

# Linear Time Algorithms for Individually Fair $k$ -means via Multi-Swap Local Search

Beirong Cui, Qilong Feng\*, Junyu Huang

School of Computer Science and Engineering, Central South University, Changsha 410083, China

## Abstract

Fair clustering has attracted increased attention in recent years. In this work, we study the individually fair clustering problem in Euclidean space. While single-swap local search methods have achieved near-linear running time and constant approximation guarantees, their performance often depends on the aspect ratio of the dataset (the ratio between the diameter and the minimum interpoint distance of the dataset). How to apply multi-swap local search while obtaining linear running time with better approximation ratio is still a challenging task. To address this, we introduce a collaborative initialization framework for that integrates greedy with sampling techniques. This framework eliminates the dependence on the aspect ratio and produces a constant-factor bicriteria approximation in linear time. In contrast to the current state-of-the-art near-linear time algorithm, which requires a restrictive assumption about the relationship between optimal centers and cluster centroids, we propose a multi-swap local search algorithm that provides an improved approximation guarantee. Our method runs in linear time with high probability and does not rely on the aforementioned assumption. We validate our theoretical results through extensive experiments on both real-world and synthetic datasets, including large-scale benchmarks with up to 100 million points. Our empirical evaluation demonstrates superior performance in terms of clustering quality and computational efficiency, along with scalability under varying parameter settings.

## Introduction

Clustering is a fundamental problem in unsupervised learning and has been extensively studied for decades. Among various clustering objectives,  $k$ -means is one of the most widely used in practice. Given a dataset  $P \subseteq \mathbb{R}^d$ , the  $k$ -means problem seeks to find at most  $k$  centers  $C \subseteq \mathbb{R}^d$  that minimize the sum of squared Euclidean distances from each point to its nearest center.

The standard  $k$ -means problem has been extensively studied over the past decades, leading to numerous algorithmic advances and practical solutions (Arthur and Vassilvitskii 2006; Bahmani et al. 2012; Cohen-Addad et al. 2022). A significant line of work focuses on developing constant-factor approximation algorithms with linear or near-linear runtime

in the data size (Aggarwal, Deshpande, and Kannan 2009; Arthur and Vassilvitskii 2006; Beretta et al. 2023; Huang et al. 2024; Lattanzi and Sohler 2019). However, in many real-world applications optimizing clustering cost alone is insufficient. Fairness has become a key concern in how data points are assigned to clusters. For instance, in placing public facilities such as hospitals (Vakilian and Yalciner 2022), it is important to ensure that no individual is too far from their nearest facility. In such scenarios, individual fairness becomes crucial: it guarantees that every person has reasonable access to a nearby center, thereby preventing systemic inequities that may arise under standard clustering objectives. Incorporating such fairness constraints not only promotes equity but also enhances the ethical and societal reliability of algorithmic systems.

Motivated by individual fairness, (Mahabadi and Vakilian 2020) introduced the individually fair  $k$ -clustering problem. Given a dataset  $P$  of  $n$  points in a metric space, each point  $p \in P$  is associated with a fair radius  $\delta(p)$ , defined as the smallest radius such that the ball of radius  $\delta(p)$  centered at  $p$  contains at least  $n/k$  points from  $P$ . This definition captures the idea that every individual should have a cluster center within a distance that reflects the local density of the data. The goal of individually fair clustering is to select at most  $k$  centers such that every point is assigned to a center within its fair radius. In this setting, (Jung, Kannan, and Lutz 2020) established a fundamental hardness result: in Euclidean spaces of dimension  $d > 1$ , no algorithm can achieve a fairness approximation ratio better than  $\sqrt{2}$ , highlighting the intrinsic difficulty of satisfying fairness constraints exactly. To enable approximation algorithms, (Mahabadi and Vakilian 2020) formulated individually fair clustering as a bi-objective optimization problem, aiming to simultaneously minimize the clustering cost while satisfying individual fairness constraints. (Negahbani and Chakrabarty 2021) addressed the  $k$ -means setting with a polynomial-time  $(8, 4)$ -bicriteria approximation algorithm that combines linear programming relaxation with randomized rounding techniques. A more general framework was introduced in (Vakilian and Yalciner 2022), achieving a  $(16^p + \epsilon, 3)$ -bicriteria approximation for fair  $k$ -clustering in polynomial time, where  $p$  is a parameter determined by the clustering objective. Using a multi-swap local search strategy, (Mahabadi and Vakilian 2020) proposed a  $(84, 7)$ -bicriteria approximation algo-

\*Corresponding author: csufeng@mail.csu.edu.cn  
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

rithm for the individually fair  $k$ -median problem. However, this algorithm relies on enumeration to construct swap pairs at each local search step, which results in a high polynomial runtime of  $O(n^5 k^5 d \log(n\Delta))$ , which limits scalability for large datasets. Building on this line of work, (Bateni et al. 2024) introduced a single-swap local search algorithm (SSLS), achieving an  $(O(1), O(1))$ -bicriteria approximation in  $O(ndk^2 \log(n\Delta))$  time. While this represents a step toward efficiency, several limitations remain:

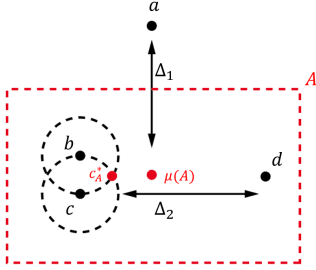


Figure 1: Failure of the Centroid Assumption

- (1) The number of local search iterations in SSLS (Bateni et al. 2024) depends heavily on the initial solution. Since initialization relies on the aspect ratio  $\Delta$ , the runtime has a logarithmic dependence on  $\Delta$ , leading to  $O(ndk^2 \log(n\Delta))$  running time. In worst-case instances,  $\Delta$  can be as large as  $2^{n^{o(1)}}$  (Cohen-Addad, Mirrokni, and Zhong 2022), potentially making SSLS polynomial in time and limiting scalability. Moreover,  $\Delta$  may be arbitrarily large in practice (Huang et al. 2023; Nguyen, Nguyen, and Jones 2022). To the best of our knowledge, no existing algorithm for individually fair  $k$ -means achieves an  $O(1)$ -approximation in near-linear or linear time while maintaining fairness. Thus, designing efficient methods for computing high-quality initial solutions in linear time remains a key challenge.
- (2) Under individual fairness constraints, a key structural assumption in the analysis of SSLS no longer holds. Specifically, the cost decomposition identity  $\text{cost}(P^*, c) = |P^*| \cdot \rho^2(c, c^*) + \text{cost}(P^*, c^*)$  (see Section 2 for notation) for an optimal cluster  $P^*$ , any point  $c$ , and its optimal center  $c^*$ , depends on  $c^*$  coinciding with the centroid  $\mu(P^*)$ . This condition is generally violated under individual fairness. To illustrate, consider the example in Figure 1, where  $P = \{a, b, c, d\}$  and  $k = 2$ . Assume  $\Delta_1 \gg \Delta_2 \gg \rho(b, c)$ . The optimal clustering assigns  $a$  to one cluster and  $A = \{b, c, d\}$  to the other, with the optimal center  $c_A^*$  placed at the intersection of fairness-enforced balls rather than near  $\mu(A)$ . As  $\Delta_2$  grows, the separation between  $c_A^*$  and  $\mu(A)$  increases, breaking the centroid-based cost decomposition. This breaks a core analytical tool of SSLS, undermining its performance and limiting its applicability in fair clustering settings.
- (3) The approximation guarantees of SSLS involve large hidden constants, with at least 2000 for the clustering cost and at least 6 for the fairness violation (relying on the

assumption). Multi-swap approach (Mahabadi and Vakilian 2020) has so far lacked comparable runtime performance despite their better approximation ratios. Bridging this gap by achieving both strong theoretical guarantees and linear running time remains a challenge in the development of scalable fair clustering algorithms.

To address these challenges, we propose a novel multi-swap local search framework that achieves a  $(62, 7)$ -bicriteria approximation in linear time.

Firstly, as shown in (Mahabadi and Vakilian 2020), the cost of a worst-case  $k$ -means solution can reach  $n\Delta^2$ , which introduces a  $\log(n\Delta)$  factor in the runtime. To remove this dependence, we propose a collaborative initialization framework that combines greedy selection with sampling. This approach achieves an  $(O(1), 4)$ -bicriteria approximation in linear time  $O(ndk + \text{poly}(k))$ , independent of the aspect ratio  $\Delta$ . As a result, it provides stronger theoretical guarantees and significantly improves practical efficiency.

Secondly, since the cost decomposition identity does not generally hold under individual fairness constraints, we generalize it to an inequality that applies to any set  $P \subseteq \mathbb{R}^d$  and any two points  $c_1, c_2 \in \mathbb{R}^d$ :  $\text{cost}(P, c_1) \leq 2|P| \cdot \rho^2(c_1, c_2) + 2 \text{cost}(P, c_2)$ . This minor relaxation allows our algorithm to be applied in general Euclidean settings, thereby enhancing its applicability and generalization.

Finally, performing multi-swap local search under individual fairness requires a sampling-based approach, as enumeration is computationally prohibitive. However, sampling raises two challenges: how to select valid swap pairs that include points near optimal centers while satisfying fairness constraints, and how to combine multiple pairs to achieve improvements beyond single swaps. To address the first, we introduce a greedy construction of anchor points and a novel center adjustment mechanism that replaces unfair candidate centers with suitable anchors, preserving fairness while reducing cost. For the second, we design an efficient algorithm that generates a small set of candidate multi-swap pairs from the sampled points, with size independent of  $n$ . We further demonstrate that, under certain conditions, this set contains at least one valid multi-swap that improves the solution in each iteration.

Our contributions are summarized as follows:

- We propose a collaborative initialization framework for individually  $k$ -means problem that combines greedy and sampling methods, achieving an  $(O(1), 4)$ -bicriteria approximation in  $O(ndk + \text{poly}(k))$  time.
- We propose a linear-time multi-swap local search algorithm for individually fair  $k$ -means, achieving a  $(62, 7)$ -bicriteria approximation in  $O(nd \cdot \text{poly}(k))$  time with constant probability. The algorithm removes a key assumption in SSLS and outperforms its  $(2000, 6)$ -bicriteria approximation, which requires  $O(ndk^2 \log(n\Delta))$  time.
- Extensive performance tests and parameter sensitivity analyses demonstrate the effectiveness of our proposed algorithms on both real-world and synthetic datasets.

## Preliminaries

In this paper, we consider a dataset  $P \subseteq \mathbb{R}^d$  of size  $n$ , where  $d$  denotes the dimensionality of the space. We let  $k$  and  $t$  denote the number of clusters and the size of swap pairs, respectively. For any two points  $p, q \in \mathbb{R}^d$ , we define their Euclidean distance as  $\rho(p, q)$ . This definition extends to sets as follows: the distance between two sets  $P, Q \subseteq \mathbb{R}^d$  is given by  $\rho(P, Q) := \inf_{x \in P, y \in Q} \rho(x, y)$ . For any positive integer  $l$ , we write  $[l]$  to denote the set  $\{1, 2, \dots, l\}$ . The centroid of a set  $P$  is defined as  $\mu(P) = \frac{1}{|P|} \sum_{p \in P} p$ . Given a set of clustering centers  $C \subseteq \mathbb{R}^d$  with  $|C| = k$ , the clustering cost is defined as  $\text{cost}(P, C) := \sum_{p \in P} \min_{c \in C} \rho^2(p, c)$ . To formalize the notion of individual fairness in  $k$ -means clustering, we proceed by defining the fair radius and the concept of fairness violation approximation.

**Definition 1 (Fair Radius).** *The fair radius function  $\delta : \mathbb{R}^d \rightarrow \mathbb{R}^+$  is defined as  $\delta(a) = \min \{r \in \mathbb{R}^+ : |B(a, r) \cap P| \geq \lceil \frac{n}{k} \rceil\}$ , where  $B(a, r) = \{x \in \mathbb{R}^d : \rho(x, a) \leq r\}$  denotes the closed ball of radius  $r$  centered at  $a$ . We refer to  $\delta(a)$  as the fair radius of point  $a$ .*

**Definition 2 ( $\alpha$ -Fairness).** *Given a parameter  $\alpha > 0$ , a set of centers  $C \subseteq \mathbb{R}^d$  is said to be  $\alpha$ -fair if for every point  $p \in P$ ,  $\rho(p, C) \leq \alpha \cdot \delta(p)$ .*

According to (Jung, Kannan, and Lutz 2020), in Euclidean spaces of dimension  $d > 1$ , there exists a point set  $P$  such that for any center set  $C$ , there is at least one point  $p \in P$  satisfying  $\rho(p, C) \geq \sqrt{2} \delta(p)$ . Moreover, determining the optimal fairness parameter  $\alpha$  for a given point set and number of clusters  $k$  is NP-hard. This suggests that, for the individually fair  $k$ -means problem, the most viable approach is to seek a bicriteria approximation that trades off between fairness violation and clustering cost. In the following, we assume  $\alpha$  is a predefined parameter, and we assume there exists at least one center set  $C \subseteq \mathbb{R}^d$  that is  $\alpha$ -fair.

**Definition 3 ( $(\beta, \eta)$ -Bicriteria Approximation).** *An algorithm is a  $(\beta, \eta)$ -bicriteria approximation for the individually fair  $k$ -means problem if, for any dataset  $P \subseteq \mathbb{R}^d$ , the solution  $C_{\text{ALG}}$  (a set of  $k$  clustering centers) satisfies the following two properties:*

1. *Cost Approximation:*  $\text{cost}(P, C_{\text{ALG}}) \leq \beta \cdot \text{cost}(P, C_{\text{OPT}})$ , where  $C_{\text{OPT}}$  denotes the optimal center set minimizing the  $k$ -means cost;
2. *Fairness Relaxation:*  $C_{\text{ALG}}$  is  $(\eta \cdot \alpha)$ -fair, meaning  $\rho(p, C_{\text{ALG}}) \leq \eta \cdot \alpha \cdot \delta(p)$  for all  $p \in P$ .

For brevity, we define  $\text{ALG} = \text{cost}(P, C_{\text{ALG}})$  and  $\text{OPT} = \text{cost}(P, C_{\text{OPT}})$ .

Let  $\mathcal{P}^* = \{P_1^*, P_2^*, \dots, P_k^*\}$  denote the optimal clustering partition of  $P$ , with corresponding optimal centers  $C^* = \{c_1^*, c_2^*, \dots, c_k^*\}$ . Let  $C = \{c_1, c_2, \dots, c_k\}$  be a set of centers produced by an algorithm, inducing a partition  $\mathcal{P} = \{P_1, P_2, \dots, P_k\}$ , where each  $P_i$  consists of all points in  $P$  assigned to center  $c_i$  under the  $k$ -means clustering. To facilitate analysis, we define two mappings:  $M : c_i \mapsto P_i$  and  $M^* : c_i^* \mapsto P_i^*$ , for  $i = 1, 2, \dots, k$ . We refer to these as  $M_c$  and  $M_{c^*}$ , respectively, when emphasizing the

center being mapped. For subsets  $I \subseteq C$  and  $O \subseteq C^*$ , we denote  $M_I^* = \bigcup_{c^* \in O} M_{c^*}^*$  as the union of optimal clusters corresponding to centers in  $O$ , and  $M_I = \bigcup_{c \in I} M_c$  as the union of clusters from centers in  $I$ .

The following lemma is a well-known decomposition of the  $k$ -means cost and is standard in clustering analysis:

**Lemma 1.** *For any point set  $P \subseteq \mathbb{R}^d$  and any point  $c \in \mathbb{R}^d$ ,*

$$\text{cost}(P, c) = |P| \cdot \rho^2(c, \mu(P)) + \text{cost}(P, \mu(P)).$$

## Multi-Swap Local Search Algorithm

In this section, we present a multi-swap local search algorithm for the individually fair  $k$ -means problem that achieves a  $(62, 7)$ -bicriteria approximation. A key challenge is sampling valid multi-swap pairs that reduce clustering cost while preserving individual fairness. Traditional approaches, such as  $D^2$ -sampling (Lattanzi and Sohler 2019), ignore fairness and thus result in fairness violations. In multi-swap settings, interactions between center replacements and cluster reassignments grow more complex. Critically, even if individual swaps are fairness-preserving in isolation, their simultaneous execution may violate fairness for certain points. Therefore, a valid multi-swap must be carefully designed to ensure that the combined effect of all center replacements does not violate fairness constraints for any point. To address these challenges, we propose a center adjustment mechanism, which ensures that each iteration improves the clustering cost while maintaining bounded fairness violation.

The full multi-swap local search procedure is formalized in Algorithm 1, which proceeds in three stages. In the first stage (Steps 1–2), we construct an anchor set  $S_0$  using a greedy strategy (Algorithm 2) that selects centers with minimal fair radii, ensuring the initial center set  $C$  satisfies individual fairness. In the second stage (Steps 4–7), we apply  $D^2$ -sampling to generate a candidate set  $Q_0$ , prioritizing locations near optimal centers to achieve significant cost reductions. However, since  $D^2$ -sampling depends only on distance, incorporating these points directly may violate fairness. To resolve this, the third stage (Steps 8–11, Algorithm 3) introduces a center adjustment mechanism: using  $S_0$  and  $Q_0$ , it constructs a candidate set  $F$  of valid multi-swap operations, which includes at least one multi-swap that simultaneously reduces clustering cost and preserves fairness. As a result, each iteration reduces the cost by a multiplicative factor of  $\Theta(1 - 1/k)$  with constant probability, leading to a  $(62, 7)$ -bicriteria approximation.

In classical  $k$ -means, the optimal center of a cluster  $P^*$  is equal to its centroid, i.e.,  $c^* = \mu(P^*)$ . However, under individual fairness constraints, centers may deviate from the centroid to ensure every point lies within its fair radius. This misalignment invalidates centroid-based assumption. To strengthen the theoretical foundation of our algorithm, we extend Lemma 1 using the following inequality:

**Lemma 2.** *For any point set  $P \subseteq \mathbb{R}^d$  and any  $c_1, c_2 \in \mathbb{R}^d$ ,*

$$\text{cost}(P, c_1) \leq 2|P| \cdot \rho^2(c_1, c_2) + 2 \cdot \text{cost}(P, c_2).$$

This inequality supports the definition of a key construct the core, which consists of points close to a near-optimal

---

**Algorithm 1: Individually Fair Multi-Swap Local Search (IFMSLS)**


---

**Input:** A dataset  $P \subseteq \mathbb{R}^d$ , number of clusters  $k$ , rounds parameter  $T$ , swap pair size  $t$ , parameters  $\varepsilon, \theta, \alpha$ .  
**Output:** A  $\theta$ -feasible center set  $C \subseteq \mathbb{R}^d$  with  $|C| \leq k$ .  
1: Initialize  $S_0 \leftarrow \text{GAS}(P, k, \alpha, 2)$ .  
2: Initialize centers  $C \leftarrow S_0$ , then randomly select  $k - |S_0|$  points from  $P$  and add to  $C$ .  
3: **for**  $i = 1 : T$  **do**  
4:   Initialize  $Q_0 \leftarrow \emptyset$ .  
5:   **for**  $j = 1 : t$  **do**  
6:     Sample  $p \in P$  with probability  $\frac{\text{cost}(p, C)}{\text{cost}(P, C)}$  and add  $p$  to  $Q_0$ .  
7:     **for**  $(Q, O) \in 2^{Q_0} \times 2^C$  s.t.  $|Q| = |O| \geq 1$  **do**  
8:        $F \leftarrow \text{CA}((Q, O), S_0, \theta, C)$ .  
9:       **for**  $(Q', O) \in F$  **do**  
10:         **if**  $\text{cost}(P, C \setminus O \cup Q') \leq (1 - \frac{\varepsilon}{k}) \cdot \text{cost}(P, C)$  **then**  
11:            $C \leftarrow C \setminus O \cup Q'$ .  
12: **return**  $C$ .

---



---

**Algorithm 2: Greedy Anchor Selection (GAS)**


---

**Input:** Dataset  $P \subseteq \mathbb{R}^d$ , number of clusters  $k$ , fairness parameter  $\alpha$ , coverage expansion factor  $\gamma$ .  
**Output:** Anchor set  $S_0 \subseteq P$  with individual fairness guarantees.  
1: Initialize anchor set:  $S_0 \leftarrow \emptyset$   
2: Initialize covered points:  $Z \leftarrow \emptyset$   
3: **while**  $Z \neq P$  **do**  
4:    $c \leftarrow \arg \min_{p \in P \setminus Z} \delta(p)$   
5:    $S_0 \leftarrow S_0 \cup \{c\}$   
6:    $Z \leftarrow Z \cup \{p \in P : \rho(p, c) \leq \gamma \alpha \cdot \delta(p)\}$   
7: **return**  $S_0$

---

center and can be efficiently sampled using  $D^2$ -sampling (Step 6 in Algorithm 1).

**Definition 4 (Core).** For a set  $E \subseteq \mathbb{R}^d$ , a point  $c$ , and  $\beta \geq 1$ , the core of  $E$  with respect to  $c$  is  $\text{core}(E, c, \beta) = E \cap B(c, R_\beta(E, c))$ , where  $R_\beta(E, c) = \sqrt{\beta \cdot \frac{\text{cost}(E, c)}{|E|}}$ .

The following lemma provides a critical sampling guarantee: if the current cost of an optimal cluster  $P^*$  accounts for a significant fraction of the optimal cost, a large fraction of this cost is concentrated in its core. This ensures that  $D^2$ -sampling selects points from the core with high probability.

**Lemma 3.** Let  $P^* \in \mathcal{P}^*$  be an optimal cluster with center  $c^*$ . If  $\text{cost}(P^*, C) \geq 2(2 + 4\varepsilon) \cdot \text{cost}(P^*, c^*)$ , then the cost of the  $\beta'$ -core of  $P^*$  under  $C$ , where  $\beta' = (1 + \varepsilon)^2$ , satisfies

$$\text{cost}(\text{core}(P^*, c^*, \beta'), C) \in \Omega(\varepsilon^3) \cdot \text{cost}(P^*, C^*).$$

To leverage the core structure effectively, we introduce a framework for constructing valid multi-swap pairs based on the relationship between the current center set  $C$  and an anchor set  $S_0$ . Each  $a \in S_0$  is an anchor point, with an associated anchor zone  $B(a, \theta \alpha \cdot \delta(a))$ , where  $\theta$  is a parameter

---

**Algorithm 3: Center Adjustment (CA)**


---

**Input:** Candidate swap pair  $(Q, O)$ , anchor set  $S_0$ , fairness slack parameter  $\theta$ , current center set  $C$ .  
**Output:** Set of  $\theta$ -feasible refined swap pairs  $F$ .  
1: Initialize feasible swap set:  $F \leftarrow \emptyset$   
2:  $A \leftarrow \{a \in S_0 : (C \setminus O \cup Q) \cap B(a, \theta \alpha \cdot \delta(a)) = \emptyset\}$   
3: **for** each  $Q' \subseteq (Q \cup A)$  such that  $|Q'| = |Q|$  **do**  
4:   **if**  $C \setminus O \cup Q'$  is  $\theta$ -feasible **then**  
5:      $F \leftarrow F \cup \{(Q', O)\}$   
6: **return**  $F$

---

specified in the algorithm. We say that the center set  $C$  is  $\theta$ -feasible with respect to  $S_0$  if, for every anchor point  $a \in S_0$ , the intersection  $C \cap B(a, \theta \alpha \cdot \delta(a))$  is non-empty. That is, each anchor zone contains at least one center from  $C$ . This condition is maintained as an invariant in our local search, guaranteeing that no anchor point is left poorly covered. The following lemma formalizes the term ‘‘anchor’’, shows that feasibility with respect to  $S_0$  serves as both a necessary and sufficient condition for fairness, up to constant factors.

**Lemma 4.** Let  $C$  be a set of  $k$  centers and  $S_0 \subseteq P$  be an anchor set constructed by Algorithm 2. Then:

1. If  $C$  is  $\theta$ -feasible, then  $C$  is  $((\theta + \gamma) \cdot \alpha)$ -fair.
2. Conversely, if  $C$  is  $(\theta \alpha)$ -fair, then  $C$  is  $\theta$ -feasible.
3. Moreover, if  $\gamma \geq 2$ , there exists an anchor set  $S_0$  of size at most  $k$  such that the above implications hold.

The lemma suggests a simple construction for  $\gamma$ -fair centers: when  $\gamma \geq 2$ , the anchor set  $S_0$  can be extended to a full set of  $k$  centers by adding  $k - |S_0|$  random points from  $P$ , yielding an initial  $\gamma$ -fair solution. To improve clustering quality, we use  $D^2$ -sampling (Step 6) to select candidate centers near optimal centers. However, directly swapping in these points may lead to severe fairness violations, as traditional nearest-point-based swap constructions fail to satisfy individual fairness constraints. To address the risk of fairness violations during center updates, we revise the standard capture relationship between optimal centers and the current solution by incorporating the anchor zone structure. This ensures that swaps preserve  $\theta$ -feasibility, thereby maintaining individual fairness throughout the local search process. We begin by classifying optimal centers based on their relationship to the anchor set  $S_0$  as follows.

**Definition 5 (Anchor and Unconstrained Optimal Centers).** A point  $c^* \in C^*$  is an anchor optimal center if it is the closest center in  $C^*$  to some point  $a \in S_0$ . In this case, we write  $a_{c^*} = a$  and say  $c^*$  is anchored at  $a$ . The set of all such centers is denoted  $A^*$ . The remaining optimal centers,  $U^* = C^* \setminus A^*$ , are called unconstrained optimal centers.

This restricted rule incorporates the relationship between centers and anchor points. We now construct the multi-swap pairs intuitively, with full formal details provided in the full version. Since optimal clusters contributing little to the current cost are unlikely to be sampled effectively via  $D^2$ -sampling, to avoid ineffective swaps, we focus on optimal clusters with cost at least  $\varepsilon \cdot \text{ALG}/k$ . The candidate multi-swap set  $H$  is built in two steps. First, for each current cen-

ter capturing between one and  $t$  such optimal centers, we let  $I$  be those centers and  $O$  be  $|I|$  lonely centers, i.e., centers capturing none, and add  $(I, O)$  to  $H$ . Second, for each center that captures more than  $t$  optimal centers, we process each captured  $c^*$  by selecting one unused center  $c_0$  and adding the swap  $(\{c^*\}, \{c_0\})$  to  $H$ . All such pairs  $(I, O)$  form the swap set  $H$ .

To guide the selection of effective swaps from the candidate set  $H$ , we classify each  $(I, O) \in H$  as good or bad based on its potential cost reduction.

**Definition 6** (Good and Bad Swap Pairs). *A multi-swap pair  $(I, O) \in H$  is good if  $\text{cost}(M_I^*, C) > (4 + \varepsilon) \cdot \text{cost}(M_I^*, C^*) + \text{Reassign}(I, O) + \frac{\varepsilon}{k} \cdot \text{cost}(P, C)$ , and bad otherwise, where  $\text{Reassign}(I, O) = \text{cost}(M_O \setminus M_I^*, C \setminus O) - \text{cost}(M_O \setminus M_I^*, C)$ .*

The intuition is that a good swap targets optimal clusters  $I$  whose current cost under  $C$  is significantly higher than their optimal cost, justifying the swap. For any good swap pair  $(I, O)$ , we can show that replacing  $O$  with  $I$  leads to a substantial decrease in total cost:  $\text{cost}(P, C \setminus O \cup I) < (1 - \frac{\varepsilon}{k}) \cdot \text{cost}(P, C) - (3 + \varepsilon) \cdot \text{cost}(M_I^*, C^*)$ . This ensures a cost reduction of at least a factor of  $1 - \varepsilon/k$ . The additional term  $(3 + \varepsilon) \cdot \text{cost}(M_I^*, C^*)$  allows us to select points from the  $\beta'$ -core of each optimal cluster, rather than requiring exact optimal centers, which occur with zero probability under  $D^2$ -sampling. This relaxation is essential for achieving a non-trivial success probability while still guaranteeing significant cost improvement. The following theorem formalizes this guarantee.

**Theorem 1.** *Let  $(I, O)$  be a good swap pair, where  $I = \{c_1^*, c_2^*, \dots, c_l^*\}$  with  $l \leq t$ . For any strongly improving set  $Q = \{q_1, q_2, \dots, q_l\}$ , defined by  $q_i \in \text{core}(M_{c_i^*}, c_i^*, \beta')$  for  $i = 1, 2, \dots, l$ , we have*

$$\text{cost}(P, C \setminus O \cup Q) \leq \left(1 - \frac{\varepsilon}{k}\right) \cdot \text{cost}(P, C).$$

While a strongly improving set  $Q$  of candidate centers can significantly reduce clustering cost, directly replacing  $O$  with  $Q$  may violate individual fairness. To preserve fairness, we introduce a center adjustment mechanism (Algorithm 3) that transforms  $Q$  into a new set  $Q'$  such that:  $|Q'| = |Q|$ ,  $C \setminus O \cup Q'$  is  $\theta$ -feasible (and thus  $((\theta + \gamma)\alpha)$ -fair by Lemma 4), and the cost reduction is preserved.

Let  $A \subseteq S_0$  be the anchor points whose anchor zones become empty after removing  $O$  (Step 2 in Algorithm 3). Our goal is to construct a set  $Q' \subseteq Q \cup A$  of size  $|Q|$  that preserves  $\theta$ -feasibility. The following proposition shows that, under mild conditions, candidate points from cores are compatible with anchor zones:

**Proposition 1.** *Let  $\theta \geq 2$ , and let  $c^* \in A^*$  that is anchored by  $a_{c^*}$ . If  $a_{c^*} \notin \text{core}(M_{c^*}, c^*, \beta')$ , then*

$$\text{core}(M_{c^*}, c^*, \beta') \subseteq B(a_{c^*}, \theta\alpha \cdot \delta(a_{c^*})).$$

Intuitively, when  $\theta \geq 2$ , the core lies within the anchor zone unless the anchor itself is in the core. Thus, sampling from the core ensures the new center lies in the anchor zone, preserving  $\theta$ -feasibility. If instead  $a_{c^*}$  is in the core, we can

explicitly include it in  $Q'$  via enumeration. In both cases, the resulting swap achieves a cost reduction by a factor of at least  $1 - \varepsilon/k$  while maintaining  $\theta$ -feasibility, and hence  $((\theta + \gamma)\alpha)$ -fairness by Lemma 4. A full proof and extension to multi-swap settings are given in the full version. Combining all components, we obtain the following theorem, which establishes the correctness and effectiveness of the center adjustment mechanism (Algorithm 3).

**Theorem 2.** *Let  $C$  be a  $\theta$ -feasible set of centers with  $\theta \geq 2$ . For every good swap pair  $(I, O) \in H$ , there exists a set  $Q' \subseteq Q \cup A$  such that  $|Q'| = |O|$  and:*

1.  $C \setminus O \cup Q'$  is  $\theta$ -feasible,
2.  $\text{cost}(P, C \setminus O \cup Q') \leq \left(1 - \frac{\varepsilon}{k}\right) \cdot \text{cost}(P, C)$ .

We now analyze the overall performance of Algorithm 1. The initial center set  $C$ , constructed in Step 2, may have cost as high as  $n\Delta^2$ . However, each successful iteration reduces the cost by a factor of  $1 - \varepsilon/k$ . By Theorem 2, this process maintains  $\theta$ -feasibility throughout. With careful parameter selection and analysis, we arrive at the main result:

**Theorem 3.** *Algorithm 1 solves the individually fair  $k$ -means problem with constant probability, producing a solution  $C$  that is:*

1. a  $(62, 7)$ -bicriteria approximation, i.e.,  $\text{cost}(P, C) \leq 62 \cdot \text{OPT}$  and  $C$  is  $7\alpha$ -fair,
2. computable in  $O(nd \cdot \text{poly}(k) \cdot \log(n\Delta))$  time.

## Fast Initialization Independent of Aspect Ratio

In the previous section, we presented a greedy initialization algorithm that achieves a constant-factor approximation in near-linear time,  $O(nd \cdot \text{poly}(k) \cdot \log(n\Delta))$ , even from an arbitrarily poor initial solution. While efficient in  $n$  and  $d$ , this runtime depends logarithmically on the aspect ratio  $\Delta$ , which can be large in practice. As shown in Lemma 4, Algorithm 2 produces a  $\gamma\alpha$ -fair solution but provides no cost guarantee. In contrast,  $D^2$ -oversampling (Aggarwal, Deshpande, and Kannan 2009) yields  $O(k)$  centers with constant expected approximation ratio, but without fairness bounds. A naive combination would result in  $O(k)$  centers, which is incompatible with the requirement of selecting exactly  $k$  centers in individually fair  $k$ -means.

To overcome the limitations of existing initialization methods, we propose a collaborative initialization framework (Algorithm 4) with three stages. First, we build a candidate set  $P_0$  of size  $O(k)$  by combining centers from (Steps 1-3): the anchor-based fair selection (Algorithm 2), ensuring  $\gamma\alpha$ -fairness, and  $D^2$ -oversampling (Algorithm 5), providing a constant-factor cost approximation in expectation. This set  $P_0$  simultaneously satisfies the fairness constraint and captures low-cost structure. Second, we reduce  $P$  to a weighted instance  $(P_0, \omega)$  by assigning each point to its nearest center in  $P_0$  (Steps 4-5), which enables clustering in time independent of  $n$  and  $\Delta$ . Finally, we show that  $P_0$  contains subset of size  $k$  that is both  $\gamma\alpha$ -fair and achieves a constant-factor cost approximation (Step 6), and that such a subset can be found in  $O(\text{poly}(k))$  time. This yields an  $(O(1), 4)$ -bicriteria approximation, effectively balancing fairness, clustering quality, and efficiency. The following lemma, based on the analysis of  $D^2$ -sampling in (Aggarwal, Deshpande, and Kannan

---

**Algorithm 4: Collaborative Initialization Framework (CIF)**

---

**Input:** A dataset  $P \subseteq \mathbb{R}^d$ , number of clusters  $k$ , parameters  $\alpha, \gamma$ , a weighted algorithm  $\mathcal{B}$  that outputs a 1-feasible solution  $C$  with  $O(1)$  clustering quality approximation.  
**Output:** An initial center set  $C_0$  with  $(O(1), 4)$ -bicriteria approximation.  
1:  $S_0 \leftarrow \text{GAS}(P, k, \alpha, 2)$ .  
2:  $S_1 \leftarrow \text{DOS}(P, k)$ .  
3:  $P_0 \leftarrow S_0 \cup S_1$ .  
4: Let  $\sigma(p) \in P_0$  be the nearest point to  $p$  in  $P_0$ .  
5: For each  $p_0 \in P_0$ , let  $\omega(p_0)$  denotes the number of points in  $P$  assigned to  $p_0$ , that is,  $\omega(p_0) = |\sigma^{-1}(p_0)|$ .  
6:  $C_0 \leftarrow \mathcal{B}(P_0, \omega, S_0, k)$ .  
7: **return**  $C_0$ .

---

---

**Algorithm 5:  $D^2$ -Oversampling (DOS)**

---

**Input:** A dataset  $P \subseteq \mathbb{R}^d$ , number of clusters  $k$ .  
**Output:** A sample set  $S_1$  of size  $O(k)$ .  
1: Initialize  $S_1 \leftarrow \emptyset$ .  
2: Randomly select a point  $p$  from  $P$  and add it to  $S_1$ .  
3: **for**  $i = 1 : O(k)$  **do**  
4:   Sample a point  $p \in P$  with probability  $\frac{\text{cost}(p, S_1)}{\text{cost}(P, S_1)}$ .  
5:    $S_1 \leftarrow S_1 \cup \{p\}$ .  
6: **return**  $S_1$ .

---

2009), establishes that Algorithm 5 produces a candidate set  $S_1$  with a constant-factor approximation to the optimal  $k$ -means cost, with constant probability.

**Lemma 5.** (Aggarwal, Deshpande, and Kannan 2009) *Let  $S_1 \subseteq P$  be the set of  $O(k)$  points selected in Algorithm 5. Then, with probability at least 0.03,  $\text{cost}(P, S_1) \leq 20 \cdot \text{OPT}_\infty$ , where  $\text{OPT}_\infty$  denotes the optimal  $k$ -means clustering cost. Furthermore, Algorithm 5 runs in  $O(ndk)$  time.*

Since the standard  $k$ -means problem is a relaxation of the individually fair variant, we have  $\text{OPT}_\infty \leq \text{OPT}$ . Thus, with constant probability,  $\text{cost}(P, P_0) \leq \text{cost}(P, S_1) \leq 20 \cdot \text{OPT}_\infty \leq 20 \cdot \text{OPT}$ . After constructing the candidate set  $P_0$  with size  $O(k)$ , we reduce the problem to clustering within  $P_0$  by assigning each point in  $P$  to its nearest center in  $P_0$ . This instance compression enables all subsequent steps to run in time independent of  $n$ , effectively eliminating dataset size from the overall time complexity. The following lemma justifies this strategy by showing that  $P_0$  contains a subset of exactly  $k$  centers that forms a constant-factor approximation to the optimal fair clustering cost.

**Lemma 6.** *For each  $c_i^* \in C^*$ , let  $c'_i \in P_0$  be the closest point in  $P_0$  to  $c_i^*$ , and  $C' = \{c'_1, \dots, c'_k\}$ . Then  $\text{cost}(P, C') \leq O(1) \cdot \text{OPT}$ .*

We now analyze the fairness of  $C'$ . Each anchor  $a \in S_0$  defines an anchor zone  $B(a, \theta\alpha \cdot \delta(a))$ , and for any optimal center  $c^*$  assigned to  $a$ , we have  $c^* \in B(a, \alpha \cdot \delta(a))$ . Let  $c' \in C'$  be the closest point in  $P_0$  to  $c^*$ . By the triangle inequality,  $\rho(c', a) \leq \rho(c', c^*) + \rho(c^*, a) \leq \rho(c', c^*) + \alpha \cdot \delta(a) \leq 2\alpha \cdot \delta(a)$ , which implies  $c' \in B(a, 2\alpha \cdot \delta(a))$ , so  $C'$  is 2-feasible.

By Lemma 4, 2-feasibility implies  $4\alpha$ -individual fairness. Combined with Lemma 6, this yields an  $(O(1), 4)$ -bicriteria approximation. Having established the existence of such a solution within  $P_0$ , the final step is to efficiently compute a  $k$ -center subset  $C_0 \subseteq P_0$  that is  $4\alpha$ -fair and achieves a constant-factor approximation on the weighted instance.

The following lemma, adapted from (Vakilian and Yalciner 2022), provides a key tool for extracting a high-quality fair solution from the candidate set.

**Lemma 7.** (Vakilian and Yalciner 2022) *Let  $C_0^*$  be the optimal 1-feasible solution for the weighted  $k$ -means problem on  $P_0$ , with cost  $\text{cost}_\omega(P_0, C) := \sum_{p \in P_0} \omega(p) \cdot \rho^2(p, C)$ , where  $\omega(p)$  is the weight of  $p \in P_0$  (assigned in Step 5 of Algorithm 4). Then there exists an algorithm that computes a 1-feasible solution  $C_0 \subseteq P_0$  runs in  $\text{poly}(k)$  time such that  $\text{cost}_\omega(P_0, C_0) \in O(1) \cdot \text{cost}_\omega(P_0, C_0^*)$ .*

Since  $C_0^*$  is optimal and  $C'$  is feasible, we have  $\text{cost}_\omega(P_0, C_0) \in O(1) \cdot \text{cost}_\omega(P_0, C')$ . We now state the main theorem for the collaborative initialization framework:

**Theorem 4.** *With constant probability, Algorithm 4 computes an  $(O(1), 4)$ -bicriteria approximate solution to the individually fair  $k$ -means problem in  $O(ndk + \text{poly}(k))$  time.*

We now revisit Algorithm 1. By replacing the initial center set, which is originally constructed in Step 2 using a method that depends on  $\log(n\Delta)$ , with the output of Algorithm 4, we eliminate the dependence on the aspect ratio  $\Delta$ . Since the new initialization runs in  $O(ndk + \text{poly}(k))$  time, we obtain the following improved guarantee:

**Theorem 5.** *With constant probability, Algorithm 1, when initialized with Algorithm 4, computes a  $(62, 7)$ -bicriteria approximate solution to the individually fair  $k$ -means problem in  $O(nd \cdot \text{poly}(k))$  time.*

## Experiment

In this section, we present empirical results of our algorithms on real-world and synthetic datasets, evaluated on a server with 72 Intel Xeon Gold 6230 CPUs and 500 GB RAM.

**Datasets.** We evaluate on diverse benchmark datasets used in prior work (Mahabadi and Vakilian 2020; Negahbani and Chakrabarty 2021; Bateni et al. 2024): Adult ( $n = 32,561, d = 6$ ), Bank ( $n = 45,211, d = 3$ ), Diabetes ( $n = 101,766, d = 2$ ), Gowalla ( $n = 100,000, d = 2$ ), Skin ( $n = 245,054, d = 4$ ), Shuttle ( $n = 58,000, d = 9$ ), and Covtype ( $n = 581,012, d = 54$ ). To assess scalability, we include two large-scale datasets: HIGGS ( $n = 10^7, d = 18$ ) and Learn ( $n = 10^8, d = 128$ ) (Jégou et al. 2011; Bateni et al. 2024). All datasets are standard normalized following (Bateni et al. 2024).

**Metrics.** We evaluate performance using three key metrics: the clustering cost defined as  $\sum_{p \in P} \rho^2(p, C)$ , the maximum bound ratio which measures fairness violation as  $\max_{p \in P} \frac{\rho^2(p, C)}{\delta(p)}$  where lower values indicate better fairness, and the running time in seconds to assess efficiency.

**Baselines.** We compare against: EMSLS (Mahabadi and Vakilian 2020) (enumeration-based local search), LP (Negahbani and Chakrabarty 2021) and its efficient variant

Dataset	Method	Cost	Fairness	Time	Dataset	Method	Cost	Fairness	Time
bank	EMSLS	3.234E+03	1.52	2428.3	diabetes	EMSLS	1.485E+03	<b>1.05</b>	2013.24
	LP	4.514E+03	1.18	22880.90		LP	1.575E+03	1.29	291.99
	LPS	4.838E+03	<b>1.16</b>	37.69		LPS	1.471E+03	1.29	3.01
	SSLS	3.289E+03	1.61	6.83		SSLS	1.733E+03	2.00	3.60
	MSLS-G	3.127E+03	1.44	<b>1.82</b>		MSLS-G	1.440E+03	1.14	<b>1.70</b>
	MSLS-W	<b>3.028E+03</b>	1.53	4.18		MSLS-W	<b>1.423E+03</b>	1.18	2.71
adult	EMSLS	8.984E+03	1.26	2762.54	shuttle	EMSLS	1.097E+04	1.80	2720.54
	LP	9.973E+03	<b>1.15</b>	25517.47		LP	1.975E+04	1.14	13160.33
	LPS	9.782E+03	1.19	280.98		LPS	1.963E+04	<b>1.12</b>	45.66
	SSLS	8.629E+03	1.31	23.21		SSLS	1.677E+04	1.95	4.89
	MSLS-G	8.686E+03	1.37	<b>1.76</b>		MSLS-G	1.088E+04	1.72	<b>1.66</b>
	MSLS-W	<b>8.620E+03</b>	1.28	2.89		MSLS-W	<b>1.059E+04</b>	1.98	3.20
gowalla	EMSLS	7.395E+02	1.47	2128.54	skin	EMSLS	1.682E+03	1.60	2187.42
	LP	9.334E+02	1.17	11367.70		LP	1.808E+03	<b>1.02</b>	4493.82
	LPS	9.990E+02	<b>1.15</b>	9.78		LPS	2.060E+03	1.13	9.26
	SSLS	7.938E+02	2.33	18.20		SSLS	1.695E+03	2.23	10.08
	MSLS-G	7.178E+02	1.69	<b>1.63</b>		MSLS-G	1.518E+03	1.75	<b>1.79</b>
	MSLS-W	<b>7.107E+02</b>	1.87	3.60		MSLS-W	<b>1.516E+03</b>	1.85	3.14

Table 1: Comparison on different datasets, with 5000 points sampled from each dataset

Dataset	Method	Cost	Fairness	Time	Dataset	Method	Cost	Fairness	Time
HIGGS	EMSLS	-	-	> 24h	Learn	EMSLS	-	-	> 24h
	LP	-	-	> 24h		LP	-	-	> 24h
	LPS	-	-	> 24h		LPS	-	-	> 24h
	SSLS	-	-	> 24h		SSLS	-	-	> 24h
	MSLS-G	3.0322E+08	1.10	<b>3396.12</b>		MSLS-G	1.2816E+10	<b>1.00</b>	<b>63147.94</b>
	MSLS-W	<b>2.9966E+08</b>	<b>1.09</b>	3783.84		MSLS-W	<b>1.2703E+10</b>	1.01	75122.80

Table 2: Comparison on large scale datasets

LPS (linear programming), SSLS (Bateni et al. 2024) (single-swap local search), and two variants of our method: MSLS-G (greedy initialization, Algorithm 2) and MSLS-W (weighted instance initialization, Algorithm 4).

**Experimental Setup.** Following the experimental setting in (Bateni et al. 2024), we execute all algorithms on each dataset 10 times, fix the number of clusters at  $k = 10$ , and set the parameters  $\varepsilon, \theta$  and  $\gamma$  as  $\varepsilon = 1/100$  and  $\theta = \gamma = 2$ . The swap pair size is set to  $t = 2$  for better scalability. For local search algorithms (SSLS, MSLS-G, and MSLS-W), we set the number of local search iterations to  $T = 500$ . To further accelerate our algorithms, we propose a distance-preserving structure that stores distances to the  $t + 1$  nearest centers each after center updates. This structure enables efficient computation of clustering cost after swap without requiring full dataset enumeration.

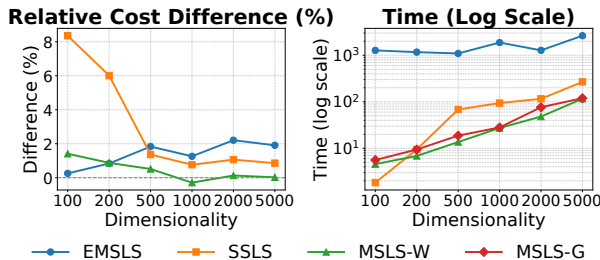


Figure 2: Comparison results with varying dimensions

**Results.** As shown in Table 1, our algorithms achieve superior performance in both clustering cost and running time. MSLS-W consistently delivers the best clustering quality

across most datasets, while MSLS-G is the fastest in terms of runtime. Both MSLS-G and MSLS-W achieve fairness guarantees comparable to the best baselines, with MSLS-W offering significant improvements in efficiency. MSLS-G and MSLS-W achieving higher clustering quality and substantially reduced runtimes compare to SSLS. Notably, MSLS-G not only runs the fastest but also improves clustering cost over all baselines. On average, MSLS-W reduces clustering cost by 5.3%, 23.8%, 25.7%, and 14.0% compared to EMSLS, LP, LPS, and SSLS, respectively. Moreover, MSLS-G is 6.5 times faster than SSLS on average.

On large-scale instances, the results in Table 2 demonstrate that MSLS-G and MSLS-W successfully complete within 24 hours on HIGGS ( $n = 10,000,000$ ) and Learn ( $n = 100,000,000$ ), while SSLS fails to finish within the time limit. Additionally, Figure 2 illustrates performance in high-dimensional settings from 100 to 5000 dimensions on synthetic dataset SYN ( $n = 5,000$ ). Our methods exhibit solid scalability as dimensionality increases, confirming their scalability in demanding scenarios.

## Conclusion

This paper proposes an effective collaborative initialization framework for individually fair  $k$ -means, which enables a linear-time multi-swap local search algorithm. Our approach removes a key assumption from prior work and achieves better solutions both theoretically and in practice. Extensive experiments demonstrate its effectiveness. A promising direction for future work is to design linear-time algorithms that further reduce fairness violations, bringing practical performance closer to theoretical optima.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62432016, 62172446) and the Central South University Research Program of Advanced Interdisciplinary Studies(No.20230YJC023). This work was also carried out in part using computing resources at the High Performance Computing Center of Central South University.

## References

- Aggarwal, A.; Deshpande, A.; and Kannan, R. 2009. Adaptive sampling for  $k$ -means clustering. In *International Workshop on Approximation Algorithms for Combinatorial Optimization*, 15–28. Springer.
- Arthur, D.; and Vassilvitskii, S. 2006.  $k$ -means++: The advantages of careful seeding. Technical report, Stanford.
- Bahmani, B.; Moseley, B.; Vattani, A.; Kumar, R.; and Vassilvitskii, S. 2012. Scalable  $k$ -means++. *arXiv preprint arXiv:1203.6402*.
- Batani, M.; Cohen-Addad, V.; Epasto, A.; and Lattanzi, S. 2024. A Scalable Algorithm for Individually Fair  $k$ -means Clustering. In *International Conference on Artificial Intelligence and Statistics*, 3151–3159. PMLR.
- Beretta, L.; Cohen-Addad, V.; Lattanzi, S.; and Parotsidis, N. 2023. Multi-swap  $k$ -means++. *Advances in Neural Information Processing Systems*, 36: 26069–26091.
- Cohen-Addad, V.; Green Larsen, K.; Saulpic, D.; Schwiegelshohn, C.; and Sheikh-Omar, O. A. 2022. Improved Coresets for Euclidean  $k$ -Means. *Advances in Neural Information Processing Systems*, 35: 2679–2694.
- Cohen-Addad, V.; Mirrokni, V.; and Zhong, P. 2022. Massively Parallel  $k$ -Means Clustering for Perturbation Resilient Instances. In *International Conference on Machine Learning*, 4180–4201. PMLR.
- Huang, J.; Feng, Q.; Huang, Z.; Xu, J.; and Wang, J. 2023. Fast algorithms for distributed  $k$ -clustering with outliers. In *International Conference on Machine Learning*, 13845–13868. PMLR.
- Huang, J.; Feng, Q.; Huang, Z.; Xu, J.; and Wang, J. 2024. Linear time algorithms for  $k$ -means with multi-swap local search. *Advances in Neural Information Processing Systems*, 36.
- Jégou, H.; Tavenard, R.; Douze, M.; and Amsaleg, L. 2011. Searching in one billion vectors: re-rank with source coding. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 861–864. IEEE.
- Jung, C.; Kannan, S.; and Lutz, N. 2020. Service in your neighborhood: Fairness in center location. *Foundations of Responsible Computing (FORC)*.
- Lattanzi, S.; and Sohler, C. 2019. A better  $k$ -means++ algorithm via local search. In *International Conference on Machine Learning*, 3662–3671. PMLR.
- Mahabadi, S.; and Vakilian, A. 2020. Individual fairness for  $k$ -clustering. In *International conference on machine learning*, 6586–6596. PMLR.
- Negahbani, M.; and Chakrabarty, D. 2021. Better Algorithms for Individually Fair  $k$ -Clustering. *Advances in Neural Information Processing Systems*, 34: 13340–13351.
- Nguyen, H. L.; Nguyen, T.; and Jones, M. 2022. Fair range  $k$ -center. *arXiv preprint arXiv:2207.11337*.
- Vakilian, A.; and Yalciner, M. 2022. Improved approximation algorithms for individually fair clustering. In *Proceeding of the 25th International Conference on Artificial Intelligence and Statistics*, 8758–8779.