

---

# A Subquadratic Time Approximation Algorithm for Individually Fair $k$ -Center

---

Matthijs Ebbens

Nicole Funk

Jan Höckendorff  
University of Cologne

Christian Sohler

Vera Weil

## Abstract

We study the  $k$ -center problem in the context of individual fairness. Let  $P$  be a set of  $n$  points in a metric space and  $r_x$  be the distance between  $x \in P$  and its  $\lceil n/k \rceil$ -th nearest neighbor. The problem asks to optimize the  $k$ -center objective under the constraint that, for every point  $x$ , there is a center within distance  $r_x$ . We give bicriteria  $(\beta, \gamma)$ -approximation algorithms that compute clusterings such that every point  $x \in P$  has a center within distance  $\beta r_x$  and the clustering cost is at most  $\gamma$  times the optimal cost. Our main contributions are a deterministic  $O(n^2 + kn \log n)$  time  $(2, 2)$ -approximation algorithm and a randomized  $O(nk \log(n/\delta) + k^2/\varepsilon)$  time  $(10, 2 + \varepsilon)$ -approximation algorithm, where  $\delta$  denotes the failure probability. For the latter, we develop a randomized sampling procedure to compute constant factor approximations for the values  $r_x$  for all  $x \in P$  in subquadratic time; we believe this procedure to be of independent interest within the context of individual fairness.

## 1 INTRODUCTION

Clustering is the process of partitioning a given data set into subsets such that, ideally, the elements in each subset are similar to each other and elements in different subsets are less similar. It is a basic approach in unsupervised learning with applications in, for example, signal processing [Lloyd, 1982], community detection in networks [Zhang et al., 2007][Bota et al., 2015], outlier detection [Chawla and Gionis, 2013] and data summarization [Kleindessner et al., 2019].

For many clustering problems each cluster has a corresponding center that can be viewed as a representative of all points contained in the given cluster. Data points closer to the center are naturally better represented than those that are far away. Ideally, we would like every data point to be close to its representative center, but this is usually impossible to realize. One way to mitigate this is to use clustering objectives such as  $k$ -median or  $k$ -means clustering that optimize the average (squared) distance to the cluster center. But optimizing the average distance will often lead to some data points being less well represented than they could be. In applications where data points model humans, such a behaviour may be undesirable and raises the question how to obtain a clustering that is acceptable for everyone. This problem is addressed in a recent paper by Jung et al. [Jung et al., 2020] that introduced the notion of individual fairness in the context of clustering.

Individual fairness can be thought of as a clustering constraint that ensures that every data point is assigned to a center that is not much worse than what one could hope for in a certain average-type setting: If one partitions a data set with  $n$  points in  $k$  clusters then on average a cluster has  $n/k$  data points [Jung et al., 2020]. Thus, individual fairness is defined to be the constraint that every data point has a center among its  $\lceil n/k \rceil$ -th nearest neighbors.

Individually fair clustering was mainly studied for  $k$ -median and  $k$ -means clustering in metric spaces and generalizations to  $l_p$ -norm type distance functions [Mahabadi and Vakilian, 2020] [Negahbani and Chakrabarty, 2021] [Vakilian and Yalçiner, 2022] [Bateni et al., 2024]. This also led to results on the individually fair  $k$ -center problem, which we will call the  $\alpha$ -fair  $k$ -center problem.

In this paper, we show that the properties of the  $\alpha$ -fair  $k$ -center problem can be exploited to achieve a simple algorithm that improves upon the current state of the art. We then give another algorithm that achieves a randomized constant bicriteria approxima-

tion in  $O(kn \log(n/\delta))$  time where  $\delta > 0$  is a parameter that bounds the failure probability. To do so, we develop a procedure to approximate the distances to the  $\lceil n/k \rceil$ -th nearest neighbors of each data point in that running time. We believe that this procedure may find further applications in the context of individual fairness.

## 2 RELATED WORK

For the  $k$ -center problem there is a  $O(nk)$  time 2-approximation by Gonzalez [Gonzalez, 1985] and it is known that the problem is NP-complete to approximate with a factor less than 2 [Hochbaum and Shmoys, 1985].

The notion of individual fairness we consider in this paper was introduced by Jung et al. [Jung et al., 2020] for  $k$ -clustering in general, but also studied previously in other contexts such as priority clustering [Plesník, 1987] and metric embedding [Chan et al., 2006][Charikar et al., 2010]. In fact, individually fair clustering is a special case of priority clustering [Bajpai et al., 2021]. In the context of individual fairness, a clustering of size  $k$  is  $\alpha$ -fair for a set of points  $P$  and some  $\alpha > 0$  if for each point  $p$  in  $P$  there is a center within  $\alpha$  times the distance to its  $\lceil |P|/k \rceil$ -th nearest neighbor in  $P$ . Jung et al. [Jung et al., 2020] showed that solving  $k$ -clustering under individual fairness with  $\alpha = 1$  is NP-hard in any metric space. They additionally designed an algorithm that computes a  $k$ -center clustering that is 2-fair but lacks any guarantee on the cost. In the context of individually fair  $k$ -clustering, previous work studied algorithms for arbitrary  $\ell_p$  norm cost functions including  $k$ -center,  $k$ -median and  $k$ -means as special cases [Mahabadi and Vakilian, 2020] [Negahbani and Chakrabarty, 2021] [Vakilian and Yalçiner, 2022] [Bateni et al., 2024]. We start by stating the results for the case of  $k$ -center. The first result with guarantees on both cost and fairness was by Mahabadi and Vakilian [Mahabadi and Vakilian, 2020], who developed a  $\tilde{O}(k^5 n^5)$ -time local search ( $O(\log n)$ , 7)-approximation algorithm that computes a  $7\alpha$ -fair solution with cost at most  $O(\log n)$  times the optimal cost, for the case  $\alpha \geq 1$ . Further work was made by Negahbani and Chakrabarty [Negahbani and Chakrabarty, 2021] utilizing LP rounding to develop a  $\tilde{O}(kn^4)$ -time  $(2+\varepsilon, 8)$ -approximation for  $\alpha = 1$  and  $\varepsilon > 0$ . The most recent result with respect to bicriteria approximation is by Vakilian and Yalçiner [Vakilian and Yalçiner, 2022], who proposed a  $\tilde{O}(n^2)$ -time  $(3+\varepsilon, 3)$ -approximation for any  $\alpha \geq 1$  and  $\varepsilon > 0$ . Han et al. [Han et al., 2023] consider a relaxed version of individual fair  $k$ -center by allowing outliers and propose a 4-

approximation for this variant. Next we discuss the results for other  $\ell_p$ -norms. The algorithm of Mahabadi and Vakilian [Mahabadi and Vakilian, 2020] achieves a  $(84, 7)$ -approximation algorithm for  $\alpha$ -fair  $k$ -median. Negahbani and Chakrabarty [Negahbani and Chakrabarty, 2021] designed a  $(2^{p+2}, 8)$ -approximation, which yields cost approximations 8 for  $k$ -median ( $p = 1$ ) and 16 for  $k$ -means ( $p = 2$ ) objective functions. Further results by Vakilian and Yalçiner [Vakilian and Yalçiner, 2022] are a polynomial time  $(16^p + \varepsilon, 3)$ -approximation for general  $p$  and a  $(7.081 + \varepsilon, 3)$ -approximation for  $k$ -median. Finally, for  $\alpha = 1$ , Bateni et al. [Bateni et al., 2024] proposed an  $\tilde{O}(nk^2)$  time  $(O(1), 6)$ -approximation algorithm for  $k$ -median and  $k$ -means. Additional studies include coresets for individually fair  $k$ -clustering by Chhaya et al. [Chhaya et al., 2022] and alternative notions to individual fairness like a feature based approach by Kar et al. [Kar et al., 2021] and an approach through similarity sets by Chakrabarti et al. [Chakrabarti et al., 2022].

## 3 OUR CONTRIBUTIONS

In this work, we give a  $O(n^2 + kn \log n)$  time  $(2, 2)$ -bicriteria approximation algorithm for the individually  $\alpha$ -fair metric  $k$ -center problem, improving upon the current state of the art  $(3, 3)$ -approximation algorithm by Vakilian and Yalçiner [Vakilian and Yalçiner, 2022].

**Result 1** (Simplified restatement of Theorem 8). *Algorithm  $(2, 2)$ -ApproxFairCenter is a  $(2, 2)$ -bicriteria approximation algorithm for the individually  $\alpha$ -fair (discrete) metric  $k$ -center problem with  $O(n^2 + kn \log n)$  running time, where  $n$  is the number of input points.*

As the running time of this algorithm is dominated by the computation of the  $\lceil n/k \rceil$ -th nearest neighbors of each data point, we develop a randomized approximation algorithm for these distances under the assumption that  $k$  is smaller than  $n/6$  (if this is not the case we can compute all distances to the  $\lceil n/k \rceil$ -th nearest neighbors exactly in  $O(nk)$  time).

**Result 2** (Simplified restatement of Lemma 11). *Let  $(P, d)$  be a finite metric space. Algorithm ApproxFairRadii computes with probability at least  $1 - \delta$  for all  $p \in P$  a 5-approximation to the distance of the  $\lceil n/k \rceil$ -th nearest neighbor. The running time of the algorithm is  $O(nk \log(n/\delta))$ .*

Using the approximated distances, we modify the  $(2, 2)$ -approximation algorithm to obtain a randomized approximation algorithm with subquadratic running time if  $k$  is asymptotically smaller than  $n$ .

**Result 3** (Simplified restatement of Theorem 13). *Let  $\delta, \varepsilon > 0$ . Algorithm  $(10, 2 + \varepsilon)$ -ApproxFairCenter is a randomized algorithm that computes with probability at least  $1 - \delta$  a  $(10, 2 + \varepsilon)$ -bicriteria approximation for the individually  $\alpha$ -fair (discrete) metric  $k$ -center problem. The running time of the algorithm is  $O(nk(\log(n/\delta)) + k^2/\varepsilon)$ , where  $n$  is the number of input points.*

## 4 PRELIMINARIES

In this section, we will establish the necessary notation and introduce the  $k$ -center problem with individual fairness. Throughout this work  $(P, d)$  will denote an arbitrary discrete metric space with  $n = |P|$ .  $P$  will be given as input and we assume that one can compute  $d(x, y)$  in constant time for any  $x, y \in P$ . For a set  $C \subseteq P$  and some point  $p \in P$  we write  $d(p, C) := \min_{c \in C} d(p, c)$  and set  $d(p, \emptyset) := \infty$ . We denote the cost of a set  $C \subseteq P$  by  $\text{cost}(P, C) := \max_{p \in P} d(p, C)$ . The definition of the unconstrained  $k$ -center problem is then as follows.

**Definition 1** (discrete metric  $k$ -center problem). *Given a discrete metric space  $(P, d)$  and a target number of clusters  $k \in \mathbb{N}$ , find a set  $C \subseteq P$  of size  $k$  such that  $\text{cost}(P, C)$  is minimized.*

The individual fairness constraint is modelled in the following way [Jung et al., 2020].

**Definition 2.** (fairness radius) *Let  $k \in \mathbb{N}$ ,  $p \in P$  and let  $B(p, r) := \{q \in P \mid d(p, q) \leq r\}$  be the ball centered at  $p$  with radius  $r \in \mathbb{R}_{\geq 0}$ . The  $k$ -fair radius of  $p$  is defined as*

$$r_k(p) := \inf\{r \in \mathbb{R}_{\geq 0} \mid |B(p, r)| \geq n/k\}.$$

**Definition 3.** (individually fair centers) *For  $\alpha \in \mathbb{R}_{>0}$  and  $k \in \mathbb{N}$ , a set  $C \subseteq P$  of centers is called  $(\alpha, k)$ -fair, if  $d(p, C) \leq \alpha r_k(p)$  for all  $p \in P$ . If in addition  $|C| \leq k$ , then  $C$  is called a feasible solution for the  $\alpha$ -fair  $k$ -center problem.*

When the parameters  $\alpha$  and  $k$  are clear from the context we call a feasible solution to the  $\alpha$ -fair  $k$ -center problem also simply a *feasible solution*.

**Definition 4** ( $\alpha$ -fair  $k$ -center problem). *Given a discrete metric space  $(P, d)$ , a target number of clusters  $k \in \mathbb{N}$ , and a fairness parameter  $\alpha \in \mathbb{R}_{>0}$ , find an  $(\alpha, k)$ -fair set  $C \subseteq P$  of size  $k$  that minimizes  $\text{cost}(P, C)$ .*

Note that there does not necessarily exist a feasible solution for all choices of  $\alpha > 0$  in the  $\alpha$ -fair  $k$ -center problem [Jung et al., 2020]. In this case, we define the cost of an optimal solution as infinity. However, it is known that for  $\alpha \geq 2$  there is always a feasible solution

[Jung et al., 2020]. We will consider the following bicriteria approximation.

**Definition 5** ( $(\beta, \gamma)$ -approximation for  $\alpha$ -fair  $k$ -center). *Given  $k \in \mathbb{N}$ ,  $\alpha \in \mathbb{R}_{>0}$  and let  $\Delta^*$  be the cost of an optimal solution for the  $\alpha$ -fair  $k$ -center problem. A set  $C \subseteq P$  of size  $k$  is a  $(\beta, \gamma)$ -approximation for the  $\alpha$ -fair  $k$ -center problem, if the following two conditions hold:*

1.  $\text{cost}(P, C) \leq \beta \Delta^*$ ,
2.  $C$  is  $(\gamma \alpha, k)$ -fair.

Note that, since  $\alpha$  is part of the input, one can reduce the  $k$ -center problem to the  $\alpha$ -fair  $k$ -center problem, by choosing a suitable  $\alpha$  for given input points s.t. any set of  $k$  centers is  $(\alpha, k)$ -fair. By the hardness results on individual fairness [Jung et al., 2020] and approximability of  $k$ -center [Hochbaum and Shmoys, 1985], a  $(2, 2)$ -approximation algorithm for  $\alpha$ -fair  $k$ -center is the best one can hope for in polynomial time unless P equals NP.

## 5 (2, 2)-APPROXIMATION ALGORITHM

In this section, we present and analyze our deterministic bicriteria approximation algorithm for the  $\alpha$ -fair  $k$ -center problem. The first algorithm for individually fair clustering by Jung et al. [Jung et al., 2020] computes a  $(2, k)$ -fair solution by considering points  $p \in P$  in order, non-decreasingly sorted by their fairness radii  $r_k(p)$ , and choosing  $p$  as center, if  $B(p, r_k(p))$  does not intersect the ball  $B(q, r_k(q))$ , i.e.  $d(p, q) > r_k(p) + r_k(q)$ , for all previously chosen centers  $q \in P$ .

This approach is equivalent to the algorithm for finding a  $(k/n)$ -density net by Chan et al. [Chan et al., 2006], the metric embedding algorithm of Charikar et al. [Charikar et al., 2010] and the 2-approximation algorithm for priority  $k$ -center by Plesník [Plesník, 1987][Bajpai et al., 2021]. Vakilian and Yalçiner [Vakilian and Yalçiner, 2022] later changed the condition of choosing  $p$  as a center to  $d(p, q)$  having to be at least  $2\alpha r_k(p) \geq 2\alpha(r_k(p) + r_k(q))$  for all previously chosen centers  $q \in P$ .

We adapt this approach by guessing an upper bound  $\Delta$  on the cost of an optimal solution and modify the condition to  $d(p, q) \geq 2 \min\{\alpha r_k(p), \Delta\}$ , as done in Algorithm 1.

Besides the fairness parameter  $\alpha > 0$  and the number of centers  $k$ , Algorithm 1 also receives a cost value  $\Delta \in \mathbb{R}$  and function  $r : P \rightarrow \mathbb{R}_{\geq 0}$  as input. The algorithm is guaranteed to always compute a  $(2\alpha, k)$ -fair set  $C \subseteq P$  with cost at most  $2\Delta$  if  $r = r_k$ . If there

exists a feasible solution for  $\alpha$ -fair  $k$ -center with cost at most  $\Delta$ , then it additionally holds that  $|C| \leq k$ . Assume that we know the optimal value  $\Delta^*$  for the  $\alpha$ -fair  $k$ -center problem, then we can apply Algorithm 1 to obtain a  $(2, 2)$ -approximation. To find the value  $\Delta^*$  we exploit the fact that there always exist two input points  $p, q \in P$  with  $d(p, q) = \Delta^*$ .

---

**Algorithm 1:** FairCenter
 

---

**Input:** Number of centers  $k \in \mathbb{N}$ , fairness parameter  $\alpha \in \mathbb{R}_{>0}$ , cost value  $\Delta \in \mathbb{R}$ , function  $r : P \rightarrow \mathbb{R}_{\geq 0}$

**Output:** Set  $C \subseteq P$

- 1 Let  $P = \{p_1, \dots, p_n\}$  be sorted s.t.  
 $r(p_1) \leq r(p_2) \leq \dots \leq r(p_n)$
  - 2  $C := \emptyset$
  - 3 **for**  $i := 1$  **to**  $n$  **do**
  - 4     **if**  $d(p_i, C) > 2 \min\{\alpha r(p_i), \Delta\}$  **then**
  - 5          $C := C \cup \{p_i\}$
  - 6 **return**  $C$
- 

**Lemma 6.** Let  $k \in \mathbb{N}$ ,  $\alpha \in \mathbb{R}_{>0}$ ,  $\Delta \in \mathbb{R}_{>0}$  and  $r : P \rightarrow \mathbb{R}$  s.t.  $r(p) \geq r_k(p)$  for all  $p \in P$ . If there exists a feasible solution for  $\alpha$ -fair  $k$ -center with cost at most  $\Delta$ , then FairCenter( $k, \alpha, \Delta, r$ ) returns a set  $C$  of size at most  $k$ .

*Proof.* Let  $(P, d)$  be the input instance and assume that there exists a feasible solution  $C^*$  for the  $\alpha$ -fair  $k$ -center problem with cost at most  $\Delta$  and let  $C = \{c_1, \dots, c_\ell\}$  be the set returned by FairCenter( $k, \alpha, \Delta, r$ ). In order to prove the lemma, we will construct a set of  $\ell$  disjoint balls (one for each center) such that every ball contains at least one center from  $C^*$ . Since  $|C^*| \leq k$  this will imply the result.

Assume that the centers  $c_1, \dots, c_\ell$  are ordered increasingly by the time at which they were included in  $C$ . Note that, since  $\alpha > 0$  and  $P$  is sorted non-decreasingly with respect to  $r$ , for  $1 \leq i \leq j \leq \ell$ ,  $\alpha r(c_i) \leq \alpha r(c_j)$  holds. Now define  $r'(c_i) := \min\{\alpha r(c_i), \Delta\}$ . Since  $c_1, \dots, c_\ell$  were included in  $C$  by Line 4 of Algorithm 1, they satisfy

$$d(c_j, c_i) > 2r'(c_j) \geq r'(c_i) + r'(c_j)$$

for  $1 \leq i < j \leq \ell$ , which implies that

$$B(c_i, r'(c_i)) \cap B(c_j, r'(c_j)) = \emptyset.$$

We now observe that  $C^*$  must have at least one center in each of these balls. Indeed, assume that  $B(c_i, r'(c_i)) \cap C^*$  is empty for some  $c_i \in C$ . Then  $d(c_i, C^*) > r'(c_i) = \min\{\alpha r(c_i), \Delta\} \geq \min\{\alpha r_k(c_i), \Delta\}$ . Thus,  $c_i$  either violates the fairness constraint of  $C^*$  or the solution  $C^*$  has cost more than

$\Delta$ , which is a contradiction. Since the balls are disjoint and each ball contains at least one point of  $C^*$ , we conclude that  $|C| \leq k$ .  $\square$

**Lemma 7.** Let  $k \in \mathbb{N}$ ,  $\alpha \in \mathbb{R}_{>0}$  and  $\Delta \in \mathbb{R}_{>0}$ . If there exists a feasible solution for  $\alpha$ -fair  $k$ -center with cost at most  $\Delta$ , then FairCenter( $k, \alpha, \Delta, r_k$ ) returns a  $(2\alpha, k)$ -fair set of at most  $k$  points with cost at most  $2\Delta$ .

*Proof.* Let  $C$  be the set returned by FairCenter( $k, \alpha, \Delta, r_k$ ). By Lemma 6 we know that  $|C| \leq k$ . The if-clause in Line 4 ensures that  $d(p, C) \leq 2 \min\{\alpha r_k(p), \Delta\} \leq 2\Delta$  for all  $p \in P$ , which concludes the proof.  $\square$

---

**Algorithm 2:** (2, 2)-ApproxFairCenter
 

---

**Input:** Number of centers  $k \in \mathbb{N}$ , fairness parameter  $\alpha \in \mathbb{R}_{>0}$

**Output:** Subset of  $P$  of size at most  $k$ , or failure message

- 1 Let  $D$  be set of pairwise distances in  $P$
  - 2 Compute  $r_k(p)$  for  $p \in P$
  - 3 Do a binary search over  $D$  to find smallest  $\Delta$   
s.t.  $|FairCenter(k, \alpha, \Delta, r_k)| \leq k$
  - 4 **if**  $\Delta$  does not exist **then**
  - 5     **return** no feasible solution exists
  - 6 **else**
  - 7     **return** FairCenter( $k, \alpha, \Delta, r_k$ )
- 

**Theorem 8.** Let  $k \in \mathbb{N}$  and  $\alpha \in \mathbb{R}_{>0}$ . If there exists a feasible solution for  $\alpha$ -fair  $k$ -center, Algorithm 2 computes a  $(2, 2)$ -approximation for  $\alpha$ -fair  $k$ -center in time  $O(n^2 + kn \log n)$ .

*Proof.* Observe that the cost of an optimal solution for  $\alpha$ -fair  $k$ -center is defined by the distance of two points in  $P$ . Let  $\Delta^*$  be the cost of an optimal solution, then by Lemma 7 the set  $C$  returned by FairCenter( $k, \alpha, \Delta^*, r_k$ ) is a  $(2, 2)$ -approximation for  $\alpha$ -fair  $k$ -center (note that we can include arbitrary additional points in  $C$  until  $|C| = k$ , if  $|C| < k$ ). To find the distance  $\Delta^*$  that is passed to Algorithm 1, we first compute in  $O(n^2)$  time for all points  $p, q \in P$  their distances  $d(p, q)$ . For each input point we can then compute its  $k$ -fair radii  $r_k(p)$  in  $O(n)$  time (e.g. Theorem 9.3 in [Cormen et al., 2022]). Additionally we sort  $P$  in non-decreasing order with respect to the  $k$ -fair radii, which takes  $O(n \log n)$  time. We apply a binary search on the pairwise distances to find the optimal  $\Delta^*$  and keep track of the subset of  $D$  that can contain the optimal solution. In each recursion step the median can be computed in linear time (e.g. Theorem 9.3 in [Cormen et al., 2022]) without sorting the values a priori. We invoke  $O(\log n)$  queries to

Algorithm 1 by doing so, where we increase the cost value if  $|C| > k$  and decrease otherwise. If the binary search terminates without finding a cost value such that  $|C| \leq k$  then no feasible solution exists. A call of Algorithm 1 takes  $O(nk)$  time, since we computed the  $k$ -fair radii and sorted  $P$  beforehand. The running time of the binary search on  $1 \leq t \leq n^2$  elements is  $T(t) = O(t) + O(nk) + T(t/2) \in O(n^2 + kn \log n)$ . This results in a total running time of  $O(n^2 + kn \log n)$ , which concludes the proof.  $\square$

Notice that all results in this section are under the assumption that a feasible solution exists. We now remark what happens if this is not the case. From Lemma 1 we can derive that if  $\text{FairCenter}(k, \alpha, \Delta, r_k)$  returns a set of size larger than  $k$  then there does not exist a feasible solution for  $\alpha$ -fair  $k$ -center clustering with cost at most  $\Delta$ . On the other hand, it may be the case that there does not exist a feasible solution but a set  $C$  of size at most  $k$  is returned. Since for all  $p \in P$ ,  $d(p, C) \leq 2 \min\{\alpha r_k(p), \Delta\}$ ,  $C$  is then a feasible solution for  $2\alpha$ -fair  $k$ -center with cost at most  $2\Delta$ .

## 6 FAST $(10, 2 + \varepsilon)$ -APPROXIMATION

Next, we give a subquadratic time  $(10, 2 + \varepsilon)$ -approximation algorithm. The running time bottleneck of the procedure discussed in Theorem 8 is the computation of the candidate cost values and the fair radii, both of which take  $\Theta(n^2)$  time. To overcome this, we show in Section 6.1 that a small set containing a good approximation of the optimal cost can be obtained using Gonzalez's algorithm [Gonzalez, 1985] and we approximate the fair radii using uniform sampling in Section 6.2.

Since the randomized algorithm presented in this section has a running time in  $\tilde{O}(kn)$ , we get subquadratic time under the natural assumption that  $k$  is asymptotically smaller than  $n$ . Furthermore, in this section we assume the pairwise distances between points in  $P$  to be distinct, which can be achieved by an arbitrarily small perturbation of the input points.

### 6.1 COST VALUE CANDIDATES

We start by discussing how to efficiently compute a small set of candidate values that contains a good approximation for the optimal cost of the given  $\alpha$ -fair  $k$ -center instance. To do so, we compute  $k$  centers that are a constant approximation for the unconstrained  $k$ -center problem and show that the optimal cost of  $\alpha$ -fair  $k$ -center either relates to the cost of unconstrained  $k$ -center, or to the distance between two of our computed centers. We use the algorithm by Gonzalez for

this procedure [Gonzalez, 1985]. This algorithm finds a 2-approximation for the optimal  $k$ -center cost by starting with an arbitrary point as first center and then greedily selecting the point as new center, which is furthest away from all previously chosen centers.

Algorithm 3 describes our procedure to find a list of cost value candidates.

---

#### Algorithm 3: CostCandidates

---

**Input:** Number of centers  $k \in \mathbb{N}$ , precision parameter  $\varepsilon \in \mathbb{R}_{>0}$

**Output:** Set  $L \subset \mathbb{R}$

- 1 Let  $C_G$  be the set of points returned by Gonzalez's algorithm
  - 2  $L := \{\frac{1}{2}(1 + \varepsilon)^j \max_{p \in P} d(p, C_G) \mid j \in \{1, \dots, \lceil \log_{1+\varepsilon}(16) \rceil\}\}$
  - 3 **for**  $(g_1, g_2) \in C_G^2$  with  $g_1 \neq g_2$  **do**
  - 4      $L := L \cup \{\frac{1}{2}(1 + \varepsilon)^j d(g_1, g_2) \mid j \in \{1, \dots, \lceil \log_{1+\varepsilon}(3) \rceil\}\}$
  - 5 **return**  $L$
- 

**Lemma 9.** *Let  $k \in \mathbb{N}$  and  $\alpha, \varepsilon \in \mathbb{R}_{>0}$ . Assume that there exists a solution for the  $\alpha$ -fair  $k$ -center problem and let  $\Delta_\alpha^*$  be the cost of an optimal solution.  $\text{CostCandidates}(k, \varepsilon)$  computes in  $O(kn + k^2/\varepsilon)$  time a set  $L$  of size  $O(k^2/\varepsilon)$  such that there exists  $\ell \in L$  with  $\Delta_\alpha^* \leq \ell \leq (1 + \varepsilon)\Delta_\alpha^*$ .*

*Proof.* Let  $\text{OPT}$  be an optimal set of centers for  $k$ -center (without fairness constraints) and  $\Delta^* := \max_{p \in P} d(p, \text{OPT})$ . Let  $C_G$  be the set of centers found by Gonzalez's algorithm and  $\Delta := \max_{p \in P} d(p, C_G)$ . It holds that  $\Delta^* \leq \Delta \leq 2\Delta^*$  and that  $C_G$  can be computed in  $O(kn)$  time as shown by Gonzalez [Gonzalez, 1985].

For each pair of distinct centers  $g_1, g_2 \in C_G$ , let

$$L(g_1, g_2) := \{\frac{1}{2}(1 + \varepsilon)^j d(g_1, g_2) \mid j \in \{1, \dots, \lceil \log_{1+\varepsilon}(3) \rceil\}\}.$$

Observe that the union of the intervals  $[(1 + \varepsilon)^{-1}\ell, \ell]$  over all  $\ell \in L(g_1, g_2)$  completely covers the interval  $[\frac{1}{2}d(g_1, g_2), \frac{3}{2}d(g_1, g_2)]$ .

Furthermore, let

$$L_0 := \{\frac{1}{2}(1 + \varepsilon)^j \Delta \mid j \in \{1, \dots, \lceil \log_{1+\varepsilon}(16) \rceil\}\}.$$

Similarly, the union of the intervals  $[(1 + \varepsilon)^{-1}\ell, \ell]$  over all  $\ell \in L_0$  completely covers the interval  $[\frac{1}{2}\Delta, 8\Delta]$ . Finally, let

$$L := L_0 \cup \bigcup_{g_1, g_2 \in C_G, g_1 \neq g_2} L(g_1, g_2)$$

as it is constructed in Algorithm 3.

We will show that  $L$  has the desired properties. The bounds on running time and size are clear from the construction. To prove that there exists  $\ell \in L$  satisfying  $\Delta_\alpha^* \leq \ell \leq (1 + \varepsilon)\Delta_\alpha^*$ , we consider two cases.

First, assume that

$$\Delta^* \leq \Delta_\alpha^* \leq 8\Delta^*.$$

Since

$$\frac{1}{2}\Delta \leq \Delta^* \leq \Delta$$

it follows that

$$\frac{1}{2}\Delta \leq \Delta_\alpha^* \leq 8\Delta.$$

Then there exists  $j \in \{1, \dots, \lceil \log_{1+\varepsilon}(16) \rceil\}$  such that

$$\frac{1}{2}(1 + \varepsilon)^{j-1}\Delta \leq \Delta_\alpha^* \leq \frac{1}{2}(1 + \varepsilon)^j\Delta,$$

thus,

$$\Delta_\alpha^* \leq \frac{1}{2}(1 + \varepsilon)^j\Delta \leq (1 + \varepsilon)\Delta_\alpha^*.$$

Since  $\frac{1}{2}(1 + \varepsilon)^j\Delta \in L$ , this proves the claim.

Second, assume that  $\Delta_\alpha^* > 8\Delta^*$ . Consider  $p \in P$  and  $c_\alpha \in \text{OPT}_\alpha$  that realize  $\Delta_\alpha^*$ , i.e.,  $d(p, c_\alpha) = \Delta_\alpha^*$ . Furthermore, let  $g(p)$  and  $g(c_\alpha)$  be the centers in  $C_G$  closest to  $p$  and  $c_\alpha$ , respectively. We have  $d(p, g(p)) \leq 2\Delta^*$  and  $d(c_\alpha, g(c_\alpha)) \leq 2\Delta^*$ , so by the triangle inequality

$$\begin{aligned} d(g(p), g(c_\alpha)) &\leq d(g(p), p) + d(p, c_\alpha) + d(c_\alpha, g(c_\alpha)) \\ &\leq \Delta_\alpha^* + 4\Delta^* \\ &\leq \frac{3}{2}\Delta_\alpha^* \end{aligned}$$

where the last inequality follows from  $\Delta_\alpha^* > 8\Delta^*$ . On the other hand,

$$\begin{aligned} d(g(p), g(c_\alpha)) &\geq d(p, c_\alpha) - d(p, g(p)) - d(c_\alpha, g(c_\alpha)), \\ &\geq \Delta_\alpha^* - 4\Delta^*, \\ &\geq \frac{1}{2}\Delta_\alpha^*, \end{aligned}$$

where the last inequality again follows from  $\Delta_\alpha^* > 8\Delta^*$ . Hence,

$$\frac{1}{2}d(g(p), g(c_\alpha)) \leq \Delta_\alpha^* \leq \frac{3}{2}d(g(p), g(c_\alpha)).$$

Then there exists  $j \in \{1, \dots, \lceil \log_{1+\varepsilon}(3) \rceil\}$  such that

$$\frac{1}{2}(1 + \varepsilon)^{j-1}d(g(p), g(c_\alpha)) \leq \Delta_\alpha^* \leq \frac{1}{2}(1 + \varepsilon)^jd(g(p), g(c_\alpha)),$$

thus,

$$\Delta_\alpha^* \leq \frac{1}{2}(1 + \varepsilon)^jd(g(p), g(c_\alpha)) \leq (1 + \varepsilon)\Delta_\alpha^*.$$

Since

$$\frac{1}{2}(1 + \varepsilon)^jd(g(p), g(c_\alpha)) \in L(g(p), g(c_\alpha)) \subseteq L,$$

this proves the claim in the second case, which concludes the proof.  $\square$

## 6.2 APPROXIMATING FAIR RADII

Next we discuss how to efficiently compute good approximations for the  $k$ -fair radii of the points in  $P$ . The procedure first samples a small subset of the input  $S$  such that with constant probability it holds that for every point  $p \in P$  there exists a point  $q \in S$  with  $d(p, q) \in (r_{3k}(p), r_k(p)]$ . The pseudocode of this subprocedure is stated in Algorithm 4. Note that a similar result was proven by Czumaj and Sohler in the context of approximating the cost of  $k$ -nearest neighbor graphs (See lemma 3.1 in [Czumaj and Sohler, 2024]).

---

### Algorithm 4: FairSampling

---

**Input:** Number of centers  $k \in \mathbb{N}$  with  $k \leq n/6$ , failure probability  $\delta \in (0, 1)$

**Output:** Function  $r' : P \rightarrow \mathbb{R}$

```

1  $s := 36k \lceil \ln(2n/\delta) \rceil$ 
2  $t := 27 \lceil \ln(2n/\delta) \rceil$ 
3 Let  $S$  be a subset of size  $s$  drawn uniformly at
   random with replacement from  $P$ 
4 for  $p \in P$  do
5   | Let  $x$  be  $t$ -nearest neighbor of  $p$  in  $S$ 
6   |  $r'(p) := d(p, x)$ 
7 return  $r'$ 
```

---

**Lemma 10.** *Given  $k \in \mathbb{N}$  such that  $k \leq n/6$  and  $\delta \in (0, 1)$ . Algorithm 4 returns in time  $O(nk \log(n/\delta))$  with probability at least  $1 - \delta$  for all  $p \in P$  a value  $r'(p) \in (r_{3k}(p), r_k(p)]$ .*

*Proof.* Let  $S$  be a sample taken uniformly at random (with replacement) from  $P$  of size  $s = 36k \lceil \ln(2n/\delta) \rceil$  as in Algorithm 4. We start by showing that  $r'(p) \in (r_{3k}(p), r_k(p)]$  for all  $p \in P$  with probability at least  $1 - \delta$ . For  $p \in P$  and  $q \in S$  let  $X_{pq}^k$  be the indicator random variable that is 1 if the sample  $q$  is in  $B(p, r_k(p))$  and 0 otherwise. Furthermore, let  $X_p^k = \sum_{q \in S} X_{pq}^k$  be the random variable indicating the number of points of  $S$  in  $B(p, r_k(p))$ . We define  $X_{pq}^{3k}$  and  $X_p^{3k}$  similarly by replacing  $r_k(p)$  by  $r_{3k}(p)$ .

Observe that  $r'(p) \in (r_{3k}(p), r_k(p)]$  if and only if  $X_p^{3k} < t$  and  $X_p^k \geq t$ , since  $r'(p)$  is defined as the distance from  $p$  to its  $t$ -nearest neighbor. Therefore,  $r'(p) \in (r_{3k}(p), r_k(p)]$  does not hold only if  $X_p^{3k} \geq t$  or  $X_p^k < t$ .

First, we give an upper bound for  $\Pr(X_p^{3k} \geq t)$ . By definition,  $|B(p, r_{3k}(p))| = \lceil \frac{n}{3k} \rceil$ , so

$$\frac{n}{3k} \leq |B(p, r_{3k}(p))| \leq \frac{n}{3k} + 1 \leq \frac{n}{2k},$$

since  $k \leq n/6$ . Each  $q \in S$  is sampled uniformly at random from  $P$ , so  $\frac{1}{3k} \leq \Pr(X_{pq}^{3k} = 1) \leq \frac{1}{2k}$ . By

linearity of expectation,  $E[X_p^{3k}] = sPr(X_{pq}^{3k} = 1)$ , so  $12\lceil \ln(2n/\delta) \rceil \leq E[X_p^{3k}] \leq 18\lceil \ln(2n/\delta) \rceil$ . Then

$$\begin{aligned} Pr(X_p^{3k} \geq t) &= Pr(X_p^{3k} \geq 27\lceil \ln(2n/\delta) \rceil), \\ &\leq Pr(X_p^{3k} \geq \tfrac{3}{2}E[X_p^{3k}]), \end{aligned}$$

from which it follows by a Chernoff bound (e.g. Theorem 4.4 in [Mitzenmacher and Upfal, 2017]) that

$$Pr(X_p^{3k} \geq t) \leq e^{-E[X_p^{3k}]/12}.$$

Since  $E[X_p^{3k}] \geq 12\lceil \ln(2n/\delta) \rceil$ , we conclude that

$$Pr(X_p^{3k} \geq t) \leq e^{-\lceil \ln(2n/\delta) \rceil} \leq \frac{\delta}{2n}.$$

Second, we give an upper bound for  $Pr(X_p^k < t)$ . By a similar reasoning as before, we see that  $\frac{1}{k} \leq Pr(X_{pq}^k = 1) \leq \frac{3}{2k}$  and  $36\lceil \ln(2n/\delta) \rceil \leq E[X_p^k] \leq 54\lceil \ln(2n/\delta) \rceil$ . Then

$$\begin{aligned} Pr(X_p^k < t) &= Pr(X_p^k < 27\lceil \ln(2n/\delta) \rceil), \\ &\leq Pr(X_p^k < \tfrac{3}{4}E[X_p^k]), \end{aligned}$$

so it follows by a Chernoff bound (e.g. Theorem 4.5 in [Mitzenmacher and Upfal, 2017]) that

$$Pr(X_p^k < t) \leq e^{-E[X_p^k]/32}.$$

Since  $E[X_p^k] \geq 36\lceil \ln(2n/\delta) \rceil$ , we conclude that

$$Pr(X_p^k < t) \leq e^{-\lceil \ln(2n/\delta) \rceil} \leq \frac{\delta}{2n}.$$

Combining both upper bounds, we see that  $Pr(r'(p) \notin (r_{3k}(p), r_k(p))) \leq \delta/n$ . It follows from a union bound over all  $p \in P$  that the probability that  $r'(p) \notin (r_{3k}(p), r_k(p))$  for at least one  $p \in P$  is at most  $\delta$ , which proves the desired claim.

As for the running time, drawing the sample set  $S$  takes  $O(k \log(n/\delta))$  time. Finding the  $t$ -nearest neighbor of  $p \in P$  in  $S$  can be done in  $O(k \log(n/\delta))$  time (e.g. Theorem 9.3 in [Cormen et al., 2022]), yielding a total running time of  $O(nk \log(n/\delta))$ .  $\square$

Notice that the radii computed by Algorithm 4 are not necessarily constant factor approximations of the fairness radii. We can however use them to compute 3-approximations of the actual  $k$ -fair radii as described in Algorithm 5. Algorithm 5 considers the input points in non-decreasing order with respect to the corresponding values  $r'$  computed by Algorithm 4. Assuming Algorithm 4 to be successful, we argue that the following yields approximate fairness radii efficiently: The algorithm maintains a set  $C$  of points for

which we compute the fairness radii exactly. Initially  $C$  is empty. Whenever a new point  $p$  is considered, we check whether  $B(p, r'(p))$  intersects any of the balls  $B(q, r'(q))$  defined by the points  $q \in C$ . If this is not the case, we add  $p$  to  $C$ . Otherwise, we set the radius value of  $p$  to  $d(p, q) + r_k(q)$ . Since the balls defined by the  $r'$  values contain sufficiently many points and are disjoint, we can argue that the case of computing the  $k$ -fair radii explicitly does not happen too often.

---

**Algorithm 5:** ApproxFairRadii

---

**Input:** Number of centers  $k \in \mathbb{N}$  with  $k \leq n/6$ , failure probability  $\delta \in (0, 1)$

**Output:** Function  $\tilde{r} : P \rightarrow \mathbb{R}$ , or a failure message

```

1 Let  $r'$  be the function returned by FairSampling( $k, \delta$ )
2 Organize  $P$  in a min-heap  $H$  with respect to  $r'$ 
3  $p := \text{extractMin}(H)$ 
4 Compute  $r_k(p)$  exactly and set  $\tilde{r}(p) := r_k(p)$ 
5  $C := \{p\}$ 
6 while  $H \neq \emptyset$  do
7    $p := \text{extractMin}(H)$ 
8   if  $\exists q \in C : r'(p) + r'(q) \geq d(p, q)$  then
9      $\tilde{r}(p) := d(p, q) + \tilde{r}(q)$ 
10  else
11    Compute  $r_k(p)$  exactly and set  $\tilde{r}(p) := r_k(p)$ 
12     $C := C \cup \{p\}$ 
13    if  $|C| > 3k$  then
14      return Fail
15 return  $\tilde{r}$ 

```

---

**Lemma 11.** *Given  $k \in \mathbb{N}$  with  $k \leq n/6$  and  $\delta \in (0, 1)$ . Then Algorithm 5 computes in  $O(kn \log(n/\delta))$  time with probability at least  $1 - \delta$  for all  $p \in P$  a value  $\tilde{r}(p)$  s.t.  $r_k(p) \leq \tilde{r}(p) \leq 5r_k(p)$ .*

*Proof.* By Lemma 10,  $r'(p) \in (r_{3k}(p), r_k(p))$  for all  $p \in P$  with probability at least  $1 - \delta$ , so assume that this is the case for the proof of correctness. The definition of  $k$ -fair radius implies the following claim.

**Claim 12.** *For all  $p, q \in P$ ,  $r_k(p) \leq d(p, q) + r_k(q)$ .*

Consider an arbitrary iteration of the algorithm and let  $p$  be the point extracted from the min-heap. Note that for all  $q \in C$ ,  $\tilde{r}(q) = r_k(q)$ . Let (i) be the case that there exists a point  $q \in C$  such that  $d(p, q) \leq r'(p) + r'(q)$ . This corresponds to Line 8 of the algorithm. Note that  $r'(q) \leq r'(p)$  for all  $q \in C$ , since these points

were considered before  $p$ . Then

$$\begin{aligned}
\tilde{r}(p) &= d(p, q) + \tilde{r}(q) \\
&= d(p, q) + r_k(q), \\
&\leq 2d(p, q) + r_k(p), \\
&\leq 2(r'(p) + r'(q)) + r_k(p), \\
&\leq 4r'(p) + r_k(p), \\
&\leq 5r_k(p),
\end{aligned}$$

where the first inequality holds by Claim 12. Let (ii) be the other case when there is no such point in  $C$ . This corresponds to Line 10 of the algorithm. Here, we trivially get  $\tilde{r}(p) = r_k(p)$ .

Let us consider the running time. By Lemma 10, the procedure  $\text{FairSampling}(k, \delta)$  takes  $O(nk \log(n/\delta))$  time. Setting up the min-heap takes  $O(n)$  time and extracting a point takes  $O(\log n)$  time. Consider an arbitrary iteration of the algorithm. Observe that the running time of Case (i) is  $O(|C|)$ , since we search  $C$  for an intersection. Case (ii) takes  $O(n \log(n))$  time, since we compute the  $k$ -fair radius exactly. Note that the number of iterations where Case (ii) happens corresponds directly to the size of  $C$ . It therefore remains to bound the size of  $C$ .

Assume that  $\text{FairSampling}(k, \delta)$  is successful, i.e., that  $r'(p) \in (r_{3k}(p), r_k(p)]$  for all  $p \in P$ . Then the ball  $B(p, r'(p))$  contains at least  $n/3k$  points for all  $p \in P$ , which implies that for all  $p \in P$  there are at least  $n/3k$  points  $q$  in  $P$  such that  $r'(p) + r'(q) \geq d(p, q)$ . Since the Balls  $B(q, r'(q))$  of points  $q \in C$  are pairwise disjoint by definition of  $C$ , we get that whenever a point  $p$  is included to  $C$  through Case (ii) there exist at least  $n/3k$  points for which Case (i) happens. This implies that Case (ii) can only happen up to  $3k$  times, so  $|C| \leq 3k$ . In total, we therefore incur a running time of  $O(nk \log(n/\delta))$ .

Now consider the case that the procedure  $\text{FairSampling}$  failed, i.e., there exists  $p \in P$  such that  $r'(p) \notin (r_{3k}(p), r_k(p)]$ . If  $r'(p) \leq r_{3k}(p)$ , we cannot argue that Case (ii) happens at most  $3k$  times during the algorithm. It is therefore checked in Line 13 if  $|C| > 3k$ , which implies that  $\text{FairSampling}$  failed and the algorithm terminates after only executing  $O(k)$  iterations of Case (i). The running time is therefore also  $O(nk \log(n/\delta))$  for the failure case.  $\square$

### 6.3 SUBQUADRATIC TIME APPROXIMATION ALGORITHM

We can now state and prove the main result of this work.

**Theorem 13.** *Given parameters  $k \in \mathbb{N}$ ,  $\alpha, \varepsilon \in \mathbb{R}_{>0}$  and  $\delta \in (0, 1)$ . If there exists a feasible solution for*

---

#### Algorithm 6: $(10, 2 + \varepsilon)$ -ApproxFairCenter

---

**Input:** Number of centers  $k \in \mathbb{N}$ , fairness parameter  $\alpha \in \mathbb{R}_{>0}$ , failure probability  $\delta \in (0, 1)$

**Output:** Subset of  $P$  of size at most  $k$ , or failure message

```

1 if  $k > n/6$  or  $k^2/\varepsilon > n^2 \log n$  then
2   | return  $(2, 2)$ -ApproxFairCenter( $P, k, \alpha$ )
3  $\tilde{r} \leftarrow \text{ApproxFairRadii}(k, \delta)$ 
4 if  $\tilde{r} = \text{Fail}$  then
5   | return  $\text{Fail}$ 
6  $L \leftarrow \text{CostCandidate}(k, \varepsilon/2)$ 
7 Binary search over  $L$  to find smallest  $\Delta$  s.t.
   |  $|\text{FairCenter}(k, \alpha, \Delta, \tilde{r})| \leq k$ 
8 if  $\Delta$  does not exist then
9   | return no feasible solution exists
10 else
11   | return  $\text{FairCenter}(k, \alpha, \Delta, r_k)$ 

```

---

*$\alpha$ -fair  $k$ -center Algorithm 6 computes with probability at least  $1 - \delta$  a  $(10, 2 + \varepsilon)$ -approximation for the  $\alpha$ -fair  $k$ -center problem. The running time of the algorithm is in  $O(kn \log(n/\delta) + k^2/\varepsilon)$  time.*

*Proof.* Assume that  $k \leq n/6$  and  $k^2/\varepsilon \leq n^2 \log n$ , otherwise the proof follows from Theorem 8.

We first execute  $\text{ApproxFairRadii}(k, \delta)$ . By Lemma 11,  $\text{ApproxFairRadii}(k, \delta)$  outputs values  $\tilde{r}(p)$  such that  $r_k(p) \leq \tilde{r}(p) \leq 5r_k(p)$  for all  $p \in P$  with probability at least  $1 - \delta$ .

Next we execute  $\text{CostCandidate}(k, \varepsilon/2)$ . By Lemma 9, the set  $L$  computed by  $\text{CostCandidates}(k, \varepsilon/2)$  contains  $\Delta \in L$  with  $\text{cost}(P, C^*) \leq \Delta \leq (1 + \varepsilon/2) \text{cost}(P, C^*)$ , where  $C^*$  is an optimal solution for  $\alpha$ -fair  $k$ -center.

Let  $C$  be the solution returned by  $\text{FairCenter}(k, \alpha, \Delta', \tilde{r})$  for an arbitrary  $\Delta' \geq \Delta$ . By Lemma 6, we know that in this case  $|C| \leq k$ . Thus, the binary search stops with a value  $\Delta' \leq \Delta$ . The if-clause in Line 4 of Algorithm  $\text{FairCenter}$  ensures that

$$\begin{aligned}
d(p, C) &\leq 2 \min\{\alpha \tilde{r}(p), \Delta'\} \\
&\leq 2 \min\{\alpha 5r_k(p), (1 + \varepsilon/2) \text{cost}(P, C^*)\}
\end{aligned}$$

This implies that  $C$  is a  $(10, 2 + \varepsilon)$ -approximation for  $\alpha$ -fair  $k$ -center.

Finally, we analyze the running time of this procedure. By Lemma 10,  $\text{ApproxFairRadii}(k, \delta)$  takes  $O(kn \log(n/\delta))$  time. By Lemma 9,  $\text{CostCandidates}(k, \varepsilon/2)$  requires  $O(kn + k^2/\varepsilon)$  time. Sorting  $P$  takes  $O(n \log n)$  time. Analogous to



Algorithm 2 one can implement the binary search to have runtime  $O(k^2/\varepsilon + kn \log(k/\varepsilon))$ . Under our assumptions the procedure takes  $O(kn \log(n/\delta) + k^2/\varepsilon)$  time in total.  $\square$

## 7 Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 459420781.

## References

- [Bajpai et al., 2021] Bajpai, T., Chakrabarty, D., Chekuri, C., and Negahbani, M. (2021). Revisiting priority  $k$ -center: Fairness and outliers. In *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 21:1–21:20.
- [Bateni et al., 2024] Bateni, 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*, pages 3151–3159. PMLR.
- [Bota et al., 2015] Bota, A., Kresz, M., and Zavalnij, B. (2015). Adaptations of the  $k$ -means algorithm to community detection in parallel environments. *International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)*, page 299–302.
- [Chakrabarti et al., 2022] Chakrabarti, D., Dickerson, J. P., Esmaeili, S. A., Srinivasan, A., and Tsepenekas, L. (2022). A new notion of individually fair clustering:  $\alpha$ -equitable  $k$ -center. In *25th International conference on artificial intelligence and statistics (AISTATS 2022)*, pages 6387–6408. PMLR.
- [Chan et al., 2006] Chan, T.-H. H., Dinitz, M., and Gupta, A. (2006). Spanners with slack. In *European Symposium on Algorithms (ESA 2006)*, pages 196–207. Springer.
- [Charikar et al., 2010] Charikar, M., Makarychev, K., and Makarychev, Y. (2010). Local global tradeoffs in metric embeddings. *SIAM Journal on Computing*, 39(6):2487–2512.
- [Chawla and Gionis, 2013] Chawla, S. and Gionis, A. (2013).  $k$ -means: A unified approach to clustering and outlier detection. In *Proceedings of the 2013 SIAM International Conference on Data Mining (SDM 2013)*, pages 189–197.
- [Chhaya et al., 2022] Chhaya, R., Dasgupta, A., Choudhari, J., and Shit, S. (2022). On coresets for fair regression and individually fair clustering. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics (AISTATS 2022)*, volume 151 of *Proceedings of Machine Learning Research*, pages 9603–9625.
- [Cormen et al., 2022] Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2022). *Introduction to algorithms*. MIT press.
- [Czumaj and Sohler, 2024] Czumaj, A. and Sohler, C. (2024). Sublinear time approximation of the cost of a metric-nearest neighbor graph. *SIAM Journal on Computing*, 53(2):524–571.
- [Gonzalez, 1985] Gonzalez, T. F. (1985). Clustering to minimize the maximum intercluster distance. *Theoretical computer science*, 38:293–306.
- [Han et al., 2023] Han, L., Xu, D., Xu, Y., and Yang, P. (2023). Approximation algorithms for the individually fair  $k$ -center with outliers. *Journal of Global Optimization*, 87(2):603–618.
- [Hochbaum and Shmoys, 1985] Hochbaum, D. S. and Shmoys, D. B. (1985). A best possible heuristic for the  $k$ -center problem. *Mathematics of operations research*, 10(2):180–184.
- [Jung et al., 2020] Jung, C., Kannan, S., and Lutz, N. (2020). Service in Your Neighborhood: Fairness in Center Location. In *1st Symposium on Foundations of Responsible Computing (FORC 2020)*, volume 156 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 5:1–5:15. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [Kar et al., 2021] Kar, D., Kosan, M., Mandal, D., Medya, S., Silva, A., Dey, P., and Sanyal, S. (2021). Feature-based individual fairness in  $k$ -clustering. *arXiv preprint arXiv:2109.04554*.
- [Kleindessner et al., 2019] Kleindessner, M., Awasthi, P., and Morgenstern, J. (2019). Fair  $k$ -center clustering for data summarization. In *International Conference on Machine Learning (ICML 2019)*, pages 3448–3457. PMLR.
- [Lloyd, 1982] Lloyd, S. (1982). Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137.
- [Mahabadi and Vakilian, 2020] Mahabadi, S. and Vakilian, A. (2020). Individual fairness for  $k$ -clustering. In *International conference on machine learning (ICML 2020)*, pages 6586–6596. PMLR.

- [Mitzenmacher and Upfal, 2017] Mitzenmacher, M. and Upfal, E. (2017). *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press.
- [Negahbani and Chakrabarty, 2021] Negahbani, M. and Chakrabarty, D. (2021). Better algorithms for individually fair  $k$ -clustering. *Advances in Neural Information Processing Systems (NeurIPS 2021)*, 34:13340–13351.
- [Plesník, 1987] Plesník, J. (1987). A heuristic for the  $p$ -center problems in graphs. *Discrete Applied Mathematics*, 17(3):263–268.
- [Vakilian and Yalçiner, 2022] Vakilian, A. and Yalçiner, M. (2022). Improved approximation algorithms for individually fair clustering. In *International conference on artificial intelligence and statistics (AISTATS 2022)*, pages 8758–8779. PMLR.
- [Zhang et al., 2007] Zhang, S., Wang, R.-S., and Zhang, X.-S. (2007). Identification of overlapping community structure in complex networks using fuzzy  $c$ -means clustering. *Physica A: Statistical Mechanics and its Applications*, 374(1):483–490.

## Checklist

1. For all models and algorithms presented, check if you include:
  - (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
  - (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
  - (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No]
2. For any theoretical claim, check if you include:
  - (a) Statements of the full set of assumptions of all theoretical results. [Yes]
  - (b) Complete proofs of all theoretical results. [Yes]
  - (c) Clear explanations of any assumptions. [Yes]
3. For all figures and tables that present empirical results, check if you include:
  - (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable]
  - (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]
  - (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]
  - (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
  - (a) Citations of the creator If your work uses existing assets. [Not Applicable]
  - (b) The license information of the assets, if applicable. [Not Applicable]
  - (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
  - (d) Information about consent from data providers/curators. [Not Applicable]
  - (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]
5. If you used crowdsourcing or conducted research with human subjects, check if you include:
  - (a) The full text of instructions given to participants and screenshots. [Not Applicable]
  - (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
  - (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]