FEATURE-BASED INDIVIDUAL FAIRNESS IN k-CLUSTERING

Debajyoti Kar
IIT Kharagpur, India
debajyoti.apeejay@gmail.com

Sourav Medya
Northwestern University, USA
sourav.medya@gmail.com

Debmalya Mandal
Columbia University, USA
dm3557@columbia.edu

Arlei Silva
Rice University, USA
arlei@rice.edu

Palash Dey
IIT Kharagpur, India
palash.dey@cse.iitkgp.ac.in

Swagato Sanyal
IIT Kharagpur, India
sanyalswagato@gmail.com

ABSTRACT

Ensuring fairness in machine learning algorithms is a challenging and important task. We consider the problem of clustering a set of points while ensuring fairness constraints. While there have been several attempts to capture group fairness in the k-clustering problem, fairness at an individual level is not well-studied. We introduce a new notion of individual fairness in k-clustering based on features that are not necessarily used for clustering. We show that this problem is NP-hard and does not admit a constant factor approximation. We then design a randomized algorithm that guarantees approximation both in terms of minimizing the clustering distance objective as well as individual fairness under natural restrictions on the distance metric and fairness constraints. Finally, our experimental results validate that our algorithm produces lower clustering costs compared to existing algorithms while being competitive in individual fairness.

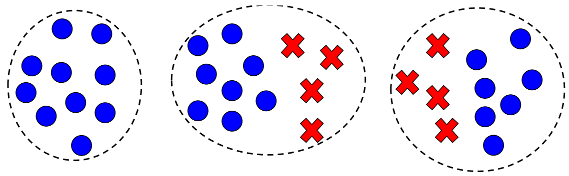
1 Introduction

Machine learning systems are increasingly being used in various societal decision making, including predicting recidivism [Ang+16; Cho17], deciding interest rates [Fus+20], and even allocating healthcare resources [Obe+19]. However, beginning with the report on bias in recidivism risk prediction [Ang+16], it has been known that such systems are often biased against certain groups of people. In recent years, various methods and definitions have been proposed for ensuring fairness in supervised learning settings, with efforts ranging from debiasing datasets [Fel+15] to explicitly encoding the fairness constraints during the training of a classifier [Aga+18].

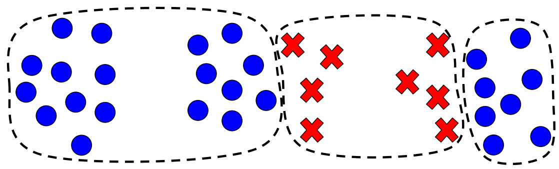
In this paper, we consider fairness in unsupervised learning, and particularly in clustering. There are two important reasons for designing a clustering that is fair with respect to different subgroups. First, clustering is often a preprocessing step to generate a new representation of the data to be used in downstream tasks. Since we want the downstream decisions to be fair, the clustering step needs to be unbiased [Abb+19]. Second, clustering is also used in various resource allocation problems, e.g. in the *facility location* problem [JKL19]. Since it is desirable that no group is disproportionately affected by such decisions, there have been an increasing interest in designing clustering algorithms that are fair with respect to different subgroups [Chi+17; Bac+19; Ber+19a; Ahm+19]. Such group-fair clustering algorithms ensure that each protected group has approximately equal presence in each cluster.

Compared to group fairness, individually fair clustering has received significantly less attention. Individually fair clustering is motivated by the *facility location problem* where the goal is to open k facilities while minimizing the total transportation cost between individuals and their nearest facility. If we choose k facilities (or centers) uniformly at random, then each point x could expect one of its nearest n/k neighbors to be one of such facilities. This led Jung et al.; Mahabadi et al. [JKL19; MV20] to consider the following notion of individual fairness in clustering. For a point x, let $r(x)$ be the radius such that the ball of radius $r(x)$ centered at x has at least n/k points. An individually fair clustering guarantees that for every x, a cluster center is chosen from the $r(x)$-neighborhood of x.

Although individually fair clustering [JKL19] provides guarantees for each point, there are two main drawbacks of such a notion. First, the existing definition does not exactly reflect the original premise of individual fairness suggested by Dwork et al. [Dwo+12], which requires that similar individuals should receive similar decisions. In the context of clustering, this means that two points x and y that are very similar (i.e. similar features) should be clustered similarly.



(a) Vanilla clustering



(b) Individually fair clustering

Figure 1: Vanilla versus individually fair clustering. Vanilla clustering violates individual fairness for the red points as each such point has less than m_v (say 6) points in its cluster.

However, the definition proposed by Jung et al. [JKL19] does not provide such guarantee, as points similar to a given point x could be different from the points within a radius of $r(x)$ from x.

The second drawback of the existing definition [JKL19] is that providing individual fairness for every point x does not necessarily result in a group-fair clustering. Consider an instance where a protected group g has size n/k and all points belonging to that group lie in a small neighborhood. In that case, it is easy to satisfy individual fairness for all points in group g by picking a single cluster center from that group. However, for such a solution, the members of group g are not represented in the other clusters. The apparent disconnect between individual fairness and group fairness in clustering arises because the former is motivated by facility location problem, while the latter tries to provide equal representation of different groups.

Proposed Definition of Individual Fairness. In order to address the two drawbacks highlighted above, we propose a new notion of individual fairness in clustering. First, motivated by the original definition of individual fairness in supervised learning [Dwo+12], we introduce a feature-based notion of individual fairness. We say that two individuals are similar if at least γ-fraction of their features match. Now, for each individual v, let $C(v)$ denote the cluster v is assigned to. Then our feature-based individually fair clustering requires that $C(v)$ also contains at least m_v individuals that are similar to v. This guarantees that a point v is not isolated in its own cluster and has a desired representation (or participation) from points similar to it.

Our notion of individual fairness guarantees that similar individuals (in terms of features) often share similar clusters. In that sense, this notion of individual fairness is closer to the original definition introduced by Dwork et al. [Dwo+12] and avoids various problems of the definition by Jung et al.; Mahabadi et al. [JKL19; MV20]. First, if one converts the cluster centers into representations to be used in downstream tasks, then similar individuals get similar representations and hence similar decisions. Furthermore, if the cluster centers are converted to facility locations then each individual v is guaranteed to have at least m_v similar individuals using the same center, and therefore, our definition promotes a desired level of representation (or participation), often found in group-fair clustering. Figure 1 illustrates an example of individually fair clustering with $m_v = 6$.

Contributions. Our main contributions are as follows:

- ▷ We propose a new definition of individual fairness in clustering based on how individuals are similar in terms of their features. Our definition guarantees that each individual has a desired level of representation of similar individuals in his/her own cluster.
- ▷ We show that minimizing the clustering cost subject to the new notion of individual fairness is NP-hard, and also not approximable within a factor δ for any $\delta > 0$.
- ▷ We then design a randomized algorithm that provides an additive approximation cost while guaranteeing fairness within a multiplicative factor with high probability.
- ▷ Our experiments on several standard datasets show that our approach produces up to ten times less cost in clustering than an existing method with a different notion of individual fairness while ensuring individual fairness for more than 85% points.

Related Work. Chierichetti et al. [Chi+17] first introduced the problem of fair clustering with disparate impact constraints and their goal was to ensure that all the protected groups have approximately equal representation in every cluster. Several follow-up works [Ber+19b; RS18] studied different generalizations of the fair clustering problem. Furthermore, several papers [Bac+19; SSS18; HJV19] have proposed procedures to scale fair clustering to a large number of points. Although we consider individual fairness, our work is related to Bera et al. [Ber+19a], which shows that a ρ-approximation to the vanilla clustering problem can be converted to a $(\rho + 2)$-approximate solution to fair clustering with bounded (and often negligible) violation of fairness constraints.

Our paper is mainly concerned with individual fairness, which was first defined by Dwork et al. [Dwo+12] in the context classification, and requires that similar individuals should be treated similarly. For clustering problems, such notion of individual fairness was first defined by Jung et al. [JKL19]. They studied individual fairness in terms of the guarantee a randomly chosen set of k points must satisfy. Informally, an individually fair clustering guarantees that for each point x, a cluster center is chosen from a certain neighborhood of x. Mahabadi et al. [MV20] design a bicriteria approximation algorithm for solving individually fair k-means and k-median problems. Their algorithms guarantee that not only the fairness constraints are approximately satisfied, but also the objective is approximately maximized. Recently, Vakilian et al. [VY21] designed improved bicriteria algorithms for general ℓ_p-norm costs.

The definition of individual fairness in Jung et al. [JKL19] was mainly motivated by fairness in the facility location problem. Recently, Kleindessner et al. [KAM20] considered a different notion of individual fairness in clustering, where the goal is to ensure that each point, is closer to the points in its own cluster than the points in any other cluster. Our proposed definition can be seen as a way to capture these two notions, as we consider feature-based similarity, as well as guaranteed representation for each point.

Here, we focus on the ℓ_p-norm cost for clustering, which is just the sum of ℓ_p-distances of each point from its corresponding cluster center. However, several papers did consider other objectives in the context of group-fair clustering [Ahm+19; GSV21; KAM19]. Finally, our focus is on fair clustering algorithms, and there is an extensive literature on fair algorithms unsupervised [Sam+18; Kle+19] and supervised learning more broadly [Aga+18; Don+18; Cel+19] . The coverage of these algorithms is out of scope of the paper, and we refer the interested reader to the following excellent surveys: Chouldechova et al.; Selbst et al. [CR18; Sel+19] and Mehrabi et al. [Meh+21].

2 Preliminaries

We first introduce some necessary notations. Let V be a set of n points $V = \{1, 2, \dots, n\}$. We denote $\{S_1, S_2, \dots, S_q\}$ as a set of q features, where S_i is the set of values for the i-th feature. We denote the tuple of q features of the point i by $X_i = (X_i^\ell)_{\ell\in[q]}$. We write $C = (C_i)_{i\in[\ell]}$ to denote a clustering (i.e. partition) of the set V and $(c_i)_{i\in[\ell]}$ to denote the corresponding cluster centers. Given a clustering C and a point v, let $\phi(v, C)$ be the cluster center assigned to the point v. When the clustering C is clear from the context, we use $\phi(v)$ to denote the cluster center assigned to the vertex v. We are also given a distance function $d : V \times V \to \mathbb{R}$ that measures the distance between any pair of points. Given a clustering C, we can measure its cost through the distance metric d. In particular, we will be interested in measuring the sum of the p-th powers of distances from each point to its cluster center for $p \in \mathbb{N} \cup \{0\}$:

$$\mathrm{Cost}(C) = \sum_{v\in V} d(v, \phi(C, v))^p. \tag{1}$$

We assume that the distance function d depends on some features of the points but don't assume any relationship between those features and the ones used for fairness.

We now introduce our definition of individual fairness in clustering. To the best of our knowledge, all the existing notions of individual fairness in clustering only depend on the distance-based neighborhood of each point. In contrast, our definition of individual fairness builds upon the features of individual points that are not necessarily used for clustering. In order to define the feature-based notion of fairness, we first define a similarity measure based on the features.

Definition 1 (γ**-similarity).** *For a parameter $\gamma \in [0, 1]$, we say two points $i, j \in V, i \neq j$ are γ-similar if they match in at least γ fraction of features, i.e. $|\{\ell \in [q] : X_i(\ell) = X_j(\ell)\}| \geqslant \gamma q$. We assume that a point is not γ-matched with itself.*

For any point v, we use $\Gamma(v)$ to denote the set of points in V that are γ-similar to v. With this notion of similarity, we introduce our definition of individually fair clustering.

Definition 2 (Individual Fairness in Clustering)**.** *Given a set V of n points along with a q-length feature vectors $X_v = (x_v^1, \dots, x_v^q)$ for every point $v \in V$, a similarity parameter $\gamma \in [0, 1]$, an integer tuple $(m_v)_{v\in V}$, and an integer*

k, we say that a clustering $(C_i)_{i\in[\ell]}$ ($\ell \leqslant k$) is $(m_v)_{v\in V}$-individually fair if it satisfies the following constraint for every point $v \in V$:

$$|\{u : u \in \Gamma(v) \text{ and } \phi(u) = \phi(v)\}| \geqslant m_v \tag{2}$$

The fairness constraint (2) says that at least m_v points that are γ-similar to point v must belong to the cluster of v. Our main goal is to compute a clustering $(C_i)_{i\in[\ell]}$ of V into ℓ ($\leqslant k$) clusters and corresponding cluster centers (or facilities[1]) $(c_i)_{i\in[\ell]}$ that is individually fair for every point and minimizes the clustering cost e.g. sum of the p-th powers of distances from cluster centers for some $p \in \mathbb{N} \cup \{0\}$. Formally, our INDIVIDUALLY FAIR CLUSTERING problem is defined as follows.

Definition 3 (INDIVIDUALLY FAIR CLUSTERING (**IFC**))**.** *The input is a set V of n points along with a q-length feature vector $X_v = (x_v^1, \ldots, x_v^q)$ for every point $v \in V$, a similarity parameter $\gamma \in [0, 1]$, an integer tuple $(m_v)_{v\in V}$, a set F of potential facilities. The objective is to open a subset $S \subseteq F$ of at most k facilities, and find an assignment $\phi : V \longrightarrow S$ to minimize Cost(C, ϕ) satisfying the fairness constraints* (2).

The classical clustering problem, which we call VANILLA CLUSTERING, is the same as the IFC problem except for the fairness requirements in (2).

3 Results

In this section, we present our main technical results in two directions. First, we provide several hardness results to show that the general INDIVIDUALLY FAIR CLUSTERING (IFC) problem is hard even if one considers approximation. Then we contrast the hardness results by developing randomized approximation algorithms for various special cases of the IFC problem.

3.1 Hardness Results

In order to prove hardness results, we consider the decision version of the IFC problem, where the goal is to find a clustering whose cost is below a certain threshold. Note that, it is always possible to find a (trivial) individually fair clustering by one cluster containing all the points. However, the cost of such a fair clustering could be high, and we ask whether it is possible to beat the cost of such trivially fair clustering. As there can be multiple trivial fair clustering (depending on the cluster center chosen), we naturally pick the one minimizing the cost as the benchmark.

Definition 4 (TRIVIALLY FAIR CLUSTERING)**.** *Given a set V of n points along with q-length feature vector $X_v = (x_v^1, \ldots, x_v^q)$ for every point $v \in V$, the trivially fair clustering puts all points in one cluster and picks the point as cluster center which minimizes the cost:*

$$\min_{k\in F} \sum_{v\in V} d(v, k)^p.$$

We show that it is NP-complete to compute if there exists any clustering better than TRIVIALLY FAIR CLUSTERING. For that, we exhibit a reduction from SATISFACTORY-PARTITION which is known to be NP-complete [BTV06].

Definition 5 (SATISFACTORY-PARTITION)**.** *Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and an integer λ_v for every vertex $v \in \mathcal{V}$, compute if there exists a partition $(\mathcal{V}_1, \mathcal{V}_2)$ of $\mathcal{V}$ such that*

i) $\mathcal{V}_1, \mathcal{V}_2 \neq \emptyset$
ii) For every $i \in [2]$ and every $v \in \mathcal{V}_i$, the number of neighbors of v in $\mathcal{V}_i$ is at least λ_v.

We denote an arbitrary instance of SATISFACTORY-PARTITION *by $(\mathcal{G}, (\lambda_v)_{v\in\mathcal{G}})$.*

Theorem 1. *It is* NP*-complete to decide whether an instance of* INDIVIDUALLY FAIR CLUSTERING *admits a clustering of cost less than the* TRIVIALLY FAIR CLUSTERING *even when there are only 2 facilities.*

Proof. The problem clearly belongs to NP. We now exhibit a reduction from SATISFACTORY-PARTITION. Let $(\mathcal{G} = (\mathcal{V} = \{v_1, v_2, \ldots, v_n\}, \mathcal{E}), (\lambda_v)_{v\in\mathcal{V}})$ be an arbitrary instance of SATISFACTORY-PARTITION. In our INDIVIDUALLY FAIR CLUSTERING instance, we have two facilities l and r and the set of points $U = \{u_1, u_2, \ldots, u_n\}$. We define the distances as

$$d(u_i, l) = \begin{cases} (\lceil \frac{n}{2} \rceil + \beta)^{1/p} & 1 \leqslant i \leqslant \lceil \frac{n}{2} \rceil + 1 \\ \beta^{1/p} & \lceil \frac{n}{2} \rceil + 2 \leqslant i \leqslant n \end{cases}$$

[1]We use cluster center and facility interchangeably.

$$d(u_i, r) = \begin{cases} \beta^{1/p} & 1 \leqslant i \leqslant \lfloor \frac{n}{2} \rfloor \\ (\lceil \frac{n}{2} \rceil + \beta + 1)^{1/p} & \lfloor \frac{n}{2} \rfloor + 1 \leqslant i \leqslant n \end{cases}$$

for any $\beta \geqslant \frac{\lceil \frac{n}{2} \rceil + 1}{2}$. We can easily verify that the following properties are satisfied:

i) The distances satisfy the triangle inequality.
ii) $\sum_{v \in V} d(v, l)^p = \sum_{v \in V} d(v, r)^p = A$ (say).
iii) For any $X \subset V, X \neq \emptyset, X \neq V$, we have $\sum_{v \in X} d(v, l)^p \neq \sum_{v \in X} d(v, r)^p$. To see this, let $s = |X \cap \{u_1, u_2, \ldots, u_{\lceil \frac{n}{2} \rceil + 1}\}|$ and $t = |X \cap \{u_{\lfloor \frac{n}{2} \rfloor + 1}, u_{\lfloor \frac{n}{2} \rfloor + 2}, \ldots, u_n\}|$ so that $s \leqslant \lceil \frac{n}{2} \rceil + 1$ and $t \leqslant \lceil \frac{n}{2} \rceil$. Then $\sum_{v \in X} d(v, l)^p = s \lceil \frac{n}{2} \rceil + \beta |X|$ and $\sum_{v \in X} d(v, r)^p = t(\lceil \frac{n}{2} \rceil + 1) + \beta |X|$. Thus $\sum_{v \in X} d(v, l)^p = \sum_{v \in X} d(v, r)^p$ would imply $s \lceil \frac{n}{2} \rceil = t(\lceil \frac{n}{2} \rceil + 1)$ and therefore either $s = t = 0$ or $s = \lceil \frac{n}{2} \rceil + 1$ and $t = \lceil \frac{n}{2} \rceil$. In the former case $X = \emptyset$ while in the latter case $X = V$, a contradiction.

We now describe the feature vector. For every edge $e \in \mathcal{E}$, we have a feature θ_e. A point $u_i, i \in [n]$ has value 1 for θ_e if the edge e is incident on the vertex v_i in $\mathcal{G}$; otherwise has value 0 for θ_e. Finally, we set the similarity parameter $\gamma = \frac{1}{m}$. We observe that two points $u_i, u_j, i, j \in [n]$ are γ-similar if and only if there is an edge between v_i and v_j in $\mathcal{G}$. Finally, we define $m_{v_i} = \lambda_{v_i}$ for every $i \in [n]$.

We claim that the SATISFACTORY-PARTITION instance is a yes instance if and only if there exists a fair clustering of U with cost $< A$. The "if" part follows directly, since any fair clustering of U with cost $< A$ must be non-trivial and fairness ensures that the corresponding partition of $\mathcal{G}$ is satisfactory.

For the "only if" part, let $(X, \bar{X})$ be a non-trivial satisfactory partition of $\mathcal{G}$. Let ϕ_1 denote the assignment that assigns all corresponding vertices in X to l and all corresponding vertices in $\bar{X}$ to r, and ϕ_2 denote the assignment that assigns all vertices in X to r and all vertices in $\bar{X}$ to l. Thus,

$$\text{Cost}(\phi_1) = \sum_{v \in X} d(v, l)^p + \sum_{v \in \bar{X}} d(v, r)^p$$

$$\text{Cost}(\phi_2) = \sum_{v \in X} d(v, r)^p + \sum_{v \in \bar{X}} d(v, l)^p$$

Thus, $\text{Cost}(\phi_1) + \text{Cost}(\phi_2) = 2A$. Now, it cannot be the case that $\text{Cost}(\phi_1) = \text{Cost}(\phi_2) = A$, which would then imply that one of the assignments ϕ_1 or ϕ_2 must have cost $< A$ and we are done. Suppose to the contrary that $\text{Cost}(\phi_1) = A$. Thus, $\sum_{v \in X} d(v, l)^p + \sum_{v \in \bar{X}} d(v, r)^p = \sum_{v \in X} d(v, r)^p + \sum_{v \in \bar{X}} d(v, r)^p$ and therefore $\sum_{v \in X} d(v, l)^p = \sum_{v \in X} d(v, r)^p$, a contradiction. □

Given the NP-completeness result, we explore the possibility of approximation for the INDIVIDUALLY FAIR CLUSTERING (IFC) problem. However, the next theorem shows that IFC is inapproximable within factor δ for any $\delta > 0$.

Theorem 2. *Distinguishing between instances of the IFC problem having zero and non-zero optimal costs is NP-complete even when there are 2 facilities. Hence, for any computable function δ, there does not exist a δ-approximation algorithm for IFC unless P=NP.*

Proof. Let $\mathcal{A}$ be a deterministic polynomial time algorithm that distinguishes between instances of IFC with 0 and non-zero optimal costs. We use $\mathcal{A}$ to build an algorithm for SATISFACTORY-PARTITION. Let $(\mathcal{G} = (\mathcal{V} = \{v_1, v_2, \ldots, v_n\}, \mathcal{E}), (\lambda_v)_{v \in \mathcal{V}})$ be an instance of SATISFACTORY-PARTITION. We create $n-1$ instances $I_1, I_2, \ldots, I_{n-1}$ of IFC as follows: the set of points is $U = \{u_1, \ldots, u_n\}$ and $m_{v_i} = \lambda_{v_i}$ for every $i \in [n]$ for all the instances; for instance $I_i, i \in [n]$, we introduce 2 facilities l and r and define distances as follows:

$$d(u_j, l) = \begin{cases} 1 & j = 1 \\ 0 & j \in \{2, 3, \ldots, n\} \end{cases}$$

$$d(u_j, r) = \begin{cases} 1 & j = i + 1 \\ 0 & j \in \{1, 2, \ldots, n\} \setminus \{i + 1\} \end{cases}$$

We now run $\mathcal{A}$ on each of these instances. If $\mathcal{A}$ returns 0 on any instance, return yes for the SATISFACTORY-PARTITION instance; otherwise we return no for the SATISFACTORY-PARTITION instance.

Clearly, the above algorithm runs in polynomial time. We now prove its correctness. If $\mathcal{G}$ does not have a non-trivial satisfactory partition, then $\mathcal{A}$ returns 1 in every run. Let $(X, \bar{X})$ be a satisfactory partition of $\mathcal{G}$. Without loss of generality, we assume that we have $v_1 \in X$. Let $v_j \in \bar{X}$, for some $j \in \{2, 3, \dots, n\}$. Then clearly $\text{OPT}(I_{j-1}) = 0$ (assigning all the corresponding vertices in X to the cluster center r and all other vertices to the cluster center l). Thus, the algorithm is correct. □

Note that the distances in the above reduction do not satisfy triangle inequality. If we insist that distances must satisfy triangle inequality, then we have the following (weaker than Theorem 2) inapproximability result.

Theorem 3. *There does not exist any FPTAS for the IFC problem when the distances in the input satisfy triangle inequality unless* $\mathsf{P} = \mathsf{NP}$.

Proof. Let $\mathcal{A}$ be an FPTAS for IFC. Similar to the proof of Theorem 2, we create $n-1$ instances of SATISFACTORY-PARTITION where instance I_i is as follows: the set of points is $U = \{u_1, \dots, u_n\}$ and $m_{v_i} = \lambda_{v_i}$ for every $i \in [n]$ for all the instances; for instance $I_i, i \in [n]$, we introduce 2 facilities l and r and define distances as follows:

$$d(u_j, l) = \begin{cases} (1+\beta)^{1/p} & j = 1 \\ \beta^{1/p} & j \in \{2, 3, \dots, n\} \end{cases}$$

$$d(u_j, r) = \begin{cases} (1+\beta)^{1/p} & j = i+1 \\ \beta^{1/p} & j \in \{1, 2, \dots, n\} \setminus \{i+1\} \end{cases}$$

where β is any constant $\geqslant 1/2$. The algorithm runs $\mathcal{B}$ on each of the above instances with approximation parameter $\varepsilon = \frac{1}{2n\beta}$. If $\mathcal{B}$ returns a solution of cost less than $1+n\beta$ on any instance, return yes for the SATISFACTORY-PARTITION instance; otherwise we return no for the SATISFACTORY-PARTITION instance.

Clearly, the cost of the trivial partition is $1 + n\beta$. Thus, if $\mathcal{G}$ does not have a non-trivial satisfactory partition, then $\mathcal{B}$ must always return the trivial assignment of cost $1 + n\beta$ for all instances. If $\mathcal{G}$ has a satisfactory partition $(X, \bar{X})$, then as in Theorem 2, there exists an instance with optimal cost $n\beta$. Thus, the solution returned by $\mathcal{B}$ will have cost at most $n\beta(1 + \frac{1}{2n\beta}) < 1 + n\beta$. Hence, the algorithm is correct. □

3.2 Algorithmic results

Given the strong inapproximability results in the previous section, we aim to develop approximation algorithms for INDIVIDUALLY FAIR CLUSTERING under suitable conditions. First, we develop an approximation algorithm for INDIVIDUALLY FAIR ASSIGNMENT (Theorem 4). Next, we show how to obtain an algorithm for INDIVIDUALLY FAIR CLUSTERING of similar guarantee (Theorem 5). Bera et al. Bera et al. [Ber+19a] designed an approximation algorithm for group fair clustering from an algorithm for group fair assignment. We follow a similar approach for individual fairness.

One of the main ingredients of our technical results is the INDIVIDUALLY FAIR ASSIGNMENT problem, which, given a set of k potential cluster centers, determines an assignment of the points i.e. which point should be assigned to which cluster center. Formally, it is defined as follows:

Definition 6 (INDIVIDUALLY FAIR ASSIGNMENT (**IFA**))**.** *Given a set V of n points along with a q-length feature vector $X_v = (x_v^1, \dots, x_v^q)$ for every point $v \in V$, a similarity parameter $\gamma \in [0, 1]$, an integer tuple $(m_v)_{v \in V}$, and a set $F = \{f_1, \dots, f_k\}$ of k facilities, an $(m_v)_{v \in V}$-fair assignment finds the optimal cost-minimizing assignment satisfying the fairness constraints (2).*

Our Algorithm, LP-FAIR: Algorithm 1 describes our randomized approximation algorithm for IFA. The linear program (LP) described in Inequality (3) is a relaxation of the IFA problem. It has a variable x_{v,f_k} for each vertex v and facility f_k. In a (integral) "solution" the variable x_{v,f_k} takes value 1 if and only if the point v is assigned to the facility f_k. After solving the LP, Algorithm 1 determines the assignment ϕ by we assigning point v to f_k with probability x^*_{v,f_k}. Finally, the above procedure is repeated $\log_{1+\delta} n$ times and the assignment with the lowest cost is returned to boost the success probability.

The next theorem presents probabilistic approximation guarantees provided by Algorithm 1.

Theorem 4. *For any $\varepsilon, \delta > 0$, there exists a randomized algorithm for IFA running in time polynomial in n and $\frac{1}{\delta}$, that outputs a solution of cost at most $(1+\delta)$OPT where each vertex v has at least $\frac{m_v}{k}(1-\varepsilon)$ neighbours assigned to the same facility as itself with high probability, if $m_v = \Omega(\frac{k \log n}{\varepsilon^2})$, $\forall v \in V$.*

Algorithm 1: LP-FAIR, Algorithm for **IFA**

Input: $(V, (X_v)_{v \in V}, \gamma, (m_v)_{v \in V}, k)$, and δ.
1: **for** $t = 1, 2, \ldots, T = \log_{1+\delta} n$ **do**
2: Solve the following LP to get solution $x_t^\star$.

$$\begin{aligned} \min_x & \sum_{v \in V} \sum_{f_k \in F} d(v, f_k)^p \cdot x_{v,f_k} \\ \text{s.t.} & \sum_{u \in N(v)} x_{u,f_k} \geqslant m_v \cdot x_{v,f_k} \ \forall v \in V, f_k \in F \\ & \sum_{f_k \in F} x_{v,f_k} = 1 \ \forall v \in V \\ & x_{v,f_k} \geqslant 0 \ \forall v \in V, f_k \in F \end{aligned} \tag{3}$$

3: **for** each $v \in V$ **do**
4: Set $\phi_t(v) = f_k$ with probability $x^\star_{t,v,f_k}$.
5: **end for**
6: **end for**
7: **return** Assignment $\phi^\star$ with the minimum cost.

Proof. Let $x^\star$ be a solution to the linear program 3. Algorithm 1 assigns point v to f_k with probability x^*_{v,f_k}. We now prove the quality of this solution.

Fix a point v. Let X be the random variable denoting the number of neighbours that are assigned to the same facility as v. For every $u \in N(v)$, let X_u be the indicator random variable indicating whether u and v are assigned to the same facility. Thus,

$$E[X_u] = \sum_{f_k \in F} x^*_{v,f_k} x^*_{u,f_k}$$

So we have,

$$E[X] = \sum_{u \in N(v)} E[X_u] = \sum_{f_k \in F} (x^*_{v,f_k} \sum_{u \in N(v)} x^*_{u,f_k}) \geqslant m_v \sum_{f_k \in F} {x^*_{v,f_k}}^2 \geqslant m_v/k$$

Now using Chernoff bound,

$$\Pr[v \text{ has at most } \frac{m_v}{k}(1-\varepsilon) \text{ neighbours}] \leqslant e^{-\frac{\varepsilon^2}{2}\frac{m_v}{k}} \leqslant \frac{1}{n^2}$$

And using union bound,

$$\Pr[\exists \text{ a vertex that has at most } \frac{m_v}{k}(1-\varepsilon) \text{ neighbours}] \leqslant \frac{1}{n}$$

Also, clearly the expected cost of the computed solution is at most OPT. Hence, using Markov's inequality,

$$\Pr[\text{cost of computed solution is } \geqslant (1+\delta)\text{OPT}] \leqslant \frac{1}{1+\delta}$$

As we repeat the above algorithm $T = \frac{\log n}{\log(1+\delta)}$ times and output the solution with minimum cost, we have

$$\Pr[\text{cost of the computed solution is } \geqslant (1+\delta)\text{OPT}] \leqslant \frac{1}{n}$$

Also, by union bound, the probability that there exists a vertex having at most $\frac{m_v}{k}(1-\varepsilon)$ neighbours in one of the T solutions is at most $\frac{T}{n}$. □

We next show a method for obtaining an approximation algorithm for INDIVIDUALLY FAIR CLUSTERING from an approximation algorithm for INDIVIDUALLY FAIR ASSIGNMENT in a black box fashion.

Algorithm 2: Algorithm for INDIVIDUALLY FAIR CLUSTERING

Input: $(V, (X_v)_{v\in V}, \gamma, (m_v)_{v\in V}, k)$
1: Solve clustering problem $(V, (X_v)_{v\in V}, \gamma, (m_v)_{v\in V}, k)$ using any vanilla algorithm ignoring fairness constraints. Let $\left((C_i)_{i\in[\ell]}, (c_i)_{i\in[\ell]}\right)$ be the output of the clustering algorithm.
2: $F \leftarrow \{c_1, \ldots, c_\ell\}$
3: Run algorithm for INDIVIDUALLY FAIR ASSIGNMENT on $(V, (X_v)_{v\in V}, \gamma, (m_v)_{v\in V}, F)$. Let ϕ be the output.
4: **return** $(\phi^{-1}(c_i))_{i\in[\ell]}$ (ignore $\phi^{-1}(c_j)$ if $\phi^{-1}(c_j)$ is the empty set for some $j \in [\ell)$

Theorem 5. *If the distances satisfy the triangle inequality, then existence of an ρ-approximation algorithm for* VANILLA CLUSTERING *and an α-approximation algorithm for* INDIVIDUALLY FAIR ASSIGNMENT *with λ-multiplicative violations for some $\lambda \geqslant 1$ imply the existence of an $\alpha(\rho+2)$-approximation algorithm for* INDIVIDUALLY FAIR CLUSTERING *with λ-multiplicative violation.*

Proof. Let S^* be the optimal set of facilities opened and ϕ^* be the optimal assignment in the input INDIVIDUALLY FAIR CLUSTERING instance. Let S be the set of facilities returned by the vanilla k-clustering problem and ϕ the assignment. For each $f^* \in S^*$, let us define $\text{nrst}(f^*) = \arg\min_{f\in S} d(f, f^*)$. Consider the assignment ϕ' over the set of facilities S that assigns each vertex v to $\text{nrst}(\phi^*(v))$. We claim that ϕ' is a fair assignment (please see Claim 1).

For any vertex v, let $\phi(v) = f$, $\phi'(v) = f'$ and $\phi^*(v) = f^*$. Thus $d(v, f') \leqslant d(v, f^*) + d(f^*, \text{nrst}(f^*)) \leqslant d(v, f^*) + d(f^*, f) \leqslant 2d(v, f^*) + d(v, f)$. Then, since l_p is a monotone norm, $l_p(S, \phi') \leqslant 2l_p(S^*, \phi^*) + l_p(S, \phi) \leqslant (\rho+2)OPT$. Now, since we have an α-approximation algorithm for INDIVIDUALLY FAIR ASSIGNMENT, the solution returned by the algorithm will have cost at most $\alpha \cdot l_p(S, \phi') \leqslant \alpha(\rho+2)\text{OPT}$. □

Claim 1. *ϕ' is an individually fair assignment.*

Proof. Consider any $v \in V$. Let T_v denote the set of vertices assigned to $\phi^*(v)$. Since ϕ^* is an individually fair assignment, the number of neighbours of v in T_v is at least m_v. Now all vertices in T_v are assigned to the facility $\text{nrst}(\phi^*(v))$. Thus the number of neighbours of v in the assignment ϕ' is also at least m_v. Hence ϕ' is an individually fair assignment. □

4 Experiments

In this section, we provide an experimental evaluation of our proposed LP-based algorithm along with two baselines using two different cost functions: **k-median** and **k-means**.

Algorithms: We evaluate the following three algorithms:

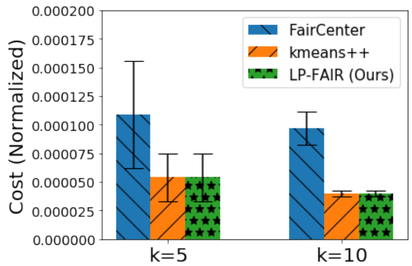
▷ **Our LP-based approach (LP-FAIR):** Our proposed LP-based procedure (LP-FAIR), which provides probabilistic approximation guarantees (Algorithm 1). Our implementation is based on Algorithm 2, as discussed in our Algorithmic Results section.
▷ **FairCenter [JKL19]**: This is Algorithm 2 in their paper, which ensures fairness based on the existence of a cluster center nearby. Notice that this algorithm has a different fairness criteria than ours.
▷ **k-means++**: We use sklearn implementation of k-means based on Elkan's algorithm [Elk03] and k-means++ [AV07] for initialization. Note that this approach does not guarantee individual fairness.

Performance measures: We evaluate the algorithms described above using the following metric:

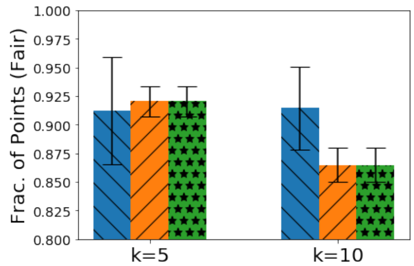
▷ **Normalized Cost:** Clustering cost (Equation 1) normalized by the cost of trivially fair clustering (Definition 4). The normalization removes the effect of dataset-dependent feature distributions and makes it easier to compare the results across datasets: Normalized Cost$(A) = \frac{Cost(A)}{Cost(\text{TRIVIALLY FAIR CLUSTERING})}$.
▷ **Fraction of Fair Points:** This denotes the fraction of points that satisfy individual fairness. Higher is better.

Datasets: We use three datasets from the UCI repository in our experiments.[2] These datasets have also been used by previous work [Ber+19a; MV20; Chi+17]. We consider the following attributes for distance and γ-similarity (fairness):

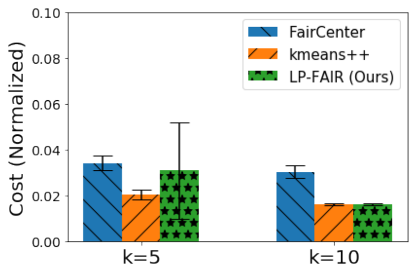
[2] https://archive.ics.uci.edu/ml/datasets



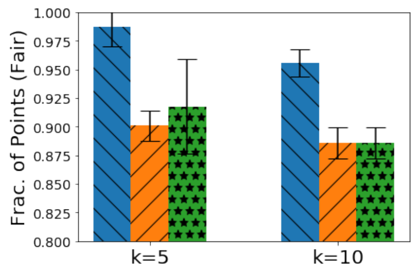
(a) Cost



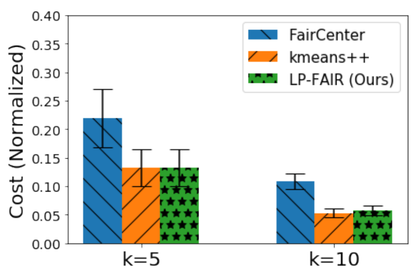
(b) Fair Points



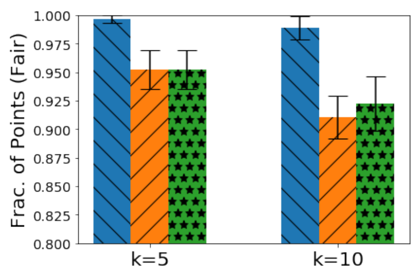
(c) Cost



(d) Fair Points



(e) Cost



(f) Fair Points

Figure 2: **k-median:** The results produced by all methods on (a-b) Bank, (c-d) Census, and (e-f) Diabetes datasets both in terms of Cost (lower is better) and fraction of points that are individually fair (higher is better). Our method, LP-FAIR produces up to 10 times lower cost than FairCenter.

▷ **Bank:** Data from customers of Portuguese bank. The distance features are “balance” and “duration”. The γ-similarity features are “job” and “age”.
▷ **Census:** US Census data from 1994. Distance features are “age”, “eductationnum” and “hoursperweek”. The γ-similarity features are “occupation” and “education”.
▷ **Diabetes:** Data from diabetes patients from 130 hospitals in the USA from 1999 to 2008. The distance features are “age” and “number-emergency”. The γ-similarity features are “gender”, “race”, and “admission-type-id”.

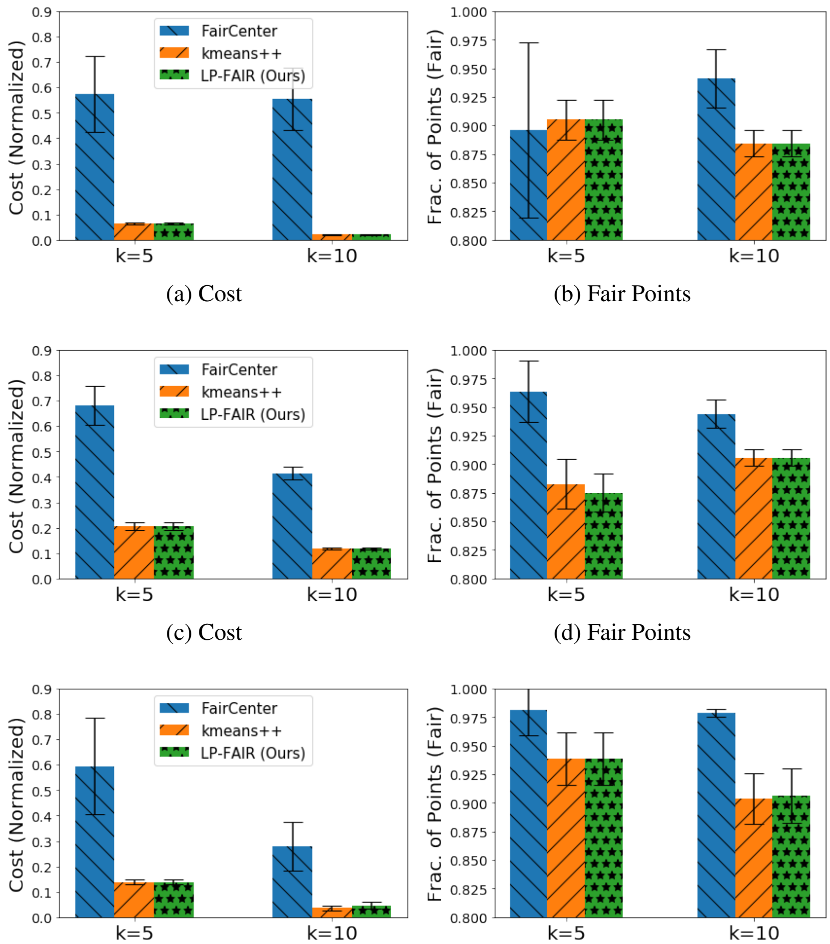
Other settings: We choose 500 points from each dataset randomly for all the experiments. The experiments are run for at least five times. Means and variances have shown based on these repetitions. Let Γ_v denote the number of γ-matched points of v in the entire dataset, the initial m_v for a node v is set as $\frac{\theta}{k} * |\Gamma_v|$. Unless specified otherwise, we set θ as 0.5 in all experiments.

4.1 Results for k-median

In this section, we present the results for the k-median cost function. Figure 2 shows the results produced by all three algorithms. Our method LP-FAIR has up to 10 times lower cost than the fair baseline FairCenter and consistently clusters more than 85% points in an individually fair manner. LP-FAIR also produces an equal or larger number of fair points than k-means++ while being comparable in cost. Note that the cost (or distance) is normalized against the cost in trivially fair clustering. Results on the Census dataset have small values of normalized cost for all three algorithms—indicating the high magnitude of the cost produced by the trivially fair clustering.

4.2 Results for k-means

This section presents the results for the k-means cost function. Figure 3 shows the results produced by the three algorithms. Consistent with the results for k-median, our method LP-FAIR has a significantly lower cost than the fair baseline FairCenter and consistently clusters more than 85% points fairly according to our individual fairness notion. Compared with k-means++, LP-FAIR clusters an equal or larger number of points fairly while achieving similar normalized cost.



(a) Cost (b) Fair Points

(c) Cost (d) Fair Points

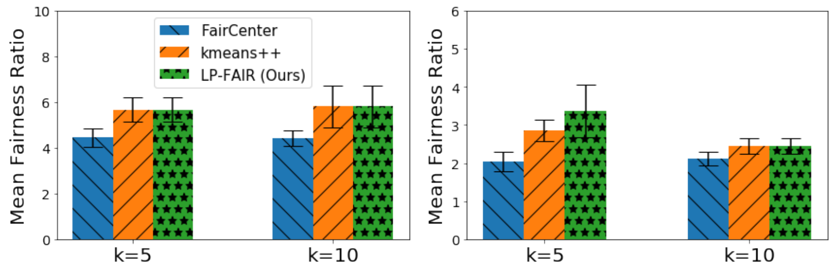
(e) Cost (f) Fair Points

Figure 3: **k-means:** The results produced by all methods on (a-b) Bank, (c-d) Census, and (e-f) Diabetes datasets both in terms of Cost (lower is better) and fraction of points that are individually fair (higher is better).

4.3 Another performance measure: fairness ratio

In the previous sections, we have applied the fraction of fair points as an evaluation metric for fairness. Here, we consider the extent to which individual fairness is satisfied and evaluate the algorithms based on the fairness ratio. More specifically, let α_v denote the ratio between the number of γ-similar points (Definition 1) of v in the same cluster and the value of m_v. The mean α is the average value of α_v, $\forall v \in V$. Thus, the higher the value of α, the better. Though FairCenter produces more fair points than LP-FAIR in the results from the previous sections, Figure 4 shows that, for both k-median and k-means objectives, LP-FAIR achieves a better fairness ratio.

Discussion A few key observations from the experiments are as follows: **(1)** Our method, LP-FAIR produces lower cost or distance than FairCenter in all settings. **(2)** LP-FAIR produces clusters that are individually fair to more than 85% points. Even though LP-FAIR is inferior than FairCenter in terms of number of fair points, our approach outperforms FairCenter in terms of fairness ratio. **(3)** Though LP-FAIR and k-means++ achieve similar performance, LP-FAIR provides probabilistic approximation guarantees.



(a) k-means (b) k-median

Figure 4: **Fairness Ratio:** The results produced by all methods on bank dataset for (a) k-median and (b) k-means.

5 Conclusion

In this paper, we have studied the k-clustering problem with individual fairness constraints. Our notion of individual fairness is defined in terms of a feature-based similarity among points, and guarantees that each point will have a pre-defined number of similar points in their cluster. We have provided an algorithm with probabilistic approximation for optimizing the cluster distance as well as ensuring fairness. Finally, the experimental results have shown that our proposed algorithm can produce up to 10 times lower clustering cost than previous work.

There are many interesting directions for future work. Analogous to existing definition of individually fair classification [Dwo+12], one can consider a general metric to measure the similarity between two points. Furthermore, it would be interesting to see if we could improve the bounds of Theorem 4 for some special cases of the problem e.g., when the number of features is small or the set of values for features is small.

References

[Abb+19] Mohsen Abbasi, Sorelle A Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. "Fairness in representation: quantifying stereotyping as a representational harm". In: *Proceedings of the 2019 SIAM International Conference on Data Mining*. 2019, pp. 801–809 (cit. on p. 1).

[Aga+18] Alekh Agarwal, Alina Beygelzimer, Miroslav Dudik, John Langford, and Hanna Wallach. "A reductions approach to fair classification". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 60–69 (cit. on pp. 1, 3).

[Ahm+19] 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*. 2019, pp. 267–275 (cit. on pp. 1, 3).

[Ang+16] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. "Machine bias". In: *ProPublica, May* 23.2016 (2016), pp. 139–159 (cit. on p. 1).

[AV07] David Arthur and Sergei Vassilvitskii. "K-Means++: The Advantages of Careful Seeding". In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. 2007, pp. 1027–1035 (cit. on p. 8).

[Bac+19] Arturs Backurs, Piotr Indyk, Krzysztof Onak, Baruch Schieber, Ali Vakilian, and Tal Wagner. "Scalable fair clustering". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 405–413 (cit. on pp. 1, 3).

[Ber+19a] Suman Bera, Deeparnab Chakrabarty, Nicolas Flores, and Maryam Negahbani. "Fair algorithms for clustering". In: *Advances in Neural Information Processing Systems*. 2019, pp. 4955–4966 (cit. on pp. 1, 3, 6, 8).

[Ber+19b] Ioana Oriana Bercea, Martin Groß, Samir Khuller, Aounon Kumar, Clemens Rösner, Daniel R Schmidt, and Melanie Schmidt. "On the Cost of Essentially Fair Clusterings." In: *APPROX-RANDOM*. 2019 (cit. on p. 3).

[BTV06] Cristina Bazgan, Zsolt Tuza, and Daniel Vanderpooten. "The satisfactory partition problem". In: *Discrete applied mathematics* 154.8 (2006), pp. 1236–1245 (cit. on p. 4).

[Cel+19] L Elisa Celis, Lingxiao Huang, Vijay Keswani, and Nisheeth K Vishnoi. "Classification with fairness constraints: A meta-algorithm with provable guarantees". In: *Proceedings of the conference on fairness, accountability, and transparency*. 2019, pp. 319–328 (cit. on p. 3).

[Chi+17] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. "Fair clustering through fairlets". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, pp. 5036–5044 (cit. on pp. 1, 3, 8).

[Cho17] Alexandra Chouldechova. "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments". In: *Big data* 5.2 (2017), pp. 153–163 (cit. on p. 1).

[CR18] Alexandra Chouldechova and Aaron Roth. "The frontiers of fairness in machine learning". In: *arXiv preprint arXiv:1810.08810* (2018) (cit. on p. 3).

[Don+18] Michele Donini, Luca Oneto, Shai Ben-David, John Shawe-Taylor, and Massimiliano Pontil. "Empirical risk minimization under fairness constraints". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. 2018, pp. 2796–2806 (cit. on p. 3).

[Dwo+12] 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*. 2012, pp. 214–226 (cit. on pp. 1–3, 11).

[Elk03] Charles Elkan. "Using the Triangle Inequality to Accelerate K-Means". In: *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*. 2003, pp. 147–153 (cit. on p. 8).

[Fel+15] Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. "Certifying and removing disparate impact". In: *proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 2015, pp. 259–268 (cit. on p. 1).

[Fus+20] Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. "Predictably unequal? the effects of machine learning on credit markets". In: *The Effects of Machine Learning on Credit Markets (October 1, 2020)* (2020) (cit. on p. 1).

[GSV21] Mehrdad Ghadiri, Samira Samadi, and Santosh Vempala. "Socially fair k-means clustering". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021, pp. 438–448 (cit. on p. 3).

[HJV19] Lingxiao Huang, Shaofeng Jiang, and Nisheeth Vishnoi. "Coresets for Clustering with Fairness Constraints". In: *Advances in Neural Information Processing Systems* 32 (2019), pp. 7589–7600 (cit. on p. 3).

[JKL19] Christopher Jung, Sampath Kannan, and Neil Lutz. "A center in your neighborhood: Fairness in facility location". In: *arXiv preprint arXiv:1908.09041* (2019) (cit. on pp. 1–3, 8).

[KAM19] Matthäus Kleindessner, Pranjal Awasthi, and Jamie Morgenstern. "Fair k-center clustering for data summarization". In: *International Conference on Machine Learning*. 2019, pp. 3448–3457 (cit. on p. 3).

[KAM20] Matthäus Kleindessner, Pranjal Awasthi, and Jamie Morgenstern. "A notion of individual fairness for clustering". In: *arXiv preprint arXiv:2006.04960* (2020) (cit. on p. 3).

[Kle+19] Matthäus Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. "Guarantees for spectral clustering with fairness constraints". In: *International Conference on Machine Learning*. 2019, pp. 3458–3467 (cit. on p. 3).

[Meh+21] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. "A survey on bias and fairness in machine learning". In: *ACM Computing Surveys (CSUR)* 54.6 (2021), pp. 1–35 (cit. on p. 3).

[MV20] Sepideh Mahabadi and Ali Vakilian. "Individual fairness for k-clustering". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 6586–6596 (cit. on pp. 1–3, 8).

[Obe+19] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting racial bias in an algorithm used to manage the health of populations". In: *Science* 366.6464 (2019), pp. 447–453 (cit. on p. 1).

[RS18] Clemens Rösner and Melanie Schmidt. "Privacy Preserving Clustering with Constraints". In: *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2018 (cit. on p. 3).

[Sam+18] Samira Samadi, Uthaipon Tantipongpipat, Jamie Morgenstern, Mohit Singh, and Santosh Vempala. "The price of fair PCA: one extra dimension". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. 2018, pp. 10999–11010 (cit. on p. 3).

[Sel+19] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. "Fairness and abstraction in sociotechnical systems". In: *Proceedings of the conference on fairness, accountability, and transparency*. 2019, pp. 59–68 (cit. on p. 3).

[SSS18] Melanie Schmidt, Chris Schwiegelshohn, and Christian Sohler. "Fair coresets and streaming algorithms for fair k-means clustering". In: *arXiv preprint arXiv:1812.10854* (2018) (cit. on p. 3).

[VY21] Ali Vakilian and Mustafa Yalçıner. "Improved Approximation Algorithms for Individually Fair Clustering". In: *arXiv preprint arXiv:2106.14043* (2021) (cit. on p. 3).