Welfare-Centric Clustering

Claire Jie Zhang[‡]
University of Washington

Seyed A. Esmaeili[‡]
University of Chicago

Jamie Morgenstern
University of Washington
Amazon AWS AI/ML

Abstract

Fair clustering has traditionally focused on ensuring equitable group representation or equalizing group-specific clustering costs. However, Dickerson et al. (2025) recently showed that these fairness notions may yield undesirable or unintuitive clustering outcomes and advocated for a welfare-centric clustering approach that models the utilities of the groups. In this work, we model group utilities based on both distances and proportional representation and formalize two optimization objectives based on welfare-centric clustering: the Rawlsian (Egalitarian) objective and the Utilitarian objective. We introduce novel algorithms for both objectives and prove theoretical guarantees for them. Empirical evaluations on multiple real-world datasets demonstrate that our methods significantly outperform existing fair clustering baselines.

1 INTRODUCTION

Algorithmic decision making has proliferated across various domains, from personalized advertising, recommender systems, and hiring to even more consequential and critical applications such as recidivism prediction, loan approval, and kidney exchange Tejaswini et al. (2020); Arun et al. (2016); Purohit et al. (2019); Berk et al. (2021); Liu et al. (2011); Berk and Bleich (2013); Awasthi and Sandholm (2009); Aziz et al. (2021); McElfresh et al. (2019); Raghavan et al. (2020). This widespread use has led to many documented examples where intuitive and agreed upon notions of fairness have been violated Kearns and Roth (2019); O'neil (2017). As a result fairness has become a major consideration in machine learning and algorithm design.

Clustering, a fundamental and classical problem in both machine learning and operations research, has received significant attention as a case study for operationalizing various notions of fairness Chierichetti et al. (2017); Bercea et al. (2018); Bera et al. (2019); Ahmadian et al. (2019); Ghadiri et al. (2021); Abbasi et al. (2021); Makarychev and Vakilian (2021); Jung et al. (2019); Chen et al. (2019). These works consider standard clustering objectives (e.g., k-means or k-median) and solve them subject to a constraint on an operationalization of some fairness notion. For example, the most prominent notion in fair clustering is that of *proportional mixing* where the standard clustering objective is minimized subject to each cluster having equitable representation of each demographic group.[*]

While proportional mixing as well as other fairness notions in clustering are well-motivated[†], it has recently been shown that the application of these notions can lead to clustering outcomes which are unintuitive and result in poor solutions both globally and for each group separately Dickerson et al. (2025). Take as an example the clustering instance in Figure 1 where the red and blue colors denote two different demographic groups. In this instance, the top two groupings of points are far from each other relative to the other points: clustering them with any points outside those groupings leads to high distance costs for those points. So, any proportionally mixed clustering will lead to large distances for all of the points in those top two

[*]Equal contribution.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

[*]For example, if the dataset consists of demographic groups A, B, and C with proportions 50%, 30%, and 20%, respectively, then for each cluster the percentage of points from groups A, B, and C should be around 50%, 30%, and 20%, respectively.

[†]Specifically, for proportional mixing, since each group is proportionally represented in each cluster, this would ensure that any outcome assigned purely by cluster assignment will affect each group proportionally.

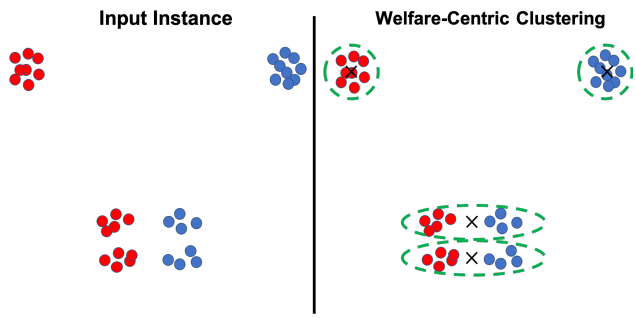


Figure 1: Instance and its welfare optimal clustering.

groupings.[‡] As noted in Dickerson et al. (2025) and shown in Figure 1, a *welfare-centric* clustering which quantifies the utilities of the red points and blue points separately and optimizes for them jointly would overcome these issues and produce clusterings where the upper groupings would be clustered separately and the lower groupings would cluster together. While the instance presented in the figure is simplified, it is representative of many settings where balancing the proportional mixing and the distances should be handled more carefully and where utility-aware clusterings are much more preferable. In this paper, we give a formalization of welfare-centric clustering and introduce algorithms with provable guarantees for it.

Our Contributions and Outline. In Section 2, we review the related works and demonstrate the need for a clustering notion based on utilities and welfare. In Section 3, we formalize the notions for group utility and introduce two formal notions of welfare. Specifically, we introduce the Rawlsian (Egalitarian) objective as well as the Utilitarian objective. In Section 4, we introduce our novel algorithms for both objectives. Interestingly, although our algorithms for both objectives share a similar format, the details of each have to be carefully adjusted to optimize its own objective. In Section 5, we implement our algorithms and show superior performance over a collection of baselines on real world datasets. Due to space limits all of the proofs are pushed to the Appendix.

2 RELATED WORK

As mentioned the most prominent notion in fair clustering is that of proportional color mixing where each cluster is constrained to have proportional representations of the groups. This notion was originally introduced in Chierichetti et al. (2017) and was then generalized to satisfy arbitrary lower and upper bounds as well as various clustering objectives Bercea et al. (2018); Bera et al. (2019); Kleindessner et al. (2019); Knittel et al. (2023). Esmaeili et al. (2021) noted that satisfying proportional color mixing might come at an unbounded increase to the clustering cost and therefore introduced the ***F**air **C**lustering under a **B**ounded **C**ost* (FCBC) problem where an upper bound on the clustering cost is given and then proportional color mixing is essentially maximized across the clusters. Although, this notion could mix the bottom clusters in Figure 1 it still leads to inequitable solutions where one group may unreasonably have a strictly lower utility than the other. This happens since FCBC is only using the clustering cost instead of formulating a welfare function based on clear utility assignments. See Appendix F for a detailed discussion. Additionally, Hakim et al. (2024) extends Esmaeili et al. (2021) by finding the Pareto frontier of the clustering cost vs the proportional violations. However, similar to Esmaeili et al. (2021) it has the same issues since it only considers the clustering cost instead of the welfare. Further, the algorithm of Hakim et al. (2024) has a run-time of $\Omega(n^k)$ where n is the number points and k is the number of clusters, making it highly impractical.

Another important notion is that of *socially fair* clustering Abbasi et al. (2021); Ghadiri et al. (2021); Makarychev and Vakilian (2021) where the average clustering cost for each group is found and then the maximum average across all groups is minimized. While this notion is well-motivated, it completely ignores the proportional representation of the groups within the formed clustering. For example, in the instance of Figure 1, the optimal socially fair clustering would not form clusters where the groups are mixed. This is a shortcoming that our framework overcomes since the average clustering cost of a group as well as its proportional representation are optimized for simultaneously by including them in the utility.

A related line of work studies probabilistic fair clustering, where group memberships are uncertain and represented probabilistically rather than deterministically Esmaeili et al. (2020). While this changes the modeling of protected attributes, it does not by itself address the issue that motivates our welfare-centric approach: a clustering can satisfy a probabilistic version of proportional mixing and still allocate distance costs inequitably across groups. In that sense, probabilistic fair clustering and our framework are orthogonal: the former changes the representation of group membership, whereas the latter changes the objective to

[‡]In general, shorter distances to centers are more favorable. This holds if the application was a machine learning or an operations research application. In the former, shorter distances correspond to better representation by the center and in the latter they correspond to shorter travel distance to the center.

explicitly optimize welfare.

Another complementary line of work studies individual-fairness in clustering, inspired by the principle that similar individuals should be treated similarly Dwork et al. (2012). For example, Jung et al. (2019) and Mahabadi and Vakilian (2020) define fairness at the level of points by requiring each point to have access to a center within an appropriate radius, or more generally by comparing the service quality received by similar individuals. These formulations typically do not use group labels and do not directly model trade-offs between different groups' outcomes. In contrast, our framework aggregates group-level disutilities, combining distance and proportional-representation terms through a social-welfare objective. This makes our approach more suitable for settings where the goal is to reason explicitly about distributional equity across protected groups rather than similarity-based fairness across individuals.

Finally, there has been significant work in supervised learning that emphasized the importance of formalizing utility/welfare and having them as the main objective to optimize for instead of imposing simple constraints such as demographic parity Heidari et al. (2018); Chen and Hooker (2021); Chohlas-Wood et al. (2024); Mladenov et al. (2020); Hu and Chen (2020); Cousins (2021); Heidari et al. (2019); Rosenfeld and Xu (2025). Despite the significant attention welfare-based formulations have received in supervised learning, they have been largely ignored in clustering. The exception being the recent work of Dickerson et al. (2025). In particular, Dickerson et al. (2025) actually introduced an example similar to Figure 1 and proved that proportional mixing and socially fair clustering would in fact produce strictly less welfare in comparison to a welfare-centric approach. In this paper, we formulate welfare in a manner largely similar to Dickerson et al. (2025) with the main difference being that we formulate utility at the level of the group instead of the individual and that we define a specific form to calculate the proportional mixing (or violation of proportional mixing). Additionally, we note that while Dickerson et al. (2025) pointed out the importance of considering welfare in clustering, it gave no algorithms or even heuristics for obtaining welfare-centric clusterings.

3 PROBLEM STATEMENT

We are given a set of points $\mathcal{C}$ representing individuals to be clustered with $|\mathcal{C}| = n$. Each point $j \in \mathcal{C}$ has a color $h \in \mathcal{H}$ where each color is associated with a specific demographic group. We assume that each point belongs to exactly one group. We therefore, define the color assignment function $\chi : \mathcal{C} \to \mathcal{H}$ that assigns points to colors, i.e., $\chi(j)$ is the color of point j. The subset of individuals with color h is denoted by $\mathcal{C}^h$, and the total number of such individuals is $n_h = |\mathcal{C}^h|$. Accordingly, the proportion of color h in the dataset is $r_h = \frac{n_h}{n}$. Additionally, we have a distance function $d(.,.)$ which defines a metric over the space.

A clustering solution is represented by the pair (S, ϕ), where S is the set of selected centers and $\phi : \mathcal{C} \to S$ is an assignment function that maps each point to a center, thereby forming the clusters. The distance between a point j and its assigned center is $d(j, \phi(j))$. Of particular interest to us are the k-median and k-means clustering objectives which can be formally written as:

$$\min_{S,\phi} \sum_{j \in \mathcal{C}} d^p(j, \phi(j)) \tag{1}$$

Where in the above we set $p = 1$ and $p = 2$ for the k-median and k-means objectives, respectively. Note that the assignment function ϕ is trivial in the above objectives since one would simply assign each point to its closest center to achieve a minimum clustering cost. However, that is not the case when constraints are imposed on the clustering objectives.[§]

Additionally, we use C_i to denote the set of points assigned to center $i \in S$, i.e., C_i is the ith cluster. Further, $C_i^h = C_i \cap \mathcal{C}^h$, i.e., the subset of those points with color h in C_i. With this notation, we may instead write the clustering cost as $\sum_{i \in S} \sum_{j \in C_i} d^p(j, i)$.

Before introducing our problem we formally introduce the following two clustering objectives which will be relevant to us. The first is *socially fair clustering* (Abbasi et al., 2021; Ghadiri et al., 2021; Makarychev and Vakilian, 2021) which is defined as:

Socially Fair Clustering:

$$\min_{S,\phi} \max_{h \in \mathcal{H}} \frac{1}{|\mathcal{C}^h|} \sum_{j \in \mathcal{C}^h} d^p(j, \phi(j)) \tag{2}$$

This objective amounts to minimizing the maximum average clustering cost across the groups. The second clustering objective is *weighted clustering* (Har-Peled and Mazumdar, 2004; Charikar et al., 1999) where each point $j \in \mathcal{C}$ has a specific weight $w_j \in \mathbb{R}_{>0}$ associated with it, formally we have:

Weighted Clustering:

$$\min_{S,\phi} \sum_{j \in \mathcal{C}} w_j \; d^p(j, \phi(j)) \tag{3}$$

Note that unlike socially fair clustering, weighted clustering is well-defined even if the points have no color assignments.

[§]For example, if we constraint each cluster to have a maximum number of points Li (2017); Chuzhoy and Rabani (2005), then some points maybe assigned to further away centers if their closest center has too many points assigned to it.

Our approach is focused on optimizing welfare (optimizing an aggregate of the average group utilities). First, clearly in a clustering solution, the distance between a point and its assigned center will have a great effect on the disutility[¶], with shorter distances leading to smaller disutilities. Therefore, for a given solution (S, ϕ) and a given group $h \in \mathcal{H}$, we define $\mathcal{D}_h$ as follows:

$$\mathcal{D}_h(S, \phi) = \sum_{j \in \mathcal{C}^h} d^p(j, \phi(j)) \tag{4}$$

Additionally, since the proportional representation of a group in a given clustering has a significant impact on the disutility as well[‖], we define $\mathcal{V}_h(S, \phi)$ as follows:

$$\mathcal{V}_h(S, \phi) = \sum_{i \in S} |C_i| \cdot \nu(h, i) \tag{5}$$

Where $\nu(h, i)$ is the amount of proportional violation color h experiences in cluster i. Formally, $\nu : \mathcal{H} \times S \to \mathbb{R}_{\geq 0}$ and is defined as follows:

$$\nu(h, i) = \max\left\{\frac{|C_i^h|}{|C_i|} - (r_h + \alpha_h),\ (r_h - \beta_h) - \frac{|C_i^h|}{|C_i|},\ 0\right\} \tag{6}$$

where α_h and β_h are upper and lower proportional violations respectively. To understand (6), note that if color h is proportionally represented, i.e., $r_h + \alpha_h \geq \frac{|C_i^h|}{|C_i|} \geq r_h - \beta_h$, then $\nu(h, i) = 0$. On the other hand, if the h is over-represented, i.e., $\frac{|C_i^h|}{|C_i|} > r_h + \alpha_h$, then would have $\nu(h, i) = \frac{|C_i^h|}{|C_i|} - (r_h + \alpha_h)$. Similarly, $\nu(h, i) = (r_h - \beta_h) - \frac{|C_i^h|}{|C_i|}$ when there is an under-representation. Note that α_h and β_h are there to relax the violation values for the upper and lower proportions desired for color h, respectively. For example, if we set $\alpha_h \geq 1 - r_h$ and $\beta_h = r_h$, then we only obtain positive violations if color h is under-represented and never when it is over-represented.

It is also important to note that in (5) the amount of violation a group (color) experiences in cluster i is weighted by the cluster size $|C_i|$. This choice is well-motivated, since one would expect that a group would not be affected significantly by an disproportional representation in a cluster if the cluster size is not significant.

[¶]Since smaller distance leads to better utility, we find it more natural to use disutilities instead of utilities. Therefore we will minimize an aggregate of the disutilities instead of maximizing it.

[‖]As noted earlier, this is the main motivation in the notion of proportional mixing.

Based on the above, we define the average disutility of a group D_h as:

$$D_h(S, \phi) = \frac{\lambda \cdot \mathcal{D}_h(S, \phi) + (1 - \lambda) \cdot \mathcal{V}_h(S, \phi)}{|\mathcal{C}^h|} \tag{7}$$

where we have $\lambda \in [0, 1]$. Equation (7) is a very natural assignment for a group's disutility, simply stating that distance and proportional violation should be minimized simultaneously. The role of λ is to decide the weighting between the distance and the proportional violation.

All we have left is to set an objective that minimizes an aggregate of the disutilities. The two most standard objectives from welfare economics Brandt et al. (2016) are the *Rawlsian* (egalitarian) objective $R = \max_{h \in \mathcal{H}} D_h(S, \phi)$ and the *Utilitarian* objective $U(S, \phi) = \sum_{h \in \mathcal{H}} D_h(S, \phi)$, therefore we will focus on minimizing one of the following two objectives:

$$\textbf{Rawlsian:} \quad \min_{S,\phi} R(S, \phi) = \min_{S,\phi} \max_{h \in \mathcal{H}} D_h(S, \phi) \tag{8}$$

$$\textbf{Utilitarian:} \quad \min_{S,\phi} U(S, \phi) = \min_{S,\phi} \sum_{h \in \mathcal{H}} D_h(S, \phi) \tag{9}$$

Clearly, for the case where $\lambda = 1$ the Rawlsian objective R is equal to the socially fair objective (2). Further, for $\lambda = 1$ it is not difficult to see that the the Utilitarian objective U would be a special instance of weighted clustering (3) where for each point $j \in \mathcal{C}$ we would have $w_j = \frac{1}{|C^{\chi(j)}|}$ where $\chi(j)$ is the color of point j.

Other aggregators of the group disutilities are also possible. For example, after mapping disutilities to utilities, one could define a Nash-social-welfare-type objective using the product of group utilities. We leave the algorithmic study of such variants to future work.

As a final notional note, we will use $\mathrm{OPT_R}$ and $\mathrm{OPT_U}$ for the optimal values of the Rawlsian and Utilitarian objectives, respectively.

4 ALGORITHMS AND THEORETICAL GUARANTEES

It is not difficult to see that ordinary clustering is a special case of both the Rawlsian and the Utilitarian objectives. Therefore, since both the k-median and k-means problems are NP-hard Bhattacharya et al. (2020); Charikar et al. (1999); Awasthi et al. (2015), it follows immediately that both the Rawlsian and Utilitarian objectives are NP-hard as well. We therefore design approximation algorithms for them that would

be guaranteed to run in polynomial instead of exponential time. Our approach is modular and is similar for both objectives. Specifically, for both objectives we start by running an algorithm to find the set of centers and then having fixed the set of centers we find an assignment from points to centers. Although such a "two stage" approach of choosing centers and then solving for the assignment was followed before in fair clustering Bercea et al. (2018); Bera et al. (2019); Esmaeili et al. (2020, 2021), the center selection is done using a *vanilla* clustering algorithm and the rounding can be done mostly through a standard structure of a min-cost max-flow network. In our case, this approach has to be significantly adjusted. Specifically, as will be shown in the next two subsections, center selection has to be done using socially fair and weighted clustering algorithms for the Rawlsian and Utilitarian objectives, respectively. Further, the min-cost max-flow rounding has to be adjusted significantly for the Rawlsian objective. For a more clear presentation, we will set $\gamma_p = 2^{p-1}(2^{p-1}+1)$ and $\gamma_{p'} = 2^{2(p-1)}$.

4.1 Algorithm for the Rawlsian Objective: RawlsianAlg

Algorithm 1 RawlsianAlg

1: **Input:** Set of points $\mathcal{C}$, number of centers k .
2: $S_{\mathbf{SF}} = \text{SociallyFairAlg}(\mathcal{C}, k)$.
3: Solve LP (10) to obtain the solution $\mathbf{x}^{\text{frac}}$.
4: Set $\mathbf{x}^{\text{integ}} = \text{Rawlsian-Rounding}(\mathbf{x}^{\text{frac}})$ and form the assignment ϕ using $\mathbf{x}^{\text{integ}}$.
5: **Return** $(S_{\mathbf{SF}}, \phi)$.

Algorithm 1 shows the pseudo-code of RawlsianAlg which is our approximation algorithm for the Rawlsian objective. The set of centers $S_{\mathbf{SF}}$ is set to the output of a SociallyFairAlg which can be any α-approximation algorithm for the socially fair objective. The intuition behind this choice lies in the fact that the Rawlsian objective is $\min_{h\in\mathcal{H}} D_h = \frac{\lambda\,\mathcal{D}_h + (1-\lambda)\,\mathcal{V}_h}{|\,\mathcal{C}^h\,|}$, which ignoring the term $(1-\lambda)\,\mathcal{V}_h$ is identical to a λ multiple of the socially fair objective (2). In fact a formal connection can be proved. Specifically, if $S_{\mathbf{SF}}$ is the set of centers chosen by an α-approximation algorithm for the socially fair objective then there exists an assignment ϕ' that leads to a solution with a bounded increase to the approximation factor α for the Rawlsian objective. The lemma below gives the formal statement.

Lemma 4.1. *Using the set of centers $S_{\mathbf{SF}}$ chosen by an α-approximation algorithm for the socially fair objective, there exists an assignment ϕ' such that $(S_{\mathbf{SF}}, \phi')$ leads to a $(\gamma_p + \gamma'_p\alpha)$-approximate solution for the Rawlsian objective.*

Note that the above lemma leads to an approximation factor of $(2+\alpha)$ and $(6+4\alpha)$ for $p=1$ and $p=2$, respectively.

It is important to note that the above lemma is non-constructive, i.e., although it establishes the existence of such an assignment ϕ' it does not give a procedure for finding it. Nevertheless, the lemma will be essential to constructing the assignment algorithm and proving guarantees for it.

It is actually still NP-hard to optimize the Rawlsian objective even with the fixed set of centers $S_{\mathbf{SF}}$ (see Appendix C for details). Our approach to finding the assignment function ϕ starts by solving the following linear program (LP):

$$\min \quad z \tag{10a}$$

subject to

$$\forall h \in \mathcal{H}: \ \frac{1}{|\,\mathcal{C}^h\,|}\left(\lambda \sum_{j\in\mathcal{C}^h}\sum_{i\in S_{\mathbf{SF}}} d^p(i,j)\,x_{ij} + (1-\lambda)\sum_{i\in S_{\mathbf{SF}}} t_{i,h}\right) \le z \tag{10b}$$

$$\forall i \in S_{\mathbf{SF}}, \forall h \in \mathcal{H}: (r_h - \beta_h)\cdot \sum_{j\in\mathcal{C}} x_{ij} - \sum_{j\in\mathcal{C}^h} x_{ij} = u_{i,h} \tag{10c}$$

$$\forall i \in S_{\mathbf{SF}}, \forall h \in \mathcal{H}: \sum_{j\in\mathcal{C}^h} x_{ij} - (r_h + \alpha_h)\cdot \sum_{j\in\mathcal{C}} x_{ij} = o_{i,h} \tag{10d}$$

$$\forall i \in S_{\mathbf{SF}}, \forall h \in \mathcal{H}: u_{i,h} \le t_{i,h}, \quad o_{i,h} \le t_{i,h}, \quad 0 \le t_{i,h} \tag{10e}$$

$$\forall j \in \mathcal{C}: \sum_{i\in S_{\mathbf{SF}}} x_{ij} = 1, \quad x_{ij} \in [0,1] \tag{10f}$$

The LP is more easily interpreted by considering integral values (in $\{0,1\}$) of the variables x_{ij} instead of the fractional values (in $[0,1]$). In particular, x_{ij} decides if point $j \in \mathcal{C}$ is assigned to center $i \in S_{\mathbf{SF}}$ and since each point must be assigned to exactly one center, the constraints of (10f) must hold. Further, when there is an under-representation of color h in cluster i, the variable $u_{i,h}$ in (10c) would equal the amount of under-representation (i.e., $|C_i|\,\nu(h,i)$). On the other hand, if there is no under-representation of color h in cluster i then we would have $u_{i,h} \le 0$. A similar comment can be made about $o_{i,h}$ (10d) for the over-representation of color h in cluster i. Therefore, the variable $t_{i,h}$ is introduced and in constraint (10e) we have $t_{i,h}$ lower bounded by 0, $u_{i,h}$ and $o_{i,h}$ ensuring that it is equal to the value of $|C_i|\,\nu(h,i)$ under all possible scenarios. Moreover, the left hand side of constraint (10b) equals the disutility D_h of group h and we have the variable z lower bounded by the disutilities of all groups. This

ensures that the objective $\min z$ is minimizing the maximum disutility, i.e., the Rawlsian objective. In fact, the LP is exactly a relaxation of of the integer program that would find the optimal Rawlsian assignment given the centers $S_{\mathbf{SF}}$. Based on the above the following lemma must hold.

Lemma 4.2. *Denoting the optimal value of LP* (10) *by* OPT_R^{LP} *then* $\mathrm{OPT}_R^{LP} \leq (\gamma_p + \gamma'_p \alpha)\,\mathrm{OPT}_R$.

The essential point behind the above lemma is that the LP includes all possible assignments to the centers $S_{\mathbf{SF}}$ and would thus include the assignment ϕ' from Lemma 4.1 leading to a cost of at most $(\gamma_p + \gamma'_p \alpha)\,\mathrm{OPT}_{\mathrm{R}}$.

Although we can obtain the optimal LP solution (assignment) $\mathbf{x}^{\mathrm{frac}}$, the values in the entries x_{ij}^{frac} would in general be fractional instead of integral and therefore cannot be used directly. One of the standard rounding methods used in fair clustering is based on constructing a min-cost max-flow network Bercea et al. (2018); Esmaeili et al. (2021); Dickerson et al. (2024). Essentially, in this constructed network lower and upper bounds on arcs are set to ensure that the total amount of points from each group in each cluster do not vary from the fractional LP solution. Further, the costs on the arcs are set to the distance values between the points and the centers, ensuring that an assignment that minimizes the total distance aggregated across points from different colors is found. However, since the Rawlsian objective minimizes the maximum disutility across the colors instead of aggregating them as a sum, we modify the min-cost max-flow scheme. Figure 2 shows a schematic of our rounding which we call Rawlsian-Rounding. In particular, the rounding is done $|\mathcal{H}|$ many times, each time over the points of one color in isolation from the rest of the points in the dataset. The full details of the network and Rawlsian-Rounding are given in Appendix D.1. Denoting the quantities associated with the LP solution with the superscript ^{frac} and those of the integral solution with ^{integ}, the following lemma can be established for Rawlsian-Rounding.

Lemma 4.3. *Let* $\mathbf{x}^{integ} =$ Rawlsian-Rounding$(\mathbf{x}^{frac})$ *then the following holds.*

$$\forall h \in \mathcal{H}: \ \mathcal{D}_h^{\mathrm{integ}} \leq \mathcal{D}_h^{\mathrm{frac}} \tag{11}$$

$$\forall i \in S_{\boldsymbol{SF}}, \ \forall h \in \mathcal{H}: \left\lfloor \sum_{j \in \mathcal{C}^h} x_{ij}^{\mathrm{frac}} \right\rfloor \leq \sum_{j \in \mathcal{C}^h} x_{ij}^{\mathrm{integ}} \leq \left\lceil \sum_{j \in \mathcal{C}^h} x_{ij}^{\mathrm{frac}} \right\rceil \tag{12}$$

In the above, Inequality (11) ensures that the value of $\mathcal{D}_h$ does not increase in the rounded integral solution for any color. Further, Inequality (12) ensures that the change in the total amount of color assigned to each cluster cannot vary by more than 1 (since it is between the floor and the ceiling). Based on the above, the following lemma can be established.

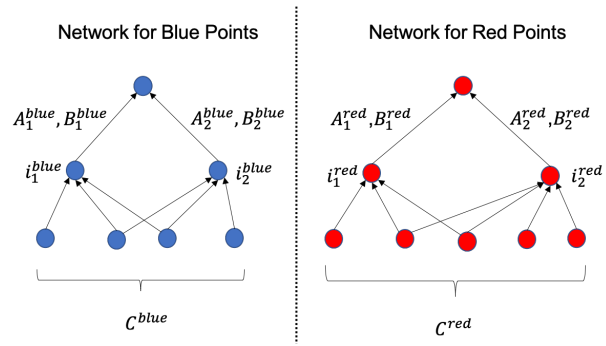


Figure 2: Rawlsian-Rounding for the case with two centers $S = \{i_1, i_2\}$. For any color h and center i we set flow lower and upper bound on the arcs between the centers and the sink to $A_i^h = \left\lfloor \sum_{j \in \mathcal{C}^h} x_{ij}^{\mathrm{frac}} \right\rfloor$, $B_i^h = \left\lceil \sum_{j \in \mathcal{C}^h} x_{ij}^{\mathrm{frac}} \right\rceil$, respectively. Other arcs, have a flow upper bound of 1 and a lower bound of 0. See Appendix D.1 for full details.

Lemma 4.4. *The rounded Rawlsian solution has a cost of at most* $\mathrm{OPT}_R^{LP} + c_R$, *where*

$$c_R = \left(\frac{|\mathcal{H}| + 1}{\min_{h \in \mathcal{H}} r_h} \right) \frac{k}{n}.$$

From Lemmas 4.2 and 4.4, we can establish the the main guarantee for RawlsianAlg.

Theorem 4.5. RawlsianAlg *(Algorithm 1) returns a solution of Rawlsian cost at most* $(\gamma_p + \gamma_{p'} \alpha)\,\mathrm{OPT}_R + c_R$.

Remark 4.6. Since the number of colors $|\mathcal{H}|$ and the ratio r_h of each color h is also a constant. Our final guarantee is $(\gamma_p + \gamma_{p'} \alpha)\,\mathrm{OPT}_{\mathrm{R}} + O(\frac{k}{n})$.

Although our final solution incurs an additional additive approximation of $O(\frac{k}{n})$ we note that it is not significant since in most applications we would have $n \gg k$, thus $\frac{k}{n}$ would be very small. Empirically, we do not observe any significant violations because of this additive factor.

4.2 Algorithm for the Utilitarian Objective: UtilitarianAlg

Our algorithm for the Utilitarian Objective is UtilitarianAlg whose pseudo-code is shown in Algorithm 2. Interestingly, although the details of UtilitarianAlg are significantly different from RawlsianAlg, the format remains very similar. In particular, UtilitarianAlg begins by assigning a weight of $w_j = \frac{1}{|\mathcal{C}^{\chi(j)}|}$ for each point $j \in \mathcal{C}$ and then it

runs WightedClusteringAlg which is an α approximation for the weighted clustering cost to find a set of centers $S_{\mathbf{W}}$. The rationale behind doing so is similar to the Rawlsian objective. Specifically, the Utilitarian objective is $\sum_{h\in\mathcal{H}} D_h = \sum_{h\in\mathcal{H}} \frac{\lambda\,\mathcal{D}_h + (1-\lambda)\,\mathcal{V}_h}{|\,\mathcal{C}^h\,|}$, which ignoring the $(1-\lambda)\,\mathcal{V}_h$ term would equal a λ weighted multiple of the weighted clustering cost of (3) with the weight of each point j set to $\frac{1}{|\,\mathcal{C}^{\chi(j)}\,|}$. Similar, to Lemma 4.1 we can establish the following lemma.

Lemma 4.7. *Using the set of centers $S_{\boldsymbol{W}}$ chosen by an α-approximation algorithm for the weighted clustering objective with $w_j = \frac{1}{|\,\mathcal{C}^{\chi(j)}\,|}$ $\forall j \in \mathcal{C}$, there exists an assignment ϕ' such that $(S_{\boldsymbol{W}}, \phi')$ leads to a $(\gamma_p + \gamma_{p'}\,\alpha)$-approximate solution for the Utilitarian objective.*

Lemma 4.7 shares significant similarities with Lemma 4.1 despite the fact that they address different objectives and are built using a different set of centers (socially fair vs weighted). In particular, the proof techniques are similar and both lemmas are non-constructive, i.e., the existence of assignment ϕ' is proved but without an algorithm for finding it. Furthermore, the increase in the approximation factor is identical and equal to $(2+\alpha)$ and $(6+4\alpha)$ for $p=1$ and $p=2$, respectively.

Having found the set of centers $S_{\mathbf{W}}$ we proceed to find the assignment. We follow a similar approach to that in the Rawlsian objective. In particular, we setup LP (13) below. The interpretation of the variables in LP (13) are similar to those in LP (10). Note however, we do not have a variable z and directly minimizes the sum of the disutilities in the objective.

$$\min \sum_{h\in\mathcal{H}} (\lambda \sum_{j\in\mathcal{C}^h} \sum_{i\in S_{\mathbf{W}}} \frac{d^p(i,j)x_{ij}}{|\,\mathcal{C}^h\,|} + (1-\lambda) \sum_{i\in S_{\mathbf{W}}} \frac{t_{i,h}}{|\,\mathcal{C}^h\,|}) \tag{13a}$$

subject to

$$\forall i \in S_{\mathbf{W}}, \forall h \in \mathcal{H} : (r_h - \beta_h) \cdot \sum_{j\in\mathcal{C}} x_{ij} - \sum_{j\in\mathcal{C}^h} x_{ij} = u_{i,h} \tag{13b}$$

$$\forall i \in S_{\mathbf{W}}, \forall h \in \mathcal{H} : \sum_{j\in\mathcal{C}^h} x_{ij} - (r_h + \alpha_h) \cdot \sum_{j\in\mathcal{C}} x_{ij} = o_{i,h} \tag{13c}$$

$$\forall i \in S_{\mathbf{W}}, \forall h \in \mathcal{H} : u_{i,h} \le t_{i,h}, \quad o_{i,h} \le t_{i,h}, \quad 0 \le t_{i,h} \tag{13d}$$

$$\forall j \in \mathcal{C} : \sum_{i\in S_{\mathbf{W}}} x_{ij} = 1, \quad x_{ij} \in [0,1] \tag{13e}$$

Based on similar reasoning we can establish a similar guarantee for the optimal value of LP (13) to Lemma 4.2 as shown below.

Lemma 4.8. *Denoting the optimal value of LP* (13) *by* OPT_U^{LP} *then* $\mathrm{OPT}_U^{LP} \le (\gamma_p + \gamma_p'\alpha)\,\mathrm{OPT}_U$.

Having found the optimal LP solution $\mathbf{x}^{\mathrm{frac}}$ we round it using the Utilitarian-Rounding min-cost max-flow network. Unlike Rawlsian-Rounding, the Utilitarian-Rounding is very similar to the standard min-cost max-flow rounding done in Bercea et al. (2018); Esmaeili et al. (2021); Dickerson et al. (2024) with the difference being that we set the cost of an arc between the vertex for a point j and a vertex for a center i to $\frac{d^p(j,i)}{|\,\mathcal{C}^{\chi(j)}\,|}$ instead of just $d^p(j,i)$ as usually done. See Appendix D.2 for the full details. Based on the construction of the the network, the following lemma can be proved.

Lemma 4.9. *Let* $\mathbf{x}^{integ} =$ Utilitarian-Rounding$(\mathbf{x}^{frac})$ *then the following holds.*

$$\sum_{h\in\mathcal{H}} \frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}_h^{integ} \le \sum_{h\in\mathcal{H}} \frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}_h^{frac} \tag{14}$$

$\forall i \in S_{\boldsymbol{W}}, \forall h \in \mathcal{H}$:

$$\left\lfloor \sum_{j\in\mathcal{C}^h} x_{ij}^{frac} \right\rfloor \le \sum_{j\in\mathcal{C}^h} x_{ij}^{integ} \le \left\lceil \sum_{j\in\mathcal{C}^h} x_{ij}^{frac} \right\rceil \tag{15}$$

$\forall i \in S_{\boldsymbol{W}}$:

$$\left\lfloor \sum_{j\in\mathcal{C}} x_{ij}^{frac} \right\rfloor \le \sum_{j\in\mathcal{C}} x_{ij}^{integ} \le \left\lceil \sum_{j\in\mathcal{C}} x_{ij}^{frac} \right\rceil \tag{16}$$

Inequality (14) guarantees that $\sum_{h\in\mathcal{H}} \frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}_h$ does not increase in the rounding. Further, Inequality (15) is identical to Inequality (12) from Rawlsian-Rounding and guarantees that the amount of points from any color assigned to each cluster in the integral solution does not vary from the fractional amount by more than 1. Interestingly, Utilitarian-Rounding has an additional inequality which is Inequality (16) that also guarantees that the integral cluster size (total number of points in the cluster) also does not vary by more than 1 from the fractional cluster size. We can now establish the following lemma.

Lemma 4.10. *The rounded Utilitarian solution has a cost of at most* $\mathrm{OPT}_U^{LP} + c_U$ *where* $c_U = \left(\frac{2k}{n} \cdot \sum_{h\in\mathcal{H}} \frac{1}{r_h}\right)$.

Based on Lemma 4.8 and Lemma 4.10, the main guarantee for UtilitarianAlg follows.

Theorem 4.11. UtilitarianAlg *(Algorithm 2) returns a solution of Utilitarian cost at most* $(\gamma_p + \gamma_{p'}\,\alpha)\,\mathrm{OPT}_U + c_U$.

Remark 4.12. Similar to the guarantee of Theorem 4.5 the number of colors $|\,\mathcal{H}\,|$ and the ratio r_h of each color h is also a constant. Therefore, c_U is a constant and our final guarantee is $(\gamma_p + \gamma_{p'}\,\alpha)\,\mathrm{OPT}_{\mathrm{R}} + O(\frac{k}{n})$.

It is worthwhile to note that Rawlsian-Rounding could have been applied instead of Utilitarian-Rounding. However, since Rawlsian-Rounding lacks the guarantee of Inequality (16) the constant c_U would have increased. Specifically, it would instead be $c_U = \left((|\mathcal{H}|+1)\frac{k}{n} \cdot \sum_{h\in\mathcal{H}} \frac{1}{r_h}\right)$ where 2 has been replaced by $(|\mathcal{H}|+1)$. Finally, we also see empirically that the final $O(\frac{k}{n})$ term does not have a significant effect.

Algorithm 2 UtilitarianAlg

1: **Input:** Set of points $\mathcal{C}$, number of centers k .
2: Set $w_j = \frac{1}{|\mathcal{C}^{x(j)}|}\ \forall j \in \mathcal{C}$.
3: $S_{\mathbf{W}} = \text{WightedClusteringAlg}(\mathcal{C}, k, \{w_j\}_{j\in\mathcal{C}})$.
4: Solve LP (13) to obtain the solution $\mathbf{x}^{\text{frac}}$.
5: Set $\mathbf{x}^{\text{integ}} = \text{Utilitarian-Rounding}(\mathbf{x}^{\text{frac}})$ and form the assignment ϕ using $\mathbf{x}^{\text{integ}}$.
6: **Return** $(S_{\mathbf{W}}, \phi)$.

Running time and dependence on subroutines. The approximation factors in our guarantees are parameterized by the factor α of the center-selection subroutine, namely a socially fair clustering algorithm for the Rawlsian objective and a weighted clustering algorithm for the Utilitarian objective. Both types of subroutines admit polynomial-time approximation algorithms in the literature. Once the centers are fixed, the remaining steps in our algorithms consist of solving the corresponding LP and performing min-cost-flow-based rounding, which are also polynomial-time. We report empirical wall-clock running times in Appendix H.10.

5 EXPERIMENTS

We evaluate the performance of RawlsianAlg and UtilitarianAlg across multiple datasets. Empirically, we observe that RawlsianAlg and UtilitarianAlg find solutions that significantly outperform the baselines in the Rawlsian and Utilitarian objectives. In the main paper, we include comparisons on three datasets for the two group setting. Additional experiments and results including those for multiple groups ($|\mathcal{H}| > 2$) are in Appendix H.

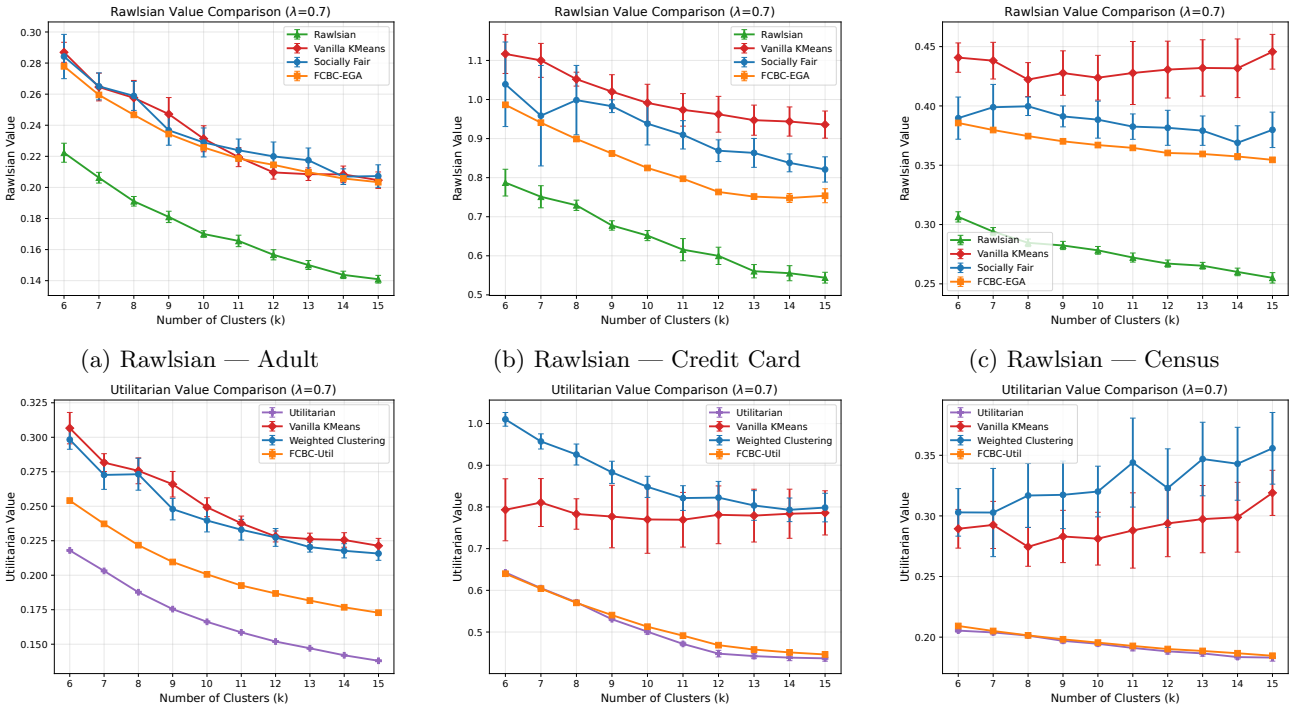
Algorithms. We compare RawlsianAlg with three other baselines: (1) Vanilla k-means; (2) Socially Fair: for the socially fair k-means objective; (3) FCBC-Rawl: the egalitarian variant in Esmaeili et al. (2021) which minimizes the maximum group fairness violation under bounded clustering cost. For UtilitarianAlg, we compare with: (1) Vanilla k-means; (2) Weight Clustering: weighted version of k-means with $w_j = \frac{1}{|\mathcal{C}^{x(j)}|}\ \forall j \in \mathcal{C}$; (3) FCBC-Util: the utilitarian variant in Esmaeili et al. (2021) which minimizes the sum of group fairness violation under bounded clustering distance. The distance measure is always the Euclidean distance. We set $\alpha_h = \beta_h = \delta r_h$, where $\delta = 0.01$. This ensures a fair comparison with FCBC algorithms which take parameter δ. Since FCBC algorithms are not designed to solve Rawlsian and Utilitarian objectives, for a fixed λ and for each number of center k, we run FCBC algorithms with a set of clustering cost upper bound parameter $\{1.05, 1.5, 2, 2.5, 3\}$ and report the minimum Rawlsian and Utilitarian objective value achieved. This extensive tuning effort makes the FCBC baselines competitive on some datasets as shown in Figure 3. See more details of the tuning procedure in Appendix H.

Hardware and Software. All experiments were conducted on a MacBook Pro (2.3GHz 2-core Intel i5, 8GB RAM) using Python 3.10. We use `KMeans` and its weighted option from `Scikit-learn` (Pedregosa et al., 2011) to find initial centers. Linear programs are solved using the CPLEX package (Nickel et al., 2022). Network-flow rounding is done via `NetworkX` (Hagberg et al., 2013). More details about hardware and implementation details can be found in Appendix H.

Datasets. We conduct experiments on four UCI datasets (Frank, 2010): **Adult**, **CreditCard**, **Census1990**** and **Bank**. We form the feature vectors by choosing a set of columns as features in each dataset. We also choose a specific column for group membership assignment. For **Adult**, **CreditCard**, and **Census1990** the features and group memberships are selected following Bera et al. (2019); Esmaeili et al. (2020) whereas for **Bank** we follow Backurs et al. (2019). **Adult**, **CreditCard**, and **Census1990** are sub-sampled such that the final ratio between the two group is 90% vs 10% and the total number of points is around $24,000$, $17,000$, and $11,000$ for each respectively.

Results. In Figure 3, we compare the Rawlsian and Utilitarian objective values for $\lambda = 0.7$. In each plot, the number of centers k is varied from 6 to 15. All curves in Figure 3 report mean objective value over multiple random seeds with 95% confidence intervals. For $\lambda = 0.7$ on sub-sampled data, we used 10 runs; for full-data experiments and all other λ values on sub-sampled data, we used 5 runs. The variance is small and the ranking of methods is stable across runs.

**Since two census datasets are included in the UCI repository, we include a link to the specific census dataset we used: `https://archive.ics.uci.edu/dataset/116/us+census+data+1990`



(a) Rawlsian — Adult (b) Rawlsian — Credit Card (c) Rawlsian — Census

(d) Utilitarian — Adult (e) Utilitarian — Credit Card (f) Utilitarian — Census

Figure 3: Comparison at $\lambda = 0.7$ across three datasets.

We observe that both RawlsianAlg and UtilitarianAlg outperform their corresponding baselines on all three datasets. In fact, in the **Adult** dataset our algorithms outperform the other baselines by significant margins. In Appendix H, we show plots for the full range of λ as well as settings with different ratios between the groups. There we observe our algorithms consistently find solutions that outperform the baselines. We also include additional plots that show in detail the clustering cost and representation violation for each group individually.

ACKNOWLEDGMENTS

JM and CJZ were supported in part by NSF ID 2045402 and a Simons Collaboration on the Theory of Algorithmic Fairness.

References

Mohsen Abbasi, Aditya Bhaskara, and Suresh Venkatasubramanian. Fair clustering via equitable group representations. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 504–514, 2021.

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

Kumar Arun, Garg Ishan, and Kaur Sanmeet. Loan approval prediction based on machine learning approach. *IOSR J. Comput. Eng*, 18(3):18–21, 2016.

Pranjal Awasthi and Tuomas Sandholm. Online stochastic optimization in the large: Application to kidney exchange. In *IJCAI*, volume 9, pages 405–411, 2009.

Pranjal Awasthi, Moses Charikar, Ravishankar Krishnaswamy, and Ali Kemal Sinop. The hardness of approximation of euclidean k-means. *arXiv preprint arXiv:1502.03316*, 2015.

Haris Aziz, Agnes Cseh, John P Dickerson, and Duncan C McElfresh. Optimal kidney exchange with immunosuppressants. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 21–29, 2021.

Arturs Backurs, Piotr Indyk, Krzysztof Onak, Baruch Schieber, Ali Vakilian, and Tal Wagner. Scalable fair clustering. In *International conference on machine learning*, pages 405–413. PMLR, 2019.

Suman Bera, Deeparnab Chakrabarty, Nicolas Flores, and Maryam Negahbani. Fair algorithms for clustering. *Advances in Neural Information Processing Systems*, 32, 2019.

Ioana O Bercea, Martin Groß, Samir Khuller, Aounon Kumar, Clemens Rösner, Daniel R Schmidt, and Melanie Schmidt. On the cost of essentially fair clusterings. *arXiv preprint arXiv:1811.10319*, 2018.

Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*, 50(1):3–44, 2021.

Richard A Berk and Justin Bleich. Statistical procedures for forecasting criminal behavior: A comparative assessment. *Criminology & Pub. Pol'y*, 12:513, 2013.

Anup Bhattacharya, Dishant Goyal, and Ragesh Jaiswal. Hardness of approximation of euclidean k-median. *arXiv preprint arXiv:2011.04221*, 2020.

Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of computational social choice*. Cambridge University Press, 2016.

Moses Charikar, Sudipto Guha, Éva Tardos, and David B Shmoys. A constant-factor approximation algorithm for the k-median problem. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 1–10, 1999.

Violet Xinying Chen and JN Hooker. Fairness through social welfare optimization. *arXiv preprint arXiv:2102.00311*, 2021.

Xingyu Chen, Brandon Fain, Liang Lyu, and Kamesh Munagala. Proportionally fair clustering. In *International conference on machine learning*, pages 1032–1041. PMLR, 2019.

Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. *Advances in neural information processing systems*, 30, 2017.

Alex Chohlas-Wood, Madison Coots, Henry Zhu, Emma Brunskill, and Sharad Goel. Learning to be fair: A consequentialist approach to equitable decision making. *Management Science*, 2024.

Julia Chuzhoy and Yuval Rabani. Approximating k-median with non-uniform capacities. In *SODA*, volume 5, pages 952–958, 2005.

Cyrus Cousins. An axiomatic theory of provably-fair welfare-centric machine learning. *Advances in Neural Information Processing Systems*, 34:16610–16621, 2021.

John Dickerson, Seyed Esmaeili, Jamie H Morgenstern, and Claire Jie Zhang. Doubly constrained fair clustering. *Advances in Neural Information Processing Systems*, 36, 2024.

John Dickerson, Seyed A Esmaeili, Jamie Morgenstern, and Claire Jie Zhang. Sok: Fair clustering: Critique, caveats, and future directions. In *2025 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pages 698–713. IEEE, 2025.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pages 214–226, 2012.

Seyed Esmaeili, Brian Brubach, Leonidas Tsepenekas, and John Dickerson. Probabilistic fair clustering. *Advances in Neural Information Processing Systems*, 33:12743–12755, 2020.

Seyed Esmaeili, Brian Brubach, Aravind Srinivasan, and John Dickerson. Fair clustering under a bounded cost. *Advances in Neural Information Processing Systems*, 34:14345–14357, 2021.

A Frank. Uci machine learning repository. irvine, ca: University of california, school of information and computer science. *http://archive. ics. uci. edu/ml*, 2010.

Mehrdad Ghadiri, Samira Samadi, and Santosh Vempala. Socially fair k-means clustering. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 438–448, 2021.

Aric Hagberg, Dan Schult, Pieter Swart, D Conway, L Séguin-Charbonneau, C Ellison, B Edwards, and J Torrents. Networkx. high productivity software for complex networks. *Webová strá nka https://networkx. lanl. gov/wiki*, 2013.

Rashida Hakim, Ana-Andreea Stoica, Christos H Papadimitriou, and Mihalis Yannakakis. The fairness-quality trade-off in clustering. *arXiv preprint arXiv:2408.10002*, 2024.

Sariel Har-Peled and Soham Mazumdar. On coresets for k-means and k-median clustering. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 291–300, 2004.

Hoda Heidari, Claudio Ferrari, Krishna Gummadi, and Andreas Krause. Fairness behind a veil of ignorance: A welfare analysis for automated decision making. *Advances in neural information processing systems*, 31, 2018.

Hoda Heidari, Michele Loi, Krishna P Gummadi, and Andreas Krause. A moral framework for understanding fair ml through economic models of equality of opportunity. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 181–190, 2019.

Lily Hu and Yiling Chen. Fair classification and social welfare. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pages 535–545, 2020.

Christopher Jung, Sampath Kannan, and Neil Lutz. A center in your neighborhood: Fairness in facility location. *arXiv preprint arXiv:1908.09041*, 2019.

Michael Kearns and Aaron Roth. *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press, 2019.

Matthäus Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. Guarantees for spectral clustering with fairness constraints. In *International conference on machine learning*, pages 3458–3467. PMLR, 2019.

Marina Knittel, Max Springer, John Dickerson, and MohammadTaghi Hajiaghayi. Fair, polylog-approximate low-cost hierarchical clustering. *Advances in Neural Information Processing Systems*, 36:44836–44846, 2023.

Shi Li. On uniform capacitated k-median beyond the natural lp relaxation. *ACM Transactions on Algorithms (TALG)*, 13(2):1–18, 2017.

Yuan Y Liu, Min Yang, Malcolm Ramsay, Xiao S Li, and Jeremy W Coid. A comparison of logistic regression, classification and regression tree, and neural networks models in predicting violent re-offending. *Journal of Quantitative Criminology*, 27:547–573, 2011.

Sepideh Mahabadi and Ali Vakilian. Individual fairness for k-clustering. In *International conference on machine learning*, pages 6586–6596. PMLR, 2020.

Yury Makarychev and Ali Vakilian. Approximation algorithms for socially fair clustering. In *Conference on Learning Theory*, pages 3246–3264. PMLR, 2021.

Duncan C McElfresh, Hoda Bidkhori, and John P Dickerson. Scalable robust kidney exchange. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1077–1084, 2019.

Martin Mladenov, Elliot Creager, Omer Ben-Porat, Kevin Swersky, Richard Zemel, and Craig Boutilier. Optimizing long-term social welfare in recommender systems: A constrained matching approach. In *International Conference on Machine Learning*, pages 6987–6998. PMLR, 2020.

Stefan Nickel, Claudius Steinhardt, Hans Schlenker, and Wolfgang Burkart. Ibm ilog cplex optimization studio—a primer. In *Decision Optimization with IBM ILOG CPLEX Optimization Studio: A Hands-On Introduction to Modeling with the Optimization Programming Language (OPL)*, pages 9–21. Springer, 2022.

Cathy O'neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown, 2017.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

Manish Purohit, Sreenivas Gollapudi, and Manish Raghavan. Hiring under uncertainty. In *International Conference on Machine Learning*, pages 5181–5189. PMLR, 2019.

Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. Mitigating bias in algorithmic hiring: Evaluating claims and practices. 2020.

Nir Rosenfeld and Haifeng Xu. Machine learning should maximize welfare, not (only) accuracy. *arXiv preprint arXiv:2502.11981*, 2025.

J Tejaswini, T Mohana Kavya, R Devi Naga Ramya, P Sai Triveni, and Venkata Rao Maddumala. Accurate loan approval prediction based on machine learning approach. *Journal of Engineering Science*, 11(4):523–532, 2020.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes] The complexity analysis for our algorithm follows what is standard in the literature. Further, Appendix H.10 includes plots showing our run-time.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes] Our code is submitted as part of the supplementary material.
2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes] We assume a metric distance function and a given partition of points into demographic groups, as specified in Section 3.
   (b) Complete proofs of all theoretical results. [Yes] The proofs of all claims, lemmas, and theorems are included in the Appendix.
   (c) Clear explanations of any assumptions. [Yes] We assume a metric distance function and a given partition of points into demographic groups, as specified in Section 3.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes] It is uploaded as part of the supplemental material.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes] Experimental details are given in the main paper in Section 5 and Appendix H.
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes] Main-text and appendix plots report mean objective values over multiple random seeds with 95% confidence intervals. For $\lambda = 0.7$ on sub-sampled data we use 10 runs; for full-data experiments and all other λ values on sub-sampled data we use 5 runs.
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes] The details of the computing infrastructure are given in Section 5.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Yes]
   (b) The license information of the assets, if applicable. [Not Applicable] The used code mostly follows standard licenses in academia, e.g., MIT License.
   (c) New assets either in the supplemental material or as a URL, if applicable. [Yes] Code is included in the supplemental material.
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Supplementary Materials: Welfare-Centric Clustering

A Useful Facts

Fact A.1. *For $p \in \{1, 2\}$ and any three arbitrary points a, b, and c with a distance measure satisfying the triangle inequality we have*

$$d^p(a, b) \leq 2^{p-1} \cdot \big(d^p(a, c) + d^p(c, b)\big) \tag{17}$$

Proof. For $p = 1$ we have $2^{p-1} = 1$ and the proof follows by the triangle inequality. For $p = 2$ we have by squaring both sides of the triangle inequality:

$$\begin{aligned} d^2(a, b) &\leq d^2(a, c) + d^2(c, b) + 2d(a, c)d(b, c) \\ &\leq 2(d^2(a, c) + d^2(c, b)) \end{aligned}$$

where in the above we used the fact that $d(a, c)d(b, c) \leq \frac{d^2(a,c)+d^2(c,b)}{2}$ by the AM-GM inequality. □

B Omitted Proofs

We restate our first lemma and give its proof.

Lemma 4.1. *Using the set of centers $S_{\boldsymbol{SF}}$ chosen by an α-approximation algorithm for the socially fair objective, there exists an assignment ϕ' such that $(S_{\boldsymbol{SF}}, \phi')$ leads to a $(\gamma_p + \gamma'_p\alpha)$-approximate solution for the Rawlsian objective.*

Proof. We denote by (S^*, ϕ^*) an optimal solution for the Rawlsian objective. $(S_{\mathbf{SF}}, \phi_{\mathbf{SF}})$ denote the solution of the α-approximation solution for the socially fair objective. We construct the assignment ϕ' as follows: for every $i^* \in S^*$, define $\text{nrst}(i^*) = \arg\min_{i \in S_{\mathbf{SF}}} d(i, i^*)$. For every $j \in \mathcal{C}$, define $\phi'(j) = \text{nrst}(\phi^*(j))$.

Before we delve into the details of the proof, we will denote by $\mathcal{D}^*_h$, $\mathcal{D}^{\mathbf{SF}}_h$, and $\mathcal{D}'_h$ the values of $\mathcal{D}_h(S^*, \phi^*)$, $\mathcal{D}_h(S_{\mathbf{SF}}, \phi_{\mathbf{SF}})$, and $\mathcal{D}_h(S_{\mathbf{SF}}, \phi')$, respectively. Similarly, for $\mathcal{V}^*_h$, $\mathcal{V}^{\mathbf{SF}}_h$, and $\mathcal{V}'_h$. Moreover, as noted earlier OPT_{R} is the optimal value of the Rawlsian objective. Further, we let OPT_{SF} be the optimal value of the socially fair objective when evaluated for the socially fair objective.

We start with the following claim.

Claim B.1.

$$\max_{h \in \mathcal{H}} \frac{\mathcal{D}^{\boldsymbol{SF}}_h}{|\,\mathcal{C}^h\,|} \quad \leq \quad \alpha \cdot \frac{1}{\lambda}\,\text{OPT}_R \tag{18}$$

Proof. Let $({S_{\mathbf{SF}}}^*, {\phi_{\mathbf{SF}}}^*)$ denote an optimal **SF**solution and let $\mathcal{D}^{\mathbf{SF}^*}_h = \mathcal{D}_h({S_{\mathbf{SF}}}^*, {\phi_{\mathbf{SF}}}^*)$. We start by showing that $\text{OPT}_{\text{SF}} \leq \frac{1}{\lambda}\,\text{OPT}_{\text{R}}$.

Suppose that was not true, i.e., $\frac{1}{\lambda}\,\text{OPT}_{\text{R}} < \text{OPT}_{\text{SF}}$. This implies that we have

$$\max_h \left(\frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}^*_h + \frac{1-\lambda}{\lambda}\,\frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{V}^*_h \right) < \max_{h'} \frac{1}{|\,\mathcal{C}^{h'}\,|}\,\mathcal{D}^{\mathbf{SF}^*}_{h'}$$

Now note that $\frac{1-\lambda}{\lambda}\frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{V}^*_h \geq 0$. This means that

$$\max_h \frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}^*_h < \max_{h'} \frac{1}{|\,\mathcal{C}^{h'}\,|}\,\mathcal{D}^{\mathbf{SF}^*}_{h'} = \text{OPT}_{\text{SF}}$$

But this implies that by using (S^*, ϕ^*) we would achieve a strictly more optimal solution for the socially fair objective which is a contradiction since $({S_{\mathbf{SF}}}^*, {\phi_{\mathbf{SF}}}^*)$ is an optimal solution for the **SF** objective. Therefore, we must have $\mathrm{OPT}_{\mathrm{SF}} \leq \frac{1}{\lambda}\,\mathrm{OPT}_{\mathrm{R}}$.

Now, using the above and since $(S_{\mathbf{SF}}, \phi_{\mathbf{SF}})$ results from an α approximation algorithm for the socially fair objective we must have $\left(\max_{h\in\mathcal{H}} \frac{\mathcal{D}_h^{\mathbf{SF}}}{|\mathcal{C}^h|}\right) \leq \alpha\,\mathrm{OPT}_{\mathrm{SF}} \leq \alpha \cdot \frac{1}{\lambda}\,\mathrm{OPT}_{\mathrm{R}}$. □

We will now prove the following claim.

Claim B.2. *$\forall h \in \mathcal{H}$ we have*

$$\mathcal{D}'_h \leq \gamma_p\,\mathcal{D}^*_h \;+\; \gamma_{p'}\,\mathcal{D}_h^{\boldsymbol{SF}} \tag{19}$$

Proof. We focus on a point j and let $i^* = \phi^*(j)$, the following must hold:

$$\begin{aligned}
d^p(j, \phi'(j)) &= d^p(j, \mathrm{nrst}(i^*)) \\
&\leq 2^{p-1}(d^p(j, i^*) + d^p(i^*, \mathrm{nrst}(i^*))) \quad \text{(By Fact A.1)} \\
&\leq 2^{p-1}(d^p(j, i^*) + d^p(i^*, \phi_{\mathbf{SF}}(j))) \quad \text{(Since nrst}(i^*)\text{ is the closest center to } i \text{ in } S_{\mathbf{SF}}) \\
&\leq 2^{p-1}(d^p(j, i^*) + 2^{p-1}(d^p(j, \phi_{\mathbf{SF}}(j)) + d^p(j, i^*))) \quad \text{(By Fact A.1)} \\
&= 2^{p-1}((2^{p-1}+1)d^p(j, i^*) + 2^{p-1}d^p(j, \phi_{\mathbf{SF}}(j))) \\
&= 2^{p-1}((2^{p-1}+1)d^p(j, \phi^*(j)) + 2^{p-1}d^p(j, \phi_{\mathbf{SF}}(j))) \\
&= \gamma_p\, d^p(j, \phi^*(j)) + \gamma_{p'}\, d^p(j, \phi_{\mathbf{SF}}(j)))
\end{aligned}$$

From the above, we get

$$\sum_{j\in\mathcal{C}^h} d^p(j, \phi'(j)) \leq \gamma_p \sum_{j\in\mathcal{C}^h} d^p(j, \phi^*(j)) + \gamma_{p'} \sum_{j\in\mathcal{C}^h} d^p(j, \phi_{\mathbf{SF}}(j))$$

Which is equivalent to

$$\mathcal{D}'_h \leq \gamma_p\,\mathcal{D}^*_h + \gamma_{p'}\,\mathcal{D}_h^{\mathbf{SF}} \tag{20}$$

□

Now we introduce the second claim for the violation.

Claim B.3. *$\forall h \in \mathcal{H}$ we have*

$$\mathcal{V}'_h \leq \mathcal{V}^*_h \tag{21}$$

Proof. First, note that $\mathcal{V}^*_h = \sum_{i^*\in S^*} |C_{i^*}|\,\nu(h, i^*)$ and $\mathcal{V}'_h = \sum_{i\in S_{\mathbf{SF}}} |C'_i|\,\nu(h, i)$ where the cluster C'_i is formed using the new assignment functio ϕ'. If the assignment assignment function ϕ' routes to center $i \in S_{\mathbf{SF}}$ one cluster from a center $i^* \in S^*$, then it follows immediately that $|C'_i|\,\nu(h, i) = |C_{i^*}|\,\nu(h, i^*)$.

We will therefore, consider the case where center $i \in S_{\mathbf{SF}}$ recieves a collection of clusters from S^*. In particular, let M denote the set of clusters routed to i by ϕ'. Let O denote the subsets of clusters in M in which color h is over-represented (strictly greater than $\alpha_h + r_h$), U denotes the subsets of clusters in M where h is under-represented (strictly less than $\alpha_h + r_h$), and E denotes the subsets of clusters in M that have no proportional violations. If the resulting proportion in cluster i is between $r_h - \beta$ and $r_h + \alpha_h$ then it follows that $\nu(h, i) = 0$ and therefore $|C'_i|\,\nu(h, i) \leq \sum_{i^*\in M} |C_{i^*}|\,\nu(h, i^*)$.

Therefore, we consider the case where the resulting cluster has an over-representation. The following must be true.

$$
\begin{aligned}
|C'_i|\,\nu(h,i) &= \left(\sum_{i^*\in M}|C_{i^*}|\right)\left(\frac{\sum_{i^*\in M}|C^h_{i^*}|}{\sum_{i^*\in M}|C_{i^*}|}-(r_h+\alpha_h)\right)\\
&= \left(\sum_{i^*\in M}|C^h_{i^*}|-(r_h+\alpha_h)\sum_{i^*\in M}|C_{i^*}|\right)\\
&= \sum_{i^*\in M}(|C^h_{i^*}|-(r_h+\alpha_h)|C_{i^*}|)\\
&= \sum_{i^*\in O}(|C^h_{i^*}|-(r_h+\alpha_h)|C_{i^*}|)+\sum_{i^*\in U}(|C^h_{i^*}|-(r_h+\alpha_h)|C_{i^*}|)+\sum_{i^*\in E}(|C^h_{i^*}|-(r_h+\alpha_h)|C_{i^*}|)\\
&\le \sum_{i^*\in O}||C^h_{i^*}|-(r_h+\alpha_h)|C^*_i||\\
&\le \sum_{i^*\in M}|C^*_i|\,\nu(h,i^*)
\end{aligned}
$$

Now if we assume that cluster i has under-representation, then the following is true.

$$
\begin{aligned}
|C'_i|\,\nu(h,i) &= \left((r_h-\beta_h)\sum_{i^*\in M}|C_{i^*}|-\sum_{i^*\in M}|C^h_{i^*}|\right)\\
&= \sum_{i^*\in M}((r_h-\beta_h)|C_{i^*}|-|C^h_{i^*}|)\\
&= \sum_{i^*\in U}((r_h-\beta_h)|C_{i^*}|-|C^h_{i^*}|)+\sum_{i^*\in O}((r_h-\beta_h)|C_{i^*}|-|C^h_{i^*}|)+\sum_{i^*\in E}|(r_h-\beta_h)|C_{i^*}|-|C^h_{i^*}||\\
&\le \sum_{i^*\in U}|(r_h-\beta_h)|C_{i^*}|-|C^h_{i^*}||\\
&\le \sum_{i^*\in M}|C_{i^*}|\,\nu(h,i^*)
\end{aligned}
$$

Since each center in S^* maps to one center in $S_{\mathbf{SF}}$, it follows that $\sum_{i\in S_{\mathbf{SF}}}|C'_i|\,\nu(h,i)\le\sum_{i^*\in S^*}|C_{i^*}|\,\nu(h,i^*)$. Therefore, it follows that $\mathcal{V}'_h\le\mathcal{V}^*_h$. □

The following claim must then hold.

Claim B.4. *$\forall h\in\mathcal{H}$ we have*

$$
D'_h \le \gamma_p\, D^*_h + \gamma_{p'}\,\lambda\frac{\mathcal{D}^{\boldsymbol{SF}}_h}{|\,\mathcal{C}^h\,|} \tag{22}
$$

Proof.

$$
\begin{aligned}
\lambda\,\mathcal{D}'_h &\le \gamma_p\,\lambda\,\mathcal{D}^*_h+\gamma_{p'}\,\lambda\,\mathcal{D}^{\mathbf{SF}}_h \qquad \text{(By Claim B.2)}\\
\implies \lambda\,\mathcal{D}'_h+(1-\lambda)\,\mathcal{V}'_h &\le \gamma_p\,\lambda\,\mathcal{D}^*_h+\gamma_{p'}\,\lambda\,\mathcal{D}^{\mathbf{SF}}_h+(1-\lambda)\,\mathcal{V}^*_h \qquad \text{(By Claim B.3)}\\
&\le \gamma_p\left(\lambda\,\mathcal{D}^*_h+(1-\lambda)\,\mathcal{V}^*_h\right)+\gamma_{p'}\,\lambda\,\mathcal{D}^{\mathbf{SF}}_h
\end{aligned}
$$

by multiplying both sides with $\frac{1}{|\,\mathcal{C}^h\,|}$ we get the inequality. □

Now we are ready to prove the lemma. Let $\hbar\in\arg\max_{h\in\mathcal{H}} D'_h$, the following is true

$$
\begin{aligned}
\max_{h\in\mathcal{H}} D'_h &= D'_{\hbar} \\
&\le \gamma_p\, D^*_{\hbar} + \gamma_{p'}\,\lambda \frac{\mathcal{D}^{\mathbf{SF}}_{\hbar}}{|\,\mathcal{C}^{\hbar}\,|} \qquad \text{(By Claim B.4)} \\
&\le \gamma_p \max_{h'\in\mathcal{H}} (D^*_{h'}) + \gamma_{p'}\cdot\lambda\cdot \max_{h''\in\mathcal{H}} \left(\frac{\mathcal{D}^{\mathbf{SF}}_{h''}}{|\,\mathcal{C}^{h''}\,|}\right) \\
&\le \gamma_p\,\mathrm{OPT_R} + \gamma_{p'}\cdot\lambda\cdot\left(\alpha\frac{1}{\lambda}\,\mathrm{OPT_R}\right) \qquad \text{(By Claim B.1)} \\
&= (\gamma_p + \gamma_{p'}\,\alpha)\,\mathrm{OPT_R}
\end{aligned}
$$

□

We restate the next lemma and give its proof.

Lemma 4.2. *Denoting the optimal value of LP* (10) *by* OPT^{LP}_R *then* $\mathrm{OPT}^{LP}_R \le (\gamma_p + \gamma'_p\alpha)\,\mathrm{OPT}_R$.

Proof. Consider some assignment $\bar{\phi} : \mathcal{C} \to S_{\mathbf{SF}}$. For a point $j \in \mathcal{C}$ set $x_{ij} = 1$ if $\bar{\phi}(j) = i$ and zero otherwise. It follows that such an assignment satisfies constraint (10f). Further, it is not difficult to see that $t_{i,h} = |C_i|\,\nu(h,i)$ for assignment $\bar{\phi}$. Moreover, the left hand side of constraint (10b) equals $D_h(S_{\mathbf{SF}}, \bar{\phi})$. Therefore, since constraint (10b) has $z \ge D_h(S_{\mathbf{SF}}, \bar{\phi})\ \forall h \in \mathcal{H}$, it follows that $\min z = \min_{h\in\mathcal{H}} D_h(S_{\mathbf{SF}}, \bar{\phi})$. This proves that any integral assignment is feasible in LP (10) and that it evaluates to the same value.

Since Lemma 4.1 establishes the existence of an assignment whose value is at most $(\gamma_p + \gamma_{p'}\,\alpha)\,\mathrm{OPT_R}$. It follows that $\mathrm{OPT^{LP}_R} \le (\gamma_p + \gamma_{p'}\,\alpha)\,\mathrm{OPT_R}$. □

We restate the next lemma and give its proof.

Lemma 4.3. *Let* $\mathbf{x}^{integ} = \textsc{Rawlsian-Rounding}(\mathbf{x}^{frac})$ *then the following holds.*

$$
\forall h \in \mathcal{H}:\ \mathcal{D}^{\mathrm{integ}}_h \le \mathcal{D}^{\mathrm{frac}}_h \tag{11}
$$

$$
\forall i \in S_{\boldsymbol{SF}},\ \forall h \in \mathcal{H}: \quad \left\lfloor \sum_{j\in\mathcal{C}^h} x^{\mathrm{frac}}_{ij} \right\rfloor \le \sum_{j\in\mathcal{C}^h} x^{\mathrm{integ}}_{ij} \le \left\lceil \sum_{j\in\mathcal{C}^h} x^{\mathrm{frac}}_{ij} \right\rceil \tag{12}
$$

Proof. We denote by $\mathbf{x}^{\mathrm{frac}}_h$ the the fractional solution restricted to the entries of points of color h, i.e., $j \in \mathcal{C}^h$. It follows by construction of the network flow from Appendix D.1, that for the network associated with each color h that $\mathbf{x}^{\mathrm{frac}}_h$ is a feasible assignment.

Since we minimize the min-cost max-flow for each color separately and since all of the capacities and demands are set to integral values, by standard network flow guarantees we can find an integral assignment that is optimal. Therefore, Inequality (11) follows.

Further, by the demand values of $\left\lfloor \sum_{j\in\mathcal{C}^h} x^{\mathrm{frac}}_{ij} \right\rfloor$ set at each vertex v^h_i and the capacity of 1 on arc (v^h_i, t^h) Inequality (12) follows. □

We restate the next lemma and give its proof.

Lemma 4.4. *The rounded Rawlsian solution has a cost of at most* $\mathrm{OPT}^{LP}_R + c_R$, *where*

$$
c_R = \left(\frac{|\,\mathcal{H}\,| + 1}{\min_{h\in\mathcal{H}} r_h}\right)\frac{k}{n}.
$$

Proof. We begin with the following claim which can be proved based on Inequality (12).

Claim B.5.

$$\forall i \in S : \sum_{j \in \mathcal{C}} x_{ij}^{frac} - |\mathcal{H}| \leq \sum_{j \in \mathcal{C}} x_{ij}^{integ} \leq \sum_{j \in \mathcal{C}} x_{ij}^{frac} + |\mathcal{H}| \tag{23}$$

Proof. First, by Inequality (12) we have

$$\forall i \in S, \forall h \in \mathcal{H} : \lfloor \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{frac}} \rfloor \leq \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{integ}} \leq \lceil \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{frac}} \rceil$$

Since $\lfloor z \rfloor \geq z - 1$ and $\lceil z \rceil \leq z + 1$ we have

$$\forall i \in S, \forall h \in \mathcal{H} : \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{frac}} - 1 \leq \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{integ}} \leq \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{frac}} + 1$$

By summing over the colors we get

$$\forall i \in S : \sum_{h \in \mathcal{H}} \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{frac}} - |\mathcal{H}| \leq \sum_{h \in \mathcal{H}} \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{integ}} \leq \sum_{h \in \mathcal{H}} \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{frac}} + |\mathcal{H}|$$

Which is

$$\forall i \in S : \sum_{j \in \mathcal{C}} x_{ij}^{\text{frac}} - |\mathcal{H}| \leq \sum_{j \in \mathcal{C}} x_{ij}^{\text{integ}} \leq \sum_{j \in \mathcal{C}} x_{ij}^{\text{frac}} + |\mathcal{H}| \tag{24}$$

□

The following claim must then be true.

Claim B.6.

$$\forall i \in S_{\boldsymbol{SF}}, \forall h \in \mathcal{H} :$$

$$\sum_{j \in \mathcal{C}^h} x_{ij}^{integ} \leq (r_h + \alpha_h) \sum_{j \in \mathcal{C}} x_{ij}^{integ} + (r_h + \alpha_h)|\mathcal{H}| + t_{i,h} + 1 \tag{25}$$

$$\sum_{j \in \mathcal{C}^h} x_{ij}^{integ} \geq (r_h - \beta_h) \sum_{j \in \mathcal{C}} x_{ij}^{integ} - (r_h - \beta_h)|\mathcal{H}| - t_{i,h} - 1 \tag{26}$$

Proof. We start with the first Inequality (25)

$$\begin{aligned}
\sum_{j \in \mathcal{C}^h} x_{ij}^{\text{integ}} &\leq \lceil \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{frac}} \rceil \qquad \text{(By Inequality (12))} \\
&\leq \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{frac}} + 1 \\
&\leq (r_h + \alpha_h) \sum_{j \in \mathcal{C}} x_{ij}^{\text{frac}} + t_{i,h} + 1 \quad \text{(By LP constraints (10d) and (10e))} \\
&\leq (r_h + \alpha_h) \left(\sum_{j \in \mathcal{C}} x_{ij}^{\text{integ}} + |\mathcal{H}| \right) + t_{i,h} + 1 \quad \text{(By Claim (B.5))} \\
&= (r_h + \alpha_h) \sum_{j \in \mathcal{C}} x_{ij}^{\text{integ}} + (r_h + \alpha_h)|\mathcal{H}| + t_{i,h} + 1
\end{aligned}$$

Similarly, for the second Inequality (26) we have

$$\begin{aligned}
\sum_{j\in\mathcal{C}^h} x_{ij}^{\text{integ}} &\geq \lfloor \sum_{j\in\mathcal{C}^h} x_{ij}^{\text{frac}} \rfloor \qquad \text{(By Inequality (12))} \\
&\geq \sum_{j\in\mathcal{C}^h} x_{ij}^{\text{frac}} - 1 \\
&\geq (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{frac}} - t_{i,h} - 1 \quad \text{(By LP constraints (10c) and (10e))} \\
&\geq (r_h - \beta_h) \left(\sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} - |\mathcal{H}| \right) - t_{i,h} - 1 \quad \text{(By Claim B.5)} \\
&= (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} - (r_h - \beta_h)|\mathcal{H}| - t_{i,h} - 1
\end{aligned}$$

□

We denote by $|C_i^{\text{integ}}|$ and $|C_i^{\text{frac}}|$ the size of the ith cluster according to the integral and fractional solutions, respectively. The following claim can be derived.

Claim B.7. *$\forall i \in S_{\boldsymbol{SF}}, \forall h \in \mathcal{H}$ we have $|C_i^{integ}|\, \nu^{integ}(h,i) \leq |C_i^{frac}|\, \nu^{frac}(h,i) + (|\,\mathcal{H}\,| + 1)$.*

Proof. First, note that by LP constraints (10c), (10d), and (10e) it follows that $|C_i^{\text{frac}}|\, \nu^{\text{frac}}(h,i) = t_{i,h}$.

Now suppose that color h in cluster i is over-represented, then it follows that

$$\begin{aligned}
|C_i^{\text{integ}}|\, \nu^{\text{integ}}(h,i) &\leq \sum_{j\in\mathcal{C}^h} x_{ij}^{\text{integ}} - (r_h + \alpha_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} \\
&\leq (r_h + \alpha_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} + (r_h + \alpha_h)|\mathcal{H}| + t_{i,h} + 1 - (r_h + \alpha_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} \qquad \text{(By Inequality (25))} \\
&= t_{i,h} + (r_h + \alpha_h)|\mathcal{H}| + 1 \\
&= |C_i^{\text{frac}}|\, \nu^{\text{frac}}(h,i) + ((r_h + \alpha_h)|\mathcal{H}| + 1) \\
&\leq |C_i^{\text{frac}}|\, \nu^{\text{frac}}(h,i) + (|\mathcal{H}| + 1) \qquad \text{(Since } r_h + \alpha_h \leq 1)
\end{aligned}$$

Similarly, if color h in cluster i is instead under-represented, then it follows that

$$\begin{aligned}
|C_i^{\text{integ}}|\, \nu^{\text{integ}}(h,i) &\leq (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} - \sum_{j\in\mathcal{C}^h} x_{ij}^{\text{integ}} \\
&\leq (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} - (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} + (r_h - \beta_h)|\mathcal{H}| + t_{i,h} + 1 \qquad \text{(By Inequality (26))} \\
&= t_{i,h} + (r_h - \beta_h)|\mathcal{H}| + 1 \\
&= |C_i^{\text{frac}}|\, \nu^{\text{frac}}(h,i) + ((r_h - \beta_h)|\mathcal{H}| + 1) \\
&\leq |C_i^{\text{frac}}|\, \nu^{\text{frac}}(h,i) + (|\mathcal{H}| + 1) \qquad \text{(Since } r_h - \beta_h \leq 1)
\end{aligned}$$

□

Now we can easily prove the following claim.

Claim B.8. *$\forall h \in \mathcal{H}$ we have:*

$$\frac{\mathcal{V}_h^{integ}}{|\,\mathcal{C}^h\,|} \leq \frac{\mathcal{V}_h^{frac}}{|\,\mathcal{C}^h\,|} + \frac{}{18}\Big(\frac{|\,\mathcal{H}\,| + 1}{r_h}\Big) \cdot \frac{k}{n}. \tag{27}$$

Proof.

$$\begin{aligned}
\frac{\mathcal{V}_h^{\text{integ}}}{|\mathcal{C}^h|} &= \frac{\sum_{i\in S_{\mathbf{SF}}} |C_i^{\text{integ}}|\,\nu^{\text{integ}}(h,i)}{|\mathcal{C}^h|} \\
&\le \frac{\sum_{i\in S_{\mathbf{SF}}} \Big(|C_i^{\text{integ}}|\,\nu^{\text{farc}}(h,i) + \big(|\mathcal{H}|+1\big)\Big)}{|\mathcal{C}^h|} \qquad \text{(By Claim B.7)} \\
&= \frac{\mathcal{V}_h^{\text{frac}}}{|\mathcal{C}^h|} + \Big(\frac{|\mathcal{H}|+1}{|\mathcal{C}^h|}\Big)\cdot k \\
&= \frac{\mathcal{V}_h^{\text{frac}}}{|\mathcal{C}^h|} + \Big(\frac{|\mathcal{H}|+1}{r_h}\Big)\cdot \frac{k}{n} \qquad \text{(Since } |\mathcal{C}^h| = r_h n\text{)}
\end{aligned}$$

□

Now we ready to prove the lemma. Specifically, for any color $h \in \mathcal{H}$ by Inequality 11 we have $\mathcal{D}_h^{\text{integ}} \le \mathcal{D}_h^{\text{frac}}$. Further, by Claim B.8 we have $\frac{\mathcal{V}_h^{\text{integ}}}{|\mathcal{C}^h|} \le \frac{\mathcal{V}_h^{\text{frac}}}{|\mathcal{C}^h|} + \Big(\frac{|\mathcal{H}|+1}{r_h}\Big)\cdot\frac{k}{n}$. Therefore we have

$$D_h^{\text{integ}} = \frac{\lambda\,\mathcal{D}_h^{\text{integ}} + (1-\lambda)\,\mathcal{V}_h^{\text{integ}}}{|\mathcal{C}^h|} \le \frac{\lambda\,\mathcal{D}_h^{\text{frac}} + (1-\lambda)\,\mathcal{V}_h^{\text{frac}}}{|\mathcal{C}^h|} + (1-\lambda)\cdot\Big(\frac{|\mathcal{H}|+1}{r_h}\Big)\cdot\frac{k}{n} \tag{28}$$

$$= D_h^{\text{farc}} + (1-\lambda)\cdot\Big(\frac{|\mathcal{H}|+1}{r_h}\Big)\cdot\frac{k}{n} \tag{29}$$

Since the above follows for each color it immediately follows that the cost is at most $\text{OPT}_R^{\text{LP}} + c_R$ where $c_R = \Big(\frac{|\mathcal{H}|+1}{\min_{h\in\mathcal{H}} r_h}\cdot\frac{k}{n}\Big)$. □

We restate the next theorem and give its proof.

Theorem 4.5. RawlsianAlg *(Algorithm 1) returns a solution of Rawlsian cost at most* $(\gamma_p + \gamma_{p'}\alpha)\,\text{OPT}_R + c_R$.

Proof. Lemma 4.1 guarantees that we can find a fractional solution whose cost is at most $(\gamma_p + \gamma_{p'}\alpha)\,\text{OPT}_\text{R}$. Further, once this solution is rounded the cost would at most be $(\gamma_p + \gamma_{p'}\alpha)\,\text{OPT}_\text{R} + c_R$ by Lemma 4.4. □

We restate the next lemma and give its proof.

Lemma 4.7. *Using the set of centers $S_{\boldsymbol{W}}$ chosen by an α-approximation algorithm for the weighted clustering objective with $w_j = \frac{1}{|\mathcal{C}^{\chi(j)}|}$ $\forall j \in \mathcal{C}$, there exists an assignment ϕ' such that $(S_{\boldsymbol{W}}, \phi')$ leads to a $(\gamma_p + \gamma_{p'}\alpha)$-approximate solution for the Utilitarian objective.*

Proof. The proof is similar to the proof of Lemma 4.1. We will denote by (S^*, ϕ^*) an optimal solution for the Utilitarian objective. $(S_{\mathbf{W}}, \phi_{\mathbf{W}})$ denote the solution of the α-approximation solution for the weighted clustering objective. We construct the assignment ϕ' as follows: for every $i^* \in S^*$, define $\text{nrst}(i^*) = \arg\min_{i\in S_{\mathbf{W}}} d(i, i^*)$. For every $j \in \mathcal{C}$, define $\phi'(j) = \text{nrst}(\phi^*(j))$.

Similar to the proof of Lemma 4.1 we define the values $\mathcal{D}_h^*$, $\mathcal{D}_h^{\mathbf{W}}$, and $\mathcal{D}_h'$ as the values of $\mathcal{D}_h(S^*, \phi^*)$, $\mathcal{D}_h(S_{\mathbf{W}}, \phi_{\mathbf{W}})$, and $\mathcal{D}_h(S_{\mathbf{W}}, \phi')$, respectively. To avoid confusion, note that $\mathcal{D}_h(S_{\mathbf{W}}, \phi')$ evaluates the distances without any weighting, i.e., $\mathcal{D}_h(S_{\mathbf{W}}, \phi') = \sum_{j\in\mathcal{C}^h} d(j, \phi_{\mathbf{W}}(j))$. Similar notion follow for $\mathcal{V}_h^*$, $\mathcal{V}_h^{\mathbf{W}}$, and $\mathcal{V}_h'$. Moreover, as noted earlier OPT_U is the optimal value of the Utilitarian objective. Further, we let $\text{OPT}_{\mathbf{W}}$ be the optimal value of the weighted clustering objective when evaluated on the weighted clustering objective.

We start with the following claim which is similar to Claim B.1.

Claim B.9.

$$\sum_{h\in\mathcal{H}} \frac{1}{|\mathcal{C}^h|}\,\mathcal{D}_h(S_{\boldsymbol{W}}, \phi_{\boldsymbol{W}}) \le \alpha\cdot\frac{1}{\lambda}\,\text{OPT}_U \tag{30}$$

Proof. Let $(S_{\mathbf{W}}{}^*, \phi_{\mathbf{W}}{}^*)$ denote an optimal solution for weighted clustering and let $\mathcal{D}_h^{\mathbf{W}^*} = \mathcal{D}_h(S_{\mathbf{W}}{}^*, \phi_{\mathbf{W}}{}^*)$. We start by showing that $\mathrm{OPT}_{\mathbf{W}} \leq \frac{1}{\lambda}\,\mathrm{OPT}_{\mathrm{U}}$.

Suppose that was not true, i.e., $\frac{1}{\lambda}\,\mathrm{OPT}_{\mathrm{U}} < \mathrm{OPT}_{\mathbf{W}}$. This implies that we have

$$\sum_{h\in\mathcal{H}} \left(\frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}_h^* + \frac{1-\lambda}{\lambda}\frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{V}_h^* \right) < \sum_{h\in\mathcal{H}} \frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}_h^{\mathbf{W}^*}$$

Since $\sum_{h\in\mathcal{H}} \frac{1-\lambda}{\lambda}\frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{V}_h^* \geq 0$, this means that

$$\sum_{h\in\mathcal{H}} \frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}_h^* < \sum_{h\in\mathcal{H}} \frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}_h^{\mathbf{W}^*} = \mathrm{OPT}_{\mathbf{W}}$$

But this implies that by using (S^*, ϕ^*) we would achieve a strictly more optimal solution for the weighted clustering objective which is a contradiction by optimality of $(S_{\mathbf{W}}{}^*, \phi_{\mathbf{W}}{}^*)$ for weighted clustering. Thus, $\mathrm{OPT}_{\mathbf{W}} \leq \frac{1}{\lambda}\,\mathrm{OPT}_{\mathrm{U}}$.

Moreover, since $(S_{\mathbf{W}}, \phi_{\mathbf{W}})$ is an α-approximation for weighted clustering, it follows that $\sum_{h\in\mathcal{H}} \frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}_h(S_{\mathbf{W}}, \phi_{\mathbf{W}}) \leq \alpha \sum_{h\in\mathcal{H}} \frac{1}{|\,\mathcal{C}^h\,|}\,\mathcal{D}_h^{\mathbf{W}^*} \leq \alpha \cdot \frac{1}{\lambda}\,\mathrm{OPT}_{\mathrm{U}}$. □

Following proofs similar to the proofs of Claims B.2, B.3, and B.4 with the socially fair solution replaced with the weighted clustering solution. The following three claims can be proved.

Claim B.10. *$\forall h \in \mathcal{H}$ we have*

$$\mathcal{D}_h' \leq \gamma_p\,\mathcal{D}_h^* \;+\; \gamma_{p'}\,\mathcal{D}_h^{\boldsymbol{W}} \tag{31}$$

Claim B.11. *$\forall h \in \mathcal{H}$ we have*

$$\mathcal{V}_h' \leq \mathcal{V}_h^* \tag{32}$$

Claim B.12. *$\forall h \in \mathcal{H}$ we have*

$$D_h' \leq \gamma_p\,D_h^* + \gamma_{p'}\,\lambda\frac{\mathcal{D}_h^{\boldsymbol{W}}}{|\,\mathcal{C}^h\,|} \tag{33}$$

Now we are ready to prove the lemma.

$$\begin{aligned}
\sum_{h\in\mathcal{H}} D_h' &\leq \gamma_p\,\big(\sum_{h\in\mathcal{H}} D_h^*\big) + \gamma_{p'}\,\lambda\Big(\sum_{h\in\mathcal{H}} \frac{\mathcal{D}_h^{\mathbf{W}}}{|\,\mathcal{C}^h\,|}\Big) \qquad \text{(By Claim B.12)}\\
&= \gamma_p\,\mathrm{OPT}_{\mathrm{U}} + \gamma_{p'}\,\lambda\Big(\sum_{h\in\mathcal{H}} \frac{\mathcal{D}_h^{\mathbf{W}}}{|\,\mathcal{C}^h\,|}\Big)\\
&\leq \gamma_p\,\mathrm{OPT}_{\mathrm{U}} + \gamma_{p'}\,\lambda\big(\alpha\frac{1}{\lambda}\,\mathrm{OPT}_{\mathrm{U}}\big) \qquad \text{(By Claim B.9)}\\
&= (\gamma_p + \gamma_{p'}\,\alpha)\,\mathrm{OPT}_{\mathrm{U}}
\end{aligned}$$

□

We restate the next lemma and give its proof.

Lemma 4.8. *Denoting the optimal value of LP* (13) *by* OPT_U^{LP} *then* $\mathrm{OPT}_U^{LP} \leq (\gamma_p + \gamma_p'\alpha)\,\mathrm{OPT}_U$.

Proof. Similar to the proof of Lemma 4.2 we can consider an assignment $\bar{\phi} : \mathcal{C} \to S_{\mathbf{W}}$ and set $x_{ij} = 1$ only if $\bar{\phi}(j) = i$ and zero otherwise. It follows that this assignment would be feasible and that $t_{i,h}$ equals the violation $|C_i|\,\nu(h,i)$ resulting from $\bar{\phi}$. Further, the objective in (13a) equals $\sum_{h\in\mathcal{H}} D_h(S_{\mathbf{W}}, \bar{\phi})$. This proves that any integral assignment in feasible in LP (13) and that the LP objective value equals the Utilitarian objective of that assignment.

Now, since Lemma 4.7 establishes the existence of an assignment of cost at most $(\gamma_p + \gamma_{p'}\,\alpha)\,\mathrm{OPT}_{\mathrm{U}}$ it follows that $\mathrm{OPT}_{\mathrm{U}}^{\mathrm{LP}} \leq (\gamma_p + \gamma_{p'}\,\alpha)\,\mathrm{OPT}_{\mathrm{U}}$

□

We restate the next lemma and give its proof.

Lemma 4.9. *Let* $\mathbf{x}^{integ} = \textsc{Utilitarian-Rounding}(\mathbf{x}^{frac})$ *then the following holds.*

$$\sum_{h\in\mathcal{H}} \frac{1}{|\mathcal{C}^h|}\mathcal{D}_h^{integ} \leq \sum_{h\in\mathcal{H}} \frac{1}{|\mathcal{C}^h|}\mathcal{D}_h^{frac} \tag{14}$$

$\forall i \in S_{\boldsymbol{W}}, \forall h \in \mathcal{H}:$

$$\left\lfloor \sum_{j\in\mathcal{C}^h} x_{ij}^{frac} \right\rfloor \leq \sum_{j\in\mathcal{C}^h} x_{ij}^{integ} \leq \left\lceil \sum_{j\in\mathcal{C}^h} x_{ij}^{frac} \right\rceil \tag{15}$$

$\forall i \in S_{\boldsymbol{W}}:$

$$\left\lfloor \sum_{j\in\mathcal{C}} x_{ij}^{frac} \right\rfloor \leq \sum_{j\in\mathcal{C}} x_{ij}^{integ} \leq \left\lceil \sum_{j\in\mathcal{C}} x_{ij}^{frac} \right\rceil \tag{16}$$

Proof. Similar to the proof of Lemma 4.3, one can see that in the network constructed in Appendix D.2 the fractional solution $\mathbf{x}^{\text{frac}}$ is feasible.

Since the capacities and demand values in the network are integral it follows that there exists an optimal integral solution. Hence, Inequality (14) follows.

Furthermore, by the demand and capacity values set inequalities (15) and (16) immediately follow. □

Lemma 4.10. *The rounded Utilitarian solution has a cost of at most* $\text{OPT}_U^{LP} + c_U$ *where* $c_U = \left(\frac{2k}{n} \cdot \sum_{h\in\mathcal{H}} \frac{1}{r_h}\right)$.

Proof. We begin with the following claim holds immediately by Inequality (16) since for any number z we have $\lfloor z \rfloor \geq z - 1$ and $\lceil z \rceil \leq z + 1$.

Claim B.13.

$$\forall i \in S: \sum_{j\in\mathcal{C}} x_{ij}^{frac} - 1 \leq \sum_{j\in\mathcal{C}} x_{ij}^{integ} \leq \sum_{j\in\mathcal{C}} x_{ij}^{frac} + 1 \tag{34}$$

The following claim can be proved.

Claim B.14.

$\forall i \in S_{\boldsymbol{W}}, \forall h \in \mathcal{H}:$

$$\sum_{j\in\mathcal{C}^h} x_{ij}^{integ} \leq (r_h + \alpha_h) \sum_{j\in\mathcal{C}} x_{ij}^{integ} + t_{i,h} + 2 \tag{35}$$

$$\sum_{j\in\mathcal{C}^h} x_{ij}^{integ} \geq (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{integ} - t_{i,h} - 2 \tag{36}$$

Proof. We start with the first Inequality (25)

$$\begin{aligned}
\sum_{j\in\mathcal{C}^h} x_{ij}^{\text{integ}} &\leq \lceil \sum_{j\in\mathcal{C}^h} x_{ij}^{\text{frac}} \rceil \qquad \text{(By Inequality (15))} \\
&\leq \sum_{j\in\mathcal{C}^h} x_{ij}^{\text{frac}} + 1 \\
&\leq (r_h + \alpha_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{frac}} + t_{i,h} + 1 \quad \text{(By LP constraints (13c) and (13d))} \\
&\leq (r_h + \alpha_h) \left(\sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} + 1 \right) + t_{i,h} + 1 \quad \text{(By Claim (B.13))} \\
&= (r_h + \alpha_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} + t_{i,h} + 2 \qquad \text{(Since } r_h + \alpha_h \leq 1\text{)}
\end{aligned}$$

Similarly, for the second Inequality (26) we have

$$\begin{aligned}
\sum_{j\in\mathcal{C}^h} x_{ij}^{\text{integ}} &\geq \lfloor \sum_{j\in\mathcal{C}^h} x_{ij}^{\text{frac}} \rfloor \qquad \text{(By Inequality (15))} \\
&\geq \sum_{j\in\mathcal{C}^h} x_{ij}^{\text{frac}} - 1 \\
&\geq (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{frac}} - t_{i,h} - 1 \quad \text{(By LP constraints (13b) and (13d))} \\
&\geq (r_h - \beta_h) \left(\sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} - 1 \right) - t_{i,h} - 1 \quad \text{(By Claim B.5)} \\
&= (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} - (r_h - \beta_h) - t_{i,h} - 1 \\
&= (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} - t_{i,h} - 2 \qquad \text{(Since } r_h + \alpha_h \leq 1\text{)}
\end{aligned}$$

□

We denote by $|C_i^{\text{integ}}|$ and $|C_i^{\text{frac}}|$ the size of the ith cluster according to the integral and fractional solutions, respectively. The following claim can be derived.

Claim B.15. *$\forall i \in S_{\boldsymbol{W}}, \forall h \in \mathcal{H}$ we have $|C_i^{integ}|\, \nu^{integ}(h,i) \leq |C_i^{frac}|\, \nu^{frac}(h,i) + 2$.*

Proof. First, note that by LP constraints (13b), (13c), and (13d) it follows that $|C_i^{\text{frac}}|\, \nu^{\text{frac}}(h,i) = t_{i,h}$.

Now suppose that color h in cluster i is over-represented, then it follows that

$$\begin{aligned}
|C_i^{\text{integ}}|\, \nu^{\text{integ}}(h,i) &\leq \sum_{j\in\mathcal{C}^h} x_{ij}^{\text{integ}} - (r_h + \alpha_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} \\
&\leq (r_h + \alpha_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} + t_{i,h} + 2 - (r_h + \alpha_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} \qquad \text{(By Inequality (35))} \\
&= t_{i,h} + 2 \\
&= |C_i^{\text{frac}}|\, \nu^{\text{frac}}(h,i) + 2
\end{aligned}$$

Similarly, if color h in cluster i is instead under-represented, then it follows that

$$\begin{aligned}
|C_i^{\text{integ}}|\, \nu^{\text{integ}}(h,i) &\leq (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} - \sum_{j\in\mathcal{C}^h} x_{ij}^{\text{integ}} \\
&\leq (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} - (r_h - \beta_h) \sum_{j\in\mathcal{C}} x_{ij}^{\text{integ}} + t_{i,h} + 2 \qquad \text{(By Inequality (36))} \\
&= t_{i,h} + +2 \\
&= |C_i^{\text{frac}}|\, \nu^{\text{frac}}(h,i) + 2
\end{aligned}$$

□

Now we can easily prove the following claim.

Claim B.16.

$$\sum_{h\in\mathcal{H}} \frac{\mathcal{V}_h^{integ}}{|\,\mathcal{C}^h\,|} \leq \sum_{h\in\mathcal{H}} \frac{\mathcal{V}_h^{frac}}{|\,\mathcal{C}^h\,|} + \Big(\frac{2k}{n} \cdot \sum_{h\in\mathcal{H}} \frac{1}{r_h} \Big) \tag{37}$$

Proof.

$$\sum_{h\in\mathcal{H}} \frac{\mathcal{V}_h^{\text{integ}}}{|\,\mathcal{C}^h\,|} = \sum_{h\in\mathcal{H}} \frac{\sum_{i\in S} |C_i^{\text{integ}}|\,\nu^{\text{integ}}(h,i)}{|\,\mathcal{C}^h\,|} \tag{38}$$

$$\leq \sum_{h\in\mathcal{H}} \frac{\sum_{i\in S}(|C_i^{\text{frac}}|\,\nu^{\text{frac}}(h,i)+2)}{|\,\mathcal{C}^h\,|} \qquad \text{(By Claim B.15)} \tag{39}$$

$$= \sum_{h\in\mathcal{H}} \frac{\mathcal{V}_h^{\text{frac}}}{|\,\mathcal{C}^h\,|} + \Big(2k\cdot\sum_{h\in\mathcal{H}}\frac{1}{|\,\mathcal{C}^h\,|}\Big) \tag{40}$$

$$= \sum_{h\in\mathcal{H}} \frac{\mathcal{V}_h^{\text{frac}}}{|\,\mathcal{C}^h\,|} + \Big(\frac{2k}{n}\cdot\sum_{h\in\mathcal{H}}\frac{1}{r_h}\Big) \qquad \text{(Since } |\,\mathcal{C}^h\,| = r_h n\text{)} \tag{41}$$

□

Now we ready to prove the lemma.

$$\begin{aligned}
\sum_{h\in\mathcal{H}} D_h^{\text{integ}} &= \sum_{h\in\mathcal{H}} \frac{\lambda\,\mathcal{D}_h^{\text{integ}} + (1-\lambda)\,\mathcal{V}_h^{\text{integ}}}{|\,\mathcal{C}^h\,|} \\
&\leq \sum_{h\in\mathcal{H}} \frac{\lambda\,\mathcal{D}_h^{\text{farc}} + (1-\lambda)\,\mathcal{V}_h^{\text{frac}}}{|\,\mathcal{C}^h\,|} + (1-\lambda)\Big(\frac{2k}{n}\cdot\sum_{h\in\mathcal{H}}\frac{1}{r_h}\Big) \qquad \text{(By Inequality (14) and Claim B.16)} \\
&= \sum_{h\in\mathcal{H}} D_h^{\text{frac}} + (1-\lambda)\Big(\frac{2k}{n}\cdot\sum_{h\in\mathcal{H}}\frac{1}{r_h}\Big) \\
&\leq \sum_{h\in\mathcal{H}} D_h^{\text{frac}} + c_U
\end{aligned}$$

Where $c_U = \Big(\frac{2k}{n}\cdot\sum_{h\in\mathcal{H}}\frac{1}{r_h}\Big)$.

□

We restate the next theorem and give its proof.

Theorem 4.11. UTILITARIANALG *(Algorithm 2) returns a solution of Utilitarian cost at most* $(\gamma_p + \gamma_p\prime\,\alpha)\,\mathrm{OPT}_U + c_U$.

Proof. The LP will return a fractional solution of cost at most $(\gamma_p + \gamma_p\prime\,\alpha)\,\mathrm{OPT_U}$ by Lemma 4.8. Further, Lemma 4.10 implies that the rounded solution from the LP will have a cost of at most $(\gamma_p + \gamma_p\prime\,\alpha)\,\mathrm{OPT_U} + c_U$. □

C NP-Hardness of the Assignment Problem

Here we show that finding the optimal Rawlsian or Utilitarian assignment ϕ given a set of centers S is NP-hard.

Following the same reduction from Exact Cover by 3-Sets as that of Esmaeili et al. (2021). See also Theorem 5.1 and Figure 4 in Esmaeili et al. (2021). Our instance will be identical. In particular, we would have $r_{\text{red}} = \frac{1}{4}$ and $r_{\text{blue}} = \frac{3}{4}$. Using our notation we would set $\alpha_h = \beta_h = 0$,$\forall h \in \{\text{red}, \text{blue}\}$. We use the same set of vertices $\mathcal{F}$ for the set of given centers. Is it immediate that the minimum disutility for the red color is $D_{\text{red}} \geq \frac{\lambda(0\cdot\frac{|\,\mathcal{C}^{\text{red}}\,|}{2} + 1\cdot\frac{|\,\mathcal{C}^{\text{red}}\,|}{2}) + (1-\lambda)\cdot 0}{|\,\mathcal{C}^{\text{red}}\,|} = \frac{1}{2}\lambda$. Similarly, for the blue color, we have $D_{\text{blue}} \geq \frac{\lambda(1\cdot|\,\mathcal{C}^{\text{blue}}\,|) + (1-\lambda)\cdot 0}{|\,\mathcal{C}^{\text{blue}}\,|} = \lambda$. We assume $\lambda > 0$. It follows straightforwardly that we have a Rawlsian objective of λ and a Utilitarian objective of $\frac{3}{2}\lambda$ if and only if Exact Cover by 3-Sets has a solution. This is proves that the assignment problem is NP-hard for both the Rawlsian and Utilitarian objectives.

D The Network Flow Rounding Scheme

Our rounding schemes build upon standard network flow rounding procedure introduced in Bercea et al. (2018). We first introduce the standard version as appeared in Bercea et al. (2018) and will then introduce our variants in their subsections.

The input to network rounding scheme is fractional solution to Linear Program, $\mathbf{x}^{\text{frac}}$ and center set S and distances $d(i,j)$, $\forall i \in S, j \in \mathcal{C}$. In standard rounding, one network is constructed. Denote the network as $G = (V, E)$ where V is the set of vertices and E the set of arcs. Components of graph G are listed below.

1. $V = \{v_j | j \in \mathcal{C}\} \cup \{v_i^h | h \in \mathcal{H}, i \in S\} \cup \{v_i | i \in S\} \cup \{t\}$
2. $E = \{(v_j, v_i^h) | x_{ij}^{\text{frac}} > 0\} \cup \{(v_i^h, v_i) | h \in \mathcal{H}, i \in S\} \cup \{(v_i, t) | i \in S\}$

Specifically, we create one vertex v_j for each point $j \in \mathcal{C}$ with demand -1. For each center $i \in S$ we define a set of vertices v_i^h for each color $h \in \mathcal{H}$, and a separate vertex v_i. Let each v_i^h has demand $\lfloor \sum_{j \in \mathcal{C}^h} x_{ij} \rfloor$, and v_i has a demand $\left(\lfloor \sum_{j \in \mathcal{C}} x_{ij} \rfloor - \sum_{h \in \mathcal{H}} \lfloor \sum_{j \in \mathcal{C}^h} x_{ij} \rfloor \right)$. We create a set of arcs (v_i^h, v_i) for each i and h. Let these arcs have cost 0. For each $x_{ij}^{\text{frac}} > 0$, we create an arc (v_j, v_i^h) for the h such that $j \in \mathcal{C}^h$. For this arc (v_j, v_i^h) we set its cost to be $d^p(i,j)$. We define a sink vertex t with demand $(|\mathcal{C}| - \sum_{i \in S} \lfloor \sum_{j \in \mathcal{C}} x_{ij} \rfloor)$ and connect t with v_i for each $i \in S$ with an arc. Arcs $\{(v_i, t) | i \in S\}$ have costs 0. We let all arcs have unit capacity.

Because all capacities and demands in this network are integral, an integral solution can be found in polynomial time. It is not difficult to see that $\mathbf{x}^{\text{frac}}$ is feasible and therefore the cost of the integral solution cannot exceed that of the fractional solution. By construction, we have the two inequalities shown in Lemma 4.9, (15) and (16).

D.1 Details of the Rawlsian Rounding

We construct the Rawlsian Rounding min-cost flow instance on fractional solution $\mathbf{x}^{\text{frac}}$ from LP (10). The network instance contains unconnected independent networks, one for each color $h \in \mathcal{H}$ as illustrated in Figure 2. We denote G^h as network of color h. $G^h = (V^h, E^h)$. Components of G^h are listed below.

1. $V^h = \{v_j | j \in \mathcal{C}^h\} \cup \{v_i^h | i \in S_{\mathbf{SF}}\} \cup t^h$
2. $E^h = \{(v_j, v_i^h) | j \in \mathcal{C}^h, i \in S_{\mathbf{SF}}\} \cup \{(v_i^h, t^h) | i \in S_{\mathbf{SF}}\}$

For each fixed color $h \in \mathcal{H}$, for each center $i \in S_{\mathbf{SF}}$ we define a set of vertices v_i^h. Each of v_i^h has demand $\lfloor \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{frac}} \rfloor$ for the corresponding center i. Note x_{ij}^{frac} are elements of solution $\mathbf{x}^{\text{frac}}$. For each point $j \in \mathcal{C}^h$, we create a vertex v_j, with demand -1. Between v_j and v_i^h, we create arc (v_j, v_i^h) if $x_{ij} > 0$. We let arcs (v_j, v_i^h) each has capacity 1. We include a sink vertex t^h which has demands $(|\mathcal{C}^h| - \sum_{i \in S_{\mathbf{SF}}} \lfloor \sum_{j \in \mathcal{C}^h} x_{ij}^{\text{frac}} \rfloor)$. We connect each of v_i^h for $i \in S_{\mathbf{SF}}$ with the sink t^h with an arc (v_i^h, t^h). These arcs have capacity 1 as well. On the set of arcs (v_j, v_i^h) we assign cost $\frac{1}{|\mathcal{C}^h|} d^p(i,j)$. For other arcs in this network, we assign cost 0.

On each color specific network constructed as described above, given the capacity, demands are integers, there exist an integral optimal min-cost flow solution. Note the fractional assignment from LP solution constrained to only entries of color h, $\mathbf{x}_h^{\text{frac}} = \{x_{ij}^{\text{frac}}\}_{j \in \mathcal{C}^h}$ constitutes a feasible fractional flow solution to the min-cost flow problem.

D.2 Details of the Utilitarian Rounding

For Utilitarian Rounding, we take fractional solution of Linear Program (13), centers $S_{\mathbf{W}}$, and distances $d(i,j)$, $\forall i \in S, j \in \mathcal{C}$ as input. We set up one network whose the components $G = (V, E)$ are the same as described as standard network setup at beginning of this section. The only difference is we set costs of arcs (v_j, v_i^h) to $\frac{d^p(i,j)}{|\mathcal{C}^{\chi(j)}|}$. Network structure remain the same. Costs of other arcs, demands of vertices, capacities of all arcs remain the same.

E Necessity of Solving the Assignment Problem

Here we start by showing that the value of the average violation $\mathcal{V}_h$ for any color $h \in \mathcal{H}$ cannot exceed 2, (see Fact E.1 below). Further, Lemmas 4.1 and 4.7 establish the existence of an assignment with a bounded approximation ratio for the Rawlsian and Utilitarian objectives, respectively. Based on the above, one might wonder if optimizing for the assignment as done in the LP and rounding steps of RawlsianAlg and UtilitarianAlg is even necessary, perhaps due to the boundedness of the average violation we would never incur a large degradation in the approximation. We show in Theorem E.2 that optimizing for the assignment is in fact necessary as otherwise the approximation ratio would be unbounded while the additive approximation would be a constant unlike our additive approximation which is on the order of $O(\frac{k}{n})$ which converges to zero as $n \to \infty$.

Fact E.1. *For any color h the normalized violation is upper bounded by 2. More specifically, $\frac{\mathcal{V}_h}{|\mathcal{C}^h|} \leq 2(1-r_h) \leq 2$.*

Proof. We set $\alpha_h = \beta_h = 0$, $\forall h \in \mathcal{H}$. Note that by definition of $\nu(h,i)$ in Eq. (6) setting $\alpha_h = \beta_h = 0$ would only lead to a higher value for $\mathcal{V}_h$ in comparison to any other values for α_h and β_h.

Now, consider a fixed color h. For a center $i \in S$ set $p_i = \frac{|C_i^h|}{|\mathcal{C}^h|}$. That is, p_i is the proportion of color h points assigned to cluster i in the given clustering. By definition $\sum_{i \in S} p_i = 1$. Denote number of points in cluster C_i having colors in $\mathcal{H}$ other than color h by $C_i^{\bar{h}}$. Accordingly, we have

$$
\begin{aligned}
\frac{\mathcal{V}_h}{|\mathcal{C}^h|} &= \frac{1}{|\mathcal{C}^h|} \sum_{i \in S} |C_i| \cdot \nu(h,i) \\
&= \frac{1}{|\mathcal{C}^h|} \sum_{i \in S} \big| |C_i| r_h - |C_i^h| \big| \\
&= \sum_{i \in S} \left| \frac{|C_i|}{|\mathcal{C}^h|} r_h - p_i \right| \\
&= \sum_{i \in S} \left| \frac{|C_i|}{|\mathcal{C}|} - p_i \right| \\
&= \sum_{i \in S} \left| \frac{|C_i^{\bar{h}}| + p_i |\mathcal{C}^h|}{|\mathcal{C}|} - p_i \right| \\
&= \sum_{i \in S} \left| \frac{|C_i^{\bar{h}}|}{|\mathcal{C}|} + \Big(\frac{|\mathcal{C}^h|}{|\mathcal{C}|} - 1 \Big) p_i \right| \\
&= \sum_{i \in S} \left| \frac{|C_i^{\bar{h}}|}{|\mathcal{C}|} + (r_h - 1) p_i \right| \\
&= \sum_{i \in S} \left| \frac{|C_i^{\bar{h}}|}{|\mathcal{C}|} - (1 - r_h) p_i \right| \\
&= (1 - r_h) \sum_{i \in S} \left| \frac{|C_i^{\bar{h}}|}{|\mathcal{C}|(1 - r_h)} - p_i \right|
\end{aligned} \tag{42}
$$

When maximizing over assignment ϕ, we essentially maximize over p_i and $|\mathcal{C}_i^{\bar{h}}|$ over $i \in S$ because other quantities in the objective are fixed. That is

$$
\max_{\phi} \frac{1}{|\mathcal{C}^h|} \sum_{i \in S} |C_i| \cdot \nu(h,i) = \max_{p_i, |C_i^{\bar{h}}|} (1 - r_h) \sum_{i \in S} \left| \frac{|C_i^{\bar{h}}|}{|\mathcal{C}|(1 - r_h)} - p_i \right|
$$

Note consider the assignment of points of colors in $\mathcal{H}$ other than h, i.e., $|C_i^{\bar{h}}|$ over $i \in S$. Define $q_i = \frac{|C_i^{\bar{h}}|}{|\mathcal{C}|(1-r_h)}$.

By definition $\sum_{i=1}^{k} q_i = 1$. With this notation, (42) can be written as

$$\begin{aligned}
\max_{p_i, |C_i^{\bar{h}}|} (1-r_h) \sum_{i \in S} \left| \frac{|C_i^{\bar{h}}|}{|\mathcal{C}|(1-r_h)} - p_i \right| &= (1-r_h) \max_{p_i, q_i} \sum_{i \in S} |q_i - p_i| \\
&= (1-r_h) \max_{\mathbf{p},\mathbf{q}} \|\mathbf{q} - \mathbf{p}\|_1 \\
&= (1-r_h) \cdot 2 \\
&\leq 2
\end{aligned}$$

where the third equality uses $\max_{\mathbf{p},\mathbf{q}} \|\mathbf{q} - \mathbf{p}\|_1 = 2$ for probability vectors $\mathbf{p}$ and $\mathbf{q}$, and the final inequality follows from $(1 - r_h) \leq 1$. □

Now we present that theorem which implies that optimizing the assignment is necessary.

Theorem E.2. *Denote by $(S_{\boldsymbol{SF}}{}^*, \phi_{\boldsymbol{SF}}{}^*)$ the optimal socially fair solution and by $(S_{\boldsymbol{W}}{}^*, \phi_{\boldsymbol{W}}{}^*)$ the optimal weighted clustering solution with $w_j = \frac{1}{|\mathcal{C}^h|}$,$\forall j \in \mathcal{C}$. Then there exists an instance where $(S_{\boldsymbol{SF}}{}^*, \phi_{\boldsymbol{SF}}{}^*)$ has a Rawlsian objective that is an unbounded multiple of the optimal Rawlsian objective plus an additive constant. Similarly, $(S_{\boldsymbol{W}}{}^*, \phi_{\boldsymbol{W}}{}^*)$ has a Utilitarian objective that is an unbounded multiple of the optimal Utilitarian objective plus an additive constant. Specifically, we have*

$$R(S_{\boldsymbol{SF}}{}^*, \phi_{\boldsymbol{SF}}{}^*) \geq \rho_1 \, \mathrm{OPT}_R + \rho_2 \tag{43}$$

$$U(S_{\boldsymbol{W}}{}^*, \phi_{\boldsymbol{W}}{}^*) \geq \rho_1 \, \mathrm{OPT}_U + \rho_2 \tag{44}$$

where ρ_1 is an arbitrarily large positive value and ρ_2 is a positive constant.

Figure 4: Instance on which socially fair solution has unbounded Rawlsian and Utilitarian objective value.

Proof. We start with the proof for the socially fair solution and the Rawlsian objective.

Consider the instance in Figure 4. In this instance, there are $(1-r)n$ blue points and rn red points where $r \in (0, 1)$. All blue points coincide with one another and all red points coincide with one another with both colors being separated by a distance of R. Further, we set $\alpha_h = \beta_h = 0$,$\forall h \in \mathcal{H}$. For this instance, with $k = 2$, the optimal socially fair solution will have two centers $S_{\mathbf{SF}}{}^*$ with one center at the blue points and another center at the red points. Clearly, $\phi_{\mathbf{SF}}{}^*$ will assign each point to its closest center. The disutilities for the blue and red groups would be:

$$D_{\text{blue}}(S_{\mathbf{SF}}{}^*, \phi_{\mathbf{SF}}{}^*) = \lambda \cdot 0 + (1-\lambda) \cdot \frac{|(1-r)(1-r)n - (1-r)n| + |(1-r)rn - 0|}{(1-r)n} = 2(1-\lambda)r$$

$$D_{\text{red}}(S_{\mathbf{SF}}{}^*, \phi_{\mathbf{SF}}{}^*) = \lambda \cdot 0 + (1-\lambda) \cdot \frac{|r(1-r)n - 0| + |r \cdot rn - rn|}{rn} = 2(1-\lambda)(1-r)$$

For simplicity, we set $r = \frac{1}{2}$, thus $R(S_{\mathbf{SF}}{}^*, \phi_{\mathbf{SF}}{}^*) = D_{\text{blue}}(S_{\mathbf{SF}}{}^*, \phi_{\mathbf{SF}}{}^*) = D_{\text{red}}(S_{\mathbf{SF}}{}^*, \phi_{\mathbf{SF}}{}^*) = (1-\lambda)$.

Consider a solution $(S_{\mathbf{SF}}{}^*, \phi')$, i.e., using the same centers as the optimal socially fair solution but with a different assignment ϕ'. Specifically, ϕ' assigns half the blue points to the right center and half the red points to the left center. It is clear that assignment ϕ' would lead to no proportional violations. The disutilities would thus be

$$D_{\text{blue}}(S_{\mathbf{SF}}{}^*, \phi') = \lambda \cdot \frac{\frac{(1-r)n}{2} R^p}{(1-r)n} + (1-\lambda) \cdot 0 = \frac{\lambda}{2} R^p$$

$$D_{\text{red}}(S_{\mathbf{SF}}{}^*, \phi') = \lambda \cdot \frac{\frac{rn}{2} R^p}{rn} + (1-\lambda) \cdot 0 = \frac{\lambda}{2} R^p$$

Thus for this solution we would have $R(S_{\mathbf{SF}}{}^*, \phi') = D_{\text{blue}}(S_{\mathbf{SF}}{}^*, \phi') = D_{\text{red}}(S_{\mathbf{SF}}{}^*, \phi') = \frac{\lambda}{2} R^p$.

Now we set $R^p = 2 \cdot \frac{1-\lambda}{\lambda} \cdot \theta$ where $\theta \in (0, 1)$. This leads to $R(S_{\mathbf{SF}}{}^*, \phi') \leq \theta(1-\lambda) < (1-\lambda) = R(S_{\mathbf{SF}}{}^*, \phi_{\mathbf{SF}}{}^*)$.

Based on the above, using $\gamma \in [0,1]$ we can write $R({S_{\mathbf{SF}}}^*, {\phi_{\mathbf{SF}}}^*)$ in terms of $R({S_{\mathbf{SF}}}^*, \phi')$ as follows:

$$R({S_{\mathbf{SF}}}^*, {\phi_{\mathbf{SF}}}^*) = \gamma \cdot \frac{R({S_{\mathbf{SF}}}^*, \phi')}{\theta} + (1-\gamma) \cdot (1-\lambda)$$

Which is equivalent to

$$R({S_{\mathbf{SF}}}^*, {\phi_{\mathbf{SF}}}^*) = \rho_1 \cdot R({S_{\mathbf{SF}}}^*, \phi') + \rho_2 \tag{45}$$

Where $\rho_1 = \frac{\gamma}{\theta}$ and $\rho_2 = (1-\gamma) \cdot (1-\lambda)$. It is clear that for $\gamma \in (0,1)$, ρ_2 is a positive constant and ρ_1 is a constant that can be made arbitrarily large by choosing an instance with sufficiently small θ.

Since $({S_{\mathbf{SF}}}^*, \phi')$ is a feasible solution for Rawlsian clustering it follows that $R({S_{\mathbf{SF}}}^*, \phi') \geq \mathrm{OPT_R}$. Based on (45) this leads to

$$R({S_{\mathbf{SF}}}^*, {\phi_{\mathbf{SF}}}^*) \geq \rho_1 \cdot \mathrm{OPT_R} + \rho_2$$

For the weighted clustering solution and the Utilitarian objective. Simply note that $({S_{\mathbf{W}}}^*, {\phi_{\mathbf{W}}}^*) = ({S_{\mathbf{SF}}}^*, {\phi_{\mathbf{SF}}}^*)$, i.e., the weighted clustering solution is equal to the socially fair solution. We would have $U({S_{\mathbf{W}}}^*, {\phi_{\mathbf{W}}}^*) = 2(1-\lambda)$. Following a the same assignment ϕ' we would have $U({S_{\mathbf{W}}}^*, \phi') = \lambda R^p$. Further, we still have $R^p = 2\frac{1-\lambda}{\lambda}\theta$ where $\theta \in (0,1)$ we have $U({S_{\mathbf{W}}}^*, \phi') < U({S_{\mathbf{W}}}^*, {\phi_{\mathbf{W}}}^*)$. Therefore, we can similarly establish the following:

$$U({S_{\mathbf{W}}}^*, {\phi_{\mathbf{W}}}^*) = \frac{\gamma}{\theta} \cdot U({S_{\mathbf{W}}}^*, \phi') + 2(1-\gamma)(1-\lambda)$$

Since $\mathrm{OPT_U} \leq U({S_{\mathbf{W}}}^*, \phi')$ then we have the final guarantee.

$$U({S_{\mathbf{W}}}^*, {\phi_{\mathbf{W}}}^*) \geq \rho_1 \cdot \mathrm{OPT_U} + \rho_2$$

With $\rho_1 = \frac{\gamma}{\theta}$ and $\rho_2 = 2(1-\gamma)(1-\lambda)$. □

F Comparing Welfare-Centric Clustering to FCBC

The FCBC formulation unlike ours is framed in terms of clustering cost not disutilities and welfare. This causes FCBC to find solutions that are acceptable in terms of the clustering cost but clearly when one considers the disutilities and welfare. To more easily demonstrate how clustering cost, i.e., $\sum_{j \in \mathcal{C}} d^p(j, \phi(j))$ would different from the Rawlsian and Utilitarian objectives, we will assume that $\lambda = 1$, so that $(1-\lambda) = 0$ and proportional violations would not incur a cost.

Consider the simple example shown in Figure 5. In that example we have $r_{\text{blue}} = \frac{2}{3}$ and $r_{\text{red}} = r_{\text{green}} = \frac{1}{6}$. There are 4 special locations in the space $\{a, b, c, o\}$. All red, blue, and green points are at locations a,b, and c, respectively. For ease of discussion we consider the k-median problem and the Hamming distance. This implies that $d(a,o) = d(b,o) = d(c,o) = 1$ and that $d(a,b) = d(a,c) = d(b,c) = 2$.

Consider the solution (S, ϕ) where $S = \{o\}$ and ϕ assigns all points to center o. Then we would have $D_{\text{red}} = D_{\text{green}} = D_{\text{blue}} = 1$. This also leads to $R(S, \phi) = 1$ and $U(S, \phi) = 3$. The clustering cost would be $\sum_{j \in \mathcal{C}} d^p(j, \phi(j)) = 1 \cdot n = n$.

Now, consider a less equitable solution (S', ϕ') where $S' = \{b\}$ and all points are assigned to center b. This would lead to $D_{\text{red}} = D_{\text{green}} = 2$ while $D_{\text{blue}} = 0$. The Rawlsian and Utilitarian objectives would become clearly higher. Specifically, we would have $R(S', \phi') = 2$ and $U(S', \phi') = 2 \cdot 2 = 4$. On the other hand, the clustering cost would drop since $\sum_{j \in \mathcal{C}} d^p(j, \phi'(j)) = 2 \cdot \frac{1}{3} \cdot n = \frac{2}{3}n < n = \sum_{j \in \mathcal{C}} d^p(j, \phi(j))$.

Since FCBC uses the clustering costs as an upper bound it would favor such solutions which are clearly inequitable. This highlights the importance of formulating the problem in terms of welfare instead of naive view of simple clustering costs.

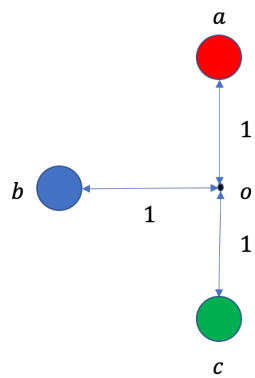


Figure 5: Clustering instance.

G Proof of Intermediate Non-Monotonicity

In Section 5 and H.7, we discussed the tuning of POF for FCBC algorithms to make them output clustering results with completitive (small) rawlsian and utilitarian objective values. We show in this section that the utilitarian objective is not monotonic POF values. Thus the tuning for FCBC algorithms cannot be done with an efficient search algorithm such as bisection.

Proposition G.1. *There exist clustering instances where the Utilitarian objective value of the solution returned by the FCBC algorithm is a non-monotonic function of its Price of Fairness (*PoF*) parameter. Specifically, there can be clustering solutions with clustering distance costs $B_0 < B_1 < B_2$ such that the corresponding welfare values U_0, U_1, U_2 satisfy $U_0 > U_1$ and $U_1 < U_2$.*

Proof. We show on the instance shown in Figure 6, the Utilitarian objective value is not monotonic in clustering cost of solutions. Consider the definition of PoF as $\text{PoF}(c) = \frac{\text{Clustering Cost of Solution c}}{\text{Clustering Cost of Optimal Agnostic Solution}}$, this suffices to show Utilitarian objective is not monotonic in PoF because the denominator Clustering Cost of Optimal Agnostic Solution is a fixed constant for any instance.

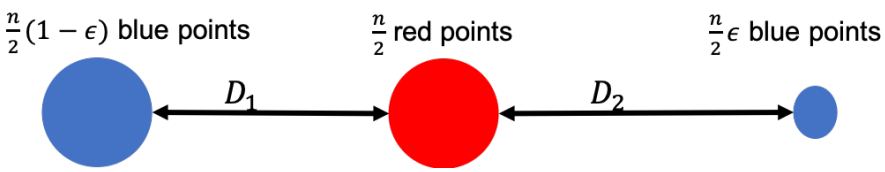


Figure 6: Clustering instance.

Recall the Utilitarian objective definition and on this instance we have $U = \frac{1}{n_R}(\lambda\,\mathcal{D}_R + (1-\lambda)\,\mathcal{V}_R) + \frac{1}{n_B}(\lambda\,\mathcal{D}_B + (1-\lambda)\,\mathcal{V}_B)$, where $\mathcal{D}$ and $\mathcal{V}$ are calculated for the two groups red and blue. Suppose $\lambda D_1 < 2(1-\lambda)$, and $\lambda D_2 > 2(1-\lambda)$ for the given fixed λ used in utilitarian objective, and allowed violation slack $\alpha_R = \beta_R = \alpha_B = \beta_B = 0$. We show there exist three solutions with distance cost $\mathcal{B}_0 < \mathcal{B}_1 < \mathcal{B}_2$ such that their corresponding Utilitarian objective value satisfy $U_0 > U_1$ and $U_1 < U_2$.

Consider a solution denoted as "Solution 0" with clustering distance cost $\mathcal{B}_0 = 0$. FCBC's only option is to form three pure clusters at the initial locations. For this solution the two distances costs are 0: $\mathcal{D}_R = 0$, $\mathcal{D}_B = 0$. The two violation costs are $\mathcal{V}_R = 0.5 \cdot n_R + 0.5 \cdot n_{B1} + 0.5 \cdot n_{B2} = 0.5 \cdot (n_R + n_{B1} + n_{B2})$. Since $n = n_R + n_{B1} + n_{B2}$, this simplifies to:$\mathcal{V}_R = 0.5 \cdot n = n/2$. The calculation for $\mathcal{V}_B$ is symmetric and also yields $n/2$. Combining these values, we get the Utilitarian objective value U_0 for Solution 0 is

$$
\begin{aligned}
U_0 &= \frac{1}{n_R}((1-\lambda)\frac{n}{2}) + \frac{1}{n_B}((1-\lambda)\frac{n}{2}) \\
&= \frac{1}{n/2}\left((1-\lambda)\frac{n}{2}\right) + \frac{1}{n/2}\left((1-\lambda)\frac{n}{2}\right) = 2(1-\lambda).
\end{aligned}
$$

Now consider another solution denoted "Solution 1" Suppose FCBC algorithm is run with distance upper bound set as $\mathcal{B}_1 = n_{B1} D_1$. With this distance budget, FCBC merges the nearby Red group and small Blue sub-group.

Distance costs for the two groups are $\mathcal{D}_R = 0$ and $\mathcal{D}_B = n_{B1}D_1$ respectively. We sum the violation contributions from each cluster for the Red group.

$$\begin{aligned}
\mathcal{V}'_R &= |\,C_{i1}\,| \cdot \left| \frac{|\,C^h_{i\,1}\,|}{|\,C_{i1}\,|} - r_{hR} \right| + |\,C_{i2}\,| \cdot \left| \frac{|\,C^h_{i\,2}\,|}{|\,C_{i2}\,|} - r_{hR} \right| \\
&= (n_R + n_{B1}) \cdot \left| \frac{n_R}{n_R + n_{B1}} - 0.5 \right| + n_{B2} \cdot \left| \frac{0}{n_{B2}} - 0.5 \right| \\
&= |n_R - 0.5(n_R + n_{B1})| + n_{B2} \cdot |-0.5| \\
&= |0.5n_R - 0.5n_{B1}| + 0.5n_{B2} \\
&= 0.5(n_R - n_{B1}) + 0.5n_{B2} \quad (\text{since } n_R > n_{B1}) \\
&= 0.5((n_{B1} + n_{B2}) - n_{B1}) + 0.5n_{B2} \quad (\text{substituting } n_R = n_{B1} + n_{B2}) \\
&= 0.5(n_{B2}) + 0.5n_{B2} \\
&= n_{B2}
\end{aligned}$$

Similarly, we sum the contributions for the Blue group.

$$\begin{aligned}
\mathcal{V}'_B &= |\,C_{i1}\,| \cdot \left| \frac{|\,C^h_{i\,1}\,|}{|\,C_{i1}\,|} - r_{hB} \right| + |\,C_{i2}\,| \cdot \left| \frac{|\,C^h_{i\,2}\,|}{|\,C_{i2}\,|} - r_{hB} \right| \\
&= (n_R + n_{B1}) \cdot \left| \frac{n_{B1}}{n_R + n_{B1}} - 0.5 \right| + n_{B2} \cdot \left| \frac{n_{B2}}{n_{B2}} - 0.5 \right| \\
&= |n_{B1} - 0.5(n_R + n_{B1})| + n_{B2} \cdot |1 - 0.5| \\
&= |0.5n_{B1} - 0.5n_R| + 0.5n_{B2} \\
&= 0.5(n_R - n_{B1}) + 0.5n_{B2} \quad (\text{since } n_R > n_{B1}) \\
&= 0.5(n_{B2}) + 0.5n_{B2} \quad (\text{again substituting } n_R - n_{B1} = n_{B2}) \\
&= n_{B2}
\end{aligned}$$

Thus Utilitarian objective value of this solution is

$$\begin{aligned}
U_1 &= \frac{1}{n_R}((1-\lambda)n_{B2}) + \frac{1}{n_B}(\lambda n_{B1}D_1 + (1-\lambda)n_{B2}) \\
&= \frac{1}{n/2}\left[\lambda n_{B1}D_1 + 2(1-\lambda)n_{B2}\right].
\end{aligned}$$

Given our condition $\lambda D_1 < 2(1-\lambda)$, it follows that $U_1 < U_0$.

Now consider another solution which we denote as "Solution 2". When given clustering distance budget $\mathcal{B}_2 = n_{B1}D_1 + mD_2 > B_1$, FCBC algorithm must start moving a few (m) Red points the long distance D_2 to improve fairness further. In this case, distance costs are $\mathcal{D}_R = mD_2$, $\mathcal{D}_B = n_{B1}D_1$. Let the new violations be $\mathcal{V}''_R$ and $\mathcal{V}''_B$. We calculate the new violation for the Red group, $\mathcal{V}''_R$:

$$\begin{aligned}
\mathcal{V}''_R &= |\,C_{i1}\,| \left| \frac{n_R - m}{|\,C_{i1}\,|} - r_R \right| + |\,C_{i2}\,| \left| \frac{m}{|\,C_{i2}\,|} - r_R \right| \\
&= (n_R - m + n_{B1}) \left| \frac{n_R - m}{n_R - m + n_{B1}} - 0.5 \right| + (m + n_{B2}) \left| \frac{m}{m + n_{B2}} - 0.5 \right| \\
&= |(n_R - m) - 0.5(n_R - m + n_{B1})| + |m - 0.5(m + n_{B2})| \\
&= |0.5(n_R - m) - 0.5n_{B1}| + |0.5m - 0.5n_{B2}| \\
&= 0.5(n_R - n_{B1} - m) + 0.5(n_{B2} - m) \quad (\text{since } n_R - n_{B1} = n_{B2} > m) \\
&= 0.5(n_{B2} - m) + 0.5(n_{B2} - m) = n_{B2} - m
\end{aligned}$$

A symmetric calculation for the Blue group shows $\mathcal{V}''_B = n_{B2} - m$.

The reduction in violation for each group when moving from Solution 1 to Solution 2 is:

$$\Delta\,\mathcal{V}_R = \mathcal{V}'_R - \mathcal{V}''_R = n_{B2} - (n_{B2} - m) = m$$
$$\Delta\,\mathcal{V}_B = \mathcal{V}'_B - \mathcal{V}''_B = n_{B2} - (n_{B2} - m) = m$$

Utilitarian objective value of Solution 2 is: $U_2 = \frac{1}{n_R}(\lambda m D_2 + (1-\lambda)\,\mathcal{V}''_R) + \frac{1}{n_B}(\lambda n_{B1} D_1 + (1-\lambda)\,\mathcal{V}''_B)$.
We show $U_2 > U_1$ given $\lambda D_2 > 2(1-\lambda)$. From $\lambda D_2 > 2(1-\lambda)$, and $m > 0$ we have

$$\frac{\lambda m D_2}{n_R} > (1-\lambda)\frac{2m}{n_R} = (1-\lambda)\left(\frac{m}{n_R} + \frac{m}{n_R}\right)$$

Now, we substitute the violation reductions we calculated ($\Delta\,\mathcal{V}_R = \Delta\,\mathcal{V}_B = m$) and the fact that $n_R = n_B$, we have

$$\frac{\lambda m D_2}{n_R} - \frac{(1-\lambda)(\mathcal{V}'_R - \mathcal{V}''_R)}{n_R} - \frac{(1-\lambda)(\mathcal{V}'_B - \mathcal{V}''_B)}{n_B} > 0$$

Since $\mathcal{D}'_R = 0$, $\mathcal{D}''_R = mD_2$, and $\mathcal{D}'_B = \mathcal{D}''_B = n_{B1}D_1$, we obtain

$$\frac{\lambda\mathcal{D}''_R + (1-\lambda)\mathcal{V}''_R}{n_R} + \frac{\lambda\mathcal{D}''_B + (1-\lambda)\mathcal{V}''_B}{n_B} > \frac{\lambda\mathcal{D}'_R + (1-\lambda)\mathcal{V}'_R}{n_R} + \frac{\lambda\mathcal{D}'_B + (1-\lambda)\mathcal{V}'_B}{n_B}.$$

Therefore, $U_2 > U_1$.

□

H Additional Experiments and Results

In this section, we give details of the experiments and show additional experiment results.

H.1 Details for Datasets Used in Experiments

Dataset	**Dimension**	**Number of Points (Full)**	**Number of Points Subsampled: 10% vs. 90%**	**Group Label**	α_h,β_h
Adult	5	32561	24211	gender	$0.01r_h$
Credit card	14	29623	17737	marital-status	$0.01r_h$
Census II (subsampled)	67	20,000	11400	gender	$0.01r_h$

Table 1: Description of the four datasets used in two group empirical evaluations.

We select the same group labels and features for **Adult**, **CreditCard**, and **Census1990** as done in Bera et al. (2019); Esmaeili et al. (2020). On **Adult**, attributes "age", "final-weight", "education-num", "capital-gain", "hours-per-week" are used as features. On **CreditCard** and **Census1990**, all attributes except group label are used as features. Gender is used as the group label in **Adult** and **Census1990**. On **CreditCard**, we use marital status.

Dataset	**Dimension**	**Number of points**	**Group Label**	α_h,β_h
Bank	3	4521	marital-status ($\mid\mathcal{H}\mid = 3$)	$0.01r_h$

Table 2: Description of dataset used in our multi-group empirical evaluation.

On **Bank**, attributes chosen to represent point in space are "age", "balance" and "duration-of-account", the same as in Backurs et al. (2019).

H.2 Implementation Details

Existing Implementation of Socially Fair and FCBC We use existing implementation of socially fair k-means from Ghadiri et al. (2021) [††] which accommodates groups with two colors but not multiple colors. Note existing code for socially fair is in written Matlab, we translate to equivalent Python implementation. We use existing code for FCBC implementation [‡‡].

Randomness in Center Selection Both our algorithms and FCBC variants uses some form of k-means to select initial centers (socially fair can be setup to runs k-means to select initial centers instead of randomly select). Thus randomness in the experimental pipeline lies in the initial centers selected by each algorithm. To minimizes the effects of selecting bad centers, we let all algorithms being compared each run k-means++ to initialize centers and select the best out of 10 sets of centers.

Translation of Parameters α, β between Our Algorithms and FCBC As mentioned in Section 5, we set $\alpha_h = \beta_h = \delta r_h$, where $\delta = 0.01$. Note the same symbol α_h and β_h refer to different concepts in paper Esmaeili et al. (2021) where FCBC algorithms were originally introduced. Denote the upper and lower bounds in Esmaeili et al. (2021) as $\bar{\alpha}_h$, $\bar{\beta}_h$. To fairly compare with FCBC algorithms, we set $\bar{\alpha}_h = r_h + \alpha_h$, $\bar{\beta}_h = r_h + \beta_h$, where α_h, β_h are parameters for our formulation.

H.3 Data Normalization

We normalize the data so that the average clustering distance under vanilla k-means is on the same order as its proportional violation. This ensures that the distance and violation terms are comparable in magnitude and therefore avoids extreme values for the trade-off parameter λ. We now introduce details of normalization factor calculation. Denote a normalization factor for each k as F^k_{Rawl} and F^k_{Util}. We calculate F^k_{Rawl} as

$$F^k_{\text{Rawl}} = \frac{\frac{1}{|\mathcal{C}|}\sum_{j\in\mathcal{C}} d^p(j,\phi(j))}{\sum_{h\in\mathcal{H}} \frac{\mathcal{V}_h(S,\phi)}{|\mathcal{C}^h|}}$$

where (S,ϕ) is vanilla k-means' solution with $p=2$, $|S|=k$. To run algorithm for intended k range $[K]$, we normalize once with $F_{\text{Rawl}} = \text{mean}(F^k_{\text{Rawl}})$, where mean is taken over the range $[K]$. Denote each data point j's feature vector from as x_j. Data is then normalized as

$$\mathbf{x}_j = \mathbf{x}_j / \sqrt{F_{\text{Rawl}}}.$$

We calculate F^k_{Util} as

$$F^k_{\text{Util}} = \frac{\sum_{h\in\mathcal{H}} \frac{1}{|\mathcal{C}^h|}\sum_{j\in\mathcal{C}^h} d^p(j,\phi(j))}{\sum_{h\in\mathcal{H}} \frac{\mathcal{V}_h(S,\phi)}{|\mathcal{C}^h|}}$$

where (S,ϕ) is k-means' solution with $p=2$, $|S|=k$. F_{Util} is similarly calculated as a mean when k is selected to be a range. Data for Utilitarian comparison are normalized as

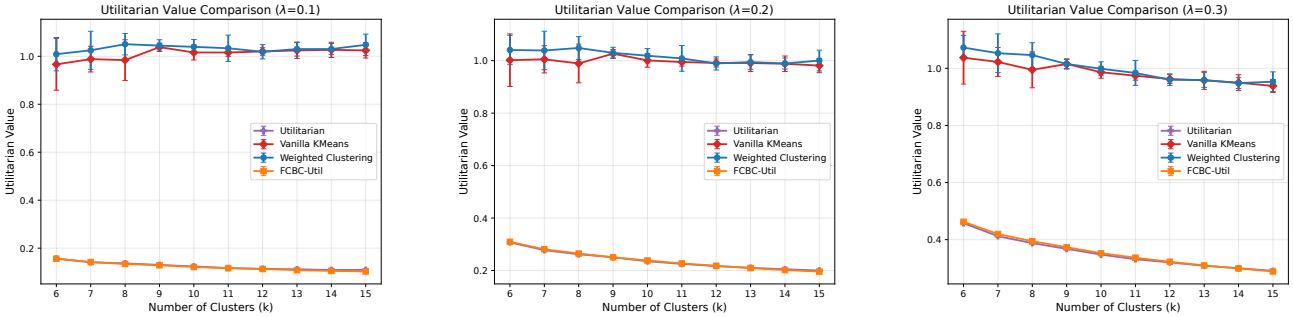
$$\mathbf{x}_j = \mathbf{x}_j / \sqrt{F_{\text{Util}}}.$$

H.4 Experimental Results in Setting with Multiple Groups

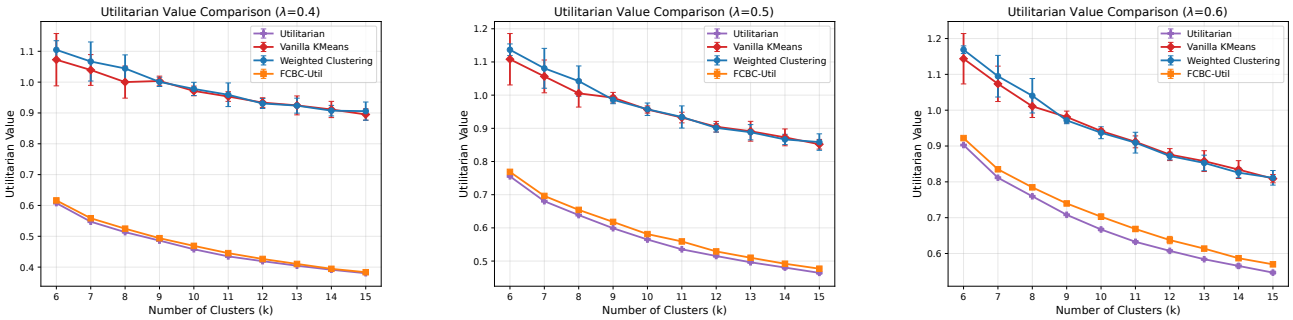
In this section, we present comparison of Utilitarian algorithm UtilitarianAlg with three other algorithms on settings with multiple groups. Below are plots for **Bank** dataset, where marital status is chosen as group label forming three groups with proportions about $12\%, 62\%$ and 26%. See Table 2 for more details about dataset size. For FCBC-Util algorithm, we set its upper bound on clustering distance cost as 1.5 times k-means vanilla distance cost. Parameter δ is set as 0.01.

[††] https://github.com/samirasamadi/SociallyFairKMeans

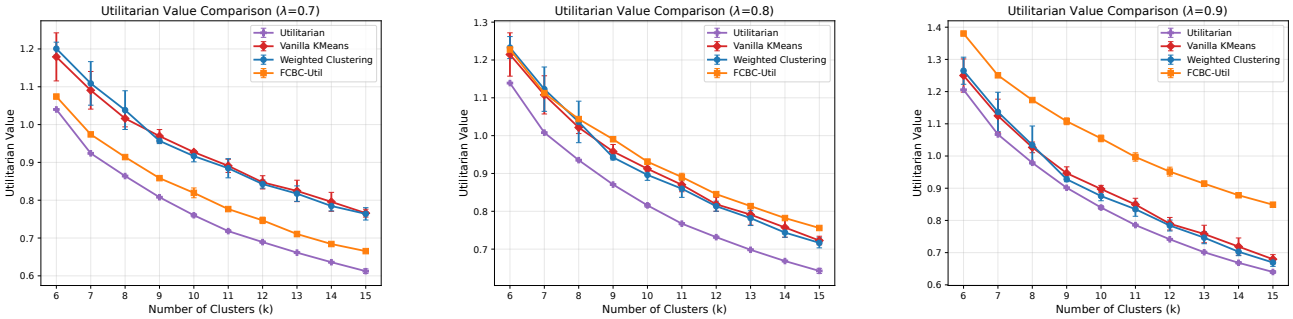
[‡‡] https://github.com/Seyed2357/Fair-Clustering-Under-Bounded-Cost



(a) Utilitarian $\lambda = 0.1$ (b) Utilitarian $\lambda = 0.2$ (c) Utilitarian $\lambda = 0.3$



(d) Utilitarian $\lambda = 0.4$ (e) Utilitarian $\lambda = 0.5$ (f) Utilitarian $\lambda = 0.6$



(g) Utilitarian $\lambda = 0.7$ (h) Utilitarian $\lambda = 0.8$ (i) Utilitarian $\lambda = 0.9$

Figure 7: Bank Data Utilitarian Value Comparison $|\mathcal{H}| = 3$

We do not include Rawlsian comparisons in the multiple-group setting, as the existing implementation of the k-means socially fair algorithm—which we use as the first step in RawlsianAlg—is designed for two groups and only provides theoretical guarantees in that setting; its extension to multiple groups relies on a heuristic without formal guarantees.

H.5 More Experimental Results in Setting with Two Groups

In the Experiment section in the main body Section 5, we presented experimental results and plots for $\lambda = 0.7$ for three subsampled datasets Adult, Credit Card and Census. In this section, we present plots for all other λ values in $\{0.1, 0.2, \ldots, 0.9\}$ for the subsampled dataset, as well as on full dataset. The proportion of groups on full **Adult** data is about 33% vs. 67%, on full **CreditCard** data is about 46% vs. 54%, and on full **Census1990** data is about 49% vs. 51%. We omit $\lambda = 0$ since it is equivalent to balancing proportions. We also omit $\lambda = 1.0$ since it is equivalent with running socially fair or weighted clustering. We observe that for all λ values, our proposed algorithm RawlsianAlg and UtilitarianAlg find clustering solutions that dominate the other algorithms' solutions on the two objectives.

Below we first present a set of plots (Figure 8, Figure 9, and Figure 10) for Rawlsian objective for $\lambda \in \{0.1, 0.2, \ldots, 0.9\}$, and a set of plots (Figure 11, Figure 12, and Figure 13) for Utilitarian objective for $\lambda \in \{0.1, 0.2, \ldots, 0.9\}$ on the subsampled data (with proportion 10% vs. 90%).

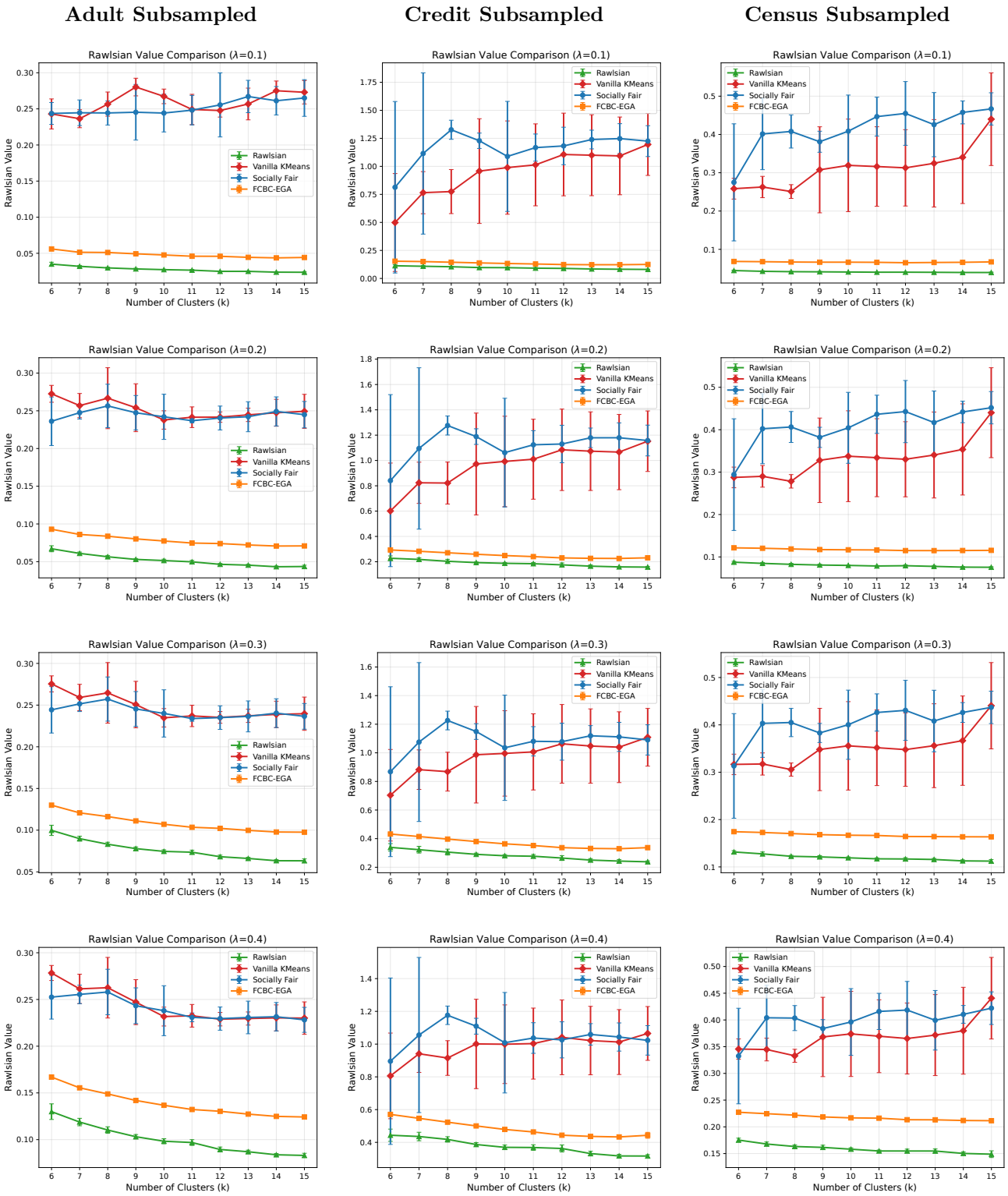


Figure 8: Rawlsian Value comparison for $\lambda = 0.1$ to 0.4 on subsampled data (10% vs. 90%).

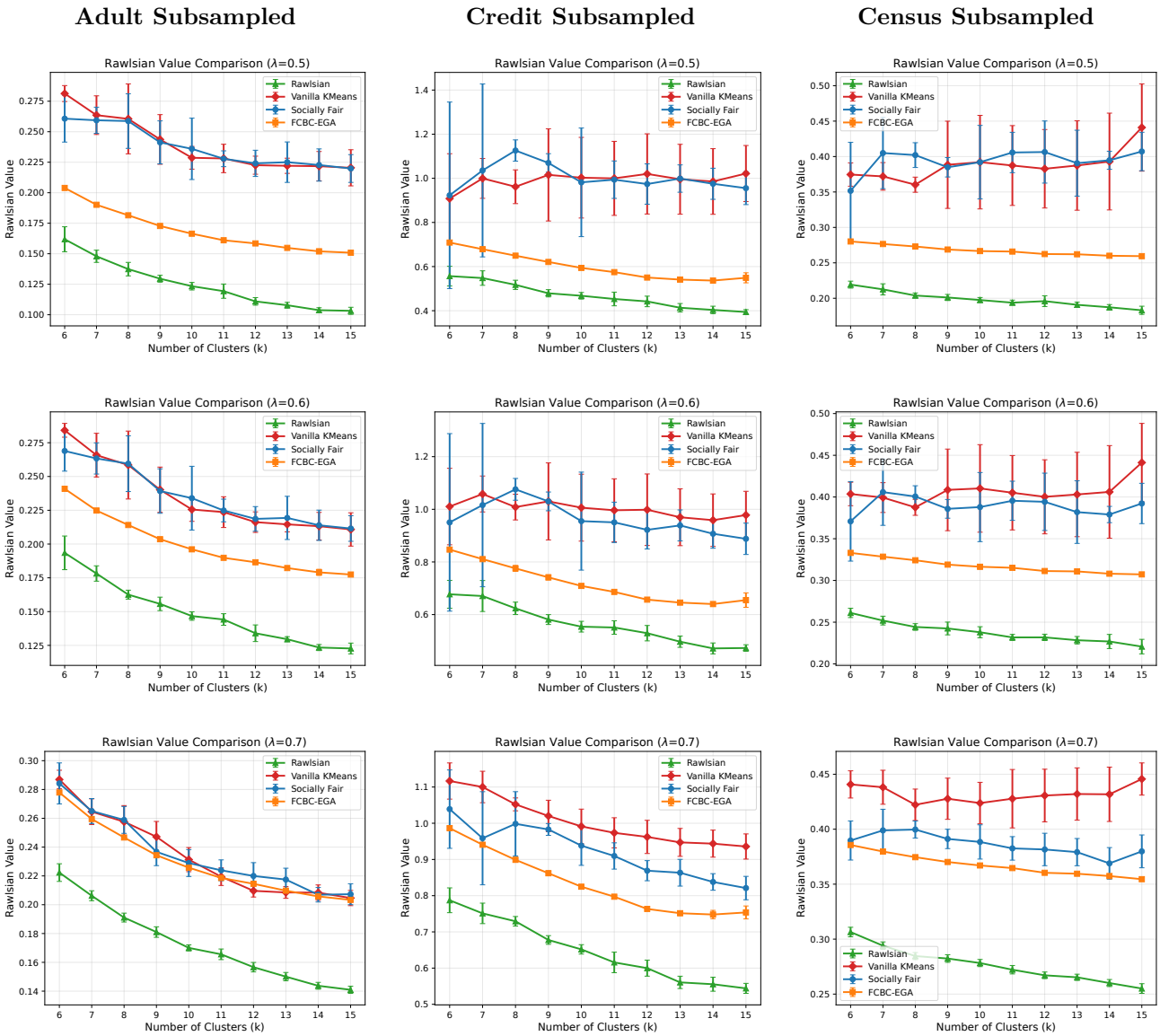


Figure 9: Rawlsian Value comparison for $\lambda = 0.5$ to 0.7 on subsampled data (10% vs. 90%).

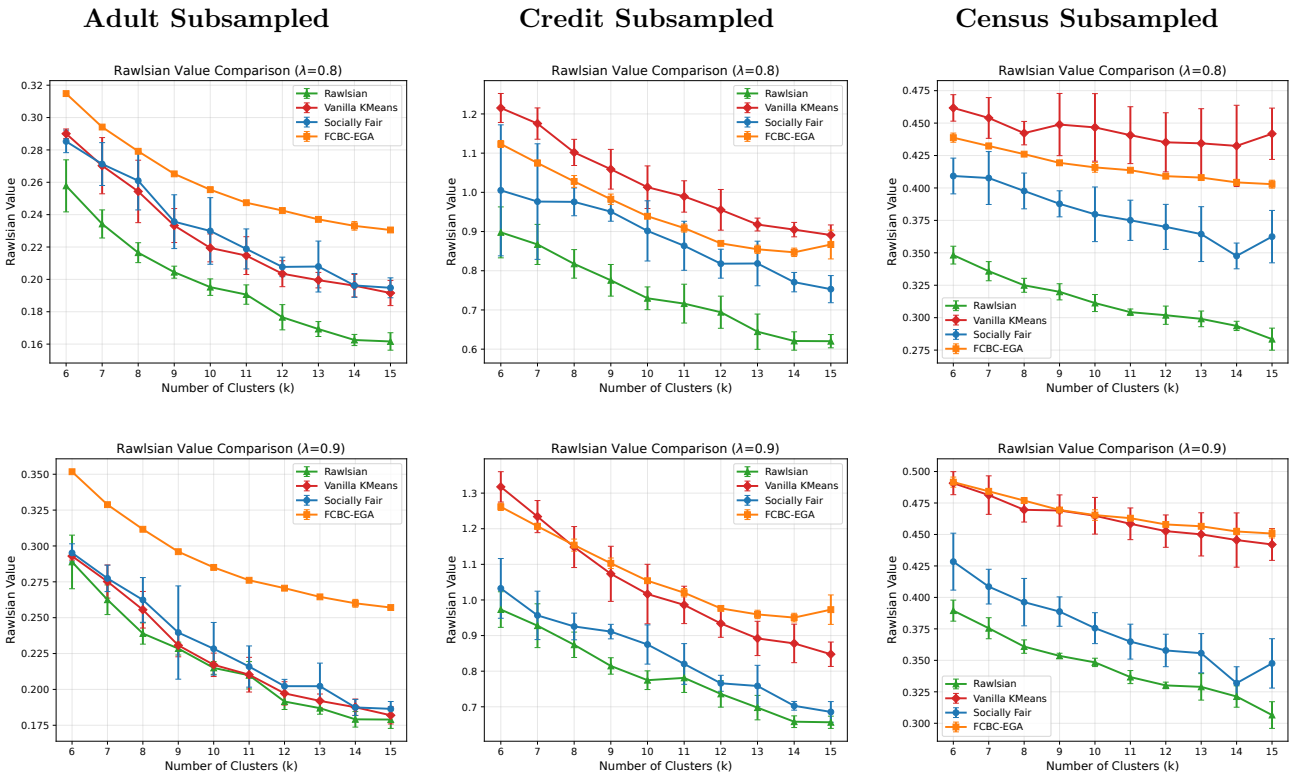


Figure 10: Rawlsian Value comparison for $\lambda = 0.8$ to 0.9 on subsampled data (10% vs. 90%).

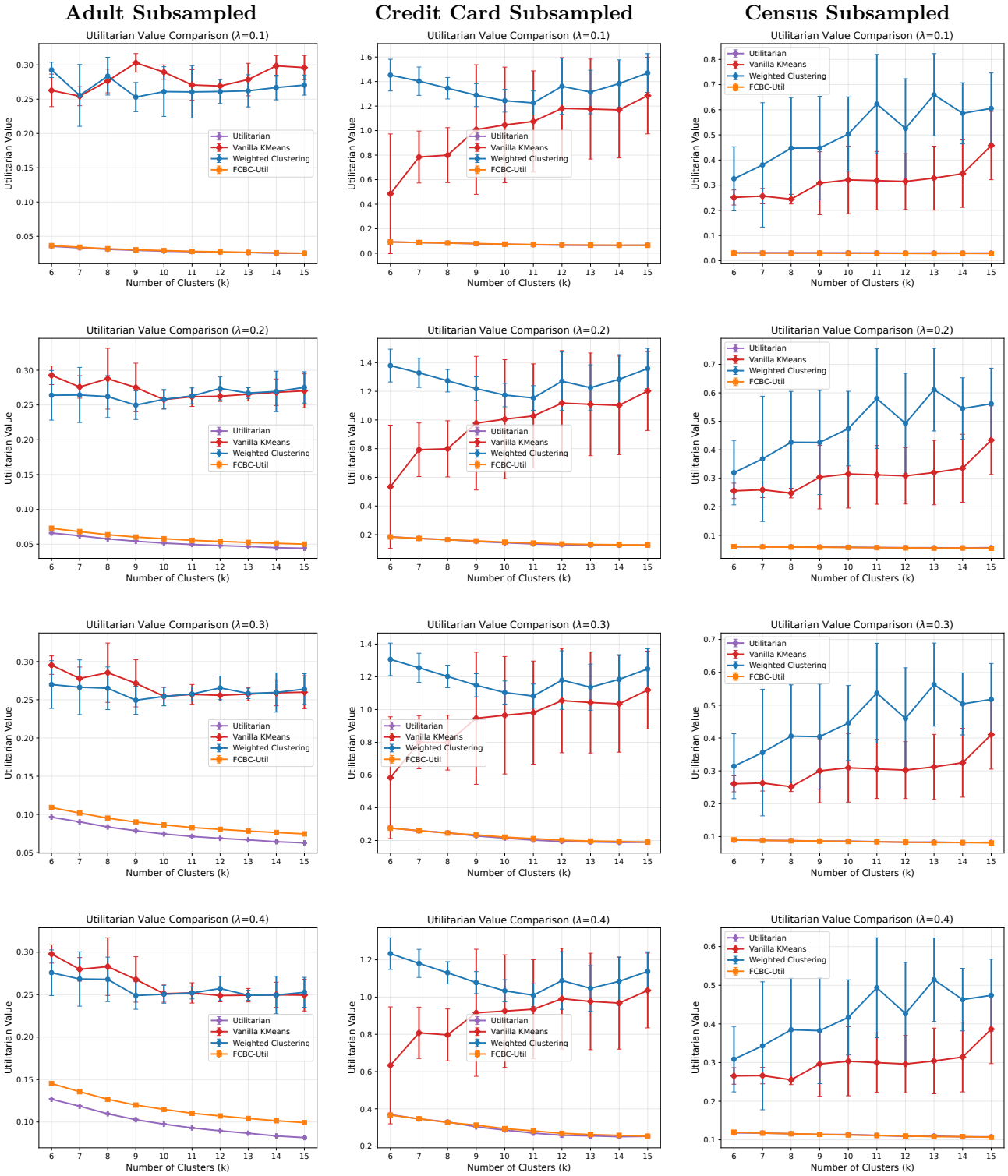


Figure 11: Utilitarian Value comparison for $\lambda = 0.1$ to 0.4 on subsampled data (10% vs. 90%).

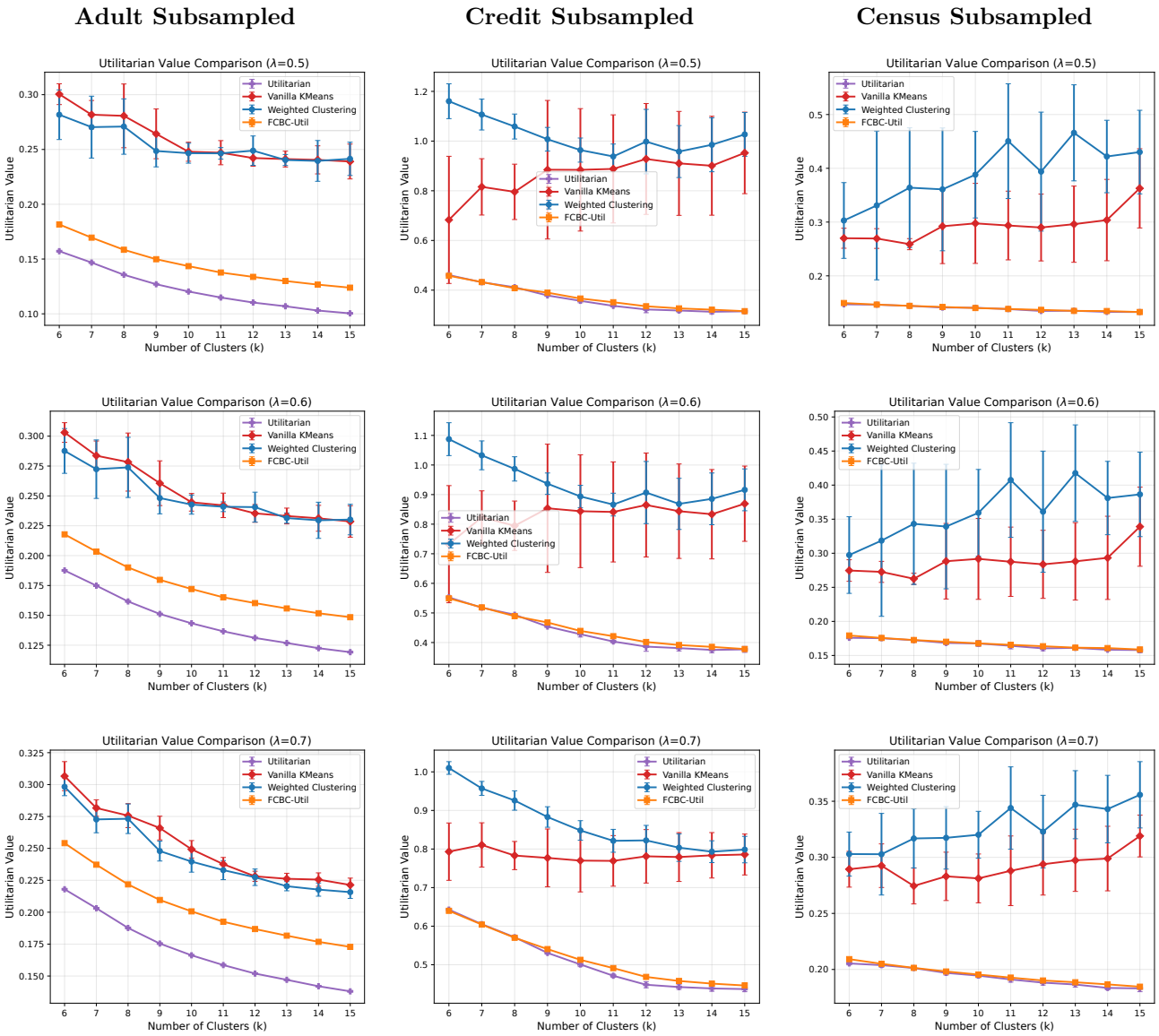


Figure 12: Utilitarian Value comparison for $\lambda = 0.5$ to 0.7 on subsampled data (10% vs. 90%).

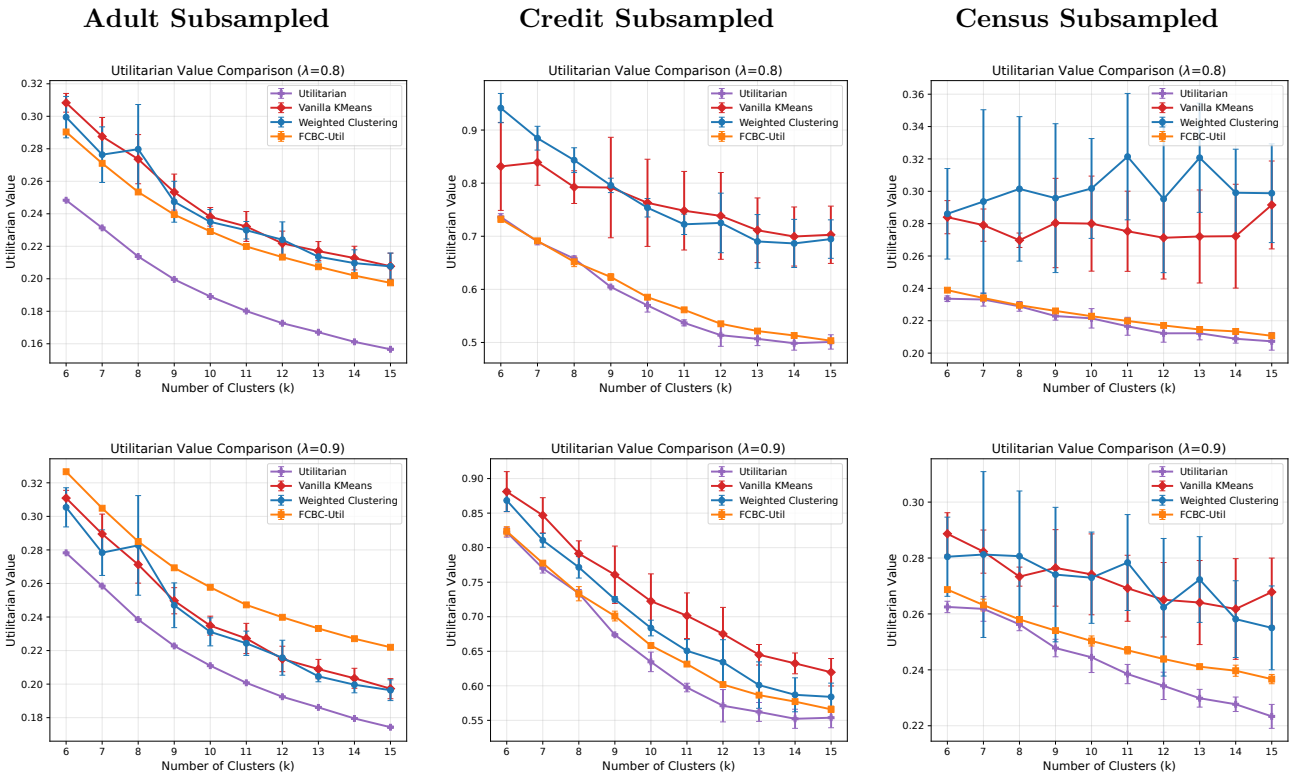


Figure 13: Utilitarian Value comparison for $\lambda = 0.8$ to 0.9 on subsampled data (10% vs. 90%).

Below we present a set of plots (Figure 14, Figure 15, and Figure 16) for Rawlsian objective for $\lambda \in \{0.1, 0.2, \ldots, 0.9\}$, and a set of plots (Figure 17, Figure 18, and Figure 19) for Utilitarian objective for $\lambda \in \{0.1, 0.2, \ldots, 0.9\}$ on the full data.

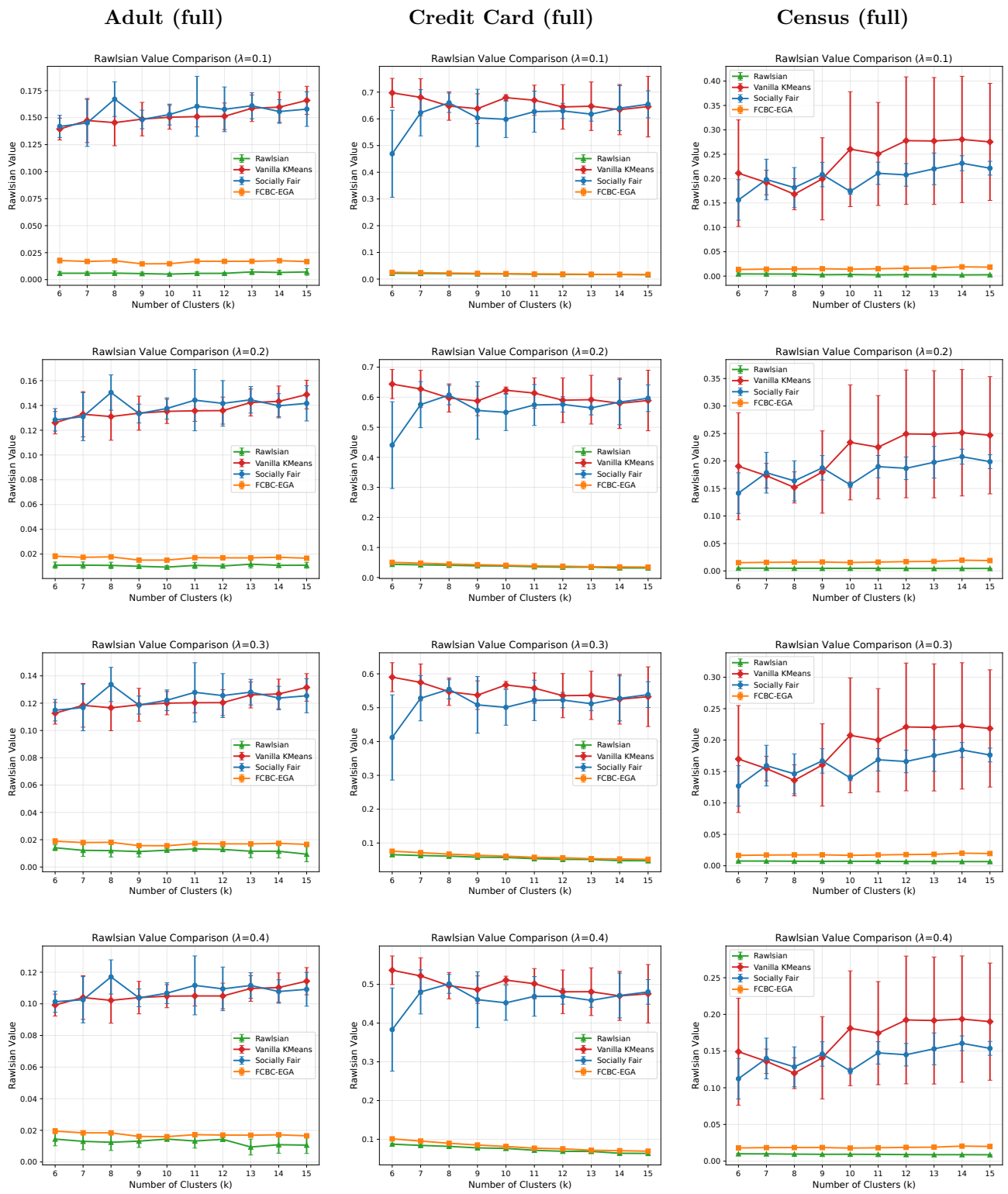


Figure 14: Rawlsian Value comparison for $\lambda = 0.1$ to 0.4 on full data.

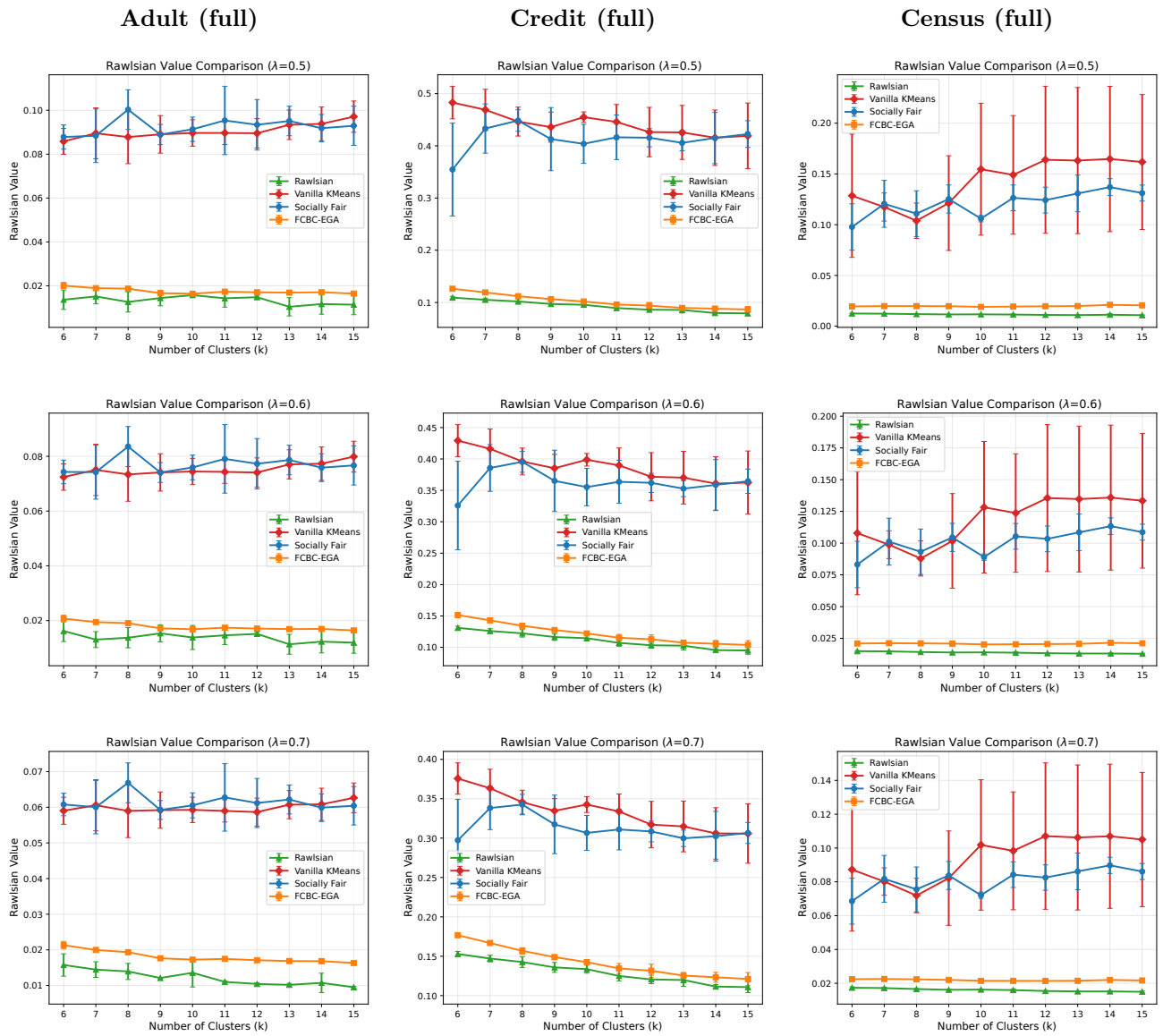


Figure 15: Rawlsian Value comparison for $\lambda = 0.5$ to 0.7 on full data.

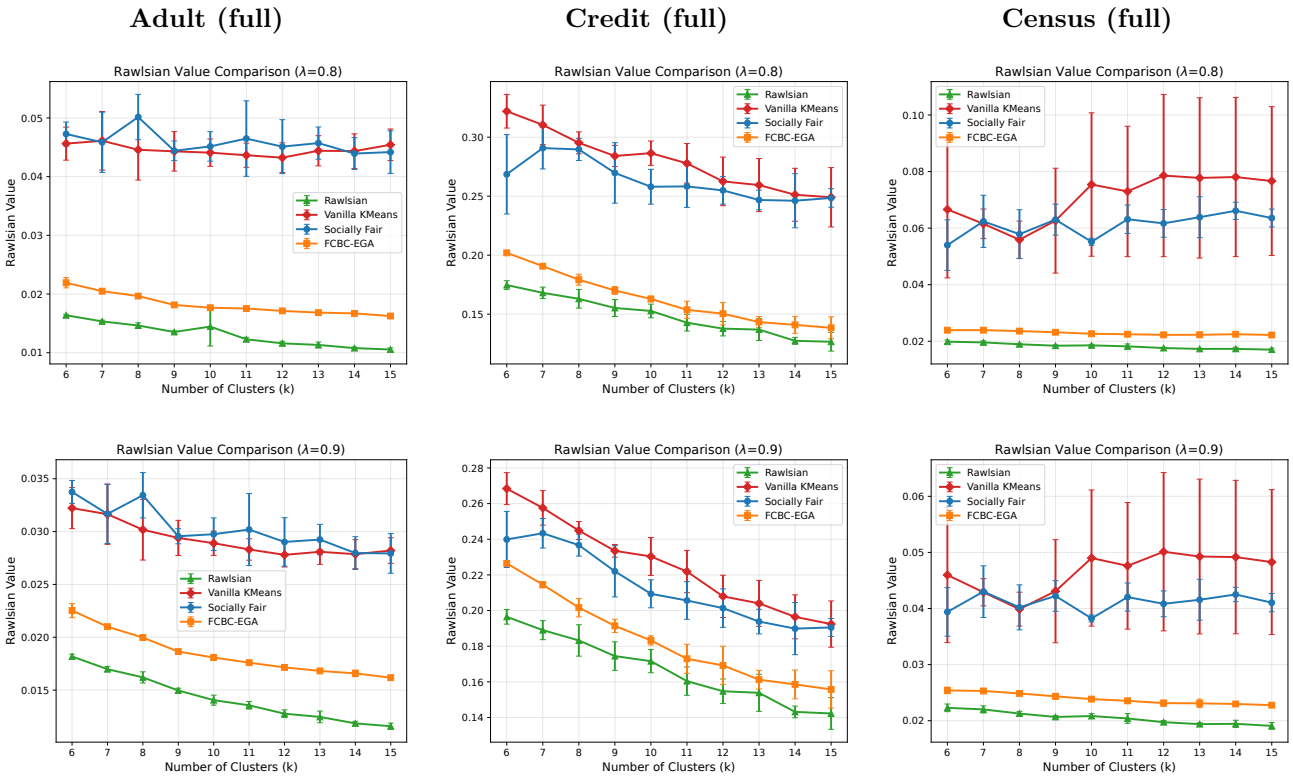


Figure 16: Rawlsian Value comparison for $\lambda = 0.8$ to 0.9 on full data.

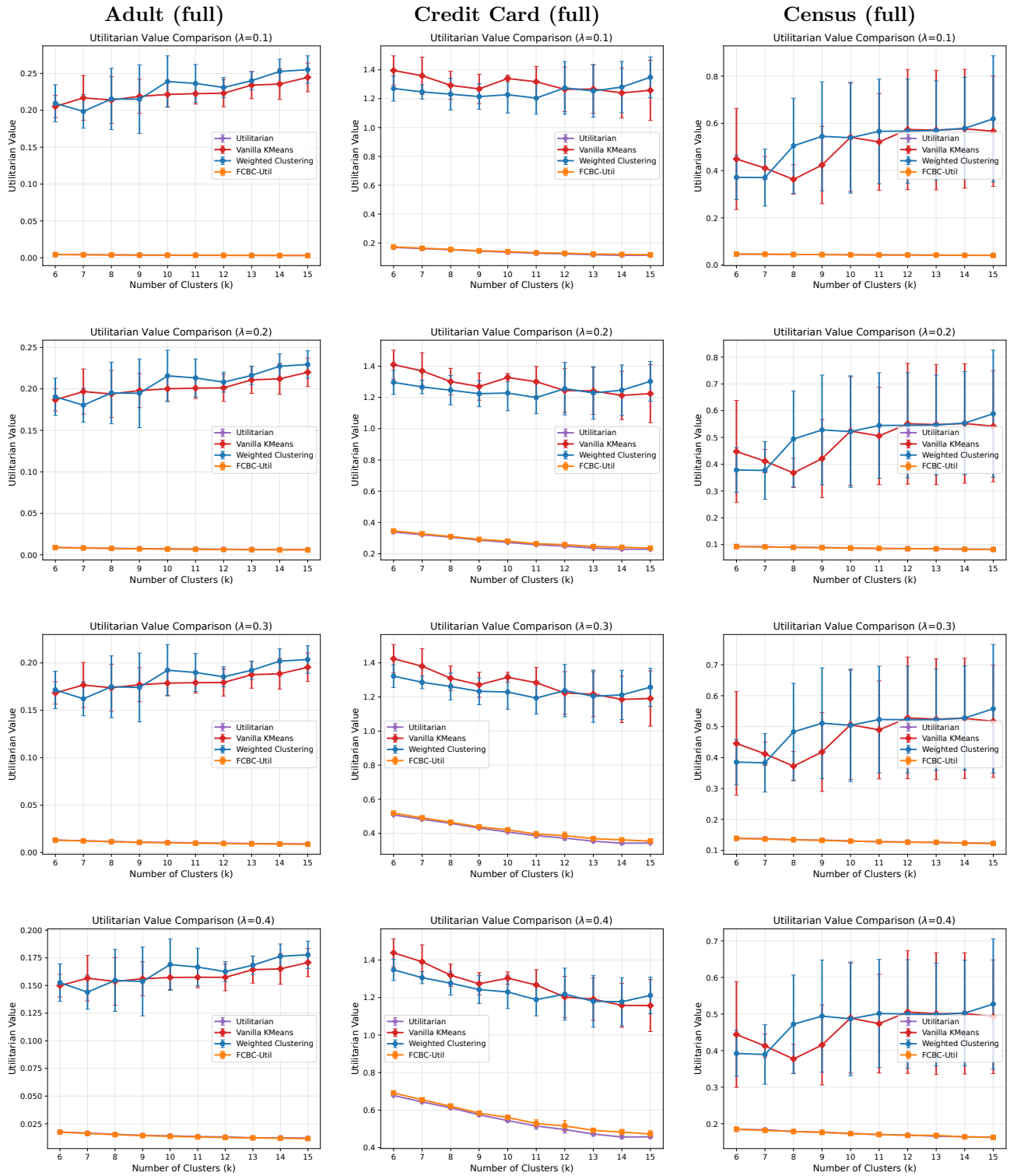


Figure 17: Utilitarian Value comparison for $\lambda = 0.1$ to 0.4 on full data.

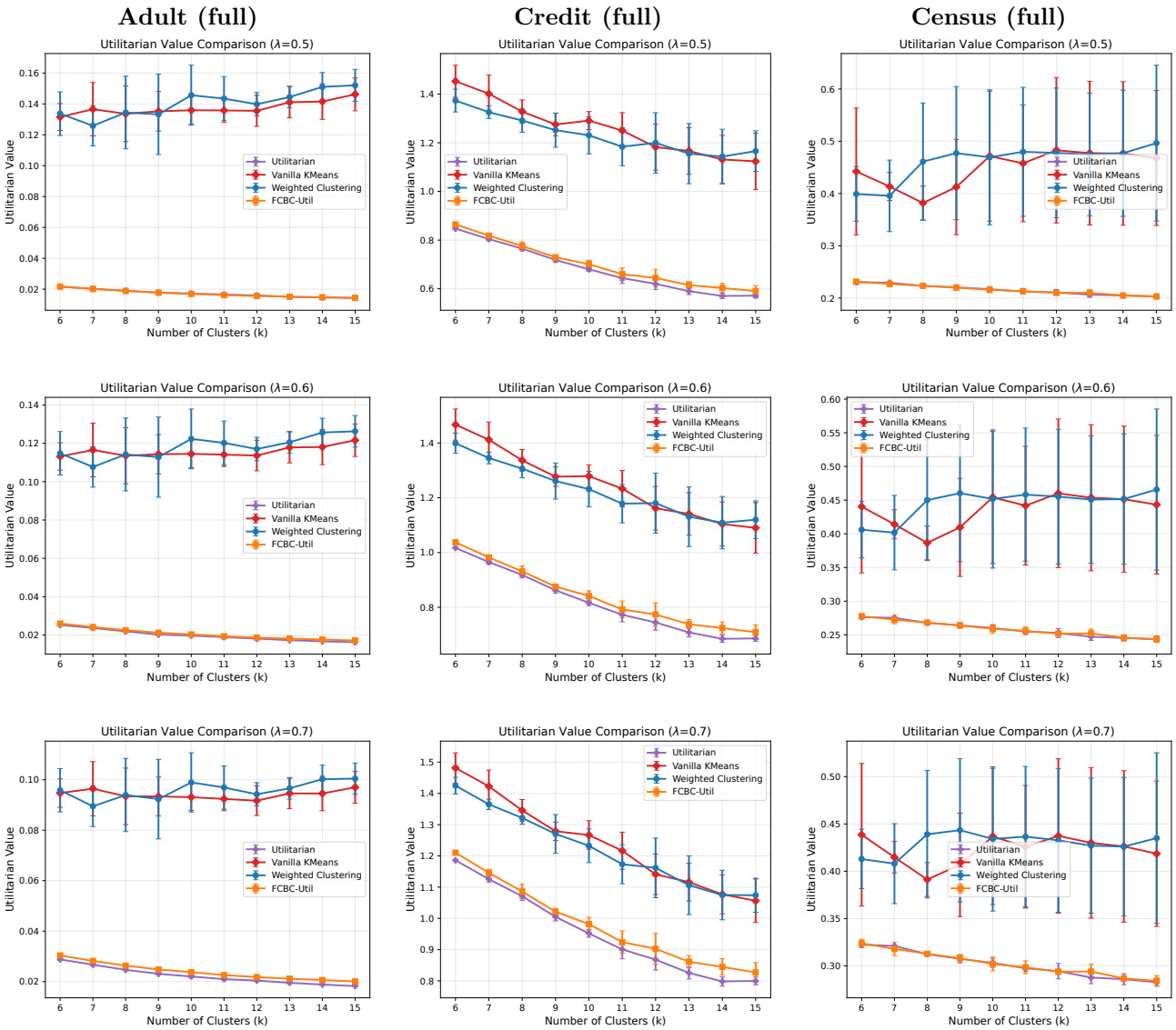


Figure 18: Utilitarian Value comparison for $\lambda = 0.5$ to 0.7 on full data.

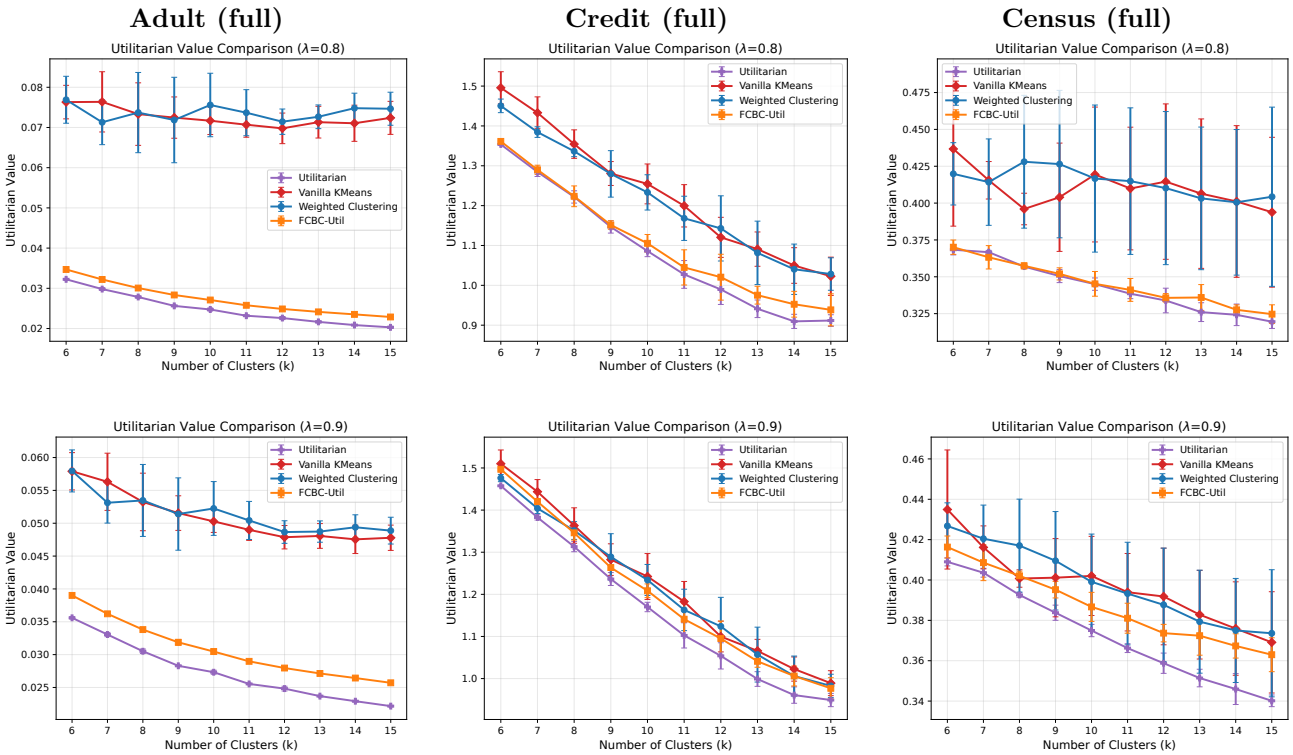


Figure 19: Utilitarian Value comparison for $\lambda = 0.8$ to 0.9 on full data.

H.6 Per-group Distance Cost and Representation Violation

In this section, we present plots comparing per group distance cost and representation violation, to highlight the differences in clustering solutions found by the algorithms being compared. We observe from plots in Figure 20 that our Rawlsian algorithm arrives at a solution with two groups having close distance costs and representation violation cost. Clustering solutions solved by Vanilla KMeans and Socially Fair have similar distance costs for both groups as our Rawlsian algorithm, but with significant disparate representation violation between the two groups. Solution solved by FCBC-Ega on the other hand has similar representation violation cost for the two groups, but disparate distance costs. We observe similar differences between the four algorithms compared for the Utilitarian objective from Figure 21.

H.7 Comparison with FCBC with other Cost Upper Bounds

In this section, we provide details of the tuning done for FCBC-Rawl and FCBC-Util. The two variants of FCBC algorithms solve constrained minimization problems where their objectives are variants of proportional representation violation, and constraint is an upper bound on allowed distance cost. To run the FCBC algorithms on any given problem instance, a user has to provide an upper bound on clustering distance cost which function as the constraint. In paper Esmaeili et al. (2021) and their implementation, this upper bound is calculated by first calculating cost of optimal vanilla (agnostic) clustering distance cost, and multiply by a constant which is denoted POF. Note POF has to be greater than 1 to allow for fair clustering solution Esmaeili et al. (2021); Dickerson et al. (2025).

As mentioned in the main text, since FCBC-Rawl and FCBC-Util do not optimize Welfare and Utilitarian objectives directly, one has to run the FCBC algorithms multiple times for each problem instance for one fixed weight parameter λ and number of centers k choice. That is, each FCBC algorithm run will have a different distance cost constraint upper bound (corresponding to a different POF value), forcing the FCBC algorithms to make trade-offs between proportional violation and distances costs this way. As shown in Section G, the objective values are not monotone with increase of distance upper bound so efficient searching algorithm such as bisection cannot be used. Note this makes FCBC algorithms not suitable for solving our objectives.

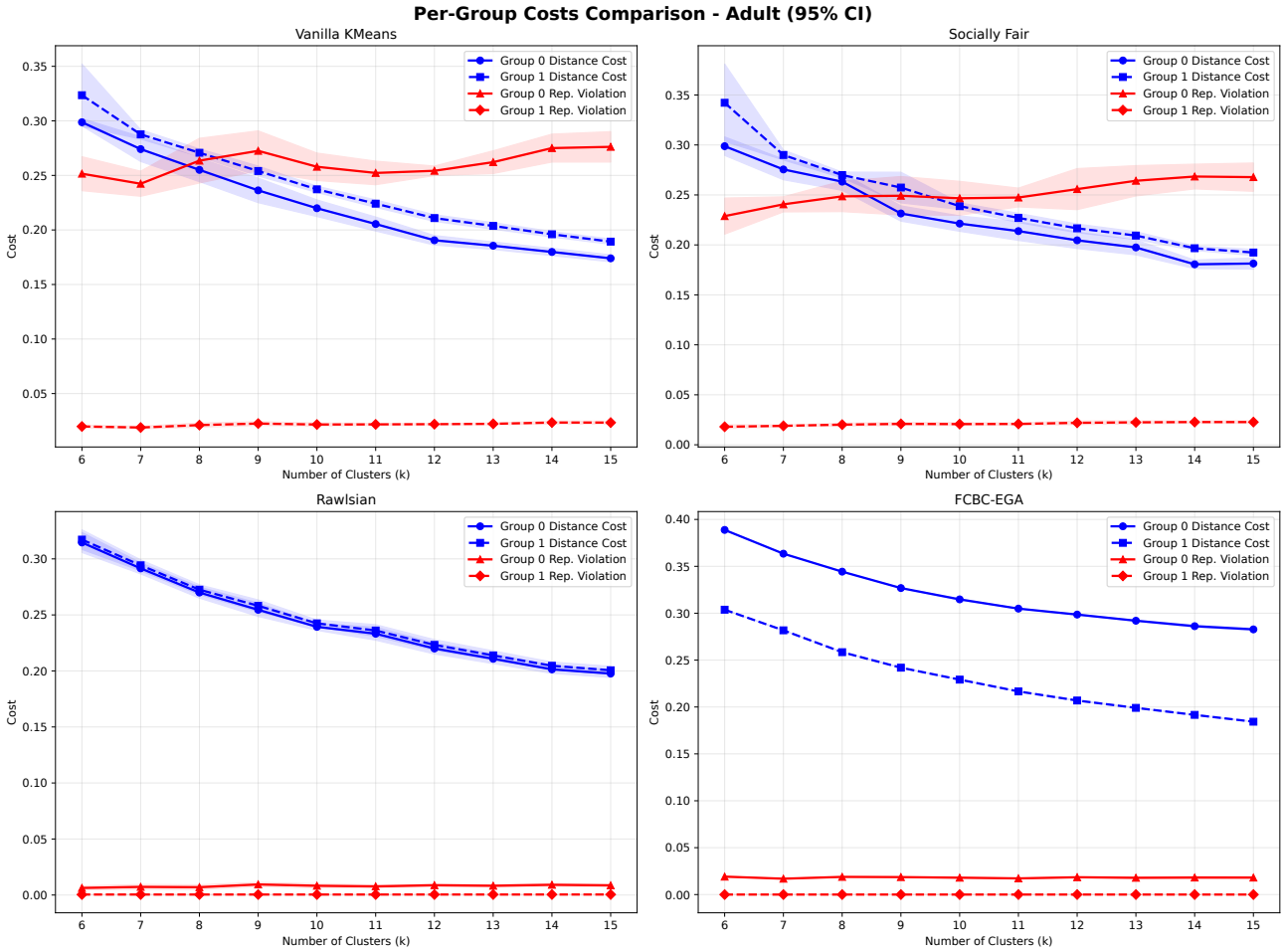


Figure 20: Per-group costs comparison. Rawlsian solution is found with $\lambda = 0.7$.

To compare the best Rawlsian value and Utilitarian value achievable by FCBC algorithms, on each data set and each λ, k choice, we run FCBC algorithms with POF $= \{1.05, 1.50, 2.00, 2.50, 3.00\}$. We show results of one such tuning effort for $\lambda = 0.7$ in Figure 22. Best Rawlsian and Utilitarian value (minimum) from those runs, represented by the red curve, are saved and compared with our algorithms Rawlsian and Utilitarian.

H.8 Performance Gap between FCBC Algorithms and Welfare Clustering Algorithms Proposed

In this section, we show how the performance gap between our algorithms RawlsianAlg and UtilitarianAlg with their FCBC counterparts being compared vary with proportion of the minority group, in a two group dataset **Adult**. Recall FCBC-Util algorithm minimizes maximum proportional violation over clusters and groups, under a given distance cost upper bound. However, unlike our formulation Rawlsian and Utilitarian, neither the proportional violation and distance cost is normalized by group's size. Given this main difference, FCBC algorithms when solving for fair solutions, will put less weights on minority groups, if the groups takes up small proportion in the whole dataset. While our formulation Rawlsian and Utilitarian always treats smaller minority group and majority group's loss equally by normalizing by their group size. Empirically we show that the smaller the minority group is, the better our algorithms RawlsianAlg and UtilitarianAlg outperform FCBC-EGA and FCBC-Util algorithms as shown in Figure 23. We show two plots measuring absolute gap in loss difference between two algorithms compared, and relative improvement. The confidence interval come from taking mean over number of centers k values ranging form 6 to 15.

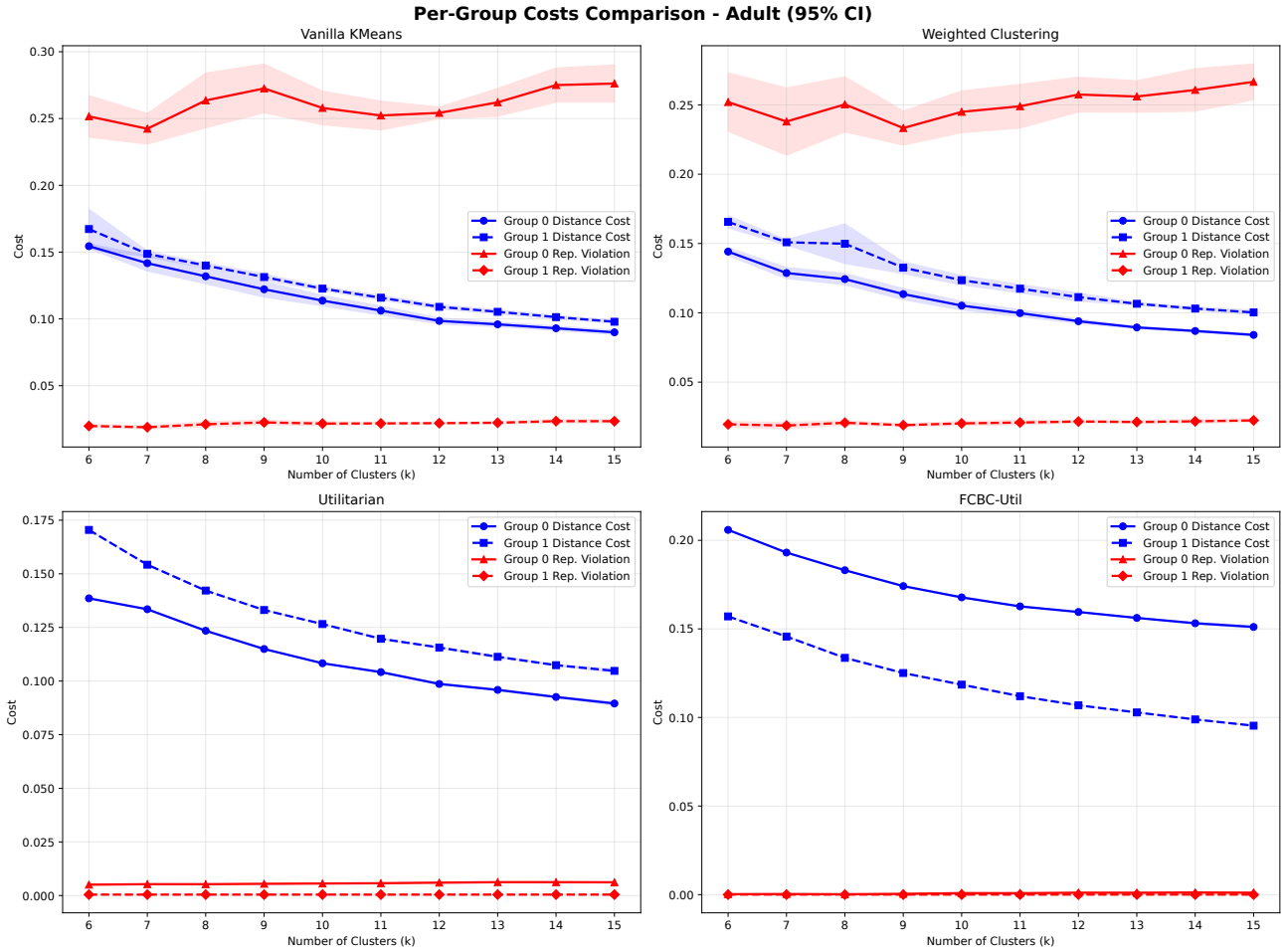


Figure 21: Per-group costs comparison. Utilitarian solution is found with $\lambda = 0.7$.

H.9 Effect of Rounding on the Rawlsian and Utilitarian

In this section, we empirically show that rounding does not change LP objective value too much for both Rawlsian and Utilitarian objectives. Here we are referring to the constants c_R and c_U that appear in Lemmas 4.4 and 4.10, respectively. Interestingly, as shown in the plots below, the change is quite small never exceeding 8×10^{-3}.

For Rawlsian, we plot difference of program (10) evaluated at $\mathbf{x}^{\text{integ}}$ and evaluated at $\mathbf{x}^{\text{frac}}$ on **Census1990** data.

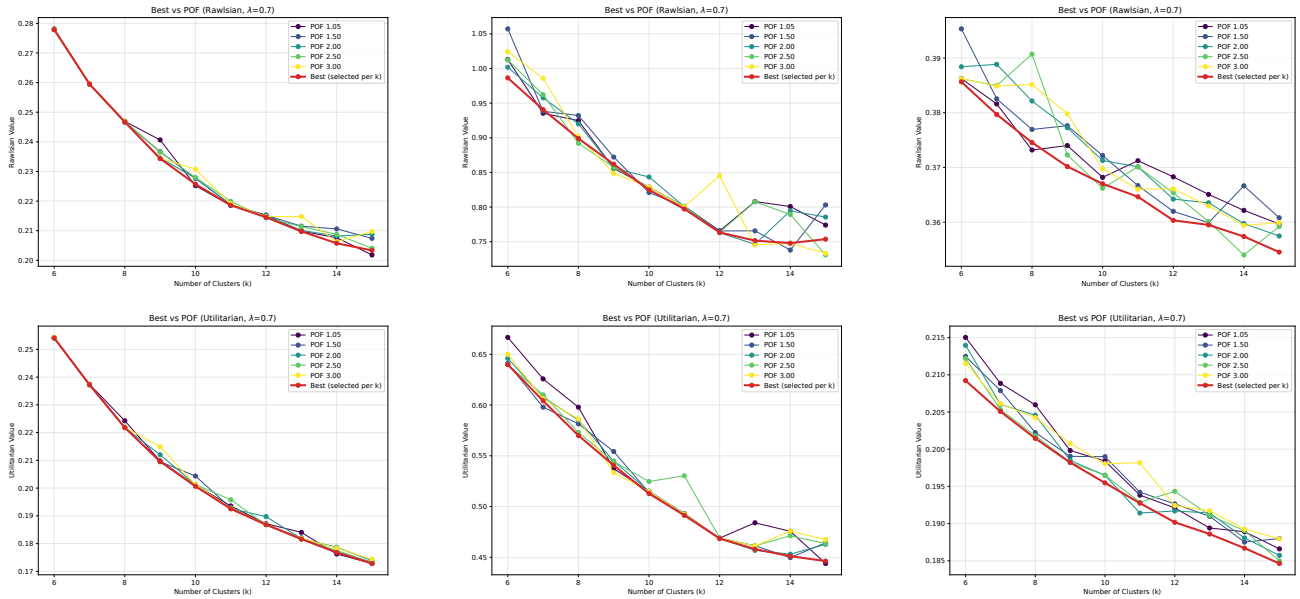


Figure 22: Rawlsiand and Utilitarian values for FCBC algorithms with different POF value on **Adult**, **CreditCard** and **Census1990**.

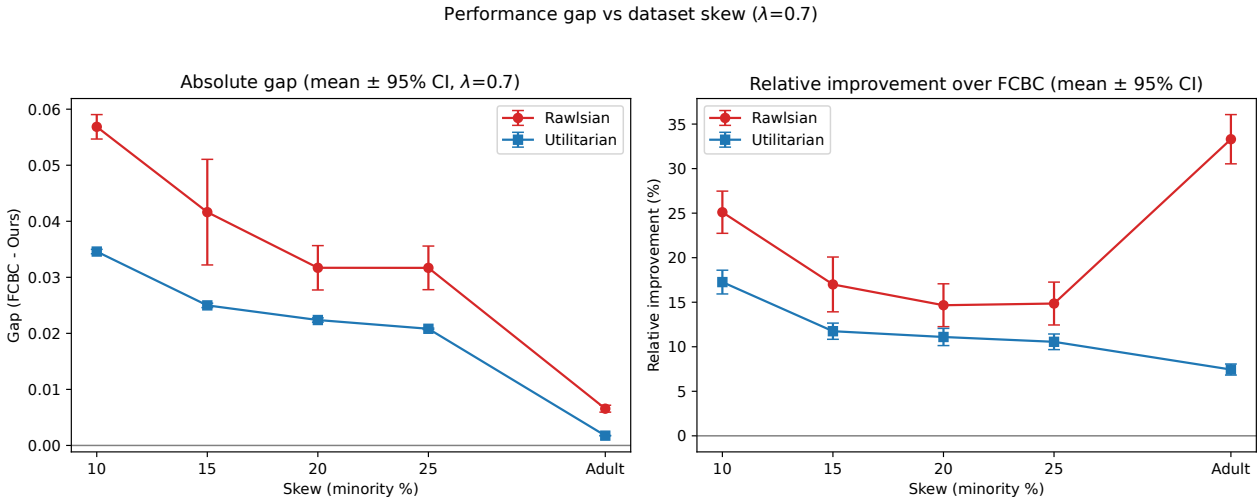


Figure 23: Performance gap between our proposed algorithms RawlsianAlg and UtilitarianAlg with FCBC algorithms vs proportion of minority group.

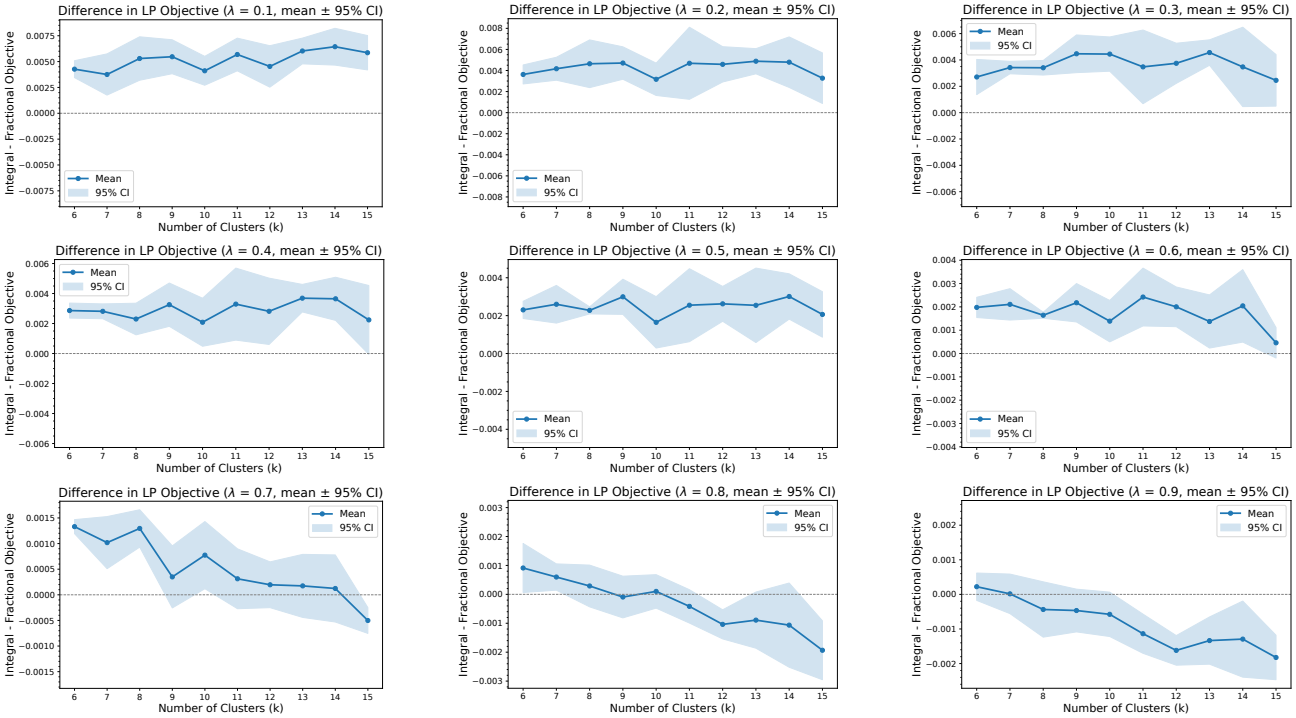


Figure 24: Integral Solution Rawlsian Value - Fractional Solution Rawlsian Value

Similarly for Utilitarian, we plot the difference of program (13) evaluated at $\mathbf{x}^{\text{integ}}$ and evaluated at $\mathbf{x}^{\text{frac}}$ on **Census1990** data.

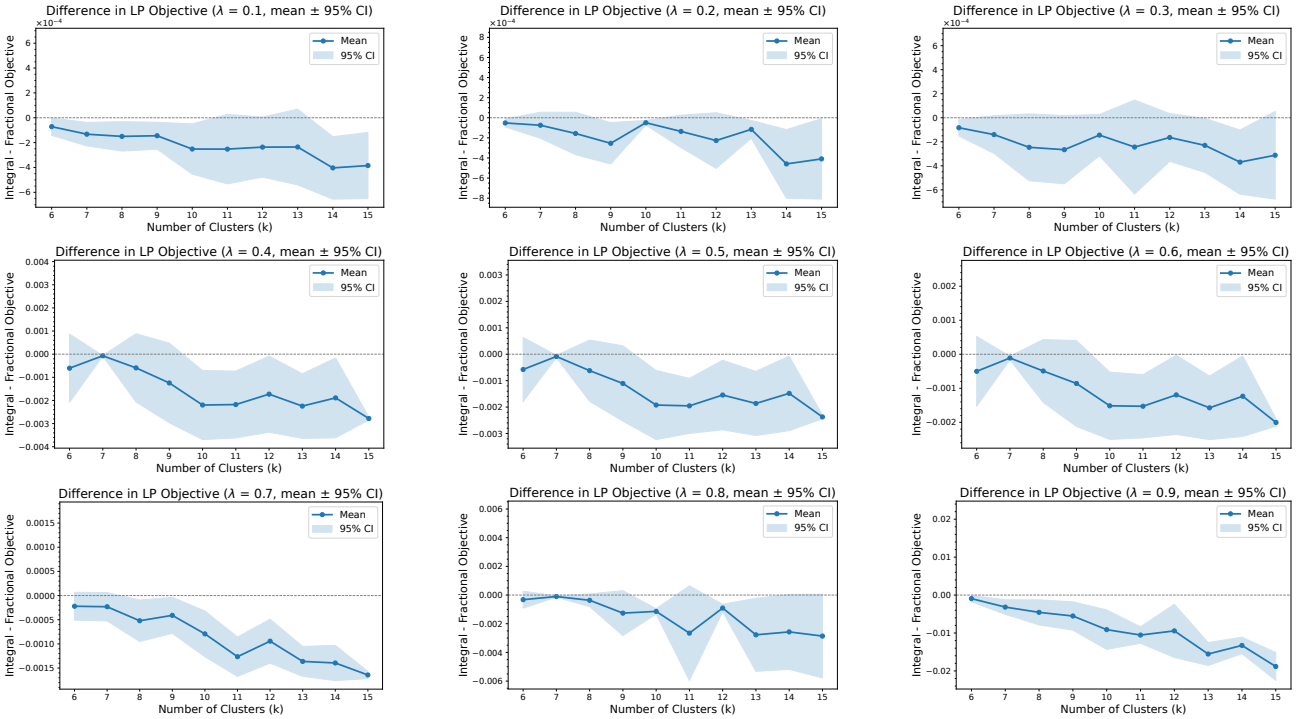


Figure 25: Integral Solution Utilitarian Value - Fractional Solution Utilitarian Value

H.10 Run Time of RawlsianAlg and UtilitarianAlg

In this section, we report the actual execution time of RawlsianAlg and UtilitarianAlg. We measure end-to-end runtime for each algorithm on sub-sampled versions of the **Census1990** dataset, with sizes ranging from 10,000 to 20,000 in increments of 1,000. Each run solves the respective objective using $k = 4$ clusters. Reported times are averaged over 5 independent runs, with standard deviation shown. All experiments were conducted on a MacBook Pro (2.3GHz 2-core Intel i5, 8GB RAM) using Python 3.10.

Figure 26 and Table 3 display the recorded runtimes. Across all sample sizes, execution times range roughly from 15 to 55 seconds, with increasing sample size generally leading to longer runtimes. The shaded regions in the plot indicate the variability across runs.

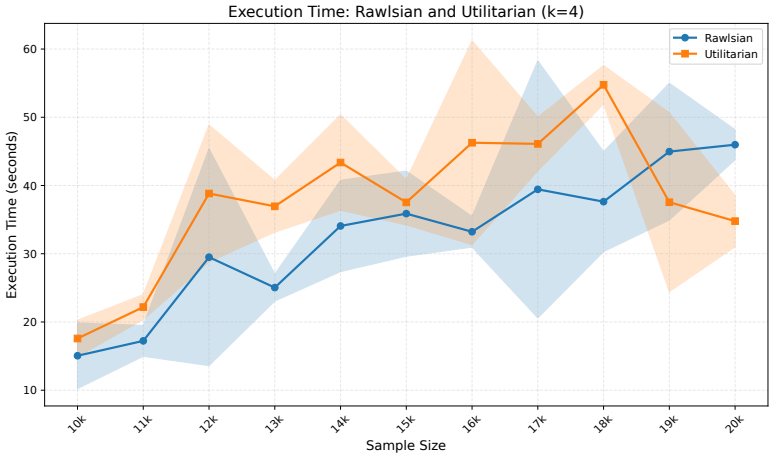


Figure 26: Runtime of RawlsianAlg and UtilitarianAlg on **Census1990** with $k = 4$ for different dataset sizes.

Size	Rawlsian Time (s)	Utilitarian Time (s)
10k	15.05 ± 4.81	17.58 ± 2.67
11k	17.23 ± 2.26	22.18 ± 1.83
12k	29.50 ± 15.90	38.83 ± 10.04
13k	25.04 ± 1.97	36.95 ± 3.78
14k	34.07 ± 6.69	43.37 ± 6.99
15k	35.88 ± 6.24	37.53 ± 3.31
16k	33.22 ± 2.25	46.27 ± 14.94
17k	39.44 ± 18.85	46.10 ± 3.93
18k	37.63 ± 7.32	54.76 ± 2.82
19k	44.97 ± 10.02	37.55 ± 13.11
20k	45.98 ± 2.20	34.78 ± 3.75

Table 3: Runtime of RawlsianAlg and UtilitarianAlg on **Census1990** with $k = 4$ for different dataset sizes.

I Limitations and Impacts

While our proposed welfare formulations provide flexible frameworks for balancing disutilities from distance and proportional representation, we emphasize that fairness is inherently a context-dependent and value-laden concept. The deployment of the proposed formulation and algorithms in practice should involve careful deliberation with relevant stakeholders to determine appropriate fairness goals and trade-offs. In particular, the choice of parameter λ, which governs the trade-off between distance and proportional representation must be guided by domain-specific considerations and the priorities of affected communities. Further, should the Rawlsian objective, which is mainly an egalitarian objective, be optimized or should the Utilitarian be optimized instead? Moreover, our formulation currently quantifies disutility primarily in terms of clustering distance and proportional group representation. In real-world settings, disutility may depend on additional factors such as downstream impacts of clustering outcomes as hinted in Dickerson et al. (2025). Additionally, even when it is clear and agreed upon that proportional group representation is important, the exact values of the proportion relaxation parameters α_h and β_h should be carefully done. It is possible that for a given application only group under-representation would matter and as a result one may set $\alpha_h \geq 1 - r_h$ for all colors $h \in \mathcal{H}$.

We believe that our proposed framework and algorithms can lead to better welfare and societal impacts if such considerations are carefully taken into account. On the other hand, a superficial approach that does not incorporate the full considerations and does carefully tune parameters such as λ, α_h, and β_h, may lead to clustering outcomes that are less preferable.