Socially Fair Clustering: Parameterized Approximation and Local Search

Aditya Anand
Computer Science and Engineering, University of Michigan, Ann Arbor, MI, USA

Yury Makarychev
Toyota Technological Institute at Chicago (TTIC), IL, USA

Liren Shan
Toyota Technological Institute at Chicago (TTIC), IL, USA

Abstract

We study the Socially Fair Clustering problem introduced by Abbasi, Bhaskara, and Venkatasubramanian [2] and by Ghadiri, Samadi, and Vempala [14], along with its extension, the (p, q)-Socially Fair Clustering problem. This problem generalizes k-median and k-means to settings where data points are partitioned into ℓ groups, and the goal is to find a fair clustering that is simultaneously good for all groups. We present several algorithms for this problem.

1. For ℓ_p-Socially Fair Clustering, we give the first constant-factor FPT-approximation parameterized by the number of groups ℓ, resolving the open question raised by Ghadiri, Singh, and Vempala [15]. Our main ingredient is a new algorithm for closing additional centers in parameterized time inspired by local search.
2. We then turn to the more general (p, q)-Socially Fair Clustering problem. The known algorithm for this problem, proposed by Chlamtáč, Makarychev, and Vakilian [13], achieves a very good approximation but is complex, slow and difficult to implement. We analyze the performance of a simple local search algorithm and show that it provides an $O(q)$-approximation in the worst case.
3. Finally, we design approximation algorithms for the facility location variant of the problem, where the number of facilities (centers) is not fixed in advance, and opening each facility incurs an opening cost. Unlike in previous work, we do not assume these opening costs are the same for all groups.

2012 ACM Subject Classification Theory of computation → Facility location and clustering; Theory of computation → Fixed parameter tractability

Keywords and phrases Socially fair clustering, Approximation algorithms, Fixed-parameter tractability, Local search, Facility location, k-median and k-means

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.3

Category APPROX

Related Version *Full Version*: https://arxiv.org/abs/2608.30026

Funding *Yury Makarychev*: NSF award ECCS-2216899

1 Introduction

1.1 Overview

Today, computers are widely used to analyze and classify data, generate recommendations, and make decisions that affect many aspects of people's lives. Traditionally, algorithms have focused on optimizing the overall quality of a solution. As a result, their decisions may be unfair to specific individuals or groups. Given the significance of these decisions, addressing fairness in algorithmic decision-making has become essential.

The k-median and k-means objectives are widely used in clustering and typically yield solutions that perform well for most data points. However, some points may be much farther from their assigned centers than others, and therefore incur disproportionately high costs. This is especially problematic when data points represent people, and those bearing the highest costs disproportionately belong to minority groups. Consider, for example, a facility location problem modeled using k-median: we wish to open k facilities (e.g., hospitals or grocery stores) in a city to serve all residents effectively. The goal is to minimize the sum of distances from each person (data point) to their nearest facility (center): $\sum_{u \in X} d(u, C)$. If a minority group resides in an isolated neighborhood, it is possible that none of the facilities will be placed nearby, leaving that group underserved.

To address this issue, [2] and [14] independently introduced the *Socially Fair Clustering* problem. This problem was also previously studied by [6] in the context of clustering under uncertainty. In the fair clustering setting, the n data points belong to ℓ groups $G_1, \dots, G_\ell \subseteq X$, representing minority or otherwise protected populations (the groups may overlap and do not have to cover X). The goal is to find a clustering that performs well for all groups simultaneously. It is convenient to assign weights to data points: for group i, let $w_i(u)$ denote the weight of point u, where $w_i(u) > 0$ if $u \in G_i$ and $w_i(u) = 0$ otherwise; one natural choice is $w_i(u) = 1/|G_i|$ for $u \in G_i$.

The ℓ_p-Socially Fair Clustering problem, referred to as ℓ_p-Clustering for brevity, asks for a set of k centers C from a set of potential centers $F \subseteq X$ that minimizes

$$\max_{i \in \{1,\dots,\ell\}} \left(\sum_{u \in X} w_i(u) d(u, C)^p \right)^{1/p}, \tag{1}$$

ensuring that the solution performs well across all groups. In other words, we compute the p-norm of distances $\left(\sum_{u \in X} w_i(u) d(u, C)^p\right)^{1/p}$ for each group G_i, and the objective aggregates these by taking a maximum (many papers define the objective as the p-th power of ours; the two formulations are essentially equivalent: an α-approximation for our objective corresponds to an α^p-approximation for the alternative, and vice versa).

A more general variant of the problem, known as (p, q)-Socially Fair Clustering with $1 \le p \le q \le \infty$, employs a different aggregation function.[1] In this formulation, we construct a vector whose i-th coordinate represents the p-norm of the distances for group i, and then take the ℓ_q norm of this vector:

$$\left(\sum_{i=1}^{\ell} \left(\sum_{u \in X} w_i(u) d(u, C)^p \right)^{q/p} \right)^{1/q}. \tag{2}$$

We refer to this problem simply as (p, q)-Clustering. When $q = \infty$, the objective in (2) reduces to that in (1); thus, ℓ_p-Clustering and (p, ∞)-Clustering are equivalent. Moreover, the $(p, \max\{p, \log \ell\})$-objective approximates the ℓ_p-objective within a constant factor. This follows from the more general fact that the ℓ_q and ℓ_r norms of a d-dimensional vector

[1] We assume that $q \ge p$, since otherwise the objective would favor *unfair* clusterings. For example, let $p = 2q$ and $\ell = 2$. Suppose there are two candidate solutions: in the first, each group has p-norm distance 1, while in the second, the p-norms of the distances for the two groups are $2^{1/p}$ and 0 (where the p-norm distance of group i is given by $\left(\sum_{u \in X} w_i(u) d(u, C)^p\right)^{1/p}$). The corresponding aggregated objectives are $(1^{1/2} + 1^{1/2})^{1/q} = 2^{1/q}$ and $(2^{1/2} + 0^{1/2})^{1/q} = 2^{1/(2q)}$. Thus, we would prefer the second solution, even though it is clearly less fair.

are within a factor of $d^{\left|\frac{1}{q}-\frac{1}{r}\right|}$ of each other, which can be shown by a simple computation. Hence in all our results for (p,q)-Clustering, we obtain results for ℓ_p-Clustering by setting $q = \max\{p, \log \ell\}$.

1.2 Known results

The problem has been extensively studied since its introduction. We begin by summarizing previously published approximation algorithms. [6] gave an $O(\log n + \log \ell)$-approximation for ℓ_1-Clustering. [2] and [14] provided $O(\ell^{1/p})$-approximation algorithms for ℓ_p-Clustering with $p \in \{1,2\}$. Additionally, [2] presented a bicriteria approximation algorithm that achieves a constant-factor approximation by opening $O(k)$ centers.

The approximation factor was later improved by [22], who obtained an $O\left(\frac{\log \ell}{\log \log \ell}\right)^{1/p}$-approximation. This approximation factor is optimal under the assumption that $NP \not\subseteq \bigcap_{\delta>0} DTIME(2^{n^\delta})$, as was shown by [10]. Finally, [13] studied the (p,q)-Clustering problem and gave an $O\left(\frac{q}{\ln(1+q/p)}\right)^{1/p}$-approximation for it. This approximation factor matches that for ℓ_p-clustering when $q = \log \ell$.

Another line of research has focused on designing fixed-parameter tractable (FPT) algorithms for ℓ_p-Clustering. In a significant result, [16] settled the approximability of ℓ_p-Fair Clustering for FPT algorithms parameterized by k by designing a $3+\varepsilon$-approximation algorithm and proving a matching lower bound of $3-\varepsilon$. They also established lower bounds of $1 + 2/e - \varepsilon$ and $\sqrt{1+8/e} - \varepsilon$ for the approximability of ℓ_1 and ℓ_2-Clustering when parameterized by both k and ℓ.

[15] observed that in practice, ℓ is typically a small constant, while k may be very large. They presented a polynomial-time algorithm achieving a $5+2\sqrt{6}$ bicriteria approximation by opening at most $k+\ell$ centers. Building on the bicriteria result mentioned above, they also designed algorithms with running times $n^{O_p(\ell^2)}$ and $k^\ell \mathrm{poly}(n)$, achieving approximation factors of $5+2\sqrt{6}+\varepsilon$ and $15+6\sqrt{6}+\varepsilon$, respectively. The latter is an FPT algorithm parameterized by both k and ℓ. They then raised the question of whether there exists a FPT algorithm achieving a constant-factor approximation parameterized solely by ℓ. This question remained open until our current work.

We also note that the problem admits a $(1+\varepsilon)$-approximation when parameterized by k for certain special classes of metric spaces, as shown by [1]. In many real-world applications, the number of centers – typically referred to as facilities in this context – is not fixed in advance. Instead, opening each facility incurs a cost. As is standard, we refer to this setting as the Facility Location problem. [2] presented a 4-approximation algorithm for the Facility Location variant of ℓ_1-Clustering, assuming homogeneous costs across groups (i.e., the cost of opening a facility is shared equally among all groups).

1.3 Other related work

The theme of fair clustering has received a lot of attention in recent years. Some other notions of fairness that have been adopted include balancedness [11], no over-representation [4], fair hierarchical clustering [3], individual fairness [20], pairwise fairness [9, 8]. Many of these incorporate constraints on clusters rather than enforcing fairness through a new objective, as is the case for Socially Fair Clustering.

1.4 Our results

1.4.1 FPT Algorithm parameterized by ℓ

In this paper, we address several important open problems in the field. First, we resolve the open question posed by [15] by presenting a constant-factor approximation FPT algorithm for ℓ_p-Clustering, parameterized by the number of groups ℓ.

▶ **Theorem 1.** *For every $\varepsilon > 0$, there exists a $(17 + 6\sqrt{6} + \varepsilon)$-approximation algorithm for ℓ_p-Clustering that runs in time $f(\ell, p, \frac{1}{\varepsilon}) \cdot n^{O(1)}$.*

We prove this result by designing a new algorithm that reduces the number of centers in fixed-parameter time.

▶ **Theorem 2.** *There exists an algorithm that, given an instance of ℓ_p-Clustering, a c-approximate solution C with $k + t$ centers, and a parameter $\varepsilon > 0$, returns a $(3c + 2 + \varepsilon)$-approximate solution with k centers. The algorithm runs in time $f(\ell, p, t, \frac{1}{\varepsilon}) \cdot n^{O(1)}$.*

Combining Theorem 2 with the following result of [15], we immediately obtain Theorem 1, thereby answering their open question.

▶ **Theorem 3** ([15])**.** *There is a bicriteria polynomial-time algorithm for ℓ_p-Clustering that finds a $(5 + 2\sqrt{6})$-approximate solution with $k + \ell$ centers.*

Additionally, in the full version of this paper, we strengthen Theorem 3 in the important case of ℓ_1-Clustering by presenting a local-search algorithm that finds a $(3+\varepsilon)$-approximation with $k + f(\ell, \frac{1}{\varepsilon})$ centers. This yields a variant of the algorithm from Theorem 1 with a $(11 + \varepsilon)$-approximation for ℓ_1-Clustering.

1.4.2 Local Search Algorithm for (p, q)-Clustering

The second theme of our paper is the design of a more efficient and implementable algorithm for (p, q)-Clustering. The algorithm in [13] solves an exponential-size convex relaxation using the round-or-cut framework, which is impractical and difficult to implement. In contrast, local search algorithms are well known for their efficiency and ease of implementation. Previously, [7] showed how to obtain a $(5 + \varepsilon)$-approximation for k-median using a single-swap local search algorithm, and a $(3 + \varepsilon)$-approximation using a multiple-swap variant. This result was generalized by [17], who gave an $O(p)$-approximation for the single-group case of ℓ_p-Clustering and showed that this analysis is tight – there are instances where the algorithm returns only an $\Omega(p)$-approximate solution.

We analyze the performance of the single-swap local search algorithm on instances of (p, q)-Clustering with an arbitrary number of groups ℓ, and show that it yields an $O(q)$-approximation. The aforementioned lower bound in [17] implies that our analysis is tight. While our approximation guarantee is worse than those of [22] and [13], it comes close in a particularly important case: ℓ_1-Clustering. For that problem, we obtain an $O(\log \ell)$-approximation by setting $q = \log \ell$, which matches the best known approximation factor of $O(\log \ell / \log \log \ell)$ up to a $\log \log \ell$ factor. Our algorithm is simpler, more implementable, and much faster than the algorithm of [13].[2]

▶ **Theorem 4.** *There is a single-swap local search algorithm for the (p, q)-Clustering problem that gives an $O(q)$-approximation, performs at most $O(kq \log n)$ iterations and runs in time $O_{p,q}(n^2 \ell k^2 \log n)$.*

---

[2] The running time in [13] is polynomial, but it is not stated explicitly.

1.4.3 (p, q)-Socially Fair Facility Location

Finally, we study the (p, q)-Socially Fair Facility Location problem, an important and practically motivated variant of (p, q)-Clustering. Building on the result of [2] for ℓ_1, we present a $(4 + \varepsilon)$-approximation algorithm for (p, q)-Socially Fair Facility Location with homogeneous costs. We then consider a more general setting with heterogeneous costs, where different groups incur different facility opening costs. In this case, we obtain an $O(q/\log q)$-approximation; in particular, for $p = 1, q = \infty$, this yields an $O(\log \ell/\log\log \ell)$-approximation[3], which almost matches the approximation hardness of $\Omega(\log^{1-\varepsilon} \ell)$ for this problem; see the full version for details.

▶ **Theorem 5.** *For Socially Fair Facility Location with group-homogeneous facility cost, there exists a polynomial-time algorithm that achieves a $(4 + \varepsilon)$-approximation for any $\varepsilon > 0$.*

▶ **Theorem 6.** *For Socially Fair Facility Location with group-heterogeneous facility cost, there exists a polynomial-time algorithm that achieves an $O(q/\log q)$-approximation.*

Organization

Section 3 gives an overview of the proofs of our results. In Section 4, we present our fixed-parameter tractable (FPT) algorithm for ℓ_p-Clustering. In Section 5, we analyze a simple single-swap local search algorithm for (p, q)-Clustering. In Section 6, we describe our algorithm for (p, q)-Socially Fair Facility Location. In the full version of the paper, we give a $(3 + \varepsilon)$-bicriteria local-search algorithm that opens $k + f(\ell)$ centers for ℓ_1-Clustering.

2 Preliminaries

This paper focuses on the ℓ_p and (p, q)-Fair Clustering Problems, referred to below simply as ℓ_p- and (p, q)-Clustering. In these problems, we are given a metric space (X, d), where the points represent data points or clients, and a set of potential center or facility locations $F \subseteq X$. Let n be the number of data points or clients. For each point $u \in X$, we are also given a vector of weights $w(u) \in \mathbb{R}^{\ell}_{\geq 0}$, where the i-th coordinate $w_i(u)$ denotes the weight for group i. We denote the number of groups by ℓ. The goal is to select a subset of k centers $C \subseteq F$ that minimizes objective (1) for ℓ_p-Clustering and (2) for (p, q)-Clustering.

To analyze these objectives, we define the ℓ_p-cost of the clustering (raised to the power p) with respect to a weight function $w_i : X \to \mathbb{R}_{\geq 0}$ and centers C as

$$\text{cost}(w_i, C) = \sum_{u \in X} w_i(u) d(u, C)^p,$$

where $d(u, C) = \min_{c \in C} d(u, c)$. We then define the cost vector induced by weights w and centers C as the ℓ-dimensional vector whose i-th coordinate is $\text{cost}(w_i, C)^{1/p}$. Then the ℓ_p-objective in (1) is the ℓ_∞-norm of the cost vector, while the (p, q)-objective in (2) is its ℓ_q-norm.

We will need the following claim.

▷ Claim 7 (see e.g. Lemma A.1 in [21]). For all non-negative real numbers a and b, and for every $p \geq 1$ and $\varepsilon > 0$, we have

$$(a + b)^p \leq (1 + \varepsilon)^{p-1} a^p + (1 + 1/\varepsilon)^{p-1} b^p,$$

[3] We recall that the $(1, \log \ell)$ objective approximates the $(1, \infty)$ objective up to a constant factor

3 Proof Overview

In this section, we offer a technical overview of the proofs of our main results. We focus primarily on the fixed-parameter tractable (FPT) approximation algorithm for ℓ_p-Clustering parameterized by the number of groups ℓ (Section 4), and then briefly describe the ideas underlying the remaining results.

3.1 FPT approximation for ℓ_p-Clustering (Section 4)

Our goal is to design a constant-factor approximation algorithm for ℓ_p-Clustering that runs in time $f(\ell) \cdot n^{O(1)}$, parameterized only by the number of groups ℓ. As a starting point, we run the algorithm of [15] (see Theorem 3) to get a $(5+2\sqrt{6})$-approximate solution with at most $k+\ell$ centers in polynomial time. In the special case of ℓ_1-Clustering, we can alternatively use the local-search-based bicriteria approximation algorithm (see the full version of the paper), which yields a better approximation. Let C denote the resulting set of $k+t$ centers, where $t=\ell$, and let c be the corresponding approximation factor.

Given such a bicriteria solution, the main challenge is to close t centers and reassign the clients assigned to them to other centers, while controlling the cost increase for all groups. We proceed in two steps. First, we show that *existentially* it is possible to close t centers C_1 and reassign every client assigned to a center c in C_1 to the center closest to c in $C \setminus C_1$ so that the cost of the resulting solution is within a constant factor of the cost of C (see Lemma 9 for the exact guarantee). The key idea is that the new connection cost for a client u can be bounded in terms of both the cost of connecting u to its originally assigned center in C_1 and the cost of connecting u to the center serving u in the optimal solution C^*. This observation is important because it allows us to reassign the clients assigned to the t closed centers to at most t centers (that are not closed), and this property is crucial for our algorithm.

However, we cannot use the procedure from Lemma 9 directly, as it requires knowledge of C^*. Thus, to obtain an FPT algorithm, we need to identify the set C_1 without access to the optimal solution. This is achieved using the technique of color coding introduced by [5]. Let $C_2 \subseteq C \setminus C_1$ be the set formed by selecting, for each $c \in C_1$, a closest center to c in $C \setminus C_1$. Note that $|C_2| \leq |C_1| = t$ (each center $c_1 \in C_1$ contributes at most one selected center; and different centers in C_1 may potentially have the same closest center in $C \setminus C_1$). We color all centers in C randomly using two colors, red and blue. Each center is colored red or blue independently with probability $1/2$. Note that with probability at least 2^{-2t}, all centers in C_1 are colored red and all centers in C_2 are colored blue. By running the algorithm $O(2^{2t}\log n)$ times, we obtain a coloring with this property with high probability. We denote the sets of red and blue centers by R and B, respectively.

Now, we introduce the notion of a Z-profile of a center c. Loosely speaking, the Z-profile is a vector in $\mathbb{R}^\ell$ that captures, for each of the ℓ groups, the reassignment cost incurred by closing center c and reassigning all clients assigned to c to the center closest to c in Z. Crucially, the C_2-profiles of all centers in C_1 fully determine the reassignment cost: the reassignment cost vector is simply the sum of the C_2-profiles of all centers in C_1. Importantly, if we were to find another set of centers in R with the same profiles, the reassignment cost would be the same. However, there are two challenges in finding such a set. First, as defined, the number of possible profiles is unbounded. Second, since we do not know the set C_2, we cannot compute the C_2-profiles of centers in R.

To resolve the first issue, we discretize the set of all possible profiles so that their number is bounded by a function of ℓ, t, p, and $1/\varepsilon$. We then "guess" the number of centers in C_1 with each possible profile. Next, we observe that the C_2-profiles and B-profiles of centers in

C_1 are equal, and thus we can use B as a proxy for C_2. We select centers in R according to our guess of the profile counts and obtain a set D_1 of t centers to close. Given the discussion above, it is not hard to see that by closing the centers in D_1 and reassigning clients to their closest centers in B, we obtain a constant-factor approximation.

3.2 Local search for (p, q)-Clustering (Section 5)

Building on the local search algorithms from [7] and [17] for k-median, k-means, and (k, p)-clustering, we design and analyze a local search algorithm for the more general (p, q)-Clustering problem. The general approach is similar to that of [7] and [17], but the analysis is more complex, as it requires fine-grained control over the distribution of swaps (see below).

The algorithm starts with an arbitrary set C of k centers and repeatedly finds and performs a swap $c \to c'$ that replaces one of the centers $c \in C$ with a center c' outside of C. The swaps are chosen to minimize the (p, q)-Clustering objective, and the algorithm terminates when no swap decreases the current potential Φ by more than $\Phi/(8k)$. To prove that the algorithm achieves an $O(q)$-approximation, we show that if the cost of the current solution exceeds the optimal cost by $\Omega(q)$ times, then there must exist an improving swap. To this end, we present a distribution over special swaps that replace centers in C with centers in an unknown optimal solution C^* (The algorithm may choose a swap from the distribution or any other swap improving the objective; we use the distribution only for the analysis). While this approach resembles that of [7] and [17], we require significantly more control over the distribution of swaps to handle the more complex objective. In particular, we use careful linear approximations for the change in the non-linear objective that allow us to show that there is indeed such a distribution over improving swaps.

To obtain these linear approximations, roughly speaking, we can obtain upper and lower bounds on the derivative of the cost function in the desired range. One can then use the mean-value theorem to find a linear approximation of the change in the objective using these bounds. The actual proof is slightly different, and uses convexity arguments along with the mean value theorem. Carefully combining this with the standard linear-objective analysis for k-Median, we show that the algorithm is an $O(q)$ approximation.

3.3 Socially Fair Facility Location (Section 6)

Finally, we study (p, q)-Socially Fair Facility Location. We consider two variants of the problem: facility location with homogeneous opening costs and facility location with heterogeneous opening costs. For the homogeneous case, we use a rounding algorithm inspired by the standard facility location algorithm of [23]. For the heterogeneous case, we replace the greedy step that selects and opens the cheapest facility within a certain ball B_u with a randomized rounding step that opens a facility in B_u with probability proportional to the target probability given by the convex relaxation. We then use Rosenthal's inequality to bound the expected ℓ_q-norm of the solution. We also give a hardness result showing that our algorithm is near-optimal for ℓ_1-Socially Fair Facility Location (the case when $p = 1$, $q = \infty$).

4 FPT algorithm for ℓ_p-socially fair clustering

As discussed in the introduction, we provide an FPT algorithm for ℓ_p-Clustering, parameterized by the number of groups ℓ. The algorithm first finds an approximate solution with $k + t$ centers using Theorem 3. This solution provides only a *bicriteria* approximation. To obtain a *true* approximation, the algorithm closes t centers – using Theorem 2.

We now proceed with the proof of Theorem 2. Given the instance of ℓ_p-Clustering on point set X and metric d, a set of centers C and an assignment $\alpha : X \to C$, we define the cost with respect to the weight function w_i and assignment α as follows:

$$\mathrm{cost}_i(\alpha) = \mathrm{cost}(w_i, \alpha) := \sum_{u \in X} w_i(u) d(u, \alpha(u))^p.$$

Note that the ℓ_p-objective function for the assignment α is then $\max_{i \in [\ell]} \mathrm{cost}_i(\alpha)^{1/p}$. Throughout this section, we fix an optimal set of centers C^* and a corresponding optimal assignment β^*. Let us define OPT to be the optimal assignment ℓ_p-objective, so that $\mathrm{OPT} = \max_i \mathrm{cost}_i(\beta^*)^{1/p}$.

First, we show that given any set C of $k+t$ centers which is a c-approximate solution, indeed there exists a subset of t centers that can be closed such that the resulting subset of k-centers is a $(3c+2)$-approximate solution. However, besides this, we show a stronger property that our algorithm requires: every client assigned to some center of C (in an optimal assignment of clients to C) that is closed in this process can now be reassigned to one of at most t open centers in C. The next key lemma proves this result. Before we state the lemma, the following definition will be useful.

▶ **Definition 8.** *Given a set of open centers C, an assignment $\alpha : X \to C$ of points to centers in C, and two disjoint subsets $C_1, C_2 \subseteq C$, we define a new assignment α' obtained from α by reassigning clients as follows. For each client $u \in X$, if $\alpha(u) \in C_1$, we find the closest center c_2 in C_2 to $c_1 \equiv \alpha(u)$ and let $\alpha'(u) = c_2$. If $\alpha(u) \notin C_1$, we set $\alpha'(u) = \alpha(u)$ (do not change the assignment).*

We then define $\mathrm{cost}(\alpha, C_1, C_2)$ to be the vector $(v_1, v_2, \ldots, v_\ell)$ with $v_i = \mathrm{cost}_i(\alpha') - \mathrm{cost}_i(\alpha)$ for each $i \in [\ell]$.

Simply put, $\mathrm{cost}_i(\alpha, C_1, C_2)$ measures the increase in cost for group (coordinate) i by re-assigning each point assigned to a center $c_1 \in C_1$ to the center in C_2 closest to c_1. We use the shorthand $\mathrm{cost}(\beta, c_1, C_2)$ to denote $\mathrm{cost}(\beta, \{c_1\}, C_2)$. If β is an optimal assignment, then $\mathrm{cost}(\beta, c_1, C_2)$ is coordinatewise non-negative, since any reassignment can only increase the cost. We also have:

$$\mathrm{cost}(\beta, C_1, C_2) = \sum_{c_1 \in C_1} \mathrm{cost}(\beta, c_1, C_2) \tag{3}$$

We formalize this nearest-neighbor redirection operation in the following lemma, showing that one can close a set C_1 of t centers and reassign all affected clients to a set C_2 of at most t receiving centers while keeping the reassignment cost controlled. The redirection operator in the definition captures exactly what happens when we close centers $C_1 \subseteq C$: for each closed center $c_1 \in C_1$, we send all of its clients to the closest open center $c_2 \in C \setminus C_1$. Let C_2 be the set of such receiving centers.

▶ **Lemma 9.** *Given a c-approximate solution C with $k+t$ centers and a corresponding optimal assignment β of clients to centers in C, there exist disjoint sets $C_1, C_2 \subseteq C$ with $|C_2| \le |C_1| = t$ such that*

- $\|\mathrm{cost}(\beta, C_1, C_2)\|_\infty \le ((3c+2)^p - c^p)\mathrm{OPT}^p$
- $C_2 = \{c_2 : c_2$ *is the closest point from* $C \setminus C_1$ *to* c_1 *for some* $c_1 \in C_1\}$.

Proof. Let $C^* = \{c_1^*, c_2^*, \ldots, c_k^*\}$ be an optimal set of centers, and let β^* be an optimal assignment of clients to the centers in C^*. For each c_i^*, mark $\gamma(c_i^*) \in C$, the closest center to c_i^* in C. Then there are at most k marked centers in C and at least t unmarked ones. Choose any subset of t unmarked centers in C to be the set C_1. For each center of C_1, we take the closest center in $C \setminus C_1$ to obtain the set C_2. Note that this means $|C_2| \le |C_1|$.

We now show that this choice of C_1, C_2 satisfies the statement of the lemma. Let $c_1 \in C_1$ and let c_2 be the closest center to c_1 in $C \setminus C_1$ (which must be in C_2). Then, for every point $u \in X$ with $\beta(u) = c_1$, we reassign u to c_2.

We wish to bound the reassignment cost of u to c_2. Let $c^* = \beta^*(u)$ be u's assigned center in C^*. By the triangle inequality,

$$\begin{aligned} d(u, c_2) &\le d(u, c_1) + d(c_1, c_2) \\ &\le d(u, c_1) + d\big(c_1, \gamma(c^*)\big) \text{ (c_2 is the closest center to c_1 in $C \setminus C_1$)} \\ &\le d(u, c_1) + d(c_1, c^*) + d\big(c^*, \gamma(c^*)\big) \\ &\le d(u, c_1) + 2d(c_1, c^*) \text{ (by definition of $\gamma(c^*)$)} \\ &\le d(u, c_1) + 2\big(d(u, c_1) + d(u, c^*)\big) \text{ (again by triangle inequality)} \\ &\le 3d(u, c_1) + 2d(u, c^*) \end{aligned}$$

It follows from Claim 7 that for every $\delta > 0$, we get the inequality

$$d(u, c_2)^p \le 3^p d(u, c_1)^p (1+\delta)^{p-1} + 2^p d(u, c^*)^p \left(1 + \frac{1}{\delta}\right)^{p-1}$$

We can now use this to bound the change in cost. By definition, it follows that the contribution of client u to $\mathrm{cost}_i(\beta, C_1, C_2)$ is $w_i(u)\big(d(u, c_2)^p - d(u, c_1)^p\big)$. Now repeating the analysis for each point u, and recalling that $c_1 = \beta(u), c^* = \beta^*(u)$, we obtain

$$\begin{aligned} \mathrm{cost}_i(\beta, C_1, C_2) \le &\sum_{u \in X} w_i(u) \Big(3^p (1+\delta)^{p-1}\, d\big(u, \beta(u)\big)^p \\ &+ 2^p \left(1 + \frac{1}{\delta}\right)^{p-1} d\big(u, \beta^*(u)\big)^p - d(u, \beta(u))^p\Big) \\ \le &\Big((3c)^p (1+\delta)^{p-1} + 2^p \left(1 + \frac{1}{\delta}\right)^{p-1} - c^p\Big)\ \mathrm{OPT}^p \quad \text{for each } i \in [\ell], \end{aligned}$$

where we use the fact that $\sum_{u \in X} w_i(u) d\big(u, \beta^*(u)\big)^p \le \mathrm{OPT}^p$ and $\sum_{u \in X} w_i(u) d\big(u, \beta(u)\big)^p \le c^p\, \mathrm{OPT}^p$, since C is a c-approximate solution.

Now we set $\delta = \frac{2}{3c}$. It then follows that

$$\mathrm{cost}_i(\beta, C_1, C_2) \le ((3c+2)^p - c^p)\, \mathrm{OPT}^p\,. \qquad ◀$$

Next, we describe how to find such sets in FPT time. We will use the color coding technique of [5]. We remark that while our algorithm is stated as a randomized algorithm for simplicity, it can be easily derandomized using standard techniques (see, for example, [12, Lemma 1]).

▶ **Lemma 10.** *Given a c-approximate solution C with $k + t$ centers and a corresponding optimal assignment β of clients to centers in C, one can find disjoint sets D_1, D_2 with $|D_2| \le |D_1| = t$ satisfying $\|\mathrm{cost}(\beta, D_1, D_2)\|_\infty \le \big((3c+2)^p - c^p + \varepsilon^p\big)\, \mathrm{OPT}^p$ in time $f(\ell, t, p, \frac{1}{\varepsilon}) n^{O(1)}$.*

Proof. Let $C_1, C_2 \subseteq C$ be the (unknown) subsets of C guaranteed by Lemma 9. We use color coding. Randomly and independently color each center in C red or blue, each with probability $1/2$. Call a coloring *good* if every center of C_1 is colored red and every center of C_2 is colored blue. Then a random coloring is good with probability at least 2^{-2t}. By independently sampling $O(2^{2t} \log n)$ random colorings, we can ensure that we enumerate a good coloring with high probability. Henceforth, assume that we are given a good coloring (we will try all the $O(2^{2t} \log n)$ colorings). Let $R \subseteq C$ be the red-colored centers, and $B \subseteq C$ be the blue-colored centers in this good coloring.

For a set $Z \subseteq C$, we say that the Z-profile of $c \in C$ is $(i_1, i_2, \ldots, i_\ell)$ if the m-th coordinate of $\operatorname{cost}(\beta, c, Z)$ is in the interval $\left[\frac{(i_m-1)\varepsilon^p}{t} \operatorname{OPT}^p, \frac{i_m \varepsilon^p}{t} \operatorname{OPT}^p\right]$ for each $m \in [\ell]$. Here $i_1, i_2 \ldots i_\ell$ are positive integers. Let $\mathcal{P}$ be the set of profiles with all $i_j \leq t(\frac{3c+2}{\varepsilon})^p$. Then $|\mathcal{P}| \leq (t(\frac{3c+2}{\varepsilon})^p)^\ell$. Note that the C_2-profile of every $c_1 \in C_1$ lies in $\mathcal{P}$. This is the case because, for every coordinate $i \in [\ell]$, (1) $\operatorname{cost}_i(\beta, c_1, C_2) \geq 0$, since β is an optimal assignment, and (2) $\operatorname{cost}_i(\beta, c_1, C_2) \leq \operatorname{cost}_i(\beta, C_1, C_2) \leq \|\operatorname{cost}(\beta, C_1, C_2)\|_\infty \leq (3c+2)^p \operatorname{OPT}^p$ by (3) and Lemma 9.

For each profile $P \in \mathcal{P}$, guess the number of centers in C_1 with C_2-profile P. Let n_P denote this number for each P. Then $\sum_{P \in \mathcal{P}} n_P = |C_1| = t$. The number of guesses we need to make is at most $|\mathcal{P}|^t \leq \left(t(\frac{3c+2}{\varepsilon})^p\right)^{\ell t}$.

Consider $c_1 \in C_1$ and let c_2 be the closest point to c_1 in $C \setminus C_1$. Note that $c_2 \in C_2$ (per item 2 in Lemma 9) and c_2 is also the closest point to c_1 in B, since $c_2 \in C_2 \subseteq B \subseteq C \setminus C_1$. Therefore, the C_2- and B-profiles of c_1 are the same. In particular, there are n_P centers with B-profile P in C_1. Since all centers in C_1 are red (recall that we assumed that the coloring is good), there are at least n_P red centers with B-profile P.

Now for each $P \in \mathcal{P}$, we choose exactly n_P red centers whose B-profile is P to form the set D_1[4]. Once we choose D_1, we let D_2 be the set of all centers $d_2 \in B$ such that d_2 is the closest center in B to some $d_1 \in D_1$. This concludes the construction of the sets D_1 and D_2.

▶ **Lemma 11.** $\|\operatorname{cost}(\beta, D_1, D_2)\|_\infty \leq ((3c+2)^p - c^p + \varepsilon^p) \operatorname{OPT}^p$.

Proof. There is a one-to-one map $f : C_1 \to D_1$ such that the C_2-profile of c_1 and the B-profile of $d_1 = f(c_1)$ are the same for every $c_1 \in C_1$.

It follows that for every $d_1 \in D_1$ and every coordinate $i \in [\ell]$,

$$\operatorname{cost}_i(\beta, d_1, D_2) = \operatorname{cost}_i(\beta, d_1, B) \leq \operatorname{cost}_i(\beta, f^{-1}(d_1), C_2) + \frac{\varepsilon^p}{t} \operatorname{OPT}^p.$$

The term $\frac{\varepsilon^p}{t} \operatorname{OPT}^p$ is exactly the length of an interval in the definition of profiles. Summing this upper bound on $\operatorname{cost}(\beta, d_1, B)$ over all $d_1 \in D_1$ and using (3), we get

$$\begin{aligned}
\operatorname{cost}_i(\beta, D_1, D_2) &= \sum_{d_1 \in D_1} \operatorname{cost}_i(\beta, d_1, D_2) \\
&\leq \sum_{d_1 \in D_1} \left(\operatorname{cost}_i(\beta, f^{-1}(d_1), C_2) + \frac{\varepsilon^p}{t} \operatorname{OPT}^p\right) \\
&\leq \operatorname{cost}_i(\beta, C_1, C_2) + \varepsilon^p \operatorname{OPT}^p \leq \left((3c+2)^p - c^p + \varepsilon^p\right) \operatorname{OPT}^p.
\end{aligned}$$ ◀

◀

Proof of Theorem 2. Let β be an optimal assignment for the center set C. Invoke Lemma 10 to obtain the sets D_1, D_2. After obtaining the sets D_1, D_2, we close the centers D_1. Since $|D_1| = t$, it follows that we obtain a solution with at most k centers.

We have $\|\operatorname{cost}(\beta)\|_\infty \leq c^p \operatorname{OPT}^p$, and by Lemma 10, the additional cost incurred is at most $\|\operatorname{cost}(\beta, D_1, D_2)\|_\infty \leq ((3c+2)^p - c^p + \varepsilon^p) \operatorname{OPT}^p$. Thus the ℓ_p-objective of the resulting solution with k centers is at most $\left((3c+2)^p - c^p + \varepsilon^p + c^p\right)^{1/p} \operatorname{OPT} \leq (3c+2+\varepsilon) \operatorname{OPT}$ where in the final step we use the inequality $(a+b)^{1/p} \leq a^{1/p} + b^{1/p}$, which holds for every $p \geq 1, a, b \geq 0$. ◀

[4] While we do not know OPT, we can guess OPT^p up to a factor of $(1+\varepsilon)$, but we omit this detail in the proof for the sake of simplicity

By Theorem 3 of [15], we can find a $c = 5 + 2\sqrt{6}$-approximate solution with $k + \ell$ centers in polynomial time. By Lemma 10, we can close ℓ centers in this bicriteria approximation solution to get a solution with k centers in time $\left(\ell(\frac{3c+2}{\varepsilon})^p\right)^{\ell^2} n^{O(1)}$. This proves Theorem 1.

5 Local search algorithm for (p, q)-clustering

5.1 Local search with scalar costs

Following [7] and [17], we first revisit the local search algorithm for (k, p)-clustering. We will later apply the results of this section to analyze the performance of the local search algorithm for (p, q)-Clustering.

Let (X, d) be a metric space, and let $w(u)$ be non-negative vertex weights. We consider one step of the local search algorithm. Let $C = \{c_1, \ldots, c_k\}$ be the current set of k centers, and $C^* = \{c_1^*, \ldots, c_k^*\}$ be a target set. Later, C^* will be an optimal solution for the (p, q)-objective, but for now, we make no assumptions about it.

We define the cost of C as

$$\mathrm{cost}(w) = \mathrm{cost}(w, C) = \sum_{u \in X} w(u) d(u, C)^p,$$

and the cost of C^* as

$$\mathrm{cost}^*(w) = \mathrm{cost}(w, C^*) = \sum_{u \in X} w(u) d(u, C^*)^p.$$

Consider a swap operation $c_a \to c_b^*$ that replaces c_a with c_b^*. We denote the resulting set of centers by

$$C_{ab} = (C \setminus \{c_a\}) \cup \{c_b^*\},$$

and its cost by $\mathrm{cost}_{ab}(w) = \mathrm{cost}(w, C_{ab}) = \sum_{u \in X} w(u) d(u, C_{ab})^p$.

Consider a probability distribution $\mathcal{D}$ on the set of swaps $c_a \to c_b^*$ (equivalently, on $[k] \times [k]$). The expected cost after a random swap drawn from $\mathcal{D}$ is $\mathbb{E}_{(a,b)\sim\mathcal{D}}[\mathrm{cost}_{ab}(w)]$.

We derive the following guarantees for local search. Conceptually, the proof closely follows the analyses in [7] and [17]. However, the bounds we require are more intricate, as we need finer control over the distribution of swaps. We prove this lemma in Section 5.4.

▶ **Lemma 12.** *Let (X, d) be a metric space, and let C and C^* be two sets of centers. There exists a probability distribution $\mathcal{D}$ over swaps such that the following holds. For every $p \geq 1$, weights w, parameters $M \geq m > 0$, and coefficients $\xi_{ab} \in [m, M]$, we have:*

1. *For $\alpha = \frac{10Mp}{m}$, the following inequality holds.*

$$\mathbb{E}_{\mathcal{D}}[\xi_{ab} \cdot (\mathrm{cost}_{ab}(w) - \mathrm{cost}(w))] \leq \frac{m}{2k}(\alpha^p \, \mathrm{cost}^*(w) - \mathrm{cost}(w)) \tag{4}$$

2. *We also have:*

$$\mathbb{E}_{\mathcal{D}}[\xi_{ab} \cdot (\mathrm{cost}_{ab}(w) - \mathrm{cost}(w))] \leq \frac{(1 + 2^{2p})M}{k}(\mathrm{cost}^*(w) + \mathrm{cost}(w)) \tag{5}$$

3. *Finally, for every swap $c_a \to c_b^*$ in the support of $\mathcal{D}$ and every $\delta_0 \in (0, 1/4)$, we have*

$$\mathrm{cost}_{ab}(w) \leq (1 + \theta_0) \, \mathrm{cost}^*(w) + (1 + \delta_0) \, \mathrm{cost}(w), \tag{6}$$

where $\theta_0 = 2^p(5p/(4\delta_0))^{p-1}$.

At a high level, the necessity for such guarantees is as follows. In k-median, the objective is linear, and the effects of the change in cost due to a single swap can be directly accounted for. However, in (p, q)-Clustering, the objective involves two layers of non-linearity (the inner and outer norms). To deal with this non-linearity, we can write the change in the objective for each swap $c_a \to c_b^*$ as a linear change with multipliers ξ_{ab} using the mean value theorem and convexity. Here the multipliers ξ_{ab} are bounded in the range $[m, M]$ for some appropriate choice of m and M. The above lemma then assumes this boundedness of ξ_{ab} and derives guarantees on the change in $\text{cost}(w)$, which in turn bounds the change in the objective.

5.2 Single swap analysis w.r.t. the (p, q)-objective

Now we consider the local search algorithm for (p, q)-Clustering – the goal is to minimize the following objective function:

$$\left(\sum_{i=1}^{\ell}\left(\sum_{u\in X} w_i(u)d(u,C)^p\right)^{q/p}\right)^{1/q} = \left(\sum_{i=1}^{\ell}\text{cost}(w_i, C)^{q/p}\right)^{1/q}$$

From now on, let C^* be an optimal solution for the (p, q)-Clustering problem. We assume that swap $c_a \to c_b^*$ is sampled from distribution $\mathcal{D}$ constructed in Lemma 12. Define $r = q/p$ and $\varphi(t) = t^r$, and

$$z_i = \text{cost}(w_i), \qquad z_i' = \text{cost}_{ab}(w_i), \qquad z_i^* = \text{cost}^*(w_i)$$

$$\Phi = \sum_{i=1}^{\ell}\varphi(z_i), \qquad \Phi_{ab} = \sum_{i=1}^{\ell}\varphi(z_i'), \qquad \Phi^* = \sum_{i=1}^{\ell}\varphi(z_i^*)$$

Note that $(\Phi^*)^{1/q}$ is the cost of the optimal solution, $\Phi^{1/q}$ is the cost of the current solution, and $\Phi_{ab}^{1/q}$ is the cost of the solution after a swap $c_a \to c_b^*$. Also, let

$$\Delta_i = \mathbb{E}\left[\varphi(z_i') - \varphi(z_i)\right].$$

Fix a coordinate i. Now we get an upper bound for $\varphi(z_i')$ by approximating φ with a piecewise linear function. Let $\hat{z}_i$ be the maximum of z_i' over all swaps in the support of $\mathcal{D}$ (we will obtain an upper bound on $\hat{z}_i$ later using Lemma 12, item 3). Let $M = \frac{\varphi(\hat{z}_i)-\varphi(z_i)}{\hat{z}_i - z_i}$ and $m = \varphi(z_i)/z_i = z_i^{r-1}$. Note that $M \le \varphi'(\hat{z}_i)$ when $\hat{z}_i > z_i$, by the Mean Value Theorem and monotonicity of φ'. Since φ is convex, we have (see Figure 1)

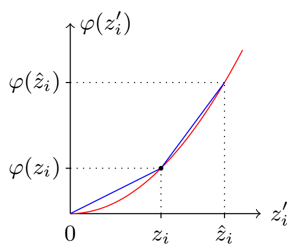

Figure 1 We approximate $\varphi(z_i')$ on the interval $[0, \hat{z}_i]$ using a piecewise linear function consisting of two segments: the first with slope m, and the second with slope M.

$$\varphi(z_i') \le \begin{cases} \varphi(z_i) + m\cdot(z_i' - z_i), & \text{if } z_i' < z_i \\ \varphi(z_i) + M\cdot(z_i' - z_i), & \text{if } z_i' \ge z_i \end{cases}$$

Define coefficients ξ_{ab} as follows

$$\xi_{ab} = \begin{cases} m, & \text{if } z_i' < z_i \\ M, & \text{otherwise} \end{cases}$$

Then $\varphi(z_i') \leq \varphi(z_i) + \xi_{ab} \cdot (z_i' - z_i)$, and we have $\Delta_i \equiv \mathbb{E}\left[\varphi(z_i') - \varphi(z_i)\right] \leq \mathbb{E}\left[\xi_{ab} \cdot (z_i' - z_i)\right]$.

We are now in a position to use Lemma 12, from which we derive the following key inequality, whose proof is deferred to Section 5.5.

$$\Delta_i \leq \frac{1}{4k}\left(\beta^q \varphi(z_i^*) - \varphi(z_i)\right) \tag{7}$$

with $\beta = O(q)$. Adding up the inequality over all coordinates $i \in [\ell]$, we get

$$\mathbb{E}\left[\Phi_{ab} - \Phi\right] = \sum_{i=1}^{\ell} \Delta_i \leq \frac{1}{4k}\left(\beta^q \Phi^* - \Phi\right).$$

In particular, this means that there is always a swap $c_a \to c_b^*$, for which

$$\Phi_{ab} - \Phi \leq \frac{1}{4k}\left(\beta^q \Phi^* - \Phi\right).$$

5.3 Putting everything together

Consider the local search algorithm that at each iteration finds a swap $c_a \to c'$ with $c_a \in C$ and $c' \in F \setminus C$ that produces the greatest decrease in the potential function Φ, where C is the current set of open centers and F is the candidate facility set. The algorithm terminates when no swap decreases the current potential Φ by more than $\frac{\Phi}{8k}$.

5.3.1 Bounding the number of iterations

Suppose that $\Phi > 2\beta^q \Phi^*$. Then, for the improving swap $c_a \to c_b^*$ identified above, we have

$$\Phi_{ab} - \Phi \leq -\frac{\Phi}{8k}$$

and

$$\Phi_{ab} - \Phi^* \leq (\Phi - \Phi^*) - \frac{\Phi}{8k} \leq \left(1 - \frac{1}{8k}\right)(\Phi - \Phi^*).$$

Thus when the algorithm stops, we must have $\Phi \leq 2\beta^q \Phi^*$. Until the algorithm stops, in every iteration, the above inequality holds. It follows that after at most $8k \log_e(\Phi/\Phi^*)$ iterations, the algorithm will find a clustering of cost at most $2^{1/q} \cdot \beta \cdot (\Phi^*)^{1/q}$.

As is standard, by truncating distances and rounding weights, we may assume that all distances and weights are integers bounded polynomially in n. Therefore, the cost of any clustering is polynomial in n; hence, $\log_e(\Phi/\Phi^*)$ is bounded by $O(q \log n)$. We conclude that the algorithm returns a solution of cost at most $O(q)$ OPT in at most $O(qk \log n)$ iterations.

5.3.2 Running time per iteration

In each iteration, we maintain the current set of centers C. For each point u, we maintain its closest and second-closest centers $s_1(u)$ and $s_2(u)$ in C. At the beginning of every iteration, these quantities can be computed in time $O(nk)$. We can compute the change in cost due to a swap $a \to b$ for each point u using $s_1(u), s_2(u)$ and the distances $d(u,a), d(u,b)$. Then we add these costs for each group in time $O_{p,q}(n\ell)$ and re-compute the potential. Thus, for every swap $a \to b$, we spend time $O_{p,q}(n\ell)$ and there are $O(nk)$ swaps, giving a per-iteration running time of $O(n^2k\ell)$. Since there are $O(kq\log n)$ iterations, the running time is at most $O_{p,q}(n^2\ell k^2 \log n)$. This concludes the proof of Theorem 4.

5.4 Proof of Lemma 12

In this Section, we prove Lemma 12.

Proof.

Defining the distribution of swaps. For each center $c_b^* \in C^*$, let $\gamma(c_b^*)$ denote the nearest center in C (see Figure 2). We partition the centers in C based on the number of preimages under γ:

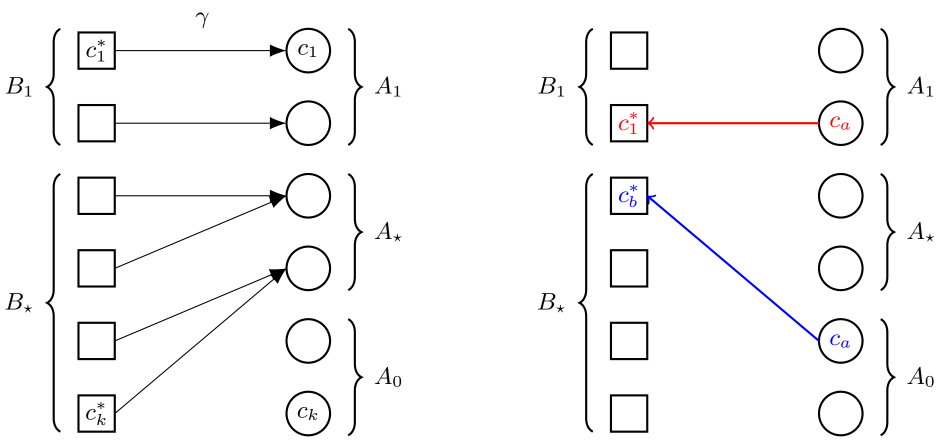

Figure 2 The left figure shows the map γ that maps every optimal center c_i^* to the closest current center c_j. This map defines the partition of the optimal centers into sets B_1 and $B_\star$ and the current centers into sets A_0, A_1, and $A_\star$. The right figure illustrates two possible types of swaps $c_a \to c_b$.

- A_1: centers with exactly one preimage,
- A_0: centers with no preimages,
- $A_\star$: centers with two or more preimages.

Define $B_1 = \gamma^{-1}(A_1)$ and $B_\star = \gamma^{-1}(A_\star)$. We observe:

- γ is a one-to-one map between B_1 and A_1,
- γ maps $B_\star$ onto $A_\star$, with every $c \in A_\star$ having at least two preimages.

Denote $s = |A_1|$. From the observations above, we get $|B_1| = |A_1| = s$ and $|A_\star| \le |B_\star|/2 = (k-s)/2$. Thus, $|A_0| = k - |A_1| - |A_\star| \ge (k-s)/2$. Now we are ready to define a probability distribution $\mathcal{D}$ over swaps.

1. With probability $\frac{s}{k}$, select $c_b^* \in B_1$ uniformly at random and set $c_a = \gamma(c_b^*) \in A_1$,
2. With probability $\frac{k-s}{k}$, select c_a uniformly from A_0 and c_b^* uniformly from $B_\star$.

This yields a random swap $c_a \to c_b^*$. Note that b is uniformly distributed in $[k]$ and therefore the probability that $b = b_0$ is $1/k$ for every fixed $b_0 \in [k]$. In general, a is not uniformly distributed; however, for every fixed $a_0 \in [k]$, the probability that $a = a_0$ is at most $\max(\frac{s}{k} \cdot \frac{1}{|A_1|}, \frac{k-s}{k} \cdot \frac{1}{|A_0|}) \le \max(1/k, 2/k) = 2/k$:

$$\Pr\{a = a_0\} \le 2/k \qquad \text{and} \qquad \Pr\{b = b_0\} = 1/k \tag{8}$$

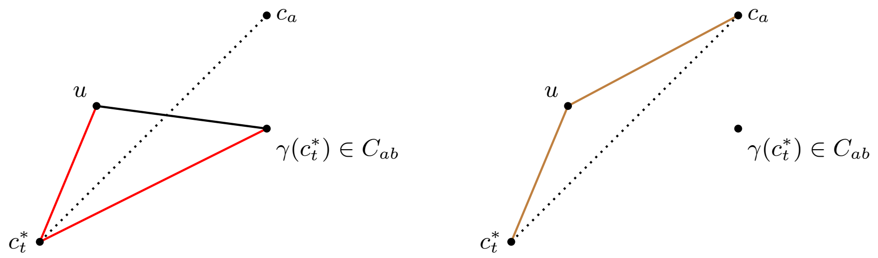

Figure 3 We assign point u to $\gamma(c_t^*)$, which is guaranteed to be in C_{ab}. The distance from u to $\gamma(c_t^*)$ is upper bounded by $d(u, c_t^*) + d(c_t^*, \gamma(c_t^*))$. Further, $d(c_t^*, \gamma(c_t^*)) \le d(c_t^*, c_a)$, since $\gamma(c_t^*)$ is the closest current center to c_t^*. In turn, $d(c_t^*, c_a)$ is upper bounded by $d(c_t^*, u) + d(u, c_a)$.

We will need below that for all $t \ne b$,

$$\gamma(c_t^*) \ne c_a. \tag{9}$$

Indeed, in item 1, $\gamma^{-1}(c_a) = \{c_b^*\}$; in item 2, $\gamma^{-1}(c_a) = \varnothing$. In either case, $c_t^* \notin \gamma^{-1}(c_a)$.

Upper bounding connection cost

Let $x_u = d(u, C)$ denote the distance from u to the closest center in the current solution C, $x'_u = d(u, C_{ab})$ the updated distance after the swap, and $x_u^* = d(u, C^*)$ the distance to the closest center in C^*. Let V_a be the Voronoi cell of c_a in C, and V_b^* the Voronoi cell of c_b^* in C^*. For $u \in V_b^*$,

$$x_u'^p - x_u^p \le x_u^{*p} - x_u^p.$$

For $u \in V_a \setminus V_b^*$, assume $u \in V_t^*$ for some $t \ne b$. By (9), $\gamma(c_t^*) \ne c_a$; thus, $\gamma(c_t^*) \in C_{ab}$ and

$$\begin{aligned} x'_u = d(u, C_{ab}) &\le d(u, \gamma(c_t^*)) \le d(u, c_t^*) + d(c_t^*, \gamma(c_t^*)) \le d(u, c_t^*) + d(c_t^*, c_a) \\ &\le d(u, c_t^*) + \big(d(c_t^*, u) + d(u, c_a)\big) = 2x_u^* + x_u. \end{aligned}$$

Here, we used that $d(c_t^*, \gamma(c_t^*)) \le d(c_t^*, c_a)$, since $\gamma(c_t^*)$ is the nearest center in C to c_t^*.

Applying Claim 7 with $\delta = (1+\varepsilon)^{p-1} - 1$ and $\theta = 2^p(1 + 1/\varepsilon)^{p-1}$, we get

$$x_u'^p \le (x_u + 2x_u^*)^p \le (1+\delta)x_u^p + \theta x_u^{*p}.$$

Thus,

$$x_u'^p - x_u^p \le \delta x_u^p + \theta x_u^{*p}.$$

Finally, if $u \notin V_a \cup V_b^*$, then $x'_u \le x_u$. We have,

$$\begin{aligned} \mathrm{cost}_{ab}(w) - \mathrm{cost}(w) &\le \sum_{u \in V_b^*} w_u \left(x_u'^p - x_u^p\right) + \sum_{u \in V_a \setminus V_b^*} w_u \left(x_u'^p - x_u^p\right) && (10) \\ &\le \sum_{u \in V_b^*} w_u \left(x_u^{*p} - x_u^p\right) + \sum_{u \in V_a} w_u \left(\delta x_u^p + \theta x_u^{*p}\right). && (11) \end{aligned}$$

We now derive an upper bound for $\xi_{ab} \cdot (\mathrm{cost}_{ab}(w) - \mathrm{cost}(w))$. Since $0 < m \le \xi_{ab} \le M$, we can replace ξ_{ab} with M in positive terms and with m in negative terms. We have:

$$\xi_{ab} \cdot (\mathrm{cost}_{ab}(w) - \mathrm{cost}(w)) \le \sum_{u \in V_b^*} w_u \left(M x_u^{*p} - m x_u^p\right) + M \sum_{u \in V_a} w_u \left(\delta x_u^p + \theta x_u^{*p}\right)$$

Now we assume that the swap $c_a \to c_b^*$ is sampled from the distribution $\mathcal{D}$ defined above. We get

$$\begin{aligned}&\mathbb{E}\left[\xi_{ab}\cdot(\mathrm{cost}_{ab}(w)-\mathrm{cost}(w))\right]\\&\quad\le\sum_{u\in X}\Pr\{u\in V_b^*\}\cdot w_u\left(Mx_u^{*p}-mx_u^p\right)+M\sum_{u\in X}\Pr\{u\in V_a\}\cdot w_u\left(\delta x_u^p+\theta x_u^{*p}\right)\end{aligned}$$

From 8, we get $\Pr\{u\in V_b^*\}=\frac{1}{k}$ and $\Pr\{u\in V_a\}\le\frac{2}{k}$ and thus

$$k\mathbb{E}\left[\xi_{ab}\cdot(\mathrm{cost}_{ab}(w)-\mathrm{cost}(w))\right]\le(1+2\theta)M\,\mathrm{cost}^*(w)-(m-2\delta M)\,\mathrm{cost}(w). \tag{12}$$

We will use (12) to derive (4), (5), and (6). We will use different values of parameters ε, δ, and θ in these derivations. We will need the following lemma.

▶ **Lemma 13.** *For every $p\ge 1$ and every value of $\delta_0\in(0,1/4)$, there exists $\varepsilon>0$ such that $\delta=(1+\varepsilon)^{p-1}-1\le\delta_0$ and $\theta\le 2^p\left(\frac{5p}{4\delta_0}\right)^{p-1}$.*

Proof. If $p=1$, we let $\varepsilon=1$, resulting in $\delta=0$ and $\theta=2$, and we are done. If $p>1$, we choose ε so that $\delta=\delta_0$. Applying the Mean Value Theorem, we get

$$\varepsilon=(1+\delta_0)^{1/(p-1)}-1\ge\delta_0\cdot\min_{t\in(0,\delta_0)}\frac{(1+t)^{1/(p-1)-1}}{p-1}\ge\frac{4}{5}\cdot\frac{\delta_0}{p-1},$$

where we used that $1+t\le 1+\delta_0\le 5/4$ and $1/(p-1)-1\ge -1$. Therefore,

$$\theta=2^p(1+1/\varepsilon)^{p-1}\le 2^p\left(\frac{5(p-1)+4\delta_0}{4\delta_0}\right)^{p-1}\le 2^p\left(\frac{5p}{4\delta_0}\right)^{p-1}.$$ ◀

First, we establish the upper bound (4). We choose parameters according to Lemma 13 with $\delta_0=m/(4M)$. Then $\theta\le 2^p(5pM/m)^{p-1}$. Since $0.5\cdot 2^p(5Mp/m)^{p-1}\ge 2^{p-1}\ge 1$, we have

$$(1+2\theta)M\le 2.5\cdot 2^p\left(\frac{5Mp}{m}\right)^{p-1}\cdot\frac{M}{m}\cdot m\le\frac{1}{2}\left(\frac{10Mp}{m}\right)^p m.$$

We conclude that

$$\mathbb{E}\left[\xi_{ab}\cdot(\mathrm{cost}_{ab}(w)-\mathrm{cost}(w))\right]\le\frac{m}{2k}\left(\left(\frac{10Mp}{m}\right)^p\mathrm{cost}^*(w)-\mathrm{cost}(w)\right).$$

Now we prove (5). We set $\varepsilon=1$; then $\theta=2^{2p-1}$ and $\delta=2^{p-1}-1$. From (12), we get

$$\begin{aligned}&k\mathbb{E}\left[\xi_{ab}\cdot(\mathrm{cost}_{ab}(w)-\mathrm{cost}(w))\right]\\&\le(1+2\theta)M\,\mathrm{cost}^*(w)+2\delta M\,\mathrm{cost}(w)\le(1+2^{2p})M(\mathrm{cost}(w)+\mathrm{cost}^*(w))\end{aligned}$$

Finally, for every swap $c_a\to c_b^*$ in the support of $\mathcal{D}$, we get the following upper bound from (11),

$$\begin{aligned}\mathrm{cost}_{ab}(w)-\mathrm{cost}(w)&\le\sum_{u\in V_b^*}w_u\left(x_u^{*p}-x_u^p\right)+\sum_{u\in V_a}w_u\left(\delta x_u^p+\theta x_u^{*p}\right)\\&\le\sum_{u\in X}w_ux_u^{*p}+\sum_{u\in X}w_u\left(\delta x_u^p+\theta x_u^{*p}\right)\le(1+\theta)\,\mathrm{cost}^*(w)+\delta\,\mathrm{cost}(w).\end{aligned}$$

Choosing ε according to Lemma 13, we get the desired bound. ◀

5.5 Proof of Inequality 7

Applying Lemma 12, we get for $\alpha = 10Mp/m$,

$$\Delta_i \le \frac{m}{2k}\left(\alpha^p z_i^* - z_i\right) \tag{13}$$

$$\Delta_i \le \frac{(1+2^{2p})M}{k}(z_i + z_i^*). \tag{14}$$

From Lemma 12, item 3, with $\delta_0 = 1/(5r)$ and

$$\theta_0 = 2^p\left(\frac{25pr}{4}\right)^{p-1},$$

we get

$$\widehat{z}_i \le (1+\theta_0)z_i^* + (1+\delta_0)z_i \le \left(1+\frac{1}{5r}\right)z_i + (1+\theta_0)z_i^*. \tag{15}$$

Fix a threshold $T = \max\left(\frac{5r(1+\theta_0)}{4},\ 2(10eq)^p\right) = O(q)^p$. Consider two cases.

Current cost z_i is small: $z_i \le Tz_i^*$

From (15) and $1+\theta_0 \le \frac{4}{5r}T$,

$$\hat{z}_i \le \frac{6}{5}Tz_i^* + \frac{4}{5}Tz_i^* = 2Tz_i^*.$$

Accordingly, when $\hat{z}_i > z_i$, $M \le \varphi'(\hat{z}_i) = r\hat{z}_i^{r-1} \le r(2Tz_i^*)^{r-1}$. Also, $m \le M$. By (14)

$$\begin{aligned}\Delta_i &\le \frac{(4^p+1)M}{k}(z_i^* + z_i) = \frac{(4^p+1)}{k}\left(M\cdot(z_i^* + z_i + \frac{m}{M}z_i) - m\cdot z_i\right)\\ &\le \frac{(4^p+1)}{k}((2T+1)Mz_i^* - mz_i) \le \frac{1}{2k}\left(4^{p+2}\cdot r\cdot(2T)^r\varphi(z_i^*) - \varphi(z_i)\right).\end{aligned}$$

Letting $\beta = (2\cdot 4^{p+2}\cdot r\cdot(2T)^r)^{1/q} = O(q)$, we get

$$\Delta_i \le \frac{1}{4k}\left(\beta^q\varphi(z_i^*) - \varphi(z_i)\right).$$

Current cost z_i is large: $z_i \ge Tz_i^*$

We show that $2\alpha^p \le T$. From (15) and $1+\theta_0 \le \frac{4}{5r}T$, we have

$$\hat{z}_i \le \left(1+\frac{1}{5r}+\frac{1+\theta_0}{T}\right)z_i \le \left(1+\frac{1}{r}\right)z_i.$$

Hence, when $\hat{z}_i > z_i$, $M \le \varphi'(\hat{z}_i) = r\hat{z}_i^{r-1} \le erz_i^{r-1}$ and $\alpha \le 10p\cdot(M/m) \le 10p\cdot(er) = 10eq$. Using that $z_i \ge Tz_i^* \ge 2(10eq)^p z_i^* \ge 2\alpha^p z_i^*$ and $\varphi(z_i) = mz_i$, we get from (13),

$$\Delta_i \le \frac{m}{2k}(\alpha^p z_i^* - z_i) \le -\frac{m}{2k}\cdot\frac{z_i}{2} = -\frac{1}{4k}\varphi(z_i).$$

6 Socially fair facility location

In this section, we consider two variants of the Socially Fair Facility Location problem and provide approximation algorithms for them.

We are given a set of n points X and a metric d on it, forming a metric space (X, d), as well as a set of possible facility locations F. There are ℓ groups on X represented by a weight function $w : X \to \mathbb{R}^{\ell}_{\geq 0}$ where $w_i(u) \geq 0$ denotes the weight of client u in group i. Each facility $c \in F$ is associated with a facility opening cost $f(c)$. We consider the following two settings with different types of facility opening costs: (1) group-homogeneous facility cost; (2) group-heterogeneous facility cost.

Group-homogeneous facility costs

Each facility $c \in F$ has a scalar opening cost $f(c) \geq 0$, paid equally across groups. The objective is to choose $C \subseteq F$ minimizing

$$\mathrm{FL}_{\mathrm{hom}}(w, f, C) := \left(\sum_{i=1}^{\ell} \mathrm{cost}(w_i, C)^{q/p} \right)^{1/q} + \sum_{c \in C} f(c),$$

where $\mathrm{cost}(w_i, C) = \sum_{u \in X} w_i(u)\, d(u, C)^p$.

[2] studied this objective with $p = 1$ and $q = \infty$ and gave a 4-approximation algorithm for it.

Group-heterogeneous facility costs

Each facility $c \in F$ has a vector opening cost $f(c) \in \mathbb{R}^{\ell}_{\geq 0}$, where group i pays $f_i(c)$ to open c.[5]

In this case, the objective is[6]

$$\mathrm{FL}_{\mathrm{het}}(w, f, C) \;:=\; \left(\sum_{i=1}^{\ell} \left(\mathrm{cost}(w_i, C)^{1/p} + \sum_{c \in C} f_i(c) \right)^{q} \right)^{1/q}.$$

We get the following results for homogeneous and heterogeneous facility costs.

▶ **Theorem 5.** *For Socially Fair Facility Location with group-homogeneous facility cost, there exists a polynomial-time algorithm that achieves a $(4 + \varepsilon)$-approximation for any $\varepsilon > 0$.*

▶ **Theorem 6.** *For Socially Fair Facility Location with group-heterogeneous facility cost, there exists a polynomial-time algorithm that achieves an $O(q/\log q)$-approximation.*

We now provide the approximation algorithm for the group-heterogeneous setting.

[5] We interpret $f_i(c)$ as a group-specific *fixed* charge for opening facility c (e.g., taxes/subsidies or administrative overhead). Thus, group i pays $f_i(c)$ whenever c is opened, regardless of whether any group-i client is assigned to c; this keeps opening costs dependent only on the opened set C, as in standard fixed-charge facility location.

[6] If the heterogeneous opening costs are *group-uniform*, i.e., $f_i(c) = f(c)$ for all $i \in [\ell]$ and $c \in F$, then the objective becomes $\left(\sum_{i=1}^{\ell} \left(\mathrm{cost}(C, w_i)^{1/p} + \sum_{c \in C} f(c) \right)^q \right)^{1/q}$. We refer to this as the *group-uniform cost* variant. This objective is generally *different* from the group-homogeneous facility costs above. The two coincide when $q = \infty$, but they differ for finite q. By using our heterogeneous convex relaxation with $f_i(c) = f(c)$ and applying the rounding scheme used for the homogeneous costs, we obtain an $(8 + \varepsilon)$-approximation for this group-uniform cost variant.

Proof of Theorem 6. We use a relaxation of the problem and then use a randomized rounding procedure to get an integer solution. We consider the following relaxation

$$
\begin{aligned}
\min \quad & H \\
\text{s.t.} \quad & \sum_{i=1}^{\ell} z_i \le H^q \\
& z_i \ge \left(\sum_u w_i(u) \sum_{c \in F} d(u,c)^p \cdot x_{u,c} \right)^{\frac{q}{p}} && \forall i \in [\ell] && (16) \\
& z_i \ge \left(\sum_{c \in F} y_c \cdot f_i(c) \right)^q && \forall i \in [\ell] && (17) \\
& z_i \ge \sum_{c \in F} y_c \cdot f_i(c)^q && \forall i \in [\ell] && (18) \\
& \sum_{c \in F} x_{u,c} \ge 1 && \forall u \in X && (19) \\
& x_{u,c} \le y_c && \forall u \in X,\ \forall c \in F \\
& 0 \le y_c \le 1 && \forall c \in F \\
& 0 \le x_{u,c} \le 1 && \forall u \in X,\ \forall c \in F
\end{aligned}
$$

We can binary search on H. For any fixed value of H, the relaxation above is convex.

We then apply the following randomized rounding algorithm inspired by [19, 24]. Let (x^*, y^*, z^*, H^*) denote the fractional solution to the convex program above. The key difference is that, instead of opening the cheapest facility in B_u, we open exactly one facility c_u in B_u selected randomly with probability proportional to y_c^*. For each point $u \in X$, let $A_u = \sum_{c \in F} d(u,c)^p x_{u,c}^*$ denote the connection cost of point u given by the fractional solution. Then, we sort all points in X by their connection costs in increasing order. We iteratively perform the following steps until all points are assigned to open facilities. We first pick the unassigned point u with the smallest connection cost A_u as an anchor point. Let $B_u = \{c \in F : d(u,c)^p \le 2A_u\}$ be all facilities in a ball with radius $(2A_u)^{1/p}$ centered at u. It can be shown that the balls corresponding to distinct anchor points are disjoint. Then, we sample exactly one facility c_u in B_u with probability proportional to y_c^*. Assign all unassigned points v with distance $d(v,u) \le (2A_u)^{1/p} + (2A_v)^{1/p}$ to this new facility c_u. Let $\widehat{C}$ be the set of centers sampled by the rounding algorithm.

We now analyze the cost of this integer solution given by the rounding algorithm. Let $\hat{y}_c \in \{0,1\}$ be the indicator variable to denote whether facility c is selected in $\widehat{C}$. We first analyze the facility opening cost of this solution. By a simple averaging argument, for each iteration with anchor node u, we have $\sum_{c \in B_u} y_c^* \ge 1/2$. Since we open exactly one facility c_u in B_u with probability proportional to y_c^*, we have, for each facility $c \in B_u$,

$$
\mathbb{E}[\hat{y}_c] = \frac{y_c^*}{\sum_{c \in B_u} y_c^*} \le 2y_c^*.
$$

Consider any fixed group i. For each iteration with anchor point u, we define a random variable $Y_u = \sum_{c \in B_u} \hat{y}_c f_i(c)$ which represents the contribution to the facility cost of group i from that iteration. Since the choices made in distinct, disjoint anchor balls are independent, the collection $\{Y_u\}$ forms a set of independent non-negative random variables. By the

Rosenthal inequality (see [18]), we have

$$\mathbb{E}\left[\left(\sum_u Y_u\right)^q\right]^{1/q} \leq O\left(\frac{q}{\log q}\right) \max\left\{\left(\sum_u \mathbb{E}[Y_u^q]\right)^{1/q}, \sum_u \mathbb{E}[Y_u]\right\}.$$

Note that in each iteration, exactly one facility c_u in B_u is opened, which means exactly one random variable $\hat{y}_{c_u} = 1$ and all other $\hat{y}_c = 0$ in B_u. Thus, we have

$$\mathbb{E}[Y_u^q] = \mathbb{E}\left[\left(\sum_{c\in B_u} \hat{y}_c \cdot f_i(c)\right)^q\right] = \mathbb{E}\left[\sum_{c\in B_u} \hat{y}_c \cdot f_i(c)^q\right] \leq 2\sum_{c\in B_u} y_c^* \cdot f_i(c)^q.$$

Hence, we have $\sum_u \mathbb{E}[Y_u^q] \leq 2\sum_{c\in\bigcup_u B_u} y_c^* \cdot f_i(c)^q \leq 2z_i^*$. We also have

$$\sum_u \mathbb{E}[Y_u] = \sum_{c\in\bigcup_u B_u} f_i(c)\mathbb{E}[\hat{y}_c] \leq 2\sum_{c\in\bigcup_u B_u} f_i(c) y_c^* \leq 2(z_i^*)^{1/q} \leq O\left(\frac{q}{\log q}\right) H^*.$$

Therefore, we have

$$\mathbb{E}\left[\left(\sum_{c\in\bigcup_u B_u} \hat{y}_c f_i(c)\right)^q\right] \leq O\left(\left(\frac{2q}{\log q}\right)^q\right) z_i^*.$$

By taking the sum over all groups and applying Markov's inequality, with probability at least 1/2, we have

$$\sum_{i=1}^{\ell}\left(\sum_{c\in\bigcup_u B_u} \hat{y}_c f_i(c)\right)^q \leq O\left(\left(\frac{2q}{\log q}\right)^q\right)\sum_{i=1}^{\ell} z_i^* \leq O\left(\left(\frac{2q}{\log q}\right)^q\right)(H^*)^q$$

We then analyze the connection cost of this integer solution. If v is assigned when u is the anchor, then $A_u \leq A_v$ and $d(v, c_u) \leq d(v, u) + d(u, c_u) \leq 3(2A_v)^{1/p}$. Hence, the connection cost for each group i given by $\widehat{C}$ is at most

$$\operatorname{cost}(w_i, \widehat{C})^{1/p} = \left(\sum_{u\in X} w_i(u) \cdot d(u, \widehat{C})^p\right)^{1/p} \leq \left(3^p \cdot 2\sum_{u\in X} w_i(u) \cdot A_u\right)^{1/p} \leq 6(z_i^*)^{1/q}.$$

By combining with the facility opening cost, the facility location cost is at most

$$\begin{aligned}&\left(\sum_{i=1}^{\ell}\left(\operatorname{cost}(w_i, \widehat{C})^{1/p} + \sum_{c\in\widehat{C}} f_i(c)\right)^q\right)^{1/q}\\ \leq &\left(\sum_{i=1}^{\ell} 2^q \operatorname{cost}(w_i, \widehat{C})^{q/p} + 2^q\left(\sum_{c\in\widehat{C}} f_i(c)\right)^q\right)^{1/q} \leq O\left(\frac{q}{\log q}\right)\cdot H^*.\end{aligned}$$

Therefore, by computing a $(1+\varepsilon)$-approximate solution to the convex relaxation, this randomized rounding provides the desired approximation with probability at least 1/2. ◀

References

1 Fateme Abbasi, Sandip Banerjee, Jarosław Byrka, Parinya Chalermsook, Ameet Gadekar, Kamyar Khodamoradi, Dániel Marx, Roohani Sharma, and Joachim Spoerhase. Parameterized approximation schemes for clustering with general norm objectives. In *Proceedings of the Symposium on Foundations of Computer Science (FOCS)*, pages 1377–1399, 2023. doi:10.1109/FOCS57990.2023.00085.

2 Mohsen Abbasi, Aditya Bhaskara, and Suresh Venkatasubramanian. Fair clustering via equitable group representations. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 504–514, 2021. doi:10.1145/3442188.3445913.

3 Sara Ahmadian, Alessandro Epasto, Marina Knittel, Ravi Kumar, Mohammad Mahdian, Benjamin Moseley, Philip Pham, Sergei Vassilvitskii, and Yuyan Wang. Fair hierarchical clustering. *Advances in Neural Information Processing Systems*, 33:21050–21060, 2020.

4 Sara Ahmadian, Alessandro Epasto, Ravi Kumar, and Mohammad Mahdian. Clustering without over-representation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 267–275, 2019. doi:10.1145/3292500.3330987.

5 Noga Alon, Raphael Yuster, and Uri Zwick. Color-coding. *Journal of the ACM (JACM)*, 42(4):844–856, 1995. doi:10.1145/210332.210337.

6 Barbara Anthony, Vineet Goyal, Anupam Gupta, and Viswanath Nagarajan. A plant location guide for the unsure: Approximation algorithms for min-max location problems. *Mathematics of Operations Research*, 35(1):79–101, 2010. doi:10.1287/MOOR.1090.0428.

7 Vijay Arya, Naveen Garg, Rohit Khandekar, Adam Meyerson, Kamesh Munagala, and Vinayaka Pandit. Local search heuristics for k-median and facility location problems. *SIAM Journal on Computing*, 33(3):544–562, 2004. doi:10.1137/S0097539702416402.

8 Sayan Bandyapadhyay, Tianzhi Chen, Zachary Friggstad, and Mahya Jamshidian. A constant-factor approximation for pairwise fair k-center clustering. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 43–57. Springer, 2025. doi:10.1007/978-3-031-93112-3_4.

9 Sayan Bandyapadhyay, Eden Chlamtáč, Zachary Friggstad, Mahya Jamshidian, Yury Makarychev, and Ali Vakilian. A polynomial-time approximation for pairwise fair k-median clustering. In *The 29th International Conference on Artificial Intelligence and Statistics*, 2026.

10 Sayan Bhattacharya, Parinya Chalermsook, Kurt Mehlhorn, and Adrian Neumann. New approximability results for the robust k-median problem. In *Scandinavian Workshop on Algorithm Theory*, pages 50–61. Springer, 2014. doi:10.1007/978-3-319-08404-6_5.

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

12 Rajesh Chitnis, Marek Cygan, MohammadTaghi Hajiaghayi, Marcin Pilipczuk, and Michał Pilipczuk. Designing fpt algorithms for cut problems using randomized contractions. *SIAM Journal on Computing*, 45(4):1171–1229, 2016. doi:10.1137/15M1032077.

13 Eden Chlamtáč, Yury Makarychev, and Ali Vakilian. Approximating fair clustering with cascaded norm objectives. In *Proceedings of the Symposium on Discrete Algorithms (SODA)*, pages 2664–2683, 2022. doi:10.1137/1.9781611977073.104.

14 Mehrdad Ghadiri, Samira Samadi, and Santosh Vempala. Socially fair k-means clustering. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, pages 438–448, 2021. doi:10.1145/3442188.3445906.

15 Mehrdad Ghadiri, Mohit Singh, and Santosh S Vempala. Constant-factor approximation algorithms for socially fair k-clustering. *arXiv preprint arXiv:2206.11210*, 2022. doi:10.48550/arXiv.2206.11210.

16 Dishant Goyal and Ragesh Jaiswal. Tight fpt approximation for socially fair clustering. *Information Processing Letters*, 182:106383, 2023. doi:10.1016/J.IPL.2023.106383.

17 Anupam Gupta and Kanat Tangwongsan. Simpler analyses of local search algorithms for facility location. *arXiv preprint arXiv:0809.2554*, 2008. arXiv:0809.2554.

18 William B Johnson, Gideon Schechtman, and Joel Zinn. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *The Annals of Probability*, pages 234–253, 1985.

19 Adam Kasperski, Adam Kurpisz, and Paweł Zieliński. Approximability of the robust representatives selection problem. *Operations Research Letters*, 43(1):16–19, 2015. doi:10.1016/J.ORL.2014.10.007.

20 Sepideh Mahabadi and Ali Vakilian. Individual fairness for k-clustering. In *International conference on machine learning*, pages 6586–6596. PMLR, 2020. URL: http://proceedings.mlr.press/v119/mahabadi20a.html.

21 Konstantin Makarychev, Yury Makarychev, and Ilya Razenshteyn. Performance of johnson–lindenstrauss transform for k-means and k-medians clustering. *SIAM Journal on Computing*, 52(2):269–297, 2022.

22 Yury Makarychev and Ali Vakilian. Approximation algorithms for socially fair clustering. In *Conference on Learning Theory*, pages 3246–3264. PMLR, 2021. URL: http://proceedings.mlr.press/v134/makarychev21a.html.

23 David B Shmoys, Éva Tardos, and Karen Aardal. Approximation algorithms for facility location problems. In *Proceedings of the Symposium on Theory of Computing*, pages 265–274, 1997.

24 Aravind Srinivasan. Approximation algorithms via randomized rounding: a survey. *Series in Advanced Topics in Mathematics, Polish Scientific Publishers PWN*, pages 9–71, 1999.