Micha and Nisarg Shah. Proportionally fair clustering revisited. In *Proceedings of the 47th International Colloquium on Automata, Languages, and Programming*, pages 85:1–85:16, 2020.

[Negahbani and Chakrabarty, 2021] Maryam Negahbani and Deeparnab Chakrabarty. Better algorithms for individually fair k-clustering. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, pages 13340–13351, 2021.

[Pellizzoni *et al.*, 2025] Paolo Pellizzoni, Andrea Pietracaprina, and Geppino Pucci. Fully dynamic clustering and diversity maximization in doubling metrics. *ACM Transactions on Knowledge Discovery from Data*, 19(4):1–45, 2025.

[Thejaswi *et al.*, 2022] Suhas Thejaswi, Ameet Gadekar, Bruno Ordozgoiti, and Michal Osadnik. Clustering with fair-center representation: Parameterized approximation algorithms and heuristics. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1749–1759, 2022.

[Wu *et al.*, 2024] Xiaoliang Wu, Qilong Feng, Ziyun Huang, Jinhui Xu, and Jianxin Wang. New algorithms for distributed fair k-center clustering: almost accurate as sequential algorithms. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, pages 1938–1946, 2024.

[Wu *et al.*, 2025] Xiaoliang Wu, Qilong Feng, Junyu Huang, and Jianxin Wang. Parameterized approximation algorithm for doubly constrained fair clustering. In *Proceedings of the 34th International Joint Conference on Artificial Intelligence*, pages 592–600, 2025.

[Zhang *et al.*, 2024] Zhen Zhang, Xiaohong Chen, Limei Liu, Jie Chen, Junyu Huang, and Qilong Feng. Parameterized approximation schemes for fair-range clustering. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, pages 60192–60211, 2024.