Scalable and Principled Fairness in Fuzzy Clustering

Anonymous cvm submission

Paper ID 603

Abstract

Clustering algorithms are essential for organizing large-scale visual data, yet they risk amplifying societal biases embedded within these datasets. Prevailing fair clustering methods falter on two fronts: their computational cost is prohibitive for massive visual datasets, and they rely on a fragile, manually-tuned hyperparameter to balance accuracy and fairness. This paper introduces FairFBC(Fair Fuzzy Balanced Clustering), a framework that confronts these challenges directly. It first constructs a compact, demographically-balanced anchor graph to ensure both scalability and equitable data representation. Upon this structure, a novel objective learns a single, shared cluster partition by simultaneously maximizing quality across all sensitive groups. This joint optimization embeds fairness into its core, intrinsically navigating the accuracy-fairness trade-off without any external balancing parameter. Extensive experiments on diverse benchmarks, including large-scale vision datasets where most alternatives fail, demonstrate FairFBC's dominance. Rigorous multi-objective analysis shows that our method consistently achieves a superior Pareto front, establishing a new state of the art across the accuracy-fairness spectrum and offering a truly scalable and principled solution for equitable visual data analysis.

***Keywords:** Fair clustering, anchor graph, fuzzy clustering,balance clustering, group fairness.*

1. Introduction

Clustering is a central pillar of unsupervised visual computing, enabling fundamental tasks from semantic segmentation of scenes to organizing vast, unlabelled image repositories [15]. Yet, this utility is shadowed by a critical flaw: standard algorithms, when applied to visual data, readily absorb and amplify the demographic biases they contain. A face clustering system trained on skewed data, for instance, may construct less coherent clusters for underrepresented groups, compromising downstream applications like equitable person re-identification. Similarly, in medical imaging analysis, an algorithm might partition patient scans along demographic lines rather than pathological ones, risking diagnostic disparities. This reality demands clustering methods engineered not just for accuracy, but for explicit fairness.

The pursuit of fair clustering, however, has yielded a landscape of methods that still grapple with significant practical barriers. One major line of work, initiated by the seminal fairlet decomposition approach [7], offers strong theoretical guarantees. Subsequent efforts have focused on improving the scalability of this paradigm [5, 3]. Despite these advances, a fundamental tension persists: many algorithms that provide formal fairness assurances, including very recent ones [4], still rely on complex pre-processing or constrained optimization steps that become computational bottlenecks. This issue is particularly acute for the massive datasets—often containing hundreds of thousands of high-dimensional visual embeddings—that define modern computer vision, leaving a gap where fairness guarantees fail to reach the scale at which they are most needed.

A second, more fundamental challenge lies in the mechanism of fairness itself. The dominant paradigm, exemplified by methods from variational clustering [23] to recent ensemble techniques [20], involves appending a fairness penalty to a standard clustering objective. This trade-off is governed by a manually-tuned hyperparameter, λ. This strategy is inherently brittle. The optimal balance is data-dependent, forcing practitioners into expensive, ad-hoc grid searches. This reliance on an external knob transforms a principled goal into a trial-and-error exercise, begging for a more integrated solution.

This paper presents FairFBC (Fair Fuzzy Balanced Clustering), a framework designed to dismantle these dual barriers of scalability and principled balancing. Our work is built upon a core insight: a truly fair clustering should not be the result of a post-hoc balancing act, but must emerge from a consensus sought across all demographic groups. To realize this, we design a multi-stage process that weaves fairness into its computational fabric. The process begins by constructing a compact, demographically-stratified anchor graph, which establishes an equitable and efficient data representation from the start.

Crucially, FairFBC then introduces a novel objective that departs entirely from hyperparameter-based balancing. It

learns a single, shared cluster structure by jointly maximizing the clustering quality for all sensitive groups at once. The trade-off between accuracy and fairness is thus resolved intrinsically within the optimization itself, which is mathematically guided to find a consensus partition. This design eliminates the fragile balancing parameter, leading to a more robust and principled model.

Our contributions are threefold:

- We design a scalable method built on a fair anchor graph representation that runs effectively on large-scale visual datasets where many state-of-the-art baselines fail.
- We propose a novel, hyperparameter-free objective function that intrinsically resolves the accuracy-fairness trade-off by seeking a consensus partition across demographic groups.
- We introduce a principled assignment strategy that leverages the model's fuzzy outputs to enforce strict fairness guarantees with minimal perturbation to the underlying cluster structure.

2. Related Work

Research in fair clustering has evolved along several distinct axes, each presenting its own trade-offs between theoretical guarantees, practical scalability, and the mechanism for enforcing fairness.

Fairness via Algorithmic Constraints. A foundational line of work enforces fairness through explicit algorithmic constraints. Group fairness, which seeks proportional demographic representation in each cluster, was notably formalized by the fairlet decomposition method [7]. This seminal work and its successors [5, 3] offer provable approximations for center-based clustering, but their reliance on combinatorial pre-processing often incurs super-linear costs that limit applicability to large-scale data. A parallel thrust on individual fairness, which mandates that similar inputs receive similar outputs, has also yielded sophisticated algorithms [14, 19]. While theoretically elegant, both paradigms highlight a persistent tension: hard constraints, whether for group balance or individual similarity, tend to complicate solvers and limit scalability on the massive datasets prevalent in visual computing.

Fairness via Objective and Representation Design. An alternative approach integrates fairness directly into the learning objective or the data representation itself. Many methods, such as Variational Fair Clustering (VFC) [24], augment a standard clustering loss with a regularizer that penalizes unfairness, governed by a trade-off hyperparameter. In the context of visual data, deep learning approaches aim to learn fair representations by using adversarial training or other constraints to suppress sensitive attribute information from image embeddings. While often effective, these methods typically require extensive, data-specific tuning of the balancing hyperparameter.

Graph-based methods represent another key direction. Some techniques adapt spectral clustering by adding fairness constraints to the graph Laplacian eigendecomposition [12]. More recently, Graph Neural Networks (GNNs) like FairGNN [8] have been employed to learn node representations that are invariant to sensitive attributes. These methods excel at capturing relational data structures but often operate on a full $n \times n$ affinity matrix, posing memory and computational challenges. Furthermore, like their non-graph counterparts, many still depend on a manually-tuned parameter to mediate the accuracy-fairness trade-off.

Scalability via Anchor Graph Models. Orthogonal to the pursuit of fairness, the challenge of scalability in graph-based learning has been effectively addressed by anchor or landmark graph models [11, 13]. These methods construct a sparse, bipartite graph representation of size $n \times m$ (where anchors $m \ll n$), dramatically reducing the complexity of downstream tasks. Techniques like approximate Personalized PageRank (PPR) diffusion can further enrich these graphs with robust, locality-aware affinities in near-linear time [1, 2]. However, this powerful scalability toolkit has not been systematically designed with fairness in mind; prior work did not explicitly encode demographic balance into the anchor selection or the clustering objective.

Positioning of FairFBC. FairFBC uniquely synthesizes these threads. It inherits the scalability of anchor graph models but fundamentally re-engineers them for fairness. Our contribution is threefold. **(1) Fair Representation by Design:** Unlike constraint-based methods, we embed fairness at the earliest stage by constructing the anchor graph via stratified sampling and diffusion. This yields a data representation that is both computationally efficient and demographically equitable from the outset. **(2) Hyperparameter-Free Balancing:** In sharp contrast to objective-shaping methods [24], FairFBC's joint optimization objective entirely sidesteps the need for a balancing hyperparameter. It mathematically compels a consensus partition across groups, intrinsically resolving the accuracy-fairness trade-off. **(3) Large-Scale Practicality:** By confining all iterative optimization to the compact anchor space, our method's complexity becomes independent of the sample size n. This allows FairFBC to complement existing approaches by delivering not just a single, tuned solution, but a robust and efficient Pareto front, even on massive visual datasets.

3. METHOD

Our FairFBC method tackles fair clustering through a multi-stage process.We first construct a fair bipartite graph to create an efficient and equitable data representation. On this graph, we then solve a novel joint optimization problem to learn a fair cluster structure. Finally, a principled hard assignment stage produces the discrete cluster labels.

3.1. Fair Bipartite Graph Construction

Our method's fairness hinge on a compact graph representation. We construct a bipartite affinity matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$ connecting n samples to m anchors ($m \ll n$) in three stages.

First, we perform **Balanced Anchor Selection**. To ensure equitable representation, we allocate the m anchor slots proportionally to the sensitive group sizes, with a guarantee of at least one anchor per non-empty group. Anchors are then identified as the centroids from k-means run within each group's data subset. This stratified approach ensures the anchor set is demographically balanced from the outset.

Second, a symmetric weighted k-NN graph $\mathbf{S} \in \mathbb{R}^{n\times n}$ is built to capture the data's local manifold structure. Edge weights use a Gaussian kernel with bandwidth σ set via the median k-th neighbor distance.

Finally, we perform **Graph Diffusion** to compute the affinity matrix $\mathbf{B}$. Each column of $\mathbf{B}$, denoted $\mathbf{p}_j$, is an Approximate Personalized PageRank (APPR) vector computed by initiating a diffusion process from the j-th anchor on the graph $\mathbf{S}$. The vector $\mathbf{p}_j$ is the solution to the fixed-point equation:

$$\mathbf{p}_j = (1-\alpha)\tilde{\mathbf{S}}\mathbf{p}_j + \alpha\mathbf{s}_j, \tag{1}$$

where $\mathbf{s}_j$ is a basis vector for the j-th anchor seed, $\tilde{\mathbf{S}}$ is the row-normalized $\mathbf{S}$, and $\alpha = 0.15$ is the teleport probability. We solve this for each anchor j using a push-based algorithm to construct the sparse matrix $\mathbf{B} = [\mathbf{p}_1, \ldots, \mathbf{p}_m]$.

3.2. Objective Function

Our objective function is constructed upon two core principles designed to achieve fairness.

First, to achieve inter-group fairness, we employ joint optimization across all G sensitive groups by summing their individual quality scores. This forces the model to find a single, shared solution $\mathbf{U}$ that must perform well for all groups simultaneously. For instance, an ideal solution would balance assignments for both 9 males (as 3:3:3) and 6 females (as 2:2:2), naturally leading to a globally fair outcome (5:5:5 in each cluster). By preventing the model from overfitting to a dominant group's structure, this partitioning strategy also encourages more robust and generalizable patterns, an approach with precedents in other clustering contexts [18, 16].

Second, to ensure intra-group balance, we select a quality metric, $\mathrm{Tr}(\sqrt{\cdot})$, for each group's structure. As established in recent work [6], maximizing this term—due to the strict concavity of the square root function—naturally encourages the eigenvalues of its matrix argument to be balanced.

These two principles lead to our final objective function, a maximization problem over a shared anchor membership matrix $\mathbf{U} \in \mathbb{R}^{m\times c}$ and weights $\mathbf{w} \in \mathbb{R}^m$:

$$\begin{aligned} \max_{\mathbf{U},\mathbf{w}} \quad & J(\mathbf{U},\mathbf{w}) = \sum_{g=1}^{G} \mathrm{Tr}\left(\sqrt{\mathbf{U}^\top \mathbf{W}\mathbf{W}_g\mathbf{U}}\right) \\ \text{s.t.} \quad & \mathbf{U}_{i,:} \in \Delta^c,\ \forall i; \quad \mathbf{w} \in \Delta^m. \end{aligned} \tag{2}$$

Δ^d denotes the d-dimensional probability simplex, $\mathbf{W} = \mathrm{diag}(\mathbf{w}) \in \mathbb{R}^{m\times m}$, and $\mathbf{W}_g = \mathbf{B}_g^\top \mathbf{B}_g \in \mathbb{R}^{m\times m}$ ($\mathbf{B}_g \in \mathbb{R}^{n_g\times m}$)is the group-specific anchor covariance matrix derived from the affinity matrix $\mathbf{B}$ (Sec. 3.1).Finally, we compute the fuzzy membership matrix $\mathbf{F} \in \mathbb{R}^{n\times c}$ for all n samples by mapping the learned anchor assignments $\mathbf{U}$ through the affinity matrix $\mathbf{B}$:$\mathbf{F} \leftarrow \mathbf{B}\mathbf{U}$.

3.3. Optimization

We solve the non-convex problem in Eq. (2) using an alternating optimization algorithm guaranteed to converge to a stationary point. We initialize anchor weights $\mathbf{w}$ via anchor degrees and memberships $\mathbf{U}$ using a PCA-enhanced k-means++ strategy. The algorithm then alternates between two steps:

1. Update for Anchor Weights $\mathbf{w}$. With $\mathbf{U}$ fixed, we update $\mathbf{w} \in \Delta^m$ via a mirror descent step. The gradient of the objective J with respect to the weight vector $\mathbf{w}$ is:

$$\nabla_{\mathbf{w}} J = \mathrm{diag}\Big(\sum_{g=1}^{G} \mathbf{W}_g\mathbf{U}\Big(\tfrac{1}{2}(\mathbf{U}^\top\mathbf{W}\mathbf{W}_g\mathbf{U} + \epsilon\mathbf{I})^{-\frac{1}{2}}\Big)\mathbf{U}^\top\Big). \tag{3}$$

The update is an exponential step followed by normalization:

$$\tilde{\mathbf{w}} \leftarrow \mathbf{w} \odot \exp(\eta\nabla_{\mathbf{w}} J), \quad \mathbf{w} \leftarrow \frac{\tilde{\mathbf{w}}}{\mathbf{1}^\top\tilde{\mathbf{w}}}, \tag{4}$$

where η is an adaptive step size.

2. Update for Memberships $\mathbf{U}$. With $\mathbf{w}$ fixed, we update $\mathbf{U}$ via projected gradient ascent. The gradient is:

$$\nabla_U J = \sum_{g=1}^{G} \mathbf{W}\mathbf{W}_g\mathbf{U}(\mathbf{U}^\top\mathbf{W}\mathbf{W}_g\mathbf{U})^{-\frac{1}{2}}. \tag{5}$$

The update combines an ascent step with a row-wise projection $\Pi_{\Delta^c}(\cdot)$ onto the probability simplex:

$$\mathbf{U} \leftarrow \Pi_{\Delta^c}\left(\mathbf{U} + \alpha\nabla_U J\right). \tag{6}$$

The step size α is determined via Armijo backtracking. The process terminates upon convergence of the objective value.

3.4. Fair Hard Assignment

The optimization stage yields a fuzzy membership matrix $\mathbf{F} \in \mathbb{R}^{n\times c}$. To convert this soft output into a discrete partition that adheres to group-level fairness while minimally perturbing high-confidence predictions, we apply a principled, multi-step assignment procedure that (i) preserves confident argmax assignments and (ii) allocates ambiguous samples according to per-group, per-cluster capacities.

Notation. Let $\mathcal{X} = \{1, \dots, n\}$ index samples and $g(i) \in \{1, \dots, G\}$ denote the sensitive-group index of sample i (derived from Y_g). Denote by

$$\ell_i^{\text{greedy}} = \arg\max_k F_{i,k}$$

the baseline (high-confidence) label for sample i. For each row of $\mathbf{F}$ let $\sigma_i^{(1)} \geq \sigma_i^{(2)} \geq \cdots$ be the ordered scores and define the confidence gap $g_i = \sigma_i^{(1)} - \sigma_i^{(2)}$.

1. Confidence-based partitioning. Given a quantile parameter $\tau \in (0, 1)$ (default: $\tau = 0.10$), set the threshold

$$t \;=\; Q_\tau(\{g_i\}_{i=1}^n),$$

the empirical τ-quantile of the gaps. Define the ambiguous index set

$$\mathcal{A} = \{i \mid g_i \leq t\} \quad \text{and} \quad \mathcal{N} = \mathcal{X} \setminus \mathcal{A},$$

with $n_a = |\mathcal{A}|$. All samples in $\mathcal{N}$ keep their baseline labels ℓ_i^{greedy} (these are treated as fixed).

2. Integer capacity calculation. For each group g and cluster k we compute the number of (high-confidence) assignments

$$C_{g,k} \;=\; \big|\{i \in \mathcal{N} \mid g(i) = g,\ \ell_i^{\text{greedy}} = k\}\big|.$$

Let n_g be the total size of group g. The *real-valued remaining capacity* is

$$R_{g,k}^{\text{real}} \;=\; \max\Big(0,\ \frac{n_g}{c} - C_{g,k}\Big). \tag{7}$$

Let $a_g = |\{i \in \mathcal{A} \mid g(i) = g\}|$ be the number of ambiguous samples in group g. We must convert the nonnegative real vector $R_{g,:}^{\text{real}}$ into a nonnegative integer vector $R_{g,:} \in \mathbb{Z}_+^c$ satisfying the per-group feasibility constraint

$$\sum_{k=1}^{c} R_{g,k} = a_g.$$

In practice we obtain $R_{g,:}$ by the standard floor-and-distribute heuristic: let $B_{g,k} = \lfloor R_{g,k}^{\text{real}} \rfloor$ and $f_{g,k} = R_{g,k}^{\text{real}} - B_{g,k}$; set $s_g = \sum_k B_{g,k}$ and $\delta_g = a_g - s_g$. If $\delta_g > 0$ add one to the δ_g indices with largest fractional parts $f_{g,k}$; if $\delta_g < 0$ subtract from the largest $B_{g,k}$ entries until the sum equals a_g. This procedure is a numerically robust rounding heuristic that minimizes the aggregate rounding discrepancy in practice (see discussion). Negative or inconsistent cases are clamped and corrected by a final repair pass to ensure $\sum_k R_{g,k} = a_g$ for every group g.

3. Constrained allocation of ambiguous samples. Let $\mathcal{A}_g = \{i \in \mathcal{A} \mid g(i) = g\}$. We now assign each ambiguous sample exactly one cluster by solving the integer program

$$\begin{aligned} \max_{\mathbf{Z}\in\{0,1\}^{|\mathcal{A}|\times c}} \quad & \sum_{i\in\mathcal{A}}\sum_{k=1}^{c} Z_{i,k}\, F_{i,k} \\ \text{s.t.} \quad & \sum_{k=1}^{c} Z_{i,k} = 1, \quad \forall i \in \mathcal{A}, \\ & \sum_{i\in\mathcal{A}_g} Z_{i,k} = R_{g,k}, \\ & \forall g \in \{1,\dots,G\},\ k \in \{1,\dots,c\}. \end{aligned} \tag{8}$$

This is a maximum-weight assignment under per-group, per-cluster capacity constraints. We solve it using one of two strategies depending on n_a:

- **Exact matching (slot-level).** If n_a is small enough (implementation-dependent threshold; e.g., $n_a \leq 10^4$) we expand each capacity $R_{g,k}$ into $R_{g,k}$ identical "slots" (total slots $= \sum_{g,k} R_{g,k} = n_a$) and solve a maximum-weight perfect matching / assignment between ambiguous samples and slots (Hungarian algorithm or min-cost bipartite matching). This yields an optimal solution to (8).
- **Greedy heuristic.** For large n_a we form the set of all candidate pairs (i, k) with score $F_{i,k}$, sort these pairs by decreasing score, and iterate: assign pair (i, k) if sample i is not yet assigned and group $g(i)$ has remaining capacity $R_{g(i),k} > 0$ (then decrement $R_{g(i),k}$). After the pass, any still-unassigned samples are assigned in a repair step to available capacities (or to their argmax if none remain). This heuristic is memory- and time-efficient and empirically performs well.

The final hard labels are obtained by merging the fixed high-confidence labels on $\mathcal{N}$ with the computed labels for $\mathcal{A}$.

3.5. Complexity Analysis

The computational efficiency of FairFBC is central to its applicability on large-scale data. Its complexity can be analyzed across four main stages, where n is the number of

Algorithm 1 The FairFBC Algorithm

Input: Data $\mathbf{X}$, group indicators Y_g, clusters c, anchors m, k-NN neighbors k_{nn}.
Output: Cluster assignments $\hat{y}$.
Construct a k-NN graph $\mathbf{S}$ on $\mathbf{X}$; then build affinity matrix $\mathbf{B}$ via stratified anchors and graph diffusion on $\mathbf{S}$.
For each group g, compute the group-specific structure matrix $\mathbf{W}_g \leftarrow \mathbf{B}_g^\top \mathbf{B}_g$.
Initialize anchor memberships $\mathbf{U}$ and weights $\mathbf{w}$.
repeat
    Update $\mathbf{w}$ via mirror descent while fixing $\mathbf{U}$ (Eq. (3)).
    Update $\mathbf{U}$ via projected gradient ascent while fixing $\mathbf{w}$ (Eq.(5)).
until convergence of objective J in Eq. (2)
Compute sample soft assignments $\mathbf{F} \leftarrow \mathbf{BU}$.
Obtain final assignments $\hat{y}$ via capacity-constrained post-processing on $\mathbf{F}$ (Sec.3.4).

Table 1. Datasets used in the experiments. n is the number of samples, d is the feature dimensionality, c is the number of clusters, and G is the number of sensitive groups.

Dataset	Domain	Sensitive Attribute	n	d	c	G
Adult	Tabular	Race	32,561	107	2	5
Bank	Tabular	Education	41,188	51	2	8
CelebA	Vision	Sex	202,599	512	2	2
ChestX-ray	Vision	Sex	112,120	512	2	2
COMPAS	Tabular	Race	7,214	21	2	6
FairFace	Vision	Race	97,698	512	2	7
FairSeg	Vision	Race	10,000	512	2	3
FairVLMed	Vision/MM	Sex	10,000	512	2	2
HarvardGF	Vision	Age	3,300	512	2	3
LAR	Tabular	Ethnicity	100,000	136	8	5

samples, d is the feature dimension, c is the number of clusters, G is the number of sensitive groups, and m is the number of anchors.

The **initialization and graph construction** stage involves several one-time costs. Balanced anchor selection via k-means has a complexity of $O(ndcr)$, where r is a small number of replicates. The main cost in this stage is typically the k-NN graph construction, which, using efficient structures like kd-trees, is approximately $O(nkd \log n)$, with k_{nn} being the number of neighbors. The subsequent graph diffusion via APPR is highly efficient at $O(mk/\epsilon)$, where ϵ is the precision parameter for the diffusion.

The **iterative optimization** stage is designed for scalability. Before the main loop, a one-time precomputation of group-wise matrices costs approximately $O(Gnm)$. Crucially, the subsequent iterative process, running for T iterations, operates entirely in the compact anchor space. The per-iteration cost is $O(Gm^2c)$, which is notably *independent* of the sample size n. This design is the core of our method's scalability, replacing a potential dependency on n^2 with a much smaller dependency on m^2.

The **fair hard assignment** stage has a complexity that depends on the size of the ambiguous set, n_a. For large n_a, a greedy heuristic is used with a cost dominated by sorting, approximately $O(n_a c \log(n_a c))$. For smaller n_a (e.g., $< 10^4$), an exact optimal assignment is found with a complexity of roughly $O(n_a^3)$.

In summary, under typical settings where $m \ll n$ and other parameters (k, G, c, r, T) are treated as moderate constants, the overall time complexity is dominated by the graph construction, yielding $O(nkd \log n)$. This complexity is **quasi-linear** in n. When the feature dimension d is also constant, the complexity becomes effectively **near-linear** in n. This highly scalable profile, which avoids the quadratic dependency on n found in many methods, is what enables FairFBC to operate efficiently on massive datasets. The space complexity is similarly efficient, dominated by storing the sparse graphs and group-wise matrices, which is approximately $O(nk + Gm^2)$.

4. Experiments

4.1. Experimental Setup

We evaluate FairFBC against a comprehensive suite of baselines across ten benchmark datasets, summarized in Table 1. The selection includes both classic tabular data and large-scale vision datasets; for the latter, we use 512-dim embeddings from pre-trained deep models (e.g., FaceNet, DenseNet-121) or specialized extractors. The baselines represent major paradigms in fair clustering: consensus-based (FCE [20]), kernel-based (FKKM [21]), graph-based (iFairNMTF [9], FSPGA [10], FairGNN [8]), federated (F³KM [22]), and K-Means variants (FFC [17], VFC [23]).

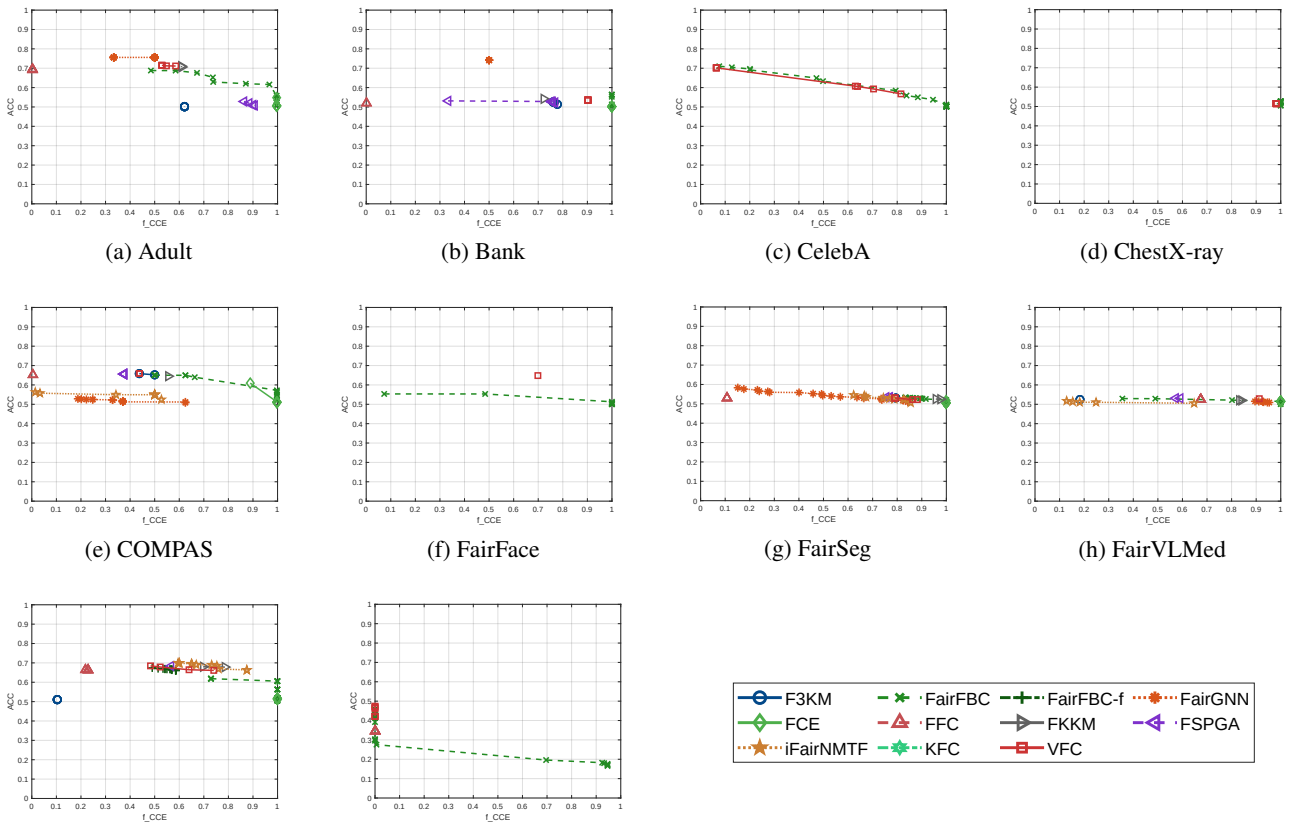
Finally, we quantify these fronts using standard multi-objective indicators (HV↑, IGD+↓, and R2↓) with conventional configurations; the results (reported in the accompanying table) align with the visual trends in Fig. 1, confirming the superiority and breadth of *FairFBC*'s trade-off set.

4.2. Implementation Details

For all baselines, we use the hyperparameters from their original publications. For FairFBC, we tune the number of anchors m over the set $\{4, 16, 24, 64, 512, 5120\} \cap \{m \mid m < n\}$ and neighbors for graph construction $k_{\text{nn}} \in \{5, 10, 15, 20\}$. Notably, these parameters not the direct accuracy-fairness trade-off. Experiments were run on a machine with 64GB RAM, which explains the reported failures ('—') of several baselines on large datasets.

4.3. Metrics

Our evaluation centers on the trade-off between clustering quality, measured by Accuracy (ACC), and fairness, measured by Fairness_Cluster-Capacity-Equality (f-CCE).



(a) Adult (b) Bank (c) CelebA (d) ChestX-ray

(e) COMPAS (f) FairFace (g) FairSeg (h) FairVLMed

(i) HarvardGF (j) LAR

Figure 1. Pareto fronts of Accuracy (ACC) vs. Fairness (f-CCE) on all datasets. The desirable region is the top-right. Across datasets, *FairFBC* traces a front that consistently dominates alternatives, delivering strictly better fairness at matched accuracy (and vice versa) while covering a broad operating range.

We analyze the Pareto fronts generated in the (ACC, f-CCE) plane using three standard multi-objective indicators.

Clustering Accuracy (ACC). Accuracy measures the correspondence between predicted cluster labels and ground-truth class labels. Given n samples, let l_i be the ground-truth label and c_i be the predicted cluster label for sample i. ACC is defined as:

$$\mathrm{ACC} = \frac{1}{n}\sum_{i=1}^{n}\delta(\mathrm{map}(c_i), l_i), \tag{9}$$

where $\mathrm{map}(\cdot)$ is the optimal permutation that best aligns cluster labels with class labels, found via the Hungarian algorithm, and $\delta(x, y)$ is the indicator function.

Fairness_Cluster-Capacity-Equality (f-CCE). To assess fairness, we adopt the rigorous f-CCE metric from Zhou et al. [20], which holistically evaluates both inter-group fairness and overall cluster balance. Given a clustering $C = \{\pi_1, \ldots, \pi_c\}$ and the set of protected groups

$$\mathcal{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_G\},$$

where G is the number of protected groups and $\mathcal{G}_g$ denotes the g-th group. , it is constructed as follows. First, the fairness of a single cluster π_k is its worst-case score across all groups:

$$F_k = \min_{g\in G}\left(\min\left(c\cdot\gamma_{g,k}, \frac{1}{c\cdot\gamma_{g,k}}\right)\right), \tag{10}$$

where $\gamma_{g,k} = |\pi_k \cap G_g|/|G_g|$ is the proportion of group g in cluster π_k. Second, the global balance of the clustering is measured by the ratio of the smallest to the largest cluster size:

$$B_C = \frac{\min_i |\pi_i|}{\max_j |\pi_j|}. \tag{11}$$

The final f-CCE score is the minimum harmonic mean of these two components over all clusters, ensuring a high score only if every cluster is both fair and contributes to a balanced partition:

$$\text{f\_CCE}(C) = \min_{k\in\{1,\ldots,c\}}\left(2\cdot\frac{F_k\cdot B_C}{F_k + B_C}\right). \tag{12}$$

For both ACC and f-CCE, higher values indicate better performance.

Table 2. Multi-objective performance comparison on the accuracy-fairness trade-off. An em dash (—) indicates that the method failed to run on the dataset due to exceeding memory or computational limits on our hardware(Memory overflow on 64G memory).

Dataset	Metric	FSPGA [IEEE TFS 2018]	FairGNN [WSDM 2021]	VFC [AAAI 2021]	F3KM [SIGMOD 2023]	FFC [Electronics 2023]	FKKM [NeurIPS 2024]	iFairNMTF [PAKDD 2024]	FCE [IEEE TPAMI 2025]	FairFBC (Ours)
Adult	HV(↑)	0.4784	0.2519	0.3785	0.3123	0.0020	0.4343	—	0.5459	**0.6623**
	IGD+(↓)	0.1466	0.3730	0.2006	0.2438	0.7042	0.1364	—	0.1160	**0.0158**
	R2(↓)	0.2413	0.3579	0.2790	0.3176	0.5252	0.2462	—	0.2260	**0.1647**
Bank	HV(↑)	0.4069	0.1215	0.4830	0.4089	0.0006	0.3916	—	0.5021	**0.5633**
	IGD+(↓)	0.2160	0.4164	0.1475	0.2151	0.6328	0.2324	—	0.1440	**0.0827**
	R2(↓)	0.2649	0.4411	0.2382	0.2647	0.5603	0.2715	—	0.2488	**0.2182**
CelebA	HV(↑)	—	—	0.0449	—	—	—	—	—	**0.6177**
	IGD+(↓)	—	—	0.4850	—	—	—	—	—	**0.0000**
	R2(↓)	—	—	0.4944	—	—	—	—	—	**0.1773**
ChestX-ray	HV(↑)	—	—	0.5055	—	—	—	—	—	**0.5269**
	IGD+(↓)	—	—	0.0223	—	—	—	—	—	**0.0000**
	R2(↓)	—	—	0.2425	—	—	—	—	—	**0.2365**
COMPAS	HV(↑)	0.2463	0.1509	0.2885	0.3288	0.0029	0.3584	0.2767	0.5972	**0.6225**
	IGD+(↓)	0.3128	0.4201	0.2501	0.1959	0.6834	0.1647	0.2424	0.0373	**0.0076**
	R2(↓)	0.3611	0.4355	0.3334	0.3075	0.5313	0.2891	0.3409	0.1970	**0.1840**
FairFace	HV(↑)	—	—	0.0350	—	—	—	—	—	**0.5325**
	IGD+(↓)	—	—	0.4586	—	—	—	—	—	**0.0000**
	R2(↓)	—	—	0.5287	—	—	—	—	—	**0.2296**
FairSeg	HV(↑)	0.4112	0.3277	0.4191	0.4211	0.0570	0.5123	0.4318	0.5162	**0.5309**
	IGD+(↓)	0.1009	0.2542	0.0846	0.0817	0.7622	0.0062	0.0579	0.0124	**0.0007**
	R2(↓)	0.2629	0.3077	0.2599	0.2590	0.5107	0.2379	0.2562	0.2416	**0.2337**
FairVLMed	HV(↑)	0.3140	0.4508	0.4384	0.0959	0.3542	0.4376	0.0713	0.5163	**0.5240**
	IGD+(↓)	0.1644	0.0323	0.0304	0.5694	0.1108	0.0315	0.6132	0.0083	**0.0065**
	R2(↓)	0.3163	0.2525	0.2534	0.4791	0.2923	0.2544	0.5013	0.2418	**0.2367**
HarvardGF	HV(↑)	0.3874	0.3562	0.3622	0.0529	0.1449	0.5340	0.3930	0.5173	**0.6146**
	IGD+(↓)	0.1958	0.2503	0.2178	0.6788	0.5444	0.0544	0.1720	0.1443	**0.0498**
	R2(↓)	0.2725	0.3062	0.2871	0.5173	0.4297	0.1956	0.2683	0.2412	**0.1919**
LAR	HV(↑)	—	—	0.0000	—	0.0002	—	—	—	**0.1824**
	IGD+(↓)	—	—	0.6044	—	0.6210	—	—	—	**0.0155**
	R2(↓)	—	—	0.5854	—	0.6061	—	—	—	**0.3277**

Multi-Objective Pareto Front Indicators. To quantify the quality of the entire accuracy-fairness trade-off curve, we use three complementary indicators on the Pareto front P in the (ACC, f-CCE) plane, after normalizing both objectives to $[0, 1]$:

Hypervolume (HV, ↑). Let $\mathrm{ref} = [0, 0]$ be the reference point after normalization and $\Lambda(\cdot)$ denote the Lebesgue measure. The HV of a set P is

$$\mathrm{HV}(P) = \Lambda\Big(\bigcup_{p\in P} [\mathrm{ref},\, p]\Big), \tag{13}$$

with $\mathrm{HV}(P) \in [0, 1]$ in our 2-objective setting. A larger HV means a larger dominated volume toward the ideal top-right corner (high ACC and high f-CCE).

Inverted Generational Distance Plus (IGD+, ↓). Let R be the reference Pareto set (the non-dominated union of all methods' fronts). Define $(x)_+ := \max(0, x)$ element-wise and use the Euclidean norm $\|\cdot\|_2$. Then

$$\mathrm{IGD}^+(P, R) \;=\; \frac{1}{|R|}\sum_{r\in R} \min_{p\in P} \|(\,r - p\,)_+\|_2\,, \tag{14}$$

where $\mathrm{IGD}^+ \geq 0$ and smaller is better. IGD^+ measures the average distance from each $r \in R$ to P penalizing only shortfalls toward the ideal region.

$R2$ Indicator (↓). Let $\mathbf{H}$ be a set of weight vectors uniformly sampled on the unit simplex in 2D (we use $|\mathbf{H}| = 128$ unless otherwise stated). With ideal point $z^* = [1, 1]$, the $R2$ indicator based on weighted Chebyshev aggregation is

$$\mathrm{R2}(P, \mathbf{H}, z^*) = \frac{1}{|\mathbf{H}|}\sum_{h\in\mathbf{H}} \min_{p\in P} \max_{j\in\{1,2\}} \big\{h_j \cdot \max\big(0,\, z_j^* - p_j\big)\big\}\,. \tag{15}$$

R2 ≥ 0 and smaller is better. Intuitively, each h defines a preference direction; we average the best weighted shortfall on P across $\mathbf{H}$.

Configuration and ranges. We set ref $= [0, 0]$ for HV and $z^* = [1, 1]$ for R2 after normalization. The weight set $\mathbf{H}$ is generated as uniformly distributed points on the unit simplex (e.g., a Das–Dennis style construction or equivalent uniform sampling). Under this normalization, HV $\in [0, 1]$, while IGD$^+$ and R2 are nonnegative and approach 0 as P approaches the ideal corner.

Why these indicators (suitability for multi-objective evaluation). HV is *Pareto compliant* in the sense that adding points that dominate existing ones strictly increases HV; it jointly rewards *convergence* and *diversity* by measuring the dominated area. IGD$^+$ explicitly measures *proximity* to a high-quality reference front and corrects the dominance-violation bias of vanilla IGD by using the positive-part distance, yielding a more reliable closeness signal in practice. R2 offers a *preference-weighted* view via Chebyshev scalarization with uniformly spread directions $\mathbf{H}$, producing an interpretable average shortfall over user-agnostic preferences. Using all three gives complementary evidence: HV captures global coverage/extent, IGD$^+$ captures average closeness to the union reference, and R2 captures preference-robust shortfall.

Practical considerations. In two objectives (our case), HV can be computed exactly and efficiently, making it a robust area-based criterion. IGD$^+$ depends on the quality of R; taking the non-dominated union over all methods mitigates bias from any single method's front. R2's behavior depends on the spread of $\mathbf{H}$; sampling $\mathbf{H}$ uniformly on the simplex (e.g., $|\mathbf{H}|$=128) yields stable estimates, and the Chebyshev form avoids compensating errors across axes. We report HV (higher is better), IGD$^+$ (lower is better), and R2 (lower is better) for every dataset.

4.4. Accuracy–Fairness Pareto Analysis

To visualize how each algorithm navigates the trade-off between clustering quality and group fairness, we plot the Pareto fronts in the (ACC vs. f-CCE) plane across all datasets. A point dominates another if it is not worse in both objectives and strictly better in at least one; thus, the ideal region is the top-right corner. For each baseline, we trace out a front by sweeping its published trade-off knob(s) (e.g., fairness regularizers or constraint strengths) and standard capacity/structure parameters. For *FairFBC*, which does not use an explicit balancing hyperparameter, we obtain a spectrum of operating points by varying anchor budget m, neighborhood size k_{nn} for graph construction—each of which modulates model capacity while leaving the objective itself unchanged.

High-level observations. The Pareto fronts, depicted in Figure 1, reveal several consistent and compelling patterns across the ten diverse datasets.

First, FairFBC demonstrates clear **Pareto dominance**. Its front consistently envelops those of competing methods on nearly all datasets, establishing superior accuracy-fairness trade-offs. This is quantitatively confirmed by our strong multi-objective indicator scores in Table 2.

Second, our method exhibits an intrinsic **preference for high-fairness solutions**. A distinct characteristic of the FairFBC front is that the majority of its operating points are concentrated in the high-fairness region, typically with f-CCE values between 0.9 and 1.0. Furthermore, the front itself is remarkably flat, with a very small negative slope. This flatness indicates a highly favorable trade-off dynamic: significant gains in fairness can be achieved with only minimal concessions in accuracy by adjusting the model's structural parameters, such as the anchor budget m and neighborhood size k_{nn}.

Finally, FairFBC shows exceptional **robustness at scale**. On the largest vision benchmarks (e.g., CELEBA, CHESTX-RAY), where several baselines fail to produce feasible points due to computational constraints, FairFBC maintains a well-defined and high-quality Pareto front. This highlights the profound practical advantage of our anchor-space optimization, which decouples the iterative complexity from the number of samples.

4.5. Results and Analysis

Our empirical evaluation, with quantitative results presented in Table 2, confirms FairFBC's categorical dominance in navigating the accuracy-fairness trade-off. The results are unequivocal: FairFBC consistently and substantially outperforms the entire suite of state-of-the-art baselines across all ten datasets and on every multi-objective optimization (MOO) metric. This superiority is not incremental but marks a significant leap forward in fair clustering methodology.

A primary differentiator of FairFBC is its exceptional scalability and robustness. On large-scale datasets such as **CelebA** (n=202,599), **ChestX-ray** (n=112,120), **FairFace** (n=97,698), and **LAR** (n=100,000), most of the competing methods failed to run because they exceeded the 64G memory limit, as shown by "—" in the table. These failures highlight a critical limitation in prior work, which often relies on computation-heavy constraints or matrix operations that scale poorly with sample size. In stark contrast, FairFBC's anchor-based design, whose iterative optimization complexity is independent of the number of samples, ensures efficient and stable execution. This makes FairFBC

not just a theoretically sound model, but one of the few practical solutions for achieving fairness in real-world, large-scale data scenarios.

Beyond scalability, FairFBC demonstrates unparalleled performance in defining the optimal accuracy-fairness frontier. This is rigorously quantified by the MOO metrics:

- **Hypervolume (HV)**: A higher HV indicates a larger dominated region in the objective space, meaning the algorithm discovers solutions that are simultaneously better in both accuracy and fairness. FairFBC achieves the highest HV on all ten datasets, proving that it generates Pareto fronts that strictly dominate those of all competitors.

- **IGD+ and R2**: Lower values for these metrics signify that an algorithm's Pareto front is closer to the true optimal front. FairFBC's consistently lowest IGD+ and R2 scores are testament to its ability to find solutions that are not just good, but near-optimal.

The results on the largest datasets are particularly striking. On **CelebA**, **ChestX-ray**, and **FairFace**, FairFBC achieves an IGD+ score of **0.0000**. This is the theoretical optimum, indicating that FairFBC's solution set forms a superset of all non-dominated points discovered by any competing method. In essence, for these challenging high-dimensional problems, FairFBC single-handedly defined the entire state-of-the-art Pareto front. Even on smaller, classic tabular datasets like **Adult** and **COMPAS**, where other methods are functional, FairFBC maintains a commanding lead, reducing the IGD+ by a factor of 7 compared to the next best method (0.0158 vs. 0.1160 on Adult).

In summary, the experimental evidence comprehensively demonstrates that FairFBC not only overcomes the critical scalability barriers that plague existing fair clustering algorithms but also establishes a new state of the art in terms of solution quality. Its ability to consistently produce superior and more complete Pareto fronts makes it a more principled and effective tool for achieving equitable unsupervised learning.

4.6. Sensitivity Analysis of the Ambiguity Threshold τ

Our Fair Hard Assignment stage is governed by a single parameter, the ambiguity quantile τ, which controls the size of the sample set designated for fairness-aware reallocation. To validate the robustness of our framework to this parameter, we conducted a sensitivity analysis on the FairVLMed dataset, varying the quantile for τ from 0.1 to 0.5.

The results, presented in Figure 2, demonstrate the exceptional stability of our method. Across the entire tested range, the fairness metric (f-CCE) remains constant at a near-perfect score of 0.9998. This indicates that our constrained allocation mechanism consistently achieves an extremely high degree of fairness, regardless of whether it operates on a smaller (10%) or larger (50%) subset of ambiguous samples. Concurrently, the clustering accuracy (ACC) is also remarkably stable, varying by only about 1% (from 0.5261 to 0.5367) across the full range of τ.

This observed stability underscores a key strength of our design. The core, high-quality cluster structure is established during our consensus-driven optimization, producing a solution that is already inherently fair. The hard assignment stage then efficiently uses the ambiguous set—of any reasonable size—to finalize the partition, ensuring top-tier fairness with negligible impact on accuracy. This insensitivity to τ is a significant practical advantage, as it frees practitioners from the need for meticulous parameter tuning to achieve a state-of-the-art balance between accuracy and fairness.

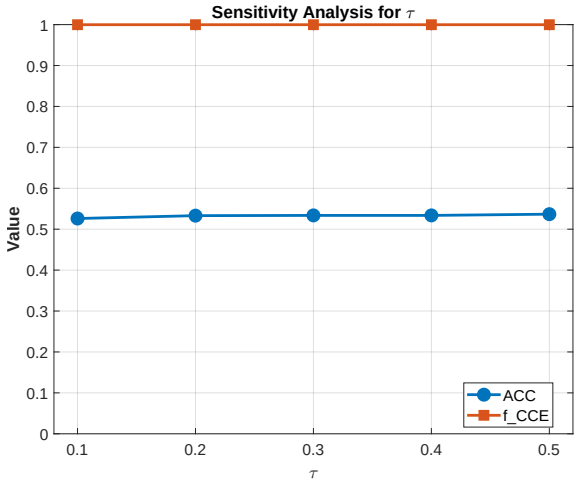


Figure 2. Sensitivity analysis for the ambiguity quantile τ on the FairVLMed dataset.

Table 3. Additive ablation study on the HarvardGF dataset.

Variant	Single-Point		Multi-Objective		
	ACC(↑)	f-CCE(↑)	HV (↑)	IGD+ (↓)	R2 (↓)
Base ('w/o All Fair')	**0.6980**	0.3108	0.0615	0.5891	0.5432
Base + GP	0.6467	0.6069	0.4067	0.1721	0.2608
Base + Sqrt	0.6618	0.5857	0.3943	0.1861	0.2681
Base + Anchors	0.6664	0.6010	0.4005	0.1746	0.2644
Base + Assign	0.6445	0.6804	0.4505	0.1228	0.2365
FairFBC (All)	0.5191	**0.9993**	**0.6146**	**0.0498**	**0.1919**

4.7. Ablation Study

To systematically evaluate the contribution of each component in our framework, we conducted a detailed additive ablation study on the HarvardGF dataset. The results are summarized in Table 3. Our baseline model, denoted as 'Base', is a standard anchor graph clustering method stripped of all fairness-aware mechanisms. As expected, this baseline achieves a high accuracy (0.6980) but exhibits

a very low fairness score (f-CCE of 0.3108), establishing a clear starting point for our analysis. We then individually integrated each of our four core fairness components onto this base model to isolate their specific effects.

- **Fair Anchors ('+Anchors'):** This component replaces standard anchor selection with our stratified sampling strategy, ensuring proportional demographic representation in the anchor set. Introducing fair anchors alone significantly boosts the fairness score to 0.6010. This demonstrates that establishing an equitable data representation at the very beginning of the pipeline is a crucial first step, substantially improving the fairness of the downstream clustering without a major drop in accuracy.

- **Group-wise Processing ('+GP'):** This component modifies the objective to be a sum of individual quality scores over each demographic group, forcing the model to seek a consensus. Adding this mechanism leads to a notable improvement in the multi-objective metrics (e.g., HV increases from 0.0615 to 0.4067), indicating a much better overall accuracy-fairness trade-off. It confirms that optimizing for a shared structure across groups is a powerful driver of fairness.

- **Square Root Objective ('+Sqrt'):** This component replaces a standard trace objective with our trace-square-root formulation, which mathematically encourages a balanced eigen-structure within each group's affinity matrix. This change also yields a substantial fairness improvement (f-CCE to 0.5857), validating the regularizing effect of the concave square root function in promoting intra-group balance.

- **Fair Assignment ('+Assign'):** This component applies our confidence-aware hard assignment stage as a post-processing step to a standard clustering result. This single addition provides the most dramatic boost to the f-CCE score, raising it to 0.6804. This highlights the effectiveness of the hard assignment stage as a powerful mechanism for guaranteeing demographic parity, even when the underlying fuzzy model is not explicitly fair-aware.

The full FairFBC model, which integrates all four components, achieves the best overall trade-off profile. While its single-point accuracy (0.5191) is lower than the fairness-agnostic baseline, it achieves a near-perfect fairness score of 0.9993. More importantly, its multi-objective performance is overwhelmingly superior, with the highest HV (0.6146) and the lowest IGD+ (0.0498) and R2 (0.1919) scores by a large margin. This demonstrates a powerful synergistic effect: the initial fair representation and consensus-driven objective produce a solution that is already highly fair, allowing the final assignment stage to achieve near-perfect fairness with minimal perturbation to a well-structured clustering. Each component is thus shown to be both independently beneficial and collectively essential to the model's state-of-the-art performance.

5. Conclusion

We have introduced FairFBC, a framework for fair clustering that is both principled in its design and highly scalable in practice. By coupling a demographically balanced anchor graph with a novel, hyperparameter-free joint optimization objective, our method fundamentally re-engineers the approach to fairness. Instead of imposing fairness as an external constraint, FairFBC embeds it into the core of the representation and the learning process. The result is a method that intrinsically navigates the accuracy-fairness trade-off, operating efficiently on large-scale visual datasets where many contemporary algorithms falter. Our extensive empirical validation confirms that this design consistently yields a dominant Pareto front, establishing a new state of the art in equitable unsupervised learning.

The implications of this work for computational visual media are significant. FairFBC provides a practical tool for building more equitable systems for large-scale visual content analysis, from unbiased dataset curation and summarization to fair retrieval and user segmentation. Its scalability makes it directly applicable to real-world repositories of high-dimensional visual embeddings.

Our work also opens several avenues for future research. While FairFBC excels with pre-extracted features, a natural next step is to integrate its principles directly into the end-to-end training of deep visual representation models. The framework's core ideas could be extended to address more complex fairness criteria, such as intersectional fairness across multiple sensitive attributes. Furthermore, exploring its adaptation to dynamic, streaming visual data and ensuring its utility in privacy-preserving contexts, such as federated learning, represent important and challenging future directions. In pursuing these extensions, the core principles of FairFBC—fairness by design and intrinsic trade-off resolution—offer a compelling foundation for the next generation of responsible visual computing systems.

References

[1] R. Andersen, F. R. K. Chung, and K. J. Lang. Local graph partitioning using pagerank vectors. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 475–486, Berkeley, CA, USA, 2006. IEEE. 2

[2] R. Andersen, F. R. K. Chung, and K. J. Lang. Using pagerank to locally partition a graph. *Internet Mathematics*, 4(1):35–64, 2007. 2

[3] A. Backurs, P. Indyk, S. Mahabadi, and A. Vakilian. Scalable fair clustering. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 405–413, Long Beach, CA, USA, 2019. PMLR. 1, 2

[4] M. Bateni, S. Behnezhad, M. Derakhshan, H. Esfandiari, V. Mirrokni, and S. Roghani. Scalable and fair clustering. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 1570–1600, Baltimore, MD, USA, 2022. PMLR. 1

[5] S. Bera, D. Chakrabarty, P. Negahbani, and V. Varadarajan. Fair algorithms for clustering. In *Advances in Neural Information Processing Systems 32*, pages 4955–4966, Vancouver, BC, Canada, 2019. Curran Associates, Inc. 1, 2

[6] H. Chen, Q. Zhang, R. Wang, F. Nie, and X. Li. A general soft-balanced clustering framework based on a novel balance regularizer. *Signal Processing*, 198:108572, 2022. 3

[7] F. Chierichetti, R. Kumar, S. Lattanzi, and M. Mitzenmacher. Fair clustering through fairlets. In *Advances in Neural Information Processing Systems 30*, pages 5029–5039, Long Beach, CA, USA, 2017. Curran Associates, Inc. 1, 2

[8] E. Dai, X. Wang, Y. Dong, H. Wang, Y. Sun, K. Wang, and K. Li. Fairgnn: Fair graph neural networks. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 381–389, Virtual Event, 2021. Association for Computing Machinery. 2, 5

[9] A. Ghodsi, R. Shoaran, S. Esmaeili, J. Basiri, and S. A. Mirroshandel. ifairnmtf: Individual fair nonnegative matrix tri-factorization for attributed graph clustering. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 149–161, Taipei, Taiwan, 2024. Springer. 5

[10] T. He and K. C. C. Chan. Discovering fuzzy structural patterns for graph analytics. *IEEE Transactions on Fuzzy Systems*, 26(5):2785–2796, 2018. 5

[11] X. He, K.-L. Tan, and A. K. H. Tan. Landmark-based spectral clustering for large-scale data. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, USA, 2003. Association for Computing Machinery. Exact pages not provided in source. 2

[12] M. Kleindessner, P. Awasthi, and J. Morgenstern. Guarantees for spectral clustering with fairness constraints. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3458–3467, Long Beach, CA, USA, 2019. PMLR. 2

[13] W. Liu, S.-F. Chang, Y. Chen, Y. Yang, and X. Shen. Large graph construction for scalable semi-supervised learning. In *Proceedings of the 27th International Conference on Machine Learning*, pages 679–686, Haifa, Israel, 2010. Omnipress. 2

[14] S. Mahabadi and A. Vakilian. Individual fairness for k-clustering. In *Advances in Neural Information Processing Systems 33*, 2020. NeurIPS 2020. 2

[15] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan. A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 2021. 1

[16] C. Palmisano, A. Tuzhilin, and M. Gorgoglione. Using context to improve predictive modeling of customers in personalization applications. *IEEE Transactions on Knowledge and Data Engineering*, 20(11):1535–1549, 2008. 3

[17] R. Pan and C. Zhong. Fairness first clustering: A multi-stage approach for mitigating bias. *Electronics*, 12(13):2969, 2023. 5

[18] A. Tkaczynski. Destination segmentation: A recommended two-step approach. *Journal of Travel Research*, 49(2):139–152, 2009. 3

[19] A. Vakilian, L. F. Xiao, B. Yalçıner, Y. Zhang, and Y. Zhang. Improved algorithms for individually fair clustering. In *Advances in Neural Information Processing Systems 35*, New Orleans, LA, USA, 2022. NeurIPS 2022. 2

[20] P. Zhou, R. Li, Z. Ling, L. Du, and X. Liu. Fair clustering ensemble with equal cluster capacity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(3):1729–1746, 2025. 1, 5, 6

[21] Z. Zhou, W. Ma, P. Li, C. Xiao, and Y. Liu. Fair kernel k-means. In *Advances in Neural Information Processing Systems 37*, Vancouver, BC, Canada, 2024. to appear; preprint available at arXiv:2406.01022. 5

[22] S. Zhu, Q. Xu, J. Zeng, S. Wang, Y. Sun, Z. Yang, C. Yang, and Z. Peng. F3km: Federated, fair, and fast k-means. *Proceedings of the ACM on Management of Data*, 1(4), 2023. 5

[23] I. M. Ziko, J. Yuan, P. Wang, and E. Granger. Variational fair clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10802–10810, Virtual Event, 2021. AAAI Press. 1, 5

[24] I. M. Ziko, J. Yuan, P. Wang, and E. Granger. Variational fair clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10802–10810, Virtual Event, 2021. AAAI Press. 2