IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 27, 2021, accepted September 17, 2021, date of publication September 20, 2021,
date of current version September 29, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3114099

An Overview of Fairness in Clustering

ANSHUMAN CHHABRA[1], KARINA MASALKOVAITĖ[2], AND PRASANT MOHAPATRA[1], (Fellow, IEEE)
[1]Department of Computer Science, University of California, Davis, CA 95616, USA
[2]Department of Chemical Engineering, University of California, Davis, CA 95616, USA

Corresponding author: Anshuman Chhabra (chhabra@ucdavis.edu)

ABSTRACT Clustering algorithms are a class of unsupervised machine learning (ML) algorithms that feature ubiquitously in modern data science, and play a key role in many learning-based application pipelines. Recently, research in the ML community has pivoted to analyzing the fairness of learning models, including clustering algorithms. Furthermore, research on fair clustering varies widely depending on the choice of clustering algorithm, fairness definitions employed, and other assumptions made regarding models. Despite this, a comprehensive survey of the field does not exist. In this paper, we seek to bridge this gap by categorizing existing research on fair clustering, and discussing possible avenues for future work. Through this survey, we aim to provide researchers with an organized overview of the field, and motivate new and unexplored lines of research regarding fairness in clustering.

INDEX TERMS Machine learning, clustering, fairness, fair clustering.

I. INTRODUCTION

Machine Learning (ML) has been used to tackle many important problems, many of which can have significant societal implications. Some of these problems include predicting the likelihood of prisoner recidivism [1]–[5], disbursement of bank loans [6]–[8], shortlisting candidates for job applications [9]–[13], and college admissions [14]–[16]. Since ML models train on large datasets that have been found to contain biases against both individuals and minority groups, they can further amplify biases when used in high-impact applications. This has been evidenced in many ML applications where fairness was not considered to be an evaluation criteria. Some examples are Microsoft's *Tay* online chatbot which learned from tweets and due to biased inputs started using racist slurs [17], and the COMPAS tool which predicted that a black individual is more likely to commit a crime [18] than a white individual even if both individuals are statistically similar with regards to other attributes.

To rectify models and correct for unfairness, ML researchers have recently begun to propose approaches that ensure fairness constraints are met [19]–[25]. However, defining fairness notions is not a trivial task, and is often done based on application and legal context. For example, fairness can be defined for minority protected groups (such as for ethnicity, gender, etc) [26] or for individuals (that is, similar individuals should be treated equitably) [27], and both possess certain advantages and disadvantages depending on where they are being utilized. It has been found that different notions of fairness are generally incompatible [28], [29] with one another and cannot be jointly optimized for, further compounding the difficulty of the problem.

Clustering algorithms are unsupervised ML algorithms that are widely utilized in problem settings where labels are not easily available (such as resource allocation problems). Moreover, recently, the issue of fairness for clustering has received considerable attention in the ML community, pioneered by the first work on fair clustering by Chierichetti et al [30] in 2017. However, ensuring fairness for clustering is harder than the general ML case, as labels are not present with the data, and ground-truth error rates cannot be calculated to estimate bias and unfairness. This makes both *defining* and *enforcing* fairness for clustering, challenging problems.

Due to this reason, many different fairness notions for clustering exist (for example [30]–[34]), with different research papers opting for different metrics, or proposing new ones. Furthermore, techniques for ensuring fairness constraints are met vary widely in methodology; comparisons between different fair approaches are usually made selectively, and there are no established (fairness and performance) metrics that are adopted for comparison. There are also no surveys or review articles that have been compiled for fair clustering approaches. This is in stark contrast with other ML sub-fields, where multiple surveys exist– such as for recommendation systems [35], natural language processing models [36], learning to rank models [37], and sequential decision-making approaches [38], among others.

The associate editor coordinating the review of this manuscript and approving it for publication was Ting Wang.

Therefore, we aim to bridge this gap and organize the field through this article. Our goal is to provide both existing and new researchers in fair clustering with an overview of the field, along with new insights. We categorize the myriad of approaches in fair clustering similar to other ML survey articles, and provide many different classifications for fairness notions for clustering. Our work also discusses real-world applications for fair clustering as well as datasets used for evaluating fair clustering approaches. Thus, the article can also serve as a tool for ML practitioners aiming to utilize fair clustering in their applications. To summarize, the contributions of this work are as follows:

- We provide the first survey on fair clustering that organizes the field and categorizes fair clustering approaches similar to other ML surveys.
- We classify the many different available fairness notions for clustering, provide details regarding the evaluation of fair models, and the datasets frequently used for the same.
- We discuss motivations for clustering using real-world applications to aid ML practitioners, and also provide a multitude of new research directions for the field.

The rest of the paper is structured as follows: Section II details relevant background regarding clustering and fairness in ML. Section III discusses different fairness notions employed for clustering and how they can be organized into intuitive sub-categories. Section IV describes the different approaches used to make clustering fair. Section V examines the datasets used for evaluating fair clustering, and motivates the research problems related to fair clustering through real-world applications. Section VI provides insights and analysis for future work in fair clustering, and Section VII concludes the paper.

II. PRELIMINARIES AND NOTATION

In this section, we briefly discuss the working of clustering algorithms and give an overview of the different approaches used to make ML models fair. We also detail the notation and symbols used throughout the paper.

A. CLUSTERING ALGORITHMS

Formally, a clustering algorithm $\mathcal{A}$ seeks to partition a given input dataset $X \in \mathbb{R}^{n \times m}$ into some $k \leq n$ clusters. Moreover, each sample $x \in X$ can belong to one (*hard* clustering) or more (*soft* clustering) of the k clusters, depending on the clustering objective used. Let $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ denote the output partition set obtained by running the clustering algorithm $\mathcal{A}$, where $C_i \subseteq X$, $\forall i \in [k]$. As there are no labels present for the data samples, X is both the *training* dataset and the *testing* dataset for the clustering problem. This is different from traditional supervised learning and classification tasks, where training datasets and test datasets are separate. The unsupervised nature of the clustering problem also further complicates the issue of defining and enforcing fairness, which we discuss in subsequent sections. It is important to note that most clustering objectives (such as k-means, hierarchical clustering, k-medoids, etc) are generally NP-Hard [39]–[41] and are usually solved using algorithms that approximate the optimal solution [42], [43] or through heuristic approaches [44]. For example, for the widely used k-means clustering objective [40], the expectation-maximization based Lloyd's algorithm is used [44] as a heuristic which works very well in practice.

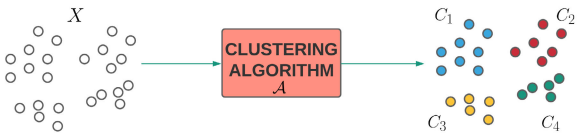

FIGURE 1. An example demonstrating the general clustering process (different colors represent different clusters).

Another distinguishing feature of clustering algorithms is that the number of clusters k could be given as an input to the learning model or obtained via the clustering optimization problem itself. For example, in center-based clustering algorithms such as k-means [44] or k-medoids [43], k is an input parameter, but hierarchical clustering algorithms [39], [45], [46] output a tree of clusters, with each level of the tree indicating a possible choice of $k \leq n$ that the user can opt for. Other algorithms such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [47] and Ordering Points To Identify the Clustering Structure (OPTICS) [48] also do not require number of clusters as input, but infer a single value for k from the dataset provided. We defer the reader to [49] for more details on different clustering algorithms.

Unless otherwise specified, we generally consider hard clustering for example scenarios in this article, that is each point can only belong to one cluster. In Fig. 1, we provide an overview of the aforementioned general clustering process. The original dataset X is provided as input to the clustering algorithm $\mathcal{A}$ and we obtain the cluster partition set $\mathcal{C} = \{C_1, C_2, C_3, C_4\}$ as output, shown in blue, red, yellow, and green respectively.

B. A BRIEF TAXONOMY OF CLUSTERING METHODS

Many different clustering methods have been proposed to partition data into meaningful clusters, and a preliminary knowledge of these is useful before delving into the numerous approaches proposed for fair clustering. For ease of understanding, we borrow from (and slightly modify) the classifications originally proposed by Xu and Wunsch [49] for differentiating data clustering methods. As a complete in-depth discussion is out of the scope of this work, we refer the reader to the surveys [49], [50] for more details on approaches for clustering data.

Clustering algorithms can be generally categorized into the following:

1) CENTER-BASED CLUSTERING

These approaches aim to partition the input dataset into clusters by minimizing an error metric between data samples assigned to a cluster, and their corresponding cluster centers. Depending on the defined error metric, cluster centers can be either the mean of the samples in the cluster (such as in k-means [44]), or the median of samples in the cluster (such as in k-medoids [43]), among many other possibilities. The most common approach for this category is k-means where the error term is defined to be the squared Euclidean distance between cluster samples and cluster centers [51]. Many different variations for k-means have been proposed that improve upon the original heuristic algorithm [52]–[54]. Other methods include k-medoids [43], Iterative Self-Organizing Data Analysis Technique (ISODATA) [55], among others.

2) HIERARCHICAL CLUSTERING

Hierarchical clustering approaches aim to partition the dataset into hierarchies, with the clustering output represented as a binary tree. The root node represents the entire dataset while the leaf nodes comprise of the singular samples of the dataset. The remaining nodes of the tree represent clusters, and in this way, a hierarchy of clusters is obtained. *Agglomerative* hierarchical clustering algorithms aim to build this tree in a *bottom-up* fashion, whereas *divisive* hierarchical clustering algorithms seek to do so in a *top-down* fashion. Some examples of agglomerative hierarchical clustering algorithms include Clustering Using Representatives (CURE) [56], Ward's method [57], Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) [58], Robust Clustering using Links (ROCK) [59], among others. For divisive hierarchical clustering, examples include the Divisive Analysis algorithm (DIANA) and Monothetic Analysis algorithm (MONA) [60]. Recently, analytical objectives for hierarchical clustering have also been proposed [39], [45], [46] which have lead to the development of more theoretically robust hierarchical clustering algorithms.

3) MIXTURE MODEL-BASED CLUSTERING

Mixture model-based clustering refers to a probabilistic clustering approach where points are assigned to clusters in a soft manner, and do not have hard memberships. Furthermore, data points are assumed to originate (and belong to) some mixture of probability distributions. In this clustering approach, the nature of distributions are generally assumed (very often to be a mixture of multivariate normal distributions). Then the clustering task transforms into finding the set of parameters for this mixture of distributions that maximize a metric such as log-likelihood (or how likely a point is determined to belong to a particular cluster). Popular clustering approaches that belong to this category are Gaussian-Mixture-Model based Expectation Maximization (GMM-EM) [61], Expectation Maximization-based Mixture program (EMMIX) [62], and AutoClass [63], among others.

4) GRAPH-BASED CLUSTERING

Graph-based clustering approaches utilize concepts from graph theory to cluster the data. This first requires translating the original dataset into a graph problem, by treating data samples as nodes/vertices in a graph, and creating edges between samples using a dissimilarity/similarity metric. The dissimilarity/similarity metric is usually defined using a distance metric between points. Then, edges can be created between nodes if points are within a certain distance threshold, often using a k-nearest-neighbor graph [64]. On obtaining a graph describing the original data, the Laplacian matrix can be obtained. Clustering using k-means (or other simple clustering algorithms) is then undertaken on the eigenvectors of the Laplacian, and the original data samples can be assigned the same cluster labels [65]. Depending on the choice of the graph Laplacian, different spectral clustering outputs can be obtained [66]. Many other graph-based clustering approaches also belong to this category, such as Clustering Identification via Identity Kernels (CLICK) [67], Delaunay Triangulation Graph based clustering (DTG) [68], among others.

5) FUZZY CLUSTERING

Fuzzy clustering algorithms consist of soft clustering approaches where data samples have *fuzzy memberships* (a grade of membership between 0 and 1) to clusters instead of binary cluster assignments. The most popular fuzzy clustering method is Fuzzy C-Means (FCM) [69]. Many improvements have been made upon FCM, including methods that more easily identify centers [70], generalize the algorithm to arbitrary distance metrics [71], reduce time complexity [72], and more. Fuzzy clustering can also be combined with hierarchical clustering, as was done in Hierarchical Unsupervised Fuzzy Clustering (HUFC) [73].

6) COMBINATORIAL SEARCH-BASED CLUSTERING

Exactly solving most clustering optimization objectives can be NP-Hard as there often exists an exponential search space of clustering solutions. Thus, the clustering problem can be reformulated as a combinatorial optimization problem, and local search approaches can be used to approximate the optimal clustering solution. Most often, due to the hardness and generality of the problem, evolutionary approaches [74] are used for the search algorithm, such as Simulated Annealing (SA) [75], Genetic Algorithms (GA) [76], etc. Clustering approaches that belong to this category include Genetically Guided Algorithm (GGA) for clustering [77], Genetic k-means Algorithm (GKA) [78], among others.

C. FAIRNESS IN ML

Fairness for ML models can be enforced/ensured in three stages of the learning pipeline [17], [79], [80]; in the 1) *before-training*, 2) *during-training*, or 3) *after-training* phase:

1) The before-training stage requires that the original data be *pre-processed* to obtain a new dataset. On training/running the unchanged ML model on this new

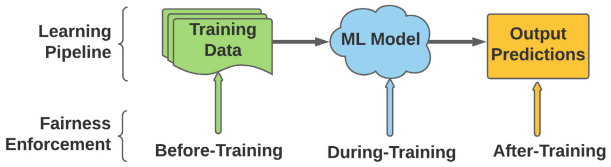

FIGURE 2. The general ML/fairness pipeline.

dataset, the output predictions will meet the fairness constraint.

2) The most common approach to improving fairness for ML models is the during-training or *in-processing* stage, where the ML model itself is modified to include the fairness constraints. This involves changing the optimization and training process such that the output predictions are fair, without changing the original dataset.
3) Finally, fairness can be enforced after-training as well, where the predictions from the original model undergo a *post-processing* procedure to compute a similar set of predictions such that they now meet the desired fairness constraints.

We detail these methodologies in the context of the learning pipeline in Fig. 2. As mentioned before, since clustering is an unsupervised learning problem where training and test datasets are the same, the diagram shown in Fig. 2 will also change accordingly to account for this. We thus discuss approaches specific to clustering in Section IV of the paper, which build upon the high-level schematic of Fig. 2. We do not discuss *how* fairness of general ML models (such as for classification, computer vision, etc.) can be measured via analytical metrics as this is outside the scope of this work. We discuss fairness metrics and notions specific to clustering in Section III of the paper, but interested readers can refer to [17] for more information on fairness notions for general ML models.

III. FAIRNESS NOTIONS FOR CLUSTERING

In this section we discuss the different notions of fairness that are generally employed for clustering. As mentioned before, fairness notions are often application specific and a particular definition might be more preferable in certain settings compared to others. For example, consider an adapted version of the application scenario provided in [31]. We have to find where to set up three ($=k$) parks for a given set of houses in a region. For this, we can use center-based clustering algorithms where each center could denote a possible park location. In this region, we have two dense city sub-regions with housing highly localized in smaller area, and a residential sub-region which encompasses large area but is less dense than the city. This scenario is shown in Fig 3. Now, if a general center-based clustering algorithm (such as k-means) was used, we would obtain a single cluster center (park) to share for the city sub-regions whereas the larger-sized suburban sub-region would get two parks.

This is unfair to the individuals living in the dense sub-regions and hence, this application requires a definition of fairness which warrants *proportionally* shared cluster centers. A fair solution (in this context) would distribute two centers for each of the two dense city sub-regions and one park for the larger/sparser sub-region. Thus, the definition of *proportionality* proposed by [31] is more suitable than other fairness notions (such as the most commonly used *balance* proposed by [30]). The former captures the idea that data samples are individuals and fairness to these individuals means being clustered in an accurate manner with regards to their dataset features and cluster centers. The latter on the other hand, aims to capture the degree to which points belonging to protected groups are represented in each output cluster. It is then evident that for the example considered above, proportionality is the better fairness notion. Note that proportional fairness is also more apt for this scenario as it does not require protected groups (in contrast with balance which explicitly requires groups) and can be tailored to the fairness requirement at the sample level.

The above example then introduces an interesting research question: *are there ways to distinguish clustering fairness notions from each other?* We answer this question in the affirmative by introducing four different classifications for fairness definitions: *group-level, individual-level, algorithm agnostic*, and *algorithm specific* fairness. Fairness notions can belong to more than one category as well. As fairness notions for clustering have not been formally categorized before, we aim for these to be a simplistic first step in doing so; many other different classifications/categorizations are possible. Subsequently, we explain each category individually and then provide existing definitions/analysis using our proposed categories for all fair clustering notions proposed to date.

A. GROUP-LEVEL NOTIONS

Group-level fairness notions are usually derived from the Disparate Impact (DI) doctrine [81] which states that no group of individuals should be adversely affected by the outcome of a decision-making system. That is, no group of individuals should be discriminated against or overtly preferred by an algorithm in terms of the output predictions made.

This category of fairness can be understood through an example. A dataset, e.g., the creditcard dataset [82], is used by the marketing division of a bank to reach out to prospective customers and offer them loans and available credit opportunities. The dataset contains features such as the potential customer's age, their education level, their weekly work hours, and their capital gains per month. The bank utilizes a clustering algorithm to find target audiences for promotional offers and uses the aforementioned attributes as input to the clustering algorithm. That is, on running the algorithm, they obtain *clusters* of people who are then (using some metrics, e.g., the education and wages-earned features) grouped together to be targeted for a particular promotion/offer. It is important to note here that people-of-color (POC) as well as

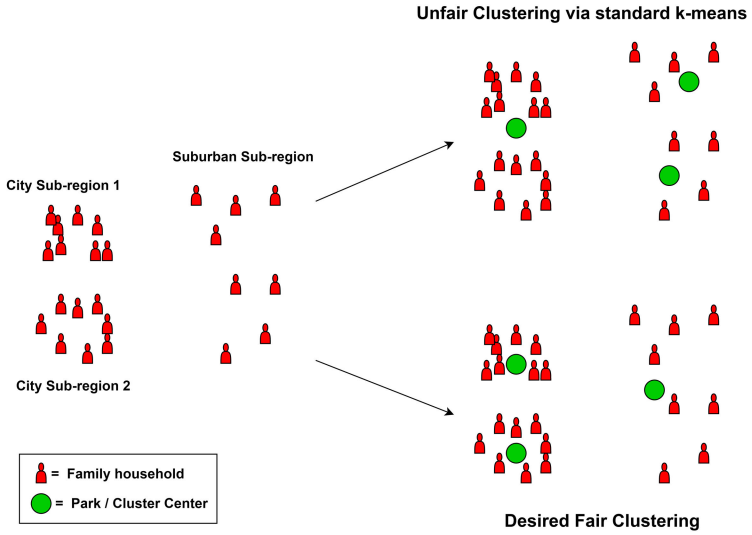

FIGURE 3. Example scenario for understanding fairness notions for clustering.

women, tend to earn lower wages than white males [83], and that POC face more adversities that lead to disparities in their education level [84] as opposed to white demographics. Now, considering these facts on the racial education divide and the wage gap, a clustering algorithm using these attributes will inherently group white households as well as men, as better candidates for better deals and offers (such as mortgages and loans). As a result, this marketing clustering algorithm has *disparate impact* on POC as well as women, as they are deprived of an opportunity of improvement. Therefore, it is important to study *protected groups* (e.g., ethnicity and gender) and the corresponding fairness in such a clustering setting. Group-level fairness measures thus aim to capture this setting in an analytical manner.

An example of a group-level measure is the balance notion first proposed by [30] and then generalized by [85]. It requires calculating the ratio between the proportion of total protected group members in the dataset and the proportion of protected group members in a cluster, and the balance of the clustering is then the minimum value obtained over all clusters and protected groups. As a result, it always lies between 0 and 1, with higher values indicating a clustering output that is more fair. We also found that balance is the most commonly used fairness notion for most research on fair clustering.

Other group-level fairness notions include *bounded representation* [86] which considers two parameters α and β which denote the allowed maximum and minimum proportions of protected group members that can be present in a cluster. Thus, through this notion no protected group members should be over or under preferred for each cluster. Another example is the Max Fairness Cost (MFC) proposed in [32] which is similar to balance but takes a user-inputted *ideal proportion* value as well. It measures the deviation of the current proportion of protected group members in a cluster from this ideal proportion using the L1 norm. We discuss other categories next and then provide a complete tabular list of group-level fairness notions in Table 1.

B. INDIVIDUAL-LEVEL NOTIONS

Individual-level fairness notions are significantly different from the group-level fairness notions. Here, we do not have any protected groups, and the goal is to ensure that *similar* individuals (samples in the dataset) are treated similarly by the ML model. That is, a clustering model abiding by individual-level fairness would cluster all individuals that are deemed similar using some dissimilarity metric in a similar manner. The proportional notion of fairness [31] discussed before is an example of an individual-level fairness notion for clustering.

Individual-level fairness for clustering has not been studied as extensively as group-level fairness, and most works only focus on facility location and center-based clustering. The differences in these individual-level fairness definitions stem from 1) how the dissimilarity metric is defined between individuals, and 2) how similarity is measured with regards to the output clustering. In [87] the authors assume that the dissimilarity metric is available as a distance metric d, and that a clustering satisfies individual-level fairness if for each individual sample in the dataset the average distance (measured using d) to samples in its cluster is less than the average distance (measured using d) to any samples in other clusters. In [34] and [88] the authors provide an alternative definition for individual-level clustering fairness: every sample in the dataset should have a center within a distance R where R is the minimum radius of the ball centered around the sample that contains at least n/k (total samples over number of clusters) samples. In [89] individual-level fairness is extended for the clustering setting from the seminal work of [26] for classification. Here, the authors consider soft clustering outputs and as the clustering is probabilistic they enforce individual-level fairness through distributional similarity of the cluster outputs. Very recently, more research has emerged on individual fairness for clustering [90]–[94], and we cover these in more detail in Section IV.

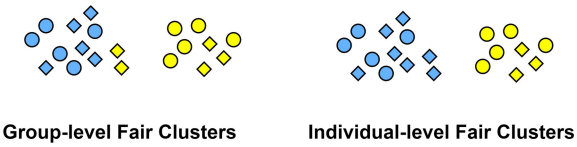

FIGURE 4. Group-level and individual-level fairness notions for clustering can have conflicting cluster assignments (colors indicate clusters and diamonds/circles indicate protected group memberships).

As mentioned previously, different fairness notions can often not be applied together. This is true for group-level and individual-level fairness notions for clustering. In particular, in [95] and [96], the authors find that forcing group-level fairness can adversely affect individual-level fairness between similar individuals. This can also be seen through a simple example shown in Fig. 4 which has been adapted from [89]. Here, different protected groups are denoted using different markers and different cluster assignments are denoted using different colors. The cluster assignments required to meet group-level fairness (for example, enforced through balance) are shown on the left and the cluster assignments to satisfy individual-level fairness are shown on the right in Fig. 4. This is because for group-level fairness each group needs to be represented in a cluster in similar proportion whereas for individual-level fairness we would like closely distanced (similar) points to be clustered together (similarly). As can be seen from the figure, these are mutually exclusive cases, hence only one notion of fairness can be enforced at a time. We provide a complete list of individual-level fairness notions in Table 1 towards the end of this section.

C. ALGORITHM AGNOSTIC NOTIONS

We also categorize fairness notions based on whether they are designed specifically for certain clustering objectives or can generalize to any given objective. Algorithm agnostic notions are generally defined for the cluster output level and can thus generalize for all clustering objectives. For example, the first proposed fairness notion balance [30], [85] discussed previously, essentially operates with cluster outputs given by any clustering algorithm. This makes it an algorithm agnostic fairness notion.

Note that any fairness notions which do not make explicit assumptions regarding clustering algorithms, but implicitly require specific clustering behavior are *not* considered as algorithm agnostic. For example, for the proportional fairness notion [31], while there is no explicit clustering algorithm mentioned in the definition, the notion requires cluster centers, thus limiting it only to center-based clustering objectives. Furthermore, both group-level and individual-level fairness notions can be algorithm agnostic. We also find that most group-level fairness notions are algorithm agnostic. Algorithm agnostic notions are tabulated towards the end of the section (Table 1).

D. ALGORITHM SPECIFIC NOTIONS

Algorithm specific fairness notions constitute fairness notions that work specifically for certain clustering objectives and algorithms. One example is the k-means *social* fairness cost, proposed by [33]. In their work, the authors define a fair clustering to be one where the average k-means cost for each protected group is minimized. While this aspect of social fairness could be extended to other learning tasks, the current work seeks to do so for k-means, making it specific to center-based clustering objectives. Other examples include proportional fairness proposed by [31] and the individual-level fairness notions of [34], [88] as they only work with center-based clustering. A full list is provided in Table 1.

E. DEFINITIONS FOR COMMONLY USED NOTIONS

In this subsection, we provide mathematical definitions for some commonly used fairness notions. However, due to the multitude of different notions proposed, we defer the list of all notions to Table 1 and provide pointers to appropriate related works that discuss and define these notions there.

We now provide technical definitions for the following fairness notions:

1) BALANCE

The group-level and algorithm agnostic fairness notion of balance was first proposed by Chierichetti *et al.* [30] for the case with 2 protected groups. It was later generalized to the multiple group case by Bera *et al.* [85]. Since then, balance has been employed as the fairness metric for most research on fair clustering [97]–[100].

Let there be m protected groups. Then, define r and r_a to be the proportion of samples of the dataset belonging to protected group b and the proportion of samples in cluster $a \in [k]$ belonging to protected group b. Then define another ratio for this cluster and protected group as $R_{a,b} = r/r_a$. The balance fairness notion is then defined over all clusters and protected groups as:

$$\min_{a\in[k],b\in[m]} \min\{R_{a,b}, \frac{1}{R_{a,b}}\}$$

As can be seen through the definition, balance lies between 0 and 1, and the higher the value, the more fair the clustering output. That is, a fair algorithm will attempt to maximize the notion of balance. This is usually done as a constraint to ensure that the balance is either lower-bounded or upper-bounded by a required pre-defined input value.

Some authors implicitly utilize the balance fairness notion but reformulate it to aid theoretical analysis. One such example is in [101] and [102]. Let there be m protected groups, and samples of dataset X in cluster a that belong to group b are denoted using the set $G_{a,b}$. Then, define for cluster a, $J_a = \min_{b\in[m]} G_{a,b}$ and $L_a = \max_{b\in[m]} G_{a,b}$. Then the reformulated notion of balance is:

$$\min_{a\in[k]} \frac{J_a}{L_a}$$

As is evident, this also outputs a value between 0 and 1, and the authors also provide theoretical analysis to show that

minimizing this notion of fairness is equivalent to minimizing the original 2-group balance notion proposed by [30].

2) SOCIAL FAIRNESS

The social fairness cost was proposed by Ghadiri *et al.* [33] for the k-means clustering objective. A similar notion of group representative fairness was developed by Abbasi *et al.* [103] for k-means and k-medians. Markarychev and Vakilian [104] generalized the social fairness problem, but here we present the k-means case as originally defined. In its current formulation, this fairness notion is algorithm specific, as it can only be used for center-based clustering.

Assume here also without loss of generality that there are m protected groups. Define the k-means clustering cost for a set of k cluster centers U and the input dataset X as $O(U, X) = \sum_{x \in X} \min_{u \in U} ||x-u||^2$. Also, let X_a denote the samples of X that belong to protected group a. Then the social fairness cost for k-means clustering becomes:

$$\max_{a \in [m]} \frac{O(U, X_a)}{|X_a|}$$

As the above notion is a cost, it needs to be minimized unlike balance which was to be maximized. That is, the lower the social fairness cost the more fair the clustering.

3) BOUNDED REPRESENTATION

The notion of bounded representation was proposed by Ahmadian *et al.* [86]. It is a group-level notion and can be defined using two parameters α and β. The fairness notion is defined through constraints that need to be imposed and met for each cluster obtained via the clustering algorithm. Let $P_{a,b}$ be the proportion of protected group $b \in [m]$ members in cluster $a \in [k]$. Then, for (α, β)- bounded representation we require that:

$$\beta \leq P_{a,b} \leq \alpha, \quad \forall\, a \in [k],\ b \in [m]$$

Essentially, unlike the other notions discussed previously, this notion is defined as a set of constraints. If all the fairness constraints for each group and cluster are met, the clustering is fair. This notion of fairness can also be defined by only considering either the upper-bound (α) or lower-bound (β) on the proportion of points. If $\alpha = \beta = 1/m$ then the notion aims to represent each group with equal proportion in the clustering output. Bounded representation has been used in conjunction with a number of clustering objectives as well [86], [105].

4) MAX FAIRNESS COST (MFC)

The MFC was defined by [32] for heuristic hierarchical agglomerative clustering algorithms. Despite this, it is an algorithm agnostic fairness notion as it works at only one level of the tree hierarchy, making it apt for any clustering algorithms with k cluster outputs. It is also a group-level notion and requires an additional parameter named the ideal proportion (I_b) defined for each protected group $b \in [m]$. Here, I_b is given by the user and provided at run-time, and can vary to account for different application requirements. Then if the proportion of group $b \in [m]$ points in cluster $a \in [k]$ are given as $P_{a,b}$, the MFC is defined as:

$$\max_{a \in [k]} \sum_{b \in [m]} |P_{a,b} - I_b|$$

The MFC is essentially the maximum of the sum of all deviations from the ideal proportion for each protected group in a cluster. The lower the MFC, the better the fairness achieved by the clustering. If the parameter I_b is set to $1/m$ then the fairness notion aims to ensure that each protected group is represented with the equal proportion in each cluster.

5) DISTRIBUTIONAL INDIVIDUAL FAIRNESS

This individual-level fairness notion was proposed by [89]. Here, a fairness similarity measure $F \in \mathbb{R}^+$ is assumed to be known that operates on a pair of samples from the dataset X. To ensure fairness, the statistical distance obtained using the f-divergence [106]–[108] for the output distributions of each pair of samples should be smaller than the distance obtained using the F metric. Also, the fairness notion is algorithm specific as it assumes cluster centers are available, limiting applicability to center-based clustering. It also assumes probabilistic clustering (a setting such as Gaussian Mixture Model based soft clustering [109]) for the problem definition. Their work extends the notion of individual fairness proposed for classification by [26].

Let U denote a k-sized cluster center set. Also let the f-divergence between the distributions V_x, V_y cast over U for pair of samples $x, y \in X \times X$ be denoted as $H_f(V_x||V_y)$. Then the distributional individual fairness notion requires that the following is met for all pairs of dataset samples $x, y \in X \times X$:

$$H_f(V_x||V_y) \leq F(x, y)$$

Note here that for the f-divergence, many possible definitions exist that can be used, such as the KL-divergence [110].

6) KLEINDESSNER *et al.* INDIVIDUAL FAIRNESS

This is another individual-level notion of fairness proposed by [87]. Unlike the previous individual-level notion, this works at the level of the clustering output $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ and hence, is algorithm agnostic. For each sample x in the dataset X, let d be a well-defined clustering distance metric and C_a be the cluster that x belongs to. Then, the fairness notion of [87] can be defined as a set of constraints for the sample x and all clusters $b \in [k], b \neq a$ as:

$$\frac{1}{|C_a| - 1} \sum_{z \in C_a} d(x, z) \leq \frac{1}{|C_b|} \sum_{z \in C_b} d(x, z)$$

If all the above constraints are met for all the individual samples in the dataset X, the clustering is deemed to be individually fair.

7) ENTROPY

Entropy is a fairness metric that was defined in [111], and has only been exclusively used for fairness in the context of deep clustering models. A distinction of deep clustering with respect to general clustering methods is that ground truth labels for each sample are known prior to training. Also, similar to balance, the higher the entropy the more fair the model. Let $N_{a,b}$ be the set containing the samples of the dataset X that belong to both the cluster $a \in [k]$ and the protected group b. Further, let n_a be the number of samples in cluster a. Then entropy is defined as follows:

$$-\sum_{a\in[k]} \frac{|N_{a,b}|}{n_a} \log \frac{|N_{a,b}|}{n_a}$$

IV. APPROACHES FOR FAIR CLUSTERING

In this section, we comprehensively discuss research to-date on fair clustering, along two dimensions: 1) the clustering objective the fairness intervention is for, and 2) what stage of the learning pipeline the intervention falls into (refer to Section II). In the first subsection that follows, we summarize all fair clustering approaches by categorizing them based on the clustering objective they employ. This includes *center-based clustering* (such as k-means, k-center, k-median), *hierarchical clustering*, *spectral clustering*, and *deep clustering* models. Since there are certain approaches that are either more general or do not belong to either of the aforementioned clustering objectives, we also have a *miscellaneous* category. We find that the most common clustering objective considered for fair clustering approaches is center-based clustering– in particular, this is one possible direction where future work can improve on (Section VI).

In the second subsection, we consider the categorization and discussion of fair clustering approaches based on what stage of the clustering pipeline the enforcement is targeting. Initially in Section II we had provided the distinctions between the pre-processing/in-processing/post-processing methodologies for general ML models. We apply this same terminology for the classification of fair clustering approaches. It is important to note that for clustering, the learning pipeline is a little different compared to traditional ML models as the training and test datasets are the same. Therefore, in the second subsection we first describe the fairness intervention stages (*pre-processing/in-processing/post-processing*) in the clustering context and then discuss categorization.

A. CLUSTERING OBJECTIVE

1) CENTER-BASED CLUSTERING

We now discuss all research on making center-based clustering fair. Also note that in fair clustering literature (and in general, for clustering), k-median(s) and k-medoids clustering are often used interchangeably to describe the latter problem. Technically, these clustering objectives are very different– k-median(s) refers to minimizing the L1 norm and cluster centers need not be *exemplars* (must be points in the original dataset), whereas for k-medoids the goal is to minimize the sum of pairwise dissimilarities defined using any distance metric, and centers need to be *exemplars*. As in other related clustering work, we will refer to latter case as k-medians, with the implicit assumption that cluster centers are exemplars. In case we discuss any deviations from this objective, we shall state it explicitly to avoid ambiguity.

Group-Level Fairness: Chierichetti *et al.* presented the first work on group-level fair clustering, specifically for the k-center and k-median clustering objectives while considering the case with only two protected groups [30]. They introduced the fairness notion of balance, which we discussed previously. To balance output clusters, they proposed the fairlet decomposition method. Fairlet decomposition is a pre-processing approach that computes fair *micro-clusters* where fairness is guaranteed. They then use the fairlet centers as a newly transformed dataset from the original. This transformed fairlet-based dataset is then provided to vanilla clustering algorithms, and hence, we obtain approximately fair clustering outputs as a result of the fairlets themselves being fair. The fairlet decomposition approach is also visually described in Fig. 5 to improve understanding. Note that fairlet decomposition can generally be used with any fairness notions but proposing efficacious approaches for computing fairlets is not a trivial task in itself.

Subsequently, Backurs *et al.* [99] improved the computational time complexity of fairlet decomposition by proposing a nearly-linear time scalable algorithm, but only for k-median clustering. Rösner and Schmidt [113] extended the fairness framework of [30] to allow for multiple protected groups and obtained a 14-approximation fair algorithm for the k-center objective.

Schmidt *et al.* [97] introduced *coresets* for fair k-means clustering, which allowed for a more scalable approach than fairlets, and also are more applicable when random-access to the dataset might not be allowed (required for fairlet decomposition). Coresets are essentially a summary of a given point set, such that they effectively approximate the cost function for any possible candidate solution and the *fair coresets* introduced in [97] aim to do this while also enforcing fairness for the case with two protected groups. Huang *et al.* [121] extended fair coresets for k-median clustering and remove the dependence of dimension for fair coreset generation in the case of k-means. Further, their approach works for multiple disjoint protected groups. Bandyapadhyay *et al.* [122] proposed the first Fixed-Parameter Tractable (FPT) time constant factor approximation algorithms for k-median and k-means while removing the dimension dependency for coreset generation. We visually describe the fair coreset approach in Fig. 5.

A number of papers expanded upon the original fairness notion of balance [30] by introducing upper and/or lower bounds to protected group membership in clusters, also previously referred to as the *bounded representation* notion. Ahmadian *et al.* [86] used only an upper bound constraint for protected group representation in clusters for fair k-center

TABLE 1. Detailed descriptions of fairness notions and their classifications.

Fairness Notion	Definition	Group-Level	Individual-Level	Algorithm Agnostic	Algorithm Specific	Clustering Algorithm	First Proposed By	Used In
Balance	minimum ratio between protected group members in a cluster and protected group members in the data set, measured over all groups and clusters	✓		✓		any	[30]	[85], [97], [98], [99], [100], [112], [105], [32], [113], [111], [114], [115], [116], [102], [117], [101]
Bounded Representation	clustering algorithm is constrained by two parameters that define the allowed maximum and minimum proportions of protected group members in a cluster	✓		✓		any	[86]	[105], [85], [86], [97], [118], [119], [120] [121] [122] [123]
Social Fairness Cost	minimum average center-based clustering costs for each protected group, ensures nearby clusters are similar and a representative exists that accurately portrays the cluster	✓			✓	center-based clustering	[33], [103]	[104], [116], [124]
Proportionality	for n samples and k clusters, any n/k points are entitled to form their own cluster if there is another center that is closer in distance for all n/k points		✓		✓	center-based clustering	[31]	[93]
Jung et al Individual Fairness	for point x in a dataset X of size n, let $r(x)$ be the minimum radius such that the ball of radius $r(x)$ centered at x has at least n/k points from X, given k clusters		✓		✓	center-based clustering	[88]	[34], [91], [92]
Diversity-Aware Fairness	requires some minimum number of cluster centers in this clustering are chosen from each protected group provided	✓			✓	center-based clustering	[125]	-
Group-Utilitarian	aims to minimize the sum of proportional violations for all the protected groups	✓		✓		any	[126]	-
Group-Egalitarian	aims to minimize the maximum proportional violations for all the protected groups	✓		✓		any	[126]	-
Group-Leximin	aims to minimize the largest proportional violations for all the protected groups in order (i.e. the maximum proportional violation first, then the second maximum proportional violation, and so on)	✓		✓		any	[126]	-
Abraham et al Group Fairness	group specific deviation of each cluster summed over all protected groups, weighted by the square of the cluster's fractional representation over the entire dataset, and summed over all the clusters	✓		✓		any	[127]	-
Fair Summaries	requires each group is represented equally in the data summary	✓			✓	k-center	[128]	[129], [130]
Distributional Individual Fairness	requires the statistical distance obtained using the f-divergence for the output distributions of each pair of samples is smaller than the distance obtained using a given F metric, which measures fairness between pairs of points		✓		✓	center-based clustering	[89]	-
Kleindessner et al Individual Fairness	ensures distance of a point from the cluster is not greater than the distance of a point from another cluster; goal to minimize		✓	✓		any	[87]	-
α-Pairwise Fairness	requires every pair of points has a probability of at most α of being assigned to different centers		✓		✓	k-center	[94]	-
β-Community Preserving Fairness	requires community points are assigned to as few different clusters as possible; an algorithm is β-community preserving if every community has probability at most β of being partitioned into more than a certain number of pre-defined clusters		✓		✓	k-center	[94]	-
Per-Point Fairness	requires the distance of a point from its center is at most α times the distance of the closest point in that cluster to the center		✓		✓	k-center	[90]	-
Aggregate Fairness	requires the distance of a point from its center is at most α times the average distance of the points in this cluster to the center		✓		✓	k-center	[90]	-
Max Fairness Cost (MFC)	uses user-inputted ideal proportion value and reduces the deviation between current and ideal proportions	✓		✓		any	[32]	-
Entropy	amount of unfairness in each cluster; goal to maximize	✓		✓		any	[111]	-
Essential Fairness	relaxed fairness notion; certain group in cluster can differ by 1 from groups' proportion in overall set	✓		✓		center-based clustering	[98]	-

with multiple protected groups present. Bera *et al.* [85] and Bercea *et al.* [98] provided approaches for more general clustering objectives that used upper and lower bound constraints on the proportion of protected group members in each cluster. The algorithm from [85] allowed for groups to overlap (for example, consider both race and gender) and they denote Δ as the number of protected groups samples can belong to simultaneously. They proposed a linear program based rounding approach that achieves a $c + 2$ approximation if the original clustering objective has a c approximation algorithm available, while incurring at most $4\Delta + 3$ additive violations to the upper and lower bound fairness constraints.

Specifically for k-center, [85] obtained a 5-approximation when centers need not be exemplars, and a 4-approximation when centers are exemplars. Harb and Shan [123] improved upon these fair k-center results of [85] by developing a faster 5-approximation algorithm for the non-exemplar case, and a better 3-approximation algorithm for the case with centers as exemplars. Jia *et al.* [120] proposed a 3-approximation algorithm for the k-center objective that allowed for multiple groups or colors. Esmaeili *et al.* [118] proposed approximation algorithms in the general setting where points are allowed to have uncertain protected group membership (that is, protected group memberships are provided as a distribution), and a sample in the dataset is assumed to only belong to one protected group at a time.

Liu and Vicente [114] introduced a stochastic approach that solves a bi-objective optimization problem and shows the trade-off between the k-means clustering objective and fairness. Their algorithm was only guaranteed to converge for smoothed problems. Esmaeili et al [126] generalized the clustering objective cost/fairness problem for k-center, k-median, and k-means and introduced new group-level fairness notions. They developed bi-criteria approximation algorithms for each notion.

Kleindessner *et al.* [128] proposed an approach to compute fair summaries for group-level fair clustering which uses k-center prototypes to summarize each group in a dataset. They provide a linear time approximation algorithm for this problem. Chiplunkar *et al.* [129] proposed improved distributed algorithms for the aforementioned fair summaries notion in the streaming setting. Jones *et al.* [130] proposed an algorithm that runs in linear time and yet achieves a 3-approximation for the fair k-center summaries problem.

Ghadiri *et al.* [33] introduced the socially fair notion which focuses on minimizing clustering cost across groups rather than constraining the proportion of protected groups in clusters. Concurrently to [33], Abbasi *et al.* [103] independently introduced a similar notion of group representation. Makarychev and Vakilian [104] presented a generalized bi-criteria approximation algorithm and generalized the socially fair clustering problem framework. Goyal and Jaiswal [124] developed an FPT time approximation algorithm for the socially fair notion. Thejaswi *et al.* [125] introduced a new notion of diversity-aware fairness, that requires each group have some minimum representation in the form of cluster centers, for the k-median objective.

Individual-Level Fairness: Chen *et al.* [31] introduced the individual level fairness notion of *proportionality* for k-center clustering that seeks to ensure points are treated equally, an important concern especially for facility placement. They showed that exact proportionally fair solutions might not always exist and provide an algorithm that achieves in the worst case a $1 + \sqrt{2}$ proportionally fair clustering solution. They also developed an approach that is $\mathcal{O}(1)$ proportionally fair and also a $\mathcal{O}(1)$ approximation for the k-medians objective of the optimal proportional fair solution. Micha and Shah [93] modified Chen's approach, developed a 2-approximation algorithm when the distance metric being used is the L2 norm, and proved the $1+\sqrt{2}$ factor was tight for other commonly used distance metrics such as the L1 norm and the L-infinity norm.

Jung *et al.* [88] introduced an individual level notion that determined a fair radius for clusters, as defined previously (Table 1), for center-based clustering objective. They developed an algorithm that achieved a 2-approximate fair k-center clustering, meaning that every point p has a center within a distance of $2r(p)$ where $r(x)$ is defined as in Table 1. Note from here on that we denote bi-criteria approximation results for the fairness notion and clustering objective using the $(.,.)$ notation. Mahabadi and Vakilian [34] confirmed Jung's results and generalized the problem, obtaining $(\mathcal{O}(1), \mathcal{O}(1))$ bi-criteria approximations for fair k-median and k-means clustering and a $(\mathcal{O}(1), \mathcal{O}(\log n))$ bi-criteria approximation for k-center. Vakilian and Yalçıner [92] improved upon the fair k-center case of [34] and improved the bi-criteria approximation from $(7, \mathcal{O}(\log n))$ to $(3, \mathcal{O}(1))$. Additionally, they provided improved bi-criteria approximations (compared to [34]) for the k-means and k-median objectives as well. Chakrabarty and Negahbani [91] also provided improved algorithms for individual fair clustering according to Jung et al's fair notion achieving an $(8, 8)$ and $(8, 4)$ bi-criteria approximations via linear program rounding for k-medians and k-means clustering respectively.

We also discuss some other work on center-based individually fair and group-level fair clustering that have recently been studied. Kleindessner *et al.* [87] introduced another individual fairness notion using a dissimilarity function that requires points be closer to points of their cluster than those of other clusters. Anderson *et al.* [89] developed fair algorithms that ensure distributional individual fairness so that similar individuals are clustered similarly. Brubach *et al.* [94] introduce two new individual fairness notions and present an algorithm for the k-means objective. More recently, Chakrabarti *et al.* [90] proposed an individual fairness notion that ensures points receive similar *quality of service* and provided algorithms for the k-center objective. Abraham *et al.* [127] introduced a fair k-means clustering algorithm for a new group-level fairness notion that is enforced at the in-processing stage of the clustering pipeline.

2) HIERARCHICAL CLUSTERING

Ahmadian *et al.* [105] and Chhabra and Mohapatra [32] concurrently proposed approaches for fair hierarchical clustering. However, both approaches have a number of different distinctions. Ahmadian *et al.* [105] proposed a fairlet decomposition approach for only (upper-bounded) bounded representation fairness, for a number of recently proposed hierarchical clustering objectives such as Dasgupta's cost [39], *value* [45], and *revenue* [46]. Due to fairlet decomposition their work constitutes a pre-processing approach. Chhabra and Mohapatra [32] on the other hand proposed an in-processing algorithm for heuristic greedy hierarchical clustering algorithms which can accommodate

any notion of fairness. Their work does not consider the newly proposed hierarchical clustering objectives such as [39] but instead focuses on traditional heuristic hierarchical agglomerative clustering used in practice. Quy *et al.* [117] utilized fairlet decomposition for making *capacitated* (clusters have some size constraints) clustering fair. They considered both hierarchical agglomerative (heuristic and greedy, similar to [32]) clustering and partition-based clustering algorithms to improve on fairness. Furthermore, as the capacitated clustering problem is relevant in an educational setting (clusters of students need both fair representation and approximately fixed sizes), they evaluate their approaches on data from school-going students.

3) SPECTRAL CLUSTERING

Kleindessner *et al.* [100] added fairness constraints (balance fairness notion) to normalized and unnormalized spectral clustering. They project the graph Laplacian onto a *fair* subspace and then perform k-means clustering on this subspace. They also gave analysis for their approach on a variant of the stochastic block model. Anagnostopoulos *et al.* [131], [132] extended the work of [100], to the densest subgraph problem.

4) DEEP CLUSTERING

The first work combining deep clustering with fairness was proposed by Wang and Davidson [102]; they introduced fairoids to represent each group and ensured centers are equally spaced from the fairoid via a discriminative deep clustering model. Fairoids allow for non-binary valued protected groups. Li *et al.* [111] developed a scalable, deep clustering model that used adversarial loss to constrain learning and ensure fairness while maintaining cluster quality. They were the first paper to use deep, fair clustering on visual datasets for visual learning. Zhang and Davidson [101] generalized the fairness constraints for deep clustering and developed a model that allowed for multiple protected groups and flexible constraints.

5) MISCELLANEOUS

Ziko *et al.* [115] developed a general variational bound-optimization framework for fair clustering. They introduce a fairness penalty term based on Kullback–Leibler (KL) divergence. The fairness penalty is used to measure and manage the trade-off between the clustering objective and fairness. Furthermore, their approach is scalable and works for large datasets.

For the graph-based correlation clustering objective, Ahmadian *et al.* [119] utilized the fairlet decomposition method. They achieve promising results for a number of different fairness constraints and find that by defining the fairlet decomposition similar to the k-median cost they obtain good approximations for fair correlation clustering.

Chhabra *et al.* [116] introduced the pre-clustering approach of adding *antidote* data points to the original dataset to improve group-level fairness. Antidote data points are dummy points that do not belong to a protected group, but when vanilla clustering is undertaken on the new dataset, the solution is more fair with respect to the original points. Their approach is general and can accommodate any fairness notions and clustering objectives. They also consider other problem settings for this work, such as in the case where clustering objectives and fairness notions are convex functions. The antidote data approach for fair clustering is visually described in Fig. 5.

While we restrict ourselves to the study of fairness in clustering algorithms, there are other related fields where fairness can be studied, such as link prediction in complex networks [133]–[136]. While an in-depth discussion of such approaches is outside the scope of this work, clustering is inherently connected to many other fields, where similar ideas of fairness can be applied.

B. PRE-PROCESSING, IN-PROCESSING, AND POST-PROCESSING APPROACHES

As mentioned before in Section II, fair approaches can be broadly classified depending on what stage of the learning pipeline the fairness is enforced in. In particular, for clustering, the same classification holds, albeit with some slight differences.

For pre-processing (or pre-clustering) based fair approaches, the fairness intervention occurs at the stage before the learning model is trained. In clustering, this means that the original dataset X is first pre-processed and then transformed to some dataset X'. When the vanilla clustering algorithm $\mathcal{A}$ is invoked on this transformed dataset, the resulting clusters obtained $\mathcal{C}_{\text{fair}}$ are fair. A schematic diagram explaining this process is shown in Fig. 6.

For in-processing (or in-clustering) based fair approaches, the fairness intervention happens as a result of changing the original learning model, to make it output only fair solutions. This is where a bulk of fair clustering approaches lie. Here, the clustering model/algorithm itself is modified from the vanilla clustering algorithm $\mathcal{A}$ to a fair clustering algorithm $\mathcal{A}'$ to make it incorporate fairness constraints in the fair solution $\mathcal{C}_{\text{fair}}$. The schematic demonstrating this is shown as Fig. 7.

Post-processing (or post-clustering) based fair approaches enforce the fairness approach after the learning model has computer initial unfair estimates. In clustering, this means that the fairness intervention occurs post the vanilla clustering process. The vanilla clustering algorithm $\mathcal{A}$ is run on the original dataset X to obtain unfair cluster solutions $\mathcal{C}$. The fairness approach then operates on $\mathcal{C}$ to obtain fair clustering outputs $\mathcal{C}_{\text{fair}}$. A lot of research works also fall into this category. The schematic explaining this is shown as Fig. 8.

We now discuss fair clustering research under this classification. Furthermore, Table 2 showcases this categorization for most of the major fair clustering papers.

1) PRE-PROCESSING APPROACHES

The concept of fairlet decomposition [30] which was used in the first work of fair clustering constitutes a pre-processing

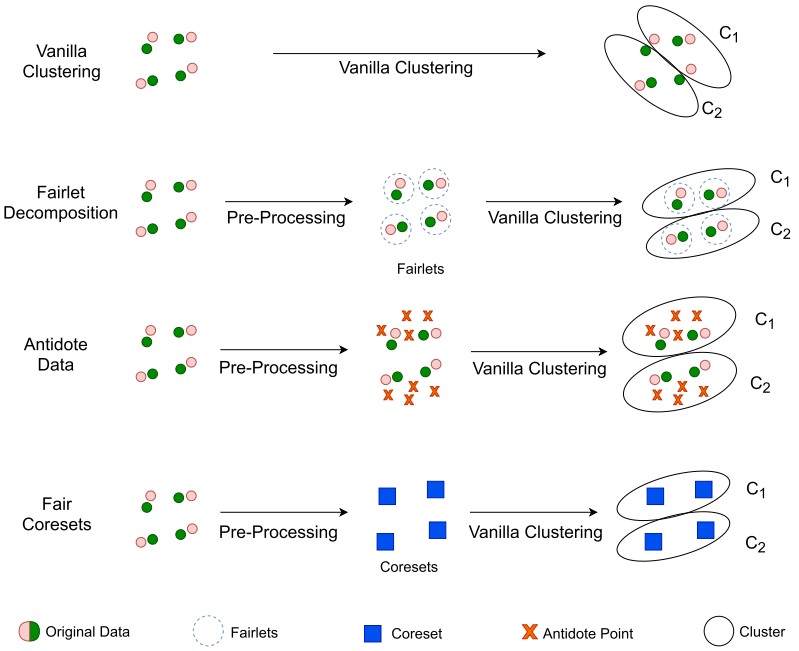

FIGURE 5. Pre-processing methods, including fairlet decomposition, fair coresets, and antidote data fair clustering approaches (green and pink colors represent protected groups).

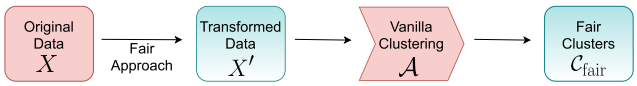

FIGURE 6. Diagram explaining pre-processing/pre-clustering fair approaches.

FIGURE 7. Diagram explaining in-processing/in-clustering fair approaches.

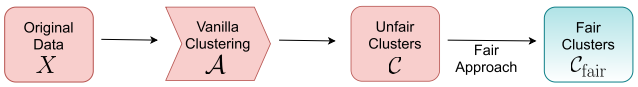

FIGURE 8. Diagram explaining post-processing/post-clustering fair approaches.

TABLE 2. Categorization of fair clustering approaches.

Classification	Significant Works
Pre-processing	[30], [99], [113], [105], [119], [97], [121], [122], [116]
In-processing	[114], [127], [115], [100], [125], [111], [101], [31], [93], [88], [91], [89], [90], [94], [102]
Post-processing	[128], [129], [130], [113], [85], [98], [86], [123], [118], [126], [34], [92], [87]

based approach. As discussed before, fairlet decomposition aims to find fairlets (or micro-clusters) within the data that meet fairness requirements. Vanilla clustering is then employed on this data leading to fair solutions. Many fair clustering works that expand upon or utilize fairlets fall into the pre-processing category: [99], [105], [113], [119]. Fair coresets are also fair representations of the dataset, that summarize the data points to ensure fairness in a more scalable manner. Introduced by [97], fair coresets were used in [121] and [122]. The antidote data approach for fair clustering [116] described before is also relevant here as it is pre-processing and augmenting the original dataset. Diagrams explaining these different pre-processing based approaches in a high-level manner are shown as part of Fig. 5.

2) IN-PROCESSING APPROACHES

In-processing approaches to fair clustering involve altering the clustering objective and algorithm itself. Often the fair algorithm optimizes between the clustering cost and fairness trade-off. Papers such as [114], [127], and [115] augmented the original algorithms with functions that measured and controlled the trade-off between fairness and clustering performance. In [100], the authors similarly adjust the spectral clustering objective to solve a minimization problem that incorporates fairness constraints. In [125], the authors developed a k-median algorithm specifically for diversity-aware fairness. In papers, [101], [102], [111] the authors constrained the deep clustering process itself, optimizing the trade-off between cluster quality and fairness through joint optimization, adversarial learning or other similar approaches.

The works by [31] and [93] also alter the clustering algorithm objectives to find individual-level proportionally fair solutions. Finally, the papers [88]–[91], and [94] also redefine the clustering objectives to make them individually fair according to the fairness notions first proposed by Jung *et al.* [88].

3) POST-PROCESSING APPROACHES

Post-processing involves modifying the clustering outputs to be fair. A vanilla clustering algorithm is first employed, and either a fair problem is separately solved or the vanilla output adjusted depending on the fairness notion. The clustering algorithm itself does not jointly optimize for the clustering cost and fairness objective, unlike methods for in-processing. Examples of post-processing approaches include those used for fair k-centers summaries– these post-process clustering centers such that every group is represented through centers equitably. This line of work was first introduced by [128] and later extended by [129] and [130]. The authors in [113] use an algorithm to maintaining fairness and privacy subsequent to first finding a non-private solution using vanilla clustering algorithms, also constituting a post-processing approach. Similarly, [85], [86], [98], [123], [34], [92], [118], [126], and [87] solve the vanilla clustering problem first and then improve fairness by proposing algorithms that change cluster assignments for points. Hence, these also constitute post-processing based fair clustering approaches.

Another post-clustering based work was by [137]. Here, the authors take as input the cluster output from a vanilla clustering algorithm, and compute a clustering close to the original, but one that meets fairness requirements. They formulate the problem as an integer linear program, and also provide theoretical results on hardness.

V. EVALUATING FAIR CLUSTERING

In this section, we discuss the aspects of fair clustering research along two facets– the datasets that are generally used for evaluation, as well as the motivations for some real-world applications. The goal here is to allow researchers to select suitable datasets for evaluation based on prior research, and also provide them with real-world use cases. These real-world scenarios can then be used for motivating theoretical problems in fair clustering, or undertaking fair clustering research with a more practical flavor.

A. DATASETS USED FOR EVALUATION

The approaches discussed in Section IV propose different methods of creating fair clustering models using different notions. The next phase is to evaluate the approach by applying it to actual data. Datasets used vary widely from paper to paper depending on the notion and overall goal, but some datasets are used more frequently than others and can be used to compare between approaches.

To serve as a guide for researchers new to the field, the datasets used in over 40 papers on fair clustering were collected in Table 3 (classical clustering approaches) and Table 4 (deep clustering models). The most common datasets used for traditional fair clustering are listed at the top of Table 3: adult [138], bank [139], creditcard [82], diabetes [140], and census [141], all of which are large datasets from the UCI ML repository [167]. Table 3 includes the name and label of each dataset, a short description, and the source paper. Most other datasets can also be found on the UCI repository. Further, the possible protected groups (such as ethnicity), that have been used in the surveyed papers are listed as well along with the dataset size. We term a dataset with over 10,000 instances as *large*. Note that some papers opt to use subsets of the datasets since their algorithms do not scale well or the running time is too long such as [30], [85], [113], [118]. For completeness, we also list all papers surveyed that use a certain dataset in the last column of the aforementioned tables.

Datasets are sometimes chosen specifically for the approach and fairness notion being proposed. For example, [103] uses North Carolina Voter information when proposing their group representation notion for facility location. Other datasets such as bank [139] and creditcard [82], with common protected groups being marital status and gender, also have fairly clear connections to the motivations behind fair clustering. Other datasets, such as iris [145], are less directly connected but can still serve as toy datasets for experimentation. We also find that the most common protected groups are gender and sex, and race. Datasets listed without specific protected groups are used in papers enforcing individual-Level fairness notions and therefore did not require a specific protected group.

Visual datasets are often used for deep clustering; these are listed in Table 4. Deep clustering, as mentioned in Section IV, differ from more traditional approaches and can learn more powerful representations. In Table 4, the datasets are described and the protected group is listed.

B. REAL-WORLD APPLICATIONS

Machine learning models have been used to assist in a vast majority of decision-making and risk assessment processes, from college admissions to online recommendation systems. For further information on the topic, Makhlouf *et al.* [80] in their paper discuss general applications of ML in decision-making processes and some existing programs where fairness should be considered. Suresh and Guttag [168] additionally show how ML models can have unintentional, damaging consequences if bias is not considered throughout the ML pipeline. Thus, in this section, real-world applications for fair clustering ML models are used to motivate further research in the field.

1) BANK LOAN DISBURSEMENT

We described a similar scenario previously in Section III for group-level notions. Clustering based models can be used to determine individuals who should receive a loan based on how likely they are to default on it. Many factors can play a role, and are often considered before disbursement, such as an applicant's education history, past payment history, past billing statements, amount of the bill paid, and age. Members of certain minority protected groups, such as women or POC, might have lower incomes due to systemic issues such as the wage gap. Furthermore, married persons

TABLE 3. Datasets used to evaluate classical fair clustering algorithms.

Dataset	Description	Source	Possible Protected Group	Size	Used In
Adult Data Set Census 1994 (Adult)	infomation from 1994 census database	[138]	sex, race, age	large	[99], [30], [85], [97], [103], [112], [105], [32], [33], [34], [89], [118], [114], [115], [126], [127], [121], [90], [91], [123], [102], [94], [116], [101], [87], [128], [130], [129]
Bank Marketing (Bank)	maketing campaign from Portugese banking institution	[139]	marital status, default, education, housing	large	[99], [30], [85], [97], [103], [112], [105], [32], [34], [89], [118], [114], [115], [121], [101], [116], [90], [123], [91]
Default of Credit Card Clients (Creditcard)	Taiwan bank customers' info and default	[82]	marital status, education, gender	large	[85], [112], [32], [33], [118], [126], [116], [90], [123], [101]
Diabetes- US 130 (Diabetes)	Diabetes data from 130 US hospitals for years 1999-2008	[140]	gender, age, race	large	[99], [30], [97], [112], [34], [89], [112], [121], [91]
US Census Data 1990 (Census)	information from 1990 census	[141]	sex, age, marital status	large	[99], [30], [85], [112], [89], [118], [115], [126], [121]
Reuter 50-50 Data Set (Reuters)	50 english texts from 50 authors	[142]	authors	small	[86], [119], [123]
Victorian Era Authorship Attribution (Victorian)	texts from 50 victorian authors	[143]	authors	large	[86], [119], [123]
120 years of Olympic history: athletes and results (Athletes)	120 years of Olympics athlete and result data	[144]	sex, sport, medal	large	[112], [121]
Iris	petal dimensions of 3 iris flower species	[145]	species	small	[31], [103]
4area	computer science researchers and their areas of study	[146]	main area of research	large	[86], [123]
KDD	sequences of TCP packets info from 1999	[147]	-	large	[31]
Pima Indians Diabetes	Pima Indians Diabetes data set about 768 diabetes patients	[148]	-	small	[31]
North Carolina Voters	voter information of 15+ years from North Carolina counties	[149]	race	large	[103]
Query	points represent bag of queries from online auction environment, private dataset from Google	Proprietary	ID of main advertiser	-	[86]
Statlog (German Credit)	1000 peoples' information and credit (good or bad)	[150]	-	small	[87]
Drug Consumption	1885 respondants' attributes and drug usage	[151]	-	small	[87]
Indian Liver Patient (ILPD)	liver and non liver patient records for 583 patients	[152]	-	small	[87]
FriendshipNet	friendships between groups of high school students	[153]	gender	small	[100]
FacebookNet	Facebook friendships between high school students	[153]	gender	small	[100]
DrugNet	network data encoding connections between drug users in Hartford, Connecticut	[154]	sex, ethnicity	small	[100]
Kinematics Word Problems (Kinematics)	161 kinematics word problems of five different types/difficulties	[155]	problem type	small	[127]
p-Median	40 test problems about solving p-median problems	[156]	-	small	[94]
Student Performance	information of two secondary education Portuguese schools' students	[157]	sex	small	[117], [130]
PISA test scores	information about students in America taking the PISA exam	[158]	gender	small	[117]
Open University Learning Analytics (OULAD)	information about students enrolled in seven courses at Open University in England	[159]	gender	small	[117]
MOOC	data of individuals enrolled in the MOOCs offered on edX by Harvard and MIT	[160]	gender	small	[117]
Labeled Faces in the Wild (LFW)	more than 13000 pictures of named faces	[161]	sex	large	[33], [116]
Web data: Amazon reviews (Amazon)	18 years worth of reviews on Amazon	[162]	item category	large	[119]

TABLE 4. Datasets used to evaluate deep fair clustering models.

Data Set	Description	Source	Possible Protected Group	Used In
Human Activity Recognition (HAR)	10299 instances with captured action features for 30 participants	[163]	participant identities	[102], [101]
MNIST-USPS	handwriting samples	[111]	source of sample	[111]
Color Reverse MNIST	color reverse of MNIST handwriting samples	[111]	original or reversed color image	[111]
Multi-task Facial Landmark (MTFL)	12,995 images of faces for facial recognition including facial landmarks, labels gender glasses smiling and head pose	[164]	with or without glasses, gender	[111]
Office-31	images of common office objects from three domain sources including different angles, lighting, backgrounds	[165]	domain source (amazon, webcam)	[111]
Daily and Sports Activity (D&S)	9,120 sensor records of human daily and sports activities	[166]	participant names	[102]

might have better credit than single persons. Vanilla clustering algorithms being used for shortlisting candidates for disbursement in an unsupervised manner that do not correct for the different sorts of bias present in data will likely cluster single people, women, and POC as higher risk and as more likely to default on their loan. Such predictions might result in fewer loans, or loans with higher interest rates, being given to protected groups, further promoting the systemic issues at hand. A well designed, fair clustering algorithm could correct for the disparate impact by requiring balance or a bounded representation, that more or less fix the proportion of protected groups in each cluster.

2) JOB SHORTLISTING

Many ML based approaches exist that parse through job candidates in order to shortlist those who should be interviewed or move onto the next application step [9]. Automating this step can reduce errors, human bias, time spent parsing applications, and allow for easy comparison between candidates [13]. Clustering algorithms can separate between accepted and rejected candidates for shortlisting based on their skill sets and other attributes, and how well they match the job requirements. Common candidate attributes include education, major, experience, skills, current location, current employment status, age, gender, etc [169]. Clustering algorithms that do not account for bias might reject POC or women and accept less qualified white men [80]. A fair clustering algorithm that requires for example, balance, for the sensitive group gender would fix the proportion of women in each cluster, assigning top qualified women from the rejected cluster to the accepted one to account for the bias. The company benefits by seeing more qualified individuals, and the applicants are not discriminated against by being rejected based on their inherent attributes.

3) COLLEGE ADMISSIONS

Clustering based ML models can be used to shortlist candidates for admission, remove definite rejects for college applications, or select those most likely to attend. Attributes considered might include GPA, leadership roles, parents' education levels, and general student information. Algorithms with unchecked bias might reject candidates based on factors that are unrelated to the candidate's ability, such as their street address [80], which can correlate to other attributes such as their socioeconomic background or race. Fair clustering algorithms that ensure individual-level fairness (Section III) could prevent individuals with approximately similar grades or leadership roles from being clustered differently based on unrelated attributes such as ethnicity.

4) FACILITY LOCATION

ML models can assist in facility location, for example in helping determine voting/polling booths, or hospital locations. As previously mentioned in Section III, regular clustering models that only consider the number of homes in an area might unfairly distribute facilities among suburban, urban, and rural areas. Fair clustering models should take into account the conditions of an area by considering other constraints, such as proportionality. Depending on the facility purpose, proportionality could ensure facilities are equally serviced [31]. Another notion, group representation, could ensure cluster centers/ facilities are fairly placed such that the centers are representative of the clusters, or each area gets its own center [103]. This could play a role in ensuring polling centers are serviced similarly and are a reasonable distance from a majority of sample locations.

5) PRISONER RECIDIVISM

ML models have been used to predict the risk/likelihood of ex-convicts re-offending to offset human bias on factors such as race [170]. Prisoner recidivism can be interpreted as a probability and could be determined by a soft clustering algorithm, in which a point can be assigned a certain proportion of each cluster– with clusters signifying either being at high risk of re-offending, or low risk. A number of factors can assist in predicting recidivism, including age and number of prior convictions [170]. However, as has been found with the COMPAS tool [18], since data used to train such algorithms might be systemically biased, the learning model could amplify bias against POC based solely on their race [79]. In such a case, a well-designed fair clustering algorithm that ensures individual fairness– that similar individuals (in terms of crimes committed and other attributes) are clustered similarly regardless of sex or race [89]– would prevent minority

protected groups members from being assigned disproportionately higher risk rates compared to non-group members with similar crime statistics.

6) RECOMMENDATION SYSTEMS

Clustering based recommendation systems have been used for many purposes, from movie recommendation [171] to distance learning course recommendations [172]. As clustering algorithms can be biased due to the data, these recommendation systems can also be biased. This could mean, for instance, giving skewed recommendations to men over women [80]. As a result, recommendation systems should be personalized for individuals, and should not be explicitly biased towards gender or ethnicity. A clustering algorithm that ensures some level of individual-level fairness could prevent certain groups from automatically receiving certain recommendations regardless of their other attributes.

7) COMMITTEE SELECTION

A final example, also presented in [125], is selecting committees that represent each group in a population. Committees are built within various communities for political, educational, fundraising, among other purposes. The goal might be to have a committee with at least one representative of each group, or have a diverse committee where every group is well represented by multiple members. A fair clustering algorithm could ensure protected groups are well represented, irrespective of individuals' ethnicity or political bias, using notions such as diversity-aware fairness [125], group representation [103], or fair summaries [128].

VI. FUTURE RESEARCH DIRECTIONS AND OPEN CHALLENGES

A. CONSIDERING ALTERNATIVE CLUSTERING OBJECTIVES

As we have seen throughout the article, and especially in Section IV, most research on fair clustering considers center-based clustering algorithms (such as k-center, k-medians, etc), and a few consider hierarchical clustering objectives and spectral clustering. However, there are a number of other clustering algorithms and objectives that have not been considered from a fairness perspective. We provide directions for research in this regard with respect to density-based clustering approaches and soft clustering methods. Furthermore, as in-clustering approaches are more popular, we consider those for this first approach.

1) DENSITY-BASED CLUSTERING

Density-based clustering algorithms use the concept of *density* or how close points are to each other in space to assign points to clusters and label points in low-density regions as *noisy* points or outliers. There are a number of different approaches that seek to perform density-based clustering, such as DBSCAN [47] and OPTICS [48]. For this task, as a first step, popular algorithms such as DBSCAN [47] and OPTICS clustering [48] could be considered. Further on, research frameworks can be extended to other density-based clustering approaches such as PreDeCon [173] and SUBCLU [174] since these share similarities with the DBSCAN approach.

In general, one can consider the following in-clustering approach to improving fairness for these clustering algorithms. First, identify a clustering objective based on the characteristics of the algorithm and application scenario. This objective allows one to eventually provide theoretical guarantees of fairness. Next, decide on how the fairness constraint is enforced, depending on the suitability to an application scenario. For example, if balance is being considered, one can consider lower bounding or upper bounding balance; if a proportion of points is being considered, bounded representation can be considered. Then, approximation algorithms can be proposed which approximate the objective. The approximation ratio obtained is the cost that the fair approximate algorithm achieves on the objective compared to the optimal value of the objective. It can also be gauged as to how much distortion is present in the fair assignment of points as compared to the original objective. Lastly, evaluation of the proposed approach using real-world datasets (as discussed in Section V) can be undertaken and fairness improvements can be analyzed.

There are other prospective research challenges associated with this problem. As most research so far has looked at center-based clustering, it is probable that fairness definitions are also designed accordingly. Thus, depending on the clustering algorithm being analyzed, alternate fairness notions can be developed and studied. For example, DBSCAN labels certain points as outliers (called noisy points) while clustering, and this might require differing notions of fairness as certain points are not being *represented* by the clustering algorithm at all now. Another prospective research direction can be to study multiple assignments to protected groups for data points. As a first step, the 2 groups case can be studied as in the seminal work of [30]. Future work can then include multiple groups, with points being assigned disjointly to each protected group. Subsequently, settings where points can be assigned to multiple protected groups at the same time can be analyzed. Finally, improvements can also be made in terms of *running-times*– while naive first approaches to providing fairness for the aforementioned clustering algorithms can have longer running times, for any practical implementation, it would be required to improve the asymptotic time complexity of their fair variants.

2) SOFT CLUSTERING

As mentioned before, much discussion and existing work have focused on hard-clustering algorithms where a data point belongs to a cluster in a binary fashion. That is, it either belongs to a cluster or it does not. However, in certain application scenarios, soft clustering is more suitable. Gaussian-mixture models [109] have been widely used in such cases, and thus could be the preliminary focus of this research direction. To estimate clustering results in a Gaussian-mixture model, an expectation–maximization (EM) algorithm [175]

is often used. EM is an iterative method to find (local) maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables. Therefore, a new research direction involves studying the fairness of such algorithms. A first approach and initial objectives could be similar to that discussed in the previous subsection on density-based clustering. One key issue is to redefine fairness in the presence of soft-clustering to reflect its probabilistic nature.

B. IMPROVED CLUSTERING PERFORMANCE ANALYSIS

Fair clustering approaches aim to improve fairness for clustering objectives by changing cluster assignments for samples in the dataset. It is well known that clustering performance is degraded as a result of improving fairness [118], [176], [177], as changing point labels to improve fairness can be contradictory with the original cluster assignments, leading to worse clustering performance. While this trade-off is well acknowledged, there is currently no standardized approach to measuring clustering performance.

Most research works measure the drop in the clustering objective over the vanilla (original/unfair) clustering objective [30], [85], [105]. However, measuring performance in this way might not be suitable in some case scenarios. Consider the following examples:

- *When Algorithm Agnostic Fairness Notions Are Used for Different Clustering Objectives:* If algorithm agnostic notions are used, but the clustering objectives are different, directly observing the values of the clustering objective after fairness enforcement would not lead to a sound comparison. For example, comparing a fair k-center cost with a fair k-means cost would not make sense. This scenario can arise when more general fair clustering approaches are being employed as in [116].
- *When Clustering Objectives Are Not Well-Defined:* This can be understood through the context of hierarchical clustering. Although recently clustering objectives for hierarchical clustering have been proposed, traditionally hierarchical clustering has been a heuristic agglomerative/divisive procedure and does not have an analytical objective to optimize. Thus, research aimed at making traditional hierarchical clustering fair [32] would not have a clustering objective which can measure the quality of the fair solution, in terms of clustering performance.

Alternatively, traditional clustering performance indicators could be used to measure clustering quality after fairness enforcement. These include the widely utilized Silhouette score [178], Calinski-Harabasz index [179], or the Davies-Bouldin index [180]. These have also been employed as a measure of clustering performance after fairness intervention in some fair clustering works [32], [116], [127]. The Silhouette score is especially appealing since it is bounded and always outputs a value between -1 and 1, making it easy to interpret. However, these metrics also have certain drawbacks– they work well only in the case with convex clusters, and might not be good indicators of performance in other case scenarios. Therefore, a future research direction for fair clustering is to investigate and propose new metrics for clustering performance specifically in the context of fairness. This would also connect the field of fair clustering with the long-standing sub-field of research on measuring clustering performance.

C. ADVERSARIAL ATTACKS AGAINST FAIRNESS

This direction for future work primarily deals with adversarial attacks on clustering algorithms that aim to degrade the fairness of a given clustering. As more and more research attempts to make clustering fair, the converse of problem in clustering also holds true. Malicious entities can seek to disrupt fairness for their personal gains and agendas. As a starting point for investigating this, it would be useful to leverage work on data poisoning for clustering in a black-box setting [181], [182]. Without changing the attack objective, the attack first proposed in [32] is especially powerful because it can be carried out without knowing the original clustering algorithm.

We can delineate a first approach for degrading fairness using the attack algorithm of [181] and for the fairness notion of *bounded representation* [86]. Let the clustering algorithm be k-means where $k = 2$. Here, for ensuring fairness each protected group's members in a cluster need to be within some minimum and maximum pre-specified proportion. In [181] details adversarial attacks where the target of the adversary is to lead to *spill-over* of as many points from one cluster to another. Thus, in the 2-way clustering setting, since this attack algorithm can change the proportion of points that belong to each cluster, we can effectively *skew* the chosen fairness metric for the outputted clustering. We defer interested readers to [181] for more details on the attack algorithm and threat model.

Subsequent to this, there are many possible directions along which fairness degrading adversarial attacks can be extended:

- *Black-Box Attacks:* Black-box attacks on clustering algorithms that disrupt fairness of the obtained clustering can be investigated. Since these are black-box attacks, the attack is powerful as it works irrespective of the choice of clustering algorithm used by the defender.
- *White-Box Attacks:* White-box attacks specific to the clustering algorithm (or its fair variant) chosen by the defender can also be investigated.
- *Other Attack Modalities and Threat Models:* Attacks when other attack modalities are considered, such as imperfect knowledge of the dataset, grey-box attacks, different fairness definitions that can be disrupted, and alternate/enhanced attack objectives, as well as costs to the adversary can also be analyzed.
- *Transferability and Other Fairness Notions:* Like in supervised learning [183], analysis can be undertaken

to observe if generated adversarial samples are transferable across algorithms, fairness definitions, and attack settings.

D. MORE APPROACHES FOR DEEP CLUSTERING

Deep clustering is the combination of deep learning paradigms to the classical clustering approaches in unsupervised learning. The approaches used are different from traditional clustering, and usually require the existence of labels in the testing phase to evaluate the deep learning models using metrics such as the Normalized Mutual Information (NMI) score [184]. In case labels are available for the ground-truth clusters, deep clustering has been shown to achieve state-of-the-art performance when compared with traditional clustering approaches such as k-means [185]. Thus, it is important to ensure fairness for these models as well, similar to traditional clustering approaches.

However, as covered in Section IV, not much research has been undertaken in this regard. To the best of our knowledge, there are only three research works covering deep fair clustering: [101], [102], [111]. Thus, an important direction for future work is to study deep clustering from a fairness perspective. Many aspects of future work exist, similar to how fairness has been studied for traditional clustering approaches.

E. ASSESSING PERCEIVED FAIRNESS

For fairness improvements in clustering with significant social impact, the evaluation stage needs to be improved to account for *perceived fairness by protected groups and individuals.* Clearly, while one may develop fair algorithms for ML based on relevant fairness costs and definitions, a fair algorithm is only beneficial if it impacts the affected community in a positive social sense. To this end, there is a lot of potential for significant research work to gauge how fair proposed algorithms are in terms of public perception. Such experiments can be carried out with special focus groups where individuals and groups (based on the protected attributes of the application at hand) directly impacted by applications where clustering algorithms are used can provide guidelines for improvement. Based on this feedback, better fairness definitions can also be proposed that are socially and practically relevant. Prior research in clustering fairness has not considered evaluation of this form, and therefore, using minority groups' feedback as an evaluation metric will lead to fairer systems along with considerable research novelty.

Another related dimension to actual perceived fairness in clustering are the *datasets* being used. Along with identifying application domains where fair clustering needs to be implemented, it is also important to obtain real-world datasets which might lead to eventual unfairness in clustering. This is important for a number of reasons: 1) obtaining empirical results for proposed algorithms on actual real-world datasets can shed light on how these algorithms perform in actual application scenarios and not on synthetic ones, and 2) doing so opens up an opportunity to understand how biases might creep in the datasets in the first place, which could lead to the development of more fair algorithms, and better fairness definitions. To do this, datasets can be obtained from actual recruiting agencies, or from universities' admission processes, and can then be used to gauge if proposed fair algorithms provide fairer results. The analytical models and algorithms can then be tuned so that they are being leveraged to induce more fairness into such real-world applications (such as admission/selection processes).

As mentioned before, perception of algorithmic fairness is an important metric for evaluation. Thus, an evaluation plan and methodology for fair clustering research should involve conducting regular meetings and focus groups. Here, proposed fair algorithms will be utilized in real-time, and minority and affected political groups will give their observations and feedback regarding its fairness. For example, users belonging to certain protected groups can be shown how the vanilla clustering algorithm performs, and then how the fair variant performs. While the fair algorithm might be better, it might still not be at an acceptable standard in terms of actual protected group members' expectations. Such constructive feedback could aid in building actual tools and algorithms that are useful to the community as a whole, and provide some real social significance. There is also a lot of scope in borrowing from similar efforts that assess perceived fairness in algorithmic decision-making systems such as [186].

F. HANDLING HIGH-DIMENSIONAL DATASETS VIA SCALABLE FAIR CLUSTERING

In general, similar to other data analysis techniques, clustering algorithms also suffer from the *curse of dimensionality* [187], and tend to perform poorly on high-dimensional datasets [188]. Moreover, the first approach for fair clustering proposed by Chierichetti *et al.* [30] was also not scalable, and could only be applied to small sized datasets. This was due to the first step involving fairlet decomposition, which has a super-quadratic running time.

While research extending this work has attempted to make fair clustering scalable, there are still many shortcomings. For example, Backurs et al [99] proposed a scalable algorithm for fairlet decomposition which runs in (almost) linear time, however, this approach is only applicable for the case with 2 protected groups. This trend is also prevalent in other fair clustering approaches proposed for other clustering algorithms. For example, the fair spectral clustering algorithms proposed by [100] do not scale well with dataset size and dimension, and even for the more general antidote data fair clustering approach [116] the authors noted that a major limitation of their work is the running time of their algorithms when applied to high-dimensional/large-scale data.

Thus, a possible future direction for research in fair clustering can aim to make the proposed fair algorithms scalable, and allow them to handle high-dimensional data. Clustering algorithms capable of handling high-dimensional data have been extensively studied in the literature [188], [189], and future research can aim to apply these techniques to the field

of fair clustering. Researchers can also aim to augment existing fair clustering approaches so as to make them scalable.

G. RELATING FAIR CLUSTERING TO CONSTRAINED CLUSTERING

The problem of constrained clustering tackles the case when additional information is known about the clustering problem, and can be used to improve the discovery of clusters [190]. This scenario arises in real-world problems where *domain specialists* can provide additional side information to aid the clustering process. In the simplest case, this can then be translated into a traditional clustering problem where we wish to impose some *instance-level* constraints on the original clustering problem [191]. While many different forms of constraints can be formulated for different clustering algorithms, we consider *must-link* and *cannot-link* constraints to motivate the connections between fair clustering and constrained clustering.

Consider individual-level fairness and assume there exists an *unbiased* domain specialist who knows that certain samples in the dataset need to belong to the same clusters (for example, a recruiter who interviewed candidates and found them to be equally suitable for a position, irrespective of their protected group attributes, such as gender or ethnicity). Conversely, the domain specialist can also provide side information indicating that two samples should not belong to the same cluster (considering the previous example, the recruiter knows that one candidate performed well in the interviews and the other did not, irrespective of their protected group memberships). Such side information about the data samples can be trivially encoded as must-link and cannot-link pairwise constraints between data samples. Then, if candidates are being shortlisted using a clustering algorithm such as k-means (similar to the job shortlisting examples considered in Section III and Section V), these must-link and cannot-link constraints can be provided as input (along with the original dataset) to a constrained k-means algorithm such as PC-KMeans [192] or COP-KMEANS [52] to enforce individual-level fairness.

In a similar fashion, even other fairness constraints (such as those enforcing group-level fairness) can be encoded with the assistance of a domain specialist. These can then be used to meet the fairness criteria using existing constrained clustering algorithms. As a future research direction, we then aim to motivate studying fair clustering from the perspective of constrained clustering, which has been extensively studied in previous work. Another important research contribution could be to provide theoretical insights into when fair clustering problems can be translated into constrained clustering problems, and the different types of constraints and fairness notions that can be used to do so.

VII. CONCLUSION

In this work, we provided the first survey on fair clustering. Initially, we discuss the relevant details regarding clustering and fairness in machine learning (Section II). Then we categorize different fairness notions used in making clustering fair (Section III) and propose intuitive classification methodologies for the same. We also organize current fair clustering literature into many sub-categories (Section IV) and provide a comprehensive overview of the field as a result. We also detail many new insights and describe possible directions for future work (Section VI). Our goal through this survey article is to add to the existing body of work on fair clustering by providing a concentrated introduction to the field, which serves useful for both researchers and industry practitioners alike.

REFERENCES

[1] R. Berk, "Accuracy and fairness for juvenile justice risk assessments," *J. Empirical Legal Stud.*, vol. 16, no. 1, pp. 175–194, Mar. 2019.
[2] R. Berk, H. Heidari, S. Jabbari, M. Kearns, and A. Roth, "Fairness in criminal justice risk assessments: The state of the art," *Sociol. Methods Res.*, vol. 50, no. 1, 2018, Art. no. 0049124118782533.
[3] A. G. Ferguson, "Big data and predictive reasonable suspicion," *U. Pa. L. Rev.*, vol. 163, no. 2, p. 327, Jan. 2015.
[4] J. B. Biddle, "On predicting recidivism: Epistemic risk, tradeoffs, and values in machine learning," *Can. J. Philosophy*, vol. 5, pp. 1–21, Jul. 2020.
[5] M. Ghasemi, D. Anvari, M. Atapour, J. S. Wormith, K. C. Stockdale, and R. J. Spiteri, "The application of machine learning to a general risk–need assessment instrument in the prediction of criminal recidivism," *Criminal Justice Behav.*, vol. 48, no. 4, pp. 518–538, Apr. 2021.
[6] J. Jagtiani and C. Lemieux, "The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform," *Financial Manage.*, vol. 48, no. 4, pp. 1009–1029, Dec. 2019.
[7] C.-F. Tsai and M.-L. Chen, "Credit rating by hybrid machine learning techniques," *Appl. Soft Comput.*, vol. 10, no. 2, pp. 374–380, Mar. 2010.
[8] W.-Y. Lin, Y.-H. Hu, and C.-F. Tsai, "Machine learning in financial crisis prediction: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 4, pp. 421–436, Jul. 2011.
[9] M. Raghavan, S. Barocas, J. Kleinberg, and K. Levy, "Mitigating bias in algorithmic hiring: Evaluating claims and practices," in *Proc. Conf. Fairness, Accountability, Transparency*, Jan. 2020, pp. 469–481.
[10] E. van den Broek, A. Sergeeva, and M. Huysman, "Hiring algorithms: An ethnography of fairness in practice," in *Proc. 40th Int. Conf. Inf. Syst. (ICIS)*, Association for Information Systems, 2020, pp. 1–9. [Online]. Available: https://scholar.googleusercontent.com/scholar.bib?q=info:iRYy5_aHqfwJ:scholar.google.com/&output=citation&scisdr=CgVB0EgzEMjY9cHyp6U:AAGBfm0AAAAAYUj0v6UQ2Oo5_tM2Mx_6PzjdCMu0Wwlr&scisig=AAGBfm0AAAAAYUj0v78DD2dYaKyGkvbO-vh7Mc5evcOk&scisf=4&ct=citation&cd=-1&hl=en&scfhb=1
[11] P. K. Roy, S. S. Chowdhary, and R. Bhatia, "A machine learning approach for automation of resume recommendation system," *Proc. Comput. Sci.*, vol. 167, pp. 2318–2327, Jan. 2020.
[12] J. L. F. M. Pombo, "Landing on the right job: A machine learning approach to match candidates with jobs applying semantic embeddings," Ph.D. dissertation, NOVA Inf. Manage. School, NOVA Univ. Lisbon, Lisbon, Portugal, 2019.
[13] G. K. Palshikar, R. Srivastava, M. Shah, and S. Pawar, "Automatic shortlisting of candidates in recruitment," in *Proc. ProfS/KG4IR/Data, Search (SIGIR)*, 2018, pp. 1–7.
[14] A. Waters and R. Miikkulainen, "GRADE: Machine learning support for graduate admissions," *AI Mag.*, vol. 35, no. 1, p. 64, Mar. 2014.
[15] N. Gupta, A. Sawhney, and D. Roth, "Will i get in? Modeling the graduate admission process for American universities," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 631–638.
[16] A. AlGhamdi, A. Barsheed, H. AlMshjary, and H. AlGhamdi, "A machine learning approach for graduate admission prediction," in *Proc. 2nd Int. Conf. Image, Video Signal Process.*, Mar. 2020, pp. 155–158.
[17] S. Caton and C. Haas, "Fairness in machine learning: A survey," 2020, *arXiv:2010.04053*. [Online]. Available: http://arxiv.org/abs/2010.04053
[18] A. W. Flores, K. Bechtel, and C. T. Lowenkamp, "False positives, false negatives, and false analyses: A rejoinder to machine bias: There's software used across the country to predict future criminals. and it's biased against blacks," *Fed. Probation*, vol. 80, no. 2, p. 38, Sep. 2016.

[19] B. Green, "'Fair'Risk assessments: A precarious approach for criminal justice reform," in *Proc. 5th Workshop Fairness, Accountability, Transparency Mach. Learn.*, 2018, pp. 1–5.

[20] S. Corbett-Davies and S. Goel, "The measure and mismeasure of fairness: A critical review of fair machine learning," 2018, *arXiv:1808.00023*. [Online]. Available: http://arxiv.org/abs/1808.00023

[21] L. T. Liu, S. Dean, E. Rolf, M. Simchowitz, and M. Hardt, "Delayed impact of fair machine learning," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 3150–3158.

[22] C. Dwork, N. Immorlica, A. T. Kalai, and M. Leiserson, "Decoupled classifiers for group-fair and efficient machine learning," in *Proc. Conf. Fairness, Accountability Transparency*, 2018, pp. 119–133.

[23] R. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork, "Learning fair representations," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 325–333.

[24] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," *Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 3315–3323, Dec. 2016.

[25] M. B. Zafar, I. Valera, M. Gomez Rodriguez, and K. P. Gummadi, "Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 1171–1180.

[26] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," in *Proc. 3rd Innov. Theor. Comput. Sci. Conf. (ITCS)*, 2012, pp. 214–226.

[27] R. Berk, H. Heidari, S. Jabbari, M. Joseph, M. Kearns, J. Morgenstern, S. Neel, and A. Roth, "A convex framework for fair regression," 2017, *arXiv:1706.02409*. [Online]. Available: http://arxiv.org/abs/1706.02409

[28] J. Kleinberg, S. Mullainathan, and M. Raghavan, "Inherent trade-offs in the fair determination of risk scores," 2016, *arXiv:1609.05807*. [Online]. Available: http://arxiv.org/abs/1609.05807

[29] R. B. Darlington, "Another look at 'cultural fairness' 1," *J. Educ. Meas*, vol. 8, no. 2, pp. 71–82, 1971.

[30] F. Chierichetti, R. Kumar, S. Lattanzi, and S. Vassilvitskii, "Fair clustering through fairlets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5029–5037.

[31] X. Chen, B. Fain, L. Lyu, and K. Munagala, "Proportionally fair clustering," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 1032–1041.

[32] A. Chhabra, V. Vashishth, and P. Mohapatra, "Fair algorithms for hierarchical agglomerative clustering," 2020, *arXiv:2005.03197*. [Online]. Available: http://arxiv.org/abs/2005.03197

[33] M. Ghadiri, S. Samadi, and S. Vempala, "Socially fair k-means clustering," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Mar. 2021, pp. 438–448. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C6&as_vis=1&q=Socially+fair+k-means+clustering&btnG=#d=gs_cit&u=%2Fscholar%3Fq%3Dinfo%3ArtRjNBgho0IJ%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den

[34] S. Mahabadi and A. Vakilian, "Individual fairness for k-clustering," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 6586–6596.

[35] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, and X. He, "Bias and debias in recommender system: A survey and future directions," 2020, *arXiv:2010.03240*. [Online]. Available: http://arxiv.org/abs/2010.03240

[36] S. Lin Blodgett, S. Barocas, H. Daumé, and H. Wallach, "Language (technology) is power: A critical survey of 'Bias' in NLP," 2020, *arXiv:2005.14050*. [Online]. Available: http://arxiv.org/abs/2005.14050

[37] M. Zehlike, K. Yang, and J. Stoyanovich, "Fairness in ranking: A survey," 2021, *arXiv:2103.14000*. [Online]. Available: http://arxiv.org/abs/2103.14000

[38] X. Zhang and M. Liu, "Fairness in learning-based sequential decision algorithms: A survey," in *Handbook of Reinforcement Learning and Control*. Cham, Switzerland: Springer, 2021, pp. 525–555.

[39] S. Dasgupta, "A cost function for similarity-based hierarchical clustering," in *Proc. 48th Annu. ACM Symp. Theory Comput.*, Jun. 2016, pp. 118–127.

[40] S. Dasgupta, "The hardness of k-means clustering," Dept. Comput. Sci. Eng., Univ. California, San Diego, CA, USA, Tech. Rep. CS2008-0916, 2008.

[41] N. Dupin, F. Nielsen, and E.-G. Talbi, "K-medoids and p-median clustering are solvable in polynomial time for a 2d Pareto front," 2018, *arXiv:1806.02098*. [Online]. Available: http://arxiv.org/abs/1806.02098

[42] R. Schrader, "Approximations to clustering and subgraph problems on trees," *Discrete Appl. Math.*, vol. 6, no. 3, pp. 301–309, Sep. 1983.

[43] L. K. P. J. Rdusseeun and P. Kaufman, "Clustering by means of medoids," in *Proc. Stat. Data Anal. Based L1 Norm Conf.*, Neuchâtel, Switzerland, Aug. 1987, pp. 405–416.

[44] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

[45] V. Cohen-Addad, V. Kanade, F. Mallmann-Trenn, and C. Mathieu, "Hierarchical clustering: Objective functions and algorithms," *J. ACM*, vol. 66, no. 4, pp. 1–42, 2019.

[46] B. Moseley, S. Vassilvtiskii, and Y. Wang, "Hierarchical clustering in general metric spaces using approximate nearest neighbors," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2021, pp. 2440–2448.

[47] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, vol. 96, no. 34, Aug. 1996, pp. 226–231. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C6&as_vis=1&q=A+density-based+algorithm+for+discovering+clusters+in+large+spatial+databases+with+noise&btnG=#d=gs_cit&u=%2Fscholar%3Fq%3Dinfo%3A-KybkyxcGYIJ%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den

[48] H.-P. Kriegel, P. Kröger, J. Sander, and A. Zimek, "Density-based clustering," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 1, no. 3, pp. 231–240, May/Jun. 2011.

[49] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.

[50] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Ann. Data Sci.*, vol. 2, no. 2, pp. 165–193, 2015.

[51] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, Oakland, CA, USA, vol. 1, 1967, pp. 281–297.

[52] K. Wagstaff, C. Cardie, S. Rogers, and S. Schrödl, "Constrained k-means clustering with background knowledge," in *Proc. 18th Int. Conf. Mach. Learn. (ICML)*. San Francisco, CA, USA: Morgan Kaufmann, 2001, pp. 577–584.

[53] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient K-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.

[54] M.-C. Su and C.-H. Chou, "A modified version of the K-means algorithm with a distance based on cluster symmetry," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 674–680, Jun. 2001.

[55] G. H. Ball and D. J. Hall, "A clustering technique for summarizing multivariate data," *Behav. Sci.*, vol. 12, no. 2, pp. 153–155, Mar. 1967.

[56] S. Guha, R. Rastogi, and K. Shim, "CURE: An efficient clustering algorithm for large databases," *ACM SIGMOD Rec.*, vol. 27, no. 2, pp. 73–84, 1998.

[57] J. H. Ward, Jr., "Hierarchical grouping to optimize an objective function," *J. Amer. Statist. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.

[58] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An efficient data clustering method for very large databases," *ACM SIGMOD Rec.*, vol. 25, no. 2, pp. 103–114, Jun. 1996.

[59] S. Guha, R. Rastogi, and K. Shim, "ROCK: A robust clustering algorithm for categorical attributes," *Inf. Syst.*, vol. 25, no. 5, pp. 345–366, 2000.

[60] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, vol. 344. Hoboken, NJ, USA: Wiley, 2009.

[61] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, May 2011.

[62] G. J. McLachlan and D. Peel, "The EMMIX algorithm for the fitting of normal and t-components," *J. Stat. Softw.*, vol. 4, no. 2, pp. 1–14, 1999.

[63] P. C. Cheeseman and J. C. Stutz, "Bayesian classification (AutoClass): Theory and results," *Adv. Knowl. Discovery Data Mining*, vol. 180, pp. 153–180, Feb. 1996.

[64] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[65] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 849–856.

[66] F. Bach and M. Jordan, "Learning spectral clustering," *Adv. Neural Inf. Process. Syst.*, vol. 16, no. 2, pp. 305–312, 2004.

[67] R. Sharan and R. Shamir, "Click: A clustering algorithm with applications to gene expression analysis," in *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, vol. 8, 2000, p. 16.

[68] J.-S. Cherng and M.-J. Lo, "A hypergraph based clustering algorithm for spatial data sets," in *Proc. IEEE Int. Conf. Data Mining*, Nov. 2001, pp. 83–90.
[69] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY, USA: Springer, 2013. [Online]. Available: https://www.springer.com/gp/book/9781475704525
[70] R. R. Yager and D. P. Filev, "Approximate clustering via the mountain method," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, no. 8, pp. 1279–1284, Aug. 1994.
[71] R. J. Hathaway, J. C. Bezdek, and Y. Hu, "Generalized fuzzy c-means clustering strategies using L_p norm distances," *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 5, pp. 576–582, Oct. 2000.
[72] J. F. Kolen and T. Hutcheson, "Reducing the time complexity of the fuzzy c-means algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 10, no. 2, pp. 263–267, Apr. 2002.
[73] A. B. Geva, "Hierarchical unsupervised fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 6, pp. 723–733, Dec. 1999.
[74] D. B. Fogel, "An introduction to simulated evolutionary optimization," *IEEE Trans. Neural Netw.*, vol. 5, no. 1, pp. 3–14, Jan. 1994.
[75] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
[76] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press, 1998.
[77] L. O. Hall, I. B. Ozyurt, and J. C. Bezdek, "Clustering with a genetically optimized approach," *IEEE Trans. Evol. Comput.*, vol. 3, no. 2, pp. 103–112, Jul. 1999.
[78] K. Krishna and M. N. Murty, "Genetic K-means algorithm," *IEEE Trans. Syst. Man, Cybern. B, Cybern.*, vol. 29, no. 3, pp. 433–439, Jun. 1999.
[79] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Comput. Surveys*, vol. 54, no. 6, pp. 1–35, Jul. 2021.
[80] K. Makhlouf, S. Zhioua, and C. Palamidessi, "On the applicability of ML fairness notions," 2020, *arXiv:2006.16745*. [Online]. Available: http://arxiv.org/abs/2006.16745
[81] G. Rutherglen, "Disparate impact under title VII: An objective theory of discrimination," *Virginia Law Rev.*, vol. 73, no. 7, p. 1297, Oct. 1987.
[82] I.-C. Yeh and C.-H. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2473–2480, Mar. 2009.
[83] (2020). *Racial, Gender Wage Gaps Persist in U.S. Despite Some Progress*. Accessed: Jun. 1, 2021. [Online]. Available: https://www.pewresearch.org/fact-tank/2016/07/01/racial-gender-wage-gaps-persist-in-u-s-despite-some-progress/
[84] (2020). *7 Findings That Illustrate Racial Disparities in Education*. Accessed: Jun. 1, 2021. [Online]. Available: https://www.brookings.edu/blog/brown-center-chalkboard/2016/06/06/7-findings-that-illustrate-racial-disparities-in-education/
[85] S. Bera, D. Chakrabarty, N. Flores, and M. Negahbani, "Fair algorithms for clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 4954–4965.
[86] S. Ahmadian, A. Epasto, R. Kumar, and M. Mahdian, "Clustering without over-representation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 267–275.
[87] M. Kleindessner, P. Awasthi, and J. Morgenstern, "A notion of individual fairness for clustering," 2020, *arXiv:2006.04960*. [Online]. Available: http://arxiv.org/abs/2006.04960
[88] C. Jung, S. Kannan, and N. Lutz, "A center in your neighborhood: Fairness in facility location," 2019, *arXiv:1908.09041*. [Online]. Available: http://arxiv.org/abs/1908.09041
[89] N. Anderson, S. K. Bera, S. Das, and Y. Liu, "Distributional individual fairness in clustering," 2020, *arXiv:2006.12589*. [Online]. Available: http://arxiv.org/abs/2006.12589
[90] D. Chakrabarti, J. P. Dickerson, S. A. Esmaeili, A. Srinivasan, and L. Tsepenekas, "A new notion of individually fair clustering: α-equitable k-center," 2021, *arXiv:2106.05423*. https://arxiv.org/abs/2106.05423
[91] D. Chakrabarty and M. Negahbani, "Better algorithms for individually fair k-clustering," 2021, *arXiv:2106.12150*. [Online]. Available: http://arxiv.org/abs/2106.12150
[92] A. Vakilian and M. Yalçıner, "Improved approximation algorithms for individually fair clustering," 2021, *arXiv:2106.14043*. [Online]. Available: http://arxiv.org/abs/2106.14043
[93] E. Micha and N. Shah, "Proportionally fair clustering revisited," in *Proc. 47th Int. Colloq. Automata, Lang., Program. (ICALP)*. Wadern, Germany: Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020, pp. 1–16.
[94] B. Brubach, D. Chakrabarti, J. Dickerson, S. Khuller, A. Srinivasan, and L. Tsepenekas, "A pairwise fair and community-preserving approach to k-center clustering," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1178–1189.
[95] M. Kearns, S. Neel, A. Roth, and Z. S. Wu, "Preventing fairness gerrymandering: Auditing and learning for subgroup fairness," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2564–2572.
[96] M. Kearns, S. Neel, A. Roth, and Z. S. Wu, "An empirical study of rich subgroup fairness for machine learning," in *Proc. Conf. Fairness, Accountability, Transparency*, Jan. 2019, pp. 100–109.
[97] M. Schmidt, C. Schwiegelshohn, and C. Sohler, "Fair coresets and streaming algorithms for fair k-means," in *Proc. Int. Workshop Approximation Online Algorithms*. Cham, Switzerland: Springer, Sep. 2019, pp. 232–251. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C6&as_vis=1&q=Fair+coresets+and+streaming+algorithms+for+fair+k-means&btnG=#d=gs_cit&u=%2Fscholar%3Fq%3Dinfo%3AemZQBW9SJbwJ%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den
[98] I. O. Bercea, M. Groß, S. Khuller, A. Kumar, C. Rösner, D. R. Schmidt, and M. Schmidt, "On the cost of essentially fair clusterings," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*. Wadern, Germany: Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2019.
[99] A. Backurs, P. Indyk, K. Onak, B. Schieber, A. Vakilian, and T. Wagner, "Scalable fair clustering," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 405–413.
[100] M. Kleindessner, S. Samadi, P. Awasthi, and J. Morgenstern, "Guarantees for spectral clustering with fairness constraints," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3458–3467.
[101] H. Zhang and I. Davidson, "Deep fair discriminative clustering," 2021, *arXiv:2105.14146*. [Online]. Available: http://arxiv.org/abs/2105.14146
[102] B. Wang and I. Davidson, "Towards fair deep clustering with multi-state protected variables," 2019, *arXiv:1901.10053*. [Online]. Available: http://arxiv.org/abs/1901.10053
[103] M. Abbasi, A. Bhaskara, and S. Venkatasubramanian, "Fair clustering via equitable group representations," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Mar. 2021, pp. 504–514.
[104] Y. Makarychev and A. Vakilian, "Approximation algorithms for socially fair clustering," 2021, *arXiv:2103.02512*. [Online]. Available: http://arxiv.org/abs/2103.02512
[105] S. Ahmadian, A. Epasto, M. Knittel, R. Kumar, M. Mahdian, B. Moseley, P. Pham, S. Vassilvitskii, and Y. Wang, "Fair hierarchical clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1–28.
[106] I. Csiszar, "Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten," *Magyer Tud. Akad. Mat. Kutato Int. Koezl.*, vol. 8, pp. 85–108, 1964. [Online]. Available: https://ci.nii.ac.jp/naid/10006737982/en/
[107] T. Morimoto, "Markov processes and the H-theorem," *J. Phys. Soc. Jpn.*, vol. 18, no. 3, pp. 328–331, 1963.
[108] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Roy. Statist. Soc. B, Methodol.*, vol. 28, no. 1, pp. 131–142, 1966.
[109] G. J. McLachlan and K. E. Basford, *Mixture Models: Inference and Applications to Clustering*, vol. 38. New York, NY, USA: M. Dekker, 1988.
[110] J. M. Joyce, *Kullback-Leibler Divergence*. Berlin, Germany: Springer, 2011, pp. 720–722.
[111] P. Li, H. Zhao, and H. Liu, "Deep fair clustering for visual learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9070–9079.
[112] M. Böhm, A. Fazzone, S. Leonardi, and C. Schwiegelshohn, "Fair clustering with multiple colors," 2020, *arXiv:2002.07892*. [Online]. Available: http://arxiv.org/abs/2002.07892
[113] C. Rösner and M. Schmidt, "Privacy preserving clustering with constraints," in *Proc. 45th Int. Colloq. Automata, Lang., Program. (ICALP)*. Wadern, Germany: Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2018.
[114] S. Liu and L. N. Vicente, "A stochastic alternating balance k-means algorithm for fair clustering," 2021, *arXiv:2105.14172*. [Online]. Available: http://arxiv.org/abs/2105.14172
[115] I. M. Ziko, E. Granger, J. Yuan, and I. B. Ayed, "Variational fair clustering," 2019, *arXiv:1906.08207*. [Online]. Available: http://arxiv.org/abs/1906.08207

[116] A. Chhabra, A. Singla, and P. Mohapatra, "Fair clustering using antidote data," 2021, *arXiv:2106.00600*. [Online]. Available: http://arxiv.org/abs/2106.00600
[117] T. Le Quy, A. Roy, G. Friege, and E. Ntoutsi, "Fair-capacitated clustering," 2021, *arXiv:2104.12116*. [Online]. Available: http://arxiv.org/abs/2104.12116
[118] S. Esmaeili, B. Brubach, L. Tsepenekas, and J. Dickerson, "Probabilistic fair clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1–13.
[119] S. Ahmadian, A. Epasto, R. Kumar, and M. Mahdian, "Fair correlation clustering," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2020, pp. 4195–4205.
[120] X. Jia, K. Sheth, and O. Svensson, "Fair colorful k-center clustering," in *Integer Programming and Combinatorial Optimization* (Lecture Notes in Computer Science), vol. 12125, D. Bienstock and G. Zambelli, Eds. Cham, Switzerland: Springer, 2020. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-45771-6_17, doi: 10.1007/978-3-030-45771-6_17.
[121] L. Huang, H.-C. S. Jiang, and K. N. Vishnoi, "Coresets for clustering with fairness constraints," in *Proc. NeurIPS*, 2019, pp. 7587–7598.
[122] S. Bandyapadhyay, F. V. Fomin, and K. Simonov, "On coresets for fair clustering in metric and Euclidean spaces and their applications," 2020, *arXiv:2007.10137*. [Online]. Available: http://arxiv.org/abs/2007.10137
[123] E. Harb and H. S. Lam, "KFC: A scalable approximation algorithm for k- center fair clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1–14.
[124] D. Goyal and R. Jaiswal, "Tight FPT approximation for socially fair clustering," 2021, *arXiv:2106.06755*. [Online]. Available: http://arxiv.org/abs/2106.06755
[125] S. Thejaswi, B. Ordozgoiti, and A. Gionis, "Diversity-aware k-median: Clustering with fair center representation," 2021, *arXiv:2106.11696*. [Online]. Available: http://arxiv.org/abs/2106.11696
[126] S. A. Esmaeili, B. Brubach, A. Srinivasan, and J. P. Dickerson, "Fair clustering under a bounded cost," 2021, *arXiv:2106.07239*. [Online]. Available: http://arxiv.org/abs/2106.07239
[127] S. S. Abraham, P. Deepak, and S. S. Sundaram, "Fairness in clustering with multiple sensitive attributes," in *Proc. EDBT*, 2020, pp. 1–12.
[128] M. Kleindessner, P. Awasthi, and J. Morgenstern, "Fair k-center clustering for data summarization," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3448–3457.
[129] A. Chiplunkar, S. Kale, and S. N. Ramamoorthy, "How to solve fair k-center in massive data models," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1877–1886.
[130] M. Jones, H. Nguyen, and T. Nguyen, "Fair k-centers via maximum matching," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 4940–4949.
[131] A. Anagnostopoulos, L. Becchetti, A. Fazzone, C. Menghini, and C. Schwiegelshohn, "Spectral relaxations and fair densest subgraphs," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 35–44.
[132] A. Anagnostopoulos, L. Becchetti, A. Fazzone, C. Menghini, and C. Schwiegelshohn, "Principal fairness: Removing bias via projections," 2019, *arXiv:1905.13651*. [Online]. Available: http://arxiv.org/abs/1905.13651
[133] L. Lü and T. Zhou, "Link prediction in weighted networks: The role of weak ties," *Europhys. Lett.*, vol. 89, no. 1, p. 18001, Jan. 2010.
[134] K.-K. Shang, T.-C. Li, M. Small, D. Burton, and Y. Wang, "Link prediction for tree-like networks," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 29, no. 6, Jun. 2019, Art. no. 061103.
[135] L. Linyuan, P. Liming, Z. Tao, Z. Yi-Cheng, and H. E. Stanley, "Toward link predictability of complex networks," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 8, pp. 2325–2330, 2015.
[136] K.-K. Shang, M. Small, X.-K. Xu, and W.-S. Yan, "The role of direct links for link prediction in evolving networks," *Europhys. Lett.*, vol. 117, no. 2, p. 28002, Jan. 2017.
[137] I. Davidson and S. R. Ravi, "Making existing clusterings fairer: Algorithms, complexity results and insights," in *Proc. AAAI*, 2020, pp. 3733–3740.
[138] R. Kohavi, "Scaling up the accuracy of Naive-Bayes classifiers: A decision-tree hybrid," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, vol. 96, 1996, pp. 202–207.
[139] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decis. Support Syst.*, vol. 62, no. 1, pp. 22–31, Jun. 2014.
[140] B. Strack, J. P. DeShazo, C. Gennings, J. L. Olmo, S. Ventura, K. J. Cios, and J. N. Clore, "Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records," *BioMed Res. Int.*, vol. 2014, Apr. 2014, Art. no. 781670.
[141] C. Meek, B. Thiesson, and D. Heckerman, "The learning curve method applied to clustering," in *Proc. Int. Workshop Artif. Intell. Statist.*, 2001, pp. 196–202.
[142] J. Houvardas and E. Stamatatos, "N-gram feature selection for authorship identification," in *Artificial Intelligence: Methodology, Systems, and Applications* (Lecture Notes in Computer Science), vol. 4183, J. Euzenat and J. Domingue, Eds. Berlin, Germany: Springer, 2006. [Online]. Available: https://link.springer.com/chapter/10.1007/11861461_10, doi: 10.1007/11861461_10.
[143] E. Stamatatos, "A survey of modern authorship attribution methods," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 3, pp. 538–556, 2009.
[144] (2018). *120 Years of Olympic History: Athletes and Results*. [Online]. Available: https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results
[145] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
[146] (2019). *D. Ataset. 4*. [Online]. Available: https://dblp.uni-trier.de/xml/
[147] S. J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. K. Chan, "Cost-based modeling for fraud and intrusion detection: Results from the JAM project," in *Proc. DARPA Inf. Survivability Conf. Expo.*, vol. 2, Jan. 2000, pp. 130–144.
[148] (2016). *Pima Indians Diabetes Dataset*. [Online]. Available: https://www.kaggle.com/uciml/pima-indians-diabetes-database
[149] *North Carolina Voters Dataset*. [Online]. Available: https://www.ncsbe.gov/results-data/voter-registration-data
[150] O. Ekin, P. L. Hammer, A. Kogan, and P. Winter, "Distance-based classification methods," *Inf. Syst. Oper. Res.*, vol. 37, no. 3, pp. 337–352, Aug. 1999.
[151] E. Fehrman, A. K. Muhammad, E. M. Mirkes, V. Egan, and A. N. Gorban, "The five factor model of personality and evaluation of drug consumption risk," in *Data Science* (Studies in Classification, Data Analysis, and Knowledge Organization), F. Palumbo, A. Montanari, and M. Vichi, Eds. Cham, Switzerland: Springer, 2017. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-55723-6_18, doi: 10.1007/978-3-319-55723-6_18.
[152] B. V. Ramana, M. S. P. Babu, and N. Venkateswarlu, "A critical comparative study of liver patients from USA and INDIA: An exploratory analysis," *Int. J. Comput. Sci. Issues*, vol. 9, no. 3, p. 506, 2012.
[153] R. Mastrandrea, J. Fournet, and A. Barrat, "Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys," *PLoS ONE*, vol. 10, no. 9, Sep. 2015, Art. no. e0136497.
[154] M. R. Weeks, S. Clair, S. P. Borgatti, K. Radda, and J. J. Schensul, "Social networks of drug users in high-risk sites: Finding the connections," *AIDS Behav.*, vol. 6, no. 2, pp. 193–206, 2002.
[155] *Kinematics Word Problems*. Accessed: Sep. 1, 2021. [Online]. Available: https://github.com/savithaabraham/Datasets
[156] *P-Median Uncapacitated Dataset*. Accessed: Sep. 1, 2021. [Online]. Available: http://people.brunel.ac.uk/~mastjjb/jeb/orlib/pmedinfo.html
[157] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," in *Proc. 5th Annu. Future Bus. Technol. Conf.*, A. Brito and J. Teixeira, Eds. Porto, Portugal: EUROSIS, 2008, pp. 5–12. [Online]. Available: https://repositorium.sdum.uminho.pt/handle/1822/8024
[158] *Programme for International Student Assessment*. Accessed: Sep. 1, 2021. [Online]. Available: https://www.oecd.org/pisa/data/
[159] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open university learning analytics dataset," *Sci. Data*, vol. 4, no. 1, pp. 1–8, Dec. 2017.
[160] S. Kumar, X. Zhang, and J. Leskovec, "Predicting dynamic embedding trajectory in temporal interaction networks," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 1269–1278.
[161] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," in *Proc. Workshop Faces 'Real-Life' Images, Detection, Alignment, Recognit.*, Oct. 2008. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C6&as_vis=1&q=Labeled+faces+2036+in+the+wild%3A+A+database+forstudying+face+recognition+in+unconstrained+2037+environments&btnG=#d=gs_cit&u=%2Fscholar%3Fq%3Dinfo%3AsmKz1UTpLF0J%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den
[162] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: Understanding rating dimensions with review text," in *Proc. 7th ACM Conf. Recommender Syst.*, Oct. 2013, pp. 165–172.
[163] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proc. Esann*, vol. 3, 2013, p. 3.

[164] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in *Computer Vision—ECCV 2014* (Lecture Notes in Computer Science), vol. 8694, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-10599-4_7, doi: 10.1007/978-3-319-10599-4_7.
[165] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Computer Vision—ECCV 2010* (Lecture Notes in Computer Science), vol. 6314, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Berlin, Germany: Springer, 2010. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-15561-1_16, doi: 10.1007/978-3-642-15561-1_16.
[166] K. Altun, B. Barshan, and O. Tunçel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognit.*, vol. 43, no. 10, pp. 3605–3620, Oct. 2010.
[167] D. Dua and C. Graff, "UCI machine learning repository," School Inf. Comput. Sci., Univ. California, Irvine, CA, USA, 2019. [Online]. Available: https://archive.ics.uci.edu/ml/citation_policy.html
[168] H. Suresh and J. V. Guttag, "A framework for understanding sources of harm throughout the machine learning life cycle," 2019, *arXiv:1901.10002*. [Online]. Available: http://arxiv.org/abs/1901.10002
[169] A. Gupta and D. Garg, "Applying data mining techniques in job recommender system for considering candidate job preferences," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2014, pp. 1458–1465.
[170] J. Dressel and H. Farid, "The accuracy, fairness, and limits of predicting recidivism," *Sci. Adv.*, vol. 4, no. 1, Jan. 2018, Art. no. eaao5580.
[171] M. Ahmed, M. T. Imtiaz, and R. Khan, "Movie recommendation system using clustering and pattern recognition network," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2018, pp. 143–147.
[172] S. B. Aher and L. M. R. J. Lobo, "Combination of machine learning algorithms for recommendation of courses in E-learning system based on historical data," *Knowl.-Based Syst.*, vol. 51, pp. 1–14, Oct. 2013.
[173] C. Bohm, K. Kailing, H.-P. Kriegel, and P. Kroger, "Density connected clustering with local subspace preferences," in *Proc. 4th IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2004, pp. 27–34.
[174] K. Kailing, H.-P. Kriegel, and P. Kröger, "Density-connected subspace clustering for high-dimensional data," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2004, pp. 246–256.
[175] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.
[176] I. Caragiannis, C. Kaklamanis, P. Kanellopoulos, and M. Kyropoulou, "The efficiency of fair division," *Theory Comput. Syst.*, vol. 50, no. 4, pp. 589–610, May 2012.
[177] D. Bertsimas, V. F. Farias, and N. Trichakis, "The price of fairness," *Oper. Res.*, vol. 59, no. 1, pp. 17–31, 2011.
[178] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. 1, pp. 53–65, 1987.
[179] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Stat., Theory Methods*, vol. 3, no. 1, pp. 1–27, 1974.
[180] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.
[181] A. Chhabra, A. Roy, and P. Mohapatra, "Suspicion-free adversarial attacks on clustering algorithms," in *Proc. AAAI*, 2020, pp. 1–8.
[182] A. E. Cinà, A. Torcinovich, and M. Pelillo, "A black-box adversarial attack for poisoning clustering," 2020, *arXiv:2009.05474*. [Online]. Available: http://arxiv.org/abs/2009.05474
[183] N. Papernot, P. McDaniel, and I. Goodfellow, "Transferability in machine learning: From phenomena to black-box attacks using adversarial samples," 2016, *arXiv:1605.07277*. [Online]. Available: http://arxiv.org/abs/1605.07277
[184] N. X. Vinh, J. Epps, and J. Bailey, "Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance," *J. Mach. Learn. Res.*, vol. 11, pp. 2837–2854, Jan. 2010.
[185] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long, "A survey of clustering with deep learning: From the perspective of network architecture," *IEEE Access*, vol. 6, pp. 39501–39514, 2018.
[186] R. Binns, M. V. Kleek, M. Veale, U. Lyngs, J. Zhao, and N. Shadbolt, "'it's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2018, p. 377.
[187] R. E. Bellman, *Adaptive Control Processes*. Princeton, NJ, USA: Princeton Univ. Press, 2015.
[188] M. Steinbach, L. Ertöz, and V. Kumar, "The challenges of clustering high dimensional data," in *New Directions in Statistical Physics*. Berlin, Germany: Springer, 2004, pp. 273–309. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-662-08968-2_16#citeas
[189] H. P. Kriegel, P. Kröger, and A. Zimek, "Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering," *ACM Trans. Knowl. Discovery Data*, vol. 3, no. 1, pp. 1–58, Mar. 2009.
[190] S. Basu, I. Davidson, and K. Wagstaff, *Constrained Clustering: Advances in Algorithms, Theory, and Applications*. Boca Raton, FL, USA: CRC Press, 2008.
[191] K. Wagstaff and C. Cardie, "Clustering with instance-level constraints," *AAAI/IAAI*, vol. 1097, pp. 577–584, Jun. 2000.
[192] S. Basu, A. Banerjee, and R. J. Mooney, "Active semi-supervision for pairwise constrained clustering," in *Proc. Int. Conf. Data Mining*. Philadelphia, PA, USA: SIAM, Apr. 2004, pp. 333–344.

ANSHUMAN CHHABRA received the B.Eng. degree in electronics and communication engineering from the University of Delhi, India. He is currently pursuing the Ph.D. degree with the University of California, Davis.

His research interests include unsupervised learning, fair machine learning, adversarial attacks against learning models, and the applicability of machine learning to other fields of computer science, such as wireless networks. He received the UC Davis Graduate Student Fellowship, in 2018, and has completed internships at ESnet, Lawrence Berkeley National Laboratory, USA, in 2017, and Max Planck Institute for Software Systems, Germany, in 2020.

KARINA MASALKOVAITĖ received the B.Sc. degree in chemical engineering from the University of California, Davis, in 2021.

During her bachelor's degree, she participated as an Undergraduate Researcher doing computational and theory-based work for an Organic Electronics Laboratory. Her research interests include machine learning, molecular simulations, and documentation development. Her awards and honors include being a recipient of the Wasson Honors Program, receiving the Chemical Engineering Senior Design Award, and being a member of Tau Beta Pi (Engineering Honors Society).

PRASANT MOHAPATRA (Fellow, IEEE) received the Ph.D. degree from Penn State University, in 1993.

He is serving as the Vice Chancellor for Research with the University of California, Davis, where he is currently a Distinguished Professor with the Department of Computer Science. Prior to this, he was the Department Chair of Computer Science, from 2007 to 2013. His research has been funded through grants from the National Science Foundation, U.S. Department of Defense, U.S. Army Research Laboratory, Intel Corporation, Siemens, Panasonic Technologies, Hewlett Packard, Raytheon, and EMC Corporation. He has published more than 350 articles in reputed conferences and journals on these topics. His research interests include wireless networks, mobile communications, cybersecurity, and internet protocols. He is a fellow of AAAS. He received the Outstanding Engineering Alumni Award, in 2008. He also received the Outstanding Research Faculty Award from the College of Engineering, University of California at Davis, and the HP Labs Innovation Awards, from 2011 to 2013.