



A Technique for Obtaining True Approximations for k -Center with Covering Constraints

Georg Anegg, Haris Angelidakis, Adam Kurpisz, and Rico Zenklusen^(✉)

Department of Mathematics, ETH Zurich, Zurich, Switzerland
{ganegg,angelidc,kurpiza,ricoz}@ethz.ch

Abstract. There has been a recent surge of interest in incorporating fairness aspects into classical clustering problems. Two recently introduced variants of the k -Center problem in this spirit are Colorful k -Center, introduced by Bandyapadhyay, Inamdar, Pai, and Varadarajan, and lottery models, such as the Fair Robust k -Center problem introduced by Harris, Pensyl, Srinivasan, and Trinh. To address fairness aspects, these models, compared to traditional k -Center, include additional covering constraints. Prior approximation results for these models require to relax some of the normally hard constraints, like the number of centers to be opened or the involved covering constraints, and therefore, only obtain constant-factor pseudo-approximations. In this paper, we introduce a new approach to deal with such covering constraints that leads to (true) approximations, including a 4-approximation for Colorful k -Center with constantly many colors—settling an open question raised by Bandyapadhyay, Inamdar, Pai, and Varadarajan—and a 4-approximation for Fair Robust k -Center, for which the existence of a (true) constant-factor approximation was also open.

We complement our results by showing that if one allows an unbounded number of colors, then Colorful k -Center admits no approximation algorithm with finite approximation guarantee, assuming that $P \neq NP$. Moreover, under the Exponential Time Hypothesis, the problem is inapproximable if the number of colors grows faster than logarithmic in the size of the ground set.

1 Introduction

Along with k -Median and k -Means, k -Center is one of the most fundamental and heavily studied clustering problems. In k -Center, we are given a finite metric

G. Anegg and R. Zenklusen—Research supported by Swiss National Science Foundation grant 200021_184622.

H. Angelidakis and R. Zenklusen—This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 817750).

A. Kurpisz—Research supported by Swiss National Science Foundation grant PZ00P2_174117.

© Springer Nature Switzerland AG 2020

D. Bienstock and G. Zambelli (Eds.): IPCO 2020, LNCS 12125, pp. 52–65, 2020.

https://doi.org/10.1007/978-3-030-45771-6_5

space (X, d) and an integer $k \in [|X|] := \{1, \dots, |X|\}$, and the task is to find a set $C \subseteq X$ with $|C| \leq k$ minimizing the maximum distance of any point in X to its closest point in C . Equivalently, the problem can be phrased as covering X with k balls of radius as small as possible, i.e., finding the smallest radius $r \in \mathbb{R}_{\geq 0}$ together with a set $C \subseteq X$ with $|C| \leq k$ such that $X = B(C, r) := \bigcup_{c \in C} B(c, r)$, where $B(c, r) := \{u \in X : d(c, u) \leq r\}$ is the ball of radius r around c .

k -Center, like most clustering problems, is computationally hard; actually it is NP-hard to approximate to within any constant below 2 [18]. On the positive side, various 2-approximations [12, 16] have been found, and thus, its approximability is settled. Many variations of k -Center have been studied, most of which are based on generalizations along one of the following two main axes:

- (i) which sets of centers can be selected, and
- (ii) which sets of points of X need to be covered.

The most prominent variations along (i) are variations where the set of centers is required to be in some down-closed family $\mathcal{F} \subseteq 2^X$. For example, if centers have non-negative opening costs and there is a global budget for opening centers, Knapsack Center is obtained. If \mathcal{F} are the independent sets of a matroid, the problem is known as Matroid Center. The best-known problem type linked to (ii) is Robust k -Center. Here, an integer $m \in [|X|]$ is given, and one only needs to cover any m points of X with k balls of radius as small as possible. Research on k -Center variants along one or both of these axes has been very active and fruitful, see, e.g., [7, 9, 10, 17]. In particular, a recent elegant framework of Chakrabarty and Negahbani [8] presents a unifying framework for designing best possible approximation algorithms for all above-mentioned variants.

All the above variants have in common that there is a single covering requirement; either all of X needs to be covered or a subset of it. Moreover, they come with different kinds of packing constraints on the centers to be opened as in Knapsack or Matroid Center. However, the desire to address fairness in clustering, which has received significant attention recently, naturally leads to multiple covering constraints. Here, existing techniques only lead to constant-factor pseudo-approximations that violate at least one constraint, like the number of centers to be opened. In this work, we present techniques for obtaining (true) approximations for two recent fairness-inspired generalizations of k -Center along axis (ii), namely

- (i) γ -Colorful k -Center, as introduced by Bandyapadhyay et al. [3], and
- (ii) Fair Robust k -Center, a lottery model introduced by Harris et al. [15].

γ -Colorful k -Center (γ CkC) is a fairness-inspired k -Center model imposing covering constraints on subgroups. It is formally defined as follows:¹

¹ The version introduced in [3] requires X_1, \dots, X_γ to partition X . However, γ CkC readily reduces to the more restrictive model in [3] by replacing an element with q colors by q elements on the same location with each having a single color.

Definition 1 (γ -Colorful k -Center (γ CkC) [3]). Let $\gamma, k \in \mathbb{Z}_{\geq 1}$, (X, d) be a finite metric space, $X_\ell \subseteq X$ for $\ell \in [\gamma]$, and $m \in \mathbb{Z}_{\geq 0}^\gamma$. The γ -Colorful k -Center problem (γ CkC) asks to find the smallest radius $r \in \mathbb{R}_{\geq 0}$ together with centers $C \subseteq X$, $|C| \leq k$ such that

$$|B(C, r) \cap X_\ell| \geq m_\ell \quad \forall \ell \in [\gamma] .$$

Such a set of centers C is called a (γ CkC) solution of radius r .

The name stems from interpreting each set X_ℓ for $\ell \in [\gamma]$ as a color assigned to the elements of X_ℓ . In particular, an element can have multiple colors or no color. In words, the task is to open k centers of smallest possible radius such that, for each color $\ell \in [\gamma]$, at least m_ℓ points of color ℓ are covered. Hence, for $\gamma = 1$, we recover the Robust k -Center problem.

We briefly contrast γ CkC with related fairness models. A related class of models that has received significant attention also assumes that the ground set is colored, but requires that each cluster contains approximately the same number of points from each color. Such variants have been considered for k -Median, k -Means, and k -Center, e.g., see [2, 4, 5, 11, 23] and references therein. γ CkC differentiates itself from the above notion of fairness by not requiring a per-cluster guarantee, but a global fairness guarantee. More precisely, each color can be thought of as representing a certain group of people (demographic), and a global covering requirement is given per demographic. Also notice the difference with the well-known Robust k -Center problem, where a feasible solution might, potentially, completely ignore a certain subgroup, resulting in a heavily unfair treatment. γ CkC addresses this issue.

The presence of multiple covering constraints in γ CkC, imposed by the colors, hinders the use of classical k -Center clustering techniques, which, as mentioned above, have mostly been developed for packing constraints on the centers to be opened. An elegant first step was done by Bandyapadhyay et al. [3]. They exploit sparsity of a well-chosen LP (in a similar spirit as in [15]) to obtain the following pseudo-approximation for γ CkC: they efficiently compute a solution of twice the optimal radius by opening at most $k + \gamma - 1$ centers. Hence, up to $\gamma - 1$ more centers than allowed may have to be opened. Moreover, [3] shows that in the Euclidean plane, a significantly more involved extension of this technique allows for obtaining a $(17 + \varepsilon)$ -approximation for $\gamma = O(1)$. Unfortunately, this approach is heavily problem-tailored and does not even extend to 3-dimensional Euclidean spaces. This naturally leads to the main open question raised in [3]:

Does γ CkC with $\gamma = O(1)$ admit an $O(1)$ -approximation, for any finite metric?

Here, we introduce a new approach that answers this question affirmatively.

Together with additional ingredients, our approach also applies to Fair Robust k -Center, which is a natural lottery model introduced by Harris et al. [15]. We introduce the following generalization thereof that can be handled with our techniques, which we name *Fair γ -Colorful k -Center problem (Fair γ CkC)*. (The Fair Robust k -Center problem, as introduced in [15], corresponds to $\gamma = 1$.)

Definition 2 (Fair γ -Colorful k -Center (Fair γ CkC)). *Given is a γ CkC instance on a finite metric space (X, d) together with a vector $p \in [0, 1]^X$. The goal is to find the smallest radius $r \in \mathbb{R}_{\geq 0}$, together with an efficient procedure returning a random γ CkC solution $C \subseteq X$ of radius r such that*

$$\Pr[u \in B(C, r)] \geq p(u) \quad \forall u \in X .$$

Hence, Fair γ CkC is a generalization of γ CkC, where each element $u \in X$ needs to be covered with a prescribed probability $p(u)$. The Fair Robust k -Center problem, i.e., Fair γ CkC with $\gamma = 1$, is indeed a fairness-inspired generalization of Robust k -Center, since Robust k -Center is obtained by setting $p(u) = 0$ for $u \in X$. One example setting where the additional fairness aspect of Fair γ CkC compared to γ CkC is nicely illustrated, is when k -Center problems have to be solved repeatedly on the same metric space. The introduction of the probability requirements p allows for obtaining a distribution to draw from that needs to consider all elements of X (as prescribed by p), whereas classical Robust k -Center likely ignores a group of badly-placed elements. We refer to Harris et al. [15] for further motivation of the problem setting. They also discuss the Knapsack and Matroid Center problem under the same notion of fairness.

For Fair Robust k -Center, [15] presents a 2-pseudo-approximation that slightly violates both the number of points to be covered and the probability of covering each point. More precisely, for any constant $\varepsilon > 0$, only a $(1 - \varepsilon)$ -fraction of the required number of elements are covered, and element $u \in X$ is covered only with probability $(1 - \varepsilon)p(u)$ instead of $p(u)$. It was left open in [15] whether a true approximation may exist for Fair Robust k -Center.

1.1 Our Results

Our main contribution is a method to obtain 4-approximations for variants of k -Center with unary encoded covering constraints on the points to be covered. We illustrate our technique in the context of γ CkC, affirmatively resolving the open question of Bandyapadhyay et al. [3] about the existence of an $O(1)$ -approximation for constantly many colors (without restrictions on the underlying metric space).

Theorem 3. *There is a 4-approximation for γ CkC running in time $|X|^{O(\gamma)}$.*

In a second step we extend and generalize our technique to Fair γ CkC, which, as mentioned, is a generalization of γ CkC. We show that Fair γ CkC admits a $O(1)$ -approximation, which neither violates covering nor probabilistic constraints.

Theorem 4. *There is a 4-approximation for Fair γ CkC running in time $|X|^{O(\gamma)}$.*

We complete our results by showing inapproximability for γ CkC when γ is not bounded. This holds even on the real line (1-dimensional Euclidean space).

Theorem 5. *It is NP-hard to decide whether γCkC on the real line admits a solution of radius 0. Moreover, unless the Exponential Time Hypothesis fails, for any function $f : \mathbb{Z}_{\geq 0} \rightarrow \mathbb{Z}_{\geq 0}$ with $f(n) = \omega(\log n)$, no polynomial time algorithm can distinguish whether γCkC on the real line with $\gamma = f(|X|)$ admits a solution of radius 0.*

Hence, assuming the Exponential Time Hypothesis, γCkC is not approximable (with a polynomial-time algorithm) if the number of colors grows faster than logarithmic in the size of the ground set. Notice that, for a logarithmic number of colors, our procedures run in quasi-polynomial time.

1.2 Outline of Main Technical Contributions and Paper Organization

We introduce two main technical ingredients. The first is a method to deal with additional covering constraints in k -Center problems, which leads to Theorem 3. For this, we combine polyhedral sparsity-based arguments as used by Bandyapadhyay et al. [3], which by themselves only lead to pseudo-approximations, with dynamic programming to design a round-or-cut approach. Round-or-cut approaches, first used by Carr et al. [6], leverage the ellipsoid method in a clever way. In each ellipsoid iteration they either separate the current point from a well-defined polyhedron P , or round the current point to a good solution. The rounding step may happen even if the current point is not in P . Round-or-cut methods have found applications in numerous problem settings (see, e.g., [1, 8, 13, 19–22]). The way we employ round-or-cut is inspired by a powerful round-or-cut approach of Chakrabarty and Negahbani [8] also developed in the context of k -Center. However, their approach is not applicable to k -center problems as soon as multiple covering constraints exist, like in γCkC .

Our second technical contribution first employs LP duality to transform lottery-type models, like Fair γCkC , into an auxiliary problem that corresponds to a weighted version of k -center with covering constraints. We then show how a certain type of approximate separation over the dual is possible, by leveraging the techniques we introduced in the context of γCkC , leading to a 4-approximation.

Even though Theorem 4 is a strictly stronger statement than Theorem 3, we first prove Theorem 3 in Sect. 2, because it allows us to give a significantly cleaner presentation of some of our main technical contributions. In Sect. 3, we then focus on the additional techniques needed to deal with Fair γCkC , by reducing it to a problem that can be tackled with the techniques introduced in Sect. 2.

Due to space constraints, various proofs are deferred to the full version of the paper, including the proof of our hardness result, Theorem 5.

2 A 4-approximation for γCkC for $\gamma = O(1)$

In this section, we prove Theorem 3, which implies a 4-approximation algorithm for γCkC with constantly many colors. We assume $\gamma \geq 2$, since $\gamma = 1$ corresponds to Robust k -Center, for which an (optimal) 2-approximation is known [7, 15].

We present a procedure that for any $r \in \mathbb{R}_{\geq 0}$ returns a solution of radius $4r$ if a solution of radius r exists. This implies Theorem 3 because the optimal radius is a distance between two points. Hence, we can run the procedure for all possible pairwise distances r between points in X and return the best solution found. Hence, we fix $r \in \mathbb{R}_{\geq 0}$ in what follows. We denote by \mathcal{P} the following canonical relaxation of γCkC with radius r :

$$\mathcal{P} = \left\{ (x, y) \in [0, 1]^X \times [0, 1]^X \mid \begin{array}{l} \sum_{v \in X} y(v) \leq k \\ \sum_{v \in B(u, r)} y(v) \geq x(u) \quad \forall u \in X \\ \sum_{u \in X_\ell} x(u) \geq m_\ell \quad \forall \ell \in [\gamma] \end{array} \right\}. \quad (1)$$

Integral points $(x, y) \in \mathcal{P}$ correspond to solutions of radius r , where y indicates the opened centers and x indicates the points that are covered. We denote by $\mathcal{P}_I := \text{conv}(\mathcal{P} \cap (\{0, 1\}^X \times \{0, 1\}^X))$ the integer hull of \mathcal{P} .

Our algorithm is based on the round-or-cut framework, first used in [6]. The main building block is a procedure that rounds a point $(x, y) \in \mathcal{P}$ to a radius $4r$ solution under certain conditions. It will turn out that these conditions are always satisfied if $(x, y) \in \mathcal{P}_I$. If they are not satisfied, then we can prove that $(x, y) \notin \mathcal{P}_I$ and generate in polynomial time a hyperplane separating (x, y) from \mathcal{P}_I . This separation step now becomes an iteration of the ellipsoid method, employed to find a point in \mathcal{P}_I , and we continue with a new candidate point (x, y) . Schematically, the whole process is described in Fig. 1.

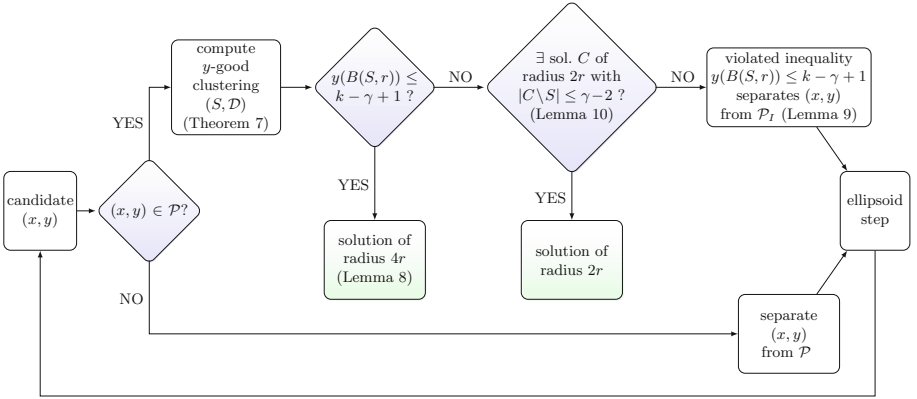


Fig. 1. An iteration of the ellipsoid method.

On a high level, we realize our round-or-cut procedure as follows. First, we check whether $(x, y) \in \mathcal{P}$ and return a violated constraint if this is not the case. If $(x, y) \in \mathcal{P}$, we partition the metric space, based on a natural greedy heuristic introduced by Harris et al. [15]. This gives a set of centers $S = \{s_1, \dots, s_q\}$ with corresponding clusters $\mathcal{D} = \{D_1, \dots, D_q\}$. We now exploit a technique by Bandyapadhyay et al. [3], which implies that if $y(B(S, r)) \leq k - \gamma + 1$, then one

can leverage sparsity arguments in a simplified LP to obtain a radius $4r$ solution that picks centers only within S . We then turn to the case where $y(B(S, r)) > k - \gamma + 1$. At this point, we show that one can efficiently check whether there exists a solution of radius $2r$ that opens at most $\gamma - 2$ centers outside of S . This is achieved by guessing $\gamma - 2$ centers and using dynamic programming to find the remaining $k - \gamma + 2$ centers in S . If no such radius $2r$ solution exists, we argue that any solution of radius r has at most $k - \gamma + 1$ centers in $B(S, r)$, proving that $y(B(S, r)) \leq k - \gamma + 1$ is an inequality separating (x, y) from \mathcal{P}_I .

We now give a formal treatment of each step of the algorithm described in Fig. 1. Given a point $(x, y) \in \mathbb{R}^X \times \mathbb{R}^X$, we first check whether $(x, y) \in \mathcal{P}$, and, if not, return a violated constraint of \mathcal{P} . Such a constraint separates (x, y) from \mathcal{P}_I because $\mathcal{P}_I \subseteq \mathcal{P}$. Hence, we may assume that $(x, y) \in \mathcal{P}$.

We now use a clustering technique by Harris et al. [15] that, given $(x, y) \in \mathcal{P}$, allows for obtaining what we call a y -good clustering (S, \mathcal{D}) , defined as follows.²

Definition 6 (y -good clustering). *Let $(x, y) \in \mathcal{P}$. A tuple (S, \mathcal{D}) , where the family $\mathcal{D} = \{D_1, \dots, D_q\}$ partitions X and $S = \{s_1, \dots, s_q\} \subseteq X$ with $s_i \in D_i$ for $i \in [q]$, is a y -good clustering if:*

- (i) $d(s_i, s_j) > 4r \quad \forall i, j \in [q], i \neq j$,
- (ii) $D_i \subseteq B(s_i, 4r) \quad \forall i \in [q]$, and
- (iii) $\sum_{i \in [q]} \min\{1, y(B(s_i, r))\} \cdot |D_i \cap X_\ell| \geq m_\ell \quad \forall \ell \in [\gamma]$.

The clustering procedure of [15] was originally introduced for Robust k -Center and naturally extends to γ CkC (see [3]). For completeness, we describe it in Algorithm 1. Contrary to prior procedures, we compute a y -good clustering whose centers have pairwise distances of strictly more than $4r$ (instead of $2r$ as in prior work). This large separation avoids overlap of radius $2r$ balls around centers in S , and allows us to use dynamic programming (DP) to build a radius $2r$ solution with centers in S under certain conditions. However, it is also the reason why get a 4-approximation if the DP approach cannot be applied.

Algorithm 1: Compute y -good clustering, given $(x, y) \in \mathcal{P}$

```

 $U \leftarrow X$ ;  $i \leftarrow 0$ ;  $S \leftarrow \emptyset$ ;  $\mathcal{D} \leftarrow \emptyset$ ;
while  $U \neq \emptyset$  do
  |  $i \leftarrow i + 1$ ;  $s_i \leftarrow \operatorname{argmax}_{u \in U} \{x(u)\}$ ;  $D_i \leftarrow U \cap B(s_i, 4r)$ ;
  |  $S \leftarrow S \cup \{s_i\}$ ;  $\mathcal{D} \leftarrow \mathcal{D} \cup \{D_i\}$ ;  $U \leftarrow U \setminus B(s_i, 4r)$ ;
end
return  $(S, \mathcal{D})$ 

```

Theorem 7 ([3, 15]). *For $(x, y) \in \mathcal{P}$, Algorithm 1 computes a y -good clustering (S, \mathcal{D}) in polynomial time.*

² As the name suggests, the properties of a y -good clustering do not depend on x . Hence, we could equivalently define the clustering for any $y \in \mathbb{R}^X$ that lies in the projection of \mathcal{P} onto the last $|X|$ coordinates, i.e., the ones corresponding to y .

Theorem 7 as well as the following lemma follow from the results in [3].

Lemma 8 ([3]). *Let $(x, y) \in \mathcal{P}$ and (S, \mathcal{D}) be a y -good clustering. Then, if $y(B(S, r)) \leq k - \gamma + 1$, a solution of radius $4r$ can be found in polynomial time.*

We are left with the case $y(B(S, r)) > k - \gamma + 1$. If $(x, y) \in \mathcal{P}_I$, then there must exist a solution $C_1 \subseteq X$ of radius r with $|C_1 \cap B(S, r)| > k - \gamma + 1$. Hence, C_1 has at most $\gamma - 2$ centers outside of $B(S, r)$. We observe that if such solution C_1 exists, then there must be a solution C_2 of radius $2r$ with all centers being within S , except for $\gamma - 2$ many. This is formalized in the following lemma (which states the contrapositive of the mentioned implication because this form is slightly more convenient later).

Lemma 9. *Let $S \subseteq X$ with $d(s, s') > 4r$ for all $s \neq s' \in S$, $\beta \in \mathbb{Z}_{\geq 0}$. If no radius $2r$ solution $C_2 \subseteq X$ satisfies $|C_2 \setminus S| \leq \beta$, then $|C_1 \cap B(S, r)| \leq k - \beta - 1$ for any radius r solution C_1 .*

Proof. Assume there is a solution C_1 of radius r where $|C_1 \cap B(S, r)| \geq k - \beta$. Let $A = C_1 \cap B(S, r)$. For each $p \in A$, let $\phi(p) \in S$ be the unique point in S such that $p \in B(\phi(p), r)$; $\phi(p)$ is well defined because $d(s, s') > 4r$ for every $s \neq s' \in S$. Let $C_2 = \phi(A) \cup (C_1 \setminus A)$. Then $|C_2| \leq |\phi(A)| + |C_1 \setminus A| \leq |A| + |C_1 \setminus A| \leq k$. Moreover, as $d(p, \phi(p)) \leq r$ for every $p \in A$, we conclude that $B(C_1, r) \subseteq B(C_2, 2r)$. Thus, C_2 is a feasible solution of radius $2r$. \square

Hence, if $y(B(S, r)) > k - \gamma + 1$ and $(x, y) \in \mathcal{P}_I$, then there is a solution C_2 of radius $2r$ with $|C_2 \setminus S| \leq \gamma - 2$. The motivation for considering solutions of radius $2r$ with all centers in S except for constantly many (if $\gamma = O(1)$) is that such solutions can be found efficiently via dynamic programming. This is possible because the centers in S are separated by distances strictly larger than $4r$, which implies that radius $2r$ balls centered at points in S do not overlap. Hence, there are no interactions between such balls. This is formalized below.

Lemma 10. *Let $S \subseteq X$ with $d(s, s') > 4r$ for all $s, s' \in S$ with $s \neq s'$, and $\beta \in \mathbb{Z}_{\geq 0}$. If a radius $2r$ solution $C \subseteq X$ with $|C \setminus S| \leq \beta$ exists, then we can find such a solution in time $|X|^{O(\beta + \gamma)}$.*

Proof. Suppose there is a solution $C \subseteq X$ of radius $2r$ with $|C \setminus S| \leq \beta$. The algorithm has two components. We first guess the set $Q := C \setminus S$. Because $|Q| \leq \beta$, there are $|X|^{O(\beta)}$ choices. Given Q , it remains to select at most $k - |Q|$ centers $W \subseteq S$ to fulfill the color requirements. Note that for any $W \subseteq S$, the number of points of color $\ell \in [\gamma]$ that $B(W, 2r)$ covers on top of those already covered by $B(Q, 2r)$ is $|(B(W, 2r) \setminus B(Q, 2r)) \cap X_\ell| = \sum_{w \in W} |(B(w, 2r) \setminus B(Q, 2r)) \cap X_\ell|$, where equality holds because centers in W are separated by distances strictly larger than $4r$, and thus $B(W, 2r)$ is the disjoint union of the sets $B(w, 2r)$ for $w \in W$. Hence, the task of finding a set $W \subseteq S$ with $|W| \leq k - |Q|$ such that $Q \cup W$ is a solution of radius $2r$ can be phrased as finding a feasible solution to the following binary program:

$$\sum_{s \in S} z(s) \cdot |(B(s, 2r) \setminus B(Q, 2r)) \cap X_\ell| \geq m_\ell - |B(Q, 2r) \cap X_\ell| \quad \forall \ell \in [\gamma]$$

$$\sum_{s \in S} z(s) \leq k - |Q| \tag{2}$$

$$z \in \{0, 1\}^S.$$

As this is a binary linear program with $\gamma + 1$ constraints with coefficients in $\{0, \dots, |X|\}$, it can be solved by standard dynamic programming techniques in $|X|^{O(\gamma)}$ time.³ As the dynamic program is run for $|X|^{O(\beta)}$ many guesses of Q , we obtain an overall running time of $|X|^{O(\beta+\gamma)}$, as claimed. \square

This completes the last ingredient for an iteration of our round-or-cut approach as shown in Fig. 1. In summary, assuming $y(B(S, r)) > k - \gamma + 1$ (for otherwise Lemma 8 leads to a solution of radius $4r$) we use Lemma 10 to check whether there is a radius $2r$ solution C_2 with $|C_2 \setminus S| \leq \gamma - 2$. If this is the case, we are done. If not, Lemma 9 implies that every radius r solution C_1 fulfills $|C_1 \cap B(S, r)| \leq k - \gamma + 1$. Hence, every point $(\bar{x}, \bar{y}) \in \mathcal{P}_I$ satisfies $\bar{y}(B(S, r)) \leq k - \gamma + 1$. However, this constraint is violated by (x, y) , and so it separates (x, y) from \mathcal{P}_I . Thus, we proved that the process described in Fig. 1 is a valid round-or-cut procedure that runs in polynomial time.

Corollary 11. *There is a polynomial-time algorithm that, given a point $(x, y) \in \mathbb{R}^X \times \mathbb{R}^X$, either returns a γ CkC solution of radius $4r$ or an inequality separating (x, y) from \mathcal{P}_I .*

We can now prove the main theorem.

Proof (of Theorem 3). We run the ellipsoid method on \mathcal{P}_I for each of the $O(|X|^2)$ candidate radii r . For each r , the number of ellipsoid iterations is polynomially bounded as the separating hyperplanes have encoding length at most $O(|X|)$ (see Theorem 6.4.9 of [14]). To see this, note that all generated hyperplanes are either inequalities defining \mathcal{P} or inequalities of the form $y(B(S, r)) \leq k - \gamma + 1$. For the correct guess of r , \mathcal{P}_I is non-empty and the algorithm terminates by returning a radius $4r$ solution. Hence, if we return the best solution among those computed for all guesses of r , we have a 4-approximation. \square

3 The Lottery Model of Harris et al. [15]

Let (X, d) be a Fair γ CkC instance, and let $\mathcal{F}(r)$ be the family of sets of centers satisfying the covering requirements with radius r , i.e.,

$$\mathcal{F}(r) := \{C \subseteq X \mid |C| \leq k \text{ and } |B(C, r) \cap X_\ell| \geq m_\ell \quad \forall \ell \in [\gamma]\}.$$

Note that a radius r solution for Fair γ CkC defines a distribution over the sets in $\mathcal{F}(r)$. Given r , such a distribution exists if and only if the following (exponential-size) linear program $\text{PLP}(r)$ is feasible (with $\text{DLP}(r)$ being its dual):

³ For any ordering $S = \{s_1, \dots, s_q\}$ of the elements in S , the DP successively computes for any prefix $\{s_1, \dots, s_i\}$ all possible left-hand side values for the $\gamma + 1$ constraints that can be achieved if $\text{supp}(z) := \{s \in S : z(s) > 0\} \subseteq \{s_1, \dots, s_i\}$. This is a straightforward extension of classical DPs for binary knapsack problems.

$$\begin{array}{ll}
\text{PLP}(r) : \min 0 & \text{DLP}(r) : \max \sum_{u \in X} p(u)\alpha(u) - \mu \\
\sum_{\substack{C \in \mathcal{F}(r): \\ u \in B(C,r)}} \lambda(C) \geq p(u) \quad \forall u \in X & \sum_{u \in B(C,r)} \alpha(u) \leq \mu \quad \forall C \in \mathcal{F}(r) \\
\sum_{C \in \mathcal{F}(r)} \lambda(C) = 1 & \alpha \in \mathbb{R}_{\geq 0}^X \\
\lambda \in \mathbb{R}_{\geq 0}^{\mathcal{F}(r)} & \mu \in \mathbb{R} .
\end{array}$$

Clearly, if $\text{PLP}(r)$ is feasible, then its optimal value is 0. Observe that $\text{DLP}(r)$ always has a feasible solution (the zero vector) of value 0. Thus, by strong duality, $\text{PLP}(r)$ is feasible if and only if the optimal value of $\text{DLP}(r)$ is 0. We note now that $\text{DLP}(r)$ is scale-invariant, meaning that if (α, μ) is feasible for $\text{DLP}(r)$ then so is $(t\alpha, t\mu)$ for $t \in \mathbb{R}_{\geq 0}$. Hence, $\text{DLP}(r)$ has a solution of strictly positive objective value if and only if $\text{DLP}(r)$ is unbounded. We thus define the following polyhedron $\mathcal{Q}(r)$, which contains all solutions of $\text{DLP}(r)$ of value at least 1:

$$\mathcal{Q}(r) := \left\{ (\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R} \mid \sum_{u \in X} p(u)\alpha(u) \geq \mu + 1, \sum_{u \in B(C,r)} \alpha(u) \leq \mu \quad \forall C \in \mathcal{F}(r) \right\}.$$

As discussed, we have the following.

Lemma 12. *$\mathcal{Q}(r)$ is empty if and only if $\text{PLP}(r)$ is feasible.*

The main lemma that allows us to obtain our result is the following.

Lemma 13. *There is a polynomial-time algorithm that, given a point $(\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R}$ satisfying $\sum_{u \in X} p(u)\alpha(u) \geq \mu + 1$ and a radius $r \geq 0$, either certifies that $(\alpha, \mu) \in \mathcal{Q}(r)$, or outputs a set $C \in \mathcal{F}(4r)$ with $\sum_{u \in B(C,4r)} \alpha(u) > \mu$.*

In words, Lemma 13 either certifies $(\alpha, \mu) \in \mathcal{Q}(r)$ or returns a hyperplane separating (α, μ) from $\mathcal{Q}(4r)$. Its proof leverages techniques introduced in Sect. 2, and we sketch it in Appendix A. Using Lemma 13, we can now prove Theorem 4.

Proof (of Theorem 4). As noted, there are polynomially many choices for the radius r , for each of which we run the ellipsoid method to check emptiness of $\mathcal{Q}(4r)$ as follows. Whenever there is a call to the separation oracle for a point $(\alpha, \mu) \in \mathbb{R}^X \times \mathbb{R}$, we first check whether $\alpha \geq 0$ and $\sum_{u \in X} p(u)\alpha(u) \geq \mu + 1$. If one of these constraints is violated, we return it as separating hyperplane. Otherwise, we invoke the algorithm of Lemma 13. The algorithm either returns a constraint in the inequality description of $\mathcal{Q}(4r)$ violated by (α, μ) , which solves the separation problem, or certifies $(\alpha, \mu) \in \mathcal{Q}(r)$. If, at any iteration of the ellipsoid method, the separation oracle is called for a point (α, μ) for which Lemma 13 certifies $(\alpha, \mu) \in \mathcal{Q}(r)$, then Lemma 12 implies $\text{PLP}(r)$ is infeasible. Thus, there is no solution to the considered Fair γCkC instance of radius r . Hence, consider from now on that the separation oracle always returns a separating hyperplane, in which case the ellipsoid method certifies that $\mathcal{Q}(4r) = \emptyset$ as

follows. Let $\mathcal{H} \subseteq \mathcal{F}(4r)$ be the family of all sets $C \in \mathcal{F}(4r)$ returned by Lemma 13 through calls to the separation oracle. Then, the following polyhedron:

$$\mathcal{Q}_{\mathcal{H}}(4r) = \left\{ (\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R} \mid \sum_{u \in X} p(u)\alpha(u) \geq \mu + 1, \sum_{u \in B(C, 4r)} \alpha(u) \leq \mu \forall C \in \mathcal{H} \right\},$$

containing $\mathcal{Q}(4r)$, is empty. As the encoding length of any constraint in the inequality description of $\mathcal{Q}(4r)$ is polynomially bounded in the input, the ellipsoid method runs in polynomial time (see Theorem 6.4.9 of [14]). In particular, the number of calls to the separation oracle, and thus $|\mathcal{H}|$, is polynomially bounded.

As $\mathcal{Q}(4r) \subseteq \mathcal{Q}_{\mathcal{H}}(4r) = \emptyset$, Lemma 12 implies that $PLP(4r)$ is feasible. More precisely, because $\mathcal{Q}_{\mathcal{H}}(4r) = \emptyset$, the linear program obtained from $DLP(4r)$ by replacing $\mathcal{F}(4r)$, which parameterizes the constraints in $DLP(4r)$, by \mathcal{H} , has optimal value equal to zero. Hence, its dual, which corresponds to $PLP(4r)$ where we replace $\mathcal{F}(4r)$ by \mathcal{H} is feasible. As this feasible linear program has polynomial size, because $|\mathcal{H}|$ is polynomially bounded, we can solve it efficiently to obtain a distribution with the desired properties. \square

A Appendix

In this appendix we discuss the proof of Lemma 13. Due to space constraints, we only sketch the proofs of some auxiliary lemmas; formal proofs are deferred to the full version.

The desired separation algorithm requires us to find a solution for a γCkC instance with an extra covering constraint; the procedure of Sect. 2 generalizes to handle this extra constraint. We follow similar steps as in Fig. 1.

Let $(\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R}$ be a point satisfying $\sum_{u \in X} p(u)\alpha(u) \geq \mu + 1$, let $r \geq 0$, and, moreover, let $\mathcal{F}^{\alpha, \mu}(r) := \{C \in \mathcal{F}(r) \mid \sum_{u \in B(C, r)} \alpha(u) > \mu\}$. We have to decide whether $\mathcal{F}^{\alpha, \mu}(4r) = \emptyset$ and, if not, find a set $C \in \mathcal{F}^{\alpha, \mu}(4r)$. We claim that there is a polynomially encoded $\varepsilon > 0$, such that this is equivalent to finding $C \in \mathcal{F}(4r)$ with $\sum_{u \in B(C, 4r)} \alpha(u) \geq \mu + \varepsilon$, or deciding that no such C exists. The next standard result guarantees that such an $\varepsilon > 0$ can be computed efficiently.

Lemma 14. *Let $(\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R}$. Then one can efficiently compute an $\varepsilon > 0$ with encoding length $O(L)$, where L is the encoding length of (α, μ) , such that the following holds: For any $C \in \mathcal{F}(r)$, we have $\sum_{u \in B(C, r)} \alpha(u) > \mu$ if and only if $\sum_{u \in B(C, r)} \alpha(u) \geq \mu + \varepsilon$.*

Proof. The tuple (α, μ) consists of $|X| + 1$ rationals $\{p_i/q_i\}_{i \in [N]}$, with $p_i \in \mathbb{Z}$ and $q_i \in \mathbb{Z}_{>0}$. Let $\Pi = \prod_{i \in [N]} q_i$. Note that if $\sum_{u \in B(C, r)} \alpha(u) > \mu$, then $\sum_{u \in B(C, r)} \alpha(u) - \mu \geq \frac{1}{\Pi}$. Thus, we set $\varepsilon = 1/\Pi$. Moreover $\log \Pi = \sum_{i \in [N]} \log q_i$, and so the encoding length of ε is $O(L)$. \square

Let $\mathcal{P}^{\alpha,\mu}$ be the following modified relaxation of γCkC , defined for given $(\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R}$, where the polytope \mathcal{P} is defined for a fixed radius r , as in Sect. 2 (see (1)):

$$\mathcal{P}^{\alpha,\mu} := \left\{ (x, y) \in \mathcal{P} \mid \sum_{u \in X} \alpha(u)x(u) \geq \mu + \varepsilon \right\}.$$

Let $\mathcal{P}_I^{\alpha,\mu} := \text{conv}(\mathcal{P}^{\alpha,\mu} \cap (\{0, 1\}^X \times \{0, 1\}^X))$ be the integer hull of $\mathcal{P}^{\alpha,\mu}$. We now state the following straightforward lemma, whose proof is an immediate consequence of the definitions of the corresponding polytopes and Lemma 14.

Lemma 15. *Let $(\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R}$ be such that $\sum_{u \in X} p(u)\alpha(u) \geq \mu + 1$ and $\mathcal{P}_I^{\alpha,\mu} = \emptyset$. Then $(\alpha, \mu) \in \mathcal{Q}(r)$.*

We now state Lemma 16, a modified version of Lemma 8. Its proof is analogous to the one of Lemma 8 and is deferred to the full version. The only difference is that the auxiliary polytope on which we exploit sparsity has one additional constraint, the one corresponding to $\sum_{u \in X} \alpha(u)x(u) \geq \mu + \varepsilon$, which explains the shift by one unit in the condition on $y(B(S, r))$.

Lemma 16. *Let $(\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R}$, let $(x, y) \in \mathcal{P}^{\alpha,\mu}$, and let (S, \mathcal{D}) be a y -good clustering. If $y(B(S, r)) \leq k - \gamma$, a set $C \in \mathcal{F}^{\alpha,\mu}(4r)$ can be found in polynomial time.*

If $y(B(S, r)) \leq k - \gamma$, then Lemma 16 leads to a set $C \in \mathcal{F}(4r)$ that satisfies $\sum_{u \in B(S, 4r)} \alpha(u) \geq \mu + \varepsilon$; this gives a constraint separating (α, μ) from $\mathcal{Q}(4r)$.

It remains to consider the case $y(B(S, r)) > k - \gamma$. As in Sect. 2, we can either find a set $C_2 \in \mathcal{F}^{\alpha,\mu}(2r)$ or certify that every $C_1 \in \mathcal{F}^{\alpha,\mu}(r)$ satisfies $|C_1 \cap B(S, r)| \leq k - \gamma$.

Lemma 17. *Let $(\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R}$, $S \subseteq X$ with $d(s, s') > 4r$ for all $s, s' \in S$ with $s \neq s'$, and $\beta \in \mathbb{Z}_{\geq 0}$. If no $C_2 \in \mathcal{F}^{\alpha,\mu}(2r)$ satisfies $|C_2 \setminus S| \leq \beta$, then $|C_1 \cap B(S, r)| \leq k - \beta - 1$ for any $C_1 \in \mathcal{F}^{\alpha,\mu}(r)$.*

Proof (Sketch). As in Lemma 9, we prove that any $C_1 \in \mathcal{F}^{\alpha,\mu}(r)$ satisfying $|C_1 \cap B(S, r)| > k - \beta - 1$ can be transformed into a set $C_2 \in \mathcal{F}^{\alpha,\mu}(2r)$ with $|C_2 \setminus S| \leq \beta$. \square

Lemma 18. *Let $(\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R}$, $S \subseteq X$ with $d(s, s') > 4r$ for all $s, s' \in S$ with $s \neq s'$, and $\beta \in \mathbb{Z}_{\geq 0}$. If there exists a set $C \in \mathcal{F}^{\alpha,\mu}(2r)$ with $|C \setminus S| \leq \beta$, then we can find such a set in time $|X|^{O(\beta+\gamma)}$.*

Proof (Sketch). As in the proof of Lemma 10, we first guess up to β centers $Q = X \setminus S$. For each of those guesses, we then consider the binary program (2) with objective function $\sum_{s \in S} z(s) \cdot \alpha(B(s, 2r))$ to be maximized. For the guess $Q = C \setminus S$, the characteristic vector $\chi^{C \cap S}$ is feasible for this binary program, implying that the optimal centers $Z \subseteq S$ chosen by the binary program fulfill $Z \cup Q \in \mathcal{F}^{\alpha,\mu}(2r)$. \square

Corollary 19. *Let $(\alpha, \mu) \in \mathbb{R}_{\geq 0}^X \times \mathbb{R}$. There is a polynomial-time algorithm that, given $(x, y) \in \mathbb{R}^X \times \mathbb{R}^X$, either returns a set $C \in \mathcal{F}^{\alpha, \mu}(4r)$ or returns a hyperplane separating (x, y) from $\mathcal{P}_I^{\alpha, \mu}$.*

Proof. If $(x, y) \notin \mathcal{P}^{\alpha, \mu}$, we return a violated constraint separating (x, y) from $\mathcal{P}_I^{\alpha, \mu}$. Hence we may assume $(x, y) \in \mathcal{P}^{\alpha, \mu}$. Since $\mathcal{P}^{\alpha, \mu} \subseteq \mathcal{P}$, we can use Theorem 7 to get a y -good clustering (S, \mathcal{D}) . If $y(B(S, r)) \leq k - \gamma$, Lemma 16 gives a set $C \in \mathcal{F}^{\alpha, \mu}(4r)$. So, assuming $y(B(S, r)) > k - \gamma$, we use Lemma 18 to check whether there is $C_2 \in \mathcal{F}^{\alpha, \mu}(2r)$ with $|C_2 \setminus S| \leq \gamma - 1$. If this is the case, we are done. If not, Lemma 17 implies that every $C_1 \in \mathcal{F}^{\alpha, \mu}(r)$ fulfills $|C_1 \cap B(S, r)| \leq k - \gamma$. Hence, every point $(\bar{x}, \bar{y}) \in \mathcal{P}_I^{\alpha, \mu}$ satisfies $\bar{y}(B(S, r)) \leq k - \gamma$. However, this constraint is violated by (x, y) , and it thus separates (x, y) from $\mathcal{P}_I^{\alpha, \mu}$. \square

Proof (Proof of Lemma 13). We use the ellipsoid method to check emptiness of $\mathcal{P}_I^{\alpha, \mu}$. Whenever the separation oracle gets called for a point $(x, y) \in \mathbb{R}^X \times \mathbb{R}^X$, we invoke the algorithm of Corollary 19. If the algorithm returns at any point a set $C \in \mathcal{F}^{\alpha, \mu}(4r)$, then C corresponds to a constraint in the inequality description of $\mathcal{Q}(4r)$ violated by (α, μ) . Otherwise, the ellipsoid method certifies that $\mathcal{P}_I^{\alpha, \mu} = \emptyset$, which implies $(\alpha, \mu) \in \mathcal{Q}(r)$ by Lemma 15. Note that the number of iterations of the ellipsoid method is polynomial as the hyperplanes used by the procedure above have encoding length $O(\text{poly}(|X|))$ (see Theorem 6.4.9 of [14]). \square

References

1. An, H.C., Singh, M., Svensson, O.: LP-based algorithms for capacitated facility location. *SIAM J. Comput.* **46**(1), 272–306 (2017)
2. Backurs, A., Indyk, P., Onak, K., Schieber, B., Vakilian, A., Wagner, T.: Scalable fair clustering. In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pp. 405–413 (2019)
3. Bandyapadhyay, S., Inamdar, T., Pai, S., Varadarajan, K.R.: A constant approximation for colorful k -center. In: *Proceedings of the 27th Annual European Symposium on Algorithms (ESA)*, pp. 12:1–12:14 (2019)
4. Bera, S.K., Chakrabarty, D., Flores, N., Negahbani, M.: Fair algorithms for clustering. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS)*, pp. 4955–4966 (2019)
5. Bercea, I.O., et al.: On the cost of essentially fair clusterings. In: *Proceedings of the 22nd International Conference on Approximation Algorithms for Combinatorial Optimization Problems (APPROX/RANDOM)*, pp. 18:1–18:22 (2019)
6. Carr, R.D., Fleischer, L.K., Leung, V.J., Phillips, C.A.: Strengthening integrality gaps for capacitated network design and covering problems. In: *Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 106–115 (2000)
7. Chakrabarty, D., Goyal, P., Krishnaswamy, R.: The non-uniform k -center problem. In: *Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming (ICALP)*, pp. 67:1–67:15 (2016)
8. Chakrabarty, D., Negahbani, M.: Generalized center problems with outliers. *ACM Trans. Algorithms* **15**(3), 1–14 (2019)

9. Charikar, M., Khuller, S., Mount, D.M., Narasimhan, G.: Algorithms for facility location problems with outliers. In: Proceedings of the 12th Annual Symposium on Discrete Algorithms (SODA), pp. 642–651 (2001)
10. Chen, D.Z., Li, J., Liang, H., Wang, H.: Matroid and knapsack center problems. *Algorithmica* **75**(1), 27–52 (2016)
11. Chierichetti, F., Kumar, R., Lattanzi, S., Vassilvitskii, S.: Fair clustering through fairlets. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), pp. 5029–5037 (2017)
12. Gonzalez, T.F.: Clustering to minimize the maximum intercluster distance. *Theor. Comput. Sci.* **38**, 293–306 (1985)
13. Grandoni, F., Kalaitzis, C., Zenklusen, R.: Improved approximation for tree augmentation: saving by rewiring. In: Proceedings of the 50th ACM Symposium on Theory of Computing (STOC), pp. 632–645 (2018)
14. Grötschel, M., Lovász, L., Schrijver, A.: *Geometric Algorithms and Combinatorial Optimization*, vol. 2. Springer, Heidelberg (2012). <https://doi.org/10.1007/978-3-642-78240-4>
15. Harris, D.G., Pensyl, T., Srinivasan, A., Trinh, K.: A lottery model for center-type problems with outliers. *ACM Trans. Algorithms* **15**(3), 36:1–36:25 (2019)
16. Hochbaum, D.S., Shmoys, D.B.: A best possible heuristic for the k -center problem. *Math. Oper. Res.* **10**(2), 180–184 (1985)
17. Hochbaum, D.S., Shmoys, D.B.: A unified approach to approximation algorithms for bottleneck problems. *J. ACM* **33**(3), 533–550 (1986)
18. Hsu, W., Nemhauser, G.L.: Easy and hard bottleneck location problems. *Discrete Appl. Math.* **1**(3), 209–215 (1979)
19. Levi, R., Lodi, A., Sviridenko, M.: Approximation algorithms for the capacitated multi-item lot-sizing problem via flow-cover inequalities. *Math. Oper. Res.* **33**(2), 461–474 (2008)
20. Li, S.: Approximating capacitated k -median with $(1 + \epsilon)k$ open facilities. In: Proceedings of the 27th Annual ACM Symposium on Discrete Algorithms (SODA), pp. 786–796 (2016)
21. Li, S.: On uniform capacitated k -median beyond the natural LP relaxation. *ACM Trans. Algorithms* **13**(2), 22:1–22:18 (2017)
22. Nutov, Z.: On the tree augmentation problem. In: Proceedings of the 25th Annual Symposium on Algorithms (ESA), pp. 61:1–61:14 (2017)
23. Rösner, C., Schmidt, M.: Privacy preserving clustering with constraints. In: Proceedings of the 45th International Colloquium on Automata, Languages, and Programming (ICALP), pp. 96:1–96:14 (2018)