

---

# Fair Clustering Under a Bounded Cost

---

**Seyed Esmaili**  
University of Maryland  
esmaeili@cs.umd.edu

**Brian Brubach**  
Wellesley College  
bb100@wellesley.edu

**Aravind Srinivasan**  
University of Maryland  
asriniv1@umd.edu

**John P. Dickerson**  
University of Maryland  
johnd@umd.edu

## Abstract

Clustering is a fundamental unsupervised learning problem where a dataset is partitioned into clusters that consist of nearby points in a metric space. A recent variant, *fair* clustering, associates a *color* with each point representing its group membership and requires that each color has (approximately) equal representation in each cluster to satisfy group fairness. In this model, the cost of the clustering objective increases due to enforcing fairness in the algorithm. The relative increase in the cost, the “price of fairness,” can indeed be unbounded. Therefore, in this paper we propose to treat an upper bound on the clustering objective as a constraint on the clustering problem, and to maximize equality of representation subject to it. We consider two fairness objectives: the group utilitarian objective and the group egalitarian objective, as well as the group leximin objective which generalizes the group egalitarian objective. We derive fundamental lower bounds on the approximation of the utilitarian and egalitarian objectives and introduce algorithms with provable guarantees for them. For the leximin objective we introduce an effective heuristic algorithm. We further derive impossibility results for other natural fairness objectives. We conclude with experimental results on real-world datasets that demonstrate the validity of our algorithms.

## 1 Introduction

Machine learning algorithms are increasingly being applied to settings that directly influence human lives. This has spurred a growing *fair machine learning* community [9], which develops machine learning algorithms that are made to satisfy certain fairness criteria. Choosing an appropriate definition of fairness—and even deciding *if* explicitly defining fairness is appropriate to begin with—is a morally-laden and application-specific decision [28, 43]. We make no normative statements here; rather, we focus on a commonly-used and often legally-backed family of fairness definitions—*group fairness*—in the context of *clustering*, arguably the most fundamental unsupervised learning problem.

A recent group-membership fairness definition, called *fair clustering* in the literature, has received significant interest [20, 10, 11, 2, 7, 31, 22, 8, 3, 33, 1]. In fair clustering, each point has a color that designates its group membership, and a clustering objective such as  $k$ -median or  $k$ -means is given. The goal is to find a clustering that minimizes the objective subject to the constraint that each cluster has each color represented within some pre-specified proportions. For example, there may be two colors, red and blue, and the constraint could require per-color representation between 40% and 60%.

An acknowledged fact in fair clustering—and, indeed, in many allocation and matching settings—is that the fairness (e.g., proportion) constraint could cause degradation in the clustering objective [12, 19]. A point may be assigned to a further away center (cluster) to satisfy the proportion constraint [20].

The degradation in the objective due to the imposed fairness constraint is called the *price of fairness* (PoF), mathematically defined as  $\text{PoF} = (\text{cost of fair solution}) / (\text{cost of agnostic solution})$ .

Unlike some examples in the literature [12, 21], the price of fairness in the case of fair clustering is unbounded, as seen in Figure 1. By enforcing a form of group fairness requiring an even split across colors in each cluster, a fair clustering algorithm would perform arbitrarily poorly as the two groups of points separate in space, while a “color-blind” algorithm would remain unchanged (bottom-left and bottom-right of Figure 1, respectively). The possibly unbounded increase in the clustering cost (unbounded price of fairness) indicates that fair clustering can yield clusters consisting of points that are far apart in the metric space instead of combining nearby points—often the main motivation behind clustering in machine learning and data analysis. Furthermore, the legal notion of disparate impact does not force an organization to output a fair clustering if it can justify an unfair one due to “business necessity,” i.e., potential loss in quality [45, 44]. This possible conflict between the clustering objective and the fairness constraint indicates the need for fair clustering algorithms that operate in a setting where the clustering cost cannot exceed a pre-set upper bound.

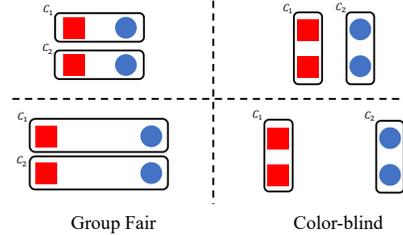


Figure 1: Comparison between group fair (left) and color-blind (right) clustering. Unlike color-blind clusters, group fair clusters may combine faraway points (bottom-left).

**Our Contributions.** In this paper, we address fair clustering under an exogenous threshold on the clustering objective. We formulate the problem mathematically in a general setting that captures all of the traditional  $k$ -clustering objectives, i.e.  $k$ -center,  $k$ -median, and  $k$ -means. Throughout, we focus on two general formulations of group fairness: GROUP-UTILITARIAN and GROUP-EGALITARIAN, along with the GROUP-LEXIMIN objective which generalizes the traditional GROUP-EGALITARIAN definition. We show that these objectives lead to problems that are NP-hard in general. Further, assuming  $P \neq NP$  we derive lower bounds on the additive approximation of any polynomial time algorithm for the GROUP-UTILITARIAN and GROUP-EGALITARIAN objectives. We provide bi-criteria approximation algorithms for the GROUP-UTILITARIAN and GROUP-EGALITARIAN objectives in which the constraint has a bounded violation and the objective is bounded from the optimal value by an additive error. For the GROUP-LEXIMIN objective we provide an effective heuristic. Further, we consider other possibly more “flexible” fairness objectives, but demonstrate inapproximability results for them. Finally, we test the performance of our algorithms on a collection of datasets and see that we obtain good solutions with low “fairness violations.” We note that due to the page limit, all proofs are placed in Appendix A.

## 2 Related Work

The metric clustering problems  $k$ -center,  $k$ -median, and  $k$ -means are fundamental in unsupervised learning and operations research. All are NP-hard with a long line of research on approximation algorithms. For  $k$ -center, two distinct algorithms achieve a 2-approximation which is tight assuming  $P \neq NP$  [26, 23, 27]. The current best approximation for  $k$ -median is a  $(2.675 + \epsilon)$ -approximation in  $n^{O((1/\epsilon) \log(1/\epsilon))}$  time [18], and for  $k$ -means, there is a 6.357-approximation [4].

The *fair* clustering problem with group fairness constraints was proposed by [20]. They studied  $k$ -center and  $k$ -median in a setting with only two colors. Followup work by [11, 10, 7, 31] gave extensions to the  $k$ -means objective, more than two colors, multi-color points (i.e., intersecting demographic groups), and scalability. Other works by [38, 15, 16] look at non-group-fairness definitions; the former investigates individual fairness, while the latter two address probabilistic fairness guarantees for pairs or communities of points. The aforementioned works optimize the clustering objective subject to fairness constraints; however, satisfying the fairness constraints may come at the expense of a significant increase in the clustering objective. Accordingly [48] and very recently [37] explored the cost/fairness tradeoff, but using a multi-objective approach. Unlike our work they do not establish approximation guarantees. Further, the fairness objectives used are different from ours, i.e. in [48] the fairness objective is penalized for proportions that are not precisely equal to the population level while in [37] it is penalized for color ratios that are not equal to 1. Moreover, in [48] the cost/fairness tradeoff is a non-monotone function of a parameter which the user must adjust, while [37] only focuses on the  $k$ -means objective and provides convergence guarantees only for a smoothed version of the original problem.

We focus on the formal tradeoff between fairness and the clustering objective. We note that the clustering objective can be replaced by the price of fairness (PoF). However, in our setting, the results are clearer if we refer to the clustering objective instead of the PoF. Ultimately a higher PoF corresponds to a weakly higher fair clustering objective and vice versa. In fact, they are multiples of one another:  $\text{PoF} = (\text{cost of fair solution}) / (\text{cost of agnostic solution})$ . This tradeoff between fairness and efficiency manifested the PoF concept, in operations research by [12] and simultaneously in computer science by [19], showing general approaches to defining and measuring it. Similar to our work, others have adapted PoF as a hard constraint in emergency response [30], organ allocation [39], and rideshare [36, 42], and a partial constraint in scarce resource allocation for kidney dialysis [29] and organ exchange [40]. We propose a framework for balancing PoF for a traditional “efficient” objective in clustering, which finds application in areas such as advertising, network analysis, and data summarization.

### 3 Preliminaries

In a clustering problem, we are given a set of points  $\mathcal{C}$  in a metric space. A distance function  $d(i, j)$  specifies the distance between each pair of points  $i, j \in \mathcal{C}$ . Furthermore,  $d$  is symmetric, non-negative, and satisfies the triangle inequality. Common clustering cost (loss) functions (e.g.,  $k$ -means or  $k$ -median) can be written as  $\min_{S:|S|\leq k, \phi} L_p^k(\mathcal{C}) = \min_{S:|S|\leq k, \phi} \left( \sum_{j \in \mathcal{C}} d^p(j, \phi(j)) \right)^{1/p}$  where  $k$  is the number of clusters,  $S$  is the set of cluster centers chosen from a candidate set of centers  $\mathcal{S}$ , and  $\phi: \mathcal{C} \rightarrow S$  is an assignment function that assigns points to cluster centers. The value  $p$  determines the type of clustering, i.e.,  $p = \infty, 1$ , and  $2$  for  $k$ -center,  $k$ -median, and  $k$ -means, respectively.

In fair clustering, each point has a color associated with it to indicate its group membership. Specifically, we have a function  $\chi: \mathcal{C} \rightarrow \mathcal{H}$  where  $\mathcal{H}$  is the set of possible colors. We denote the set of all points of color  $h$  by  $\mathcal{C}^h$ . The fair clustering problem (**FC**) [20, 10, 11, 2, 7, 31, 8] is to minimize the clustering objective while satisfying additional fairness constraints:

$$\min_{S:|S|\leq k, \phi} \left( \sum_{j \in \mathcal{C}} d^p(j, \phi(j)) \right)^{1/p} \quad (1a)$$

$$\text{s.t. } \forall i \in S, \forall h \in \mathcal{H} : \beta_h |\mathcal{C}_i| \leq |\mathcal{C}_i^h| \leq \alpha_h |\mathcal{C}_i| \quad (1b)$$

where  $\mathcal{C}_i$  is the set of points in cluster  $i$  and  $\mathcal{C}_i^h \subseteq \mathcal{C}_i$  is the subset of points in cluster  $i$  with color  $h$ .  $\beta_h$  and  $\alpha_h$  are pre-specified lower and upper proportion bounds for color  $h$ , respectively. Clearly,  $0 < \beta_h \leq \alpha_h < 1$ .

In “unfair” clustering problems, the assignment function  $\phi$  maps points to the nearest center in  $S$ , i.e.,  $\phi(j) = \text{argmin}_{i \in S} d(i, j)$  since this minimizes the objective. However, satisfying the added constraints in fair clustering may cause points to be assigned to clusters that are farther away. This motivates the fair assignment problem (**FA**), in which the set of centers  $S$  is given and the objective is to minimize the clustering cost subject to fairness constraints:

$$\min_{\phi} \left( \sum_{j \in \mathcal{C}} d^p(j, \phi(j)) \right)^{1/p} \quad (2a)$$

$$\text{s.t. } \forall i \in S, \forall h \in \mathcal{H} : \beta_h |\mathcal{C}_i| \leq |\mathcal{C}_i^h| \leq \alpha_h |\mathcal{C}_i| \quad (2b)$$

The only difference between the fair assignment (2) and fair clustering (1) problems is that  $S$  is not an optimization variable in the fair assignment problem.

### 4 Fair Clustering Under a Bounded Cost (FCBC)

The fundamental idea of fair clustering under a bounded cost (**FCBC**) is to minimize a measure of unfairness subject to an upper bound on the clustering cost:

$$\min \text{Unfairness} \quad (3a)$$

$$\text{s.t. Clustering Cost} \leq \text{Given upper bound} \quad (3b)$$

Next, we transform (3a) and (3b) above into a clear mathematical optimization problem.

**The Constraint (3b):** The clustering cost is  $\left( \sum_{j \in \mathcal{C}} d^p(j, \phi(j)) \right)^{1/p}$ . Let  $U$  denote the exogeneous upper bound on clustering cost. Then, (3b) becomes  $\left( \sum_{j \in \mathcal{C}} d^p(j, \phi(j)) \right)^{1/p} \leq U$ . Note that for the case of the  $k$ -center where  $p = \infty$ , the constraint reduces to a simpler form, specifically  $\forall j \in \mathcal{C}, d(j, \phi(j)) \leq U$ .

**The Objective (3a):** In prior work, a given clustering is considered fair if for each cluster, the proportions of each color lie within pre-specified lower and upper bounds, i.e.:  $\forall i \in S, \forall h \in \mathcal{H} : \beta_h |C_i| \leq |C_i^h| \leq \alpha_h |C_i|$ . However, bounding the clustering cost may make it impossible to have a fair feasible solution. Therefore, we instead set a measure of unfairness for each color and try to minimize this measure. Let  $\Delta_h$  denote the worst proportional violation across the clusters for a color  $h$ . Specifically, for a given clustering,  $\Delta_h \in [0, 1]$  is the minimum non-negative value such that:

$$\forall i \in S : (\beta_h - \Delta_h) |C_i| \leq |C_i^h| \leq (\alpha_h + \Delta_h) |C_i|. \quad (4)$$

Clearly, if  $\Delta_h = 0$ , then color  $h$  is within the desired proportion in every cluster. Having set  $\Delta_h$  to be a measure of the unfair treatment that group  $h$  receives, we are faced with the question of setting the fairness objective, for which there are many reasonable options. We consider two prominent and intuitive fairness objectives [14]:

$$\text{GROUP-UTILITARIAN} = \min \sum_{h \in \mathcal{H}} \Delta_h \quad , \quad \text{GROUP-EGALITARIAN} = \min \max_{h \in \mathcal{H}} \Delta_h$$

The GROUP-UTILITARIAN objective minimizes the sum of proportional violations for all of the colors, treating all points of a specific color as a single player in a game. The GROUP-EGALITARIAN objective minimizes the maximum proportional violation across the colors. We also consider a more generalized version of the GROUP-EGALITARIAN objective, the GROUP-LEXIMIN objective. Like GROUP-EGALITARIAN, the GROUP-LEXIMIN objective minimizes the maximum (worst) violation, but it goes further to minimize the second-worst violation, then the third-worst violation, and so on until no further improvement can be made. We now state the fair clustering under a bounded cost problem (FCBC):

$$\min_{S: |S| \leq k, \phi} \text{UNFAIRNESS-OBJECTIVE} \quad (5a)$$

$$\text{s.t.} \quad \left( \sum_{j \in \mathcal{C}} d^p(j, \phi(j)) \right)^{1/p} \leq U \quad (5b)$$

where the UNFAIRNESS-OBJECTIVE could equal GROUP-UTILITARIAN, GROUP-EGALITARIAN, or GROUP-LEXIMIN. Similar to the fair assignment FA problem (2), we may define the fair assignment under a bounded cost (FABC) problem as:

$$\min_{\phi} \text{UNFAIRNESS-OBJECTIVE} \quad (6a)$$

$$\text{s.t.} \quad \left( \sum_{j \in \mathcal{C}} d^p(j, \phi(j)) \right)^{1/p} \leq U \quad (6b)$$

where similarly the optimization is over the assignment function  $\phi$  while the set of centers  $S$  is fixed.

## 5 Hardness of FCBC & FABC

First, we formally state the hardness of the fair clustering FC and the fair assignment FA problems.

**Theorem 5.1.** *The fair clustering FC (1) and fair assignment FA (2) problems are NP-hard.*

We now establish the hardness of fair clustering under a bounded cost FCBC and fair assignment under a bounded cost FABC. We note that these hardness results follow for all objectives (GROUP-UTILITARIAN, GROUP-EGALITARIAN, and GROUP-LEXIMIN).

**Theorem 5.2.** *Fair clustering under a bounded cost FCBC and fair assignment under a bounded cost FABC are NP-hard.*

Although we have shown that both the fair clustering and fair assignment problems under a bounded cost are NP-hard, the reductions rely on setting the upper bound  $U$  to a small enough value, precisely that of the optimal fair clustering cost. It seems reasonable to expect both problems to transition into being polynomial time solvable if the upper bound becomes sufficiently large. We show in Section 8 that such a result is not easy to establish and would lead to a true approximation for fair clustering which is yet to be produced in the fair clustering literature for arbitrary metric spaces and arbitrary lower and upper color proportion bounds.

For a given clustering cost  $U$ , there are many clusterings (solutions) of cost not exceeding  $U$ . Let  $\mathcal{S}_U$  be the set of those solutions, i.e. if  $(S_t, \phi_t) \in \mathcal{S}_U$ , then  $(S_t, \phi_t)$  is a clustering with a cost that does not exceed  $U$ . Further, let  $L_t$  be the size of the smallest non-empty cluster<sup>1</sup> in the clustering  $(S_t, \phi_t)$ ,

<sup>1</sup>An empty cluster is a cluster with no points assigned to it. This could happen if for example the assignment function  $\phi$  does not map any point to a given center including the center itself.

then we define  $L(U)$  to be the size of the smallest cluster across all clusterings of cost not exceeding  $U$ , i.e.  $L(U) = \min_{(S_t, \phi_t) \in \mathcal{S}_U} L_t$ . Clearly, for  $U_1$  and  $U_2$  such that  $U_2 \geq U_1$ , then  $L(U_2) \leq L(U_1)$  since  $\mathcal{S}_{U_1} \subseteq \mathcal{S}_{U_2}$ . We can conclude the following fact from the definition of  $L(U)$ :

**Fact 5.1.** *For a given upper bound  $U$ , no clustering with cost less than or equal to  $U$  can have less than  $L(U)$  many points in a non-empty cluster.*

We show that the quantity  $L(U)$  plays a fundamental role. In fact, lower bounds on the additive approximation<sup>2</sup> for the proportional violations and fairness objectives are related to  $L(U)$  as shown in the following theorem:

**Theorem 5.3.** *For a given instance of the FCBC or FABC problem with an arbitrary upper bound  $U$ , unless  $P = NP$  no polynomial time algorithm can produce a solution with a cost not exceeding  $U$  that satisfies any of the following conditions: (a) The proportional violation of any color  $h \in \mathcal{H}$  is  $\Delta_h < \frac{1}{8L(U)}$ . (b) The additive approximation for the GROUP-UTILITARIAN objective is less than  $\frac{|\mathcal{H}|}{8L(U)}$ . (c) The additive approximation for the GROUP-EGALITARIAN objective is less than  $\frac{1}{8L(U)}$ .*

## 6 Algorithms for FCBC

Our main result for the FCBC problem is the following theorem which follows as a direct consequence of the guarantees of Theorems 6.2, 6.4, 6.5, 6.6, and 6.7:

**Theorem 6.1.** *For any clustering objective, given a bound  $U$  on the clustering cost, Algorithm 1 solves the fair clustering under a bounded cost FCBC problem at a cost of at most  $U' = (2 + \alpha)U$  where  $\alpha$  is the approximation ratio of the color-blind clustering algorithm. The additive approximation is  $|\mathcal{H}|(\epsilon + \frac{2}{L(U')})$  for the GROUP-UTILITARIAN objective and  $\epsilon + \frac{2}{L(U')}$  for the GROUP-EGALITARIAN objective.*

From the theorem above, it is clear that the additive approximation guarantees we have improve when the cost does not permit small clusters. Indeed, in the absence of outlier points and for reasonable values of  $k$ , small clusters are unlikely to exist. Further, empirically we verify the smallest cluster size and find that the smallest cluster size is 159 points (see Section 9.3). See Appendix C for more discussion.

We now provide our general algorithm for fair clustering under a bounded cost FCBC which we denote by  $\text{ALG-FCBC}(U, \text{UNFAIRNESS-OBJECTIVE})$  where we have made explicit reference to the dependence of  $\text{ALG-FCBC}$  on the given cost upper bound  $U$  and the desired UNFAIRNESS-OBJECTIVE which could either be the GROUP-UTILITARIAN, GROUP-EGALITARIAN, or GROUP-LEXIMIN objective.

$\text{ALG-FCBC}(U, \text{UNFAIRNESS-OBJECTIVE})$  (see Algorithm 1) involves two steps, in step **(1)**: we use a color-blind approximation algorithm to find the cluster centers  $S$ , in step **(2)**: we call the algorithm  $\text{ALG-FABC}(S, U', \text{UNFAIRNESS-OBJECTIVE})$  for the FABC problem. It should be noted that we have fed  $\text{ALG-FABC}$  the set of centers  $S$  from step **(1)**, further the cost upper bound for  $\text{ALG-FABC}$  is set to  $U' = (2 + \alpha)U$  while the UNFAIRNESS-OBJECTIVE remains unchanged. We further note that  $\text{ALG-FABC}$  will have the same clustering objective as  $\text{ALG-FCBC}$ , e.g. if  $\text{ALG-FCBC}$  is given the  $k$ -median objective so well  $\text{ALG-FABC}$ .

Clearly, from algorithm  $\text{ALG-FCBC}$  the FCBC problem is closely related to the FABC problem. In fact, we establish the following general theorem for all clustering objectives:  $k$ -center,  $k$ -median, and  $k$ -means that shows that an algorithm which solves the FABC problem with provable guarantees can be used to solve the FCBC problem with provable guarantees:

**Theorem 6.2.** *For any clustering objective and both the GROUP-UTILITARIAN and GROUP-EGALITARIAN objectives, given an algorithm that solves fair assignment under a bounded cost FABC with additive approximation  $\mu$ , the fair clustering under a bounded cost FCBC problem can be solved with an additive approximation of  $\mu$  and at a cost of at most  $(2 + \alpha)U$ , where  $\alpha$  is the approximation ratio of the color-blind clustering algorithm.*

<sup>2</sup>An algorithm for a minimization problem with additive approximation  $\mu > 0$ , returns a value for the objective that is at most  $\text{OPT} + \mu$  where  $\text{OPT}$  is the optimal value.

---

**Algorithm 1** :ALG-FCBC( $U$ , UNFAIRNESS-OBJECTIVE)

---

- 1: Choose a set of centers  $S$  by running a color-blind clustering algorithm of approximation ratio  $\alpha$ .
  - 2: Set  $U' = (2 + \alpha)U$  and call **ALG-FABC**( $S, U'$ , UNFAIRNESS-OBJECTIVE)
- 

### 6.1 Fair Assignment Under a Bounded Cost

Algorithm block 2 shows the steps of our algorithm **ALG-FABC** for the **FABC** objective. In step **(1)**: we search for the optimal proportional violations given the bound on the clustering cost  $U$  using LPs. Having found the near-optimal solution, in step **(2)**: we round the possibly fractional solution to a feasible integer solution using a network flow algorithm. We note that the details of the search done in step **(1)** depend on the objective, i.e., **GROUP-UTILITARIAN** or **GROUP-EGALITARIAN**.

---

**Algorithm 2** :ALG-FABC( $S, U$ , UNFAIRNESS-OBJECTIVE)

---

- 1: Given the UNFAIRNESS-OBJECTIVE, search for the optimal proportion violation values  $\Delta_h$  at a cost upper bound of  $U$  using the feasibility LPs of (7).
  - 2: Apply network flow rounding to the LP solution with the optimal value.
  - 3: **return** the set of centers  $S$  and the assignment function  $\phi$  (resulting from the rounded LP solution).
- 

We note that in fair assignment under a bounded cost **FABC** the set of centers  $S$  has already been chosen and the optimization is done only over the assignment  $\phi$  of points to centers. We let  $x_{ij}$  be a decision variable that equals 1 if point  $j$  is assigned to center  $i \in S$  and 0 otherwise. Note that the values of  $x_{ij}$  are a way to represent the assignment function  $\phi$ . Regardless of the objective that is being minimized, the following set of constraints must hold:

$$\sum_{i,j} d^p(i,j)x_{ij} \leq U^p \quad (7a)$$

$$\forall j \in \mathcal{C} : \sum_{i \in S} x_{ij} = 1, \quad x_{ij} \in [0, 1] \quad (7b)$$

$$\forall h \in \mathcal{H} : \Delta_h \in [0, 1] \quad (7c)$$

$$\forall h \in \mathcal{H}, \forall i \in S : (\beta_h - \Delta_h) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \leq \sum_{j \in \mathcal{C}^h} x_{ij} \leq (\alpha_h + \Delta_h) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \quad (7d)$$

For the  $k$ -center problem, the first constraint (7a) is replaced by  $\forall j \in \mathcal{C} : x_{ij} = 0$  if  $d(i, j) > U$ . Note that in the above we have  $x_{ij} \in [0, 1]$  which is a relaxation of  $x_{ij} \in \{0, 1\}$ , as the latter would result in an intractable mixed-integer program. With our variables being  $x_{ij}$  and  $\Delta_h$  it is reasonable to consider a convex optimization approach. That is, we could choose to minimize the objective **GROUP-UTILITARIAN** or the objective **GROUP-EGALITARIAN** with our set of constraints being (7). Looking at the form of the **GROUP-UTILITARIAN** and the **GROUP-EGALITARIAN** objectives, it is not difficult to see that they are linear (and therefore convex) in the parameters  $x_{ij}$  and  $\Delta_h$ , however as the following theorem shows, the constraint set (7) is not convex. In fact, either of the proportion bounds alone in constraint (7d) would lead to a non-convex set. The non-convexity of the constraint set implies that the resulting optimization problem would also be non-convex:

**Theorem 6.3.** *The constraint set (7) is not convex.*

Although the constraint set (7) is not convex, if we fix the values of  $\Delta_h$  then we clearly have a simple feasibility LP with variables  $x_{ij}$ . We therefore take an approach where for a given objective (**GROUP-UTILITARIAN** or **GROUP-EGALITARIAN**), we search for the corresponding optimal values of  $\Delta_h$  by running the feasibility LP of (7). Note that with a given set of values for  $\Delta_h$ , we can obtain the corresponding value for the **GROUP-UTILITARIAN** or **GROUP-EGALITARIAN** objectives and therefore the LP does not need an objective: a feasibility check suffices. Further, since we only use non-trivial values for  $\Delta_h \in [0, 1]$ , constraint (7c) can be omitted. Sections 6.1.1 and 6.1.2 detail how we use the feasibility LPs of (7) to obtain LP solutions that are approximately optimal (having bounded additive approximation from the optimal) for the **GROUP-UTILITARIAN** and **GROUP-EGALITARIAN** objectives, respectively. Since these resulting LP solutions could contain fractional values, i.e., it is possible to have a value  $x_{ij} \notin \{0, 1\}$ , the approximately optimal LP solution would have to be rounded to an integral solution. This rounding further degrades the approximation, but we show that this degradation is not large and can be bounded. The details of the rounding are shown

in Section 6.1.3. The search algorithms of Sections 6.1.1 and 6.1.2, followed by the rounding of Section 6.1.3, lead to an algorithm for **FABC**.

### 6.1.1 Search Algorithm for the GROUP-UTILITARIAN Objective

We are searching for the optimal proportional violations  $\Delta_h^* \in [0, 1]$  for the GROUP-UTILITARIAN objective. The first step we take is to discretize the space by a parameter  $\epsilon \in (0, 1)$ . For convenience, we set  $\epsilon = \frac{1}{r}$  where  $r \in \mathbb{Z}^+$ , i.e.,  $r$  is a positive integer. Accordingly, instead of interacting with the continuous interval  $[0, 1]$  for the proportional violations, we instead interact with  $E_\epsilon = \{\epsilon, 2\epsilon, \dots, (\frac{1}{\epsilon} - 1)\epsilon, 1\}$ , with  $|E_\epsilon| = \frac{1}{\epsilon}$ . Therefore, we have a set of  $(\frac{1}{\epsilon})^{|\mathcal{H}|}$  many possibilities for the proportional violations and we can obtain the optimal solution for the GROUP-UTILITARIAN objective through exhaustive search by checking the feasibility of LP (7) and picking the feasible combination of proportional violations which leads to the smallest value for the GROUP-UTILITARIAN objective, i.e.,  $\text{GROUP-UTILITARIAN} = \sum_{h \in \mathcal{H}} \Delta_h$ .

The above approach would take  $O((\frac{1}{\epsilon})^{|\mathcal{H}|})$  many LP runs. Therefore, we show a faster search that tries instead  $O((\frac{1}{\epsilon})^{|\mathcal{H}|-1})$ . The key to this speed up comes from the fact that for the two-color case, we only need to evaluate  $O(\frac{1}{\epsilon})$  many possibilities.

**Theorem 6.4.** *For FABC with GROUP-UTILITARIAN objective, we can use  $O((\frac{1}{\epsilon})^{|\mathcal{H}|-1})$ -many LP runs to obtain an LP solution with additive approximation  $|\mathcal{H}|\epsilon$ .*

Furthermore, for the important two-color case with *symmetric* upper and lower bounds we show a search algorithm that requires only  $O(\log \frac{1}{\epsilon})$  LP runs. The two color case with *symmetric* upper and lower bounds is that where the two colors  $h_1$  and  $h_2$  are present with proportions  $r_1$  and  $r_2$  in the dataset, and the proportion bounds are set to  $\alpha_i = r_i + \lambda_i, \beta_i = r_i - \lambda_i$  for  $i \in \{1, 2\}$  and some valid  $\lambda_1, \lambda_2 \in [0, 1]$ . The key observation for the two-color symmetric case is that the proportion of one color implies the proportion of the other; hence, we can run binary search over the set  $E_\epsilon$ .

**Theorem 6.5.** *For FABC with two colors, symmetric lower & upper bounds, and the GROUP-UTILITARIAN objective, we can use  $O(\log(\frac{1}{\epsilon}))$ -many LP runs to get a solution with an additive approximation of  $|\mathcal{H}|\epsilon = 2\epsilon$ .*

### 6.1.2 Search Algorithm for GROUP-EGALITARIAN and GROUP-LEXIMIN Objectives

For the GROUP-EGALITARIAN objective we follow the same discretization step as for the GROUP-UTILITARIAN objective. For all colors, their violation  $\Delta_h$  is set to the same value and the optimal solution is found simply by doing binary search over the set  $E_\epsilon$  by running the feasibility LP (7).

**Theorem 6.6.** *For FABC with the GROUP-EGALITARIAN objective, we can use  $O(\log(\frac{1}{\epsilon}))$ -many LP runs to get a solution with an additive approximation of  $\epsilon$ .*

We provide a heuristic algorithm for the GROUP-LEXIMIN objective; a rough sketch follows. In the first step, it obtains the GROUP-EGALITARIAN solution. Then, it proceeds by finding a color that cannot improve beyond the current optimal violation; if more than one color is found, then one of these colors is randomly picked. The algorithm then looks for the optimal violation for the remaining colors, having the violations of the previous colors fixed. These steps are followed until no color can have its proportional violation improved. See Appendix B for the full details.

### 6.1.3 The Rounding Scheme and ALG-FABC Guarantees

Having obtained the optimal LP solutions for either the GROUP-UTILITARIAN or GROUP-EGALITARIAN objectives, we now round the solutions to integral values at a bounded increase to the additive approximation. To do the rounding, we apply the network flow method of [11] (see Appendix E for details), although other rounding methods are applicable. Given the LP solution  $x_{ij}$  and its associated proportional violations  $\Delta_h$ , if we denote the rounded integral solution by  $\bar{x}_{ij}$  and  $\bar{\Delta}_h$ , then network-flow rounding guarantees the following: **(i)**  $\sum_{i,j} d^p(i,j)\bar{x}_{ij} \leq \sum_{i,j} d^p(i,j)x_{ij}$ . **(ii)**  $\forall i \in [k] : \left\lfloor \sum_{j \in \mathcal{C}} x_{ij} \right\rfloor \leq \sum_{j \in \mathcal{C}} \bar{x}_{ij} \leq \left\lceil \sum_{j \in \mathcal{C}} x_{ij} \right\rceil$ . **(iii)**  $\forall h \in \mathcal{H}, \forall i \in [k] : \left\lfloor \sum_{j \in \mathcal{C}^h} x_{ij} \right\rfloor \leq \sum_{j \in \mathcal{C}^h} \bar{x}_{ij} \leq \left\lceil \sum_{j \in \mathcal{C}^h} x_{ij} \right\rceil$ .

Property (i) ensures that the clustering objective will not increase beyond the LP value, and thus, provided the LP cost does not exceed the upper bound on the cost  $U$ , the cost of the rounded assignment will not exceed  $U$  as well. Property (ii) guarantees that the total number of points assigned to a cluster will not vary by more than 1 point. Property (iii) guarantees that the total number of points of a given color assigned to a cluster will not vary by more than 1 point. We can use the above properties along with the lower bound on the size of any cluster  $L(U)$  to establish the following theorem:

**Theorem 6.7.** *For the **FABC** problem, the rounded solution has cost of at most  $U$  and an additive approximation of: (1)  $|\mathcal{H}|(\epsilon + \frac{2}{L(U)})$  for the **GROUP-UTILITARIAN** objective and (2)  $\epsilon + \frac{2}{L(U)}$  for the **GROUP-EGALITARIAN** objective.*

Recalling the additive approximation lower bounds of Theorem 5.3 for the **FABC** problem, we see that we obtain a solution for **FABC** of cost at most  $U$  with near-optimal additive approximation. Specifically, our additive approximations for the **GROUP-UTILITARIAN** and **GROUP-EGALITARIAN** are  $\frac{2|\mathcal{H}|}{L(U)}$  and  $\frac{2}{L(U)}$  compared to their lower bounds of  $\frac{|\mathcal{H}|}{8L(U)}$  and  $\frac{1}{8L(U)}$ , respectively.

**A randomized extension.** We also briefly mention a randomized rounding algorithm's guarantees; the description of this algorithm (which is motivated by an approach of [35]) is detailed in Appendix G. This algorithm efficiently constructs a random vector  $\bar{X}$  with entries  $\bar{X}_{i,j}$  in  $\{0, 1\}$  such that: (a) Properties (ii) and (iii) hold with probability 1, and (b) the expected value  $E[\bar{X}_{i,j}]$  equals  $x_{i,j}$  for all  $(i, j)$ . This has three consequences: (b1) the fairness guarantee for each cluster and color become *better in expectation*: for all  $h \in \mathcal{H}$  and for all  $i \in S$ :  $E[\sum_{j \in \mathcal{C}^h} \bar{X}_{ij}]$  indeed lies between  $(\beta_h - \Delta_h) \left( \sum_{j \in \mathcal{C}} x_{ij} \right)$  and  $(\alpha_h + \Delta_h) \left( \sum_{j \in \mathcal{C}} x_{ij} \right)$ . (b2) The expected value of the objective function (the left-hand side in (i)) is at most the right-hand side of (i). (b3) Even if we had multiple linear objective functions, they will all be preserved in expectation.

## 7 Fairness Across the Clusters is not Possible

It is tempting to modify both the **GROUP-UTILITARIAN** and **GROUP-EGALITARIAN** (or **GROUP-LEXIMIN**) objectives to sum across the clusters instead of taking the maximum violation across the clusters. More specifically, we can replace the objectives by the following: **GROUP-UTILITARIAN-SUM**, which equals  $\sum_{h \in \mathcal{H}, i \in [k]} \Delta_h^i$ , and **GROUP-EGALITARIAN-SUM**, which

equals  $\min_{h \in \mathcal{H}} \max_{i \in [k]} \sum \Delta_h^i$ , where  $\Delta_h^i$  is the violation of color  $h$  in cluster  $i$ ; clearly the previously-considered violation  $\Delta_h$  is  $\max_{i \in [k]} \Delta_h^i$ . It can be seen that such an objective is more flexible. For

example, the maximum violations might occur in a cluster that cannot be improved within the given bound on the clustering cost, while it may be possible to improve it for other clusters. The original **GROUP-UTILITARIAN** and **GROUP-EGALITARIAN** objectives may bring no improvement in such a situation but their above modifications could. We prove a negative result. Specifically, while we were able to approximate **FABC** by small additive values for the original objectives (Theorem 6.7), for the new objectives we cannot efficiently approximate the **FABC** problems within even relatively-large additive approximations:

**Theorem 7.1.** *For **FABC**, the objectives **GROUP-UTILITARIAN-SUM** and **GROUP-EGALITARIAN-SUM** that sum across the clusters cannot be approximated in polynomial time to within an additive approximation of  $O(n^\delta)$  where  $\delta$  is a constant in  $[0, 1)$ , unless  $P = NP$ .*

## 8 Solving the Problem Optimally for a Large-Enough Cost

It seems reasonable to assume that when the upper bound on the cost  $U$  is large enough, the problem becomes solvable in polynomial time. It is not difficult to devise such guarantees for some special cases. However, in the theorem below we show that an algorithm with such a guarantee would imply a true approximation<sup>3</sup> for fair clustering. Since fair clustering has remained resistant to a true polynomial-time approximation for general metric spaces and arbitrary lower- and upper- proportion bounds [13, 8], this suggests that the problem is indeed nontrivial. Furthermore, we also show the converse, i.e., a true approximation algorithm for fair clustering would imply an exact algorithm for fair clustering under a bounded cost **FCBC**.

<sup>3</sup>A true approximation algorithm yields a solution that approximates the optimal objective value with no constraint violation, in contrast to bi-criteria algorithms which have a bounded violation in the constraints.

**Theorem 8.1.** *Suppose that there is a polynomial time algorithm which can obtain the optimal solution for FCBC for the upper bound of  $U$  if  $U \geq \alpha(\mathcal{I}) \text{OPT}_{cb}(\mathcal{I})$  where  $\mathcal{I}$  is a specific instance of FCBC and  $\text{OPT}_{cb}(\mathcal{I})$  is the optimal cost of its color-blind clustering. Then we have a true polynomial time approximation algorithm for fair clustering. Further, a true polynomial time  $\alpha'(\mathcal{I})$ -approximation algorithm for fair clustering implies that FCBC can be solved optimally in polynomial time for  $U \geq \alpha'(\mathcal{I}) \text{OPT}_{FC}(\mathcal{I})$ .*

## 9 Experiments

We validate our algorithms on datasets from the UCI repository [24]. The results here are for  $k$ -means clustering; additional experiments are in Appendix F.

**Hardware, Software, and Algorithms:** We only use commodity hardware for all experiments with our programs running on Python 3.6. For the color-blind  $k$ -means clustering, we use the  $k$ -means++ algorithm [5] which has an approximation ratio of  $O(\log k)$ . Our LPs are solved using CPLEX [32]. Scikit-learn [46] is called for subroutines such as  $k$ -means++. The network-flow rounding is handled using NetworkX [25].

**Datasets:** We use all 32,561 entries of the **Adult** dataset [34]. For the **Census1990** dataset [41], because of its large size (over 2 million points) we sub-sample the dataset to a range similar to that considered in the fair clustering literature [20, 10]; specifically we use 20,000 data points. We also use the **CreditCard** dataset [47] which has 30,000 points (results are in Appendix F). For all datasets we use the numeric attributes to assign the coordinates in the space and the distance between any two points is set to the Euclidean distance.

**Setting and Measurements:** Each color  $h \in \mathcal{H}$  has proportion  $r_h$ , i.e.,  $r_h = \frac{|c^h|}{|C|}$ . We set the upper and lower bounds for each color to  $\alpha_h = (1 + \delta)r_h$  and  $\beta_h = (1 - \delta)r_h$ . This means that each cluster should have each color with the same proportion as in the population with a possible deviation of  $\delta$ .

We first do color-blind clustering using the  $k$ -means++ algorithm. The clustering cost we obtain from the  $k$ -means++ is a proxy of the lowest possible value of the clustering cost (since the hardness of clustering forbids the calculation of the true optimal value). We gradually increase the upper bound cost from the color-blind cost to higher values and for each choice of the upper bound, we minimize either the GROUP-UTILITARIAN or GROUP-LEXIMIN objectives using our algorithms and record the objective value. For better interpretation, instead of showing the value of the upper bound, we show its ratio to the color-blind clustering, which is the PoF. Further, for all experiments we discretize the space by  $\epsilon = \frac{1}{2^7} < 0.008$ .

### 9.1 GROUP-UTILITARIAN Experiments

We use the **Adult** and **Census1990** datasets with self-reported gender (male or female) as the attribute. We note that both datasets explicitly use categorical labels for this socially-complex concept, and acknowledge that this is reductive [17]. Figure 2 shows the PoF versus the achieved GROUP-UTILITARIAN objective, with  $\delta = 0.1$ . As expected, as the price of fairness increases (higher bound on the cost), we can further minimize the proportional violations. Eventually the GROUP-UTILITARIAN objective becomes less than 0.1 and even very close to zero. We also observe that at a given cost upper bound, we can achieve lower values for the GROUP-UTILITARIAN objective when the number of clusters ( $k$ ) is lower.

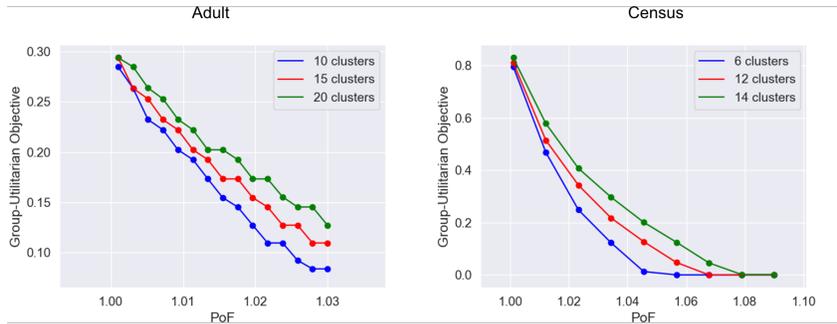


Figure 2: PoF vs the GROUP-UTILITARIAN objective for the **Adult** and **Census1990** datasets.

## 9.2 GROUP-EGALITARIAN and GROUP-LEXIMIN Experiments

We again use the **Adult** and **Census1990** datasets. However, for **Adult**, we set the fairness attribute to race which—in this dataset, and with the same inherent social caveats as the categorization of gender—has 5 groups (colors). For **Census1990**, we set the fairness attribute to age where we have three age groups.<sup>4</sup> We set  $\delta = 0.05$  and  $k = 10$  for **Adult** and  $\delta = 0.1$  and  $k = 5$  for **Census1990**. Figure 3 shows the results of our algorithm. We notice that for some colors smaller violations are harder to achieve and we need to set the maximum allowable clustering cost to larger values to reduce their violations.

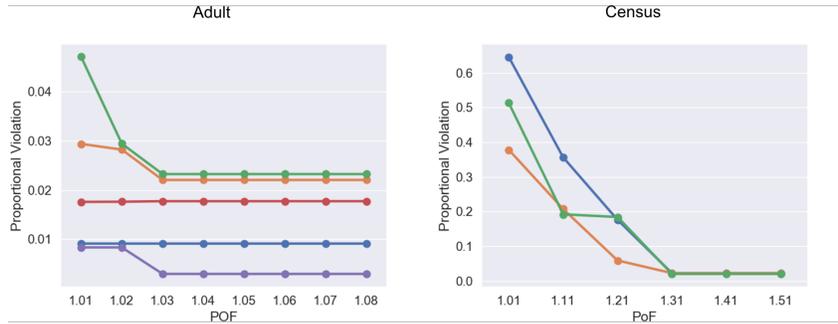


Figure 3: PoF versus the proportional violation for different groups (each colored graph is a group) in the **Adult** and **Census1990** datasets.

## 9.3 Checking the Size of the Smallest Cluster

As mentioned in Section 5 and Theorem 6.1 our approximations are dependent on the size of the smallest cluster in the solution. While it is not tractable to obtain the value of  $L(U)$  for a given  $U$ , we can still empirically check the size of the smallest cluster in the cost bounded clusterings we obtain. We note that, throughout, we do not impose any lower bound on the cluster size in our algorithm. For the above experiments we considered, we find that the minimum cluster size (across all choices of  $k$ ) are as follows: **Adult** (159 points), **Census1990** (171 points). The fact that the size of the smallest cluster is large means that we are achieving small (accurate) additive approximations with near-optimal objective values and when we obtain a large objective value it is because of how stringent the cost upper bound is.

**Acknowledgments.** Seyed Esmaeili and John Dickerson were supported in part by NSF CAREER Award IIS-1846237, NSF D-ISN Award #2039862, NSF Award CCF-1852352, NIH R01 Award NLM-013039-01, NIST MSE Award #20126334, DARPA GARD #HR00112020007, DoD WHS Award #HQ003420F0035, and a Google Faculty Research award. Aravind Srinivasan was supported in part by NSF awards CCF-1749864 and CCF-1918749, as well as research awards from Adobe, Amazon, and Google. We thank the anonymous reviewers for helpful feedback.

## References

- [1] S. Ahmadian, A. Epasto, M. Knittel, R. Kumar, M. Mahdian, B. Moseley, P. Pham, S. Vassilvitskii, and Y. Wang. Fair hierarchical clustering. *arXiv preprint arXiv:2006.10221*, 2020.
- [2] S. Ahmadian, A. Epasto, R. Kumar, and M. Mahdian. Clustering without over-representation. In *International Conference on Knowledge Discovery & Data Mining (KDD)*, 2019.
- [3] S. Ahmadian, A. Epasto, R. Kumar, and M. Mahdian. Fair correlation clustering. In *International Conference on Artificial Intelligence and Statistics*, pages 4195–4205. PMLR, 2020.
- [4] S. Ahmadian, A. Norouzi-Fard, O. Svensson, and J. Ward. Better guarantees for k-means and euclidean k-median by primal-dual algorithms. *SIAM Journal on Computing*, (0):FOCS17–97, 2019.

<sup>4</sup>**Census1990** actually has 8 age groups. For better interpretability of the results, we merge nearby groups  $\{0, 1, 2\}$ ,  $\{3, 4, 5\}$ , and  $\{7, 8\}$  to form 3 groups.

- [5] D. Arthur and S. Vassilvitskii. k-means++: The advantages of careful seeding. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1027–1035, 2007.
- [6] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristics for k-median and facility location problems. *SIAM Journal on computing*, 33(3):544–562, 2004.
- [7] A. Backurs, P. Indyk, K. Onak, B. Schieber, A. Vakilian, and T. Wagner. Scalable fair clustering. In *International Conference on Machine Learning (ICML)*, pages 405–413, 2019.
- [8] S. Bandyapadhyay, F. V. Fomin, and K. Simonov. On coresets for fair clustering in metric and euclidean spaces and their applications. *arXiv preprint arXiv:2007.10137*, 2020.
- [9] S. Barocas, M. Hardt, and A. Narayanan. *Fairness and Machine Learning*. fairmlbook.org, 2019. <http://www.fairmlbook.org>.
- [10] S. K. Bera, D. Chakrabarty, and M. Negahbani. Fair algorithms for clustering. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [11] I. O. Bercea, M. Groß, S. Khuller, A. Kumar, C. Rösner, D. R. Schmidt, and M. Schmidt. On the cost of essentially fair clusterings. In *International Conference on Approximation Algorithms for Combinatorial Optimization Problems (APPROX-RANDOM)*, 2019.
- [12] D. Bertsimas, V. F. Farias, and N. Trichakis. The price of fairness. *Operations Research*, 59(1):17–31, 2011.
- [13] M. Böhm, A. Fazzone, S. Leonardi, and C. Schwiegelshohn. Fair clustering with multiple colors. *arXiv preprint arXiv:2002.07892*, 2020.
- [14] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia. *Handbook of computational social choice*. Cambridge University Press, 2016.
- [15] B. Brubach, D. Chakrabarti, J. P. Dickerson, S. Khuller, A. Srinivasan, and L. Tsepenekas. A pairwise fair and community preserving approach to  $k$ -center clustering. In *International Conference on Machine Learning (ICML)*, 2020.
- [16] B. Brubach, D. Chakrabarti, J. P. Dickerson, A. Srinivasan, and L. Tsepenekas. Fairness, semi-supervised learning, and more: A general framework for clustering with stochastic pairwise constraints. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2021.
- [17] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*, pages 77–91, 2018.
- [18] J. Byrka, T. Pensyl, B. Rybicki, A. Srinivasan, and K. Trinh. An improved approximation for k-median, and positive correlation in budgeted optimization. In *Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 737–756. SIAM, 2014.
- [19] I. Caragiannis, C. Kaklamanis, P. Kanellopoulos, and M. Kyropoulou. The efficiency of fair division. International Workshop on Internet and Network Economics (WINE), 2009.
- [20] F. Chierichetti, R. Kumar, S. Lattanzi, and S. Vassilvitskii. Fair clustering through fairlets. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5029–5037, 2017.
- [21] J. P. Dickerson, A. D. Procaccia, and T. Sandholm. Price of fairness in kidney exchange. In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1013–1020, 2014.
- [22] S. A. Esmaili, B. Brubach, L. Tsepenekas, and J. P. Dickerson. Probabilistic fair clustering. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- [23] T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 1985.

- [24] G. Gunduz and E. Fokoue. Uci machine learning repository. *Irvine, CA: University of California, School of Information and Computer Science*, 20, 2013.
- [25] A. Hagberg, D. Schult, P. Swart, D. Conway, L. Séguin-Charbonneau, C. Ellison, B. Edwards, and J. Torrents. Networkx. high productivity software for complex networks. <https://networkx.github.io/>, 2013.
- [26] D. S. Hochbaum and D. B. Shmoys. A best possible heuristic for the  $k$ -center problem. *Math. Oper. Res.*, May 1985.
- [27] D. S. Hochbaum and D. B. Shmoys. A unified approach to approximation algorithms for bottleneck problems. *J. ACM*, May 1986.
- [28] K. Holstein, J. Wortman Vaughan, H. Daumé III, M. Dudik, and H. Wallach. Improving fairness in machine learning systems: What do industry practitioners need? In *ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1–16, 2019.
- [29] J. N. Hooker and H. P. Williams. Combining equity and utilitarianism in a mathematical programming model. *Management Science*, 58(9):1682–1693, 2012.
- [30] K. Huang, Y. Jiang, Y. Yuan, and L. Zhao. Modeling multiple humanitarian objectives in emergency response to large-scale disasters. *Transportation Research Part E: Logistics and Transportation Review*, 75:1–17, 2015.
- [31] L. Huang, S. Jiang, and N. Vishnoi. Coresets for clustering with fairness constraints. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 7587–7598. Curran Associates, Inc., 2019.
- [32] IBM. IBM CPLEX user’s manual, version 12 release 7. *IBM ILOG CPLEX Optimization*, 2016.
- [33] M. Kleindessner, S. Samadi, P. Awasthi, and J. Morgenstern. Guarantees for spectral clustering with fairness constraints. In *International Conference on Machine Learning*, pages 3458–3467. PMLR, 2019.
- [34] R. Kohavi. Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In *Kdd*, volume 96, pages 202–207, 1996.
- [35] V. S. A. Kumar, M. V. Marathe, S. Parthasarathy, and A. Srinivasan. A unified approach to scheduling on unrelated parallel machines. *J. ACM*, 56(5):28:1–28:31, 2009.
- [36] N. S. Lesmana, X. Zhang, and X. Bei. Balancing efficiency and fairness in on-demand ridesourcing. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5310–5320, 2019.
- [37] S. Liu and L. N. Vicente. A stochastic alternating balance  $k$ -means algorithm for fair clustering. *arXiv preprint arXiv:2105.14172*, 2021.
- [38] S. Mahabadi and A. Vakilian. (individual) fairness for  $k$ -clustering. In *International Conference on Machine Learning (ICML)*, 2020.
- [39] N. Mattei, A. Saffidine, and T. Walsh. Fairness in deceased organ matching. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, pages 236–242, 2018.
- [40] D. McElfresh and J. P. Dickerson. Balancing lexicographic fairness and a utilitarian objective with application to kidney exchange. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2018.
- [41] C. Meek, B. Thiesson, and D. Heckerman. The learning-curve sampling method applied to model-based clustering. *Journal of Machine Learning Research*, 2(Feb):397–418, 2002.
- [42] V. Nanda, P. Xu, K. A. Sankararaman, J. P. Dickerson, and A. Srinivasan. Balancing the tradeoff between profit and fairness in rideshare platforms during high-demand hours. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2020.

- [43] D. Saha, C. Schumann, D. C. McElfresh, J. P. Dickerson, M. L. Mazurek, and M. C. Tschantz. Measuring non-expert comprehension of machine learning fairness metrics. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2020.
- [44] Supreme Court of the United States. 13-1371 – Texas Department of Housing and Community Affairs v. The Inclusive Communities Project, Inc., Jan. 2015.
- [45] United States Senate. S. 1745 – 102nd Congress: Civil Rights Act of 1991, 1991. <https://www.govtrack.us/congress/bills/102/s1745>.
- [46] G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, et al. Scikit-learn: Machine learning in python. *J Mach Learn Res*, 12:2825–2830, 2011.
- [47] I.-C. Yeh and C.-h. Lien. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Systems with Applications*, 36(2):2473–2480, 2009.
- [48] I. M. Ziko, E. Granger, J. Yuan, and I. B. Ayed. Variational fair clustering. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

## A Omitted Proofs

In this section, we provide proofs for theorems and lemmas in the main paper. We recall Theorem 5.1:

**Theorem 5.1.** *The fair clustering FC (1) and fair assignment FA (2) problems are NP-hard.*

*Proof.* Since fair clustering problems, i.e. fair  $k$ -center, median, or means) generalize their NP-hard classical counterparts, i.e. the  $k$ -center, median, or means) clustering, it follows that fair clustering problems are also NP-hard.

The hardness of the fair assignment problem was established by [11] for  $k$ -center clustering. Here we show that fair assignment is NP-hard for  $k$ -median and  $k$ -means clustering as well.

First, following Section 4 of [11], the reduction is from the Exact Cover by 3-Sets (X3C). In Exact Cover by 3-Sets, we have a universal set of elements  $\mathcal{U}$  with  $|\mathcal{U}| = 3r$  where  $r$  is a positive integer and a set  $\mathcal{F}$  whose elements are subsets of  $\mathcal{U}$ . The problem is to decide if there exists a set  $\mathcal{F}'$  such that  $\mathcal{F}' \subseteq \mathcal{F}$  and each element in  $\mathcal{U}$  is included exactly once in one set in  $\mathcal{F}'$ .

The reduction is done by creating the following graph (see Figure 4 for an example). In the lowest level we have the elements  $e$  of the set  $\mathcal{U}$  each represented with a blue vertex. In the higher level we have the sets in  $\mathcal{F}$  each represented with a blue vertex. We draw edges between vertices in  $e \in \mathcal{U}$  and vertices in  $F \in \mathcal{F}$  if and only if the element  $e \in F$ . For set  $F$  in  $\mathcal{F}$  we add 3 auxiliary blue vertices which are connected to it through an edge. Finally, we add a set  $\mathcal{T}$  of red vertices where  $|\mathcal{T}| = \frac{|\mathcal{U}|}{3} = r$  in the highest level where each of those vertices is connected through an edge to every vertex in the set  $\mathcal{F}$ .

The distance function puts a cost of zero if the distance is between identical vertices and a cost of one between vertices connected through an edge. For vertices with no edges between them, the distance is the minimum distance found according to this graph by calculating the minimum cost path. This means that the distance between the blue auxiliary vertices and a center which is not their parent center is 3 (the path from the vertex to the associated center to an element in  $\mathcal{T}$ , then the specified center).

In fair assignment, the set of centers is already chosen. We choose the set of centers to be the elements of  $\mathcal{F}$ . Therefore, the number of centers  $k = r$ . Further, it is clear that this is a two color problem, we set the lower and upper bounds for the red color to  $\beta_{\text{red}} = \alpha_{\text{red}} = \frac{1}{4}$ . It follows that  $\beta_{\text{blue}} = \alpha_{\text{blue}} = \frac{3}{4}$ , i.e. the ratio of red to blue vertices is 1 : 3.

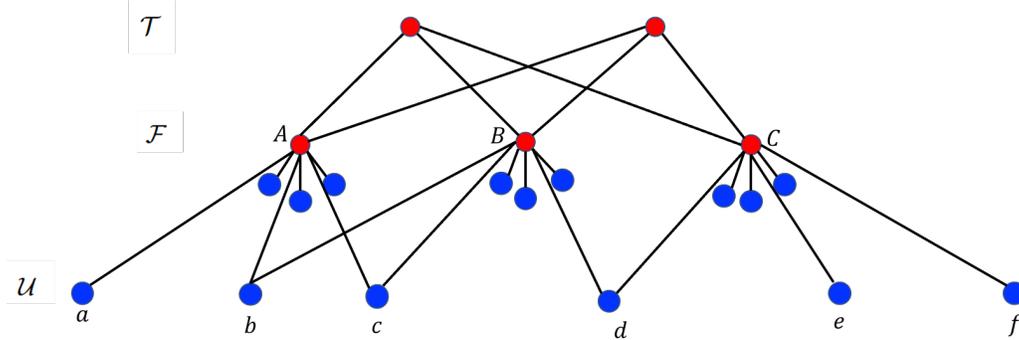


Figure 4: Figure follows the example of [11]. We show the fair assignment resulting graph, from the given Exact Cover by 3-Sets example where we have  $\mathcal{U} = \{a, b, c, d, e, f\}$  and  $\mathcal{F} = \{A = \{a, b, c\}, B = \{b, c, d\}, C = \{d, e, f\}\}$ .

We note the following lemma:

**Lemma A.1.** *Given the constructed graph with the set of centers being  $\mathcal{F}$ , the minimum clustering cost is lower bounded by 1 for the  $k$ -center problem and  $n - k$  for the  $k$ -median and  $k$ -means.*

*Proof.* First we note the following fact:

**Fact A.1.**  $\forall u, v \in G$  where  $u$  and  $v$  are distinct, we have that  $d(u, u) = d(v, v) = 0$  and  $d(u, v) \geq 1$  if  $u \neq v$ .

**$k$ -center:** Since the number of points is greater than the number of centers it follows that there exists a point  $u$  which will be assigned to another vertex  $v$  and therefore  $d(u, v) \geq 1$ .

**$k$ -median and  $k$ -means:** Denoting the assignment function (assigning vertices to centers) by  $\phi$ , the set of centers by  $S$ , and the integer  $p$  where  $p = 1$  for the  $k$ -median and  $p = 2$  for the  $k$ -means, we have that:

$$\begin{aligned}
\sum_{v \in G} d^p(v, \phi(v)) &= \sum_{v \in S} d^p(v, \phi(v)) + \sum_{v \in G-S} d^p(v, \phi(v)) \\
&\geq 0 + \sum_{v \in G-S} d^p(v, \phi(v)) \\
&\geq 0 + \sum_{v \in G-S} 1^p \\
&\geq \sum_{v \in G-S} 1 \\
&= n - k
\end{aligned}$$

where the above follows from Fact A.1. □

Therefore, we have:

**Lemma A.2.** *If there exists an exact cover, then the fair assignment problem can have a 1 : 3 red to blue vertex ratio and at a cost of 1 for the  $k$ -center and a cost of  $n - k$  for the  $k$ -median and  $k$ -means.*

*Proof.* We translate the exact cover by 3-sets solution to the constructed graph. Each chosen set in exact cover  $\mathcal{F}'$  will have the 3 corresponding elements from  $\mathcal{U}$  assigned to its center, along with its 3 auxiliary vertices and 1 vertex from  $\mathcal{T}$ . If the set was not chosen in the exact cover, then it will have only its 3 auxiliary vertices assigned to it.

This clearly matches the lower bound on the cost function from lemma (A.1) for each clustering objective. Further, it is also clear that the 1 : 3 red to blue color ratio is preserved in each cluster. □

**Lemma A.3.** *If there exists a fair assignment solution with 1 : 3 red to blue proportion and whose cost is 1 for the  $k$ -center and  $(n - k)$  for the  $k$ -median and  $k$ -means, there exists a solution to the exact cover by 3-sets problem.*

*Proof.* The costs of 1 and  $(n - k)$  for the  $k$ -center and  $k$ -median/mean respectively can only be achieved by assigning elements  $e \in \mathcal{U}$  to a center that they have an edge between. Similarly, all of the blue auxiliary vertices have to be assigned to their parent. Further to achieve the 1 : 3 red to blue ratio, a center will either choose 3 elements from  $\mathcal{U}$  and therefore has to choose an element from  $\mathcal{T}$  to satisfy the proportion. Or a center will not choose any element from  $\mathcal{U}$  and in that case it would not need to pick an element from  $\mathcal{T}$  to satisfy the proportion. □

□

Here, we recall Theorem 5.2:

**Theorem 5.2.** *Fair clustering under a bounded cost **FCBC** and fair assignment under a bounded cost **FABC** are NP-hard.*

*Proof.* The hardness of fair clustering under a bounded cost **FCBC** simply follows by setting the upper bound to  $U = \text{OPT}_{\text{FC}}$  where  $\text{OPT}_{\text{FC}}$  is the optimal value of fair clustering **FC**. An optimal solution to fair clustering would achieve the optimal value of 0 for all possible fairness objectives of **FCBC** and would have a cost  $\text{OPT}_{\text{FC}} \leq U$ .

Conversely, an optimal solution for **FCBC** would have a proportional violation of zero for all colors (therefore it is fair). Moreover, its cost would not exceed  $U = \text{OPT}_{\text{FC}}$ . Therefore, it is an optimal solution for fair clustering.

By the above, a solution is optimal for a fair clustering if and only if it is an optimal solution to the corresponding **FCBC** instance with  $U = \text{OPT}_{\text{FC}}$ . It follows that since fair clustering is NP-hard from theorem (5.1), that fair clustering under a bounded cost **FCBC** is also NP-hard.

In a similar manner, by setting  $U = \text{OPT}_{\text{FA}}$  the hardness of fair assignment under a bounded cost **FABC** can be established from the hardness of fair assignment.  $\square$

Here, we recall Theorem 5.3:

**Theorem 5.3.** *For a given instance of the **FCBC** or **FABC** problem with an arbitrary upper bound  $U$ , unless  $P = NP$  no polynomial time algorithm can produce a solution with a cost not exceeding  $U$  that satisfies any of the following conditions: (a) The proportional violation of any color  $h \in \mathcal{H}$  is  $\Delta_h < \frac{1}{8L(U)}$ . (b) The additive approximation for the **GROUP-UTILITARIAN** objective is less than  $\frac{|\mathcal{H}|}{8L(U)}$ . (c) The additive approximation for the **GROUP-EGALITARIAN** objective is less than  $\frac{1}{8L(U)}$ .*

*Proof.* We note that our derivation uses the reduction from X3C shown in the proof of theorem (5.1) and the resulting graph shown in figure (4). We start by deriving a collection of useful lemmas:

**Lemma A.4.** *If  $U = 1$  for the  $k$ -center objective or  $U = n - k$  for the  $k$ -median and  $k$ -means objectives, then  $L(U) = 4$  for all objectives. Further the only set of centers that can lead to a cost not exceeding  $U$  is  $S = \mathcal{F}$ .*

*Proof.* First it is clear that if we choose the set  $\mathcal{F}$  to be the centers, i.e.  $S = \mathcal{F}$ , then if we route each point to one of its closest centers in  $\mathcal{F}$ , then we can have for the  $k$ -center we would have a cost of 1 since every point in the graph is at most a distance 1 from a point in  $\mathcal{F}$ . Further, for the  $k$ -median and  $k$ -means objectives, the points  $\mathcal{F}$  would be routed to themselves and every other point would be routed to one of its closest centers in  $\mathcal{F}$  which is at a distance of 1, this leads to a cost of  $(0)k + (1)(n - k) = n - k$ , therefore choosing the  $\mathcal{F}$  as the set of centers we can indeed satisfy the upper bound  $U$  for all objectives.

Now, consider another set of centers  $S'$  such that  $\exists i \in S'$  and  $i \notin \mathcal{F}$ , i.e. we have at least one center not from  $\mathcal{F}$ . Let  $f$  be the point in  $\mathcal{F}$  not selected in  $S'$ . For the  $k$ -center objective with  $U = 1$ , it follows that the blue auxiliary points of  $f$  have to be made as centers since every other point is at least a distance of 2 away from them, but each auxiliary point of  $f$  is made a center, then it follows that  $|S' - \mathcal{F}| \geq 3$ , i.e. at least two more points of  $\mathcal{F}$  have not be selected as centers. We can invoke the argument again on the new auxiliary points to conclude that  $|S' - \mathcal{F}| \geq 9$ . Invoking the argument again, we will see get that  $|S' - \mathcal{F}| \geq 3k$  which is infeasible since  $|S' - \mathcal{F}| \leq |S'| \leq k$ . Therefore, for the  $k$ -center with  $U = 1$ , we must have  $S = \mathcal{F}$ . Now having proven that  $S = \mathcal{F}$  and since  $U = 1$ , it follows that the smallest cluster size is 4 formed by mapping the center in  $S = \mathcal{F}$  to itself along with its auxiliary points, i.e.  $L(U) = 4$  for the  $k$ -center.

For the  $k$ -median and  $k$ -means objectives with  $U = n - k$ , similar to the  $k$ -center it is clear that every point which has not been selected as a center must have a center at a distance of at most 1 away. If we exclude one point  $f \in \mathcal{F}$  from the set of centers, then its auxiliary points will each have to become centers to satisfy the upper bound cost of  $U = n - k$ , but this would mean that there are at least 2 more points in  $\mathcal{F}$  that have been excluded. Following an argument similar to that of the  $k$ -center, we will have that the set of required centers would be at least  $3k$  which is a contradiction. Therefore, the only possible choice of centers is  $S = \mathcal{F}$ . It follows as well that the smallest cluster size if 4 formed by mapping the center in  $S = \mathcal{F}$  to itself along with its auxiliary points, i.e.  $L(U) = 4$  for the  $k$ -median and  $k$ -means objectives.  $\square$

Further, we define  $\Delta_{\text{red}}^i$  and  $\Delta_{\text{blue}}^i$  as the red and blue violations in the  $i^{\text{th}}$  cluster, respectively. Then we have the following lemma

**Lemma A.5.** *For the two color case of the above reduction,  $\Delta_{\text{red}}^i = \Delta_{\text{blue}}^i$  and  $\Delta_{\text{red}} = \Delta_{\text{blue}}$ .*

*Proof.* for cluster  $i$ , consider the red and blue violations  $\Delta_{\text{red}}^i, \Delta_{\text{blue}}^i$  at that cluster, then we have:

$$\Delta_{\text{red}}^i = |p_{\text{red}}^i - \frac{1}{3}| = |(1 - p_{\text{blue}}^i) - (1 - \frac{2}{3})| = |\frac{2}{3} - p_{\text{blue}}^i| = \Delta_{\text{blue}}^i$$

It is clear then that  $\Delta_{\text{red}} = \max_{i \in [k]} \Delta_{\text{red}}^i = \max_{i \in [k]} \Delta_{\text{blue}}^i = \Delta_{\text{blue}}$   $\square$

The following lemma follows immediately from the above:

**Lemma A.6.** *For the two color case of the above reduction*  $\text{GROUP-UTILITARIAN} = 2\text{GROUP-EGALITARIAN}$ .

*Proof.*  $\text{GROUP-UTILITARIAN} = \Delta_{\text{red}} + \Delta_{\text{blue}} = 2\Delta_{\text{red}} = 2\text{GROUP-EGALITARIAN}$ .  $\square$

We also note the following lemma:

**Lemma A.7.** *For a given cluster  $i$  with set of points  $C_i$ , if the set of red points in the cluster  $C_i^{\text{red}}$  satisfy  $\Delta_{\text{red}}^i = |\frac{|C_i^{\text{red}}|}{|C_i|} - \frac{1}{4}| < \frac{1}{4|C_i|}$ , then cluster  $i$  has no violation.*

*Proof.* Suppose that  $|\frac{|C_i^{\text{red}}|}{|C_i|} - \frac{1}{4}| < \frac{1}{4|C_i|}$ , then it follows that  $||C_i^{\text{red}}| - \frac{1}{4}|C_i|| < \frac{1}{4}$ . Since  $|C_i|$  is an integer it follows that  $\frac{1}{4}|C_i|$  is of the form  $m, m + \frac{1}{4}, m + \frac{1}{2}$ , or  $m + \frac{3}{4}$  where  $m$  is an integer. Further since  $|C_i^{\text{red}}|$  is also an integer, the fact that  $||C_i^{\text{red}}| - \frac{1}{4}|C_i|| < \frac{1}{4}$  implies that  $|C_i^{\text{red}}| = \frac{1}{4}|C_i|$  and we have no violation for the red color in cluster  $i$ . Further, from Lemma A.5 the blue violation equals the red violation and therefore we have no violation in cluster  $i$ .  $\square$

Now we are ready to prove the main claims for the **FCBC** problem.

For the first claim, assume by contradiction that a polynomial time algorithm gave a solution of violation less than  $\frac{1}{8L}$  and cost  $\leq U$ . Now, if we consider clusters  $i$  of size  $|C_i|$  such that  $4 \leq |C_i| \leq 8$ , then it clear that since  $\Delta_{\text{red}}^i \leq \Delta_{\text{red}} \leq \frac{1}{8L(U)}$ ,  $\Delta_{\text{red}}^i \leq \frac{1}{4|C_i|}$  because  $|C_i| \leq 8 \leq 2L(U)$ , therefore there is no violation in these clusters by Lemma A.7.

Now consider a cluster of size greater than 8, (note by Lemma A.4 that  $S = \mathcal{F}$ ) because of the upper bound  $U$  such clusters could only add points for the top row set  $\mathcal{T}$  to the cluster which are all red, it clear that the more red points are added the greater the violation, if one additional red point is added, then for the best color proportions the cluster has a total of: 6 blues and 3 reds, which lead to a violation of  $|\frac{1}{3} - \frac{1}{4}| = \frac{1}{12} > \frac{1}{8L(U)} = \frac{1}{32}$ , therefore it is impossible for the algorithm to form such clusters as that would contradict the assumption that the algorithm obtains a violation  $< \frac{1}{8L(U)}$  for each color. Therefore such clusters are not possible. This means that there is no violation in any cluster and that the problem has been solved optimally which by the NP-hardness is impossible unless  $P = NP$ .

Now the two remaining claims follow easily. By definition we have that  $\text{GROUP-EGALITARIAN} = \max_{h \in \mathcal{H}} \Delta_h$ . If  $\text{GROUP-EGALITARIAN} < \frac{1}{8L(U)}$ , then it follows that  $\Delta_h < \frac{1}{8L(U)}$  for every color  $h \in \mathcal{H}$  which by the first claim cannot happen unless  $P = NP$ .

Further, by Lemma A.6  $\text{GROUP-UTILITARIAN} = 2 \text{GROUP-EGALITARIAN}$ , therefore if  $\text{GROUP-UTILITARIAN} < \frac{|\mathcal{H}|}{8L(U)}$ , then  $\text{GROUP-EGALITARIAN} < \frac{1}{8L}$ . which is impossible unless  $P = NP$ .

The same claims for the **FABC** problem can be proven by simply setting the set of centers  $S = \mathcal{F}$  and the upper bound  $U = 1$  for  $k$ -center and  $n - k$  for the  $k$ -median/means, then following similar arguments.  $\square$

**Note:** In the proof above, if we consider only the upper bound cost  $U$  and ignore the fairness objective the problem is solvable in polynomial time. For **FCBC** simply set  $S = \mathcal{F}$ , then route each point to one of its closest centers. For the **FABC** with a given  $S = \mathcal{F}$ , again simply route each points to one of its closest centers. This highlights that the hardness is not from the upper bound cost  $U$ .

Next, we recall Theorem 6.2

**Theorem 6.2.** *For any clustering objective and both the GROUP-UTILITARIAN and GROUP-EGALITARIAN objectives, given an algorithm that solves fair assignment under a bounded cost **FABC** with additive approximation  $\mu$ , the fair clustering under a bounded cost **FCBC** problem can be solved with an additive approximation of  $\mu$  and at a cost of at most  $(2 + \alpha)U$ , where  $\alpha$  is the approximation ratio of the color-blind clustering algorithm.*

*Proof.* Let  $S$  and  $\phi$  be the set of centers and assignment of the color-blind algorithm. Let  $S^*$  and  $\phi^*$  be the optimal set of centers and assignment for the fair assignment under bounded cost **FABC**. Let  $\phi'$  be an assignment that routes the vertices from their center in  $S^*$  to the nearest center in  $S$ , i.e. for a given vertex  $j$ ,  $\phi'(j) = \operatorname{argmin}_{i' \in S} d(i', \phi^*(j))$ . Based on this setting we can upper bound the objective based on the following:

$$\begin{aligned} d(j, \phi'(j)) &\leq d(j, \phi^*(j)) + d(\phi'(j), \phi^*(j)) \\ &\leq d(j, \phi^*(j)) + d(\phi(i), \phi^*(j)) \\ &\leq d(j, \phi^*(j)) + d(j, \phi^*(j)) + d(j, \phi(j)) \\ &\leq 2d(j, \phi^*(j)) + d(j, \phi(j)) \end{aligned}$$

It follows then by the triangle inequality of the  $p$ -norm and the non-negativity of the components, that  $\left(\sum_{j \in \mathcal{C}} d^p(j, \phi'(j))\right)^{1/p} \leq 2\left(\sum_{j \in \mathcal{C}} d^p(j, \phi^*(j))\right)^{1/p} + \left(\sum_{j \in \mathcal{C}} d^p(j, \phi(j))\right)^{1/p} \leq 2U + \alpha U = (2 + \alpha)U$ . Note that in the last inequality the bounded the color-blind cost as follows:  $\left(\sum_{j \in \mathcal{C}} d^p(j, \phi(j))\right)^{1/p} \leq \alpha \operatorname{OPT}_{\text{cb}} \leq \alpha U$ , where as noted the optimal color-blind cost  $\operatorname{OPT}_{\text{cb}}$  is upper bounded by  $U$ , i.e.  $\operatorname{OPT}_{\text{cb}} \leq U$  otherwise the problem would not be feasible. This proves the upper bound on the objective.

Now we establish guarantees on the proportions. For a given center  $s$  in  $S$ , let  $N(s) = \{i' \in S^* \mid s = \operatorname{argmin}_{i \in S} d(i, i')\}$ , i.e.  $N(s)$  is the set of centers in  $S^*$  routing their vertices to  $s$ . Denote the set of points assigned to cluster  $i'$  by  $\phi^{*-1}(i')$ , i.e.  $\phi^{*-1}(i') = \{j \in \mathcal{C} \mid \phi^*(j) = i'\}$ . Then for any color  $h$  we have that:

$$\begin{aligned} \min_{i' \in N(s)} \frac{\left(\sum_{j \in \phi^{*-1}(i'), \chi(j)=h} 1\right)}{|\phi^{*-1}(i')|} &\leq \\ \frac{\sum_{i' \in N(s)} \left(\sum_{j \in \phi^{*-1}(i'), \chi(j)=h} 1\right)}{\sum_{i' \in N(s)} |\phi^{*-1}(i')|} &\leq \\ \max_{i' \in N(s)} \frac{\left(\sum_{j \in \phi^{*-1}(i'), \chi(j)=h} 1\right)}{|\phi^{*-1}(i')|} & \end{aligned}$$

That is the final color proportion will be within the lower and upper proportions of the routing centers. It follows that  $\Delta_h$  does not increase for any color and that the GROUP-UTILITARIAN, GROUP-EGALITARIAN, and GROUP-LEXIMIN objectives using  $\phi'$  are not greater than that of the optimal solution.

The above facts, combined with the premise of having an algorithm that solves the fair assignment under bounded cost **FABC** with an additive violation of  $\mu$  completes the proof.  $\square$

Next, we recall Theorem 6.3

**Theorem 6.3.** *The constraint set (7) is not convex.*

*Proof.* The non-convexity of the constraint set (7) can be shown even when ignoring the upper proportionality constraint, i.e. constraint (7d) only with the lower bound. Specifically, we would have

the following constraint set:

$$\sum_{i,j} d^P(i,j)x_{ij} \leq U^P \quad (8a)$$

$$\forall j \in \mathcal{C} : \sum_{i \in \mathcal{S}} x_{ij} = 1, \quad x_{ij} \in [0, 1] \quad (8b)$$

$$\forall h \in \mathcal{H} : \Delta_h \in [0, 1] \quad (8c)$$

$$\forall h \in \mathcal{H}, \forall i \in \mathcal{S} : (\beta_h - \Delta_h) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \leq \sum_{j \in \mathcal{C}^h} x_{ij} \quad (8d)$$

Now, assume that the upper bound on the cost  $U$  is sufficiently large (this would let assignments of a high cost remain feasible). Consider the case of two colors: red and blue, with  $\beta_{\text{red}} = \beta_{\text{blue}} = \frac{1}{2}$ . Let each color constitute half the dataset, i.e.  $|\mathcal{C}^{\text{red}}| = |\mathcal{C}^{\text{blue}}| = \frac{n}{2}$ , clearly  $|\mathcal{C}| = 2|\mathcal{C}^{\text{red}}| = 2|\mathcal{C}^{\text{blue}}| = n$ . Set the number of clusters to two ( $k = 2$ ), consider the following two feasible solutions  $x_{ij}^1, \Delta_{\text{red}}^1, \Delta_{\text{blue}}^1$  and  $x_{ij}^2, \Delta_{\text{red}}^2, \Delta_{\text{blue}}^2$  with  $\Delta_{\text{blue}}^1 = \Delta_{\text{blue}}^2 = 1$ , then the following holds (note that  $\alpha = \frac{2}{3}$ ):

**For  $x_{ij}^1, \Delta_{\text{red}}^1$ :**

$$\text{cluster 1: } \sum_{j \in \mathcal{C}^{\text{red}}} x_{1j}^1 = \sum_{j \in \mathcal{C}^{\text{blue}}} x_{1j}^1 = \alpha \frac{n}{2} = \frac{2}{3} \frac{n}{2} = \frac{n}{3}$$

$$\text{cluster 2: } \sum_{j \in \mathcal{C}^{\text{red}}} x_{2j}^1 = \sum_{j \in \mathcal{C}^{\text{blue}}} x_{2j}^1 = (1 - \alpha) \frac{n}{2} = \frac{1}{3} \frac{n}{2} = \frac{n}{6}$$

$$|C_2| = \sum_{j \in \mathcal{C}} x_{2j}^1 = \frac{n}{3}$$

$$\Delta_{\text{red}}^1 = 0$$

**For  $x_{ij}^2, \Delta_{\text{red}}^2$ :**

$$\text{cluster 1: } \sum_{j \in \mathcal{C}^{\text{red}}} x_{1j}^2 = \frac{n}{2},$$

$$\sum_{j \in \mathcal{C}^{\text{blue}}} x_{1j}^2 = (1 - (\alpha + \frac{1}{n/2})) \frac{n}{2} = \frac{n}{6} - 1$$

$$\text{cluster 2: } \sum_{j \in \mathcal{C}^{\text{red}}} x_{2j}^2 = 0,$$

$$\sum_{j \in \mathcal{C}^{\text{blue}}} x_{2j}^2 = (\alpha + \frac{1}{n/2}) \frac{n}{2} = (\frac{2}{3} + \frac{1}{n/2}) \frac{n}{2} = \frac{n}{3} + 1$$

$$|C_2| = \sum_{j \in \mathcal{C}} x_{2j}^2 = \frac{n}{3} + 1$$

$$\Delta_{\text{red}}^2 = \frac{1}{2}$$

We now form a simple convex combination of the two solutions  $x_{ij} = \frac{1}{2}(x_{ij}^1 + x_{ij}^2)$ ,  $\Delta_{\text{red}} = \frac{1}{2}(\Delta_{\text{red}}^1 + \Delta_{\text{red}}^2) = \frac{1}{4}$ . Constraints (8a), (8b), and (8c) would clearly be satisfied, but if we consider constraint (8d) for the red color and the second cluster, then we have:

$$RHS = \sum_{j \in \mathcal{C}^{\text{red}}} x_{2j} = \frac{n}{12}$$

$$LHS = (\frac{1}{2} - \frac{1}{4})(\frac{n}{3} + \frac{1}{2}) = \frac{n}{12} + \frac{1}{8}$$

It is clear that  $LHS \leq RHS$  does not hold and therefore, the constraint is not satisfied for the convex combination and therefore the constraint set of the problem is indeed not convex.

A similar assignment of solutions can be used to show that the set is not convex if we were to consider only the over-representation constraint in (7d) instead.  $\square$

Next, we recall Theorem 6.4

**Theorem 6.4.** *For FABC with GROUP-UTILITARIAN objective, we can use  $O\left(\left(\frac{1}{\epsilon}\right)^{|\mathcal{H}|-1}\right)$ -many LP runs to obtain an LP solution with additive approximation  $|\mathcal{H}|\epsilon$ .*

Before we give the proof of Theorem 6.4 we introduce some observations and algorithm (3). The core of the speed up is an algorithm that runs  $O\left(\frac{1}{\epsilon}\right)$  many LPs for the two color case. Therefore given a general instance of  $|\mathcal{H}| > 2$  many colors, we simply explore all  $O\left(\left(\frac{1}{\epsilon}\right)^{|\mathcal{H}|-2}\right)$  possibilities for  $|\mathcal{H}| - 2$  colors and find the optimal value for the two excluded colors through  $O\left(\frac{1}{\epsilon}\right)$  many LP runs for each possibility, leading to a total of  $O\left(\left(\frac{1}{\epsilon}\right)^{|\mathcal{H}|-1}\right)$ .

**Algorithm for FABC with two colors and arbitrary proportion bounds:** Now we explain our algorithm for the two color case that runs at most  $O\left(\frac{1}{\epsilon}\right)$  many LPs. The algorithm utilizes two facts:

**Fact A.2.** *If the LP is feasible for  $\Delta_1$  and  $\Delta_2$ , then it is also feasible for  $\Delta'_1$  and  $\Delta'_2$  where it either holds that  $\Delta'_1 \geq \Delta_1$ ,  $\Delta'_2 \geq \Delta_2$ , or both.*

**Fact A.3.** *If the LP is feasible for  $\Delta_1$  and  $\Delta_2$ , then  $\Delta'_1 = \Delta_1 + i$  and  $\Delta'_2 = \Delta_2 - i$  result in the same value for the GROUP-UTILITARIAN objective.*

As mentioned before we discretize the set of possibilities for  $\Delta_1$  and  $\Delta_2$  in the range  $[0, 1]$  by  $\epsilon = \frac{1}{r}$  where  $r$  is a positive integer. This leads to a set of  $E_\epsilon \times E_\epsilon$  many possibilities where  $E_\epsilon = \{\epsilon, 2\epsilon, \dots, (\frac{1}{\epsilon} - 1)\epsilon, 1\}$ . This results in a two dimensional grid with each cell having a certain value for the GROUP-UTILITARIAN objective as figure 5 shows for a specific value of  $\epsilon$ .

The algorithm is shown in block (3). Note that  $LP(\Delta_1, \Delta_2)$  is a function that returns **true** if the LP is feasible for the proportional violations of  $\Delta_1$  and  $\Delta_2$  for colors 1 and 2, respectively and returns **false** otherwise. The rough sketch of the algorithm is that it starts at the top right of the grid and checks for feasibility. It then chooses smaller values of  $\Delta_1$  until it encounters an infeasible instance. Once that happens the algorithm moves diagonally, looking for a feasible instance<sup>5</sup>. Once a feasible instance is found on the diagonal, the algorithm moves vertically down until it reaches the bottom and terminates or reaches an infeasible instance and therefore goes back to exploring diagonally again. See figure 6 for a run of the algorithm.

**Lemma A.8.** *Algorithm (3) runs at most  $O\left(\frac{1}{\epsilon}\right)$  many LPs and finds a solution with additive approximation  $2\epsilon$ .*

*Proof.* First we start by bounding the total number of LP runs. Since the algorithm in each step either moves horizontally (at most  $O\left(\frac{1}{\epsilon}\right)$  steps), or diagonally (at most  $O\left(\frac{1}{\epsilon}\right)$  steps), or vertically (at most  $O\left(\frac{1}{\epsilon}\right)$  steps), the total number of LP runs is indeed  $O\left(\frac{1}{\epsilon}\right)$ . We note that the horizontal and vertical explorations can be done faster through binary search but the diagonal still has to be done through a linear sweep which makes the asymptotic number of LP runs still  $O\left(\frac{1}{\epsilon}\right)$ .

Now we prove that we obtain a values at most  $2\epsilon$  greater than the optimal. Suppose that the optimal proportional violations are  $\Delta_1^*$  and  $\Delta_2^*$ , then in the discrete grid we have  $\bar{\Delta}_1$  and  $\bar{\Delta}_2$  such that  $\bar{\Delta}_1 - \Delta_1^* \leq \epsilon$  and  $\bar{\Delta}_2 - \Delta_2^* \leq \epsilon$  which must be feasible, this follows by Fact (A.2). It is therefore clear that there is a cell in the grid with value at most  $2\epsilon$  more than the optimal. Now we show that the algorithm either finds it or finds a better solution. If  $\bar{\Delta}_2 = 1$ , then during the horizontal exploration the algorithm will pass over the  $(\bar{\Delta}_1, \bar{\Delta}_2)$  cell and therefore it will either output  $\bar{\Delta}_1 + \bar{\Delta}_2$  as the optimal value or find a smaller value.

If  $\bar{\Delta}_2 < 1$ , then by Fact (A.2) cell  $(\bar{\Delta}_1, 1)$  must be feasible. Since  $(\bar{\Delta}_1, 1)$  is feasible, then if  $(\bar{\Delta}_1 - \epsilon, 1)$  is infeasible the algorithm will switch to vertical exploration and encounter cell  $(\bar{\Delta}_1, \bar{\Delta}_2)$ . If a cell  $(\Delta_1, 1)$  with  $\Delta_1 < \bar{\Delta}_1$  is feasible, then the algorithm will continue horizontally exploring, assuming cell  $(\epsilon, \epsilon)$  is not feasible<sup>6</sup> then the algorithm eventually switches to diagonal exploration

<sup>5</sup>Note that there is no point to explore above the diagonal as these are cells that lead to higher not lower values for the GROUP-UTILITARIAN objective

<sup>6</sup>If cell  $(\epsilon, \epsilon)$  is feasible, then the algorithm will explore horizontally to the last cell, followed by vertical exploration to the last cell.

---

**Algorithm 3** Algorithm for Arbitrary Proportion Two Color Case for the GROUP-UTILITARIAN Objective

---

```

1: Set  $UTIL^* = 2, \Delta_1^* = 1, \Delta_2^* = 1$ .
2: Set direction = 0. {0 for horizontal, 1 for diagonal, 2 for vertical}
3: while (direction = 1 and  $\Delta_1 \leq 1$  and  $\Delta_2 \geq \epsilon$ ) or (direction = 2 and  $\Delta_2 \geq \epsilon$ ) do
4:   if direction = 0 then
5:      $\Delta_1 = \Delta_1 - \epsilon$ 
6:     if  $\Delta_1 = 0$  then
7:        $\Delta_1 = \Delta_1 + \epsilon$ 
8:       direction = 2
9:     else if  $LP(\Delta_1, \Delta_2) = \text{TRUE}$  then
10:       $\Delta_1^* = \Delta_1$ 
11:       $UTIL^* = \Delta_1 + \Delta_2$ 
12:     else if  $LP(\Delta_1, \Delta_2) = \text{FALSE}$  then
13:       $\Delta_1 = \Delta_1 + \epsilon$ 
14:      direction = 1 {Explore diagonally}
15:     end if
16:   else if direction = 1 then
17:      $\Delta_1 = \Delta_1 + \epsilon$ 
18:      $\Delta_2 = \Delta_2 - \epsilon$ 
19:     if  $LP(\Delta_1, \Delta_2) = \text{TRUE}$  then
20:       $UTIL^* = \Delta_1 + \Delta_2$ 
21:       $\Delta_1^* = \Delta_1$ 
22:       $\Delta_2^* = \Delta_2$ 
23:      direction = 2 {Explore vertically}
24:     end if
25:   else if direction = 2 then
26:      $\Delta_2 = \Delta_2 - \epsilon$ 
27:     if  $LP(\Delta_1, \Delta_2) = \text{TRUE}$  then
28:       $UTIL^* = \Delta_1 + \Delta_2$ 
29:       $\Delta_2^* = \Delta_2$ 
30:     else
31:      direction = 1 {Explore diagonally}
32:     end if
33:   end if
34: end while
35: return  $\Delta_1^*, \Delta_2^*, UTIL^*$ .

```

---

if this diagonal has cells of GROUP-UTILITARIAN value less than  $\bar{\Delta}_1 + \bar{\Delta}_2$ , then the algorithm has an encountered a feasible cell with value less than or equal to  $\bar{\Delta}_1 + \bar{\Delta}_2$ . If the diagonal has cells of GROUP-UTILITARIAN value equal to  $\bar{\Delta}_1 + \bar{\Delta}_2$ , then either the algorithm will encounter the cell  $(\bar{\Delta}_1, \bar{\Delta}_2)$  or it will encounter another feasible with the same value by Fact A.3. If the diagonal has cells of GROUP-UTILITARIAN value greater than  $\bar{\Delta}_1 + \bar{\Delta}_2$ , then we note that we could have a collection of optimal cells of GROUP-UTILITARIAN value less than or equal to  $\bar{\Delta}_1 + \bar{\Delta}_2$  located below the diagonal, then it should be clear that the algorithm will encounter the optimal cell with smallest  $\Delta_1$  value.  $\square$

Now we proof theorem Theorem 6.4:

*Proof of Theorem 6.4.* We have  $|\mathcal{H}|$  many proportional violation values to decide, one for each color. We do exhaustive search over the proportional violation values in the discrete set  $E_\epsilon$  for all colors except for the first two colors  $h_1$  and  $h_2$ . For a given value of proportional violations for the colors in  $\mathcal{H} - \{h_1, h_2\}$ , we find the optimal values for colors  $h_1$  and  $h_2$  by running algorithm (3). It follows by Theorem (A.8), that we would find the optimal value for  $h_1$  and  $h_2$ . Further, since finding the optimal value for the two colors  $h_1$  and  $h_2$  takes  $O(\frac{1}{\epsilon})$  many LP runs and exhaustive search over the set of colors  $\mathcal{H} - \{h_1, h_2\}$  takes  $O((\frac{1}{\epsilon})^{|\mathcal{H}|-2})$  LP runs, then the optimal value can be found in at most  $O((\frac{1}{\epsilon})^{|\mathcal{H}|-1})$  LP runs.  $\square$

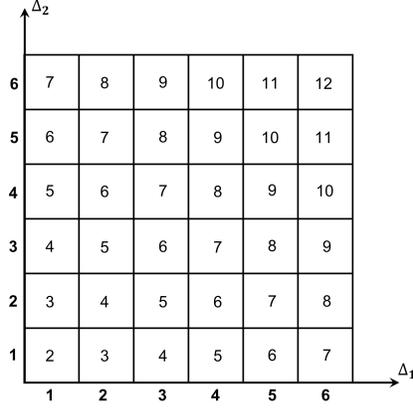


Figure 5: The two dimensional grid formed by discretizing the values of  $\Delta_1$  and  $\Delta_2$ . Here  $\epsilon = \frac{1}{6}$ . Note that each number shown is a multiple of  $\epsilon$ , i.e. the first coordinate on the ( $\Delta_1$ ) axis is  $1 \times \epsilon$ , then  $2 \times \epsilon$  and so on. For each cell we put the GROUP-UTILITARIAN objective value, which is again a multiple of  $\epsilon$ , i.e. the top right cell has a GROUP-UTILITARIAN of  $12\epsilon = 12\frac{1}{6} = 2$ . Notice how the cells on the same diagonal have the same value as stated in Fact (A.3.)

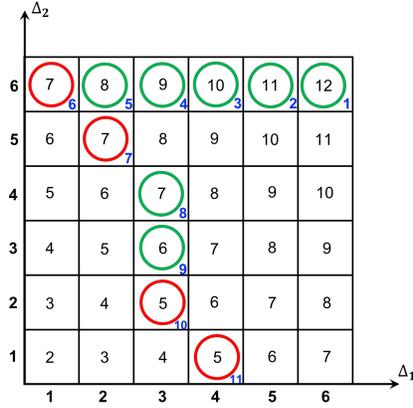


Figure 6: Here we show a possible run of the algorithm over the grid of figure 5. The circled cells are ones where the LP was run. A green circle indicates that the LP was feasible whereas a red circle indicates that the LP was infeasible. The blue number indicates the order where the LP corresponding to the cell was run, accordingly we start at the top right cell. It should be clear that the algorithm starts by horizontal exploration to the left, that it then moves to diagonal exploration which changes to vertical exploration once a feasible cell is found. Furthermore, the vertical exploration changes to diagonal once a an infeasible cell is encountered. It should be clear that the optimal value found by the algorithm is at the cell with  $\Delta_1 = \Delta_2 = 3\epsilon$  with GROUP-UTILITARIAN =  $6\epsilon$ .

Next, we recall Theorem 6.5

**Theorem 6.5.** For FABC with two colors, symmetric lower & upper bounds, and the GROUP-UTILITARIAN objective, we can use  $O\left(\log\left(\frac{1}{\epsilon}\right)\right)$ -many LP runs to get a solution with an additive approximation of  $|\mathcal{H}| \epsilon = 2\epsilon$ .

Before we prove Theorem 6.5 we point out the following definition and observations. For the two color case, our color set is  $\mathcal{H} = \{h_1, h_2\}$ . Further, we denote the proportions for color  $i$  by  $r_i$  where  $r_i = \frac{|\{j|j \in \mathcal{C}, x(j)=i\}|}{|\mathcal{C}|} = \frac{|\mathcal{C}^i|}{|\mathcal{C}|}$ . We use color  $h_1$  to denote the color with less points, i.e.  $r_1 \leq r_2$ . The upper and lower bounds we consider for each color are:  $\beta_i = (1 - \delta)r_i$  and  $\alpha_i = (1 + \delta)r_i$ . The idea behind the algorithm is that the proportions of one color imply the proportion of the other color.

**Algorithm for FABC with two colors and symmetric lower and upper proportion bounds:** Our algorithm is based on the simple observation shown in figure 7

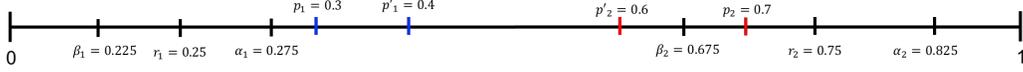


Figure 7: Proportions and bounds for two colors.  $r_1 = 0.25, r_2 = 0.75, \lambda_i = \delta r_i$  for  $i \in \{1, 2\}$  where  $\delta = 0.1$ . Notice how if color 1 violates the upper bound by having  $p_1 = 0.3$ , then we must have  $p_2 = 0.7$ , but color 2 is not violating. On the other hand, a violation for color 1 with  $p'_1 = 0.4$  implies  $p_2 = 0.6$  which causes a violation for color 2.

Without loss of generality, let  $\lambda_1 \leq \lambda_2$ , based on the observation in figure 7, we have the following lemma:

**Lemma A.9.** *If  $\Delta_1 < \lambda_2 - \lambda_1$ , then  $\Delta_2 = 0$ . If  $\Delta_1 \geq \lambda_2 - \lambda_1$ , then  $\Delta_2 = \Delta_1 - (\lambda_2 - \lambda_1)$*

*Proof.* Let color 1 have a  $\Delta_1$  violating proportion of  $p_1$ , then in some cluster  $p_1 = \alpha_1 + \Delta_1$  or  $p_1 = \beta_1 - \Delta_1$ .

Consider the case where  $p_1 = \alpha_1 + \Delta_1$ , then  $p_2 = 1 - p_1 = 1 - \alpha_1 - \Delta_1$ . Now if  $\Delta_1 < \lambda_2 - \lambda_1$ , then we have  $p_2 > 1 - \alpha_1 - (\lambda_2 - \lambda_1) = 1 - \lambda_2 + \lambda_1 - \alpha_1 = (1 - r_1) - \lambda_2 = r_2 - \lambda_2 = \beta_2$  this means that color 2 does not violate the lower bound. If we assume that color 2 violates the upper bound by an amount  $\Delta_2 > 0$ , then this would imply that  $p_1 = 1 - p_2$  and the lower violation for color 1 would be  $\beta_1 - p_1 = \beta_1 - (1 - \alpha_2 - \Delta_2) = \beta_1 - 1 + \alpha_2 + \Delta_2 = r_1 - \lambda_1 + r_2 + \lambda_2 + \Delta_2 - 1 = 1 + (\lambda_2 - \lambda_1) + \Delta_2 - 1 = (\lambda_2 - \lambda_1) + \Delta_2 > (\lambda_2 - \lambda_1)$  which is a contradiction since we assumed that  $\Delta_1 < (\lambda_2 - \lambda_1)$ .

Similarly, if  $\Delta_1 \geq (\lambda_2 - \lambda_1)$ , then we have  $\Delta_2 = \beta_2 - (1 - \alpha_1 - \Delta_1) = \beta_2 - 1 + \alpha_1 + \Delta_1 = r_2 - \lambda_2 + r_1 + \lambda_1 - 1 + \Delta_1 = \lambda_1 - \lambda_2 + \Delta_1 = \Delta_1 - (\lambda_2 - \lambda_1)$ . Now if we assume that color 2 has a violation of the upper bound by an amount  $\Delta'_2 > \Delta_1 - (\lambda_2 - \lambda_1)$ , this would imply that color 1 violates the lower bound by  $\beta_1 - p_1 = r_1 - \lambda_1 + r_2 + \lambda_2 - 1 + \Delta_2 = (\lambda_2 - \lambda_1) + \Delta_2$  which is a contradiction since  $\Delta_1 < \Delta_2 + (\lambda_2 - \lambda_1)$ , therefore color 2 cannot violate by more than  $\Delta_1 - (\lambda_2 - \lambda_1)$ .

The case of  $p_1 = \beta_1 - \Delta_1$  follows similar arguments. □

The above observations lead to the following algorithm:

---

**Algorithm 4** GROUP-UTILITARIAN Algorithm for Two Colors with Symmetric Bounds for the GROUP-UTILITARIAN Objective

---

**Input:** set of points  $\mathcal{C}$ , price of fairness  $U$ , for each color  $h \in \mathcal{H}$  lower and upper proportion values  $\beta_h, \alpha_h$ , error parameter  $\epsilon$ .

Define the set  $E_\epsilon = \{0, \epsilon, \dots, (\frac{1}{\epsilon} - 1)\epsilon\}$

Binary search  $\Delta_1$  over the set  $E_\epsilon$  by running the LP (7) ( if  $\Delta_1 < \delta(r_2 - r_1)$  then  $\Delta_2 = 0$ , otherwise set  $\Delta_2 = \Delta_1 - \delta(r_2 - r_1)$  ).

---

Now we are ready to prove Theorem 6.5.

*Proof of Theorem 6.5.* It follows from Lemma (A.9) that we can do binary search over the set  $E_\epsilon$  using  $\Delta_1$  as done in algorithm (4). Clearly, at most  $O\left(\log\left(\frac{1}{\epsilon}\right)\right)$  many LPs will be run because of binary search. Further, we know that we will find a solution at most  $\epsilon$  greater, i.e. we worst case best LP value is:  $\Delta_1^* + \epsilon, \Delta_2^* + \epsilon = (\Delta_1^* + \Delta_2^*) + 2\epsilon = \text{OPT} + 2\epsilon$ . □

Next, we recall Theorem 6.6

**Theorem 6.6.** *For FABC with the GROUP-EGALITARIAN objective, we can use  $O\left(\log\left(\frac{1}{\epsilon}\right)\right)$  many LP runs to get a solution with an additive approximation of  $\epsilon$ .*

*Proof.* For each color  $h$  set  $\Delta_h$  to the same value  $\Delta$ , do binary search over  $E_\epsilon$  using LP(8). Clearly, the final value is at most  $\epsilon$  greater than the optimal and we need at most  $O(\log(\frac{1}{\epsilon}))$  many LP runs.  $\square$

Now we introduce the following lemma:

**Lemma A.10.**  $\bar{\Delta}_h < \Delta_h + \frac{2}{L(U)}$ , i.e. rounding will increase the violation by at most  $\frac{2}{L(U)}$ .

*Proof.* Based on properties (ii) and (iii) from network flow rounding (see Section 6.1.3), we can get the following bound for the upper proportion:

$$\begin{aligned}
\sum_{j \in \mathcal{C}^h} \bar{x}_{ij} &\leq \left\lceil \sum_{j \in \mathcal{C}^h} x_{ij} \right\rceil && \text{(by property (iii))} \\
&\leq \left\lceil \min((\alpha_h + \Delta_h), 1) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \right\rceil && \text{(problem constraint)} \\
&< \min((\alpha_h + \Delta_h), 1) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) + 1 && \text{(ceiling upper bound)} \\
&\leq \min((\alpha_h + \Delta_h), 1) \left( \sum_{j \in \mathcal{C}} \bar{x}_{ij} + 1 \right) + 1 && \text{(by property (ii))} \\
&\leq \min((\alpha_h + \Delta_h), 1) \left( \sum_{j \in \mathcal{C}} \bar{x}_{ij} \right) + \min((\alpha_h + \Delta_h), 1) + 1 \\
&\leq \min((\alpha_h + \Delta_h), 1) \left( \sum_{j \in \mathcal{C}} \bar{x}_{ij} \right) + 2 && \text{(since } \min((\alpha_h + \Delta_h), 1) \leq 1) \\
&\leq (\alpha_h + \Delta_h) \left( \sum_{j \in \mathcal{C}} \bar{x}_{ij} \right) + 2
\end{aligned}$$

This implies that the new violation for the rounded solution  $\bar{\Delta}_h$  satisfies:

$$\alpha_h + \bar{\Delta}_h = \frac{\sum_{j \in \mathcal{C}^h} \bar{x}_{ij}}{\sum_{j \in \mathcal{C}} \bar{x}_{ij}} < \alpha_h + \Delta_h + \frac{2}{\sum_{j \in \mathcal{C}} \bar{x}_{ij}} \leq \alpha_h + \Delta_h + \frac{2}{L(U)}$$

Therefore, we have:

$$\bar{\Delta}_h - \Delta_h < \frac{2}{L(U)}$$

For the lower proportions, we also have:

$$\begin{aligned}
\sum_{j \in \mathcal{C}^h} \bar{x}_{ij} &\geq \left\lfloor \sum_{j \in \mathcal{C}^h} x_{ij} \right\rfloor && \text{(by property (iii))} \\
&\geq \left\lfloor \max((\beta_h - \Delta_h), 0) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \right\rfloor && \text{(problem constraint)} \\
&> \max((\beta_h - \Delta_h), 0) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) - 1 && \text{(ceiling upper bound)} \\
&\geq \max((\beta_h - \Delta_h), 0) \left( \sum_{j \in \mathcal{C}} \bar{x}_{ij} - 1 \right) - 1 && \text{(by property (ii))} \\
&\geq \max((\beta_h - \Delta_h), 0) \left( \sum_{j \in \mathcal{C}} \bar{x}_{ij} \right) - \max((\beta_h - \Delta_h), 0) - 1 \\
&\geq \max((\beta_h - \Delta_h), 0) \left( \sum_{j \in \mathcal{C}} \bar{x}_{ij} \right) - 2 && \text{(since } \max((\beta_h - \Delta_h), 0) \leq 1) \\
&\geq (\beta_h - \Delta_h) \left( \sum_{j \in \mathcal{C}} \bar{x}_{ij} \right) - 2
\end{aligned}$$

$$\beta_h - \bar{\Delta}_h = \frac{\sum_{j \in \mathcal{C}^h} \bar{x}_{ij}}{\sum_{j \in \mathcal{C}} \bar{x}_{ij}} > \beta_h - \Delta_h - \frac{2}{\sum_{j \in \mathcal{C}} \bar{x}_{ij}} \geq \beta_h - \Delta_h - \frac{2}{L(U)}$$

Therefore, we have:

$$\bar{\Delta}_h - \Delta_h < \frac{2}{L(U)}$$

□

Next, we recall Theorem 6.7:

**Theorem 6.7.** *For the FABC problem, the rounded solution has cost of at most  $U$  and an additive approximation of: (1)  $|\mathcal{H}|(\epsilon + \frac{2}{L(U)})$  for the GROUP-UTILITARIAN objective and (2)  $\epsilon + \frac{2}{L(U)}$  for the GROUP-EGALITARIAN objective.*

*Proof.* (1) For the GROUP-UTILITARIAN, by theorems (6.4) and (6.5) the LP solution has a violation of  $|\mathcal{H}| + \epsilon$ , then by lemma (A.10) and the definition of the GROUP-UTILITARIAN =  $\sum_{h \in \mathcal{H}} \Delta_h$ , the violation is at most  $|\mathcal{H}|(\epsilon + \frac{2}{L(U)})$ .

(2) For the GROUP-EGALITARIAN, by theorem (6.6) and lemma (A.10), the rounded solution would have a worst case violation of  $\epsilon + \frac{2}{L(U)}$  across the colors. □

We introduce the following lemma which is important for proving Theorem 7.1:

**Lemma A.11.** *Any polynomial time approximation algorithm for FABC for a general upper bound  $U$  must have  $\mu > 0$ , i.e. it must have a strictly greater than zero additive approximation guarantee.*

*Proof.* The proof follows from the proof of Theorem (5.2). Specifically, the proof of Theorem (5.2) shows that hard instances for FABC could have an optimal value of 0 for the GROUP-UTILITARIAN, GROUP-EGALITARIAN, and GROUP-LEXIMIN objectives, specifically when  $U = \text{OPT}_{\text{FC}}$  where  $\text{OPT}_{\text{FC}}$  is the optimal value of fair clustering. Therefore, if a polynomial time approximation algorithm with approximation ratio  $\rho \geq 1$  and additive approximation  $\mu \geq 0$  is ran over such hard instances, then it would output a solution of value  $\rho \text{OPT} + \mu = \rho(0) + \mu = \mu$ . If the algorithm has  $\mu = 0$ , then it would mean that the problem has been solved optimally which is impossible unless  $P = NP$ . Therefore,  $\mu > 0$ . □

Now we recall Theorem 7.1:

**Theorem 7.1.** *For **FABC**, the objectives **GROUP-UTILITARIAN-SUM** and **GROUP-EGALITARIAN-SUM** that sum across the clusters cannot be approximated in polynomial time to within an additive approximation of  $O(n^\delta)$  where  $\delta$  is a constant in  $[0, 1)$ , unless  $P = NP$ .*

*Proof.* By the result of Lemma A.11 we know that we can hard instances with  $\text{OPT} = 0$  and that any polynomial time algorithm should have an additive approximation  $\mu > 0$ . Further, we consider the same **X3C** reduction of Theorem 5.2 and Figure 4 for **FABC** with the centers set to the points of  $\mathcal{F}$ .

To prove Theorem 7.1, suppose by contradiction that an algorithm  $\mathcal{A}$  exists that guarantees an additive approximation of  $O(n^\delta)$  for  $\delta \in [0, 1)$ . Suppose, we are given an instance of the problem with optimal solution value of  $\text{OPT}$  and  $n$  many points. Note by Lemma A.7 if  $\Delta_{\text{red}}^i < \frac{1}{|C_i|}$ , then there is no violation. It follows that if  $\sum_{i \in [k]} (\Delta_{\text{red}}^i + \Delta_{\text{blue}}^i) < \frac{1}{4n}$  then we have no violation.

Now, create  $D$  many duplicates of the given set of points. Let the distance between the points belonging to the same duplicate be the same as in the original instance, whereas for points in different duplicates the distance is infinity. Further, let the number of centers be  $Dk$  where each duplicate has  $k$  many centers assigned at the same points as the original instance. Given the original upper bound on the clustering objective  $U$ , the new upper bound  $U'$  is set to  $U' = U$  for the  $k$ -center,  $U' = DU$  for the  $k$ -median, and  $U' = \sqrt{DU}$  for the  $k$ -means objectives.

If this modified instance is given to  $\mathcal{A}$ , then the output would have a value of at most  $\rho D \text{OPT} + c(Dn)^\delta$  for some  $c > 0$ . If  $D > \frac{1}{4^\delta - 1} c^{\frac{1}{1-\delta}} n^{\frac{1+\delta}{1-\delta}}$  (which is polynomial in  $n$ ), then the average violation across the duplicates is:

$$\begin{aligned} \frac{\rho D \text{OPT} + c(Dn)^\delta}{D} &= \rho \text{OPT} + cn^\delta D^{\delta-1} \\ &< \rho \text{OPT} + c \frac{1}{4} n^\delta c^{\frac{\delta-1}{1-\delta}} n^{\frac{(1+\delta)(\delta-1)}{1-\delta}} = \rho \text{OPT} + \frac{1}{4n} = 0 + \frac{1}{4n} \end{aligned}$$

This means that there must exist at least one duplicate for which the violation is at most  $\frac{1}{4n}$  which means that the problem has been exactly in polynomial time which is impossible unless  $P = NP$ .  $\square$

Next, we recall Theorem 8.1:

**Theorem 8.1.** *Suppose that there is a polynomial time algorithm which can obtain the optimal solution for **FCBC** for the upper bound of  $U$  if  $U \geq \alpha(\mathcal{I}) \text{OPT}_{\text{cb}}(\mathcal{I})$  where  $\mathcal{I}$  is a specific instance of **FCBC** and  $\text{OPT}_{\text{cb}}(\mathcal{I})$  is the optimal cost of its color-blind clustering. Then we have a true polynomial time approximation algorithm for fair clustering. Further, a true polynomial time  $\alpha'(\mathcal{I})$ -approximation algorithm for fair clustering implies that **FCBC** can be solved optimally in polynomial time for  $U \geq \alpha'(\mathcal{I}) \text{OPT}_{\text{FC}}(\mathcal{I})$ .*

*Proof.* Suppose  $\alpha(\mathcal{I}) \text{OPT}_{\text{cb}}(\mathcal{I}) \leq \text{OPT}_{\text{FC}}(\mathcal{I})$  where  $\text{OPT}_{\text{FC}}$  is the optimal fair clustering cost, then fair clustering is solvable in polynomial time which is impossible unless  $P = NP$  since fair clustering is NP-hard.

If  $\alpha(\mathcal{I}) \text{OPT}_{\text{cb}}(\mathcal{I}) > \text{OPT}_{\text{FC}}(\mathcal{I})$ . Then this algorithm will not have any fairness violation if we choose  $U = \alpha(\mathcal{I}) \text{OPT}_{\text{cb}}(\mathcal{I})$ , further its cost is  $\alpha(\mathcal{I}) \text{OPT}_{\text{cb}}(\mathcal{I}) \leq \alpha(\mathcal{I}) \text{OPT}_{\text{FC}}(\mathcal{I})$ . Therefore, we have a true polynomial time approximation algorithm for fair clustering with approximation ratio at most  $\alpha(\mathcal{I})$ .

Now we prove the second part. By definition the output of a true approximation for fair clustering would have no proportional violations therefore achieving the optimal value for any objective for **FCBC**. Therefore, we have an optimal algorithm for **FCBC** for  $U \geq \alpha'(\mathcal{I}) \text{OPT}_{\text{FC}}(\mathcal{I}) \geq \alpha'(\mathcal{I}) \text{OPT}_{\text{cb}}(\mathcal{I})$ .  $\square$

Further, a true approximation algorithm for fair clustering algorithm would imply an exact algorithm for fair clustering under a bounded cost **FCBC**.

**Theorem A.1.** *A true polynomial time  $\alpha(\mathcal{I})$  approximation algorithm for fair clustering implies that fair clustering under a bounded cost **FCBC** can be solved optimally in polynomial time for  $U \geq \alpha(\mathcal{I}) \text{OPT}_{\text{FC}}(\mathcal{I})$ .*

*Proof.* Since we have a true approximation, then there would be no fairness violation (optimal value). Further, this would be  $U$  such that  $U \geq \alpha(\mathcal{I}) \text{OPT}_{\text{FC}}(\mathcal{I}) \geq \alpha(\mathcal{I}) \text{OPT}(\mathcal{I})$ .  $\square$

## B Solution for the GROUP-LEXIMIN Case

Here we present the full details for the GROUP-LEXIMIN case, see algorithm (5) below.

---

### Algorithm 5 Leximin Algorithm

---

**Input:** set of points  $\mathcal{C}$ , price of fairness  $U$ , for each color  $h \in \mathcal{H}$  lower and upper proportion values  $\beta_h, \alpha_h$   
 $\mathcal{H}_{\text{fixed}} = \{\}, \Delta_{\min} = 0$   
**while**  $\mathcal{H}_{\text{fixed}} \neq \mathcal{H}$  **do**  
  **Step 0:** For each color  $h \in \mathcal{H}_{\text{fixed}}$  set its violation to its minimum found in the previous iterations  $\Delta_h^{\min}$ .  
  **Step 1:** For each color  $h \in \mathcal{H} - \mathcal{H}_{\text{fixed}}$ , set  $\Delta_h = \Delta$ .  
  Find the the minimum  $\Delta$ , such that  $\Delta < \Delta_{\min}^{q-1}$  that satisfies LP (7) using binary search over  $E_\epsilon$ , let  $\Delta_{\min}^q = \Delta + \frac{2}{L}$ .  
  **Step 2:** For the set  $h'_\ell \in \mathcal{H} - \mathcal{H}_{\text{fixed}}$ , find the minimum set of colors with violation  $\Delta_{\min}^q$  and add them to  $\mathcal{H}_{\text{fixed}}$  using LP (9).  
  **Step 3:** If in **Step 2** no color is found, then randomly pick a color from  $\mathcal{H} - \mathcal{H}_{\text{fixed}}$  and add it to  $\mathcal{H}_{\text{fixed}}$ .  
**end while**

---

LP (9) below is run once for each  $h'_\ell \in \mathcal{H} - \mathcal{H}_{\text{fixed}}$  (note lines 9c and 9e). The LP does not have an objective and amounts to a feasibility check.

$$\sum_{i,j} d^p(i,j) x_{ij} \leq U^p \quad (9a)$$

$$\forall j \in \mathcal{C} : \sum_{i \in S} x_{ij} = 1, \quad x_{ij} \in [0, 1] \quad (9b)$$

$$(9c)$$

for the given  $h'_\ell$ :

$$\left( \beta_h - \left[ \Delta_{\min}^q - \frac{2}{L} - \epsilon \right] \right) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \leq \sum_{\substack{j \in \mathcal{C}, \\ \chi(j)=h'_h}} x_{ij} \leq \left( \alpha_h + \left[ \Delta_{\min}^q - \frac{2}{L} - \epsilon \right] \right) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \quad (9d)$$

$$\forall h \in \mathcal{H} - \left( \mathcal{H}_{\text{fixed}} \cup \{h'_\ell\} \right), \forall i \in \{1, \dots, k\} : \quad (9e)$$

$$\left( \beta_h - \Delta_{\min}^q \right) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \leq \sum_{\substack{j \in \mathcal{C}, \\ \chi(j)=h}} x_{ij} \leq \left( \alpha_h + \Delta_{\min}^q \right) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \quad (9f)$$

$$\forall h \in \mathcal{H}_{\text{fixed}} : \left( \beta_h - \Delta_h^{\min} \right) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \leq \sum_{\substack{j \in \mathcal{C}, \\ \chi(j)=h}} x_{ij} \leq \left( \alpha_h + \Delta_h^{\min} \right) \left( \sum_{j \in \mathcal{C}} x_{ij} \right) \quad (9g)$$

In algorithm (5), step 1 does a binary search over the set  $E_\epsilon$  to find the minimum feasible violation  $\Delta_{\min}^q$  for the active set of colors (in the set  $\mathcal{H} - \mathcal{H}_{\text{fixed}}$ ). Step 2 finds the set of colors whose violation can be improved beyond  $\Delta_{\min}^q$ , by running an LP (see B) specific to each color in  $\mathcal{H} - \mathcal{H}_{\text{fixed}}$ . If all colors in  $\mathcal{H} - \mathcal{H}_{\text{fixed}}$  can improve, then in step 3 a random color is picked from that a set. Step 0 simply sets the violation to the optimal found value for the set of colors that can no longer be improved (the set  $\mathcal{H}_{\text{fixed}}$ ).

## C Discussion on the Size of the Smallest Cluster for a Given Cost Upper Bound

As discussed, it is clear from Theorem 6.1 that the larger the size of the smallest cluster, the better our approximation. In Section C.1 we consider examples where in the absence of outliers and for suitable values of  $k$ , we would not have small clusters. We note further that whereas the given cost is  $U$  the final approximation in Theorem 6.1 is in terms of  $\frac{1}{L(U')}$  where  $U' = (2 + \alpha)U$  with  $\alpha$  being the color-blind approximation ratio. So it is reasonable to wonder about the gap between  $L(U)$  from the lower bound of Theorem 5.3 and  $L(U')$  from Theorem 6.1. We show that the gap is example dependent, specifically in Section C.2 we show that the gap can be arbitrarily large and in Section C.3 we show that it's possible that they are precisely equal, i.e.  $L(U) = L(U')$ . We note here that color assignments have no significance.

### C.1 For suitable $k$ with no outliers $L(U)$ is large

For reasonable values of  $k$  and in the absence of outliers, provided the upper clustering cost  $U$  is not very large, we do not expect clusters of a small size, i.e.  $L(U)$  is large. In Figure 8 we choose a value of  $k$  that recovers the underlying structure. We can see that there are no small clusters. On the other had, in Figure 9 because of outliers we end up with small clusters. Moreover, if we choose a value of  $k$  greater than 3, then we notice of the clusters fragment as in Figure 10, this leads to smaller clusters but even then they are not pathologically small.

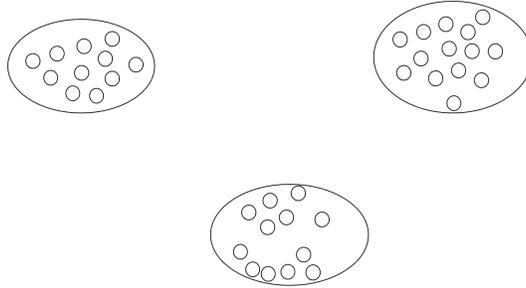


Figure 8: Here there are no outliers and we choose  $k = 3$  which recovers the clusters of the dataset.

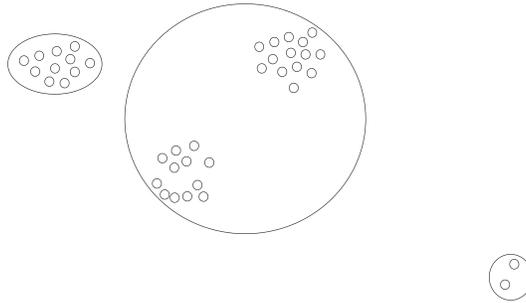


Figure 9: Because of outlier points, the smallest cluster consists of only two points.

### C.2 Example where the gap between $L(U)$ and $L(U')$ is large

See Figure 11 which is the same as that of Figure 4. Note that by Lemma A.4, for the  $k$ -center if  $U = 1$ , then  $L(U) = 4$ . But at  $U' = (2 + \alpha)U = 4$ , we can get  $L(U') = 1$ . This is not difficult to see since if we select any point from the top row  $\mathcal{T}$ , then every other point in the graph is at a distance of at most 2.

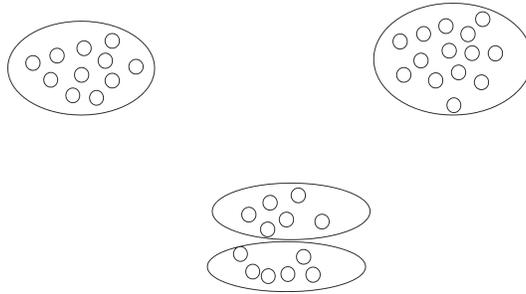


Figure 10: Although it can be argued that  $k = 3$  is more suitable for this example. Setting  $k = 4$  does not lead to very small clusters.

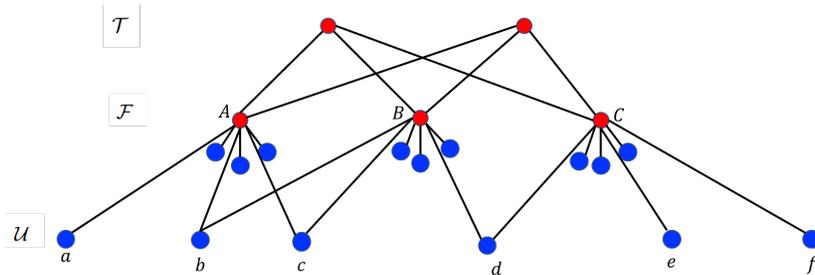


Figure 11: This is an example where the change in the size of the smallest cluster is very large. At  $U = 1$ , we have  $L(U) = 4$ . However, at  $U' = 4$ , we have  $L(U') = 1$ .

### C.3 Example where $L(U') = L(U)$

See Figure 12 with  $k = 3$  and the distance between the clusters  $R$  is sufficiently large, the size of the smallest cluster does not decrease with increasing cost for reasonable values of  $U$ . For the  $k$ -center, if we have  $U = r$ , then  $U' = (\alpha + 2)r$  where  $r$  is the maximum radius of the “true” clusters. If  $R > (2 + \alpha)r$  then we would not have small clusters.



Figure 12: In this example because of the large separation between the clusters, we would have  $L(U) = L(U')$  for reasonable values of  $U$ .

## D Tradeoff between the Clustering Cost Bound and Fairness

It is worth asking if we can obtain a clear tradeoff between the upper clustering cost bound and the achievable fairness (whether using the GROUP-EGALITARIAN or GROUP-UTILITARIAN objectives). In the two sections bellow, we show that in general this is not achievable. In some examples, the tradeoff is effectively a step function whereas in others it gradually increases. Throughout, we consider two clustering ( $k = 2$ ) with the  $k$ -center objective and set  $\alpha_{\text{red}} = \alpha_{\text{blue}} = \beta_{\text{red}} = \beta_{\text{blue}} = \frac{1}{2}$ . By the above proportions, it is not difficult to see that  $\Delta_{\text{red}} = \Delta_{\text{blue}}$  and that GROUP-UTILITARIAN = 2GROUP-EGALITARIAN, simply following the proofs of Lemma A.5 and Lemma A.6, respectively. Since GROUP-UTILITARIAN = 2GROUP-EGALITARIAN, we will only discuss the GROUP-EGALITARIAN objective for simplicity. Further, it follows as well that  $\Delta_{\text{red}}, \Delta_{\text{blue}} \leq \frac{1}{2}$  and that any non-trivial color-proportional clustering should have  $\Delta_{\text{red}}, \Delta_{\text{blue}} < \frac{1}{2}$ .

### D.1 Example where the tradeoff is a step function

In the following examples (a) and (b) shown in Figure 13 the tradeoff follows a step function. Letting  $r$  be the distance between the nearby same color points and letting  $R$  be the smallest distance between points of different color. It is clear that if  $U < R$ , then  $\text{GROUP-EGALITARIAN} = \frac{1}{2}$ . However, if  $U \geq R$ , then  $\text{GROUP-EGALITARIAN} = 0$ . Note that we can make the value of  $R$  arbitrarily large as shown in (a) and (b).



Figure 13: Examples where the tradeoff between the upper cost bound and the fairness objective is abrupt.

### D.2 Example where the fairness increases gradually with higher cost upper bound

Consider the case where the points lie on a line and  $|\mathcal{C}^{\text{red}}| = |\mathcal{C}^{\text{blue}}| = \frac{|\mathcal{C}|}{2} = n'$  where  $n'$  is odd. Clearly,  $|\mathcal{C}| = n = 2n'$ . Further, let all of the first  $n'$  points be red and the remaining  $n'$  points be blue. The distance between any two consecutive points is  $r$ . Figure 14 shows an example of such a construction for  $n' = 5$ .

It is not difficult to see that for the two clustering case ( $k = 2$ ), that the optimal radius is  $(\frac{n'-1}{2})r$ , simply set the middle point of each color as the center. Note that we would have  $\text{GROUP-EGALITARIAN} = \frac{1}{2}$ , i.e. no mixing of the colors.

If we index the upper  $U$  by the following  $U = U(m) = (\frac{n'-1}{2} + m)r$  where  $m$  is a non-negative integer and  $m < \frac{n'}{4}$ , then it is not difficult to show that the achievable fairness is  $\text{GROUP-EGALITARIAN}(m) = \frac{1}{2} - \frac{2m}{n'}$ . That is, if we allow the cost to increase by  $mr$ , then we can improve the  $\text{GROUP-EGALITARIAN}$  objective by  $\frac{2m}{n'}$ .

For example, in Figure 14, for  $U = U(0) = 2r$ , then we cannot mix the color and  $\text{GROUP-EGALITARIAN} = \frac{1}{2}$ . If on the other hand, we choose  $U = U(1) = 3r$ , then for the red color we can have a representation of  $\frac{2}{5}$ , leading to  $\text{GROUP-EGALITARIAN}(1) = \frac{1}{2} - \frac{2}{5}$ .

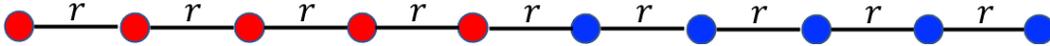


Figure 14: In the above example the tradeoff between the clustering cost and the fairness objective is gradual.

## E Network Flow Rounding

Here, we summarize the network flow rounding due to [11] as applied to our problem. Recall that it gives the following guarantees:

- (i)  $\sum_{i,j} d^p(i,j) \bar{x}_{ij} \leq \sum_{i,j} d^p(i,j) x_{ij}$ .
- (ii)  $\forall i \in \{1, \dots, k\} : \left\lceil \sum_{j \in \mathcal{C}} x_{ij} \right\rceil \leq \sum_{j \in \mathcal{C}} \bar{x}_{ij} \leq \left\lceil \sum_{j \in \mathcal{C}} x_{ij} \right\rceil$
- (iii)  $\forall h \in \mathcal{H}, \forall i \in \{1, \dots, k\} : \left\lceil \sum_{j \in \mathcal{C}^h} x_{ij} \right\rceil \leq \sum_{j \in \mathcal{C}^h} \bar{x}_{ij} \leq \left\lceil \sum_{j \in \mathcal{C}^h} x_{ij} \right\rceil$

The first guarantee essentially states that we can round the fractional LP assignments to an integral solution without increasing the cost given that we solved the LP for a fixed set of centers. However, proving this depends on the objective. For  $k$ -center, assigning a point  $j$  to any center  $i$  with  $x_{ij} > 0$

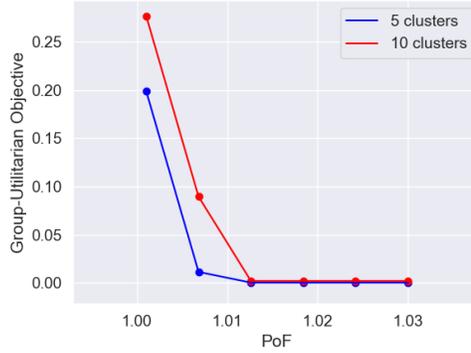


Figure 15: PoF versus the GROUP-UTILITARIAN objective for the **CreditCard** dataset with  $\delta = 0.05$ .

will not affect the cost of the solution with respect to the clustering objective. This is because  $k$ -center only minimizes the maximum radius of any cluster and we only allow  $x_{ij} > 0$  if  $d^p(i, j)$  is at most the maximum radius. On the other hand, a fractional solution to the LP for  $k$ -median or  $k$ -means may fractionally assign part of a point to a nearby center and part to a faraway center to improve fairness. A rounded integral solution that assign the point wholly to the farther center could potentially increase the cost. The proof due to [11] defines these objectives as reassignable ( $k$ -center) and separable ( $k$ -median and  $k$ -means) and focuses on the more challenging case of separable objectives.

We create a min-cost flow instance with unit capacities to solve this problem as follows.

- For each center  $i \in S$ , we define a subset of points  $V_i$  with a point  $v_i^h$  for each color  $h \in \mathcal{H}$  and a main point  $v_i^0$ . Each  $v_i^h$  has a balance of  $-\left\lfloor \sum_{j \in \mathcal{C}^h} x_{ij} \right\rfloor$  and  $v_i^0$  has a balance of  $-\left(\left\lfloor \sum_{j \in \mathcal{C}} x_{ij} \right\rfloor - \sum \left\lfloor \sum_{j \in \mathcal{C}^h} x_{ij} \right\rfloor\right)$ . The arc set of these points is  $(v_i^h, v_i^0)$  for each color  $h \in \mathcal{H}$  and each arc has cost 0.
- For each point  $j \in \mathcal{C}$ , we define a point  $v_j$  with balance of 1. For each  $x_{ij} > 0$  we add an arc  $(v_j, v_i^h)$  with cost  $d^p(i, j)$  where  $h$  is the color of  $j$ .
- Finally, we add a sink  $t$  with balance  $-(|\mathcal{C}| - \sum_{i \in S} \left\lfloor \sum_{j \in \mathcal{C}} x_{ij} \right\rfloor)$ . For each center  $i$ , we add an arc  $(v_i^0, t)$  with cost 0.

Note that all of the capacities, costs, and balances are integral and that the LP solution translates to a feasible flow. Thus, we can find an integral flow solution with cost at most that of LP solution and it is easy to see that this can then be translated back to an integral assignment. Also, note that our flow solution almost preserves the fairness of the LP solution. The additive error of 1 for the second and third guarantees above arise from taking the floor (e.g.,  $\left\lfloor \sum_{j \in \mathcal{C}^h} x_{ij} \right\rfloor$ ) to have integrality.

## F Additional Experimental Results

In this section, we report additional experimental result on the  $k$ -means and  $k$ -median objectives. We also use the **CreditCard** dataset from the UCI repository [24] with all 30,000 data points. The fairness attribute we use is marriage where we merge groups 2 (single) and 3 (other) into one group to have a binary attribute.

### F.1 Additional $k$ -means Results

#### F.1.1 GROUP-UTILITARIAN Objective

Here we add results for the GROUP-UTILITARIAN for the **CreditCard** dataset with  $\delta = 0.05$ , see Figure 15. We see that we can achieve better proportional violations for the smaller value of  $k$ .

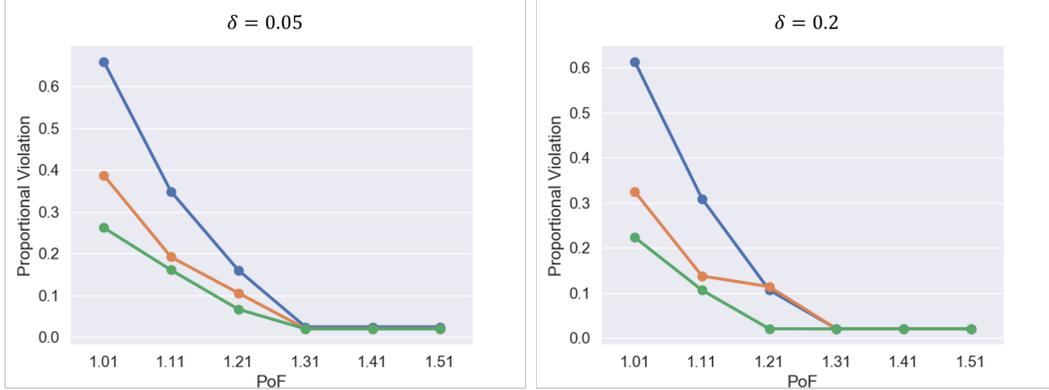


Figure 16: PoF versus the GROUP-LEXIMIN objective for the **Census1990** dataset for different values of  $\delta$ .

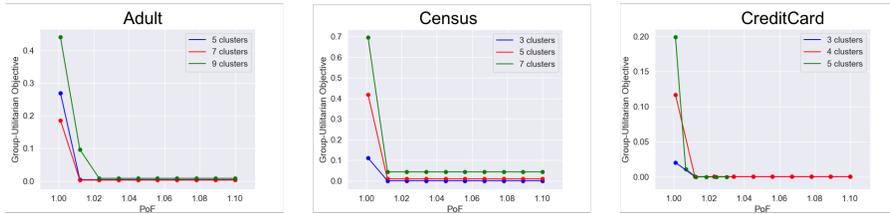


Figure 17: PoF versus the GROUP-UTILITARIAN objective for the **Adult**, **Census1990**, and **CreditCard** datasets with for the  $k$ -median objective.

### F.1.2 GROUP-LEXIMIN Objective

Figure 16 shows the results for GROUP-LEXIMIN on the **Census1990** dataset where we have  $k = 5$ . We set  $\delta = 0.05$  and  $\delta = 0.2$ , we see similar behaviour to that in Figure 3. We also notice that for higher values of  $\delta$  (more relaxed proportion bounds) some colors are able to achieve smaller proportional violation which we expect.

## F.2 $k$ -median Results

For the color-blind implementation of the  $k$ -median objective we follow [10, 22] and use the 5-approximation of [6] with modified  $D$ -sampling [5]. Because we are interested to see the behaviour of the algorithm and since the color-blind approximation is time consuming we sub-sample all datasets to 1,000 points.

### F.2.1 GROUP-UTILITARIAN Objective

Figure 17 shows the performance on the different datasets for the GROUP-UTILITARIAN objective. Note that we observe a similar trend as in figure 2 where it is easier to minimize the proportional violations when the number of clusters is lower. Note that we set  $\delta = 0.1$  for all datasets.

### F.2.2 GROUP-LEXIMIN Objective

Figure 18 shows the results on the GROUP-LEXIMIN objective on the **Census1990** dataset with  $k = 5$  and  $\delta = 0.1$ . The behaviour is very similar to the  $k$ -means.

## F.3 Checking the Size of the Smallest Cluster

Here we check the size of the smallest cluster again, for Section F.1 the smallest cluster is of size 176 for the **CreditCard** dataset and 168 for the **Census1990** dataset.

In Section F.2 where the datasets have been sub-sampled to 1,000 points. The smallest cluster size found is 27, 15, and 24 for the **Adult**, **Census1990**, and **CreditCard** datasets, respectively.

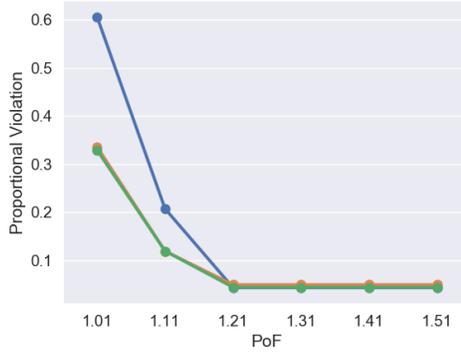


Figure 18: PoF versus the GROUP-LEXIMIN objective for the *Census1990* dataset with  $k = 5$  and  $\delta = 0.1$  for the  $k$ -median objective.

It is clear from these experiments that size of the smallest cluster is large and therefore the approximations we obtain are not far from the optimal.

## G Details of the Randomized Extension

We give a sketch of the algorithm, and defer some of the proof and details to the full version. Suppose the LP returns a solution  $x_{i,j} \in [0, 1]$  for all  $(i, j)$ , satisfying the following equalities for some non-negative quantities  $a_i$  and  $b_{i,h}$ :

1.  $\forall j \in \mathcal{C}: \sum_i x_{ij} = 1.$
2.  $\forall i \in \{1, \dots, k\}: \sum_{j \in \mathcal{C}} x_{ij} = a_i.$
3.  $\forall h \in \mathcal{H}, \forall i \in \{1, \dots, k\}: \sum_{j \in \mathcal{C}^h} x_{ij} = b_{i,h}.$
4.  $\forall (i, j): x_{i,j} \in [0, 1].$

We will sketch a randomized algorithm that gradually rounds the  $x_{i,j}$ 's to the eventual  $\bar{x}_{i,j}$  in a polynomial number of iterations. Let  $Y_{i,j,t}$  denote the (random) value of  $x_{i,j}$  at the end of iteration  $t$ ; initially, we deterministically have  $Y_{i,j,0} = x_{i,j}$ . Let  $A_{i,t}$  and  $B_{i,h,t}$  be random variables such that

- $\forall i \in \{1, \dots, k\}: A_{i,t} = \sum_{j \in \mathcal{C}} Y_{i,j,t}.$
- $\forall h \in \mathcal{H}, \forall i \in \{1, \dots, k\}: B_{i,h,t} = \sum_{j \in \mathcal{C}^h} Y_{i,j,t}.$

Let us now describe iteration  $t \geq 1$ , which operates on the values  $Y_{i,j,t-1}$  and probabilistically modifies them to the  $Y_{i,j,t}$ . Fix the values  $Y_{i,j,t-1}$  to be some arbitrary  $y_{i,j,t-1}$ , and define  $a_{i,t-1} = A_{i,t-1}$ ,  $b_{i,h,t-1} = B_{i,h,t-1}$ . We will maintain the following **five invariants**:

- (I1)  $\forall j \in \mathcal{C}: \sum_i Y_{i,j,t} = 1$  with probability one.
- (I2)  $\forall i \in \{1, \dots, k\}: \lfloor a_{i,t-1} \rfloor \leq A_{i,t} \leq \lceil a_{i,t-1} \rceil$  with probability one.
- (I3)  $\forall h \in \mathcal{H}, \forall i \in \{1, \dots, k\}: \lfloor b_{i,h,t-1} \rfloor \leq B_{i,h,t} \leq \lceil b_{i,h,t-1} \rceil$  with probability one.
- (I4)  $E[Y_{i,j,t}] = y_{i,j,t-1}.$
- (I5)  $Y_{i,j,t} \in [0, 1]$  with probability one.

In particular, we have the following key properties:

- if  $a_{i,t-1}$  is an integer, then  $A_{i,t} = a_{i,t-1}$  with probability one;
- if  $b_{i,h,t-1}$  is an integer, then  $B_{i,h,t} = b_{i,h,t-1}$  with probability one; and
- if  $y_{i,j,t-1}$  is an integer (which will be 0 or 1), then  $Y_{i,j,t} = y_{i,j,t-1}$  with probability one.

Our strategy is to show that there is a way of maintaining our invariants above, while making at least one *more*  $A_{i,t}$ ,  $B_{i,h,t}$ , or  $Y_{i,j,t}$  integral at the end of iteration  $t$ . Since there is only a polynomial number of these terms and since we are done when all the  $Y_{i,j,t}$ 's are integers, our proof of correctness will then be complete by a simple induction on  $t$ .

Briefly, iteration  $t$  starts with the following constraint system with variables  $z_{i,j}$ , initialized with the feasible solution  $z_{i,j} = y_{i,j,t-1}$ :

$$(C1) \quad \forall j \in \mathcal{C} : \sum_i z_{ij} = 1.$$

$$(C2) \quad \forall i \in \{1, \dots, k\} \text{ such that } a_{i,t-1} \text{ is an integer} : \sum_{j \in \mathcal{C}} z_{ij} = a_{i,t-1}.$$

$$(C3) \quad \forall h \in \mathcal{H}, \forall i \in \{1, \dots, k\} \text{ such that } b_{i,h,t-1} \text{ is an integer} : \sum_{j \in \mathcal{C}^h} z_{ij} = b_{i,h,t-1}.$$

$$(C4) \quad \forall (i, j) \text{ such that } y_{i,j,t-1} \text{ is an integer (i.e., lies in } \{0, 1\}) : z_{i,j} = y_{i,j,t-1}.$$

Given our initialization of  $z$ , we also have that  $\forall (i, j) : z_{i,j} \geq 0$ .

We prove below via a careful counting argument that the system (C1)-(C4) of the form  $Az = v$  is an under-determined system. Thus, there is a nonzero vector  $r$  that is efficiently computable such that  $Ar = 0$ . We then calculate certain positive scalars  $u_1$  and  $u_2$ , and probabilistically transition from the vector  $y = (y_{i,j,t-1})$  to the random vector  $Y = (Y_{i,j,t})$  as follows:

- with probability  $u_2/(u_1 + u_2)$ , set  $Y = y + u_1 r$ ;
- with the remaining probability of  $u_1/(u_1 + u_2)$ , set  $Y = y - u_2 r$ .

Briefly,  $u_1$  and  $u_2$  are chosen positive and just large enough so that we maintain our five invariants, while making at least one *more*  $A_{i,t}$ ,  $B_{i,h,t}$ , or  $Y_{i,j,t}$  integral at the end of iteration  $t$ .

**Proof that (C1)-(C4) is an under-determined system.** Let us call pair  $(i, j)$  *rounded* if  $y_{i,j,t-1}$  lies in  $\{0, 1\}$ , and *floating* otherwise (i.e., if  $y_{i,j,t-1}$  lies in  $(0, 1)$ ). Let  $R$  and  $F$  respectively denote the sets of rounded and floating pairs.

We next remove two types of redundant constraints from our system (C1)-(C4):

**(R1)** Given the constraints (C4), we can remove any constraint in (C1), (C2), or (C3) in which all pairs  $(i, j)$  that appear in the LHS of the constraint, lie in  $R$ : such constraints are redundant given (C4), and are removed.

**(R2)** If the constraint for  $i$  in (C2)—if it appears in (C2)—is linearly dependent on the constraints for  $(i, \cdot)$  in (C3),<sup>7</sup> then we say that  $i$  is *(C2)-redundant* and remove the constraint for  $i$  in (C2).

Clearly, removal of these redundant equalities does not change our system. *From now on, (C1)-(C4) refers to the reduced system after the removal of these redundant constraints.*

We next define certain numbers of floating indices after the removals in (R1) and (R2). If the constraint for  $j$  appears in (C1), let  $N_1(j)$  denote the number of floating pairs  $(\cdot, j)$  in the LHS of this constraint. The second number of floating indices is defined in a slightly-more-refined manner: if the constraint for  $i$  appears in (C2), let  $N_2(i)$  denote the number of floating pairs  $(i, \cdot)$  in the LHS of this constraint, *that do not appear in the LHS of any of the constraints for  $(i, \cdot)$  in (C3)*. Next, if the constraint for  $(i, h)$  appears in (C3), let  $N_3(i, h)$  denote the number of floating pairs  $(i, \cdot)$  in the LHS of this constraint.

We next make three useful observations. Note first that by (R1), we have  $N_1(j)$  and  $N_3(i, h)$  are positive; in fact, since the constants in the RHS of (C1) and (C3) are all *integers*, a moment's thought reveals that each of  $N_1(j)$  and  $N_3(i, h)$  is *at least two*. Second, we claim that each  $N_2(i)$  is at least two as well. Indeed, since the constraint for  $i$  appears in (C2), we must have that this constraint is linearly independent of (in particular, is not the sum of) the constraints for  $(i, \cdot)$  in (C3); hence, the LHS of the constraint for  $i$  in (C2) must have at least one variable not covered by the constraints for  $(i, \cdot)$  in (C3), implying that  $N_2(i)$  is positive as well. The fact that the constants in the RHS of (C2) are integers, again implies that  $N_2(i) \geq 2$ . Third, suppose  $a_{i,t-1}$  is *not* an integer,

<sup>7</sup>This simply means here that the constraint for  $i$  in (C2) is the sum of the constraints for  $(i, \cdot)$  in (C3), since the latter are all supported on pairwise-disjoint sets of variables.

in which case we will say  $i$  is “(C2)-not-integral.” In this case, it is easy to see that there must exist some  $h$  such that  $z_{i,h}$  *does not* appear in the LHS of any of the constraints (C2) and (C3).

We are finally ready to prove that (C1)-(C4) is underdetermined. Note that the number of constraints in (C4) is  $|R|$ , and that the total number of variables is  $|R| + |F|$ . Let  $n_1$ ,  $n_2$ , and  $n_3$  denote the respective numbers of constraints in (C1), (C2), and (C3). Recall again that these (C1)-(C3) refer to the system after applying the removal steps (R1) and (R2).

Since each variable in (C1) appears only once in (C1) and since  $N_1(j) \geq 2$ , we get that

$$|F| \geq 2n_1. \quad (10)$$

Similarly, from the consideration of (C2) and (C3) along with the observation above about indices  $i$  that are (C2)-not-integral, we get

$$|F| \geq 2(n_2 + n_3) + 1 \text{ if there is a (C2)-not-integral } i \quad (11)$$

$$\geq 2(n_2 + n_3) \text{ otherwise} \quad (12)$$

Thus, by averaging the above and using the fact that  $|F|$  is an integer, we obtain

$$|F| \geq n_1 + n_2 + n_3 + 1 \text{ if there is a (C2)-not-integral } i \quad (13)$$

$$\geq n_1 + n_2 + n_3 \text{ otherwise} \quad (14)$$

Now, if our system is *not* under-determined, we must have that the total number of constraints, which is  $n_1 + n_2 + n_3 + |R|$ , is at least as large as the total number of variables  $|F| + |R|$ . We see from (13) that this is impossible if there is a (C2)-not-integral  $i$ . Thus we may assume that there is **no** (C2)-not-integral  $i$ , and hence that the number of constraints exactly equals the number of variables. We will next show that this system is linearly dependent, which, from the fact that the number of constraints equals the number of variables, will show that the system is under-determined. Assume for convenience that no constraint in (C2) for any  $i$  was removed in (R2) (if not, we simply replace this constraint by the sum of the constraints for  $(i, \cdot)$  in (C3), in the following argument). Then, a moment’s reflection shows that the sum of the constraints in (C1) equals the sum of the constraints in (C2), yielding the desired linear dependence.