Efficient and Stable Fully Dynamic Facility Location

Sayan Bhattacharya
Department of Computer Science
University of Warwick
Coventry, CV47AL, United Kingdom
s.bhattacharya@warwick.ac.uk

Silvio Lattanzi
Google Research
silviol@google.com

Nikos Parotsidis
Google Research
nikosp@google.com

Abstract

We consider the classic facility location problem in fully dynamic data streams, where elements can be both inserted and deleted. In this problem, one is interested in maintaining a stable and high quality solution throughout the data stream while using only little time per update (insertion or deletion). We study the problem and provide the first algorithm that at the same time maintains a constant approximation and incurs polylogarithmic amortized recourse per update. We complement our theoretical results with an experimental analysis showing the practical efficiency of our method.

1 Introduction

Clustering is a fundamental problem in unsupervised learning with many practical applications in community detection, spam detection, image segmentation and many others. A central problem in this space with a long and rich history in computer science and operations research Korte and Vygen [2018] is the facility location problem(which can also be seen as the Lagrangian relaxation of the classic k–median clustering). Due to its scientific importance and practical applications the problem has been studied extensively and several approximation algorithms are known for the problem Guha and Khuller [1999], Jain et al. [2003], Li [2013]. In addition, the problem has also been extensively studied in various computational model as the streaming model Indyk [2004], Lammersen and Sohler [2008], Czumaj et al. [2013], the online model Meyerson [2001], Fotakis [2008] and dynamic algorithm model Cygan et al. [2018], Goranci et al. [2018], Cohen-Addad et al. [2019], Guo et al. [2020] and many more.

Real world applications nowadays often process evolving data-sets, i.e. social networks continuously evolve in time, videos and pictures are constantly uploaded and taken down from media platforms, news article and blog posted are uploaded or taken down and so on so for. For this reason, it is important to design algorithms that are able to maintain a stable and high quality solution and that at the same time can process updates efficiently. As a consequence, the dynamic and online model have been extensively studied and several interesting results are known for classic learning problems Lattanzi and Vassilvitskii [2017], Chan et al. [2018], Cohen-Addad et al. [2019], Jaghargh et al. [2019], Lattanzi et al. [2020], Fichtenberger et al. [2021], Guo et al. [2021]. In this paper, we extend this line of work by studying the classic facility location problem in the fully dynamic setting.

Problem definition. The input to our problem consists of a collection F of *facilities* and a collection D of *clients* in a general metric space. For all $i \in F, j \in D$, we let d_{ij} denote the distance between facility i and client j[1]. These distances satisfy triangle inequality. Let f_i denote the *opening cost* of facility $i \in F$. Our goal is to open a subset of facilities $F' \subseteq F$ and then connect every client $j \in D$ to some open facility $i_j \in F'$, so as to minimise the objective $\sum_{i \in F'} f_i + \sum_{j \in D} d_{i_j,j}$. The first sum

[1]For simplicity, we assume that all the distances between facility i and client j are polynomially bounded.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

in the objective denotes the sum of the opening costs of the facilities in F', whereas the second sum denotes the sum of the connection costs for the clients. Throughout the rest of this paper, we will let $m = |F|$ and $n = |D|$ respectively denote the number of facilities and clients.

We consider a *dynamic setting*, where the input to the problem keeps changing via a sequence of updates. Each update inserts or deletes a client in the metric space. Throughout this sequence of updates, we wish to maintain an approximately optimal solution to the current input instance with small *recourse*. The recourse of the algorithm is defined as follows. Whenever we open a new facility or close an already open facility, we incur one unit of *facility-recourse*. Similarly, whenever we reassign a client from one facility to another, we incur one unit of *client-recourse*. The recourse of the algorithm after an update is the sum of facility-recourse and client-recourse. Note that the recourse of an algorithm naturally captures the consistency of the solution throughout the input updates. So having small recourse is particularly important for real world application where a clustering is served to the user or used in a machine learning pipeline Lattanzi and Vassilvitskii [2017]. Our main result in this paper is summarized in the theorem below.

Theorem 1. *There is a deterministic $O(1)$-approximation algorithm for fully dynamic facility location problem that has $O(\log m)$ amortized recourse, under the natural assumption that the distances between facilities and clients and the facility opening-costs are bounded by some polynomial in m.*

Our Technique. To the best of our knowledge, this is the first algorithm for fully dynamic facility location that can handle non-uniform facility costs and maintains a $O(1)$ approximation with polylogarithmic recourse per update. To obtain our result, we consider a *relaxed* version of the natural greedy algorithm in the static setting. This relaxed greedy algorithm proceeds in iterations. In each iteration the algorithm picks a *cluster*, which specifies a set of currently unassigned clients and the facility they will get assigned to, that has approximately minimum average marginal cost. The algorithm stops when every client gets assigned to some facility. In the dynamic setting, we ensure that the solution maintained by our algorithm always corresponds to an output of the relaxed greedy algorithm (assuming the relaxed-greedy algorithm receives the current input instance). To be a bit more precise, we formulate a set of invariants which capture the workings of the relaxed greedy algorithm, and fix the invariants using a simple heuristic whenever one or more of them get violated in the dynamic setting. The approximation guarantee now follows from the observation that the relaxed greedy algorithm returns a $O(1)$ approximation in the static setting. The recourse bound, on the other hand, follows from a careful token-based argument that is inspired by the analysis of a fully dynamic greedy algorithm for minimum set cover Gupta et al. [2017]. In addition, using standard data structures it is easy to show that the amortized *update time* of our algorithm is $\tilde{O}(m)$. We note that this update time is near-optimum, as each arriving client needs to reveal its distances to the facilities, which also holds for the uniform facility opening cost as pointed out in Cohen-Addad et al. [2019][2].

We also complement our theoretical analysis with an in-depth study of the performance of our algorithm showing that our algorithm is very efficient and effective in practice.

Related Works. Facility location and dynamic algorithms have been extensively studied in the algorithm and machine learning literature. Here we give an overview of closely related results.

Facility Location. The classic offline metric facility location problem has been extensively studied (see Korte and Vygen [2018]). The best-known approximation for the problem algorithm gives a 1.488 approximation Li [2013]. The problem is known to be NP-hard, and it cannot be approximated within a factor of 1.463 unless **NP** $\subseteq$ **DTIME**$(n^{\log\log(n)})$ Guha and Khuller [1999].

Dynamic algorithm for clustering and facility location. Recently several dynamic algorithms have been introduced for clustering and facility location problems Lattanzi and Vassilvitskii [2017], Chan et al. [2018], Cygan et al. [2018], Goranci et al. [2018], Cohen-Addad et al. [2019], Henzinger et al. [2020], Guo et al. [2020], Goranci et al. [2021], Fichtenberger et al. [2021]. The recourse for clustering problems was studied in Lattanzi and Vassilvitskii [2017], then several papers followed up with interesting results on recourse and dynamic algorithms Chan et al. [2018], Cohen-Addad et al. [2019], Henzinger et al. [2020], Fichtenberger et al. [2021].

Efficient algorithm for the fully-dynamic facility location problem are known either when the doubling dimension of the problem is bounded Goranci et al. [2018, 2021] or when the facility have uniform

[2] In some works, some assumptions on the form of the input are make, which allow avoiding an $\Omega(m)$ factor, see e.g. Guo et al. [2020] where each arriving client reports its nearest facility uppon arrival.

cost Cygan et al. [2018], Cohen-Addad et al. [2019]. When the facilities can have arbitrary non-uniform costs in a general metric space, prior to our work the best known fully dynamic algorithm achieved a $O(\log |F|)$ approximation with polylogarithmic recourse Guo et al. [2020].

2 Our Dynamic Algorithm

In Section 2.1, we present an intuitive overview of our algorithm. We formally describe our algorithm in Section 2.2 and Section 2.3, and analyze its recourse and approximation guarantee in Section 2.4.

2.1 An Informal Overview of Our Algorithm

We start by recalling a natural greedy algorithm for the facility location problem in the *static setting*. This (static) algorithm is known to return a constant approximation for the problem Jain et al. [2003]. The algorithm proceeds in *rounds*. Throughout the duration of the algorithm, let $F^* \subseteq F$ and $D^* \subseteq D$ respectively denoted the set of opened facilities and the set of clients that are already assigned to some open facility. Initially, before the start of the first round, we set $F^* \leftarrow \emptyset$ and $D^* \leftarrow \emptyset$. The next paragraph describes what happens during a given round of this algorithm.

inline (??) Say that a *cluster* C is an ordered pair (i, A), where $i \in F$ and $A \subseteq D$. A cluster is either *satellite* or *critical*. If a cluster $C = (i, A)$ is satellite (resp. critical), then its *cost* is given by $cost(C) := \sum_{j \in A} d_{ij}$ (resp. $cost(C) := f_i + \sum_{j \in A} d_{ij}$). Thus, the only difference between an satellite cluster and a critical cluster is that the cost of the former (resp. latter) does *not* (resp. does) include the opening cost of the concerned facility. The *average cost* of a cluster $C = (i, A)$ is defined as $cost_{avg}(C) := cost(C)/|A|$. Let $\mathcal{C}_{ord}(F^*, D^*)$ be the set of all satellite clusters $C = (i, A)$ with $i \in F^*$ and $A \subseteq D \setminus D^*$, and let $\mathcal{C}_{crit}(F^*, D^*)$ be the set of all critical clusters $C = (i, A)$ with $i \in F \setminus F^*$ and $A \subseteq D \setminus D^*$. In the current round, we pick a cluster $C = (i, A) \in \mathcal{C}_{crit}(F^*, D^*) \cup \mathcal{C}_{ord}(F^*, D^*)$ with minimum *average cost*, set $F^* \leftarrow F^* \cup \{i\}$ and $D^* \leftarrow D^* \cup A$, and assign all the clients $j \in A$ to the facility i. At this point, if $D^* = D$, then we terminate the algorithm. Otherwise, we proceed to the next round.

Intuitively, in each round the above greedy algorithm picks a cluster with minimum average *cost*, which is defined in such a way that takes into account the opening cost of a facility the first time some client gets assigned to it. Let $\mathcal{C}$ denote the collection of all clusters picked by this algorithm, across all the rounds. Note: (1) Every client $j \in A$ belongs to exactly one cluster in $\mathcal{C}$. (2) Let $\mathcal{C}(i) \subseteq \mathcal{C}$ denote the subset of clusters in $\mathcal{C}$ which contain a given facility $i \in F$. If $\mathcal{C}(i) \neq \emptyset$, then exactly one of the clusters in $\mathcal{C}(i)$ is critical. We refer to a collection $\mathcal{C}$ which satisfy these two properties as a *clustering*. It is easy to check that a clustering $\mathcal{C}$ defines a valid solution to the concerned instance of the facility location problem, where the objective value of the solution is given by $obj(\mathcal{C}) := \sum_{C \in \mathcal{C}} cost(C)$.

Note that if the above algorithm picks an *satellite* cluster $C = (i, A)$ in a given round, then w.l.o.g. we can assume that $|A| = 1$. This is because if $|A| > 1$, then we can only decrease the average cost of C by deleting from A all but the one client that is closest to facility i. Accordingly, from now on we will only consider those satellite clusters that have exactly one client.

Our dynamic algorithm is based on a *relaxation* of the above static greedy algorithm, where in each round we have the flexibility of picking a cluster with *approximately* minimum average cost. The output of this relaxed greedy algorithm corresponds to what we call a *nice clustering* (see Section 2.2 for a formal definition). We now present a high-level, informal overview of our dynamic algorithm.

Preprocessing: Upon receiving an input (F, D) at preprocessing, we run the relaxed greedy algorithm which returns a nice clustering $\mathcal{C}$. We assign each cluster $C \in \mathcal{C}$ to an (not necessarily positive) integer *level* $\ell(C) \in \mathbb{Z}$ such that $cost_{avg}(C) = \Theta(2^{\ell(C)})$. Thus, the level of a cluster encodes its average cost upto a constant multiplicative factor. Define the level of a client $j \in C$ to be $\ell(j) := \ell(C)$.

Handling an update: When a client j gets deleted from D, we simply delete the concerned client from the cluster in $\mathcal{C}$ it appears in. Similarly, when a client j gets inserted into D, we arbitrarily assign it to any open facility i by creating an satellite cluster $(C = (i, \{j\})$ at the minimum possible level $k \geq \Theta(\log d_{ij})$. At this point, we check if the clustering $\mathcal{C}$ being maintained by our algorithm still remains *nice*, i.e., whether $\mathcal{C}$ can correspond to an output of the relaxed greedy algorithm on the current input. If the answer is yes, then our dynamic algorithm is done with handling the current update. Thus, for the rest of this discussion, assume that the clustering $\mathcal{C}$ is no longer *nice*.

It turns out that in this event we can always identify a cluster $C = (i, A)$ (not necessarily part of $\mathcal{C}$) and a level $k \in \mathbb{Z}$ such that: $cost_{avg}(C) \leq O(2^k)$ and $\ell(j) > k$ for all clients $j \in A$. We refer to such a cluster $C = (i, A)$ as a *blocking cluster*. Intuitively, the existence of a blocking cluster is a certificate that the current clustering $\mathcal{C}$ is not nice. This is because the relaxed greedy algorithm constructs the levels in a bottom-up manner, in increasing order of average costs of the clusters that get added to the solution in successive rounds. Accordingly, the relaxed greedy algorithm would have added the cluster C to its solution at level $\leq k$ before proceeding to level $k + 1$, and this contradicts the fact that a client $j \in A$ appears at level $> k$ in the current clustering $\mathcal{C}$.

As long as there is a blocking cluster $C = (i, A)$ at some level k, we update the current clustering $\mathcal{C}$ by calling a subroutine FIX-BLOCKING(C, k) which works as follows: (1) it adds the cluster C to level k, and (2) it removes each client $j \in A$ from the cluster in $\mathcal{C}$ it belonged to prior to this step.

Next, note that because of step (2) above, some cluster $C' = (i', A')$ might lose one or more clients j as they move from C' to the newly formed cluster C, and this might increase the average cost of the cluster C'. Thus, we might potentially end up in a situation where a cluster $C' = (i', A')$ has $cost_{avg}(C') \gg 2^{\ell(C')}$. As long as this happens to be the case, we call a subroutine FIX-LEVEL(C) whose job is to increase the level of the affected cluster C' by one.

Our algorithm repeatedly calls the subroutines FIX-BLOCKING$(.,.)$ and FIX-LEVEL$(.)$ until there is no blocking cluster and every cluster $C \in \mathcal{C}$ has $cost_{avg}(C) = \Theta(2^{\ell(C)})$. At this point, we are guaranteed that the current clustering is nice and we are done with processing the concerned update.

The approximation ratio of our algorithm follows from the observation that a nice clustering corresponds to the output of the relaxed greedy algorithm in the static setting, which in turn gives a $O(1)$-approximation. We bound the amortized recourse by a careful token-based argument.

2.2 Nice clustering

In Section 2.1, we explained that a *nice clustering* corresponds to an output of the relaxed greedy algorithm in the static setting. We now give a formal definition of this concept by introducing four invariants. Specifically, we define a clustering $\mathcal{C}$ to be nice iff it satisfies Invariants 1, 2, 3 and 4.

Every cluster $C \in \mathcal{C}$ is assigned a (not necessarily positive) integer *level* $\ell(C) \in \mathbb{Z}$. For any cluster $C = (i, A) \in \mathcal{C}$ and any client $j \in A$, we define the level of the client j to be $\ell(j) := \ell(C)$.

Invariant 1. *For every cluster $C \in \mathcal{C}$, we have $cost_{avg}(C) < 2^{\ell(C)}$.*

The relaxed greedy algorithm constructs these levels in a "bottom-up" manner: it assigns the relevant clusters to a level k before moving on to level $k + 1$. This holds because the algorithm picks the clusters in (approximately) increasing order of their average costs. Also, before picking an satellite cluster $C = (i, \{j\})$, the algorithm ensures that facility i is open. This leads to the invariant below.

Invariant 2. *Consider any facility $i \in F$ with $\mathcal{C}(i) \neq \emptyset$. Then the unique critical cluster $C^* \in \mathcal{C}(i)$ satisfies: $\ell(C^*) \leq \ell(C)$ for all $C \in \mathcal{C}(i)$.*

The next invariant captures the fact that in an output of the relaxed greedy algorithm, if a client $j \in D$ gets assigned to a facility $i \in F$ then $\ell(j) \geq \Theta(\log d_{ij})$.

Invariant 3. *For every cluster $C = (i, A) \in \mathcal{C}$ and every client $j \in A$, we have $\ell(j) \geq \kappa^*_{ij}$, where κ^*_{ij} is the unique level s.t. $2^{\kappa^*_{ij}-4} \leq d_{ij} < 2^{\kappa^*_{ij}-3}$.*

Finally, we formulate an invariant which captures the fact that the relaxed greedy algorithm picks the clusters in an (approximately) increasing order of their average costs. Towards this end, we define the notion of a *blocking cluster*, as described below.

Definition 1. *Consider any clustering $\mathcal{C}$, any level $k \in \mathbb{Z}$, and any cluster $C = (i, A)$ that does* not *necessarily belong to $\mathcal{C}$. The cluster C is a* blocking cluster *at level k w.r.t. $\mathcal{C}$ iff three conditions hold. (a) We have $cost_{avg}(C) < 2^{k-3}$. (b) For every $j \in A$, we have $\ell(j) > k \geq \kappa^*_{ij}$, where κ^*_{ij} is the unique level s.t. $2^{\kappa^*_{ij}-4} \leq d_{ij} < 2^{\kappa^*_{ij}-3}$. (c) If C is an satellite cluster, then there is a critical cluster $C^* \in \mathcal{C}(i)$ with $\ell(C^*) \leq k$. Else if C is a critical cluster, then $k \leq \ell(C')$ for all $C' \in \mathcal{C}(i)$.*

Intuitively, if $C = (i, A)$ is a blocking cluster at level k w.r.t. $\mathcal{C}$, then the relaxed greedy algorithm should have formed the cluster C at some level $k' < k$ before proceeding to construct the subsequent levels which currently contain all the clients in A. This leads us to the invariant below.[3]

Invariant 4. *There is* no *blocking cluster w.r.t. the clustering $\mathcal{C}$ at any level $k \in \mathbb{Z}$.*

Remark. Consider a cluster $C = (i, A) \in \mathcal{C}$ at level $\ell(C) = k$. If $cost_{avg}(C) \ll 2^k$, then it is not difficult to show that there exists a subset of clients $A' \subseteq A$ such that $C' = (i, A')$ is a blocking cluster at level $k - 1$. This, along with Invariant 1, implies that $cost_{avg}(C) = \Theta(2^{\ell(C)})$ for all $C \in \mathcal{C}$.

The next theorem holds since a nice clustering corresponds to the output of the relaxed greedy algorithm, and since the (original) greedy algorithm is known to have an approximation ratio of $O(1)$. We defer the proof of Theorem 2 to Appendix B.

Theorem 2. *Any nice clustering forms a $O(1)$-approximate optimal solution to the concerned input instance of the facility location problem.*

2.3 Description of Our Dynamic Algorithm

In the dynamic setting, we simply maintain a nice clustering. To be more specific, w.l.o.g. we assume that $D = \emptyset$ at preprocessing. Subsequently, during each update, a client gets inserted into / deleted from the set D. We now describe how our algorithm handles an update.

Handling the deletion of a client j: Suppose that client j belonged to the cluster $C = (i, A) \in \mathcal{C}$ just before getting deleted. We set $A \leftarrow A \setminus \{j\}$, and then we call the subroutine described in Figure 1.

Handling the insertion of a client j: We identify any open facility $i \in F$.[4] Let $\kappa^*_{ij} \in \mathbb{Z}$ be the unique level such that $2^{\kappa^*_{ij}-4} \leq d_{ij} < 2^{\kappa^*_{ij}-3}$. Let $C = (i, A)$ be the unique critical cluster with facility i. If $\kappa^*_{ij} \leq \ell(C)$, then we set $A \leftarrow A \cup \{j\}$. In this case the client j becomes part of the critical cluster C at level $\ell(C)$. Else if $\kappa^*_{ij} > \ell(C)$, then we create a new satellite cluster $(i, \{j\})$ at level $\ell((i, \{j\})) = \kappa^*_{ij}$, and set $\mathcal{C} \leftarrow \mathcal{C} \cup \{(i, j)\}$. Next, we call the subroutine described in Figure 1.

The subroutine FIX-CLUSTERING**(.):** The insertion / deletion of a client might lead to a violation of Invariant 4 or Invariant 1 (the procedure described in the preceding two paragraphs ensure that Invariant 2 and Invariant 3 continue to remain satisfied). The subroutine FIX-CLUSTERING(.) handles this issue. It repeatedly checks whether Invariant 4 or Invariant 1 is violated, and accordingly calls FIX-BLOCKING(C, k) or FIX-LEVEL(C). Each of these last two subroutines has the property that it does not lead to any new violation of Invariant 2 or Invariant 3. Thus, when the WHILE loop in Figure 1 terminates, all the invariants are restored and we end up with a nice clustering.

```
WHILE either Invariant 4 or Invariant 1 is violated:
    IF Invariant 4 is violated, THEN
        Find a cluster C = (i, A) that is blocking at level k (say).
        Call the subroutine FIX-BLOCKING(C, k).
    ELSE IF Invariant 1 is violated, THEN
        Find a cluster C = (i, A) (say) that violates Invariant 1.
        Call the subroutine FIX-LEVEL(C).
```

Figure 1: FIX-CLUSTERING(.).

The subroutine FIX-BLOCKING(C, k)**:** This subroutine works as follows. We first add the cluster $C = (i, A)$ to the current clustering. Towards this end, we set $\ell(C) \leftarrow k$, $\mathcal{C} \leftarrow \mathcal{C} \cup \{C\}$, and perform the following operations for all clients $j \in A$.

- Let $C'_j = (i'_j, A'_j)$ be the cluster the client j belonged to just before the concerned call to the subroutine FIX-BLOCKING(C, k). Set $A'_j \leftarrow A'_j \setminus \{j\}$.

[3] The condition $\ell(j) \geq \kappa^*_{ij}$ guarantees that C, in some sense, is a minimal blocking cluster. This is because if we remove a client violating this condition from the cluster C, then $cost_{avg}(C)$ can never become $> 2^{k-\mu}$. We will use this condition while analyzing the amortized recourse of our dynamic algorithm.

[4] If j is the first client being inserted, then there is no existing open facility. In this case, we identify the facility $i \in F$ which minimizes $d_{ij} + f_i$, create a critical cluster $C = (i, \{j\})$ and assign it to the level $k \in \mathbb{Z}$ such that $d_{ij} + f_i \in \left[2^{k-4}, 2^{k-3}\right)$.

At this point, we fork into one of two cases. *Case (a):* If C is an satellite cluster, then there is nothing further that needs to be done, and we terminate the call to this subroutine. *Case (b):* If $C = (i, A)$ is a critical cluster, then let $C' = (i, A')$ denote the unique critical cluster with facility i just before the concerned call to the subroutine. Since there has to be exactly one critical cluster with facility i, we now convert the cluster $C' = (i, A')$ into (possibly) multiple satellite clusters by performing the following operations for all clients $j \in A'$.

- Create an satellite cluster $(i, \{j\})$ at level $\ell((i, \{j\})) = \ell(C')$, and set $\mathcal{C} \leftarrow \mathcal{C} \cup \{(i, \{j\})\}$.

Finally, we remove the previous critical cluster C' from the current clustering, by setting $\mathcal{C} \leftarrow \mathcal{C} \backslash \{C'\}$, and then terminate the call to this subroutine.

The subroutine FIX-LEVEL(C)**:** Suppose that the cluster $C = (i, A)$ is currently at level $k = \ell(C)$. We first deal with the corner case where $\mathcal{C}(i) = \{C\}$ and $A = \emptyset$. In this case, we simply set $\mathcal{C} \leftarrow \mathcal{C} \setminus \{C\}$, and then terminate the call to this subroutine. Next, we check if C is a critical cluster. If the answer is yes, then the cluster C *absorbs* all the existing satellite clusters involving facility i at level k. To be more specific, for all satellite clusters $(i, \{j\})$ at level k, we set $A \leftarrow A \cup \{j\}$ and $\mathcal{C} \leftarrow \mathcal{C} \setminus \{(i, \{j\})\}$. This ensures that Invariant 2 continues to remain satisfied. Finally, at this point if it continues to be the case that $cost_{avg}(C) \geq 2^k$, then we move the cluster C up to level $k + 1$, by setting $\ell(C) \leftarrow \ell(C) + 1$. We now terminate the call to this subroutine.

2.4 Analysis of Our Dynamic Algorithm

The approximation guarantee of our dynamic algorithm follows from Theorem 2 and Theorem 3. We defer the proof of Theorem 3 to Appendix A.

Theorem 3. *The clustering $\mathcal{C}$ maintained by our dynamic algorithm always remains nice.*

For the rest of this section, we focus on bounding the amortized recourse of our algorithm. Towards this end, whenever a client moves up (resp. down) one level, we say that the algorithm performs one unit of *up-work* (resp. *down-work*).

Lemma 1. *During any sequence of t updates, the total recourse of the algorithm is at most t plus $O(1)$ times the total number of units of down-work that get performed.*

Proof. First, note that the total facility-recourse is upper bounded, within a constant multiplicative factor, by the total client-recourse. This is because if we open a new facility, then it necessarily implies that at least one client gets reassigned to the newly open facility.

Next, observe that a client j gets assigned to a new facility i only if one of the following two events occurs. (1) The client j gets inserted during an update. (2) The client j participates in a blocking cluster $C = (i, A)$ at level k (say), and the algorithm calls the subroutine FIX-BLOCKING(C, k). The total recourse incurred due to events of type (1) is at most t. On the other hand, during an event of type (2) the concerned client j moves down from its current level $\ell(j)$ to a smaller level k, as per condition (b) in Definition 1. Thus, the total recourse incurred due to events of type (2) is at most the total units of down-work performed by the algorithm. The lemma follows. □

It now remains to bound the total units of down-work performed by our algorithm. We achieve this by means of a *token based* argument, which in turn, requires us to first introduce the notion of an *epoch* of a cluster, and then bound the total units of *up-work* performed by our algorithm during an epoch.

Epoch: Observe that when a concerned cluster $C = (i, A)$ first becomes part of the clustering $\mathcal{C}$ at some level (say) k, we have $cost_{avg}(C) < 2^{k-3}$ (see condition (a) in Definition 1). Subsequently, during its entire lifetime, the level of the cluster C can only increase (this can happen because of calls to the subroutine FIX-LEVEL(C)). Accordingly, we partition the lifetime of the concerned cluster $C = (i, A)$ into multiple *epochs*, where an epoch refers to a maximal time-interval during which the level of the cluster does *not* increase. To be a bit more specific, suppose that the lifetime of cluster C consists of epochs $\{0, 1, \ldots, \lambda\}$, for some integer $\lambda \geq 0$. An epoch $r \in [0, \lambda]$ begins when the cluster $C = (i, A)$ arrives at level $k + r$, and ends when the cluster moves up from level $k + r$ to level $k + r + 1$. Let A_r denote the state of the set A just before the end of epoch r. The algorithm performs $|A_r|$ units of up-work for the cluster $C = (i, A)$ in epoch r. We will now upper bound the total units of up-work performed for the cluster C during its entire lifetime, which is given by $\sum_{r=0}^{\lambda} |A_r|$. The proof of Lemma 2 appears in Appendix C.

Lemma 2. *Consider a cluster $C = (i, A)$ that first became part of the clustering $\mathcal{C}$ at level k. During its entire lifetime, our algorithm performs at most $f_i/2^{k-2}$ units of up-work on this cluster C.*

Tokens: Recall that the opening costs of the facilities and the distances between the clients and facilities are all polynomially bounded by $m = |F|$. Hence, there are at most $O(\log m)$ levels in the clustering $\mathcal{C}$. To be more specific, there are two integral parameters $L, U \in \mathbb{Z}$, with $0 \leq U - L = O(\log m)$, such that $\ell(j) \in [L, U]$ for all clients $j \in D$ at all times. To analyse the amortised recourse of our algorithm, we now associate some *tokens* with each client in the following manner.

$$\text{For each client } j \in D, \text{ there is a token at each level } k \in [L, \ell(j)]. \tag{1}$$

Observation 1 holds since there are only $O(\log m)$ levels in the clustering $\mathcal{C}$.

Observation 1. *Whenever a client is inserted, it leads to the* creation *of at most $O(\log m)$ tokens. In contrast, whenever a client is deleted, it leads to the* release *of at most $O(\log m)$ tokens. We assume that these released tokens get* deposited *into a special* bank-account.

Observation 2. *Whenever the algorithm performs one unit of down-work, it* releases *one token. We assume that $1/2$ of this released token gets* deposited *into the special* bank-account*, and the remaining $1/2$ of the token* disappears *after paying for the cost of the down-work being performed. In contrast, whenever the algorithm performs one unit of up-work, it* withdraws *one token from the special bank-account to ensure that (1) continues to hold.*

We assume that just before preprocessing, the special bank-account had zero balance. We will show that throughout the duration of our algorithm, the balance in this bank-account always remains non-negative. Towards this end, observe that the algorithm withdraws tokens from this account only when it performs up-work, and it performs up-work only when some cluster increases its level by one. Accordingly, we focus on a cluster $C = (i, A)$ that first becomes part of the clustering $\mathcal{C}$ at time τ_0 (say). Suppose that C was assigned to level k (say) at time τ_0. Let A^* denote the state of A at time τ_0. As per condition (a) of Definition 1, at time τ_0 we have:

$$\frac{f_i}{|A^*|} \leq cost_{avg}(C) = \frac{f_i + \sum_{j \in A^*} d_{ij}}{|A^*|} < 2^{k-3}.$$

Hence, we get $|A^*| \geq f_i/2^{k-3}$. The next observation holds since every client in A^* had to move down at least one level before the cluster C is formed at level k, as per condition (b) of Definition 1.

Observation 3. *Just before a cluster C becomes part of the clustering $\mathcal{C}$ for the first time (say, at a level k), the algorithm deposits at least $f_i/2^{k-4}$ many tokens into the special bank-account.*

Lemma 2 implies that at most $f_i/2^{k-2}$ tokens get withdrawn from the bank-account because of the cluster C. By Observation 3, this is at most the number of tokens that get deposited into the same account just before the cluster C is formed at time τ_0. It follows that the bank-account never runs into negative balance. Theorem 4 now follows from Observations 1, 2 and Lemma 1.

Theorem 4. *Starting from an empty set of clients, our algorithms spends at most $O(t \log m)$ total recourse to handle a sequence of t updates. So, the amortised recourse of our algorithm is $O(\log m)$.*

2.5 Implementation of our algorithm in $\widetilde{O}(m)$ amortized update time.

In Appendix D we show how to implement our algorithm in amortized $\widetilde{O}(m)$ time per update. Here, we briefly describe how to quickly (i.e., in $\widetilde{O}(m)$ time) check whether Invariant 4 is satisfied.

For every facility i and level k, let $S(i, k)$ denote the sequence of clients that are at a level larger than k, in increasing order of their distances from the facility i. The key observation is this: If there exists a blocking cluster involving facility i at level k, then there must exist a blocking cluster (with the same facility and at the same level) whose clients form a prefix of $S(i, k)$. This has the following implication. Suppose that we explicitly maintain these sequences $\{S(i, k)\}$. Then given a specific pair (i, k), we can determine in only $\widetilde{O}(1)$ time (via a simple binary search and some standard data structures) whether or not there is a blocking cluster involving facility i at level k. Finally, note that we can indeed maintain the sequences $\{S(i, k)\}$ by paying a factor $\widetilde{O}(m)$ overhead in our update time. This is because whenever a client changes its level (or gets inserted / deleted) we need to modify at most $mL = \widetilde{O}(m)$ of these sequences $\{S(i, k)\}$, where L is the number of levels.

In Appendix D, we present a more fine tuned data structure, where we discretize the distances in powers of $(1 + \epsilon)$. However, the main idea behind the data structure is the same as described above.

3 Experimental Evaluation

Datasets We experiment with three classic datasets [5] from UCI library Dua and Graff [2017]: KDD-Cup Stolfo et al. [2000] ($311,029$ points of dimension 74) and song Bertin-Mahieux et al. [2011] ($515,345$ points of dimension 90) Census Kohavi et al. [1996] ($2,458,285$ points of dimension 68). We did not alter in any way the dimensions of the data.

For each dataset, we only keep a $5,000$ points for each experiment as suffices to show the merits and limitations of the different approaches; we also confirm that the relative behavior of the algorithms remains the same over the whole sequence of updates on one of the dataset that we consider. We consider the L_2 distances of the embedding, and add $1/5000$ to each distance to avoid having distances zero while not disturbing the structure of the instance.

Order of updates. We consider insertions and deletions of points under the sliding window model, where we order the points according to some permutation, and then we consider an interval (a.k.a., window) of $\ell = 1,000$ indices, which we slide from the beginning to the end of the ordered points and at each step form an instance containing the points that lie within the window limits; that is, as we slide the window, the new point that gets within the limits of the window is treated as an insertion, while each point that falls outside the window limits is treated as a deletion.

Choice of facilities. We consider two different processes for choosing the set of facilities and their opening cost. The first way is to pick uniformly at random 5% of the point in the instance, remove them from the set of clients, and make them facilities. In the second variant we use 50% of the points as facilities. Using these two processes we generate six families of instances: Census-5%, Census-50%, KDD-Cup-5%, KDD-Cup-50%, song-5%, and song-50%. These instances aim to cover the settings where the set of facilities is very restrictive or very broad, in order to assess the behavior of the algorithms in both of these settings.

We set the facility opening costs within a range to ensure that the problem does not become too easy; that is, we don't want to the trivially good solutions of either opening only a single, or opening all facilities. We set the facility weights as follows: for a given instance, we calculated the median distance of clients to their nearest facility, and multiplied this number by 100.

Algorithms We compare our algorithm against two main competing algorithms: 1) Following each point insertion or deletion, we rerun from scratch the offline Greedy algorithm Jain et al. [2003], which we call GREEDYOFF and 2) The state of the art $O(\log(m))$ approximation algorithm for the fully-dynamic facility location problem from Guo et al. [2020], which we call GKLX (from the authors last names). We also compare our algorithm to the baseline that simply maintains open the nearest facility of each of client. We call this baseline NEAREST FACILITY.

The behavior of our algorithm depends on two parameters: μ and ϵ. The parameter ϵ defines the base of the exponential bucketing scheme, and the parameter μ defines the level κ^*_{ij} (see Invariant 3), so that $(1+\epsilon)^{\kappa^*_{ij}-\mu-1} \leq d_{ij} < (1+\epsilon)^{\kappa^*_{ij}-\mu}$. In Section 2.2 we set $\epsilon = 1$ and $\mu = 3$ for ease of analysis. Both of these control parameters trade-off between running time, approximation, and recourse. While our theoretical part covers the case where $\mu \geq 3$ and $\epsilon = 1$, we go beyond what is proven and explore setting values $\mu \in \{1, 3\}$ and $\epsilon \in \{0.05, 1\}$. We use NICECLUSTERING($\mu = 1, \epsilon = 0.05$), NICECLUSTERING($\mu = 1, \epsilon = 1$), NICECLUSTERING($\mu = 3, \epsilon = 0.05$), NICECLUSTERING($\mu = 3, \epsilon = 1$) to refer to the different configurations of our algorithm. We note that we did not try to optimize μ and ϵ to minimize the approximation guarantee (the main goal was to get any constant) for simplicity of the presentation and the analysis, although we believe this is doable.

We defer the interested reader to Appendix D for implementation details.

Setup. All of our code is written in C++ and is available online [6]. We used a e2-standard-16 Google Cloud instance, with 16 cores, 2.20GHz Intel(R) Xeon(R) processor, and 64 GiB main memory.

[5] None of these datasets is ideal for our setting as they lack information that determines the order of insertions and deletions. However, we are not aware of any targeted datasets that are openly available. We note that our datasets and setup are typical in studies of dynamic algorithms (see, e.g., Cohen-Addad et al. [2019]).

[6] `https://github.com/google-research/google-research/tree/master/fully_dynamic_facility_location`

3.1 Results

Throughout this section, as the relative behavior of the algorithms is similar among the different datasets, we use the KDD-Cup-5% and KDD-Cup-50% datasets to analyze the performance of the algorithm, and summarize the behavior of the algorithms for the rest of the datasets (we defer the detailed description and the plots of the remaining datasets to Appendix E).

Given that our process of creating the instances contain randomized decisions, we repeat each of our experiments 10 times and report the mean relative behavior of the algorithms in Appendix F. For visualization purposes, we discuss the outcome of a single experiment; repeating the experiment multiple times does not alter significantly the relative behavior of the algorithms.

Moreover, we assessed the performance of NICECLUSTERING$(\mu = 1, \epsilon = 0.05)$, it performs identically to NICECLUSTERING$(\mu = 3, \epsilon = 0.05)$ in terms of solution cost and time, but exhibits slightly larger recourse. Hence, we omit it to avoid complicating the plots and tables. Finally, we only run GREEDYOFF on the datasets KDD-Cup-5% and KDD-Cup-50%, and only once, for illustration purposes as it is very computationally expensive to run.

Solution cost. We analyze the different algorithms in terms of the cost of their solution. Figure 2 summarizes the results for both settings where we consider 5% and 50% of the points to be facilities.

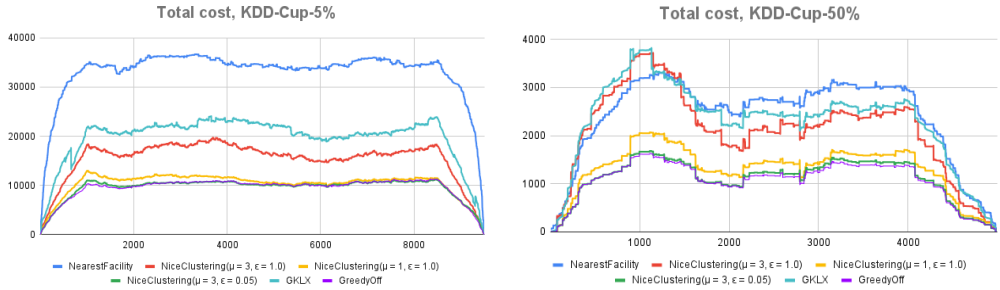

Figure 2: The solution cost by the different algorithms over the whole sequence of updates, on KDD-Cup-5% and KDD-Cup-50%.

As expected, the smaller the values of μ and ϵ in NICECLUSTERING, the better the solution quality. NICECLUSTERING$(\mu = 3, \epsilon = 0.05)$ perform almost identically to GREEDYOFF, which is an algorithm that is very computationally expensive. Over 10 repetitions, NICECLUSTERING$(\mu = 3, \epsilon = 0.05)$ performs roughly $17\% - 29\%$ better and $55\% - 67\%$ better than GKLX on KDD-Cup-5% and KDD-Cup-50%, respectively. The $> 55\%$ happens because in this particular dataset there are a few facilities with small distances to many clients, GKLX prioritizes minimizing opening costs rather than connection cost in this case. On the other datasets NICECLUSTERING$(\mu = 3, \epsilon = 0.05)$ outperforms GKLX by $14\% - 30\%$ (see Appendix E).

Recourse. Next, we study the recourse occurred by each of the algorithms that we consider on the datasets KDD-Cup-5% and KDD-Cup-50%. Figure 3 summarizes the results of this experiment.

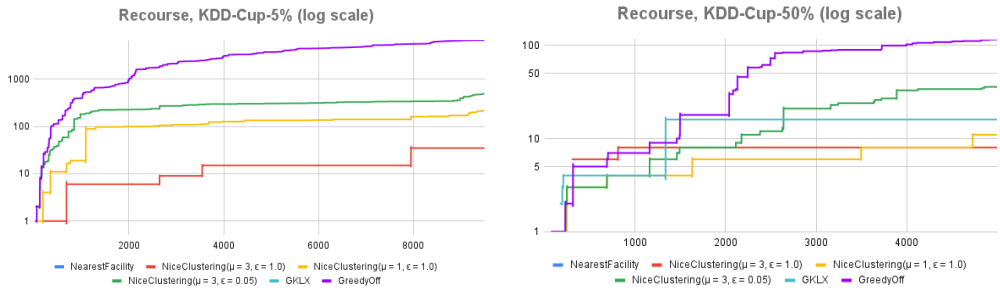

Figure 3: The cumulative recourse incurred by the algorithms that we consider over the whole sequence of updates, on KDD-Cup-5% (left), and KDD-Cup-50% (right). Missing lines imply 0 recourse throughout. The plot is in log scale.

While all algorithms, except GREEDYOFF, have limited recourse (considering the window size is $1,000$ and the length of the update sequence is $9,500$ and $5,000$ for KDD-Cup-5% and KDD-

Cup-50%, respectively), GKLX performs the best (NEARESTFACILITY trivially has no recourse). Naturally, smaller values of μ and ϵ in our algorithm imply that it gets closer GREEDYOFF (which has no recourse guarantees) in terms of cost, and this implies higher recourse. While GKLX performs better than NICECLUSTERING on KDD-Cup-5%, they are comparable on KDD-Cup-50%. Repeating the experiment 10 times show similar relative behavior.

Running time. As expected, our algorithm is slower compared to GKLX. NICECLUSTERING($\mu = 3, \epsilon = 1$) NICECLUSTERING($\mu = 1, \epsilon = 1$) are one order of magnitude slower. This is dues to the fact that GKLX, apart from the initial preprocessing phase and determining the nearest facility to each client, spends only polylogarithmic time per update as it operates on a tree structure of logarithmic depth. In fact, after the preprocessing phase, GKLX does not need to store the connections between clients and facilities; it suffices to remember the nearest facility to each client. The reduced space allows better cashing and fewer main memory accesses. On the other hand, our algorithm stores explicitly the connection costs of clients and visit them regularly. NICECLUSTERING($\mu = 3, \epsilon = 0.05$) is slower by another order of magnitude, due to the increased number of level in our algorithm. We refer to Appendix E.1 for the running time plots. Finally, GREEDYOFF is over two orders of magnitude slower compared to the slowest of the variants of our algorithm.

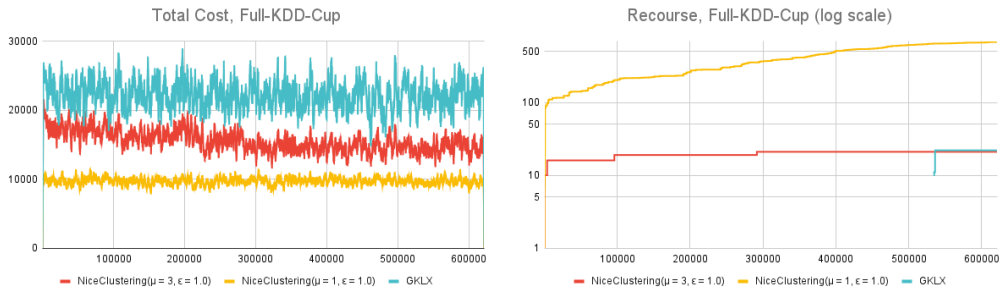

Figure 4: The solution cost by the algorithms NICECLUSTERING($\mu = 3, \epsilon = 1$), NICECLUSTERING($\mu = 1, \epsilon = 1$), GKLX over the whole sequence of updates, on Full-KDD-Cup (left), and the cumulative recourse incurred by the same algorithms, in log scale, on Full-KDD-Cup (right). Missing lines imply 0 recourse throughout.

Consistency over the long sequence of updates To showcase that the relative behavior of the algorithms is not affected by our choice to restrict each datasets to the first 5000 points, we run the algorithm on whole KDD-Cup dataset consisting of 311,029 points, where we choose 250 points as facilities (the same absolute number as in the case of KDD-Cup-5%). Similarly to the rest of the experiments, we consider insertions and deletions of points under the sliding window model (with $\ell = 1000$). We refer to this instance as Full-KDD-Cup. We compare the most competitive algorithms, that is NICECLUSTERING($\mu = 3, \epsilon = 1$), NICECLUSTERING($\mu = 1, \epsilon = 1$), and GKLX. We compare the three algorithm in terms of the total cost, and recourse; the results are shown in Figure 4. The conclusions from the previous experiments on the restricted dataset hold also in this experiment. However, we observe two behaviors. First, the performance of GKLX fluctuates more compared to NICECLUSTERING($\mu = 3, \epsilon = 1$) and NICECLUSTERING($\mu = 1, \epsilon = 1$). This can be attributed to the fact that the algorithm doesn't adapt rapidly to the changes in the instance. To a lesser extend, the behavior of NICECLUSTERING($\mu = 3, \epsilon = 1$) also fluctuates more than NICECLUSTERING($\mu = 1, \epsilon = 1$), for the same reasons. The second observation is that while NICECLUSTERING($\mu = 3, \epsilon = 1$) and GKLX incur very little recourse, while NICECLUSTERING($\mu = 1, \epsilon = 1$) continuously incurs (limited) recourse. These two observations are related. This implies that NICECLUSTERING($\mu = 1, \epsilon = 1$) adapts more rapidly to the changing dataset so that it is able to maintain a better solution.

Conclusion of Experiments. We showed that our algorithm outperforms, often significantly, in terms of solution cost the GKLX algorithm, but in some cases exhibits an increase in recourse (while still being a small absolute number) and it is also slower. We believe the our algorithm is a good fit for applications where solution quality is very important, and one wants to maintain a stable solution in reasonable time. Given that our algorithm performs almost in par with GREEDYOFF, it becomes a good choice in cost-sensitive applications on dynamic data.

References

Thierry Bertin-Mahieux, Daniel PW Ellis, Brian Whitman, and Paul Lamere. The million song dataset. 2011.

TH Hubert Chan, Arnaud Guerqin, and Mauro Sozio. Fully dynamic k-center clustering. In *Proceedings of the 2018 World Wide Web Conference*, pages 579–587, 2018.

Vincent Cohen-Addad, Niklas Oskar D Hjuler, Nikos Parotsidis, David Saulpic, and Chris Schwiegelshohn. Fully dynamic consistent facility location. *Advances in Neural Information Processing Systems*, 32, 2019.

Marek Cygan, Artur Czumaj, Marcin Mucha, and Piotr Sankowski. Online facility location with deletions. In *Proceedings 26th Annual European Symposium on Algorithms*, page 21:1–21:15, 2018.

Artur Czumaj, Christiane Lammersen, Morteza Monemizadeh, and Christian Sohler. (1+ ε)-approximation for facility location in data streams. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1710–1728. SIAM, 2013.

Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL `http://archive.ics.uci.edu/ml`.

Hendrik Fichtenberger, Silvio Lattanzi, Ashkan Norouzi-Fard, and Ola Svensson. Consistent k-clustering for general metrics. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2660–2678. SIAM, 2021.

Dimitris Fotakis. On the competitive ratio for online facility location. *Algorithmica*, 50(1):1–57, 2008.

Gramoz Goranci, Monika Henzinger, and Dariusz Leniowski. A tree structure for dynamic facility location. In *Proceedings 26th Annual European Symposium on Algorithms*, page 39:1–39:13, 2018.

Gramoz Goranci, Monika Henzinger, Dariusz Leniowski, Christian Schulz, and Alexander Svozil. Fully dynamic k-center clustering in low dimensional metrics. In *2021 Proceedings of the Workshop on Algorithm Engineering and Experiments (ALENEX)*, pages 143–153. SIAM, 2021.

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

Xiangyu Guo, Janardhan Kulkarni, Shi Li, and Jiayi Xian. On the facility location problem in online and dynamic models. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

Xiangyu Guo, Janardhan Kulkarni, Shi Li, and Jiayi Xian. Consistent k-median: Simpler, better and robust. In *International Conference on Artificial Intelligence and Statistics*, pages 1135–1143. PMLR, 2021.

Anupam Gupta, Ravishankar Krishnaswamy, Amit Kumar, and Debmalya Panigrahi. Online and dynamic algorithms for set cover. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 537–550, 2017.

Monika Henzinger, Dariusz Leniowski, and Claire Mathieu. Dynamic clustering to minimize the sum of radii. *Algorithmica*, 82(11):3183–3194, 2020.

Piotr Indyk. Algorithms for dynamic geometric problems over data streams. In *Proceedings of the thirty-sixth annual ACM Symposium on Theory of Computing*, pages 373–380, 2004.

Mohammad Reza Karimi Jaghargh, Andreas Krause, Silvio Lattanzi, and Sergei Vassilvtiskii. Consistent online optimization: Convex and submodular. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2241–2250. PMLR, 2019.

Kamal Jain, Mohammad Mahdian, Evangelos Markakis, Amin Saberi, and Vijay V Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing lp. *Journal of the ACM (JACM)*, 50(6):795–824, 2003.

Ron Kohavi et al. Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In *Kdd*, volume 96, pages 202–207, 1996.

Bernhard Korte and Jens Vygen. *Facility Location*, pages 629–665. Springer Berlin Heidelberg, Berlin, Heidelberg, 2018.

Christiane Lammersen and Christian Sohler. Facility location in dynamic geometric data streams. In *European Symposium on Algorithms*, pages 660–671. Springer, 2008.

Silvio Lattanzi and Sergei Vassilvitskii. Consistent k-clustering. In *International Conference on Machine Learning*, pages 1975–1984. PMLR, 2017.

Silvio Lattanzi, Slobodan Mitrović, Ashkan Norouzi-Fard, Jakub M Tarnawski, and Morteza Zadimoghaddam. Fully dynamic algorithm for constrained submodular optimization. *Advances in Neural Information Processing Systems*, 33:12923–12933, 2020.

Shi Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. *Information and Computation*, 222:45 – 58, 2013. ISSN 0890-5401. doi: https://doi.org/10.1016/j.ic.2012.01.007. 38th International Colloquium on Automata, Languages and Programming (ICALP 2011).

Adam Meyerson. Online facility location. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 426–431. IEEE, 2001.

S.J. Stolfo, Wei Fan, Wenke Lee, A. Prodromidis, and P.K. Chan. Cost-based modeling for fraud and intrusion detection: results from the jam project. In *Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00*, volume 2, pages 130–144 vol.2, 2000. doi: 10.1109/DISCEX.2000.821515.

Paper checklist

Q (a):Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

A: Yes. We prove the claims for theoretical guarantee, and conduct the appropriate experiments to support our claims on performance.

Q (b): Have you read the ethics review guidelines and ensured that your paper conforms to them?

A: Yes.

Q (c): Did you discuss any potential negative societal impacts of your work?

A: No. This paper studies a basic algorithmic question. We are not aware of direct applications of the facility location problem with negative societal impact. Our findings and code are made openly available for all.

Q (d): Did you describe the limitations of your work?

A: Yes. In the experimental section we discuss both strengths and weaknesses of our algorithm.

On theoretical results

Q (a): Did you state the full set of assumptions of all theoretical results?

A: yes.

Q (b): Did you include complete proofs of all theoretical results?

A: yes.

On experiments

Q (a): Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)?

A: No. We will the code is made available.

Q (b): Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)?

A: There is no training in this paper.

Q (c): Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)?

A: Yes, we repeated the experiments 10 times, and report the mean and standard deviation in the Appendix F.

Q (d): Did you include the amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)?

A: Yes. The experimental section reports the type of machine that we used.

On using existing assets.

Q (a): If your work uses existing assets, did you cite the creators?

A: Yes. We cited the datasets that we used.

Q (b): Did you mention the license of the assets?

A: No. These datasets have no individual licences in the UCI repository.

Q (c): Did you include any new assets either in the supplemental material or as a URL?

A: No. We released the code.

Q (d): Did you discuss whether and how consent was obtained from people whose data you're using/curating?

A: Yes. They the data are available at the UCI machine learning repository, and no additional licence is reported.

Q (e): Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content?

A: No. This is clear since the data are simply embeddings.

A Proof of Theorem 3

By induction hypothesis, suppose that the clustering $\mathcal{C}$ satisfies Invariant 1, Invariant 2, Invariant 4 and Invariant 3 just before a given update. While handling the update, our dynamic algorithm calls the subroutine in Figure 1. The main WHILE loop in Figure 1 immediately ensures that Invariant 1 and Invariant 4 continue to hold when our dynamic algorithm is done with processing the update.

Moving on to Invariant 2, it is easy to verify that this invariant is not violated by our dynamic algorithm at any point in time. Specifically, this invariant does not get violated by any step in Figure 1, or by the steps performed immediately before calling the subroutine described in Figure 1.

It now remains to prove that our algorithm always satisfy Invariant 3. Towards this end, fix any client $j \in D$. If the client j gets assigned to facility $i \in F$ immediately after getting inserted, then the manner in which our algorithm handles an insertion ensures that $\ell(j) \geq \kappa^*_{ij}$ at that point in time. Subsequently, consider each of the following two three of events.

(1) The client j remains assigned to the facility i, but moves down to a smaller level (say) k. This happens only if there is a blocking cluster $C = (i, A)$ with $j \in A$ at level k, and we call the subroutine FIX-BLOCKING(C, k). Hence, Definition 1 ensures that $k \geq \kappa^*_{ij}$. Thus, we continue to have $\ell(j) = k \geq \kappa^*_{ij}$ immediately after this event.

(2) The client j remains assigned to the facility i, but moves up to a larger level. Clearly, this event cannot lead to a violation of Invariant 3.

(3) The client j gets reassigned to a different facility i', by joining a cluster at level k (say) that has facility i' as its center. This happens only if the client j is part of a blocking cluster $C' = (i', A')$ at level k, with $j' \in A'$, and we call the subroutine FIX-BLOCKING(C', k). Hence, Definition 1 ensures that $k \geq \kappa^*_{i'j}$. Thus, we continue to have $\ell(j) = k \geq \kappa^*_{i'j}$ immediately after this event.

The above discussion implies that Invariant 3 continues to remain satisfied all the time, even when the algorithm is handling an update. This concludes the proof of Theorem 3.

B Proof of Theorem 2

We first recall the LP relaxation for the facility location problem:

$$\text{Minimize} \sum_i f_i \cdot y_i + \sum_j \sum_i d_{ij} \cdot x_{ij} \tag{2}$$

$$\text{s.t.} \sum_{i \in F} x_{ij} \geq 1 \qquad \text{for all } j \in D \tag{3}$$

$$x_{ij} \leq y_i \qquad \text{for all } i \in F, j \in D \tag{4}$$

$$y_i, x_{ij} \geq 0 \qquad \text{for all } i \in F, j \in D \tag{5}$$

The dual LP is given below.

$$\text{Maximize} \sum_{j \in D} \alpha_j \tag{6}$$

$$\text{s.t.} \sum_{j \in D} \beta_{ij} \leq f_i \qquad \text{for all } i \in F \tag{7}$$

$$\alpha_j - \beta_{ij} \leq d_{ij} \qquad \text{for all } i \in F, j \in D \tag{8}$$

$$\alpha_j, \beta_{ij} \geq 0 \qquad \text{for all } i \in F, j \in D \tag{9}$$

We will prove Theorem 2 using the primal-dual method. Specifically, we will show that given any nice clustering, we can come up with a pair of feasible primal and dual solutions whose objectives are within a $O(1)$ multiplicative factor of each other. For the rest of the proof, fix any nice clustering $\mathcal{C}$.

Based on the clustering $\mathcal{C}$, we construct the following natural primal solution (x, y). For all $i \in F$, we set $y_i = 1$ if facility i is open (i.e., if $\mathcal{C}(i) \neq \emptyset$); otherwise, we set $y_i = 0$. Similarly, for all

$i \in F, j \in D$, we set $x_{ij} = 1$ if the clustering $\mathcal{C}$ assigns client j to facility i; otherwise, we set $x_{ij} = 0$. It is easy to check that this gives us a feasible primal solution. Next, based on the clustering $\mathcal{C}$, we construct an (infeasible) dual solution (α, β) as described in Figure 5.

1. FOR ALL clusters $C = (i, A) \in \mathcal{C}$:
2. FOR ALL clients $j \in A$:
3. Set $\alpha_j \leftarrow 2^{\ell(C)}$.
4. FOR ALL facilities $i' \in F$:
5. Set $\beta_{ij} \leftarrow \max\big(0, \alpha_j/2^{10} - d_{ij}\big)$.

Figure 5: Constructing an (infeasible) dual solution (α, β) based on a nice clustering $\mathcal{C}$.

We now show that the primal objective at (x, y) is at most the dual objective at (α, β).

Lemma 3. *We have* $\sum_{i \in F} f_i \cdot x_i + \sum_{i \in F, j \in D} d_{ij} \cdot y_{ij} \leq \sum_{j \in D} \alpha_j$.

Proof. The cost of the primal solution (x, y) is equal to the total cost of all the clusters in the dual solution (α, β). Thus, we have:

$$\sum_{i \in F} f_i \cdot x_i + \sum_{i \in F, j \in D} d_{ij} \cdot y_{ij} = \sum_{C \in \mathcal{C}} cost(C). \tag{10}$$

Next, fix any cluster $C = (i, A) \in \mathcal{C}$, and observe that:

$$cost(C) = |C| \cdot cost_{avg}(C) \leq |C| \cdot 2^{\ell(C)} = \sum_{j \in A} \alpha_j. \tag{11}$$

The lemma now follows from (10) and (11) and the observation that each client $j \in A$ appears in exactly one cluster $C \in \mathcal{C}$. □

The next lemma shows that $(\alpha/2^{10}, \beta)$ is a feasible dual solution.

Lemma 4. *For all $j \in D$, define $\hat{\alpha}_j := \alpha_j/2^{10}$. Then $(\hat{\alpha}, \beta)$ is a feasible solution to LP (6).*

Theorem 2 follows from Lemma 3 and Lemma 4. The proof of Lemma 4 appears in Appendix B.1.

B.1 Proof of Lemma 4

Step 5 of Figure 5 ensures that the solution $(\hat{\alpha}, \beta)$ satisfies constraint (8) of the dual LP. It now remains to show that the solution $(\hat{\alpha}, \beta)$ also satisfies constraint (7) of the dual LP. This is achieved by Lemma 6. In other words, Lemma 4 follows from Lemma 6.

Lemma 5. *Consider any facility $i \in F$ and clients $j, j' \in D$ with $\alpha_j \geq d_{ij}$ and $\alpha_{j'} \geq d_{ij'}$. Then we have: $\alpha_j/2^4 \leq d_{ij} + 2\alpha_{j'}$.*

Proof. If $\alpha_j \leq \alpha_{j'}$, then the lemma trivially holds; so we assume that $\alpha_j > \alpha_{j'}$ throughout the rest of the proof. From step 3 of Figure 5, it follows that each client $j^* \in D$ appears at a level $\ell(j^*) = \log(\alpha_{j^*})$ in the clustering $\mathcal{C}$. Accordingly, since $\alpha_{j'} < \alpha_j$, we have $\ell(j') \leq \ell(j) - 1$.

Let i' denote the facility to which the client j' gets assigned in the nice clustering $\mathcal{C}$. Now we claim that $d_{i'j} \geq \alpha_j/2^4$. For the sake of contradiction, suppose $d_{i'j} < \alpha_j/2^4$. We will show that $C := (i', \{j\})$ forms an satellite blocking cluster at level $k := \ell(j) - 1$. Towards this end, we note:

- (a) $cost_{avg}(C) = d_{i'j} < \alpha_j \cdot 2^{-4} = 2^{\ell(j)-4} = 2^{k-3}$.
- (b) $\ell(j) > k \geq \kappa^*_{i'j}$, where $\kappa^*_{i'j}$ is the unique level such that $2^{\kappa^*_{i'j}-4} \leq d_{i'j} < 2^{\kappa^*_{i'j}-3}$. Here, the first inequality holds because $k := \ell(j) - 1$, whereas the second inequality holds because $2^{\kappa^*_{i'j}-4} \leq d_{i'j} < 2^{k-3}$ as per condition (a) above.
- (c) Since i' is the center of a cluster at level $\ell(j')$ and $\mathcal{C}$ is a nice clustering, there exists a critical cluster $C' \in \mathcal{C}(i')$ with $\ell(C') \leq \ell(j') \leq \ell(j) - 1 = k$.

Conditions (a), (b) and (c) imply that $C = (i', \{j\})$ is an satellite blocking cluster at level $\ell(j) - 1$, as per Definition 1. This contradicts the fact that $\mathcal{C}$ is a nice clustering. Hence, we get:

$$d_{i'j} \geq \alpha_j / 2^4. \tag{12}$$

Next, since client $j' \in D$ is assigned to facility $i' \in F$ in the nice clustering $\mathcal{C}$, by Invariant 3 we have $\kappa^*_{ij'} \leq \ell(j')$. This implies that:

$$d_{i'j'} < 2^{\kappa^*_{i'j'}} \leq 2^{\ell(j')} = \alpha_{j'}. \tag{13}$$

Applying triangle inequality, we now obtain $d_{i'j} \leq d_{i'j'} + d_{ij'} + d_{ij} \leq d_{i'j'} + \alpha_{j'} + d_{ij} \leq 2 \cdot \alpha'_j + d_{ij}$ (the last inequality follows from (13)), and $d_{ij'} \leq \alpha_{j'}$ by our starting assumption). Thus, from (12) we infer that: $\alpha_j / 2^4 \leq d_{i'j} \leq 2\alpha'_j + d_{ij}$. □

Lemma 6. *For every facility $i \in F$, we have $\sum_{j \in D} \max\left(0, \frac{\alpha_j}{2^{10}} - d_{ij}\right) \leq f_i$.*

Proof. Fix a facility $i \in F$ for the rest of the proof. Define $C_i := \left\{ j \in D \mid \frac{\alpha_j}{2^{10}} \geq d_{ij} \right\}$. Note that:

$$\sum_{j \in D} \max\left(0, \frac{\alpha_j}{2^{10}} - d_{ij}\right) = \sum_{j \in C_i} \left(\frac{\alpha_j}{2^{10}} - d_{ij}\right) \tag{14}$$

Accordingly, for the rest of the proof, we focus on upper bounding the right hand side of (14).

For ease of notations, we assume that $C_i = \{j_1, \ldots, j_q\}$ and w.l.o.g. $\alpha_{j_1} \leq \alpha_{j_2} \leq \cdots \leq \alpha_{j_q}$. Say that a client $j \in C_i$ is *big* if $d_{ij} \geq \alpha_{j_1}/2^4$ and *small* otherwise. Let $B \subseteq C_i$ and $S = C_i \setminus B$ respectively denote the set of all big and small clients in C_i. Observe that:

$$\alpha_{j_1} - d_{ij} \leq 2^4 \cdot d_{ij} \text{ for all } j \in B \implies \sum_{j \in B} (\alpha_{j_1} - d_{ij}) \leq 2^4 \cdot \sum_{j \in B} d_{ij}. \tag{15}$$

Next, we claim that:

$$\sum_{j \in S} (\alpha_{j_1} - d_{ij}) \leq 2^4 \cdot f_i + (2^4 - 1) \cdot \sum_{j \in S} d_{ij}. \tag{16}$$

For the sake of contradiction, suppose that (16) does not hold. Then we have:

$$|S| \cdot \alpha_{j_1} > 2^4 \left(f_i + \sum_{j \in S} d_{ij} \right) \implies \frac{\alpha_{j_1}}{2^4} > \frac{f_i + \sum_{j \in S} d_{ij}}{|S|}. \tag{17}$$

We now consider two possible cases.

Case 1: There is no critical cluster in $\mathcal{C}(i)$ at level $\leq k := \ell(j_1) - 1 = \log(\alpha_{j_1}) - 1$.

In this case, consider the critical cluster $C := (i, S)$. We will show that C is a blocking cluster at level k w.r.t. the clustering $\mathcal{C}$. To see why this is true, observe that:

- (a) $cost_{avg}(C) := \frac{f_i + \sum_{j \in S} d_{ij}}{|S|} < \frac{\alpha_{j_1}}{2^4} = \frac{2^{\ell(j_1)}}{2^4} = 2^{k-3}$.
- (b) Consider any client $j \in S$. Since $\alpha_j \geq \alpha_{j_1}$, we have $\ell(j) = \log(\alpha_j) \geq \log(\alpha_{j_1}) = \ell(j_1) > k$. Furthermore, since $j \in S$, we have $d_{ij} < \alpha_{j_1}/2^4 = 2^{k-3}$. This implies that $k \geq \kappa^*_{ij}$, where κ^*_{ij} is the unique level such that $2^{\kappa^*_{ij} - 4} \leq d_{ij} < 2^{\kappa^*_{ij} - 3}$. To summarize, we infer that $\ell(j) > k \geq \kappa^*_{ij}$ for all clients $j \in S$.
- (c) We have $\ell(C') > k$ for all clusters $C' \in \mathcal{C}(i)$. This follows from Invariant 2 and our assumption that there is no critical cluster in $\mathcal{C}(i)$ at level $\leq k$.

Conditions (a), (b) and (c) above imply that C is a critical blocking cluster at level k w.r.t. the clustering $\mathcal{C}$, as per Definition 1. But this contradicts the fact that $\mathcal{C}$ is a nice clustering. Thus, we conclude that (16) holds in this case.

Case 2: There is a critical cluster $C^* \in \mathcal{C}(i)$ at level $\ell(C^*) \leq k$.

In this case, let $j \in S$ be a client such that $d_{ij} = \min_{j' \in S} d_{ij'}$. Consider the satellite cluster $C := (i, \{j\})$. We will show that C is a blocking cluster at level $k := \ell(j_1) - 1$ w.r.t. the clustering $\mathcal{C}$. To see why this is true, observe that:

- (a) $cost_{avg}(C) := d_{ij} < \frac{\alpha_{j_1}}{2^4} = \frac{2^{\ell(j_1)}}{2^4} = 2^{k-3}$.
- (b) Since $\alpha_j \geq \alpha_{j_1}$, we have $\ell(j) = \log(\alpha_j) \geq \log(\alpha_{j_1}) = \ell(j_1) > k$. Furthermore, since $j \in S$, we have $d_{ij} < \alpha_{j_1}/2^4 = 2^{k-3}$. This implies that $k \geq \kappa^*_{ij}$, where κ^*_{ij} is the unique level such that $2^{\kappa^*_{ij}-4} \leq d_{ij} < 2^{\kappa^*_{ij}-3}$. To summarize, we infer that $\ell(j) > k \geq \kappa^*_{ij}$.
- (c) There is a critical cluster $C^* \in \mathcal{C}(i)$ at level $\ell(C^*) \leq k$.

Conditions (a), (b) and (c) above imply that C is an satellite blocking cluster at level k w.r.t. the clustering $\mathcal{C}$, as per Definition 1. But this contradicts the fact that $\mathcal{C}$ is a nice clustering. Thus, we conclude that (16) holds in this case as well.

To summarize, we have proved that (16) holds. Thus, from (15) and (16), we get:

$$\begin{aligned}\sum_{j \in C_i} (\alpha_{j_1} - d_{ij}) &= \sum_{j \in B} (\alpha_{j_1} - d_{ij}) + \sum_{j \in S} (\alpha_{j_1} - d_{ij}) \\ &\leq 2^4 \cdot \sum_{j \in B} d_{ij} + 2^4 \cdot f_i + (2^4 - 1) \cdot \sum_{j \in S} d_{ij} \\ &\leq 2^4 \cdot f_i + 2^4 \cdot \sum_{j \in C_i} d_{ij}. \end{aligned} \tag{18}$$

Now consider any client $j \in C_i$. Since $j, j_1 \in C_i$, by definition we have $d_{ij} \leq \alpha_j$ and $d_{ij_1} \leq \alpha_{j_1}$. Accordingly, Lemma 5 gives us:

$$\alpha_j/2^4 \leq d_{ij} + 2 \cdot \alpha_{j_1} \implies \alpha_j \leq 2^4 \cdot d_{ij} + 2^5 \cdot \alpha_{j_1}. \tag{19}$$

Summing (19) over all the clients $j \in C_i$, we get:

$$\begin{aligned}\sum_{j \in C_i} \alpha_j &\leq \sum_{j \in C_i} \left(2^4 \cdot d_{ij} + 2^5 \cdot \alpha_{j_1}\right) = (2^4 + 2^5) \sum_{j \in C_i} d_{ij} + 2^5 \sum_{j \in C_i} (\alpha_{j_1} - d_{ij}) \\ &\leq (2^4 + 2^5) \sum_{j \in C_i} d_{ij} + 2^9 f_i + 2^9 \sum_{j \in C_i} d_{ij} \qquad \text{(follows from (18))} \\ &\leq 2^{10} \sum_{j \in C_i} d_{ij} + 2^{10} f_i \implies \sum_{j \in C_i} \left(\frac{\alpha_j}{2^{10}} - d_{ij}\right) \leq f_i. \end{aligned} \tag{20}$$

The lemma now follows from (14) and (20). □

C Proof of Lemma 2

We use the notations introduced in the paragraph just before the statement of Lemma 2. Note that during its entire lifetime, our algorithm performs $\sum_{r=0}^{\lambda} |A_r|$ units of up-work on the cluster C. Accordingly, Lemma 2 follows from Claim 1.

Claim 1. *We have* $\sum_{r=0}^{\lambda} |A_r| \leq f_i/2^{k-2}$.

Proof. Consider a given round $r \in [0, \lambda]$. For all $j \in A_r$, we have $d_{ij} < 2^{\kappa^*_{ij}-3} \leq 2^{k+r-3}$, where the last inequality follows from Invariant 3 (this is because every client $j \in A_r$ is at level $k + r$ just before the end of epoch r). The cluster $C = (i, A)$ moves up to level $k + r + 1$ because $cost_{avg}(C) > 2^{k+r}$. Thus, we conclude that:

$$\frac{f_i + 2^{k+r-3} \cdot |A_r|}{|A_r|} \geq \frac{f_i + \sum_{j \in A_r} d_{ij}}{|A_r|} = cost_{avg}((i, A_r)) > 2^{k+r}.$$

Multiplying both sides of the above inequality by $|A_r|$ and then rearranging the terms, we get:

$$f_i > \left(2^{k+r} - 2^{k+r-3}\right) \cdot |A_r| \geq 2^{k+r-1} \cdot |A_r|.$$

So, we have $|A_r| \leq \frac{f_i}{2^{k+r-1}}$ for all $r \in [0, \lambda]$. This gives us: $\sum_{r=0}^{\lambda} |A_r| \leq \sum_{r=0}^{\lambda} \frac{2f_i}{2^{k+r}} \leq \frac{f_i}{2^{k-2}}$. □

D Implementation details

In this section we describe how to implement our algorithm to perform only $\tilde{O}(m)$ amortized time per update, and also the cases where we deviate from the straightforward implementation as described in Section 2.3.

Client insertions. Following the insertion of a client j, we first try to connect it to its nearest open facility i (by either adding it to the critical cluster centered at i, or create the satellite cluster $(i, \{j\})$), instead trying to connect it to any open facility $i \in F$ as described in Section 2.3.

Efficient implementation of Routine FIX-CLUSTERING(.)**.** We describe how to implement the routine so that, it requires $\tilde{O}(m)$ amortized time per update. In particular, we describe an implementation that takes $\tilde{O}(m)$ per update, plus $O(m)$ time per each level change of a client. The latter is still withing the $\tilde{O}(m)$ amortized bound, as each client changes levels $O(\log m)$ times on average (see Section 2.4).

First we need an efficient mechanism to detect whether there is a blocking cluster violating invariant 4, or if there exists a cluster violating invariant 1. Then, we are going to describe how we implement the subroutines FIX-BLOCKING(C, k) and FIX-LEVEL(C) that restore the invariants of the algorithm. We describe our data structure as if we had a parameter ϵ that defines the base of the exponential bucketing scheme, despite setting it to $\epsilon \leftarrow 1$ in Section 2.3.

Checking for violations of invariants 4 and 1. Notice, that by simply maintaining the average cost of each critical cluster one can simply iterate over all critical clusters (which are $O(m)$) and check whether Invariant 1 is violated. Satellite clusters need to be checked only at their creation time, as their set of clients and subsequently their average cost does not change. Hence, we don't need to iterate over the satellite clusters at each iteration inside FIX-CLUSTERING(.). In fact, each satellite cluster is checked iteratively until it is placed at the right level following its creation (either by inserting a new client, or by converting a critical cluster into an satellite one), and the time for all of these check is accounted in the bound of the up-work and down-work, which is $O(\log m)$ amortized per update.

Checking for violations of Invariant 4 efficiently requires an additional data structure. Conceptually, for each facility i, we maintain a two-dimensional matrix $\mathcal{W}$ of counters of dimensions $\log_{(1+\epsilon)}(\Delta) \times (\ell_{max} - \ell_{min})$, where ℓ_{max}, ℓ_{min} the maximum and minimum level that are non-empty at any point in time Δ the ratio of the largest distance over the smallest distance between a client and a facility. The entry $\mathcal{W}[x, y]$ contains the number of clients that are at distance in the interval $[(1+\epsilon)^{x-1}, (1+\epsilon)^x)$ from facility i and are currently placed at level y. This matrix has size at most $O(\log^2_{(1+\epsilon)}(\Delta))$ (which is $O(\log^2(m))$ as we assume Δ is polynomially bounded by m), but can be sparsified by maintaining a two dimensional hash table containing only the entries with positive counter values, while still maintaining constant time access with high probability. Although we implement the sparse version of this data structure, for ease of description we assume that we maintain the full matrix $\mathcal{W}$. Using the matrix $\mathcal{W}$ for facility i, we check whether there exists a blocking cluster at level ℓ centered at a facility i as follows. We iterate over all entries of $\mathcal{W}[x, y]$ such that $y > \ell$, and in increasing order of the x coordinate. Each time we visit an entry $\mathcal{W}[x, y]$ we check whether the following inequality holds:

$$\frac{f_i + \left(\sum_{x'=x_{min}}^{x} \sum_{y'=\ell+1}^{y_{max}} W[x', y'] \cdot (1+\epsilon)^{y'}\right)}{\left(\sum_{x'=x_{min}}^{x} \sum_{y'=\ell+1}^{y_{max}} 1\right)} \leq (1+\epsilon)^{\ell},$$

and if it does then we determine that facility i forms a blocking cluster at level ℓ. By maintaining appropriate running summations, we can iterate over the appropriate entries of the matrix, each time testing the aforementioned inequality, in time linear in the size of the matrix. Notice that due to the bucketisation of the distances in powers of $(1+\epsilon)$, we lose another factor of $(1+\epsilon)$ in the approximation; that is we say that a cluster is blocking at level ℓ only if its average cost is below $(1+\epsilon)^{(\ell-1)}$ instead of if its average cost is below $(1+\epsilon)^{\ell}$, which increases the approximation by another $(1+\epsilon)$ factor. Finally, notice that to maintain the matrix $\mathcal{W}$ of each facility updated, we only

need to spend constant time each time a client changes levels (as the distance between a facility and a client does not change). That is, for each client changing levels, we spend $O(m)$ time to update the matrices of all facilities. A client can change levels at most $O(\log m)$ times, as described in Section 2.4. Hence, one can check whether there exists a blocking cluster in amortized time $\tilde{O}(m)$ over the sequence of all updates. In order to identify the actual blocking cluster one can do that in time proportional to the size of the blocking cluster: One needs to maintain the neighbors of each facility in increasing distance at each level and visit them in increasing distance from the facility and only at the levels that are eligible (that is, levels greater then ℓ).

Implementing subroutines FIX-BLOCKING(C, k) **and** FIX-LEVEL(C)**.** The subroutines FIX-BLOCKING(C, k) and FIX-LEVEL(C) perform the necessary changes to the maintained clustering to restore any violations to the invariants of the algorithm. The only two scenarios that causes a client to be assigned to a different cluster at the same level are 1) the case where operation FIX-LEVEL(C) increases the level of a critical cluster, which causes the critical cluster to merge with other satellite clusters of the destination level, and 2) the case where a critical cluster gets converted into a set of satellite clusters at the same level. In both of these scenarios, we can charge the work to the time spent creating the satellite clusters by insertions of new (which is $O(1)$ per update), plus the time of creating the critical clusters which then get converted into satellite clusters and is at most $O(\log m)$ amortized per udpate. In all other cases, whenever a client changes the cluster to which it is assigned, it also changes levels. As discussed in Section 2.4, the number of changes of levels of clients is $O(\log m)$ amortized. This implies that the amortized time spent in the subroutines FIX-BLOCKING(C, k) and FIX-LEVEL(C) is $O(\log m)$ per update.

E Missing experiments

E.1 Missing experiments for KDD-Cup dataset

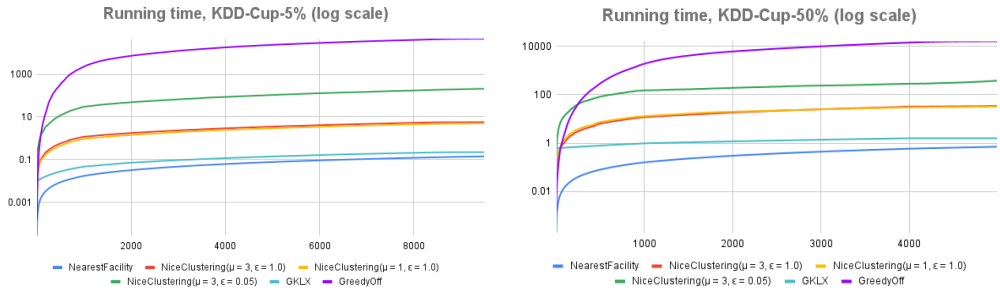

Figure 6: The running time of the algorithms that we consider, in seconds, on the datasets KDD-Cup-5% (left) and KDD-Cup-50% (right). The plot is in log scale.

E.2 Experiments for Census dataset

In terms of cost, NICECLUSTERING$(\mu = 3, \epsilon = 0.05)$ algorithm performs overall the best on the Census-5% and Census-50% instances. In particular, it maintains solutions that are 20%-30% better compared to GKLX on average over the whole sequence of updates on the Census datasets. On the other hand NICECLUSTERING$(\mu = 1, \epsilon = 1)$ performs roughly 10% worse than NICECLUSTERING$(\mu = 3, \epsilon = 0.05)$, but better than GKLX. NICECLUSTERING$(\mu = 3, \epsilon = 1)$ runs roughly 70%-80% worse than NICECLUSTERING$(\mu = 3, \epsilon = 0.05)$, and has the worse performance in terms of cost overall. The behavior of the algorithms, for a single run, is summarized in Figure 7.

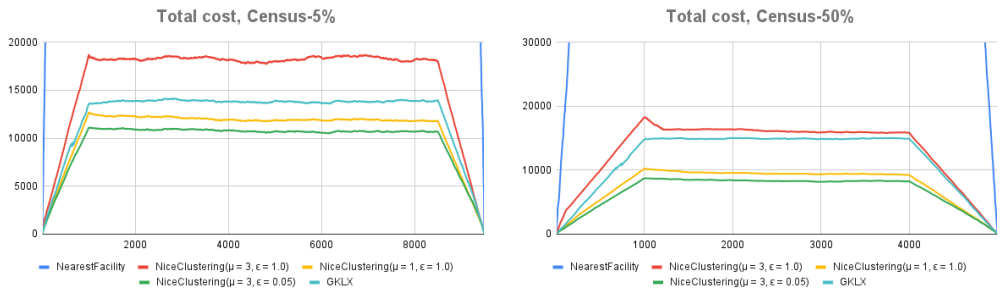

Figure 7: The total cost incurred by the algorithms that we consider, on the datasets Census-5% (left) and Census-50% (right). The full plot of the cost of NEARESTFACILITY is cut off to distinguish the performance of the rest of the algorithms, and is incomparable the rest of the algorithms.

In terms of recourse, NICECLUSTERING$(\mu = 3, \epsilon = 0.05)$ exhibits higher recourse than GKLX, by nearly two orders of magnitude on Census-5%, while GKLX11 incurs around $10\times$ recourse compared to GKLX. In all cases, recourse remains relatively low in absolute value (around 100), especially considering the lengths of the update sequence and the instance sizes. The behavior of the algorithms, for a single run, is summarized in Figure 8.

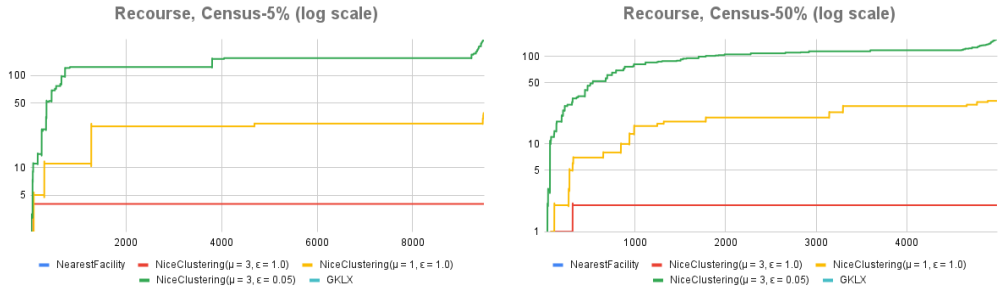

Figure 8: The cumulative recourse incurred by the algorithms that we consider, on the datasets Census-5% (left) and Census-50% (right). Missing lines imply 0 recourse throughout. The plot is in log scale.

The running time of the algorithms on the Census instances is very similar to that on the KDD-Cup instances, as described in the main body of the paper.

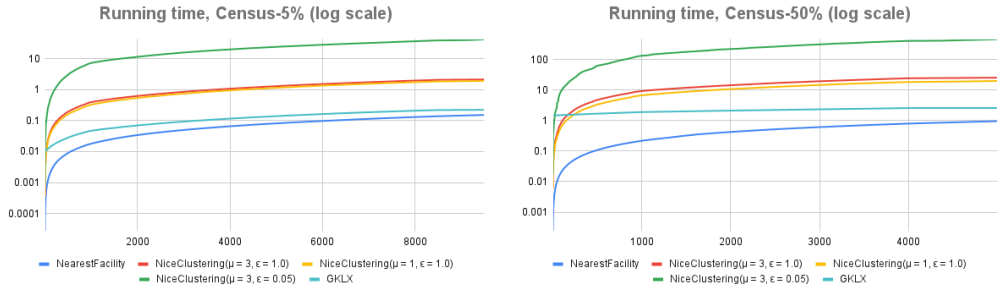

Figure 9: The running time of the algorithms that we consider, in seconds, on the datasets Census-5% (left) and Census-50% (right). The plot is in log scale.

E.3 Experiments for song dataset

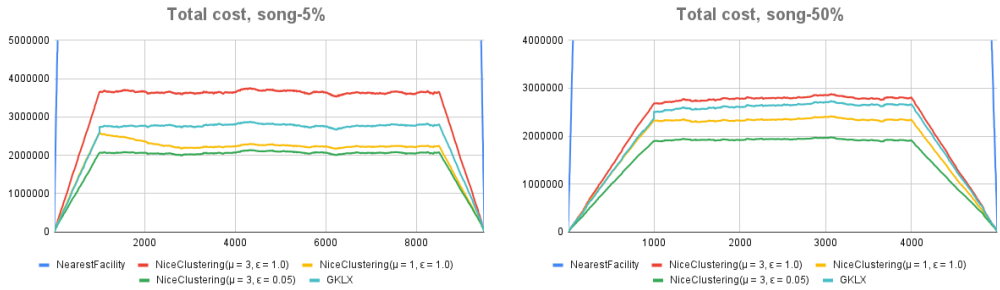

Figure 10: The total cost incurred by the algorithms that we consider, on the datasets song-5% (left) and song-50% (right). The full plot of the cost of NEARESTFACILITY is cut off to distinguish the performance of the rest of the algorithms, and is incomparable the rest of the algorithms.

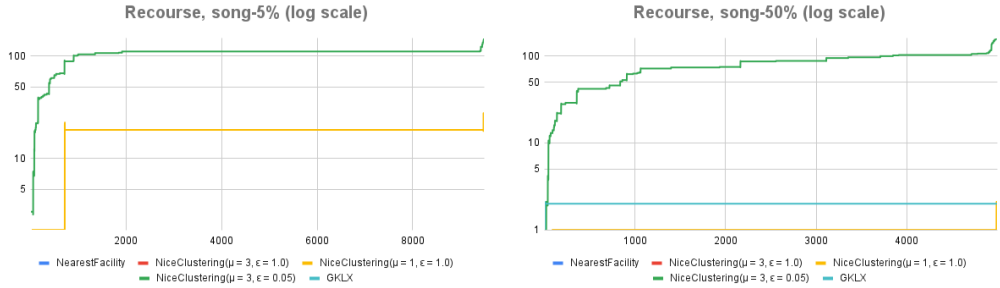

Figure 11: The cumulative recourse incurred by the algorithms that we consider, on the datasets song-5% (left) and song-50% (right). Missing lines imply 0 recourse throughout. The plot is in log scale.

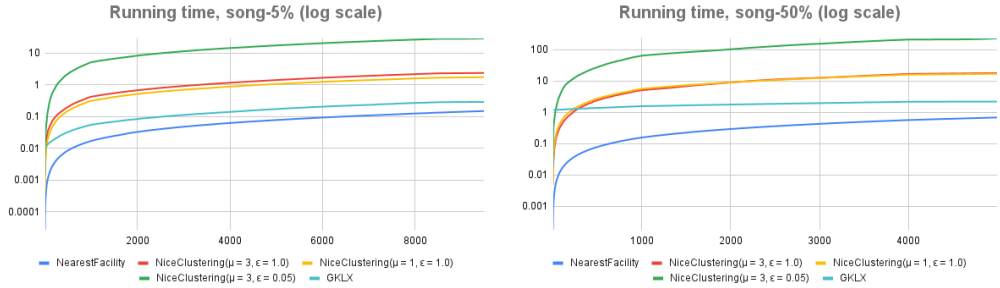

Figure 12: The running time of the algorithms that we consider, in seconds, on the datasets song-5% (left) and song-50% (right). The plot is in log scale.

F Comparing the relative behavior of the algorithms over multiple runs

Comparison of algorithms In order to compare the relative behavior over multiple repetitions of each experiment that we consider, we proceed as follows. Consider the case where we want to compare the cost of the solution produced by algorithm $\mathcal{A}_1$ to the cost of the solution produced by algorithm $\mathcal{A}_2$. For each dataset that we consider and each repetition of an experiment we average the ratio $cost_i(\mathcal{A}_1)/cost_i(\mathcal{A}_2)$, for all $1 \leq i \leq max_u$, where $cost_i(\mathcal{A})$ the cost of the solution produced by algorithm $\mathcal{A}$ on the instance following the i-th update, and max_u the total number

of updates. Denote this average by $\phi(\mathcal{A}_1, \mathcal{A}_2) = \left(\sum_{i=1}^{max_u} \frac{cost_i(\mathcal{A}_1)}{cost_i(\mathcal{A}_2)}\right) / max_u$. This gives us the relative behavior of the two algorithms with respect to the cost of the solution that they produce, on average, over the whole sequence of updates, for a single repetition.

We use the same methodology to compare the cumulative recourse of two algorithms, which we spell out for clarity. For each dataset that we consider and each repetition of an experiment we average the ratio $recourse_i(\mathcal{A}_1)/recourse_i(\mathcal{A}_2)$, for all $1 \leq i \leq max_u$, where $recourse_i(\mathcal{A})$ the cumulative recourse of the solution maintained by algorithm $\mathcal{A}$ following the i-th update. Denote this average by $\psi(\mathcal{A}_1, \mathcal{A}_2) = \left(\sum_{i=1}^{max_u} \frac{recourse_i(\mathcal{A}_1)+1}{recourse_i(\mathcal{A}_2)+1}\right) / max_u$. We add $+1$ to both the nominator and denominator to avoid division with 0.

To aggregate the behavior over multiple repetitions, we report the mean and the standard deviation of $\phi(\mathcal{A}_1, \mathcal{A}_2)$ and $\psi(\mathcal{A}_1, \mathcal{A}_2)$ over 10 repetitions of the experiment.

F.1 Results

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	1.27 (±0.41)	1.71 (±0.29)	1.78 (±0.28)	1.79 (±0.28)	1.30 (±0.32)
NC$(3,1)$	1.01 (±0.00)	1	1.50 (±0.28)	1.58 (±0.29)	1.58 (±0.29)	1.10 (±0.21)
NC$(1,1)$	0.67 (±0.04)	0.72 (±0.10)	1	1.05 (±0.02)	1.05 (±0.02)	0.74 (±0.05)
NC$(3,0.05)$	0.64 (±0.04)	0.69 (±0.10)	0.96 (±0.01)	1	1.00 (±0.00)	0.71 (±0.06)
NC$(1,0.05)$	0.63 (±0.04)	0.69 (±0.10)	0.95 (±0.01)	1.00 (±0.00)	1	0.71 (±0.06)
GKLX	0.92 (±0.09)	0.98 (±0.13)	1.39 (±0.12)	1.45 (±0.15)	1.46 (±0.16)	1

Table 1: Relative behavior of every pair of algorithms that we consider, in terms of the cost of the solution that they produce. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\phi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset KDD-Cup-5%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	0.03 (±0.01)	0.04 (±0.02)	0.03 (±0.01)	0.02 (±0.01)	0.42 (±0.28)
NC$(3,1)$	78.70 (±32.30)	1	2.61 (±0.78)	1.42 (±0.78)	1.36 (±0.84)	30.20 (±30.70)
NC$(1,1)$	35.00 (±18.60)	0.68 (±0.41)	1	0.58 (±0.26)	0.56 (±0.28)	17.40 (±23.30)
NC$(3,0.05)$	183.00 (±196.00)	2.93 (±3.29)	5.80 (±0.00)	1	1.02 (±0.27)	82.10 (±157.00)
NC$(1,0.05)$	158.00 (±124.00)	2.54 (±2.06)	5.92 (±0.29)	1.14 (±0.29)	1	55.80 (±84.30)
GKLX	9.06 (±5.35)	0.14 (±0.07)	0.34 (±0.12)	0.17 (±0.12)	0.15 (±0.11)	1

Table 2: Relative behavior of every pair of algorithms that we consider, in terms of the cumulative recourse that they incur. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\psi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset KDD-Cup-5%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	4.00 (±5.63)	4.82 (±5.68)	5.18 (±5.82)	5.20 (±5.81)	1.76 (±2.01)
NC$(3,1)$	0.86 (±0.00)	1	1.72 (±0.14)	1.96 (±0.17)	1.97 (±0.17)	0.78 (±0.15)
NC$(1,1)$	0.50 (±0.02)	0.62 (±0.04)	1	1.14 (±0.02)	1.15 (±0.02)	0.46 (±0.07)
NC$(3,0.05)$	0.44 (±0.02)	0.54 (±0.03)	0.88 (±0.01)	1	1.01 (±0.00)	0.40 (±0.06)
NC$(1,0.05)$	0.43 (±0.02)	0.54 (±0.03)	0.88 (±0.01)	0.99 (±0.00)	1	0.40 (±0.06)
GKLX	1.19 (±0.27)	2.18 (±2.43)	3.13 (±2.66)	3.47 (±2.71)	3.49 (±2.71)	1

Table 3: Relative behavior of every pair of algorithms that we consider, in terms of the cost of the solution that they produce. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\phi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset KDD-Cup-50%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	0.27 (±0.23)	0.22 (±0.05)	0.17 (±0.06)	0.14 (±0.04)	0.37 (±0.27)
NC$(3,1)$	7.25 (±4.56)	1	1.46 (±0.65)	0.94 (±0.65)	0.74 (±0.44)	1.97 (±2.11)
NC$(1,1)$	7.42 (±1.40)	2.04 (±2.10)	1	0.67 (±0.25)	0.51 (±0.16)	1.84 (±2.36)
NC$(3,0.05)$	17.50 (±8.94)	3.82 (±3.15)	2.23 (±0.00)	1	0.85 (±0.25)	3.39 (±3.99)
NC$(1,0.05)$	23.00 (±9.53)	5.27 (±4.56)	2.80 (±0.88)	1.54 (±0.88)	1	3.88 (±4.45)
GKLX	21.80 (±15.20)	4.55 (±5.74)	2.71 (±1.12)	1.66 (±1.12)	1.21 (±1.01)	1

Table 4: Relative behavior of every pair of algorithms that we consider, in terms of the cumulative recourse that they incur. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\psi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset KDD-Cup-50%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	4.80 (±0.72)	7.24 (±0.40)	7.91 (±0.38)	8.00 (±0.38)	6.15 (±1.30)
NC$(3,1)$	0.23 (±0.00)	1	1.57 (±0.17)	1.73 (±0.20)	1.75 (±0.19)	1.33 (±0.25)
NC$(1,1)$	0.15 (±0.00)	0.66 (±0.05)	1	1.10 (±0.02)	1.11 (±0.02)	0.85 (±0.13)
NC$(3,0.05)$	0.13 (±0.00)	0.60 (±0.05)	0.91 (±0.02)	1	1.01 (±0.00)	0.77 (±0.11)
NC$(1,0.05)$	0.13 (±0.00)	0.59 (±0.05)	0.90 (±0.01)	0.99 (±0.00)	1	0.76 (±0.11)
GKLX	0.19 (±0.04)	0.83 (±0.21)	1.28 (±0.41)	1.40 (±0.44)	1.41 (±0.44)	1

Table 5: Relative behavior of every pair of algorithms that we consider, in terms of the cost of the solution that they produce. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\phi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset Census-5%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	0.43 (±0.31)	0.08 (±0.03)	0.01 (±0.00)	0.01 (±0.00)	0.57 (±0.29)
NC$(3,1)$	4.99 (±3.69)	1	0.33 (±0.04)	0.05 (±0.04)	0.03 (±0.02)	2.64 (±2.42)
NC$(1,1)$	19.10 (±10.50)	7.98 (±7.22)	1	0.15 (±0.06)	0.08 (±0.03)	11.24 (±10.02)
NC$(3,0.05)$	142.00 (±33.90)	63.09 (±58.76)	9.44 (±0.00)	1	0.61 (±0.11)	72.04 (±48.93)
NC$(1,0.05)$	268.00 (±75.00)	113.88 (±108.74)	16.13 (±0.41)	1.84 (±0.41)	1	135.79 (±105.36)
GKLX	9.95 (±7.62)	3.48 (±3.88)	0.64 (±0.04)	0.07 (±0.04)	0.04 (±0.03)	1

Table 6: Relative behavior of every pair of algorithms that we consider, in terms of the cumulative recourse that they incur. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\psi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset Census-5%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	4.82 (±0.53)	7.25 (±0.33)	7.86 (±0.16)	7.94 (±0.17)	6.28 (±0.65)
NC$(3,1)$	0.22 (±0.00)	1	1.54 (±0.14)	1.69 (±0.18)	1.70 (±0.18)	1.35 (±0.21)
NC$(1,1)$	0.15 (±0.00)	0.67 (±0.05)	1	1.09 (±0.02)	1.10 (±0.02)	0.87 (±0.08)
NC$(3,0.05)$	0.13 (±0.00)	0.61 (±0.05)	0.92 (±0.02)	1	1.01 (±0.00)	0.80 (±0.06)
NC$(1,0.05)$	0.13 (±0.00)	0.61 (±0.05)	0.91 (±0.02)	0.99 (±0.00)	1	0.79 (±0.06)
GKLX	0.17 (±0.01)	0.79 (±0.11)	1.18 (±0.14)	1.28 (±0.13)	1.30 (±0.13)	1

Table 7: Relative behavior of every pair of algorithms that we consider, in terms of the cost of the solution that they produce. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\phi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset Census-50%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	0.30 (±0.28)	0.07 (±0.03)	0.01 (±0.01)	0.01 (±0.00)	0.29 (±0.29)
NC$(3,1)$	8.28 (±6.88)	1	0.42 (±0.10)	0.09 (±0.10)	0.07 (±0.08)	2.43 (±4.62)
NC$(1,1)$	21.35 (±7.78)	5.10 (±4.33)	1	0.19 (±0.12)	0.14 (±0.10)	5.19 (±7.03)
NC$(3,0.05)$	167.01 (±102.33)	62.30 (±95.60)	9.33 (±0.00)	1	0.69 (±0.12)	33.70 (±29.64)
NC$(1,0.05)$	274.23 (±141.82)	95.60 (±132.00)	15.59 (±0.42)	1.65 (±0.42)	1	44.07 (±36.83)
GKLX	17.18 (±8.94)	4.38 (±3.61)	0.97 (±0.09)	0.14 (±0.09)	0.09 (±0.06)	1

Table 8: Relative behavior of every pair of algorithms that we consider, in terms of the cumulative recourse that they incur. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\psi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset Census-50%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	6.95 (±1.14)	8.70 (±0.60)	9.77 (±0.53)	9.99 (±0.45)	7.84 (±0.77)
NC$(3,1)$	0.16 (±0.00)	1	1.30 (±0.20)	1.46 (±0.18)	1.51 (±0.21)	1.17 (±0.19)
NC$(1,1)$	0.13 (±0.00)	0.79 (±0.08)	1	1.14 (±0.03)	1.17 (±0.03)	0.90 (±0.04)
NC$(3,0.05)$	0.11 (±0.00)	0.70 (±0.05)	0.88 (±0.02)	1	1.03 (±0.01)	0.80 (±0.04)
NC$(1,0.05)$	0.11 (±0.00)	0.68 (±0.06)	0.86 (±0.02)	0.98 (±0.01)	1	0.78 (±0.04)
GKLX	0.14 (±0.01)	0.89 (±0.12)	1.12 (±0.06)	1.27 (±0.09)	1.30 (±0.09)	1

Table 9: Relative behavior of every pair of algorithms that we consider, in terms of the cost of the solution that they produce. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\phi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset song-5%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	0.92 (±0.23)	0.30 (±0.35)	0.02 (±0.01)	0.01 (±0.00)	0.32 (±0.23)
NC$(3,1)$	2.00 (±3.00)	1	0.39 (±0.02)	0.03 (±0.02)	0.02 (±0.02)	0.40 (±0.31)
NC$(1,1)$	7.26 (±3.65)	6.34 (±3.52)	1	0.10 (±0.05)	0.06 (±0.03)	1.68 (±0.77)
NC$(3,0.05)$	118.00 (±49.10)	106.11 (±54.66)	35.90 (±0.00)	1	0.66 (±0.26)	30.45 (±29.23)
NC$(1,0.05)$	222.00 (±100.00)	210.56 (±114.13)	55.97 (±1.22)	2.23 (±1.22)	1	44.09 (±31.32)
GKLX	11.60 (±4.50)	10.29 (±5.33)	2.81 (±0.07)	0.12 (±0.07)	0.07 (±0.04)	1

Table 10: Relative behavior of every pair of algorithms that we consider, in terms of the cumulative recourse that they incur. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\psi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset song-5%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	6.35 (±1.35)	8.33 (±0.52)	9.70 (±0.42)	9.89 (±0.44)	7.96 (±0.59)
NC$(3,1)$	0.18 (±0.00)	1	1.38 (±0.26)	1.64 (±0.37)	1.67 (±0.37)	1.32 (±0.26)
NC$(1,1)$	0.13 (±0.00)	0.76 (±0.09)	1	1.18 (±0.03)	1.20 (±0.03)	0.96 (±0.03)
NC$(3,0.05)$	0.11 (±0.00)	0.65 (±0.08)	0.85 (±0.02)	1	1.02 (±0.01)	0.81 (±0.03)
NC$(1,0.05)$	0.11 (±0.00)	0.63 (±0.08)	0.84 (±0.02)	0.98 (±0.01)	1	0.80 (±0.03)
GKLX	0.14 (±0.00)	0.79 (±0.10)	1.05 (±0.03)	1.24 (±0.05)	1.26 (±0.05)	1

Table 11: Relative behavior of every pair of algorithms that we consider, in terms of the cost of the solution that they produce. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\phi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset song-50%. ± 0.00 means the standard deviation is less than 0.005.

	NF	NC$(3,1)$	NC$(1,1)$	NC$(3,0.05)$	NC$(1,0.05)$	GKLX
NF	1	0.99 (±0.04)	0.41 (±0.34)	0.02 (±0.01)	0.01 (±0.00)	0.43 (±0.19)
NC$(3,1)$	1.12 (±0.36)	1	0.42 (±0.01)	0.02 (±0.01)	0.01 (±0.00)	0.43 (±0.19)
NC$(1,1)$	5.38 (±3.83)	5.19 (±3.79)	1	0.06 (±0.05)	0.03 (±0.02)	2.43 (±2.95)
NC$(3,0.05)$	134.83 (±82.51)	132.00 (±81.40)	46.79 (±0.00)	1	0.59 (±0.17)	63.67 (±97.23)
NC$(1,0.05)$	239.92 (±108.17)	232.00 (±98.70)	83.56 (±0.62)	2.05 (±0.62)	1	99.59 (±132.35)
GKLX	9.20 (±4.21)	8.84 (±3.74)	3.12 (±0.06)	0.10 (±0.06)	0.05 (±0.02)	1

Table 12: Relative behavior of every pair of algorithms that we consider, in terms of the cumulative recourse that they incur. For a single entry in this table, let $\mathcal{A}_1$ (resp., $\mathcal{A}_2$) be the algorithm representing in the row (resp., column) of the entry. The entry reports the mean and standard deviation of $\psi(\mathcal{A}_1, \mathcal{A}_2)$, over 10 repetitions, for the dataset song-50%. ± 0.00 means the standard deviation is less than 0.005.