Privacy Preserving Clustering with Constraints

Clemens Rösner
Department of Theoretical Computer Science, University of Bonn, Germany
roesner@cs.uni-bonn.de

Melanie Schmidt
Department of Theoretical Computer Science, University of Bonn, Germany
melanieschmidt@uni-bonn.de

Abstract

The k-center problem is a classical combinatorial optimization problem which asks to find k centers such that the maximum distance of any input point in a set P to its assigned center is minimized. The problem allows for elegant 2-approximations. However, the situation becomes significantly more difficult when constraints are added to the problem. We raise the question whether general methods can be derived to turn an approximation algorithm for a clustering problem with some constraints into an approximation algorithm that respects one constraint more. Our constraint of choice is privacy: Here, we are asked to only open a center when at least ℓ clients will be assigned to it. We show how to combine privacy with several other constraints.

2012 ACM Subject Classification Theory of computation → Facility location and clustering

Keywords and phrases Clustering, k-center, Constraints, Privacy, Lower Bounds, Fairness

Digital Object Identifier 10.4230/LIPIcs.ICALP.2018.96

Related Version A full version of the paper is available at [33], `https://arxiv.org/abs/1802.02497`.

1 Introduction

Clustering is a fundamental unsupervised learning task: Given a set of objects, partition them into clusters, such that objects in the same cluster are well matched, while different clusters have something that clearly differentiates them. The three classical clustering objectives studied in combinatorial optimization are *k-center*, *k-median* and *facility location*. Given a point set P, k-center and k-median ask for a set of k centers and an assignment of the points in P to the selected centers that minimize an objective. For k-center, the objective is the maximum distance of any point to its assigned center. For k-median, it is the sum of the distances of all points to their assigned center (this is called connection cost). Facility location does not restrict the number of centers. Instead, every center (here called facility) has an opening cost. The goal is to find a set of centers such that the connection cost plus the opening cost of all chosen facilities is minimized. In the unconstrained versions each point will be assigned to its closest center. With the addition of constraints a different assignment is often necessary in order to satisfy the constraints.

A lot of research has been devoted to developing approximation algorithms for these three. The earliest success story is that of k-center: Gonzalez [21] as well as Hochbaum and Shmoys [23] gave a 2-approximation algorithm for the problem, while Hsu and Nemhauser [24] showed that finding a better approximation is NP-hard.

Since then, much effort has been made to approximate the other two objectives. Typically, facility location will be first, and transferring new techniques to k-median poses additional challenges. Significant techniques developed during the cause of many decades are LP rounding

45th International Colloquium on Automata, Languages, and Programming (ICALP 2018).
Editors: Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella;
Article No. 96; pp. 96:1–96:14
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

techniques [11, 34], greedy and primal dual methods [25, 26], local search algorithms [6, 29], and, more recently, the use of pseudo-approximation [32]. The currently best approximation ratio for facility location is 1.488 [31], while the best lower bound is 1.463 [22]. For k-median, the currently best approximation algorithm achieves a ratio of $2.675+\epsilon$ [9], while the best lower bound is $1 + \frac{2}{e} \approx 1.736$ [25].

While the basic approximability of the objectives is well studied, a lot less is known once constraints are added to the picture. Constraints come naturally with many applications of clustering, and since machine learning and unsupervised learning methods become more and more popular, there is an increasing interest in this research topic. It is one of the troubles with approximation algorithms that they are often less easy to adapt to a different scenario than some easy heuristic for the problem, which was easier to understand and implement in the first place. Indeed, it turns out that adding constraints to clustering often requires fundamentally different techniques for the design of approximation algorithms and is a very new challenge altogether.

A good example for this is the *capacity* constraint: Each center c is now equipped with a capacity $u(c)$, and can only serve $u(c)$ points. This natural constraint is notoriously difficult to cope with; indeed, the standard LP formulations for the problems have an unbounded integrality gap. Capacitated k-center was first approximated with uniform upper bounds [8, 28]. Local search provides a way out for facility location, leading to 3- and 5-approximations for uniform [1] and non-uniform capacities [7], and preprocessing together with involved rounding proved sufficient for k-center to obtain a 9-approximation [15, 5]. However, the choice of techniques that turned out to work for capacitated clustering problems is still very limited, and indeed *no* constant factor approximation is known to date for k-median.

And all the while, new constraints for clustering problems are proposed and studied. In *private* clustering [3], we demand a lower bound on the number of points assigned to a center. As stated in [2, 3] this ensures a certain anonymity and is motivated through the need to obtain data privacy. The more general form where each cluster has an individual lower bound is called clustering *with lower bounds* [4]. *Fair* clustering [14] assumes that points have a protected feature (like gender), modeled by a color, and that we want clusters to be fair in the sense that the ratios between points of different colors is the same for every cluster. Clustering *with outliers* [12] assumes that our data contains measurement errors and searches for a solution where a prespecified number of points may be excluded from the cost computation. Other constraints include fault tolerance [27], matroid or knapsack constraints [13], must-link and cannot-link constraints [35], diversity [30] and chromatic clustering constraints [18, 19].

The abundance of constraints and the difficulty to adjust methods for all of them individually asks for ways to *add* a constraint to an approximation algorithm in an oblivious way. Instead of adjusting and reproving known algorithms, we would much rather like to take an algorithm as a black box and ensure that the solution satisfies one more constraint in addition. This is a challenging request. We start the investigation of such add-on algorithms by studying the privacy constraint. Indeed, we develop a method to add the privacy constraint to approximation algorithms for constraint k-center problems. That means that we use an approximation algorithm as a subroutine and ensure that the final solution will additionally respect a given lower bound. The method has to be adjusted depending on the constraint, but it is oblivious to the underlying approximation algorithm used for that constraint.

This works for the basic k-center problem (giving an algorithm for the private k-center problem), but we also show how to use the method when the underlying approximation algorithm is for k-center with outliers, fair k-center and capacitated k-center. We also demonstrate that our method suffices to approximate *strongly private* k-center, where we assume a protected feature like in fair clustering, but instead of fairness, now demand that a minimum number of points of each color is assigned to each open center to ensure anonymity for each class individually.

Our Technique. The general structure of the algorithm is based on standard thresholding [23], i.e., the algorithm tests all possible thresholds and chooses the smallest for which it finds a feasible solution. For each threshold, it starts with the underlying algorithm and computes a non private solution. Then it builds a suitable network to shift points to satisfy the lower bounds. The approximation ratio of the method depends on the underlying algorithm and on the structure of this network.

The shifting does not necessarily work right away. If it does not produce a feasible solution, then using the max flow min cut theorem, we obtain a set of points for which we can show that the clustering uses too many clusters (and can thus not satisfy the lower bounds). The algorithm then recomputes the solution in this part. Depending on the objective function, we have to overcome different hurdles to ensure that the recomputation works in the sense that it a) makes sufficient progress towards finding a feasible solution and b) does not increase the approximation factor. The process is then iterated until we find a feasible solution.

Results. We obtain the following results for multiple combinations of privacy with other constraints. Note that our definition of k-center (see §2) distinguishes between the set of points P and the set of possible center locations L. This general case is also called the *k-supplier problem*, while classical k-center often assumes that $P = L$. Our reductions can handle the general case (with a slight increase in approximation ratio); whether the resulting algorithm is then for k-center or k-supplier thus depends on the evoked underlying algorithm.

- We obtain a 4-approximation for private k-center with outliers (5 for the supplier version). This matches the best known bounds [3] ([4] for the supplier version (this also holds for non-uniform lower bounds)).
- We compute an 11-approximation for private capacitated k-center (i.e., centers have a lower bound *and* an upper bound), and a 8-approximation for private uniform capacitated k-center (where the upper bounds are uniform, as well). The best known bounds for these two problems are 9 and 6 [17]. For the supplier version we obtain a 13-approximation which matches the best known bound [17] (for uniform upper bounds a 9-approximation-algorithm is known [17]).
- We achieve constant factor approximations for private capacitated/uncapacitated and fair k-center/k-supplier clustering. The approximation factor depends on the *balance* of the input point set and the type of upper bounds, it ranges between 10 in the uncapacitated case where for each color c the number of points with color c is an integer multiple of the number of points with the rarest color and 325 in the general supplier version with non-uniform upper bounds. To the best of our knowledge, all these combinations have not been studied before.
- Along the way, we propose constant factor algorithms for general cases of fair clustering. While [14] introduces a pretty general model of fairness, it only derives approximation algorithms for inputs with two colors and a balance of $1/t$ for an integer t. We achieve ratios of 14 and 15 for the general fair k-center and supplier problem, respectively.
- Finally, we propose the *strongly private k-center problem.* As in the fair clustering problem, the input here has a protected feature like gender, modeled by colors. Now instead of a fair clustering, we aim for anonymity for each color, meaning that we have a lower bound for each color. Each open center needs to be assigned this minimum number

Table 1 An overview on the approximation results that we combine with privacy.

	Vanilla	Capacities		Outlier	Fair Subset Partition	
		uniform	non-uniform		$\frac{r}{b} \in \mathbb{N}$	general
k-center	2 [23]	6 [28]	9 [5]	2 [10]	2 [14]	12 (full version Thm.22 [33])
k-supplier	3 [23]		11 [5]	3 [12]		

of points for each color. To the best of our knowledge, this problem has not been studied before; we obtain a 4-approximation as well as a 5-approximation for the supplier version.

Since our method does not require knowledge of the underlying approximation algorithm, the approximation guarantees improve if better approximation algorithms for the underlying problems are found. There is also hope that our method could be used for new, not yet studied constraints, with not too much adjustment.

Related Work. The unconstrained k-center problem can be 2-approximated [21, 23], and it is NP-hard to approximate it better [24]. The k-supplier problem can be 3-approximated [23], and this is also tight.

Capacitated k-center was first approximated with uniform upper bounds [8, 28]. Two decades after the first algorithms for the uniform case, [15] provided the first constant factor approximation for non-uniform capacities. The algorithm was improved and applied to the k-supplier problem in [5]. In contrast to capacities, *lower* bounds are less studied. The private k-center problem is introduced and 2-approximated in [3], and non-uniform lower bounds are studied in [4]. The k-center/k-supplier problem with outliers is 3-approximated in [12] alongside approximations to other robust variants of the k-center problem. The approximation factor for the k-center problem with outliers was improved to 2 in [10].

The fair k-center problem was introduced in [14]. The paper describes how to approximate the problem by using an approximation for a subproblem that we call fair subset partition problem. Algorithms for this subproblem are derived for two special cases where the number of colors is two, and the points are either perfectly balanced or the number of points of one color is an integer multiple of the number of points of the other color.

These are the constraints for which we make use of known results. We state the best known bounds and their references in Table 1. Approximation algorithms are also e.g. known for fault tolerant k-center [27] and k-center with matroid or knapsack constraints [13].

Relatively little is known about the combination of constraints. Cygan and Kociumaka [16] give a 25-approximation for the capacitated k-center problem with outliers. Aggarwal et. al [3] give a 4-approximation for the private k-center problem with outliers. Ahmadian and Swamy [4] consider the combination of k-supplier with outliers with (non-uniform) lower bounds and derive a 5-approximation. The paper also studies the k-supplier problem with outliers (without lower bounds), and the min-sum-of-radii problem with lower bounds and outliers. Their algorithms are based on the Lagrangian multiplier preserving primal dual method due to Jain and Vazirani [26].

Ding et. al [17] study the combination of capacities and lower bounds as well as capacities, lower bounds and outliers by generalizing the LP algorithms from [5] and [16] to handle lower bounds. They give results for several variations, including a 6-approximation for private capacitated k-center and a 9-approximation for private capacitated k-supplier.

Friggstad, Rezapour, Salavatipour [20] consider the combination of uniform capacities and non-uniform lower bounds for *facility location* and obtain bicriteria approximations.

Outline. In §2, we introduce necessary notation. §3 then presents our method, applied to the private k-center problem with outliers. We choose the outlier version since it is non-trivial but still intuitive and does thus give a good impression on the application of our method. In §4, we then briefly explain how to adjust the method to approximate private and fair k-center, private and capacitated k-center, k-center with all three constraints as well as the strongly private k-center problem. §5 provides a conclusion.

2 Preliminaries

Let (X, d) be a finite metric space, i.e., X is a finite set and $d : X \times X \to \mathbb{R}_{\geq 0}$ is a metric. We use $d(x, T) = \min_{y \in T} d(x, y)$ for the smallest distance between $x \in X$ and a set $T \subseteq X$. For two sets $S, T \subseteq X$, we use $d(S, T) = \min_{x \in S, y \in T} d(x, y)$ for the smallest distance between any pair $x \in S, y \in T$.

Let $P \subseteq X$ be a subset of X called *points* and let $L \subseteq X$ be a subset of X called *locations*. An instance of a private *assignment constrained* k-center problem consists of P, L, an integer $k \in \mathbb{N}$, a lower bound $\ell \in \mathbb{N}$ and possibly more parameters. Given the input, the problem is to compute a set of centers $C \subseteq L$ with $|C| \leq k$ and an *assignment* $\phi : P \to C$ of the points to the selected centers that satisfies $\ell \leq |\phi^{-1}(c)|$ for every selected center $c \in C$, and some specific *assignment restriction*. The solution C, ϕ shall be chosen such that

$$\max_{x \in P} d(x, \phi(x))$$

is minimized. Different assignment restrictions lead to different constrained private k-center problems. The *capacity* assignment restriction comes with an upper bound function $u : L \to \mathbb{N}$ for which we require $\ell \leq u(x)$ for all $x \in L$, and then demands $|\phi^{-1}(c)| \leq u(c)$. When we have $u(x) = u$ for all $x \in L$ and some $u \in \mathbb{N}$, then we say that the capacities are *uniform*, otherwise, we say they are *non-uniform*. The *fairness* assignment restriction provides a mapping $\chi : P \to Col$ of points to colors and then requires that each cluster has the same ratio between the numbers of points with different colors (see §4.2 in the full version [33] for specifics). The *strongly private k-center problem* can also be cast as a k-center problem with an assignment restriction. Again, the input now additionally contains a mapping χ of points to colors. Now the assignment is restricted to ensure that it satisfies the lower bound for the points of each color. We even consider the slight generalization where each color has its own lower bound, and call this problem the strongly private k-center problem.

An instance of the *private k-center problem with outliers* consists of P, L, an integer $k \in \mathbb{N}$, a lower bound ℓ, and a parameter o for the maximum number of outliers. The problem is to compute a set of centers $C \subseteq L$ with $|C| \leq k$ and an assignment $\phi : P \to C \cup \{out\}$, with $|\phi^{-1}(out)| \leq o$, which assigns each point to a center in C or to be an outlier. The choice of C, ϕ shall minimize

$$\max_{x \in P \setminus \phi^{-1}(out)} d(x, \phi(x)).$$

3 Private k-center with Outliers

▶ **Theorem 1.** *Assume that there exists an approximation algorithm A for the k-center problem with outliers with approximation factor α.*

Then for instances P, L, k, ℓ, o of the private k-center problem with outliers, we can compute an $(\alpha + 2)$-approximation in polynomial time.

Proof. Below, we describe an algorithm that uses a threshold graph with threshold τ. We show that for any given $\tau \in \mathbb{R}$, the algorithm has polynomial runtime and, if τ is equal to opt, the value of the optimal solution, computes an $(\alpha + 2)$-approximation. Since we know that the value of every solution is equal to the distance between a point and a location, we test all $O(|P||L|)$ possible distances for τ and return the best feasible clustering returned by any of them. The main proof is the proof of Lemma 2 below, which concludes this proof. ◀

Let us start with a general idea on how the algorithm works for a fixed τ. If $\tau < \mathsf{opt}$, then the algorithm will detect this, so assume that $\tau \geq \mathsf{opt}$ for now. We first use the approximation algorithm to obtain a clustering which does not necessarily satisfy the privacy constraint. Then we try to reassign points to establish the lower bounds. A point p may be reassigned to any cluster which contains at least one point which is within distance 2τ of p. (Note that any point in p's optimum cluster is at distance $\leq 2\tau$). If it is possible to reassign points like that and obtain a feasible clustering, we can compute such a reassignment with a maximum flow algorithm. Otherwise we find a set of points $P(V')$ for which we can show that the optimal clustering uses less clusters to cover all of them. We add all outliers to $P(V')$ to account for the fact that $P(V')$ may already contain outliers of the optimum solution. We then use the approximation algorithm on $P(V')$ to compute a new clustering with outliers. The output will contain fewer clusters or the same number of clusters and fewer outliers. We repeat the process until we find a feasible solution.

▶ **Lemma 2.** *Assume that there exists an approximation algorithm A for the k-center problem with outliers with approximation factor α. Let P, L, k, ℓ, o be an instance of the private k-center problem with outliers, let $\tau > 0$ and let opt denote the maximum radius in the optimal feasible clustering for P, L, k, ℓ, o. We can in polynomial time compute a feasible clustering with a maximum radius of at most $(\alpha + 2)\tau$ or determine $\tau < \mathsf{opt}$.*

Proof. We first use A to compute a solution without the lower bound. Let $\mathcal{C} = (C, \phi)$ be an α-approximate solution for the k-center problem with outliers on P, L, k, o. Note that $\mathcal{C}$ may contain clusters with less than ℓ points. Let $k' = |C|$ (note that $k' < k$ is possible), $C = \{c_1, \ldots, c_{k'}\}$, and let $C_1, \ldots, C_{k'}$ be the clusters that $\mathcal{C}$ induces, i.e., $C_j := \phi^{-1}(c_j)$. Finally, let $r = \max_{x \in P} d(x, \phi(x))$ be the largest distance of any point to its center. Observe that an optimal solution to the k-center problem with outliers can only have a lower objective value than the optimal solution to our problem because we only dropped a condition. Therefore, $\tau \geq \mathsf{opt}$ implies that $r \leq \alpha \cdot \mathsf{opt} \leq \alpha \cdot \tau$. If we have $r > \alpha \cdot \tau$, we return $\tau < \mathsf{opt}$.

We use $\mathcal{C}$ and τ to create a threshold graph which we use to either reassign points between the clusters to obtain a feasible solution or to find a set of points P' for which we can show that every feasible clustering with maximum radius τ uses less clusters than our current solution to cover it. In the latter case we compute another α-approximate solution which uses fewer clusters on P' and repeat the process. Note that for $\tau < \mathsf{opt}$ such a clustering does not necessarily exist, but for $\tau \geq \mathsf{opt}$ the optimal clustering provides a solution for P' with fewer clusters. If we do not find such a clustering with maximum radius at most $\alpha \cdot \tau$, we return $\tau < \mathsf{opt}$.

We show that every iteration of the process reduces the number of clusters or the number of outliers, therefore the process stops after at most $k \cdot o$ iterations. It may happen that our final solution contains much less clusters than the optimal solution (but it will be an approximate solution for the optimal solution with k centers).

We will use a network flow computation to move points from clusters with more than ℓ points to clusters with less than ℓ points. Moving a point to another cluster can increase the radius of the cluster. We only want to move points between clusters such that the radius does not increase by too much. More precisely, we only allow a point p to be moved to another cluster C_i if the distance $d(p, C_i)$ between the point and the clusters is at most 2τ. This is ensured by the structure of the network described in the next paragraph. Unless stated otherwise, when we refer to distances between a point and a cluster in the following, we mean the distance between the point and the cluster in its original state before any points have been reassigned.

Given $\mathcal{C}$ and τ, we create the threshold graph $G_\tau = (V_\tau, E_\tau)$ as follows. V_τ consists of a source s, a sink t, a node v_i for each cluster C_i, a node v_{out} for the set of outliers and a node w_p for each point $p \in P$. For all $i \in [k']$, we connect s to v_i if the cluster C_i contains more than ℓ points and set the capacity of (s, v_i) to $|C_i| - \ell$. If the cluster C_i contains fewer than ℓ points, we connect v_i with t and set the capacity of (v_i, t) to $\ell - |C_i|$. Furthermore, we connect v_i with w_p for all $p \in C_i$ and set the capacity of (v_i, w_p) to 1. We also connect s to v_{out} with capacity o and v_{out} with w_p for all $p \in \phi^{-1}(out)$ with capacity 1. Whenever a point p and a cluster C_i with $p \notin C_i$ satisfy $d(p, C_i) \leq 2\tau$ (i.e., there is a point $q \in C_i$ that satisfies $d(p, q) \leq 2\tau$), we connect w_p with v_i with capacity 1.

Formally the graph $G_\tau = (V_\tau, E_\tau)$ is defined by

$$\begin{aligned} V_\tau =& \{v_{out}\} \cup \{v_i \mid 1 \leq i \leq k'\} \cup \{w_p \mid p \in P\} \cup \{s, t\} \text{ and} && (1)\\ E_\tau =& \{(v_i, w_p) \mid p \in C_i\} \cup \{(w_p, v_i) \mid p \notin C_i \wedge d(p, C_i) \leq 2\tau\} && (2)\\ & \cup \{(v_{out}, w_p) \mid \phi(p) = out\} && (3)\\ & \cup \{(s, v_{out})\} \cup \{(s, v_i) \mid |C_i| - \ell > 0\} \cup \{(v_i, t) \mid |C_i| - \ell < 0\}. && (4) \end{aligned}$$

We define the capacity function $cap : E_\tau \to \mathbb{R}$ by

$$cap(e) = \begin{cases} \ell - |C_i|, & \text{if } e = (v_i, t) \\ |C_i| - \ell, & \text{if } e = (s, v_i) \\ o, & \text{if } e = (s, v_{out}) \\ 1 & \text{otherwise.} \end{cases} \tag{5}$$

We use $G = (V, E)$ to refer to G_τ as τ is clear from context. We now compute an integral maximum s-t-flow f on G. According to f we can reassign points different clusters.

▶ **Lemma 3.** *Let f be an integral maximal s-t-flow on G. It is possible to reassign p to C_i for all edges (w_p, v_i) with $f((w_p, v_i)) = 1$.*

The resulting solution has a maximum radius of at most $r + 2\tau$. If f saturates all edges of the form (v_i, t), then the solution is feasible.

Proof. Let $p \in C_i$. The choice of capacity 1 on (v_i, w_p) and flow conservation ensure $\sum_{(w_p, v_j) \in E} f((w_p, v_j)) \leq 1$ for p. Therefore no point would have to be reassigned to more than one cluster. Note that for every point $p \in C_i$ that would be reassigned we must have $f((v_i, w_p)) = 1$ and for every edge (v_i, w_p) with $f((v_i, w_p)) = 1$ the point p would be reassigned.

For any $1 \leq j \leq k'$, let $p \in C_i$ be any point which we want to reassign to C_j. Then we must have $(w_p, v_j) \in E$ and therefore there must be a point $q \in C_j$ with $d(p, q) \leq 2\tau$. Thus we have

$$d(p, c_j) \leq d(p, q) + d(q, c_j) \leq 2\tau + r = r + 2\tau.$$

Now assume that f saturates all edges of the form (v_i, t) and let $1 \leq i \leq k'$. If E contains the edge (v_i, t), then it can not contain the edge (s, v_i) and therefore all incoming edges of v_i

are of the form (w_p, v_i). Flow conservation then implies that the number of points reassigned to C_i minus the points reassigned away from C_i is equal to $f((v_i, t))$, which increases the number of points in C_i to ℓ.

If E contains the edge (s, v_i), then it can not contain the edge (v_i, t) and therefore all outgoing edges of v_i are of the form (v_i, w_p). Flow conservation then implies that the number of points reassigned away from C_i minus the points reassigned to C_i is equal to $f((s, v_i))$, which reduces the number of points in C_i to at least ℓ.

If E contains neither (s, v_i) nor (v_i, t), then the number of points in C_i is equal to ℓ and does not change (the points may change, but their number does not).

In all three cases C_i contains at least ℓ points after the reassignment. ◀

If f saturates all edges of the form (v_i, t) in G, then we reassign points according to Lemma 3 and return the new clustering.

Otherwise we look at the residual network G_f of f on G. Let V' be the set of nodes in G_f which can not be reached from s. V' contains all nodes representing clusters which would not satisfy the privacy constraint after the reassignment. We say cluster C_i *belongs* to V' if $v_i \in V'$, and a point $p \in C_i$ is *adjacent* to V' if $w_p \in V'$ and $v_i \notin V'$. Let $C(V')$ denote the set of clusters belonging to V'. Let $k'' = |C(V')|$. We say a point p belongs to V' if the cluster C_i with $p \in C_i$ belongs to V'. Let $P(V')$ and $P_A(V')$ denote the set of points that belong to V' and the set of points adjacent to V'.

▶ **Lemma 4.** *Any clustering on P with maximum radius at most τ that contains at least ℓ points in every cluster uses fewer than k'' clusters to cover all points in $P(V')$.*

Proof. We first observe that V' must have the following properties:

- $v_i \in V'$ and $(w_p, v_i) \in E$ implies $w_p \in V'$.
- $w_p \in V'$, $(w_p, v_i) \in E$ and $f((w_p, v_i)) > 0$ implies $v_i \in V'$.
- $w_p \in V'$ for some $p \in C_i$ and $v_i \notin V'$ implies $f((v_i, w_p)) = 1$.

The first property follows from the fact that f can only saturate (w_p, v_i) if f also saturates (v_j, w_p) for $p \in C_j$. So, either (w_p, v_i) is not saturated, which means that v_i can be reached from any vertex that reaches w_p, or (w_p, v_i) *is* saturated, which means that the only incoming edge of w_p in G_f is (v_i, w_p). In both cases, if $v_i \in V'$, then $w_p \in V'$. The second property follows since $f((w_p, v_i)) > 0$ implies $(v_i, w_p) \in E(G_f)$. The third property is true since we defined $cap((v_i, w_p)) = 1$.

This implies that a reassignment due to Lemma 3 would reassign all points adjacent to V' to clusters in $C(V')$ and moreover all reassignments from points in $P(V') \cup P_A(V')$ would be to clusters in $C(V')$. Let n_i denote the number of points that would be assigned to C_i after the reassignment. Then $|P(V')| + |P_A(V')| = \sum_{C_i \in C(V')} n_i$.

Now we argue that this sum is smaller than $k'' \cdot \ell$ by observing that each $n_i \leq \ell$ and at least one n_i is strictly smaller than ℓ.

Let C_i be a cluster with more than ℓ points after the reassignment. Then (s, v_i) is not saturated by f and v_i can be reached from s in G_f. Therefore after the reassignment no cluster $C_i \in C(V')$ would contain more than ℓ points; in other words, $n_i > \ell$ implies $C_i \notin C(V')$.

Let C_i be a cluster which would still contain fewer than ℓ points after the reassignment. This implies that f does not saturate the edge (v_i, t). Therefore t can be reached from v_i and since f is a maximum s-t flow, v_i can not be reached from s. We must have $v_i \in V'$.

Because we assumed that the reassignment does not satisfy all lower bounds, at least one such cluster has to exist. This implies

$$|P(V')| + |P_A(V')| = \sum_{C_i \in C(V')} n_i < k'' \cdot \ell.$$

Which means that the clusters in $C(V')$ and $P_A(V')$ do not contain enough points to satisfy the lower bound in k'' clusters.

By definition of G and V', for two points p, q with $p \in P(V')$ and $d(p, q) \le 2\tau$ we must have $q \in P(V') \cup P_A(V')$. Let $\mathcal{C}'$ be a clustering that abides the lower bounds and has a maximal radius of at most τ. Then every cluster C' in $\mathcal{C}'$ that contains at least one point from $P(V')$ can only contain points from $P(V') \cup P_A(V')$. Therefore $\mathcal{C}'$ must contain fewer than k'' clusters which contain at least one point from $P(V')$. ◀

If we have $\tau \ge \mathsf{opt}$, then Lemma 4 implies that the optimal solution covers all points in $P(V')$ with fewer than k'' clusters. An α-approximative solution on the point set $P(V')$ with at most $k'' - 1$ clusters which contains at most o outliers is then α-approximative for $P(V')$.

Unfortunately, we do not know how many outliers an optimal clustering has in $P(V')$. We therefore involve the outliers $\phi^{-1}(out)$ in our new computation as well. Let $o' = |\phi^{-1}(out)|$ denote the current number of outliers. We obtain the following Lemma through a counting argument.

▶ **Lemma 5.** *We call a cluster special if it contains at least one point from $P(V')$ or only contains points from $\phi^{-1}(out)$. Let $\mathcal{C}'$ be a clustering on P with a maximum radius of at most τ on all special clusters that respects the lower bounds, has at most o outliers and consists of at most k clusters out of which at most k'' are special. If $\mathcal{C}'$ has exactly k'' special clusters, then $\mathcal{C}'$ has at most $o' - 1$ outliers in $P(V') \cup \phi^{-1}(out)$.*

Proof. Assume the clustering contains exactly special k'' clusters. Each of these clusters has to contain at least ℓ points from $P(V') \cup P_A(V') \cup \phi^{-1}(out)$. We know

$$|P(V') \cup P_A(V') \cup \phi^{-1}(out)| \le |P(V') \cup P_A(V')| + o' < k''\ell + o'.$$

So there remain at most $o' - 1$ unclustered points in $P(V') \cup \phi^{-1}(out)$. ◀

Now we need to show that such a clustering exists if $\tau \ge opt$ is the case.

▶ **Lemma 6.** *If $\tau \ge opt$, then there exists a clustering $\mathcal{C}'$ on P with a maximum radius at most τ on all special clusters that respects the lower bounds, has at most o outliers and consists of at most k clusters out of which at most k'' are special.*

Proof. We look at an optimal clustering $\mathcal{C}_{opt}$. The only way $\mathcal{C}_{opt}$ can violate a condition is if it contains $k''' > k''$ special clusters. Lemma 4 implies that $\mathcal{C}_{opt}$ contains at least $k''' - k''$ clusters that contain only points in $\phi^{-1}(out)$. If all clusters in $\mathcal{C}_{opt}$ are special we know $P = P_A(V') \cup P(V') \cup \phi^{-1}(out)$. We arbitrarily select $k''' - k''$ clusters from $\mathcal{C}_{opt}$ that contain only points in $\phi^{-1}(out)$, declaring all points in them as outliers and closing the corresponding centers. This leaves us with k'' clusters which contain at least $k'' \cdot \ell$ points. Since $P = P_A(V') \cup P(V') \cup \phi^{-1}(out)$ this leaves at most $o' - 1$ outliers. Otherwise, if $\mathcal{C}_{opt}$ contains at least one cluster C which is not special, we add all outliers from $P \setminus (P_A(V') \cup P(V') \cup \phi^{-1}(out)$ to C. Again we arbitrarily select $k''' - k''$ clusters from $\mathcal{C}_{opt}$ that contain only points in $\phi^{-1}(out)$, declaring all points in them as outliers and closing the corresponding centers. By creation there are no unclustered points in $P \setminus (P_A(V') \cup P(V') \cup \phi^{-1}(out)$ and exactly k'' special clusters with radius at most τ. Therefore this clustering contains at most $o' - 1$ outliers and has at most k clusters. ◀

We now use A again to compute new solutions without the lower bound: Let $\mathcal{C}_1' = (C_1', \phi_1')$ be an α-approximate solution for the k-center problem with outliers on $P(V') \cup \phi^{-1}(out)$, L, $k'' - 1$, o and let $\mathcal{C}_2' = (C_2', \phi_2')$ be an α-approximate solution for the k-center problem with outliers on $P(V') \cup \phi^{-1}(out)$, L, k'', $o' - 1$. Let $r_i' = \max_{x \in P(V') \cup \phi^{-1}(out)} d(x, \phi_i'(x))$.

Note that in case $\tau < \mathsf{opt}$, it can happen that no such clustering exists or that we obtain $r_i' > \alpha \cdot \tau$ for both $i = 1$ and $i = 2$. We then return $\tau < \mathsf{opt}$. Otherwise for at least one $i \in \{1, 2\}$ $\mathcal{C}_i'$ must exist together with $r_i' \leq \alpha \cdot \tau$.

If $\mathcal{C}_2'$ exists and we have $r_2' \leq \alpha \cdot \tau$ we replace $C(V')$ by C_2' in $\mathcal{C}$ and adjust ϕ accordingly to obtain $\mathcal{C}_1 = (C_1, \phi_1)$ with $C_1 = (C \setminus C(V')) \cup C_2'$ and

$$\phi_1(p) = \begin{cases} \phi_2'(p) & \text{if } p \in P(V') \cup \phi^{-1}(out) \\ \phi(p) & \text{otherwise.} \end{cases} \tag{6}$$

Otherwise, if $\mathcal{C}_1'$ exists, we have $r_1' \leq \alpha \cdot \tau$ and either $\mathcal{C}_2'$ does not exist or we have $r_2' > \alpha \cdot \tau$, we analogous replace $C(V')$ by C_1' to obtain $\mathcal{C}_1$.

▶ **Lemma 7.** *If we did not return $\tau < \mathsf{opt}$, then $\mathcal{C}_1$ is a solution for the k-center problem with outliers on P, L, k, o and we have $r_1 = \max_{x \in P} d(x, \phi_1(x)) \leq \alpha \cdot \tau$.*

Proof. $\mathcal{C}$ is a solution for the k-center problem with outlier on P, L, k, o with $r < \alpha \cdot \tau$ and since we did not return $\tau < \mathsf{opt}$, we must have $r_i' \leq \alpha\tau$ for the chosen $i \in \{1, 2\}$. ◀

We iterate the previous process with the new clustering $\mathcal{C}_1$ until we either determine $\tau < \mathsf{opt}$ or the reassignment of points according to Lemma 3 yields a feasible solution. Since each iteration reduces the number of clusters or keeps the same number of clusters and reduces the number of outliers, the process terminates after at most $k \cdot o$ iterations. ◀

▶ **Corollary 8.** *We can compute a 4-approximation for instances of the private k-center problem with outliers and a 5-approximation for instances of the private k-supplier problem in polynomial time.*

Proof. Follows from Theorem 1 together with the 2-approximation for k-center with outliers in [10] and the 3-approximation for k-supplier with outliers in [12]. ◀

4 Combining Privacy with other Constraints

In this section, we take the general idea from §3 and instead of outliers use it to combine privacy with other restrictions on the clusters. Given a specific restriction $\mathcal{R}$ and an approximation algorithm A for the k-center problem with restriction $\mathcal{R}$ with approximation factor α we ask: Can we similar to §3 extend A to compute an $O(\alpha)$-approximation for the private k-center problem with restriction $\mathcal{R}$?

In §3, we made use of two properties of a clustering with outliers. In Lemma 3 we used that reassigning points to another cluster never increases the number of outliers and in Lemma 4 we used that outliers have the somewhat local property that computing a new clustering on the points V' from a subset of the clusters together with the set of outliers can not create more outliers on the remaining points.

We use this section to briefly explain how the proofs with other restriction properties differ from §3 and state our results for the general cases (for the complete proofs as well as the approximation factors for all different cases, see §4+§5 in the full version [33]).

As new restrictions we study capacities, fairness and multiple lower bounds for different types of points. For the private capacitated k-center problem, the proof is easier than for private k-center with outliers. The threshold graph $G_\tau = (V_\tau, E_\tau)$ does not need to contain nodes for the outliers anymore. It is defined by

$$\begin{aligned} V_\tau =& \{v_i \mid 1 \leq i \leq k'\} \cup \{w_p \mid p \in P\} \cup \{s, t\} \text{ and} && (7)\\ E_\tau =& \{(v_i, w_p) \mid p \in C_i\} \cup \{(w_p, v_i) \mid p \notin C_i \wedge d(p, C_i) \leq 2\tau\} && (8)\\ & \cup \{(s, v_i) \mid |C_i| - \ell > 0\} \cup \{(v_i, t) \mid |C_i| - \ell < 0\}, && (9) \end{aligned}$$

together with the capacity function $cap : E_\tau \to \mathbb{R}$

$$cap(e) = \begin{cases} \ell - |C_i|, & \text{if } e = (v_i, t)\\ |C_i| - \ell, & \text{if } e = (s, v_i)\\ 1 & \text{otherwise.} \end{cases} \qquad (10)$$

The capacities on the clusters do not influence the threshold graph; they are handled by the underlying approximation algorithm. Since we never increase the number of points of a cluster to more than ℓ, our method will not introduce any capacity violation. We obtain the following results.

▶ **Theorem 9.** *Assume that there exists an approximation algorithm A for the capacitated k-center problem with approximation factor α. Then we can compute an $(\alpha + 2)$-approximation for the private capacitated k-center problem in polynomial time.*

In order to combine privacy with fairness, we first develop an approximation algorithm for general cases of the fair k-center problem. [14] show how to solve fair k-center based on approximating a subproblem that we call *fair subset partitioning problem* (it's called fairlet decomposition in [14]), and which consists of partitioning the input into fair subsets such that the maximum diameter of any subset is minimal. They give an exact algorithm for two colors with perfect balance and a 2-approximation for two colors with balance $1/t$ for an integer t. We propose an algorithm for multiple colors and general balance values. It separates the points into the different colors, computes a capacitated clustering on the points for one of the colors and uses a matching algorithm on threshold graphs in order to add the points of the other colors.

▶ **Theorem 10.** *A 12-approximation for the fair subset partition problem can be computed in polynomial time. If $b_c = 1$ for at least one color $c \in Col$, then a 2-approximation for the fair subset partition problem can be computed in polynomial time (even if $|Col| > 2$).*

With that we obtain the first approximation algorithm for the general fair k-center problem. For the private and fair k-center problem we then use a partitioning into fair subsets and let these be the nodes of the threshold graph instead of single points. When establishing the lower bounds, we move fair subsets as a whole. Let $F = \{F_1, \ldots\}$ denote the fair subsets. Then we define $G_\tau = (V_\tau, E_\tau)$ by

$$\begin{aligned} V_\tau =& \{v_{out}\} \cup \{v_i \mid 1 \leq i \leq k'\} \cup \{f_i \mid F_i \in F\} \cup \{s, t\} \text{ and} && (11)\\ E_\tau =& \{(v_i, f_j) \mid F_j \subseteq C_i\} \cup \{(f_j, v_i) \mid F_j \cap C_i = \emptyset \wedge d(C_i, F_j) \leq 2\tau\} && (12)\\ & \cup \{(s, v_i) \mid |C_i| - \ell > 0\} \cup \{(v_i, t) \mid |C_i| - \ell < 0\} && (13) \end{aligned}$$

with the capacity function $cap : E_\tau \to \mathbb{R}$

$$cap(e) = \begin{cases} \left\lceil \frac{\ell - |C_i|}{b} \right\rceil, & \text{if } e = (v_i, t)\\ \left\lfloor \frac{|C_i| - \ell}{b} \right\rfloor, & \text{if } e = (s, v_i)\\ 1 & \text{otherwise.} \end{cases} \qquad (14)$$

We obtain the following results.

▶ **Theorem 11.** *Assume that there exists an approximation algorithm A for the fair subset partition problem with approximation factor α. Then we can compute a $(3\alpha + 4)/(3\alpha + 5)$-approximation for the private fair k-center/supplier problem in polynomial time.*

We also adjust our method to approximate the private fair and capacitated k-center problem, and combine it with our results on approximating the fair subset partitioning problem. The approximation ratio now becomes $\alpha(2\beta + 1)$, where α is the approximation ratio for the capacitated k-center problem, and β is the approximation ratio for the fair subset partitioning problem (see Lemma 29 in the full version [33]).

For the strongly private k-center problem (see §5 in the full version [33]) we create a separate threshold graph for each color and separately try to satisfy the lower bound for each color. For a color $i \in Col$ we let P^i denote the points in P with color i and let C_j^i denote the points in C_j with color i. We create the threshold graph $G_{\tau,i} = (V_{\tau,i}, E_{\tau,i})$ by

$$\begin{aligned} V_{\tau,i} =& \{v_j \mid 1 \leq j \leq k'\} \cup \{w_p \mid p \in P^i\} \cup \{s, t\} \text{ and} && (15)\\ E_{\tau,i} =& \{(v_j, w_p) \mid p \in C_j^i\} \cup \{(w_p, v_j) \mid p \in P^i \setminus C_j \wedge d(p, C_j) \leq 2\tau\} && (16)\\ & \cup \{(s, v_j) \mid |C_j \cap \chi^{-1}(i)| - \ell_i > 0\} \cup \{(v_j, t) \mid |C_j \cap \chi^{-1}(i)| - \ell_i < 0\}. && (17) \end{aligned}$$

We define the capacity functions $cap_i : E_{\tau,i} \to \mathbb{R}$ by

$$cap(e) = \begin{cases} \ell_i - |C_j \cap \chi^{-1}(i)|, & \text{if } e = (v_j, t) \\ |C_j \cap \chi^{-1}(i)| - \ell_i, & \text{if } e = (s, v_j) \\ 1 & \text{otherwise.} \end{cases} \quad (18)$$

We obtain the following result.

▶ **Theorem 12.** *Assume that there exists an approximation algorithm A for the k-center problem with approximation factor α. Then we can compute an $(\alpha + 2)$-approximation for the strongly private k-center problem in polynomial time.*

The following corollary summarizes the remaining results in the full version [33] for the different constrained private k-center problems, and fair k-center clustering.

▶ **Corollary 13.** *There are polynomial approximation algorithms that compute*

- *an 11/13-approximation for the private capacitated k-center/supplier problem (Cor. 14 + 15),*
- *a 14/15-approximation for the fair k-center/supplier problem (Cor. 23),*
- *a 40/41-approximation for the private and fair k-center/supplier problem (Cor. 25),*
- *a 225/325-approximation for the private fair capacitated k-center/supplier problem (C. 30),*
- *improved approximations for the last three problems in special cases, and*
- *a 4/5-approximation for the strong private k-center/supplier problem (Cor. 36).*

5 Conclusion and open questions

We studied k-center with capacities, fairness and outliers and have coupled these constraints with privacy, and we proposed strongly private k-center. In addition to improving the approximation guarantee of the presented coupling process, open questions include extending it to arbitrary lower bounds, and to different objective functions. In Appendix A of the full version [33], we present a bicriteria result for facility location that violates the lower bounds. For k-median, a similar procedure is not known at all so far. Finally, the question how to obliviously add other constraints than privacy is completely open, too.

References

1 Ankit Aggarwal, Anand Louis, Manisha Bansal, Naveen Garg, Neelima Gupta, Shubham Gupta, and Surabhi Jain. A 3-approximation algorithm for the facility location problem with uniform capacities. *Mathematical Programming*, 141(1-2):527–547, 2013.

2 Gagan Aggarwal, Tomas Feder, Krishnaram Kenthapadi, Rajeev Motwani, Rina Panigrahy, Dilys Thomas, and An Zhu. Approximation algorithms for k-anonymity. In *Proceedings of the International Conference on Database Theory (ICDT 2005)*, November 2005. URL: `http://ilpubs.stanford.edu:8090/645/`.

3 Gagan Aggarwal, Rina Panigrahy, Tomás Feder, Dilys Thomas, Krishnaram Kenthapadi, Samir Khuller, and An Zhu. Achieving anonymity via clustering. *ACM Transaction on Algorithms*, 6(3):49:1–49:19, 2010.

4 Sara Ahmadian and Chaitanya Swamy. Approximation algorithms for clustering problems with lower bounds and outliers. In *43rd International Colloquium on Automata, Languages, and Programming, (ICALP)*, pages 69:1–69:15, 2016.

5 Hyung-Chan An, Aditya Bhaskara, Chandra Chekuri, Shalmoli Gupta, Vivek Madan, and Ola Svensson. Centrality of trees for capacitated k-center. *Mathematical Programming*, 154(1-2):29–53, 2015.

6 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.

7 Manisha Bansal, Naveen Garg, and Neelima Gupta. A 5-approximation for capacitated facility location. In *20th Annual European Symposium on Algorithms (ESA)*, pages 133–144, 2012.

8 Judit Bar-Ilan, Guy Kortsarz, and David Peleg. How to allocate network centers. *Journal of Algorithms*, 15(3):385–415, 1993.

9 Jaroslaw Byrka, Thomas Pensyl, Bartosz Rybicki, Aravind Srinivasan, and Khoa Trinh. An improved approximation for k-median and positive correlation in budgeted optimization. *ACM Transactions on Algorithms*, 13(2):23:1–23:31, 2017.

10 Deeparnab Chakrabarty, Prachi Goyal, and Ravishankar Krishnaswamy. The non-uniform k-center problem. In *Proceedings of the 43rd International Colloquium on Automata, Languages, and Programming (ICALP)*, volume 55 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 67, 15. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.

11 Moses Charikar, Sudipto Guha, Éva Tardos, and David B. Shmoys. A constant-factor approximation algorithm for the k-median problem. *Journal of Computer and System Sciences*, 65(1):129–149, 2002.

12 Moses Charikar, Samir Khuller, David M. Mount, and Giri Narasimhan. Algorithms for facility location problems with outliers. In *Proceedings of the 12th Annual Symposium on Discrete Algorithms (SODA)*, pages 642–651, 2001.

13 Danny Z. Chen, Jian Li, Hongyu Liang, and Haitao Wang. Matroid and knapsack center problems. *Algorithmica*, 75(1):27–52, 2016.

14 Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017 (NIPS)*, pages 5036–5044, 2017.

15 Marek Cygan, MohammadTaghi Hajiaghayi, and Samir Khuller. LP rounding for k-centers with non-uniform hard capacities. In *53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 273–282, 2012.

16 Marek Cygan and Tomasz Kociumaka. Constant factor approximation for capacitated k-center with outliers. In *Proceedings of the 31st International Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 251–262, 2014.

17 Hu Ding, Lunjia Hu, Lingxiao Huang, and Jian Li. Capacitated center problems with two-sided bounds and outliers. In *Proceedings of the 15th International Symposium on Algorithms and Data Structures (WADS)*, pages 325–336, 2017.

18 Hu Ding and Jinhui Xu. Solving the chromatic cone clustering problem via minimum spanning sphere. In *Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 773–784, 2011.

19 Hu Ding and Jinhui Xu. A unified framework for clustering constrained data without locality property. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1471–1490, 2015.

20 Zachary Friggstad, Mohsen Rezapour, and Mohammad R. Salavatipour. Approximating connected facility location with lower and upper bounds via LP rounding. In *15th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT)*, pages 1:1–1:14, 2016.

21 Teofilo F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science*, 38:293–306, 1985.

22 Sudipto Guha and Samir Khuller. Greedy strikes back: Improved facility location algorithms. *Journal of Algorithms*, 31(1):228–248, 1999.

23 Dorit S. Hochbaum and David B. Shmoys. A unified approach to approximation algorithms for bottleneck problems. *Journal of the ACM*, 33(3):533–550, 1986.

24 Wen-Lian Hsu and George L. Nemhauser. Easy and hard bottleneck location problems. *Discrete Applied Mathematics*, 1(3):209–215, 1979.

25 Kamal Jain, Mohammad Mahdian, and Amin Saberi. A new greedy approach for facility location problems. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, pages 731–740, 2002.

26 Kamal Jain and Vijay V. Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and lagrangian relaxation. *Journal of the ACM*, 48(2):274–296, 2001.

27 Samir Khuller, Robert Pless, and Yoram J. Sussmann. Fault tolerant k-center problems. *Theoretical Computer Science*, 242(1-2):237–245, 2000.

28 Samir Khuller and Yoram J. Sussmann. The capacitated K-center problem. *SIAM Journal on Discrete Mathematics*, 13(3):403–418, 2000.

29 Madhukar R. Korupolu, C. Greg Plaxton, and Rajmohan Rajaraman. Analysis of a local search heuristic for facility location problems. *Journal of Algorithms*, 37(1):146–188, 2000.

30 Jian Li, Ke Yi, and Qin Zhang. Clustering with diversity. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 188–200, 2010.

31 Shi Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. *Information and Computation*, 222:45–58, 2013.

32 Shi Li and Ola Svensson. Approximating k-median via pseudo-approximation. *SIAM Journal on Computing*, 45(2):530–547, 2016.

33 Clemens Rösner and Melanie Schmidt. Privacy preserving clustering with constraints. *CoRR*, abs/1802.02497, 2018. `arXiv:1802.02497`.

34 David B. Shmoys, Éva Tardos, and Karen Aardal. Approximation algorithms for facility location problems. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing (STOC)*, pages 265–274, 1997.

35 Kiri Wagstaff, Claire Cardie, Seth Rogers, and Stefan Schrödl. Constrained k-means clustering with background knowledge. In *Proceedings of the 18th International Conference on Machine Learning (ICML)*, pages 577–584, 2001.