

Exact Ratio Preservation via Outliers for Fair k -Center Clustering

Anna Arutyunova 

Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Germany

Irina Fast 

Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Germany

Annika Hennes  

Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Germany

Carsten Krollmann 

Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Germany

Daniel R. Schmidt  

Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Germany

Melanie Schmidt  

Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Germany

Abstract

We study the k -center clustering problem under demographic fairness constraints, where the point set is partitioned into groups, and the aim is to compute clusters that exhibit a given group proportion. Previous work in this direction assumes that the entire point set already respects the desired proportions or uses relaxed notions of fairness.

In this work, we propose a model that facilitates the creation of clusters that exactly match given target ratios, even when the input point set does not. We combine the well-known fair clustering model initiated by Chierichetti, Kumar, Lattanzi, and Vassilvitskii [10] with the notion of outliers to obtain a practical combinatorial framework that provides constant-factor approximate solutions for all proportion settings from $1 : 1$ for two groups to $t_1 : t_2 : \dots : t_m$ for $m \geq 2$ groups, where $t_1, \dots, t_m$ are integers.

We implement and evaluate our algorithms, compare different variants, and provide evidence of the practicability of this approach.

2012 ACM Subject Classification Theory of computation $\rightarrow$ Facility location and clustering; Theory of computation $\rightarrow$ Approximation algorithms analysis

Keywords and phrases Fairness, k -center, approximation algorithms

Digital Object Identifier 10.4230/LIPIcs.ESA.2026.23

Related Version *Full Version*: <https://arxiv.org/abs/2607.05342>

Supplementary Material *Software (Source Code)*: <https://github.com/algo-hhu/fair-k-center-via-outliers/> [3], archived at [swh:1:dir:c5548b3049a3013bb70e93d8345358258ee89e92](https://www.swh.io/dir/c5548b3049a3013bb70e93d8345358258ee89e92)

Funding Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project 456558332 through the Emmy Noether Programme.

1 Introduction

Fair clustering is a very active line of research in clustering algorithm design, introduced by Chierichetti, Kumar, Lattanzi, and Vassilvitskii [10] in 2017. The key idea is that, when applying k -clustering, the user may specify a protected attribute such as gender, race, or another demographic variable to which the algorithm should pay particular attention. The



© Anna Arutyunova, Irina Fast, Annika Hennes, Carsten Krollmann, Daniel R. Schmidt, and Melanie Schmidt;

licensed under Creative Commons License CC-BY 4.0

34th Annual European Symposium on Algorithms (ESA 2026).

Editors: Philip Bille, Seth Pettie, and Sabine Storandt; Article No. 23; pp. 23:1–23:23



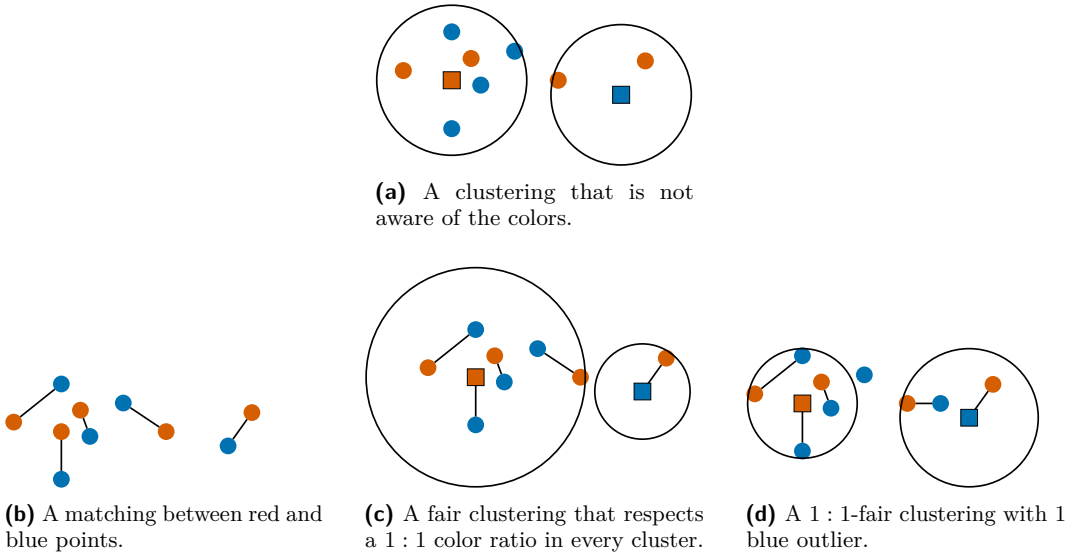
Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

goal is to ensure group-level fairness: each cluster shall contain the same proportion of attribute values (i.e., demographic groups) as observed in the overall dataset. That is, the composition of every cluster shall match the global distribution of the protected attribute.

Enforcing such fairness can be valuable in a variety of real-world settings, as it helps ensure diversity within each group. This is why the model is called *fair*: It makes sure that every group is adequately represented in each cluster. Fairness is a desired or necessary condition in many applications, e.g., representation of people with protected characteristics in committees, creating geographic zones with demographic constraints (e.g., schools [21]), allocation of scarce resources (e.g., access to childcare, or charging times for electric vehicles [24]), or preventing dominance of a single actor in ads [1].

To further explain the model, let us consider a simple base case (the general case is defined in Section 1.2): Given is a set of points P , a metric $d : P \times P \rightarrow \mathbb{R}_{\geq 0}$, the desired number of centers k , and a mapping $\gamma : P \rightarrow \{\text{blue}, \text{red}\}$ that satisfies $|\gamma^{-1}(\text{blue})| = |\gamma^{-1}(\text{red})|$, i.e., the input contains an equal number of red and blue points. Then the fair k -center problem is to partition P into k clusters where each of the clusters $C_1, \dots, C_k$ also contains an equal number of red and blue points, i.e., $|C_i \cap \gamma^{-1}(\text{blue})| = |C_i \cap \gamma^{-1}(\text{red})|$ for all $i \in [k]$. Quite intuitively, this problem can be solved via a combination of matching and unconstrained clustering: First, pair each blue point with one red point such that the maximum distance between a point and its partner is minimized (see Figure 1b). Second, compute an unconstrained k -center clustering on P and assign each pair (r, b) to the center c that minimizes $\max\{d(r, c), d(b, c)\}$ (see Figure 1c). This approach indeed yields a constant factor approximation, and it can be refined to achieve a 3-approximation for the fair k -center problem for this special two-color case with color classes of equal cardinality.



■ **Figure 1** Visualizations of fair decomposition and fair clustering without/with one outlier.

The first step is a special case of a so-called *fairlet decomposition* which decomposes P into micro clusters called fairlets which 1) exactly match a given ratio and 2) are inclusion-wise minimal under this property, i.e., no subset satisfies the ratio. When applied to multiple colors and arbitrary ratios, the combinatorial structure is more involved, but computing fairlet decompositions that exactly meet given ratios remains a combinatorially nice problem. For example [22] describe how to compute fairlet decompositions in general.

The main hurdle now is that in real-world instances, enforcing exact proportional representation is often infeasible or undesirable. Imagine that a point set P contains 999 red points and 1001 blue points. Since these numbers are coprime, the only subset of P that matches the input ratio is P itself, i. e., the only existing fairlet decomposition consists of only one fairlet containing P completely.

There are different ways out of this scenario. The path pursued so far is to relax the notion of fairness and require that the input ratio is only nearly matched, e. g., by allowing an ϵ -deviation. This idea is studied in [5] and [6] for multicolor inputs and arbitrary input ratios. The resulting algorithms achieve bicriteria guarantees, allowing a small additive violation in the fairness constraints, and are based on LP rounding. Combinatorial algorithms for these interval-based models are not known. The reliance on LP methods can be expected for this variant, as the introduction of approximate preservation disrupts the underlying combinatorial structure that fair clustering problems typically exhibit and that combinatorial approximation algorithms could exploit.

We go a different route. Our model is inspired by the well-studied k -center with outlier problem where a few points can be excluded from the clusters. For an initial motivation, consider the problem of finding a minimum cost maximum matching in a bipartite graph. If the two sides of the graph are of different cardinality, then we can only find a perfect matching for the smaller side. The larger side will have unmatched vertices, which can be viewed as outliers. The number of outliers is determined by the difference in cardinality of the two sides and the outliers are chosen such that the cost of the matching is minimized. This scenario is a special case of our model, with $m = 2$ and a target ratio of $1 : 1$. We want to generalize this idea to fair clustering. In its most general form, we have an arbitrary number of classes and allow the specification of a target ratio that might be different from the input ratio. In the fair clustering, the target ratio is achieved by excluding a certain number of points from different classes as outliers such that the ratio of the remaining points matches the target ratio.

Our model, applied to the two-color case, is as follows: Say the ratio of red to blue points in the input point set is $r : b$, where these numbers are coprime and $r < b$. Then the user can specify a target ratio $1 : s$ where $s \leq b/r$, i. e., it is achievable by excluding points from the majority color. Then our fair k -center with outliers formulation asks for $1 : s$ -fair clusters and allows the respective number of points of the second color to be marked as outliers. See Figure 1d for a $1 : 1$ -fair clustering with 1 blue outlier.

The benefit of this model is that it keeps the nice combinatorial structure of the exactly fair model (compared to the approximately fair models) but enables us to compute reasonable clusterings for any point set, not only those where the ratio of colors in the input is specifically nice. We extend known algorithms for fair k -center clustering to obtain our results. Table 1 summarizes our results for fair k -center for different cases. The first column describes the target ratio that we want to achieve. The first row is only for two colors and the important special case that the ratio is of the form $1 : t$ which allows for better approximation. Rows 2 and 3 are for arbitrary numbers of colors where row 3 allows ratios of the form $2 : 3 : 3$ while row 2 gives the easier case that is an extension of row 1. We explicitly list how expensive the fairlet decomposition is; the final clustering only increases the cost mildly.

One can view our algorithms also as a contribution to the field of combining different clustering constraints, in this case, fairness and outliers. However, note that the number of outliers is computed automatically from the target ratio, and that we never choose outliers from the minority color class.

■ **Table 1** Summary of our results on fair k -center with outliers. The final clustering is always computed by Algorithm 1, see Lemma 2.1. OPT refers to the optimal solution for the respective fairness condition.

	compute fairlets	fairlet cost	clustering cost
$1 : t$	Algorithm 3	2 OPT (Corollary B.3)	4 OPT (Theorem 2.2)
$1 : t_2 : \dots : t_m$	Algorithm 4	2 OPT (Corollary C.2)	4 OPT (Theorem C.4)
$t_1 : \dots : t_m$	Algorithm 5	12 OPT (Lemma D.5)	14 OPT (Theorem D.7)

1.1 Related work

We review work that is most related to our work and follows the seminal paper [10]. Various variants of fair clustering have been proposed that we do not discuss in detail, e. g., [2, 16, 18, 19]. Chierichetti, Kumar, Lattanzi, and Vassilvitskii [10] introduced the following model: For a point set $P = R \cup B$ consisting of red and blue points, the *balance* of a set $Y \subset P$ is the value $\min(\frac{|Y \cap R|}{|Y \cap B|}, \frac{|Y \cap B|}{|Y \cap R|}) \in [0, 1]$. The balance of a clustering is the smallest balance of any of its clusters. The focus of this model is to produce clusterings that are as balanced as possible, ideally reaching a balance of 1 (i. e., a 1 : 1 ratio). A clustering is t -fair in this model if every cluster has a balance value of at least t .

[10] also introduced the notion of fairlets. One striking observation in the paper was that for objective functions such as k -center and k -median, the computation of a cheap *fairlet decomposition*, i. e., a partitioning of P into fairlets, often becomes a clean combinatorial optimization problem. For example, if P itself is (exactly) 1-balanced, then computing a fairlet decomposition becomes a matching problem (in a suitable bipartite graph with weights depending on the objective), and constant factor approximations for computing 1-balanced clusterings can be deduced from this fact. Early follow-up works [15] and [23] provide speed-ups that increase the scalability of this approach.

Multi-color generalizations. If the protected attribute has more than two values, then a literal generalization of the balance-based model is the α -capped k -center problem introduced by [1]. There, the goal is to partition P in a way such that no single color (group) exceeds an α -fraction of any cluster. This generalization is motivated by real-world scenarios like online advertising, where keywords (points) are clustered, and advertisers (colors) should not dominate any single cluster to prevent manipulation. For two colors, the α -capped k -center problem equals the t -balanced fair k -center problem. This model focuses on limiting the majority's share in each cluster, not on matching the cluster representation to the original distribution of group proportions. Our work builds on the line of work that emphasizes accurate group representation and minority protection and uses a different generalization. The works [6, 7, 22] for the multi-color case demand that each protected group's share of each cluster must match the share in the overall population exactly. We call this model *exact representation* in Section 1.2. If the input contains the same number of points of each color, then [7] provides $O(1)$ -approximations for exact representation for various k -clustering objectives, including k -median and k -center. For preserving other input ratios $t_1 : t_2 : \dots : t_m$, $O(1)$ -approximations are only known for the k -center and k -supplier case, see [6, 22]. The best approximation ratio for fair k -center with exact representation is a 5-approximation for multiple colors [6].

Approximate representation. The works of [6] and [5] independently introduced a model for approximate representation where the ratio of color i in each cluster has to be in an interval $[\rho_i - \ell_i, \rho_i + u_i]$ where ρ_i is the input ratio and ℓ_i and u_i can be specified in the input. Both papers provide bi-criteria approximations for this model that incur a small additive fairness violation, and both results are obtained via LP rounding. Harb and Lam [14] employ a randomized approach that reduces the size of the LP in practice, providing a 3-approximation where clusters are fair only in expectation.

1.2 Definitions

Let P be a set of n points, $d: P \times P \rightarrow \mathbb{R}_{\geq 0}$ be a metric on P and $k \in \{1, \dots, n\}$ be the desired number of clusters. A k -clustering is a sequence of k pairwise disjoint sets $C_1, \dots, C_k \subset P$ and points $c_i \in C_i$ for $1 \leq i \leq k$ such that $\bigcup_{i=1}^k C_i = P$. For a given clustering instance (P, d, k) the k -center problem is to find such a k -clustering while minimizing the objective $\max_{i=1, \dots, k} \max_{p \in C_i} d(p, c_i)$. The fair k -center problem is a constrained version in which every point in the set P is assigned a color between 1 and m , and the clustering must preserve certain color ratios. We first define the input setting for such a problem.

► **Definition 1.1.** A fair clustering instance (P, d, k, m, γ) consists of a set P , a metric d on P , an upper bound on the number of clusters $k \in \mathbb{N}$, the number of colors m , and a color assignment $\gamma: P \rightarrow \{1, \dots, m\}$. We denote by $\Gamma = [m]$ the set of colors. For every $i \in \Gamma$ we denote by $H_i = \gamma^{-1}(i)$ the points in P with color i .

We assume that the colors are numbered such that $|H_1| \leq |H_2| \leq \dots \leq |H_m|$.

Fair k -clustering with outliers. Let $t_1, \dots, t_m \in \mathbb{N}$ be coprime integers that indicate the desired proportion of points from different protected groups within the clusters, i.e., a fair cluster should contain points from the m protected groups in the ratio $t_1 : t_2 : \dots : t_m$. We define:

► **Definition 1.2.** A cluster C_j is $t_1 : t_2 : \dots : t_m$ -fair for coprime integers $t_1, \dots, t_m \in \mathbb{N}$ if it satisfies for all $i \leq m$:

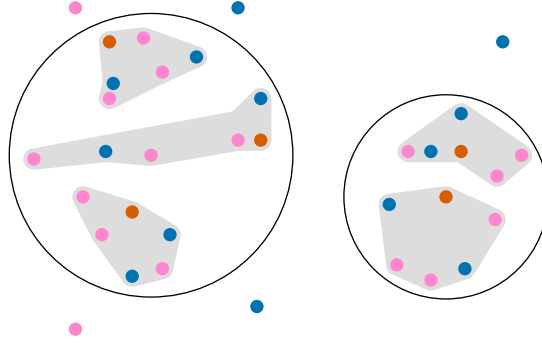
$$\frac{|C_j \cap H_i|}{|C_j|} = \frac{t_i}{\sum_{\ell \leq m} t_\ell}. \quad (1)$$

If there are no outliers, then fair clusters under this definition can only be obtained if the t_i match the input ratios. In the *exact representation model* one chooses the t_i by setting $t_i = |H_i| / \gcd(|H_1|, \dots, |H_m|)$ for all $i \leq m$. In fair k -clustering with outliers, we can also use other values for the t_i by allowing the fitting number of outliers for all but the minority color.

► **Definition 1.3.** Given an instance (P, d, k, m, γ) as defined in Definition 1.1, a $t_1 : t_2 : \dots : t_m$ -fair k -clustering with z outliers is given by a set of k centers $C = \{c_1, \dots, c_k\} \subseteq P$ and an assignment $\alpha: P \rightarrow C \cup \{\perp\}$ such that: (a) $|Z| \leq z$ where $Z = \alpha^{-1}(\perp)$, (b) every cluster $C_i := \alpha^{-1}(c_i)$ induced by $c_i \in C$ is $t_1 : t_2 : \dots : t_m$ -fair.

Equivalently, such a clustering can also directly be defined as a sequence of pairwise disjoint clusters $\mathcal{C} = \{C_1, \dots, C_k\}$ from P such that: (a) $Z = P \setminus \bigcup_{i \leq k} C_i$ has cardinality at most z , (b) C_i is $t_1 : t_2 : \dots : t_m$ -fair for all $1 \leq i \leq k$.

We will use both definitions interchangeably. Fair k -clustering with outliers can be applied to any k -clustering problem. In this paper, we initiate the study by investigating the metric k -center problem where $r(C, \alpha) = \max_{x \in P \setminus Z} d(x, \alpha(x))$ needs to be minimized.



■ **Figure 2** An optimal clustering with two pink and three blue outliers and a decomposition of the optimal clusters into fairlets.

1.2.1 General fairlets and fairlet decompositions

We now give a general definition of fairlets. A fairlet is an atomic fair cluster, i.e., a cluster that is fair and as small as possible.

► **Definition 1.4** ($t_1 : \dots : t_m$ -fairlet). Let $P = H_1 \dot{\cup} H_2 \dot{\cup} \dots \dot{\cup} H_m$ be a set of points, and $t_1, t_2, \dots, t_m \in \mathbb{N}$ with $\gcd(t_1, \dots, t_m) = 1$. A subset $f \subseteq P$ is a $t_1 : \dots : t_m$ -fairlet if $|f \cap H_i| = t_i$ for all $i \leq m$.

The concept of fairlets was introduced by [10], and generalized in subsequent papers [6, 7, 22], and our algorithms are adapted from [10] and [22]. The structural insight that we inherit from the exact representation model is that if every cluster in the optimal solution is $t_1 : \dots : t_m$ -fair, then every optimal cluster can be decomposed into fairlets, and thus P can be decomposed into a set of fairlets (see Figure 2 for an optimal $t_1 : \dots : t_m$ -fair k -center clustering with outliers with $m = 3$, $t_1 = 1$, $t_2 = 2$, $t_3 = 3$ and $k = 2$ and its decomposition into fairlets), and these fairlets are cheap (their diameter cannot exceed the diameter of an optimal cluster). Since this decomposition into cheap fairlets exists, finding it and then computing a clustering from it becomes the core of the problem, and this can be tackled via combinatorial algorithms.

For every fairlet f , we fix an *anchor* $\text{anc}(f) \in P$. This is a point that will represent the fairlet. It can be a point in the fairlet, but also outside of the fairlet. We define the *cost* of a fairlet f as $\text{cost}(f) = \max_{p \in f} d(\text{anc}(f), p)$. A *fairlet decomposition* of a set P is a maximal set of pairwise disjoint fairlets inside P . Let $\mathcal{F}$ be a fairlet decomposition. By $P(\mathcal{F}) = \bigcup_{f \in \mathcal{F}} f$, we denote the set of points covered by the fairlet decomposition. The *cost of a fairlet decomposition* is given by the maximal cost among any of its fairlets. We will also need the notion of agreement between a fairlet decomposition and a clustering.

► **Definition 1.5.** Let $\mathcal{C}$ be a fair k -center clustering with outliers and $\mathcal{F}$ be a fairlet decomposition of P . We say that $\mathcal{F}$ and $\mathcal{C}$ agree with each other iff for all $f \in \mathcal{F}$ there exists a cluster $C \in \mathcal{C}$ such that $f \subseteq C$.

We assume that the desired color ratio $t_1 : \dots : t_m$ fulfills $\gcd(t_1, \dots, t_m) = 1$. Otherwise, we can divide the t_i by their greatest common divisor. We distinguish between two cases. In the *one-sided outlier case*, we fix one minority color class from which we do not remove any outliers and the numbers of outliers in the other colors are determined by the excess of the number of points in relation to the minority color class. In this case, the outliers are a means to achieve the target ratio. In the *two-sided outlier case*, we are given a number z of outliers

we may exclude instead of it being determined by the number of points in the minority class. Thereby, we can choose outliers from all color classes. In this case, the outliers are a means to achieve the target ratio as well as to reduce the clustering cost. This variant is closer to the original idea of clustering with outliers; we give an algorithm for the 2-color case $m = 2$. In the one-sided outlier case, we further assume that $t_i|H_1| \leq t_1|H_i|$ for all $i \leq m$ and that $|H_1|/t_1 \in \mathbb{Z}$. Under these assumptions, it is possible to find a fairlet decomposition of size $|H_1|/t_1$ that fully covers H_1 , and the fairlet decomposition problem is to compute such a decomposition of minimal cost.

2 The algorithms

We extend a collection of results for fair k -center with exact representation to the case of fair k -center with outliers from [10, 22]. The general approach of our adaptation is as follows: Find a set of fairlets and so-called *anchors* and declare the points not contained in any fairlet as outliers. Then, find a set of centers among the non-outlier points (e.g., within the set of all anchors) and assign the non-outlier points to centers such that points inside the same fairlet are assigned to the same center. Outliers are never chosen from the color class that has the smallest cardinality. We exclude the minimum number of points from the other classes such that the point set achieves the target ratio. That is, the number of outliers is determined by the fairness criterion and the actual color distribution inside the given point set. (Note that in Appendix D.1 we also study a variant with two colors where outliers can be chosen from both color classes and give an 11-approximation for this scenario, but the main part of our paper does not allow choosing outliers from the minority color class.)

Algorithm 1 FAIR- k -CENTER-WITH-OUTLIERS.

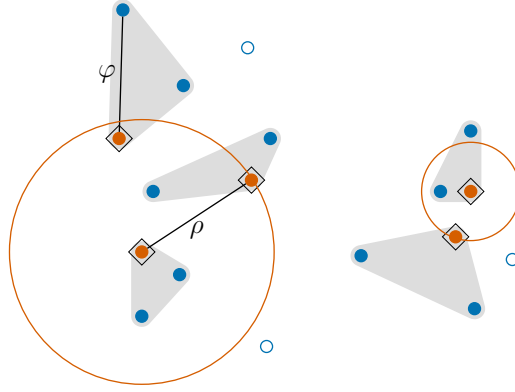
Input : Point set P , distances $d: P \times P \rightarrow \mathbb{R}_{\geq 0}$, number of clusters k
Output : Set of centers C , assignment $\alpha: P(\mathcal{F}) \rightarrow C$

```

1  $\mathcal{F}, \text{anc} \leftarrow \text{COMPUTE-FAIRLETS}(P, d, k)$ 
2  $A \leftarrow \{\text{anc}(f) \mid f \in \mathcal{F}\}$ 
3  $C \leftarrow \text{COMPUTE-CENTERS}(A, d, k)$ 
4 for  $f \in \mathcal{F}$  do
5    $c_f \leftarrow \arg \min_{c \in C} d(c, \text{anc}(f))$ 
6    $\forall p \in f: \alpha(p) \leftarrow c_f$ 
7 return  $C, \alpha$ 
```

Algorithm 1 describes the framework in detail. It uses the subroutines COMPUTE-FAIRLETS and COMPUTE-CENTERS. The latter is given in Line 3 and consists of running the well-known farthest-first traversal on A . Gonzalez showed that farthest-first traversal yields a 2-approximation for k -center that can be implemented to run in time $O(|A| \cdot k)$ [13]. In Lemma A.1 (Appendix A), we generalize the approximation result and show that the cost of the solution computed by farthest-first traversal on A is bounded by twice the cost of any valid k -center solution S on P . The difference here is that S might use points as centers that are not present in A . It is important that we run COMPUTE-CENTERS on the set of anchors, not on P (this might result in outliers being chosen as centers), and also not on the inlier points. The latter idea results in a worse approximation guarantee as shown in Section 3.1.

The subroutine COMPUTE-FAIRLETS differs depending on the case we are studying. Figure 3 illustrates the framework for 1 : 2-fairness. We want to achieve 1 : 2-fairness between the red and blue points. We use the red points as anchor points, which is convenient because



■ **Figure 3** Illustration of the clustering framework. Anchors are depicted by diamond frames, the fairlet decomposition is indicated by gray areas, and the outliers are depicted by unfilled circles. The red circles show an unconstrained clustering of the anchors. The fair clustering is obtained by assigning all points of a fairlet to the center that is closest to their anchor. The resulting fair clustering costs $\rho + \varphi$, where ρ is the cost of the unconstrained clustering of the red anchors and φ is the cost of the fairlet decomposition.

■ **Algorithm 2** COMPUTE-CENTERS [13].

Input : Set of points A , distances $d: A \times A \rightarrow \mathbb{R}_{\geq 0}$, number of clusters k
Output : Set of centers C

- 1 pick c_1 arbitrarily from A
- 2 **for** $i = 2, \dots, k$ **do**
- 3 $c_i \leftarrow \arg \max_{c \in A} d(c, \{c_1, \dots, c_{i-1}\})$

every fairlet contains exactly one red point. As we discuss later in Section 3.1, this choice is crucial for the approximation guarantee. There, we give an example on which choosing the centers from the set of all inliers merely yields an 8-approximation. In Figure 3, the fairlet decomposition excludes three blue points as outliers. We build a fair clustering by assigning each point in a fairlet to the center closest to its anchor. The fair clustering has radius $\varphi + \rho$. The analysis of the framework is done in Appendix A. We restate the cost result here.

► **Lemma 2.1.** *Assume that COMPUTE-FAIRLETS in Line 1 of Algorithm 1 returns a set of fairlets $\mathcal{F}$ such that $\varphi := \max_{f \in \mathcal{F}} \max_{p \in f} d(p, \text{anc}(f))$. Let A denote the set of anchors of the fairlets in $\mathcal{F}$. Further, assume that COMPUTE-CENTERS in Line 3 gives a set of centers $C \subseteq A$ such that $\rho := \max_{a \in A} \min_{c \in C} d(a, c)$. Then, Algorithm 1 computes a fair k -center solution on P with outliers $P \setminus P(\mathcal{F})$ and cost $\varphi + \rho$ (also see Figure 3).*

2.1 1 : t -fairness

The algorithm for this case follows [10, 22]. We have two colors, and the desired color ratio is $1 : t$ for some integer $t \geq 1$, i. e., we want to achieve $|B \cap C| = t \cdot |R \cap C|$ for all clusters C . Let us assume that $|B| \geq t|R|$. To guarantee a $1 : t$ -fair clustering, we exclude $|B| - t \cdot |R|$ blue outliers. Fairlets consist of one red and t blue points. For the special case of $t = 1$, fairlets consist of one blue and one red point each, and it suffices to compute a matching. If there are no outliers, i. e., if $|B| = |R|$, choosing the anchors carefully from the set of all points yields an approximation guarantee of 3, which we explain in detail in the full version. However, for

fair k -center with outliers, the same approach does not yield the same guarantee, but only a 6-approximation. Instead, if we limit the set of potential anchors to the red points, we are able to achieve a 4-approximation. As the approach for the $1 : 1$ -case is a special case of the $1 : t$ -case, we directly proceed with the general $1 : t$ -case. There, we have to match a red point to t blue points, and the red point is the fairlet's anchor. We identify $R = A$ for the remainder of this section, and justify this decision further in Section 3.1. The algorithm is described in Algorithm 3. To find a fairlet decomposition, it sets up a flow network

■ **Algorithm 3** COMPUTE-1 : t -FAIRLETS.

Input : Point set P , $R \subseteq P$, $B \subseteq P$, distance metric d , integer k , threshold factor τ

Output : Set of fairlets $\mathcal{F}$, anchor assignment $\text{anc} : \mathcal{F} \rightarrow R$

```

1  $\mathcal{F} \leftarrow \emptyset$ 
2  $D \leftarrow \{d(x, y) \mid x, y \in P\}$ 
3 sort  $D$  increasingly
4 for all  $\delta \in D$  do
5   construct flow network  $N_\delta = (V, E(\tau\delta), v, w, c)$ 
6    $\mathbf{f} \leftarrow$  maximum integral flow in  $N_\delta$ 
7   if  $|\mathbf{f}| = t|R|$  then
8     for all  $r \in R$  do
9        $f_r \leftarrow \{r\}$ 
10       $\text{anc}(f_r) \leftarrow r$ 
11      for all  $b$  such that  $(b, r) \in E_B^R(\tau\delta)$  do
12        if  $\mathbf{f}(b, r) = 1$  then
13           $f_r \leftarrow f_r \cup \{b\}$ 
14       $\mathcal{F} \leftarrow \mathcal{F} \cup \{f_r\}$ 
15 return  $\mathcal{F}, \text{anc}$ 

```

$N = (V, E(T), v, w, c)$ as follows: We set $V = B \cup R \cup \{v, w\}$ and $E(T) = E_v^B \cup E_B^R(T) \cup E_R^w$, where $E_v^B = \{(v, b) \mid b \in B\}$, $E_B^R(T) = \{(b, r) \mid b \in B, r \in R, d(b, r) \leq T\}$, for some threshold $T \geq 0$, and $E_R^w = \{(r, w) \mid r \in R\}$. In general, we want $T = \tau\delta$ for some guess δ of the radius of an optimal solution, and some threshold factor τ . Let OPT denote the cost of an optimal solution. As OPT equals the pairwise distance between two points, it is $\text{OPT} \in D$ and the for-loop will eventually consider $\delta = \text{OPT}$. In this iteration the threshold becomes $T = \tau \text{OPT}$. We choose τ such that we can prove that we can construct a fairlet decomposition of cost τOPT , which depends on the fairness case. For the $1 : t$ and $1 : t_2 : \dots : t_m$ cases, we set $\tau = 2$ because we can show that the cost of a fairlet decomposition that agrees with an optimal solution can be bounded by 2OPT if we use the anchor as a center, and we can find such a decomposition easily by a max flow computation. In the $t_1 : t_2 : \dots : t_m$ case, we need to set $\tau = 12$ because of a more intricate procedure to compute the fairlets. The capacities are given by

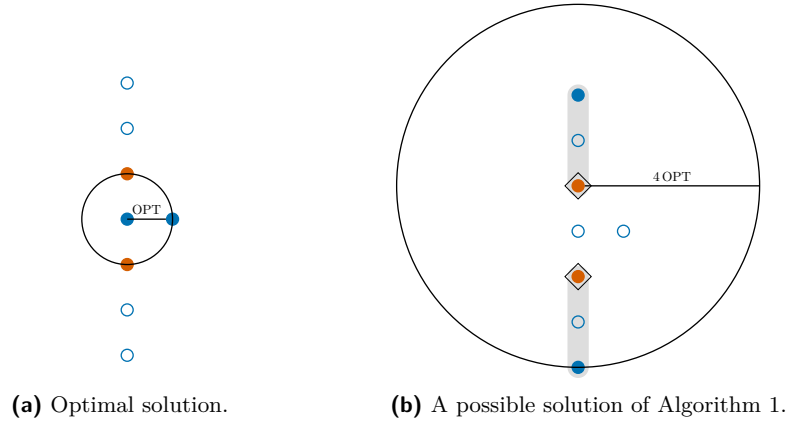
$$c(e) = \begin{cases} 1 & \text{if } e \in E_v^B \cup E_B^R(T) \\ t & \text{if } e \in E_R^w \end{cases}$$

If an edge $(b, r) \in E_B^R(T)$ carries flow, this means that b and r belong to the same fairlet.

The overview of the analysis of this method is presented in Appendix B. We state the main cost result here; note that a more general result, including the running time, is given in the next section for the related multicolor case.

► **Theorem 2.2.** *let $t \in \mathbb{N}$ be an integer and let $P = B \cup R$ be set of points consisting of blue and red points with $|B| \geq t|R|$. Algorithm 1 with COMPUTE-CENTERS = Algorithm 2 and COMPUTE-FAIRLETS = Algorithm 3 computes a center-aware 4-approximation for 1 : t -fair k -center with $|B| - t|R|$ blue outliers.*

The term *center-aware* in this theorem refers to the property that a center point is assigned to itself. This natural property is sometimes violated by algorithms for fair clustering, i. e., it can happen that a cluster can have a center which is itself not part of the cluster. We say that the approximation algorithm is center-aware if it is ensured that this cannot happen. Figure 4 illustrates a worst-case scenario that matches the 4-approximation guarantee. It demonstrates that, even in the 1 : 1 case, no better approximation can be guaranteed with this approach. Figure 4a shows an optimal solution for $k = 1$. Figure 4b illustrates a solution computed by Algorithm 1 for the same set of points. In this case, each fairlet has a diameter of 2OPT . The radius of the resulting clustering is 4OPT . This example illustrates how the choice of fairlets and outliers can affect the quality of the resulting clustering solution.



■ **Figure 4** Comparison of optimal solution and solution computed by Algorithm 1 for 1 : 1-fair k -center with $k = 1$. The gray areas indicate fairlets, and the unfilled circles are outliers. The fairlets consist only of the two filled points contained within the gray areas.

2.2 1 : t_2 : $\dots$: t_m -fairness

Now consider the scenario that P can be partitioned into $m \geq 2$ colors: $P = H_1 \dot{\cup} \dots \dot{\cup} H_m$, and we aim for a color ratio of $1 : t_2 : \dots : t_m$ for given values $t_2, \dots, t_m$. We assume that $|H_i| \geq t_i |H_1|$ so that it suffices to exclude outliers from the sets $H_2, \dots, H_m$. To achieve a fair clustering, we exclude $|H_i| - t_i |H_1|$ many points from H_i . Set $A = H_1$. Fairlets consist of one point from H_1 and t_i points from H_i for all $2 \leq i \leq m$. Similar to before, we will choose a fairlet's anchor from H_1 . We can find a fairlet decomposition by finding a $1 : t_i$ -fairlet decomposition on $H_1 \cup H_i$ for every $2 \leq i \leq m$ and gluing together fairlets that share a point in H_1 . This approach is visualized in Figure 5, formalized in Algorithm 4 and its analysis can be found in Appendix C.

► **Theorem 2.3.** *For $t_2, \dots, t_m \in \mathbb{N}$ and a fair clustering instance satisfying $t_i |H_1| \leq |H_i|$ for all $2 \leq i \leq m$, there exists a center-aware 4-approximation for 1 : t_2 : $\dots$: t_m -fair k -center with $|H_i| - t_i |H_1|$ outliers in H_i that runs in time $O(n^{2+o(1)} \log n)$.*

■ **Algorithm 4** COMPUTE-1: t_2 :...: t_m -FAIRLETS.

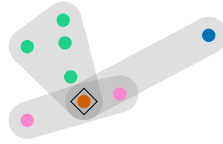
Input : Point set P , subsets $H_1, \dots, H_m \subseteq P$, distance metric d , number of clusters k , threshold factor τ

Output : Set of fairlets $\mathcal{F}$, anchor assignment $\text{anc}: \mathcal{F} \rightarrow H_1$

```

1 for  $i = 2, \dots, m$  do
2    $\mathcal{F}_i \leftarrow \text{COMPUTE-1:}t_i\text{-FAIRLETS}(P, H_1, H_i, d, k, \tau)$ 
3  $\mathcal{F} = \emptyset$ 
4 for  $h_1 \in H_1$  do
5    $f \leftarrow \bigcup_{2 \leq i \leq m} \{f' \mid f' \in F_i, h_1 \in f'\}$ 
6    $\text{anc}(f) \leftarrow h_1$ 
7    $\mathcal{F} \leftarrow \mathcal{F} \cup \{f\}$ 
8 return  $\mathcal{F}, \text{anc}$ 

```



■ **Figure 5** The gluing-process as described in Algorithm 4 for 1 : 1 : 2 : 4-fairness. We identify the anchor color (red) and start by individually computing 1 : 1-fairlets for the red and blue points, 1 : 2-red-pink fairlets, and 1 : 4-red-green-fairlets. The individual fairlets are indicated by gray areas, which overlap at one red point. The union of these fairlets forms one 1 : 1 : 2 : 4-fairlet.

The importance of Theorem 2.3 lies in its practicability and broad applicability. Computing maximum flows in practice is a well-studied problem; the approximation ratio is decent, and we can apply the result to any input point set. To do so, set $t_1 = 1$ and $t_i = \lfloor \frac{|H_i|}{|H_1|} \rfloor$, which satisfies the precondition of the theorem. This can be applied to any input point set without further parameterization.

3 General case, outliers in two colors and further considerations

The general case is of lesser practical interest since it requires the demanding computation of capacitated k -center solutions. We defer the algorithm and its analysis to Appendix D.

► **Theorem 3.1.** *For $t_1, t_2, \dots, t_m \in \mathbb{N}$ and a fair clustering instance satisfying $t_i |H_1| \leq t_1 |H_i|$ for all $2 \leq i \leq m$ and $\frac{|H_1|}{t_1} \in \mathbb{Z}$, there exists a center-aware 14-approximation for $t_1 : t_2 : \dots : t_m$ -fair k -center with $|H_i| - \frac{t_i}{t_1} |H_1|$ outliers from H_i for all $2 \leq i \leq m$ that runs in time $O(n^{2+o(1)} \log n)$.*

The aforementioned results use outliers for colors 2 to m to achieve the designated ratios. We also report a result for a variant that allows outliers in color 1. For this variant, the number of outliers z is an input parameter. In the above results, z is computed from the input point set to match the designated ratios. We study this concept only for the two-color case and balanced clusters. The algorithm and analysis can be found in Appendix D.1.

► **Theorem 3.2.** *Let $|H_2| \geq |H_1|$. For any $z \geq |H_2| - |H_1|$ there exists an 11-approximation for 1 : 1-fair k -center with z outliers from $H_1 \cup H_2$ that runs in time $O(n^5)$.*

We need the lower bound of $|H_2| - |H_1|$ on z to guarantee that a 1 : 1-fair solution exists. The clustering computed in Theorem 3.2 is not necessarily center-aware. In Appendix D.1, we also discuss how to obtain a center-aware solution, which is a 14-approximation.

3.1 Choosing centers from the inlier set

As a final note from the analysis, let us discuss our choice of computing centers from A which matches H_1 . This means that we choose centers only from the minority color class rather than from all input points. At first glance, this seems counterintuitive: We restrict the choice of centers to a smaller set. We do this because we cannot use the complete input point set since we may choose centers that are later deemed outliers. Indeed, this is quite likely since farthest-first traversal favours points that are far away. We could use the set of inliers to compute the centers. However, our inliers do not necessarily match the inliers of an optimum solution. Quite surprisingly, choosing the centers from the inliers – a larger set than the set A – yields a worse approximation ratio.

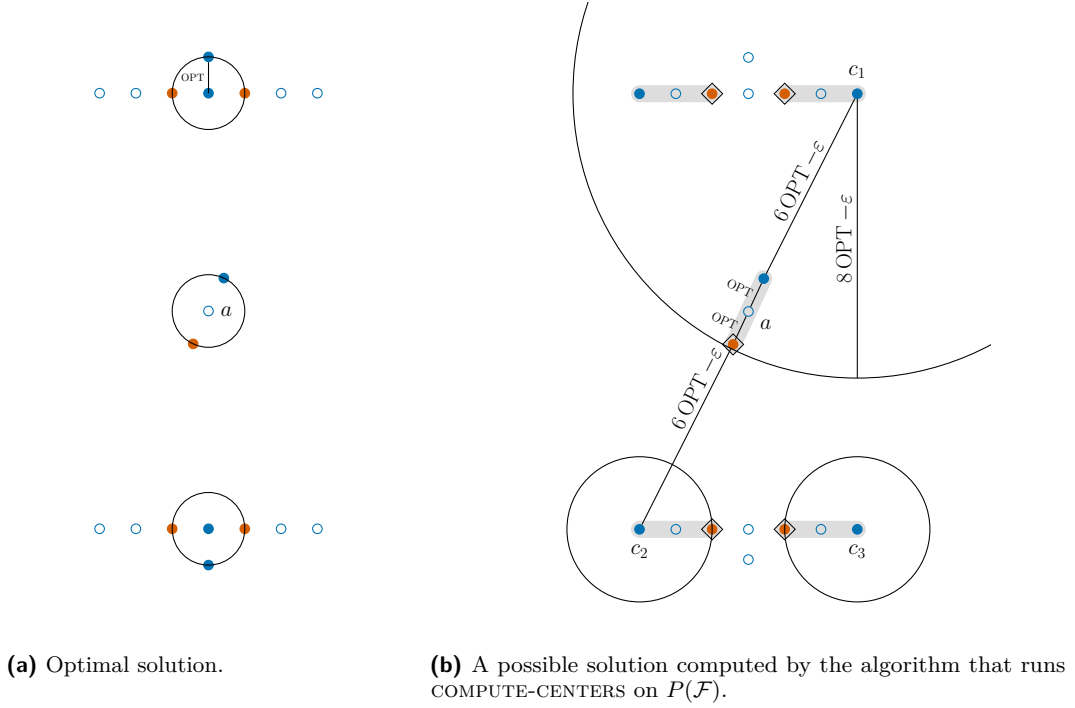


Figure 6 Comparison of an optimal solution and a solution computed by the algorithm that runs COMPUTE-CENTERS on $P(\mathcal{F})$ for 1 : 1-fair k -center with outliers for $k = 3$. Unfilled circles indicate outliers.

Assume that we use Algorithm 1, but with the slight modification that the set of centers C is computed

$$C \leftarrow \text{COMPUTE-CENTERS}(P(\mathcal{F}), d, k)$$

in Line 3. Figure 6 shows an instance on which the algorithm only yields an 8-approximation. Figure 6 (a) shows an optimal solution for $k = 3$ with radius OPT and Figure 6 (b) shows a potential solution computed by the considered algorithm variant with radius 8 OPT. The problem, as compared to the optimal solution, is two-fold. First, the fairlets are

computed. Here, we can only guarantee a cost of 2 OPT . As OPT is the smallest pairwise distance between any pair of points, the network for the max-flow computation does not differentiate between point pairs of distance OPT and distance 2 OPT . Therefore, a fairlet decomposition as depicted in Figure 6 might indeed be found. Then, the farthest-first traversal on $P(\mathcal{F})$ chooses the centers c_1, c_2 , and c_3 , in this order. In this construction, $d(c_1, a) = d(c_3, a) = d(c_2, a) = 7 \text{ OPT} - \varepsilon$, so it does not matter to which center the fairlet with anchor a is assigned. Assigning it to c_1 leads to an overall clustering cost of $8 \text{ OPT} - \varepsilon$. In the full version, we complement this by showing that the approximation factor is indeed 8.

4 Experimental Evaluation

We have argued in the previous sections that fair clustering algorithms often assume an ideal distribution of the protected attributes (e.g., a 1 : 1 male to female ratio) while realistic data often deviates from these ideal ratios. Previous work has dealt with this problem by randomly discarding input points (see, e.g., [10]) in order to establish the exact ratio in the input. We instead hypothesized that an informed choice of discarded outliers yields better results. In order to test the hypothesis, we compare the resulting clustering cost of [10] with random discarded outliers to the clustering cost resulting from our algorithms with an informed choice of outliers in computational experiments. We evaluate our fair k -center with outlier algorithms for the case that $t_1 = 1$, both for two colors and for multiple colors. Our evaluation is run on four datasets, *bank* [20], *census* [4], *diabetes* [11] and *income* [12].

■ **Table 2** Dataset parameters. n : subsample size, d : number of features, No. of test sets: number of test sets created from this dataset, Groups: The number of different attribute values of the protected attribute, Protected attribute: the name in the original dataset.

Dataset	n	d	No. of test sets	Groups	Protected attribute
Bank	1,000	7	45	2	married
Census	600	6	54	2	sex
Diabetes	1,000	15	91	2	gender
Income	1,000	9	195	2	SEX
Census-Race	600	6	54	5	race
Diabetes-Race	1,000	15	91	6	race

Our experimental setup mirrors the setup of the reference experiments in [10] where possible: We use the same datasets Bank, Census and Diabetes and the same protected attributes. Since the reference only exists for the 1 : 1 case, we generalize accordingly where appropriate. Like [10], we subdivide each dataset into numerous parts in order to create a large database for our evaluation and smooth out artifacts. The total number of test datasets is 385, see Table 2 (note that two datasets appear twice since we evaluate two different protected attributes). In order to create the test datasets, we randomized the order of the datasets and then cut it into pieces. In this fashion, all data present in the dataset is used, the expected ratio of each set matches the ratio of the overall dataset, but the individual test sets are all a bit different. Before splitting the dataset, the data is normalized: We first subtract the mean value from each attribute, then scale the values such that all attribute values lie in the interval $[-1, 1]$. We do this because the datasets contain columns with very different ranges. For each test set, we evaluate the algorithms for all $k \in \{1, 2, 5, 10, 20, 30, 40, 50, 100\}$. Since the algorithms are not randomized, each test set is evaluated once.

Algorithms. We implement Algorithm 1 for fair k -center with outliers. In the two-color $1 : t$ case, we use Algorithm 3 to compute the fairlets; in the multi-color $1 : t_2 : \dots : t_m$ case, we use Algorithm 4. Our algorithm implicitly identifies a set of z outliers during the fairlet decomposition, see below. In the two-color case, we compare ourselves to the algorithm proposed by Chierichetti et al. [10] for fair k -center without outliers. This algorithm assumes that the given point set satisfies the desired color ratio; hence, we first identify the number of points z needed to exclude from the majority class to achieve a balanced point set and subsample the points accordingly, i. e., we sample $n - z$ blue points uniformly at random and discard the other points.

Choosing the desired ratios for fair k -clustering with outliers automatically. Notice that we can *choose* the t_i for the specific input data at hand. In the exact representation model, this is done by setting $t_i = |H_i| / \gcd(|H_1|, \dots, |H_m|)$ for all $i \leq m$. This results in no outliers. We want to use outliers to allow for reasonable clusterings in cases where the input $|H_i|$ are, for example, coprime. For illustration, consider a dataset with $|H_1| = 100$ and $|H_2| = 201$. In the exact representation model without outliers, we would have $t_1 = 100$ and $t_2 = 201$ and a fairlet decomposition would just consist of one fairlet containing all points. However, the intuitively correct ratio is $1 : 2$, which would lead to much smaller fairlets of size 3 each and likely more meaningful clusterings. Removing only 1 outlier allows for such a ratio. In the following, we formalize this intuition: Recall that we assume $|H_1| \leq |H_2| \leq \dots, |H_m|$. An automatic way to choose the t_i in a beneficial and feasible way is to set

$$t_1 = 1 \text{ and } t_i = \lfloor |H_i| / |H_1| \rfloor$$

This results in a $1 : t_2 : \dots : t_m$ -fair k -clustering problem with outliers, and it can be applied to any input dataset in an automatic fashion. We always get ratios with $t_1 = 1$ in this fashion, thus we can apply the above mentioned special case algorithms. In the full version, we report the input ratios of all test sets and the ratios to which these were rounded down.

Cost comparison for two colors. We compare the clustering cost of our method to the method by Chierichetti et al. [10] in Figure 7. For each k , we compare the mean, min and max clustering cost over all subinstances. In each subinstance, we cluster to the rounded color ratio as described above. For the input of [10], we randomly sample outliers from the majority color to achieve the idealized ratio. The comparison shows the clustering cost by the more informed outlier selection of our method (blue) vs. [10] (orange). The diagrams also report the fairlet cost as it is often observed that this dominates the final cost in fair clustering, an effect that can also be seen here. As expected, our method gives lower clustering cost.

Effect of center choices on the clustering cost. In Figure 8, we demonstrate that our findings on the center choice are not merely theoretical. As opposed to Figure 7, we now restrict [10] to *only* use minority points as anchors, this forcing this algorithm to select minority points as centers. We observe a sizable decrease of the clustering cost. This is in line with our theoretical results. We observe that our method still gives better clustering cost.

Multicolor results. As in Figure 7, we evaluate how the clustering cost compares if we run our method and the method by [10] combined with random outlier selection. We cannot form as many clusters because the uneven distribution of the attribute values means that fairlets are large even after rounding the ratios. This is the case for CENSUS and DIABETES

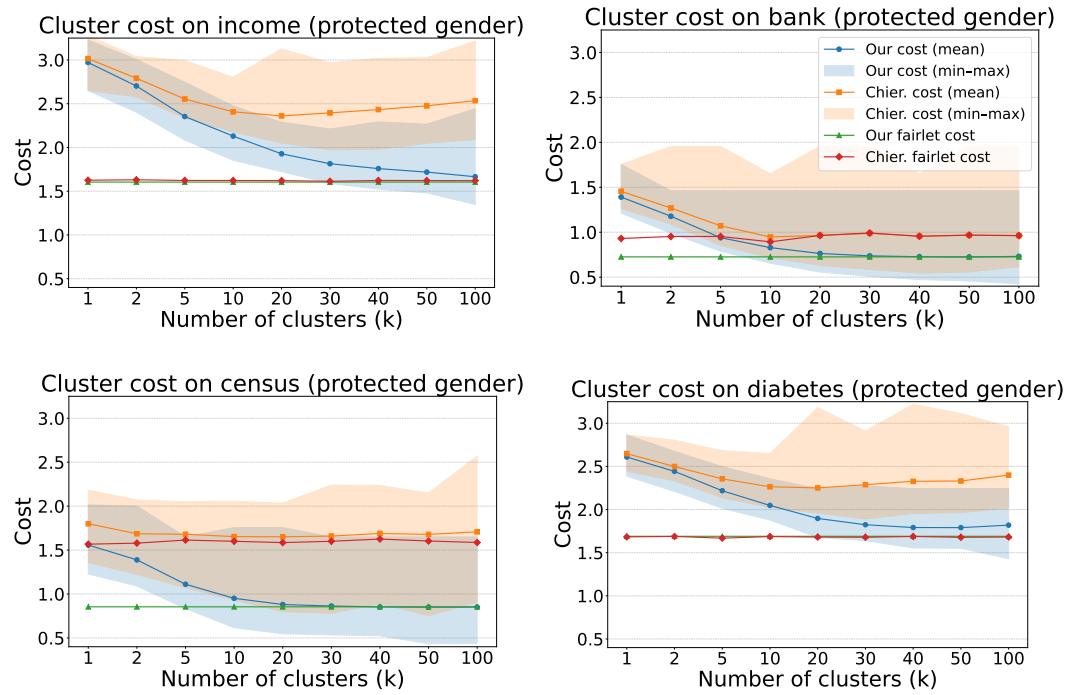


Figure 7 Comparison of the clustering cost of our method to random outlier selection.

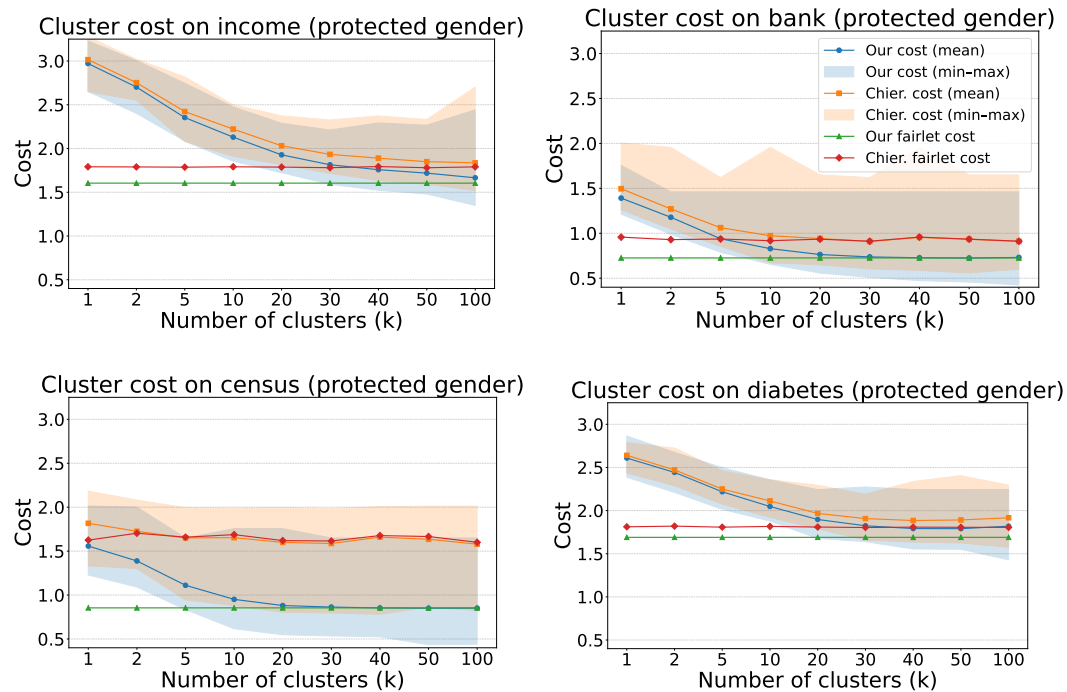
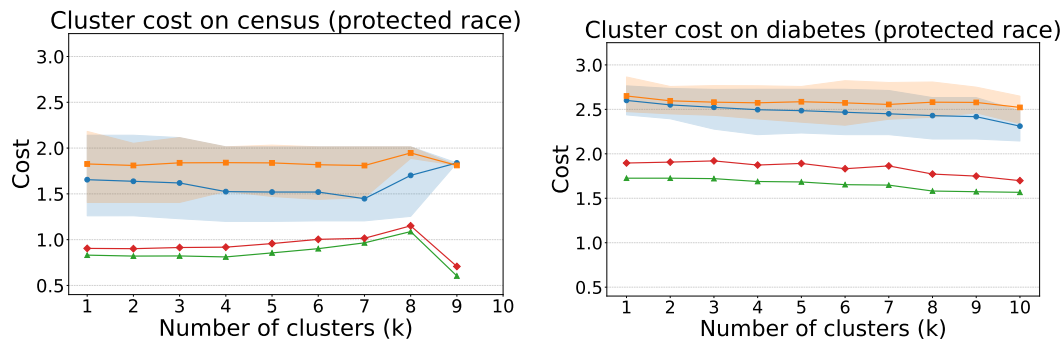


Figure 8 The effect of center choices.



■ **Figure 9** Cluster cost comparison for multiple colors.

with protected attribute RACE (also see the full version for evidence) where the minority group has very few points. This restricts the number of fairlets that we can build and the number of clusters that we can obtain. The result is depicted in Figure 9. We report the raw data for the plots in Figure 7 and 8 in the full version.

References

- 1 Sara Ahmadian, Alessandro Epasto, Ravi Kumar, and Mohammad Mahdian. Clustering without over-representation. In Ankur Teredesai, Vipin Kumar, Ying Li, Rómer Rosales, Evimaria Terzi, and George Karypis, editors, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, pages 267–275. ACM, 2019. doi:10.1145/3292500.3330987.
- 2 Georg Anegg, Haris Angelidakis, Adam Kurpisz, and Rico Zenklusen. A technique for obtaining true approximations for k -center with covering constraints. In Daniel Bienstock and Giacomo Zambelli, editors, *Integer Programming and Combinatorial Optimization - 21st International Conference, IPCO 2020, London, UK, June 8-10, 2020, Proceedings*, Lecture Notes in Computer Science, pages 52–65. Springer, 2020. doi:10.1007/978-3-030-45771-6_5.
- 3 Anna Arutyunova, Annika Hennes, and Daniel R. Schmidt. algo-hhu/fair-k-center-via-outliers. Software, DFG-456558332, swbId: swb:1:dir:c5548b3049a3013bb70e93d8345358258ee89e92 (visited on 2026-08-12). URL: <https://github.com/algo-hhu/fair-k-center-via-outliers/>. doi:10.4230/artifacts.27636.
- 4 Barry Becker and Ronny Kohavi. Adult. UCI Machine Learning Repository, 1996. DOI: <https://doi.org/10.24432/C5XW20>.
- 5 Suman Kalyan Bera, Deeparnab Chakrabarty, Nicolas Flores, and Maryam Negahbani. Fair algorithms for clustering. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d’Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 4955–4966, 2019. URL: <https://proceedings.neurips.cc/paper/2019/hash/fc192b0c0d270dbf41870a63a8c76c2f-Abstract.html>.
- 6 Ioana Oriana Bercea, Martin Groß, Samir Khuller, Aounon Kumar, Clemens Rösner, Daniel R. Schmidt, and Melanie Schmidt. On the cost of essentially fair clusterings. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019, Massachusetts Institute of Technology, Cambridge, MA, USA, September 20-22, 2019*, LIPIcs, pages 18:1–18:22. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.APPROX-RANDOM.2019.18.
- 7 Matteo Böhm, Adriano Fazzzone, Stefano Leonardi, Cristina Menghini, and Chris Schwiegelshohn. Algorithms for fair k -clustering with multiple protected attributes. *Oper. Res. Lett.*, 49(5):787–789, 2021. doi:10.1016/j.orl.2021.08.011.

- 8 Deeparnab Chakrabarty, Prachi Goyal, and Ravishankar Krishnaswamy. The non-uniform k -center problem. *ACM Trans. Algorithms*, 16(4):46:1–46:19, 2020. doi:10.1145/3392720.
- 9 Moses Charikar, Samir Khuller, David M. Mount, and Giri Narasimhan. Algorithms for facility location problems with outliers. In S. Rao Kosaraju, editor, *Proceedings of the Twelfth Annual Symposium on Discrete Algorithms, January 7-9, 2001, Washington, DC, USA*, pages 642–651. ACM/SIAM, 2001. URL: <http://dl.acm.org/citation.cfm?id=365411.365555>.
- 10 Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5029–5037, 2017. URL: <https://proceedings.neurips.cc/paper/2017/hash/978fce5bcc4eccc88ad48ce3914124a2-Abstract.html>.
- 11 John Clore, Krzysztof Cios, Jon DeShazo, and Beata Strack. Diabetes 130-US hospitals for years 1999-2008. UCI Machine Learning Repository, 2014. DOI: <https://doi.org/10.24432/C5230J>.
- 12 Frances Ding, Moritz Hardt, John Miller, and Ludwig Schmidt. Retiring adult: New datasets for fair machine learning. *Advances in Neural Information Processing Systems*, 34, 2021. URL: <https://proceedings.neurips.cc/paper/2021/hash/32e54441e6382a7fbacbbbf3c450059-Abstract.html>.
- 13 Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.*, 38:293–306, 1985. doi:10.1016/0304-3975(85)90224-5.
- 14 Elfarouk Harb and Ho Shan Lam. KFC: A scalable approximation algorithm for k -center fair clustering. In Hugo Larochelle, Marc’Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL: <https://proceedings.neurips.cc/paper/2020/hash/a6d259bfbfa2062843ef543e21d7ec8e-Abstract.html>.
- 15 Lingxiao Huang, Shaofeng H.-C. Jiang, and Nisheeth K. Vishnoi. Coresets for clustering with fairness constraints. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d’Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 7587–7598, 2019. URL: <https://proceedings.neurips.cc/paper/2019/hash/810dfbbebb17302018ae903e9cb7a483-Abstract.html>.
- 16 Xinrui Jia, Kshiteej Sheth, and Ola Svensson. Fair colorful k -center clustering. *Math. Program.*, 192(1):339–360, 2022. doi:10.1007/s10107-021-01674-7.
- 17 Samir Khuller and Yoram J. Sussmann. The capacitated k -center problem. *SIAM J. Discret. Math.*, 13(3):403–418, 2000. doi:10.1137/S0895480197329776.
- 18 Kunwoong Kim, Jihu Lee, Sangchul Park, and Yongdai Kim. Fair clustering via alignment. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu, editors, *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*, Proceedings of Machine Learning Research. PMLR / OpenReview.net, 2025. URL: <https://proceedings.mlr.press/v267/kim25t.html>.
- 19 Matthäus Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. Guarantees for spectral clustering with fairness constraints. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, Proceedings of Machine Learning Research, pages 3458–3467. PMLR, 2019. URL: <http://proceedings.mlr.press/v97/kleindessner19b.html>.
- 20 S. Moro, P. Rita, and P. Cortez. Bank Marketing. UCI Machine Learning Repository, 2012. DOI: <https://doi.org/10.24432/C5K306>.

- 21 Ariel D. Procaccia, Isaac Robinson, and Jamie Tucker-Foltz. School redistricting: Wiping unfairness off the map. In David P. Woodruff, editor, *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024*, pages 2704–2724. SIAM, 2024. doi:10.1137/1.9781611977912.97.
- 22 Clemens Rösner and Melanie Schmidt. Privacy preserving clustering with constraints. In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, Prague, Czech Republic, July 9-13, 2018*, LIPIcs, pages 96:1–96:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.ICALP.2018.96.
- 23 Melanie Schmidt, Chris Schwiegelshohn, and Christian Sohler. Fair coresets and streaming algorithms for fair k -means. In Evripidis Bampis and Nicole Megow, editors, *Approximation and Online Algorithms - 17th International Workshop, WAOA 2019, Munich, Germany, September 12-13, 2019, Revised Selected Papers*, Lecture Notes in Computer Science, pages 232–251. Springer, 2019. doi:10.1007/978-3-030-39479-0_16.
- 24 Georgios Tsaousoglou, Juan S. Giraldo, Pierre Pinson, and Nikolaos G. Paterakis. Fair and scalable electric vehicle charging under electrical grid constraints. *IEEE Trans. Intell. Transp. Syst.*, 24(12):15169–15177, 2023. doi:10.1109/TITS.2023.3311509.
- 25 Jan van den Brand, Li Chen, Richard Peng, Rasmus Kyng, Yang P. Liu, Maximilian Probst Gutenberg, Sushant Sachdeva, and Aaron Sidford. A deterministic almost-linear time algorithm for minimum-cost flow. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA, November 6-9, 2023*, pages 503–514. IEEE, 2023. doi:10.1109/FOCS57990.2023.00037.

This appendix gives an overview of the statements used in the analyses of our methods. We defer the proofs of the following statements to the full version of the paper. The full version further contains a detailed overview of the 1 : 1 without outliers case, a complementing analysis of the algorithm variant described in Section 3.1, and tables with the numerical results of our experiments.

A Analysis of the framework

► **Lemma A.1** ([13]). *Let S be the radius of a valid k -center solution on P . Then, COMPUTE-CENTERS (Algorithm 2) returns a set of centers $C \subseteq A$ such that $\max_{a \in A} \min_{c \in C} d(a, c) \leq 2S$.*

► **Lemma A.2.** *Let $\mathcal{F}$ be the fairlet decomposition computed by COMPUTE-FAIRLETS(P, d, k) in Line 1. Let $T_{\mathcal{F}}$ and T_C denote the running times of COMPUTE-FAIRLETS(P, d, k) and COMPUTE-CENTERS(A, d, k), respectively. Then, the running time of Algorithm 1 is in $O(T_{\mathcal{F}} + T_C + |\mathcal{F}| \cdot k + n)$.*

Note that we will show that the running time of the overall algorithm amounts to $O(n^{2+o(1)} \log n)$ in all one-sided outlier cases.

B Analysis of the 1 : t -fairlet decomposition

► **Lemma B.1.** *Assume that there exists a 1 : t -fairlet decomposition $\mathcal{F}$ such that $d(r, b) \leq \tau\delta$ for every $f \in \mathcal{F}$, the anchor $r \in f \cap R$ of f and every blue point $b \in f \cap B$ for some $\delta \in D$. Then, Algorithm 3 computes a 1 : t -fairlet decomposition with this property.*

► **Lemma B.2.** *Let $\mathcal{C}$ be a 1 : t -fair k -center clustering with outliers. Then there exists a 1 : t -fairlet decomposition $\mathcal{F}$ that agrees with $\mathcal{C}$.*

► **Corollary B.3.** *Let OPT_{fair} be the radius of an optimal 1 : t -fair k -center with outliers solution. Algorithm 3 computes a 1 : t -fairlet decomposition with $d(r, b) \leq 2 \text{OPT}_{\text{fair}}$ for every $f \in \mathcal{F}$, the anchor $r \in f \cap R$ of f and every blue point $b \in f \cap B$.*

► **Lemma B.4.** *The overall running time of Algorithm 3 is in $O(n^2 \log n + \frac{n^{2+o(1)}}{t} \log n)$ if it uses binary search to find the smallest $\delta \in D$ such that a maximum flow of value $t|R|$ is found in the network N_δ .*

C Analysis of the $1 : t_2 : \dots : t_m$ -fair decomposition

► **Lemma C.1.** *Let $D = \{d(p, q) \mid p, q \in P\}$ and let $\delta \in D$ such that for all $2 \leq i \leq m$, there exists a $1 : t_i$ -fairlet decomposition $\mathcal{F}_i$ of $H_1 \cup H_i$ and $d(h_1, h_i) \leq \tau\delta$ for $h_1 \in f \cap H_1$ and $h_i \in f \cap H_i$ for all $f \in \mathcal{F}$. Then, Algorithm 4 computes a $1 : t_2 : \dots : t_m$ -fairlet decomposition $\mathcal{F}$ of $H_1 \cup \dots \cup H_m$ excluding $|H_i| - t_i|H_1|$ outliers from H_i for all $2 \leq i \leq m$ such that $\max_{p \in f} d(p, h_1) \leq \tau\delta$, where $h_1 \in f \cap H_1$.*

In the final algorithm, we only need to change the way the fairlets are computed; everything else is done as in the $1 : t$ -case.

► **Corollary C.2.** *Let OPT_{fair} be the radius of an optimal $1 : t_2 : \dots : t_m$ -fair k -center with outliers solution. Algorithm 4 computes a $1 : t_2 : \dots : t_m$ -fairlet decomposition with $d(h_1, p) \leq 2 \text{OPT}_{\text{fair}}$ for every $f \in \mathcal{F}$, the anchor $h_1 \in f \cap H_1$ of f and every blue point $p \in f \cap (H_2 \cup \dots \cup H_m)$.*

► **Lemma C.3.** *Algorithm 4 (COMPUTE- $1 : t_2 : \dots : t_m$ -FAIRLETS) runs in $O(n^{2+o(1)} \log n)$.*

► **Theorem C.4.** *Algorithm 1 with $\text{COMPUTE-FAIRLETS} = \text{COMPUTE-}1 : t_2 : \dots : t_m\text{-FAIRLETS}$ (Algorithm 4) computes a 4-approximation for $1 : t_2 : \dots : t_m$ -fair k -center with $|H_i| - t_i|H_1|$ outliers in H_i in time $O(n^{2+o(1)} \log n)$.*

D Algorithm and analysis for $t_1 : t_2 : \dots : t_m$ -fair decompositions

Consider the most general case that there are m colors $H_1, H_2, \dots, H_m$ in the ratio $t_1 : t_2 : \dots : t_m$ for positive integers $t_1, t_2, \dots, t_m$. Assume that $|H_1| > 0$ is a multiple of t_1 . We want to find $C_1, \dots, C_k$ such that $t_i \cdot |H_j \cap C_\ell| = t_j \cdot |H_i \cap C_\ell|$ for all $\ell \leq k$ and all $1 \leq i, j \leq m$. Assume that $|H_i| \geq \frac{t_i}{t_1}|H_1|$ for all $2 \leq i \leq m$. A fair solution excludes $|H_i| - \frac{t_i}{t_1}|H_1|$ outliers from H_i .

► **Lemma D.1.** *Let $\mathcal{C}$ be a $t_1 : t_2 : \dots : t_m$ -fair k -center clustering with outliers. Then there exists a $t_1 : t_2 : \dots : t_m$ -fairlet decomposition $\mathcal{F}$ that agrees with $\mathcal{C}$.*

To compute $t_1 : t_2 : \dots : t_m$ -fair fairlets, we first find the anchors of the fairlets by solving an instance of the *capacitated multi- ℓ -center problem* (also called *capacitated k -center problem* with soft capacities).

► **Definition D.2** (The capacitated multi- ℓ -center problem). *Given a set of points P and numbers $\ell \in \mathbb{N}$ and $u \in \mathbb{N}$. The goal is to find a multiset of at most ℓ centers C^ℓ and an assignment $\phi : P \rightarrow C^\ell$ from points to centers such that the ℓ -center objective is minimized, while $|\phi^{-1}(c_i)| \leq u$ for all $i \leq \ell$.*

► **Lemma D.3** ([17]). *Let $\mathcal{I} = (P, \ell, u)$ be an instance of the capacitated multi- ℓ -center problem. There is an algorithm that computes a 5-approximation for $\mathcal{I}$ in time $O(n^{2+o(1)} \log n)$.*

Lemma D.3 gives an upper bound on the radii of the capacitated multi- ℓ -center clustering in terms of the cost of an optimal capacitated multi- ℓ -center solution. We need the following result to get a bound in terms of OPT_{fair} .

► **Corollary D.4.** *Let (A^*, α^*) be an optimal solution to the capacitated multi- ℓ -center problem on point set H_1 with capacity t_1 . Then, $\max_{p \in H_1} d(p, \alpha^*(p)) \leq 2 \text{OPT}_{\text{fair}}$.*

By assumption on $|H_1|$, a $t_1 : t_2 : \dots : t_m$ -fairlet decomposition covers H_1 completely. The idea is to compute a capacitated ℓ -center clustering with $\ell = \frac{|H_1|}{t_1}$ and capacity t_1 . Let (A, ℓ) be such a capacitated ℓ -center solution. The set A might contain multiple copies of points from H_1 , i.e., A is a multiset of points from H_1 . The points in A will serve as the anchors for the fairlets that we will construct as follows. We can reduce the construction of a $t_1 : \dots : t_m$ -fairlet decomposition of P to the construction of a $1 : t_1 : t_2 : \dots : t_m$ -fairlet decomposition on $P' = A \dot{\cup} H_1 \dot{\cup} H_2 \dot{\cup} \dots \dot{\cup} H_m$ where the anchors A are added to form a new color set. Notice that every $p' \in P'$ is (the copy of) a point $p \in P$. Let $d' : P' \times P' \rightarrow \mathbb{R}$ be the distance function induced by $d : P \times P \rightarrow \mathbb{R}$, where the distance between two points equals the distance between their original points. Note that d' is a pseudometric. However, this does not cause any problems, as all results so far hold in this setting as well. Algorithm 5 formalizes the fairlet computation.

■ **Algorithm 5** COMPUTE- $t_1 : t_2 : \dots : t_m$ -FAIRLETS.

Input : Point set $P = H_1 \dot{\cup} H_2 \dot{\cup} \dots \dot{\cup} H_m$, distance metric d , number of clusters k
Output : Set of fairlets $\mathcal{F}$, anchor assignment $\text{anc} : \mathcal{F} \rightarrow A$

- 1 $(A, \kappa) \leftarrow$ 5-approx. sol. for capacitated multi- $\frac{|H_1|}{t_1}$ -center on H_1 with capacity t_1
- 2 $\mathcal{F}', \text{anc}' \leftarrow$ COMPUTE- $1 : t_1 : t_2 : \dots : t_m$ -FAIRLETS($P', A, H_1, H_2, \dots, H_m, d', k, 12$)
- 3 $\mathcal{F} \leftarrow \emptyset$
- 4 $\forall f' \in \mathcal{F}' : f \leftarrow f' \setminus A, \text{anc}(f) \leftarrow \text{anc}'(f'), \mathcal{F} \leftarrow \mathcal{F} \cup \{f\}$
- 5 **return** $\mathcal{F}, \text{anc}$

► **Lemma D.5.** *Algorithm 5 computes a fairlet decomposition $\mathcal{F}$ with $|H_i| - \frac{t_i}{t_1}|H_1|$ outliers from H_i for all $2 \leq i \leq m$ such that for all $f \in \mathcal{F}$, $\max_{p \in f} d(\text{anc}(f), p) \leq 12 \text{OPT}_{\text{fair}}$, where $\text{anc}(f)$ is the anchor of fairlet f .*

► **Lemma D.6.** *Algorithm 5 (COMPUTE- $t_1 : \dots : t_m$ -FAIRLETS) runs in time $O(n^{2+o(1)} \log n)$.*

In the complete approach, Algorithm 1 uses Algorithm 2 on the multiset A and sets the final set of centers to be the underlying points from which the copies in A originated. As we only identify centers that have zero distance, the bound on the cost does not change.

► **Theorem D.7** (Reformulation of Theorem 3.1). *For $t_1, \dots, t_m \in \mathbb{N}$ and a fair clustering instance satisfying $t_i|H_1| \leq t_1|H_i|$ for all $2 \leq i \leq m$ and $\frac{|H_1|}{t_1} \in \mathbb{Z}$, Algorithm 1 with COMPUTE-FAIRLETS = COMPUTE- $t_1 : \dots : t_m$ -FAIRLETS (Algorithm 5) computes a center-aware 14-approximation for $t_1 : t_2 : \dots : t_m$ -fair k -center with $|H_i| - \frac{t_i}{t_1}|H_1|$ outliers from H_i for all $2 \leq i \leq m$ in time $O(n^{2+o(1)} \log n)$.*

D.1 1 : 1 fairness with outliers on both sides

So far, we used outliers to establish the desired fairness ratio in clusters. However, since k -center is prone to outliers, we may also want to combine fairness and outliers to satisfy fairness and simultaneously exclude points that are far away. In previous sections, the number of outliers from group H_i is always a fixed number computed from the desired color ratio and the numbers $|H_1|$ and $|H_i|$. We now want to allow more outliers so that it is possible to discard points that are far away. We consider the case of two colors, i.e., $P = R \dot{\cup} B$ with

$|R| \leq |B|$ and a given number $z \geq |B| - |R|$ of outliers. If a solution has z_r outliers in R , it has exactly $z_b = z_r + |B| - |R|$ outliers in B since otherwise the above fairness condition is violated.

We now develop an algorithm for 1 : 1-fair k -center with outliers. Similarly to the previous algorithms, we start by constructing a bipartite graph $G = (V, E)$ with $V = R \cup B$ and $E = \{\{r, b\} \mid r \in R, b \in B \text{ and } d(r, b) \leq \delta\}$, i. e., edges between any pair of a red and a blue point are added if their distance is upper bounded by δ . Let $\mathcal{F}_\delta$ denote a maximum matching in this graph. If $\mathcal{F}_\delta$ has cardinality at least $\frac{|P|-z}{2}$, we store $\mathcal{F}_\delta$, and otherwise, we discard it. Let $\delta_{\min}$ denote the smallest threshold such that $\mathcal{F}_{\delta_{\min}}$ has cardinality at least $\frac{|P|-z}{2}$. We can determine $\delta_{\min}$ via a binary search and then compute a maximum matching for all $\delta \geq \delta_{\min}$. Let $\mathcal{F} = \{\mathcal{F}_\delta \mid \delta \geq \delta_{\min}\}$ be the set of all such matchings.

■ **Algorithm 6** SET-OF-POSSIBLE-1 : 1-FAIRLETS.

Input : Point set $P = R \cup B$, distance metric d , integer k , number of outliers z

Output : Set of possible fairlet decompositions $\mathcal{F}$

```

1  $D \leftarrow \{d(x, y) \mid x, y \in P\}$ 
2  $\mathcal{F} \leftarrow \emptyset$ 
3 for all  $\delta \in D$  do
4   construct bipartite graph  $G = (R \cup B, \{\{r, b\} \mid r \in R, b \in B, d(r, b) \leq \delta\})$ 
5    $\mathcal{F}_\delta \leftarrow$  maximum matching in  $G$ 
6   if  $|\mathcal{F}_\delta| \geq \frac{|P|-z}{2}$  then
7      $\mathcal{F} \leftarrow \mathcal{F} \cup \{\mathcal{F}_\delta\}$ 
8 return  $\mathcal{F}$ 
```

► **Lemma D.8.** Let OPT denote the optimal cost of a 1 : 1-fair k -center solution with outliers. Then we have $2 \text{OPT} \geq \delta_{\min}$.

The proof also shows that there exists a fairlet decomposition $\mathcal{F}_\delta \in \mathcal{F}$ which is a maximum matching in the graph $G_{2\text{OPT}} = (R \cup B, \{\{r, b\} \mid r \in R, b \in B, d(r, b) \leq 2\text{OPT}\})$. We denote this matching by $\mathcal{F}_{2\text{OPT}}$ for convenience. For every δ we denote by $Z_\delta = P \setminus P(\mathcal{F}_\delta)$ the points which are not part of a fairlet in $\mathcal{F}_\delta$. Notice that if $|\mathcal{F}_\delta| \geq \frac{|P|-z}{2}$ we have $|Z_\delta| \leq z$. Since points from both colors can be outliers, both points of a fairlet may be outliers in the optimal solution. Since the cost of an optimal solution is not known, it is not sufficient to compute the fairlet decomposition for $\delta_{\min}$ as we see in Figure 10.



■ **Figure 10** Optimal clustering for $k = 1$, $z = 2$ shown in green. However, if we set δ too small, then we might only find the fairlets shown in gray. A clustering based on these fairlets would have to place both fairlets in the same cluster. As we could place the two points to the right arbitrarily far away, the cost of such a clustering could be arbitrarily high.

► **Lemma D.9.** *SET-OF-POSSIBLE-1 : 1-FAIRLETS (Algorithm 6) runs in time $O(n^{4+o(1)})$.*

Proof. The set of pairwise distances D has cardinality $O(n^2)$. For every $\delta \in D$, we construct a bipartite graph in $O(n^2)$ and compute a maximum matching in this graph in $O(n^{2+o(1)})$ by computing a maximum flow in the corresponding network using the algorithm by [25]. In total, this yields a running time of $O(n^{4+o(1)})$ and dominates the overall running time. ◀

■ **Algorithm 7** 1 : 1-FAIR- k -CENTER-WITH-OUTLIERS-ON-BOTH-SIDES.

Input : Set of fairlet decompositions $\mathcal{F}$
Output : 1 : 1-fair k -center solution with $\leq z$ outliers

```

1 for all  $\mathcal{F}_\delta \in \mathcal{F}$  do
2    $z_\delta \leftarrow \frac{z - |Z_\delta|}{2}$ ,  $R' \leftarrow R \setminus Z_\delta$ 
3   let  $(C_R, \alpha_R)$  be a  $k$ -supplier with outliers solution on  $(R', P, d, z_\delta, k)$ 
4    $C_\delta \leftarrow C_R$ 
5   for all  $\{r, b\} \in \mathcal{F}_\delta$  do
6      $\alpha_\delta(r) = \alpha_R(r)$ 
7      $\alpha_\delta(b) = \alpha_R(r)$ 
8   for all  $p \in Z_\delta$  do
9      $\alpha_\delta(p) = \perp$ 
10 return  $\operatorname{argmin}_{\delta \geq \delta_{\min}} \operatorname{cost}(C_\delta, \alpha_\delta)$ 

```

Similarly to Definition 1.5, we define the notion of a fairlet decomposition and a clustering that agree with each other. Notice that this definition is slightly different, as fairlets can now also be outliers.

► **Definition D.10.** Let $\mathcal{C}$ be a clustering with outliers Z and $\mathcal{F}_\delta \in \mathcal{F}$ a fairlet decomposition with $Z_\delta = P \setminus P(\mathcal{F}_\delta)$. We say that $\mathcal{F}_\delta$ and $\mathcal{C}$ agree with each other if and only if $Z_\delta \subset Z$ and for all $f \in \mathcal{F}_\delta$ either $f \subset Z$ or there exists a cluster $C \in \mathcal{C}$ such that $f \subseteq C$.

Suppose that a clustering $\mathcal{C}$ with outliers Z agrees with $\mathcal{F}_\delta \in \mathcal{F}$, then the number of fairlets from $\mathcal{F}_\delta$ which are contained in Z is upper bounded by $z_\delta = \frac{z - |Z_\delta|}{2}$. Furthermore, $\mathcal{C}$ satisfies 1 : 1 fairness since every cluster in $\mathcal{C}$ is a union of fairlets from $\mathcal{F}_\delta$. We can find a clustering $\mathcal{C}$ that agrees with $\mathcal{F}_\delta$ by reduction to k -supplier with outliers.

► **Definition D.11.** An instance of k -supplier with outliers is given by (X, F, d, z, k) , where $(X \cup F, d)$ is a metric space, and $z, k \in \mathbb{N}$. The goal is to choose a set $C \subset F$ of cardinality at most k and an assignment $\alpha: X \rightarrow C \cup \{\perp\}$ such that the set of outliers $Z = \alpha^{-1}(\perp)$ has cardinality at most z and the cost $\max_{p \in X \setminus Z} d(p, \alpha(p))$ is minimized.

We compute a k -supplier with outliers solution on points $R \setminus Z_\delta$ and centers from P with z_δ outliers. Given such a solution (C_R, α_R) we set $C_\delta = C_R$ and for every fairlet $\{r, b\} \in \mathcal{F}_\delta$ we set $\alpha_\delta(b) = \alpha_\delta(r) = \alpha_R(r)$ and $\alpha_\delta(p) = \perp$ for all $p \in Z_\delta$. Then at most $2z_\delta$ points from $P(\mathcal{F}_\delta)$ are marked as outliers and additionally $|Z_\delta|$ points. In conclusion, the number of outliers does not exceed z , and the clustering agrees with $\mathcal{F}_\delta$. By definition of α_δ , the resulting clustering $(C_\delta, \alpha_\delta)$ is also 1 : 1-fair. To bound the cost of the final clustering, we bound the cost of a k -supplier solution.

► **Lemma D.12.** Let OPT denote the optimal cost of a 1 : 1-fair k -center solution with outliers. There exists a k -supplier with outliers solution (C, α) on $(R \setminus Z_{2\text{OPT}}, P, d, z_{2\text{OPT}}, k)$ with cost at most 3OPT .

► **Lemma D.13.** *Let $(C_\delta, \alpha_\delta)$ be the solution computed by Algorithm 7 on $\mathcal{F}_\delta$ and let r be the radius of the solution (C_R, α_R) computed in Line 3. Then $(C_\delta, \alpha_\delta)$ is 1 : 1-fair, has at most z outliers and its cost is upper bounded by $r + \delta$.*

There exists a 3-approximation algorithm for k -supplier with outliers [9]. Using this to compute a k -supplier with outliers solution (C_R, α_R) on (R', P, d, z_δ, k) in Algorithm 7, we obtain the following result.

► **Observation D.14.** *Notice that the clustering computed in Algorithm 7 is not necessarily center-aware, since centers can be outliers. For the center-aware version, we can similarly prove a 14-approximation in the following way: Suppose we can ensure that $C_R \subset R \setminus Z_\delta$ in Line 3. Let $r \in C_R$ and $\{r, b\} \in \mathcal{F}_\delta$ be the fairlet containing r . We set $\alpha_\delta(r) = \alpha_\delta(b) = r$ to obtain a center-aware solution. By Lemma D.12, there exists a solution (C, α) for k -supplier with outliers on $(R \setminus Z_\delta, P \setminus Z_\delta, d, z_\delta, k)$ with cost 3 OPT . For $c \in C$ and its cluster D , we can pick an arbitrary point from $D \cap (R \setminus Z_\delta)$ as a new center. This increases the cost by 3 OPT . Thus, there exists a solution with cost 6 OPT for k -center with z_δ outliers on $R \setminus Z_\delta$. If we replace Line 3 by a 2-approximation for k -center with outliers [8], this yields a 14-approximation in total.*