FAIR DIVERSITY MAXIMIZATION VIA LOCAL SEARCH

Anonymous authors
Paper under double-blind review

ABSTRACT

Diversity maximization is a fundamental optimization problem with applications in machine learning, data summarization, information retrieval, and recommendation systems. In many such applications, the data are partitioned into groups, and the selected subset must satisfy prescribed group quotas. We study Fair Diversity Maximization: given a set of points in a metric space partitioned into m groups, the goal is to select exactly k_i points from each group i while maximizing the minimum pairwise distance among the selected points.

The best previously known approximation guarantee is $m + 1$, which grows linearly with the number of groups. We show that this dependence on m is not fundamental. We present a new local-search framework that yields a 4-approximation for any constant number of groups, with no restrictions on the metric space or on the size of the selected set. To the best of our knowledge, this is the first constant-factor approximation whose guarantee is independent of the number of groups in this general setting. Our framework maintains all group quotas exactly while progressively eliminating violations of the diversity objective. We further develop a specialized algorithm for two groups that achieves a 2-approximation, improving the previous best factor of 3. This factor is optimal: unless $\mathrm{P} = \mathrm{NP}$, no polynomial-time algorithm can achieve an approximation factor strictly better than 2, even for the unconstrained case.

1 INTRODUCTION

The increasing scale and complexity of modern datasets across domains such as machine learning have made it increasingly important to extract small yet informative data summaries. Such subset-selection problems arise whenever processing, storing, annotating, or presenting the entire dataset is computationally expensive. A basic challenge is then to decide which points are worth keeping. One natural principle is to favor diversity—choosing elements that are distinct from one another to obtain a varied and nonredundant view of the data. This simple idea has led to a broad line of work on diversity-based subset selection, with applications ranging from data summarization and information retrieval to recommendation systems, active learning, feature selection, training-data curation, and in-context demonstration selection (Ahmed, 2019; Drosou & Pitoura, 2010; Abbar et al., 2013; Kunaver & Požrl, 2017; Zadeh et al., 2017; Sener & Savarese, 2018; Liu et al., 2024; Wang et al., 2025).

In the classical diversity maximization problem, also known as max-min diversity, the goal is to select a subset of a prescribed size that maximizes the minimum pairwise distance among the selected elements. This simple objective has been studied for decades with its roots in facility-location and dispersion problems (Moon & Chaudhry, 1984; Erkut & Neuman, 1989; Erkut, 1990; Ravi et al., 1991). Max-min diversification has also been used to construct diverse molecular libraries (Ashton et al., 2002), and has recently appeared in machine-learning applications such as data subset selection and the pruning of redundant visual tokens in large multimodal and video-language models (Fahrbach et al., 2025; Alvar et al., 2025; Fan et al., 2026).

In many applications, however, diversity alone is not enough. The objects in the underlying dataset often belong to different groups, and one may want the selected subset to contain a prescribed number of elements from each of them. These groups may correspond, for example, to demographic attributes, data sources, classes, genres, geographic regions, or time periods. An unconstrained diversity objective gives no guarantee that these groups will be adequately represented: a highly diverse solution may still contain very few or even entirely omit elements from a specific group.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

This issue arises naturally in applications such as training-data selection and pruning, where one may wish to preserve the representation of different classes or subpopulations, active learning, where only a limited number of points can be labeled, recommendation, where one may wish to control the representation of different genres or content providers, and data summarization, where the selected items may need to reflect different sources, categories, or temporal periods (Rolf et al., 2021; Hassan et al., 2025; Anahideh et al., 2022; Celis et al., 2018a; Abbassi et al., 2013; Mahabadi & Trajanovski, 2023).

In this paper, we study Fair Diversity Maximization. We are given a set of n elements in a metric space, partitioned into m disjoint groups, together with integers $k_1, \ldots, k_m$. The goal is to select exactly k_i elements from group i, for each $i \in [m]$, so that the minimum distance between any two selected elements is as large as possible. Thus, the problem combines exact group-representation constraints with a max-min diversity objective.

The diversity problem was initially studied by Moon & Chaudhry (1984) and Erkut & Neuman (1989). Erkut (1990) showed that the problem is NP-hard. Subsequently, Ravi et al. (1991) gave a polynomial-time 2-approximation in any metric space and showed that no approximation factor better than 2 is possible unless $\mathrm{P} = \mathrm{NP}$. Thus, the approximability of the unconstrained problem in general metric spaces has been completely understood for more than three decades.

The picture changes substantially once exact group-representation constraints are imposed. The fair version of diversity maximization was introduced by Moumoulidou et al. (2021). They presented a 4-approximation algorithm for the case of two groups, and extended their approach to obtain a $(3m-1)$-approximation for general m. Addanki et al. (2022) subsequently improved the approximation factor to $m + 1$ when m is constant.

1.1 Our Contributions

Prior to our work, the best known polynomial-time approximation guarantee for fair diversity maximization with a constant number of groups was $m + 1$, which grows linearly with the number of groups. We show that this dependence on m is not fundamental. We present a new local-search framework that yields constant factor approximation, independent of m, for any constant number of groups, with no restrictions on the metric space or on the size of the selected set.

Breaking the dependence on the number of groups. Existing approximation guarantees degrade with the number of groups. Our main result is a constant-factor approximation for any arbitrary constant number of groups. Specifically, we obtain a polynomial-time 4-approximation, giving the first constant-factor guarantee independent of m. Our main result is summarized in the following theorem.

Theorem 1. *For any constant number of groups m, Fair Diversity Maximization admits a polynomial-time 4-approximation.*

An optimal algorithm for two groups. We further consider the important special case of two groups, where the approximation guarantee can be improved to the optimal factor of 2. This improves upon the previous best factor of 3 and matches the optimal approximation factor for the unconstrained problem.

Theorem 2. *Fair Diversity Maximization with two groups admits a polynomial-time 2-approximation.*

The approximation factor 2 is the best possible, since achieving a factor better than 2 is NP-hard, even for the unconstrained problem.

A comparison of our results with the best previously known guarantees is provided in Table 1.

1.2 Technical Overview

Going beyond the cluster-and-flow approach. Both the previous $3m - 1$ and $m + 1$ approximation algorithms follow the same general paradigm. They first organize points into proximity clusters, with the intention of selecting at most one point from each cluster, and then use a flow computation to make this selection while satisfying the prescribed group counts exactly (Moumoulidou et al.,

Table 1: Summary of the results for fair diversity maximization.

NUMBER OF GROUPS	APPROXIMATION GUARANTEE	REFERENCE
$m = O(1)$	$3m - 1$	Moumoulidou et al. (2021)
	$m + 1$	Addanki et al. (2022)
	4	This work
$m = 2$	4	Moumoulidou et al. (2021)
	3	Addanki et al. (2022)
	2 (optimal)	This work

2021; Addanki et al., 2022). For this approach to work, the clustering must satisfy two competing requirements: points chosen from different clusters must be sufficiently far apart, so that selecting at most one point from each cluster guarantees diversity; at the same time, there must remain enough clusters for such a selection to satisfy every group quota exactly. The main difficulty is to satisfy these two requirements simultaneously. In both previous algorithms, doing so leads to clusters that may contain chains of nearby points spanning as many as m groups. As a result, the diameter of a cluster can grow linearly with m, which is precisely where the m-dependence enters their approximation guarantees.

Our approach, on the other hand, avoids trying to enforce diversity and fairness simultaneously in one shot. Instead, we begin with a solution that already satisfies all group quotas and gradually improve its diversity while preserving fairness throughout. The difficulty is that even a local exchange must preserve the exact group counts, and it is not clear whether an appropriate local modification always exists. Establishing the existence of such exchanges is the main technical challenge behind our framework.

A unified local-repair framework. Both our results are obtained through a unified local-search framework. We work with two distance scales $\alpha \leq \beta$, where α is the diversity we aim to achieve, while β is a stronger separation maintained between points belonging to the same group. Starting from a fair solution satisfying this within-group separation, we repeatedly choose a selected point q that has another selected point at distance less than α. We then replace q, together with any points that conflict with the replacement, by a carefully chosen set of new points. We call such a set a *q-repair*. Informally, a q-repair must be mutually compatible and must contain enough points from every group to compensate for all points from that group that are removed. Thus, applying a repair preserves both exact fairness and the within-group separation, while strictly decreasing the number of selected points that have at least one neighbor with distance less than α. Consequently, the local search terminates with a fair solution of diversity at least α.

Technical novelty. The main technical challenge is to prove that a suitable local improvement always exists. For any constant number of groups, this can be a delicate procedure, because inserting a single point may simultaneously force the removal of points from several different groups. We capture these interactions using *repair vectors*, which record the surplus or deficit in each group caused by an exchange. We show that a minimal collection of such vectors satisfying the required balance has a simple structure and is exactly balanced. A Steinitz-type vector-balancing argument then implies that this collection has size bounded solely as a function of m. Consequently, for constant m, a suitable local improvement can be found by exhaustive search in polynomial time.

The two-group case allows for a sharper argument. Since there is only one other group whose balance needs to be restored, we can greedily build a maximal auxiliary set from that group. The maximality of this set gives us a useful saturation property. We then compare this set with an optimal solution and use the disjointness of suitable neighborhoods around optimal points to show, by a counting argument, that one of these points can be used to form a valid local improvement. This leads to the tight 2-approximation.

1.3 Other Related Work

Several works have obtained stronger guarantees by restricting the parameters, relaxing the fairness constraints, or imposing additional structure on the metric. Moumoulidou et al. (2021) obtained a 5-approximation when m is constant and $k = o(\log n)$. Addanki et al. (2022) showed that a 2-approximation is possible when the fairness constraints are required only in expectation, and gave a 6-approximation while selecting at least a $(1 - \epsilon)$ fraction of the required number of elements from every group. For Euclidean metrics, they also showed that the problem is exactly solvable in one dimension and admits a $(1 + \epsilon)$-approximation in constant dimension.

Wang et al. (2023b) considered the more general setting in which each group has lower and upper representation bounds, obtaining a $5/(1-\epsilon)$-approximation in polynomial time when $m = O(1)$ and $k = o(\log n)$. Gao & Borodin (2024) improved the approximation guarantees under strengthened or relaxed metric assumptions, while Kurkure et al. (2024) developed constant-factor near-linear-time algorithms for Euclidean instances. More recently, Adriaens & Tatti (2025) obtained improved guarantees when only a constant number of representatives is required from each group, while Zheng et al. (2026) studied exact and metaheuristic methods for Fair Max-Min Diversity.

Beyond Max-Min Diversity, several other distance-based diversity objectives have been studied extensively. A prominent example is Max-Sum Diversity, which maximizes the sum of pairwise distances among the selected elements. Ravi et al. (1991) gave a 4-approximation for this objective, which was later improved to a 2-approximation by Hassin et al. (1997). Abbassi et al. (2013) proposed a local-search algorithm achieving a tight 2-approximation under partition-matroid constraints. Borodin et al. (2017) further generalized this setting by giving a 2-approximation for maximizing the sum of a monotone submodular function and Max-Sum diversity. Ceccarello et al. (2018; 2020) subsequently developed coreset-based algorithms for diversity maximization under matroid constraints.

A separate line of work has focused on diversity maximization in large-scale and restricted computational models. Indyk et al. (2014) introduced composable coresets with constant-factor guarantees for several diversity objectives, and Aghamolaei et al. (2015) subsequently improved the guarantees for several of these objectives. Diversity maximization has also been studied in MapReduce, streaming, and sliding-window settings (Ceccarello et al., 2017; Borassi et al., 2019). For Max-Min Diversity in particular, Haqi & Zarrabi-Zadeh (2023) obtained an almost-tight $(2+\epsilon)$-approximation in the massively parallel computation model, improving the previous factor of 6 due to Indyk et al. (2014). More recently, diversity maximization has also been considered in fully dynamic settings (Pellizzoni et al., 2025).

Scalability has also been studied in the presence of group-representation constraints. Mahabadi & Trajanovski (2023) developed coresets for fair Max-Sum and sum-of-nearest-neighbor diversity, while Wang et al. (2023a) designed one-pass streaming and sliding-window algorithms for Fair Max-Min Diversity. Most recently, Ahmadvand et al. (2026) developed composable coresets for Fair Diversity Maximization.

Fairness and representation constraints have also been studied under other selection objectives and notions of fairness. Celis et al. (2018a) studied fair determinantal data summarization, while Wang et al. Wang et al. (2021) considered fair representative subset selection under submodular objectives. More recently, Li & Wang (2025) studied diversity maximization under *individual* rather than group fairness, obtaining bicriteria approximation algorithms for Max-Min and Max-Sum diversity. Related fairness constraints have also been considered in multiwinner voting and submodular maximization (Celis et al., 2018b; Halabi et al., 2020), as well as in problems such as fair k-center clustering (Chen et al., 2016; Kleindessner et al., 2019; Jones et al., 2020), fair set cover (Dehghankar et al., 2024), and fair matching (García-Soriano & Bonchi, 2020).

2 Problem Formulation and Definitions

Let (X, d) be a metric space, where X is a set of n points, and $d : X \times X \to \mathbb{R}_{\geq 0}$ is a metric distance function. The notion of diversity is defined as follows.

Definition 1 (Diversity). *Given a subset $S \subseteq X$, the* diversity *of S, denoted by $\mathrm{div}(S)$, is the minimum pairwise distance between distinct elements in S, i.e.,*

$$\mathrm{div}(S) = \min_{p,q \in S, p \neq q} d(p, q).$$

Incorporating fairness constraints, the main problem studied in this paper is formally defined as follows.

Problem 1 (Fair Diversity Maximization). *Given a set X partitioned into m disjoint groups $X_1, X_2, \ldots, X_m$ and positive integers $k_1, k_2, \ldots, k_m$, where $k = \sum_{i=1}^{m} k_i$, a subset $S \subseteq X$ is called* fair *if $|S \cap X_i| = k_i$ for all $i \in [m]$. The goal is to find a fair subset $S \subseteq X$ of size k that maximizes $\mathrm{div}(S)$.*

In the remainder of this section, we introduce the notation and tools used in our algorithms and analyses. For each point $p \in X$, we denote the index of the group containing p with $c(p)$. For a more intuitive representation, we refer to $c(p)$ as the *color* of p. Whenever a set is indexed, for example S_i, this means $S \cap X_i$, i.e., the points in S with color i.

Definition 2 (Color Separation). *A set $S \subseteq X$ is β-color separated if for every color $i \in [m]$*

$$\mathrm{div}(S_i) \geq \beta$$

For set S and point q we define $S^{-q} = S \setminus \{q\}$. We further define open balls and conflict sets as tools to enforce diversity.

Definition 3 (Open Ball). *Let $\beta \geq \alpha > 0$ be fixed parameters. For each color $i \in [m]$ and point p we define*

$$B_i(p) = \begin{cases} \{v \in X_i : d(v, p) < \alpha\}, & i \neq c(p), \\ \{v \in X_i : d(v, p) < \beta\}, & i = c(p), \end{cases}$$

letting $B(p) = \bigcup_{i \in [m]} B_i(p)$ be the union over all colors.

Definition 4 (Conflict Set). *Let $\beta \geq \alpha > 0$ be fixed parameters. Given set S and point $q \in S$, for each color $i \in [m]$ and point p we define*

$$\Gamma_i^q(p, S) = B_i(p) \cap S^{-q},$$

again letting $\Gamma^q(p, S) = \bigcup_{i \in [m]} \Gamma_i^q(p, S)$ be the union over all colors.

When it is clear from context, we may remove the S from $\Gamma_i^q(p, S)$. We may further expand these definitions to include sets of points P to denote the union, i.e., $\Gamma_i^q(P, S) = \bigcup_{p \in P} \Gamma_i^q(p, S)$. Moreover, a point $p \in S$ is called *bad* with respect to S if $|S \cap B(p)| > 1$. In simpler terms, p is called bad with respect to S if there is a point in S (other than p) that is "too close" to p. It is important to note that if S is β-color separated, a bad point may only be close to another point with a different color in S.

Throughout this paper, we denote by S^* an optimal solution, and by OPT its diversity, i.e., $\mathrm{div}(S^*) = \mathrm{OPT}$. We refer to each point in S^* as an *optimal point*.

3 Results for Fair Diversity Maximization

In this section, we present our results for Fair Diversity Maximization. We first introduce a unified framework that underlies all of our algorithms. We then use this framework to obtain a 4-approximation for any constant number of groups. Finally, we use the same framework for the case of $m = 2$ groups to obtain a tight 2-approximation.

3.1 Local-Repair Framework

Our local-repair framework is given in Algorithm 1. In this algorithm, we assume that we are given a value μ satisfying $\mu \leq \mathrm{OPT}$. We are also given two lower bounds $\alpha \leq \beta \leq \frac{\mu}{2}$.

Algorithm 1 LOCAL-REPAIR FRAMEWORK

Input: A point set X, a guess $\mu \leq \mathrm{OPT}$, two values $\frac{\mu}{c} \leq \alpha \leq \beta \leq \frac{\mu}{2}$, for some constant $c > 0$
Output: A fair subset $S \subseteq X$ with $\mathrm{div}(S) \geq \alpha$.

```
for i ∈ [m] do
    let S_i = GMM(X_i, k_i)
let S = ⋃_{i∈[m]} S_i
while there exists bad point q in S do
    P ← FINDREPAIR(S, q, α, β)
    S ← (S^{-q} \ Γ^q(P)) ∪ P
return S
```

We first construct a fair β-color separated set S by applying the GMM algorithm (Ravi et al., 1991) [1] independently to each color group to select k_i points from color i.

We then start a local repair to gradually improve the diversity of S. In each iteration, we identify a bad point $q \in S$ and replace a subset of points in S, along with q with another set of points such that the resulting solution remains fair while strictly decreasing the number of bad points. To facilitate this replacement, we introduce the notion of a q-repair set.

Definition 5 (q-repair)**.** *Let $q \in S$ be a bad point. A set P is a q-repair if.*

- $\mathrm{div}(P) \geq \alpha$,
- *P is β-color separated,*
- *For every color $i \in [m]$, $|\Gamma_i^q(P)| + \mathbf{1}_{[i=c(q)]} \leq |P_i|$*

The first two conditions ensure that no two points of P violate the required separation. Points of different colors are at distance at least α, while points of the same color are at distance at least β. The third condition ensures that inserting P preserves the group quotas. After removing q and all points in $\Gamma^q(P)$ from S and inserting P, the resulting set contains at least k_i points of every color i. This also explains the terminology *conflict set*, since the points in $\Gamma^q(P)$ are exactly the selected points that prevent P from being inserted while maintaining the required separation.

The FINDREPAIR subroutine is responsible for finding a suitable q-repair set. This is the setting-specific part of the framework, and its implementation can be adapted to the problem at hand. Specifically, the user of this framework must design a proper FINDREPAIR subroutine and appropriate parameters α and β depending on μ such that FINDREPAIR always returns a q-repair whenever $\mu \leq \mathrm{OPT}$.

Theorem 3. *Given a polynomial-time subroutine* FINDREPAIR*, a guess $\mu \leq \mathrm{OPT}$, and two appropriate values $\mu/c \leq \alpha \leq \beta \leq \mu/2$, for some constant $c > 0$, Algorithm 1 finds a fair set S with $\mathrm{div}(S) \geq \mu/c$ in polynomial time.*

The correctness of the framework follows from a simple invariant. Initially, GMM guarantees diversity at least $\mu/2 \geq \beta$ within each color group, so S is β-color separated. In each iteration, let P be the q-repair returned by FINDREPAIR. By the definition of $\Gamma^q(P)$, removing q and $\Gamma^q(P)$ eliminates every point that would conflict with a point of P.

Since P has diversity at least α and is itself β-color separated, inserting P creates no new bad points, while q is removed.

The balance condition in the definition of a q-repair guarantees that, for every color i,

$$|S_i| - |\Gamma_i^q(P)| - \mathbf{1}_{[i=c(q)]} + |P_i| \geq k_i.$$

Hence, any excess points can be removed while preserving both fairness and β-color-separation. Therefore, the number of bad points strictly decreases in every iteration. Since there are at most k bad points, the algorithm terminates after at most k iterations, at which point $\mathrm{div}(S) \geq \alpha \geq \mu/c$.

It remains to explain how the value μ is chosen. Let $D := \{d(u, v) \mid u, v \in X\}$ be the set of all pairwise distances. Since $\mathrm{OPT} \in D$, we may run Algorithm 1 for every $\mu \in D$ and return the

[1] The GMM algorithm is described in the appendix.

solution with maximum diversity. Alternatively, we can use binary search over D to find the desired solution more efficiently. Since for every guess $\mu \leq \mathrm{OPT}$ the algorithm returns a fair solution of diversity at least $\alpha \geq \mu/c$, the run with $\mu = \mathrm{OPT}$ yields a c-approximation.

3.2 4-Approximation Algorithm for any Constant m

In this section we show how to use the local-repair framework developed in the previous subsection to obtain a 4-approximation for fair diversity maximization for any constant number of groups. To this end, we need to design a FindRepair subroutine that finds a q-repair for $(\alpha, \beta) = (\frac{\mu}{4}, \frac{\mu}{2})$ in polynomial time.

We first introduce the repair vector of p.

Definition 6. *For a point p, we define* repair vector $\mathrm{rep}(p) \in \mathbb{Z}^m$ *to be*

$$\mathrm{rep}_i(p) = |B_i(p) \cap S^{-q}| - \mathbf{1}_{[i=c(q)]}.$$

The interpretation is that inserting p into S^{-q} supplies one unit of its own color while consuming one unit for every conflicting point in S that must be removed.

Lemma 7. *For any two optimal points $p_1, p_2 \in S^*$, the sets $B(p_1)$ and $B(p_2)$ are disjoint.*

Proof. Suppose there exists a $v \in B(p_1) \cap B(p_2)$. Then,

$$\mu \leq d(p_1, p_2) \leq d(p_1, v) + d(v, p_2) < \frac{1}{2}\mu + \frac{1}{2}\mu,$$

which is a contradiction. □

Lemma 8. *For the optimal set S^* we have*

$$\sum_{p \in S^*} \mathrm{rep}(p) \leq -\boldsymbol{e}^{(c(q))},$$

where e_j is the j-th unit vector, and the inequality is coordinate-wise.

Proof. According to Lemma 7 each $v \in S^{-q}$ is inside at most one $B(p)$ for $p \in S^*$. This means that for every color i,

$$\sum_{p \in S^*} |B_i(p) \cap S^{-q}| \leq |S_i^{-q}| = k_i - \mathbf{1}_{[i=c(q)]}.$$

Since $|S_i^*| = k_i$ then we must have

$$\sum_{p \in S^*} \mathrm{rep}_i(p) = \sum_{p \in S^*} |B_i(p) \cap S^{-q}| - |S_i^*| \leq -\mathbf{1}_{[i=c(q)]},$$

which is exactly $-\boldsymbol{e}^{(c(q))}$. □

This lemma shows that S^* is a q-repair for all S and q. Let $P^* \subseteq S^*$ be the smallest subset of S^* satisfying

$$\sum_{p \in P^*} \mathrm{rep}(p) \leq -\boldsymbol{e}^{(c(q))}.$$

We will refer to this inequality as the *balance inequality*. In the following lemmas, we will show that $|P^*|$ is bounded by a function of only m, independent of k and n.

Lemma 9. *Given $p \in P^*$, for every color i, we must have $|B_i(p) \cap S^{-q}| \leq 1$, and in particular, $|B_{c(p)}(p) \cap S^{-q}| = 0$.*

Proof. First suppose $i \neq c(p)$ and $|B_i(p) \cap S^{-q}| \geq 2$. Pick distinct $u, v \in B_i(p) \cap S^{-q}$. Since u, v have the same color,

$$\frac{1}{2}\mu \leq d(u, v) \leq d(u, p) + d(v, p) < \frac{1}{4}\mu + \frac{1}{4}\mu,$$

a contradiction.

Algorithm 2 FINDREPAIR GENERAL

Input: A $\frac{\mu}{2}$-color separated set S, a point $q \in S$, and two parameters $(\alpha, \beta) = (\frac{\mu}{4}, \frac{\mu}{2})$
Output: A q-repair set P, provided $\mu \leq \mathrm{OPT}$

```
for all P ⊆ X with size at most c_m do
    if P is a valid q-repair set then return P
return NONE
```

Now let $i = c(p)$ and suppose $|B_i(p) \cap S^{-q}| \geq 1$. Then all coordinates of $\mathrm{rep}(p)$ are non-negative, and hence

$$\sum_{p' \in P^* \setminus \{p\}} \mathrm{rep}(p') \leq -\mathrm{rep}(p) - \boldsymbol{e}^{(c(q))} \leq -\boldsymbol{e}^{(c(q))}.$$

Thus $P^* \setminus \{p\}$ still satisfies the balance inequality, contradicting the minimality of P^*. □

An important corollary is that for all $p \in P^*$, $B_i(p) \cap S^{-q}$ is either empty or contains a single point, and therefore, $\mathrm{rep}(p) \in \{-1, 0, 1\}^m$. We now describe a useful property of P^*: the inequality in the balance property is indeed an equality.

Lemma 10. *The minimal set P^* satisfying the balance inequality must satisfy*

$$\sum_{p \in P^*} \mathrm{rep}(p) = -\boldsymbol{e}^{(c(q))}.$$

Proof. Suppose the inequality is strict in some coordinate i. Since the vectors are integral,

$$\sum_{p \in P^*} \mathrm{rep}_i(p) \leq -1 - \mathbf{1}_{[i=c(q)]}.$$

By Lemma 9, the only negative entry of $\mathrm{rep}(p)$ is $\mathrm{rep}_{c(p)}(p) = -1$. Hence, there exists $p_0 \in P^*$ with $c(p_0) = i$.

For every $j \neq i$, $\mathrm{rep}_j(p_0) \geq 0$, so removing p_0 preserves the balance inequality in coordinate j. For coordinate i,

$$\sum_{p \in P^* \setminus \{p_0\}} \mathrm{rep}_i(p) = \sum_{p \in P^*} \mathrm{rep}_i(p) + 1 \leq -\mathbf{1}_{[i=c(q)]}.$$

Thus $P^* \setminus \{p_0\}$ still satisfies the balance inequality, contradicting the minimality of P^*. □

Finally we have enough structure to prove our claim

Lemma 11. *There exists a constant c_m (depending only on m), such that the minimal subset $P^* \subseteq S^*$ satisfying the balance inequality has at most c_m points.*

Proof. Consider the vectors $\{\mathrm{rep}(p) : p \in P^*\}$ together with $\boldsymbol{e}^{(c(q))}$. By Lemma 10, their sum is zero, and by Lemma 9, each lies in $\{-1, 0, 1\}^m$ and hence has ℓ_∞-norm at most 1.

By the Steinitz lemma, they can be ordered as $v_1, \ldots, v_t$ so that every partial sum $u_r = \sum_{j=1}^{r} v_j$ lies in $[-m, m]^m$. Thus, there are at most $(2m+1)^m$ possible values for the partial sums.

If $|P^*| > (2m+1)^m - 1$, then $t = |P^*| + 1 > (2m+1)^m$, so two partial sums coincide. The vectors between them form a nonempty zero-sum set, and so do the remaining vectors. The set not containing $\boldsymbol{e}^{(c(q))}$ therefore consists of vectors $\mathrm{rep}(p)$ for a nonempty subset $P' \subseteq P^*$ whose sum is zero. Hence,

$$\sum_{p \in P^* \setminus P'} \mathrm{rep}(p) = -\boldsymbol{e}^{(c(q))},$$

contradicting the minimality of P^*. Setting $c_m = (2m+1)^m - 1$ completes the proof. □

By Lemma 11, whenever $\mu \leq \mathrm{OPT}$, there exists a q-repair set of size at most c_m. Since m is constant, c_m is also constant, and Algorithm 2 can enumerate all subsets of size at most c_m and check in polynomial time whether any is a valid q-repair. Thus, Algorithm 2 runs in polynomial time. Plugging this algorithm into our local-repair framework yields a 4-approximation for fair diversity maximization, proving Theorem 1.

Algorithm 3 FINDREPAIR TWO-GROUPS

Input: A $\frac{\mu}{2}$-color separated set S, a point $q \in S$, and two parameters $(\alpha, \beta) = (\frac{\mu}{2}, \frac{\mu}{2})$
Output: A q-repair P, provided $\mu \leq \mathrm{OPT}$

1: rename colors so that $c(q)$ is red
2: **if** there exists a red point p outside $B(S^{-q})$ **then**
3: **return** $\{p\}$
4: let $R = \{\}$
5: **while** a point v exists outiside $B(R \cup S^{-q})$ **do**
6: $R \leftarrow R \cup \{v\}$
7: choose a red point $p \notin S^{-q}$ with $|B(p) \cap S_r^{-q}| = 0$ and $|B(p) \cap (S_b^{-q} \cup R)| \leq |R|$
8: **return** $\{p\} \cup (R \setminus B(p))$

3.3 TIGHT BOUND FOR TWO GROUPS

In this subsection, we present a simple algorithm for finding a q-repair with $(\alpha, \beta) = (\frac{\mu}{2}, \frac{\mu}{2})$. Combined with our local-repair framework, this algorithm yields a 2-approximation for fair diversity maximization with two groups. Here, we provide an overview of our algorithm. Complete lemmas and proofs are provided in the appendix.

Throughout this subsection, we use red and blue to denote the two groups for simplicity. Accordingly, we denote the corresponding subsets of X by X_r and X_b, and the associated color quotas by k_r and k_b, respectively.

Our implementation of FINDREPAIR for the two-color case is given in Algorithm 3. The algorithm takes as input $(\alpha, \beta) = (\frac{\mu}{2}, \frac{\mu}{2})$, a $\frac{\mu}{2}$-color-separated set S, and a point $q \in S$. The following lemma shows that the algorithm correctly returns a q-repair, provided $\mu \leq \mathrm{OPT}$.

Lemma 12. *If $\mu \leq \mathrm{OPT}$, then Algorithm 3 returns a valid q-repair in polynomial time.*

Proof overview. The algorithm terminates in either line 3 or line 8. If it terminates in line 3, then $P = \{p\}$ is clearly a valid q-repair set. Suppose instead that the algorithm terminates in line 8. We show that there exists a desirable point p such that $|B(p) \cap S_r^{-q}| = 0$ and $|B(p) \cap (S_b^{-q} \cup R)| \leq |R|$. The algorithm then returns $P = \{p\} \cup (R \setminus B(p))$.

By the construction of R, $B(R)$ contains no point of S^{-q}. Moreover, $B(p) \cap S^{-q}$ can contain only blue points. Thus, the points in P satisfy the required separation conditions. The choice of p and the construction of R also ensure that P satisfies the q-repair color condition. Finally, after adding p, the algorithm removes all points in $B(p)$ from R, leaving only points compatible with p. Hence, P is compatible. Therefore, Algorithm 3 returns a valid q-repair set in polynomial time. □

Combining this algorithm with our local-repair framework yields a 2-approximation for fair diversity maximization with two groups, thereby proving Theorem 2.

4 CONCLUSION

We studied Fair Diversity Maximization under exact group-representation constraints and showed that the previously known dependence on the number of groups can be avoided when m is constant. Our main result is a polynomial-time 4-approximation for any constant number of groups, with an approximation guarantee independent of m. We also obtained a tight 2-approximation for the special case of two groups, improving the previous best factor of 3. Both results are based on a unified local-repair framework that preserves the group quotas exactly while gradually improving the diversity objective. All of our results are supported by rigorous mathematical proofs.

Having eliminated the dependence on m from the approximation factor, a natural direction for future work is to determine whether the factor 4 can be further improved. In particular, our tight 2-approximation for two groups raises the question of whether the optimal factor of 2 can be achieved for any constant number of groups. Establishing such a tight approximation guarantee even for small values of m remains an intriguing open problem.

AI USE STATEMENT

We used AI-assisted tools to improve the clarity, grammar, and presentation of the manuscript. In particular, these tools were used for proofreading and rephrasing portions of the text. All mathematical results, algorithms, proofs, and claims were developed and verified by the authors. The authors take full responsibility for the content of the paper.

REFERENCES

Sofiane Abbar, Sihem Amer-Yahia, Piotr Indyk, and Sepideh Mahabadi. Real-time recommendation of diverse related articles. In *Proceedings of the 22nd International Conference on World Wide Web (WWW)*, pp. 1–12, 2013.

Zeinab Abbassi, Vahab S Mirrokni, and Mayur Thakur. Diversity maximization under matroid constraints. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 32–40, 2013.

Raghavendra Addanki, Andrew McGregor, Alexandra Meliou, and Zafeiria Moumoulidou. Improved approximation and scalability for fair max-min diversification. In *Proceedings of the 25th International Conference on Database Theory*, pp. 7:1–7:21, 2022.

Florian Adriaens and Nikolaj Tatti. Fair diversity maximization with few representatives. In *Proceedings of the 31st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 17–25, 2025.

Sepideh Aghamolaei, Majid Farhadi, and Hamid Zarrabi-Zadeh. Diversity maximization via composable coresets. In *Proceedings of the 27th Canadian Conference on Computational Geometry*, pp. 38–48, 2015.

Ali Ahmadvand, Mohammad Ansari, Mobin Razavi, and Hamid Zarrabi-Zadeh. Composable coresets for fair diversity maximization. In *Proceedings of the 38th ACM Symposium on Parallelism in Algorithms and Architectures*, pp. 561–570, 2026.

Mohiuddin Ahmed. Data summarization: a survey. *Knowledge and Information Systems*, 58(2):249–273, 2019.

Saeed Ranjbar Alvar, Gursimran Singh, Mohammad Akbari, and Yong Zhang. DivPrune: Diversity-based visual token pruning for large multimodal models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9392–9401, 2025.

Hadis Anahideh, Abolfazl Asudeh, and Saravanan Thirumuruganathan. Fair active learning. *Expert Systems with Applications*, 199:116981, 2022.

Mark Ashton, John Barnard, Florence Casset, Michael Charlton, Geoffrey Downs, Dominique Gorse, John Holliday, Roger Lahana, and Peter Willett. Identification of diverse database subsets using property-based and fragment-based molecular descriptions. *Quantitative Structure-Activity Relationships*, 21(6):598–604, 2002.

Michele Borassi, Alessandro Epasto, Silvio Lattanzi, Sergei Vassilvitskii, and Morteza Zadimoghaddam. Better sliding window algorithms to maximize subadditive and diversity objectives. In *Proceedings of the 38th ACM Symposium on Principles of Database Systems (PODS)*, pp. 254–268, 2019.

Allan Borodin, Aadhar Jain, Hyun Chul Lee, and Yuli Ye. Max-sum diversification, monotone submodular functions, and dynamic updates. *ACM Transactions on Algorithms*, 13(3):41:1–41:25, 2017.

Matteo Ceccarello, Andrea Pietracaprina, Geppino Pucci, and Eli Upfal. Mapreduce and streaming algorithms for diversity maximization in metric spaces of bounded doubling dimension. *Proceedings of the VLDB Endowment*, 10(5):469–480, 2017.

Matteo Ceccarello, Andrea Pietracaprina, and Geppino Pucci. Fast coreset-based diversity maximization under matroid constraints. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, pp. 81–89, 2018.

Matteo Ceccarello, Andrea Pietracaprina, and Geppino Pucci. A general coreset-based approach to diversity maximization under matroid constraints. *ACM Transactions on Knowledge Discovery from Data*, 14(5): 1–27, 2020.

Elisa Celis, Vijay Keswani, Damian Straszak, Amit Deshpande, Tarun Kathuria, and Nisheeth Vishnoi. Fair and diverse DPP-based data summarization. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pp. 716–725, 2018a.

L. Elisa Celis, Lingxiao Huang, and Nisheeth K. Vishnoi. Multiwinner voting with fairness constraints. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 144–151, 2018b.

Danny Z Chen, Jian Li, Hongyu Liang, and Haitao Wang. Matroid and knapsack center problems. *Algorithmica*, 75:27–52, 2016.

Mohsen Dehghankar, Rahul Raychaudhury, Stavros Sintos, and Abolfazl Asudeh. Fair set cover. In *Proceedings of the 31st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 189–200, 2024.

Marina Drosou and Evaggelia Pitoura. Search result diversification. *ACM SIGMOD Record*, 39(1):41–47, 2010.

Erhan Erkut. The discrete p-dispersion problem. *European Journal of Operational Research*, 46(1):48–60, 1990.

Erhan Erkut and Susan Neuman. Analytical models for locating undesirable facilities. *European Journal of Operational Research*, 40(3):275–291, 1989.

Matthew Fahrbach, Srikumar Ramalingam, Morteza Zadimoghaddam, Sara Ahmadian, Gui Citovsky, and Giulia DeSalvo. GIST: Greedy independent set thresholding for max-min diversification with submodular utility. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 38, pp. 46322–46349, 2025.

Ziyang Fan, Keyu Chen, Ruilong Xing, Yulin Li, Li Jiang, and Zhuotao Tian. FlashVID: Efficient video large language models via training-free tree-based spatiotemporal token merging. In *International Conference on Learning Representations (ICLR) (Poster)*, 2026.

Ashley Qianxi Gao and Allan Borodin. Fair max-min diversification in refined and relaxed metric spaces. Preprint, 2024.

David García-Soriano and Francesco Bonchi. Fair-by-design matching. *Data Mining and Knowledge Discovery*, 34:1291–1335, 2020.

Marwa El Halabi, Slobodan Mitrovic, Ashkan Norouzi-Fard, Jakab Tardos, and Jakub Tarnawski. Fairness in streaming submodular maximization: Algorithms and hardness. In *Proceedings of the 33rd Advances in Neural Information Processing Systems (NeurIPS)*, pp. 9150–9171, 2020.

Alireza Haqi and Hamid Zarrabi-Zadeh. Almost optimal massively parallel algorithms for k-center clustering and diversity maximization. In *Proceedings of the 35th ACM Symposium on Parallel Algorithms and Architectures*, pp. 239–247, 2023.

Atif Hassan, Swanand Khare, and Jiaul H. Paik. RCAP: Robust, class-aware, probabilistic dynamic dataset pruning. In *Proceedings of the Forty-first Conference on Uncertainty in Artificial Intelligence*, volume 286 of *Proceedings of Machine Learning Research*, pp. 1648–1662. PMLR, 2025.

Refael Hassin, Shlomi Rubinstein, and Arie Tamir. Approximation algorithms for maximum dispersion. *Operations research letters*, 21(3):133–137, 1997.

Piotr Indyk, Sepideh Mahabadi, Mohammad Mahdian, and Vahab S Mirrokni. Composable core-sets for diversity and coverage maximization. In *Proceedings of the 33rd ACM Symposium on Principles of Database Systems (PODS)*, pp. 100–108, 2014.

Matthew Jones, Huy Nguyen, and Thy Nguyen. Fair k-centers via maximum matching. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pp. 4940–4949, 2020.

Matthäus Kleindessner, Pranjal Awasthi, and Jamie Morgenstern. Fair k-center clustering for data summarization. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, pp. 3448–3457, 2019.

Matevž Kunaver and Tomaž Požrl. Diversity in recommender systems: A survey. *Knowledge-Based Systems*, 123:154–162, 2017.

Yash Kurkure, Miles Shamo, Joseph Wiseman, Sainyam Galhotra, and Stavros Sintos. Faster algorithms for fair max-min diversification in R^{d}. *Proceedings of the ACM on Management of Data*, 2(3):1–26, 2024.

Ruien Li and Yanhao Wang. Individually fair diversity maximization. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 38, pp. 160253–160278, 2025.

Zifan Liu, Amin Karbasi, and Theodoros Rekatsinas. TSDS: Data selection for task-specific model finetuning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pp. 10117–10147, 2024.

Sepideh Mahabadi and Stojan Trajanovski. Core-sets for fair and diverse data summarization. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 78987–79011, 2023.

I Douglas Moon and Sohail S Chaudhry. An analysis of network location problems with distance constraints. *Management Science*, 30(3):290–307, 1984.

Zafeiria Moumoulidou, Andrew McGregor, and Alexandra Meliou. Diverse data selection under fairness constraints. In *Proceedings of the 24th International Conference on Database Theory*, pp. 13:1–13:25, 2021.

Paolo Pellizzoni, Andrea Pietracaprina, and Geppino Pucci. Fully dynamic clustering and diversity maximization in doubling metrics. *ACM Transactions on Knowledge Discovery from Data*, 19(4):1–45, 2025.

Sekharipuram S Ravi, Daniel J Rosenkrantz, and Giri Kumar Tayi. Facility dispersion problems: Heuristics and special cases. In *Proceedings of the 17th Workshop on Algorithms and Data Structures*, pp. 355–366, 1991.

Esther Rolf, Theodora T. Worledge, Benjamin Recht, and Michael I. Jordan. Representation matters: Assessing the importance of subgroup allocations in training data. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9040–9051. PMLR, 2021.

Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *International Conference on Learning Representations (ICLR) (Poster)*, 2018.

Xubin Wang, Jianfei Wu, Yuan Yichen, Deyu Cai, Mingzhe Li, and Weijia Jia. Demonstration selection for in-context learning via reinforcement learning. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*, pp. 64937–64954. PMLR, 2025.

Yanhao Wang, Francesco Fabbri, and Michael Mathioudakis. Fair and representative subset selection from data streams. In *Proceedings of the Web Conference 2021*, pp. 1340–1350, 2021.

Yanhao Wang, Francesco Fabbri, Michael Mathioudakis, and Jia Li. Fair max-min diversity maximization in streaming and sliding-window models. *Entropy*, 25(7):1066, 2023a.

Yanhao Wang, Michael Mathioudakis, Jia Li, and Francesco Fabbri. Max-min diversification with fairness constraints: Exact and approximation algorithms. In *Proceedings of the 2023 SIAM International Conference on Data Mining*, pp. 91–99, 2023b.

Sepehr Zadeh, Mehrdad Ghadiri, Vahab Mirrokni, and Morteza Zadimoghaddam. Scalable feature selection via distributed diversity maximization. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, pp. 2876–2883, 2017.

He Zheng, Anna Martínez-Gavara, Rafael Martí, and Fred Glover. Advanced probabilistic tabu search for the fair max-min diversity problem. *Computers & Operations Research*, 196:107650, 2026.

A Appendix

A.1 GMM algorithm

The GMM algorithm used in Section 3.1 is presented here. This simple greedy algorithm, originally proposed by Ravi et al. (1991), repeatedly selects the point farthest from the set of points selected so far. The algorithm is given in Algorithm 4.

Algorithm 4 GMM(X, k)

Input: A point set X in a metric space and a parameter k
Output: A subset $S \subseteq X$ of k points

$S \leftarrow \{$an arbitrary point of $X\}$
for $i = 2$ **to** k **do**
    find a point $p \in X \setminus S$ maximizing $d(p, S)$
    $S \leftarrow S \cup \{p\}$
return S

It is well known that GMM provides a 2-approximation for the standard single-group diversity maximization problem. We use this property in our framework to ensure that the initial set S is β-color separated, a property that is maintained throughout our algorithm.

A.2 Lemmas and proofs for Section 3.3

Lemma 13. *If $\mu \leq \mathrm{OPT}$ and $\alpha \leq \beta = \frac{\mu}{2}$, then for any point q, we have $|B(q) \cap S^*| \leq 1$.*

Proof. Suppose, for contradiction, that there exist distinct points $p_1, p_2 \in B(q) \cap S^*$. Then

$$\mathrm{OPT} \leq d(p_1, p_2) \leq d(p_1, q) + d(q, p_2) < 2\beta = \mu \leq \mathrm{OPT},$$

which is a contradiction. □

A direct consequence of this lemma is that the sets $B(v)$ are pairwise disjoint for $v \in S^*$. Intuitively, the points in S^* are pairwise at distance at least OPT, while each ball $B(v)$ has radius at most $\frac{1}{2}$ OPT.

Lemma 14. *If $\mu \leq$ OPT, then a point p satisfying the conditions in line 7 of Algorithm 3 exists.*

Proof. Define $x = |S_b^{-q} \cup R| - k_b$. Since the condition in line 2 is not met, all points added to R are blue. Hence $x = |R|$. Moreover, since q is red, we have $|S_r^{-q}| = k_r - 1$.

We first show that there exists a red point $p \in S_r^*$ such that $|B(p) \cap S_r^{-q}| = 0$. By lemma 13, for every $q_1 \in S_r^{-q}$,

$$|B(q_1) \cap S_r^*| \leq |B(q_1) \cap S^*| \leq 1.$$

Therefore,

$$|B(S_r^{-q}) \cap S_r^*| \leq \sum_{q_1 \in S_r^{-q}} |B(q_1) \cap S_r^*| \leq k_r - 1.$$

Since $|S_r^*| = k_r$, there exists some $p \in S_r^*$ such that $|B(p) \cap S_r^{-q}| = 0$.

It remains to show that this point p satisfies $|B(p) \cap (S_b^{-q} \cup R)| \leq x$. By lemma 13, the sets $B(v) \cap (S^{-q} \cup R)$ are pairwise disjoint for $v \in S^*$. Moreover, by the construction of R, every point v of S^* has at least one point in $B(v) \cap (S^{-q} \cup R)$; otherwise, it would have been added to R. Thus,

$$\begin{aligned} k_b + k_r + x - 1 &= |S^{-q} \cup R| \\ &\geq |B(S^*) \cap (S^{-q} \cup R)| \\ &= \sum_{v \in S^*} |B(v) \cap (S^{-q} \cup R)| \\ &\geq |S^* \setminus \{p\}| + |B(p) \cap (S^{-q} \cup R)| \\ &= k_b + k_r - 1 + |B(p) \cap (S^{-q} \cup R)|. \end{aligned}$$

Hence

$$|B(p) \cap (S^{-q} \cup R)| \leq x.$$

Since $|B(p) \cap S_r^{-q}| = 0$, this implies

$$|B(p) \cap (S_b^{-q} \cup R)| \leq x = |R|.$$

Thus, p satisfies both required properties. Finally, such a point can be found in polynomial time by checking every red point. □

We are now ready to provide the complete proof of Lemma 12, stated in Section 3.3.

Proof of Lemma 12. The algorithm terminates in either line 3 or line 8. First, suppose that the algorithm terminates in line 3. Then $P = \{p\}$, where $p \notin B(S^{-q})$. Hence, $|\Gamma_i^q(P)| = 0$ for both colors i. Since p and q have the same color, we have $|P_r| = 1$, and thus the color condition for red is satisfied. The color condition for blue is immediate, since $P_b = \emptyset$ and $\Gamma_b^q(P) = \emptyset$. Moreover, since P contains only one point, it is trivially $(\frac{\mu}{2}, \frac{\mu}{2})$-compatible. Thus, P is a valid q-repair.

Now suppose that the algorithm terminates in line 8. By Lemma 14, there exists a desirable point p satisfying

$$|B(p) \cap S_r^{-q}| = 0 \qquad \text{and} \qquad |B(p) \cap (S_b^{-q} \cup R)| \leq |R|.$$

The algorithm returns

$$P = \{p\} \cup (R \setminus B(p)).$$

By the construction of R, $B(R)$ contains no point of S^{-q}. Moreover, $B(p) \cap S^{-q}$ contains only blue points. Therefore,

$$|\Gamma_r^q(P)| = 0.$$

Since $|P_r| = 1$, the red color condition follows:

$$|P_r| = 1 \geq 1 + |\Gamma_r^q(P)|.$$

For the blue color, we have

$$\begin{aligned}|P_b| &= |R \setminus B(p)| \\ &= |R| - |B(p) \cap R| \\ &= |R| + |B(p) \cap S_b^{-q}| - |B(p) \cap (S_b^{-q} \cup R)| \\ &\geq |B(p) \cap S_b^{-q}|.\end{aligned}$$

Since $p \in P$ and $R \setminus B(p)$ contains no point of $B(p)$,

$$|B(p) \cap S_b^{-q}| = |B(P) \cap S_b^{-q}| = |\Gamma_b^q(P)|.$$

Hence

$$|P_b| \geq |\Gamma_b^q(P)|,$$

and the blue color condition is satisfied.

It remains to verify compatibility. By construction, the points in R are mutually compatible and have no point of S^{-q} in their respective balls. When p is added to P, all points in $B(p)$ are removed from R. Thus, every point remaining in $R \setminus B(p)$ is compatible with p, and consequently P is $(\frac{\mu}{2}, \frac{\mu}{2})$-compatible.

Therefore, in either case, Algorithm 3 returns a valid q-repair set. All steps of the algorithm, including the required distance and neighborhood checks, can be performed in polynomial time. □