Insights into the computational complexity of the single-source capacitated facility location problem with customer preferences

Christina Büsing   Timo Gersing   Sophia Wrede

Teaching and Research Area Combinatorial Optimization, RWTH Aachen University,
Im Süsterfeld 9, 52072 Aachen, Germany.
{buesing,gersing,wrede}@combi.rwth-aachen.de

April 7, 2025

Abstract

Single-source capacitated facility location problems are well studied in the operations research literature, yet classic problems often lack practicability by disregarding the customers' perspective: An authority that assigns customers to open facilities deprives customers from choosing facilities according to their individual preferences. In reality, this can render solutions infeasible, as customers may deviate to their most preferred open facility, thereby breaking capacities of facilities serving more customers than originally planned. *Preference constraints* aim to prevent this by ensuring that each customer is served at their most preferred open facility. The corresponding problem is called *single-source capacitated facility location problem with customer preferences* (*CFLP-CP*) and is known to be strongly NP-hard.

In this paper, we provide a deeper understanding of the transition from polynomially solvable cases to strongly NP-hard cases of the CFLP-CP. In particular, we show that the effect of preference constraints on the theoretical complexity can go both ways: Some problems become harder, while others become easier. This is because preference constraints simplify the assignment of customers to facilities, while simultaneously increasing the complexity of locating facilities. We show that the type of customer preferences, e.g., strict preferences or geographically closest assignments, has a vital impact on the complexity. Notably, strict preferences, i.e., customers cannot be indifferent, allow to compute feasible solutions of the CFLP-CP in polynomial time.

Keywords: facility location, capacities, preferences, computational complexity

1 Introduction

Facility location problems (FLPs) play an essential role in operations research literature (Laporte et al., 2019, Celik Turkoglu and Erol Genevois, 2020). In their basic version, facilities providing some sort of service for customers need to be opened at potential sites and customers are assigned to these facilities. The aim is to minimise the total cost consisting of opening costs for the facilities and costs for assigning the customers to the facilities. We refer to this problem as the *uncapacitated facility location problem* (*UFLP*). Naturally, each customer is assigned to the open facility with lowest assignment cost. The class of FLPs is widely applicable to many real-world problems such as locating health care institutions (Ahmadi-Javid et al., 2017), charging stations for electrical cars (Ahmad et al., 2022), and many more (Celik Turkoglu and Erol Genevois, 2020).

In UFLPs, there are no limitations on the service a facility can provide. However, such limitations are essential for many real-world problems, for example, when looking at hospitals and the number of patients that can be treated (Mestre et al., 2015). Often, customers incur a certain demand and facilities have a certain capacity for serving customers' demands. Such problems belong to the class of *capacitated facility location problems*. In that setting, it is no longer guaranteed that each customer will be assigned to the open facility with lowest assignment cost. The problem is called *single-source* capacitated facility location problem (CFLP) if each customer has to be assigned to exactly one open facility.

Both, the CFLP and the UFLP assume that each customer can be assigned to any facility and that the customers follow these instructions. When locating health care emergency centers, however, it can be assumed that customers will seek the service of their closest open facility in an emergency. In this case, customers have individual preferences over the facilities and their assignment in the solution to the UFLP or CFLP might be in conflict with their preferences. In reality, customers then deviate to one of their most preferred open facilities. While this yields a feasible, although possibly worse, solution in the UFLP, it might turn a feasible solution of the CFLP infeasible.

We refer to the problem in which each customer has to be served at exactly one of their most preferred open facilities as the *(single-source capacitated) facility location problem with customer preferences* (CFLP-CP/UFLP-CP). If preferences are defined by assignment costs, we call the problem *(single-source capacitated) facility location problem with closest assignments* (CFLP-CA/UFLP-CA). Note, any solution for the UFLP-CA coincides with a solution for the UFLP. In this work, we study the combinatorial structures of both the CFLP-CP and the CFLP-CA. The main contribution of this paper is the observation that preference constraints can, on the one hand, ease the complexity of assigning customers to facilities and, on the other hand, increase the complexity of locating facilities. Our observation is supported by the following results:

1. Determining feasibility of the CFLP-CP can be done in polynomial time if either each customer has a strict preference ordering of the facilities or each customer has uniform demand; cf. Section 4.

2. Computing an optimal solution for the CFLP-CP with strict preferences is strongly NP-hard even if capacities are neglected and assignment costs are given by an underlying path graph; cf. Section 5.

3. Computing an optimal solution for the CFLP-CA with strict preferences is exactly weakly NP-hard if assignment costs are given by an underlying star graph, and the CFLP-CA is solvable in polynomial time if customer demands are uniform and assignment costs are given by an an underlying star graph; cf. Section 6.

4. Determining feasibility of the CFLP-CP with uniform customer demands and lower bounds on the customer demand to be served at each open facility is strongly NP-complete even though the corresponding assignment problem can be solved in polynomial time; cf. Section 7.

Contributions 1. and 3. support our claim that preference constraints can ease the complexity of assigning customers to facilities. Recall that determining feasibility for the CFLP is already strongly NP-complete independently of any structures in the assignment costs. Contributions 2. and 4. show that the introduction of customer preferences can increase the complexity of locating facilities in contrast to the complexity of locating facilities in solutions for the UFLP and CFLP, respectively: The UFLP can be solved to optimility in polynomial time if assignment costs correspond to the costs in an underlying tree graph (Mirchandani and Francis, 1990), and a feasible solution of the CFLP with uniform demands can be determined in polynomial time by opening all facilities and solving an instance of the matching problem.

The remainder of this paper is organized as follows. Section 2 provides an overview of related literature. A formal problem definition is given in Section 3, in which we also summarize used notation. In Section 4, we discuss the feasibility of the CFLP-CP with either strict preferences or uniform demands. This is followed by a discussion on the computational complexity of determining an optimal solution for the CFLP-CP with strict customer preferences and assignment costs given by an underlying path graph; cf. Section 5. In Section 6, we study the CFLP-CA with assignment costs which are given by an underlying star graph. We show that this problem is exactly weakly NP-hard if customer preferences are strict, and solvable in polynomial time if customer demands are uniform. In Section 7, we consider the CFLP-CP with uniform demands and additional lower bounds on the customer demand to be served at each open facility. If the set of open facilities is already given, we show that determining a feasible assignment of the customers can be done in polynomial time. If the set of open facilities is unknown, we show that determining feasibility is strongly NP-complete. Last but not least, we discuss our findings and future work in Section 8.

2 Related work

Facility location problems are thoroughly studied in the literature, see, e.g., Laporte et al. (2019), Celik Turkoglu and Erol Genevois (2020). The *uncapacitated facility location problem* (UFLP) poses as the fundamental problem which occurs in all facility location problems as a subproblem. From the complexity-theoretical point of view, it is easy to find a feasible solution for the UFLP but it is strongly NP-hard to find an optimal solution (Mirchandani and Francis, 1990). If the customers and facility locations are defined on an underlying graph with assignment costs equating distances in the graph, then an optimal solution can be computed efficiently for the UFLP on tree graphs (Mirchandani and Francis, 1990). Conversely to the UFLP, already finding a feasible solution for the *single-source capacitated facility location problem* (CFLP) is strongly NP-complete; this can be seen via a reduction from 3-*partition*; cf. Garey and Johnson (1979) for a definition of 3-partition. The relevance and complexity of FLPs has triggered a large number of scientific articles related to different aspects, such as exact solution methods (see, e.g., Avella and Boccia (2009), Görtz and Klose (2012), Fischetti et al. (2016)), the investigation of their polyhedral structure (see, e.g., Leung and Magnanti (1989), Aardal et al. (1995), Avella and Boccia (2009), Avella et al. (2021)) and heuristic solution methods (see, e.g., Mirchandani and Francis (1990), Korte and Vygen (2018)).

Most articles considering facility location problems with customer preferences focus on the uncapacitated case. The work of Hanjoul and Peeters (1987) is considered to be the first occurrence of preference constraints in the context of facility location problems. The authors propose an exact algorithm, which utilises a branch-and-bound procedure, as well as two heuristics for solving the UFLP-CP. Several articles focus on preprocessing strategies and valid inequalities, see, e.g., Cánovas et al. (2007), Vasilyev et al. (2010) and Vasilyev et al. (2013). A semi-Lagrangian relaxation heuristic approach is proposed by Cabezas and García (2022). To the best of our knowledge, a thorough analysis of the computational complexity of the UFLP with customer preferences has not yet been conducted.

Rojeski and ReVelle (1970), Wagner and Falkson (1975) and Gerrard and Church (1996) focus on the special case in which preferences are defined by distances, i.e., each customer prefers to be served at their closest open facility. Rojeski and ReVelle (1970) additionally consider capacities, allow to split customer demands and to expand capacities. Wagner and Falkson (1975) present models for the location of public facilities which maximise social welfare and propose a set of closest assignment constraints. Gerrard and Church (1996) review constraints for integer linear programming formulations to enforce closest assignments and to identify applications for facility location problems with closest assignment constraints. Espejo et al. (2012) theoretically compare all closest assignment constraints in the literature of discrete location theory up until 2011 and contribute a new set of constraints.

When studying CFLPs with customer preferences, there are two main approaches to dealing with overloaded facilities. In the first approach, the overloaded facility may be opened and some customers will be served at a facility they like less. In the second approach, each customer has to be served at their most preferred open facility and facilities that are overloaded can never be opened. Note that whether a facility will be overloaded depends on which other facilities are opened. Casas-Ramírez et al. (2018), Calvete et al. (2020), Polino et al. (2023) study the first approach. Here, the authors consider variations of a bilevel setting where the leader opens facilities with the aim to minimise the total sum of opening and assignment costs; the follower assigns each customer, for which a ranking of all potential facilities is known, to the open facilities with the objective to optimise the sum of achieved preference rankings of the customers. Calvete et al. (2020) compare computational results of the first and second approach to deal with overloaded facilities. The second approach is studied by Büsing et al. (2022, 2024). Büsing et al. (2024) propose preprocessing methods and cover-based inequalities. If preferences are defined by assignment costs, and assignment costs correspond to distances in an underlying path or cycle, Büsing et al. (2022) show that the CFLP-CA can be solved in polynomial time. In contrast, the CFLP is strongly NP-hard independently of the structure of assignment costs, as discussed at the beginning of this section.

The polynomial time algorithms for special cases of the CFLP-CP in which overloaded facilities cannot be opened suggest further potential arising from combining capacities and customer preferences. With this work, we aim to contribute (a) to a better understanding of the combinatorial structures of the CFLP-CP and (b) theoretical foundations which can be utilised in exact and heuristic solution approaches for the CFLP-CP.

3 Problem definition and notation

In the single-source capacitated facility location problem with customer preferences (CFLP-CP), we are given a set of customers I and a set of potential facilities J. Each customer $i \in I$ has a demand $d_i \in \mathbb{N}_0$, each potential facility $j \in J$ has a capacity $Q_j \in \mathbb{N}_0$ and opening cost $f_j \geq 0$. Assigning a customer $i \in I$ to a facility $j \in J$ causes assignment costs $c_{ij} \geq 0$. Each customer $i \in I$ has ranked all potential facilities according to their own preferences, inducing a complete, weak ordering $\leq_i$ on the potential facilities. We write $j <_i k$ if customer $i \in I$ strictly prefers potential facility $j \in J$ over potential facility $k \in J$, and $j =_i k$ indicates that i is indifferent between j and k. With this notation at hand, consider the following definition of the CFLP-CP.

Definition 1. *Let $(I, J, (d_i)_{i\in I}, (Q_j)_{j\in J}, (f_j)_{j\in J}, (c_{ij})_{i\in I, j\in J}, (\leq_i)_{i\in I})$ be given as described above. The* single-source capacitated facility location problem with customer preferences *(CFLP-CP) consists in finding a subset $F \subseteq J$ of facilities that are open and an assignment $\Lambda : I \to F$ of customers to these facilities such that*

1. *the capacity limit of each open facility is met, i.e., $\sum_{i\in\Lambda^{-1}(j)} d_i \leq Q_j$ holds for each $j \in F$,*
2. *each customer is assigned to a facility they prefer most among all open facilities, i.e., there is no $i \in I$ and $j \in F$ such that $j <_i \Lambda(i)$,*
3. *the cost of opening facilities and assigning customers is minimised, i.e., minimise $\sum_{j\in F} f_j + \sum_{i\in I} c_{i\Lambda(i)}$.*

In the following, we denote the set of facilities a customer $i \in I$ prefers most among all facilities in a set $J' \subseteq J$ with $p_i(J') = \{j \in J' | \nexists j' \in J' : j' <_i j\}$.

Facility location problems are often introduced by some underlying graph structure. We then identify facilities and customers with the nodes of the graph. Each node then obtains a demand for the corresponding customer as well as a capacity and opening cost for the corresponding facility. Facilities and customers are connected via edges in the graph, which may represent streets and have non-negative travel cost. The assignment cost of a customer i to a facility j then corresponds to the cost of a shortest path between the corresponding nodes i and j in the graph. We can use the underlying network structure to classify facility location instances. In the remainder of this paper, we will use the formulations *CFLP-CP with assignment costs which are given by an underlying graph* and *CFLP-CP on a graph* interchangeably. The same holds, of course, for all other problems studied here and for more specific graph classes.

We aim in this paper to develop a better understanding of how capacities and preferences affect the computational complexity in detail. That is, in which special cases do capacities and preferences cause the problem to be NP-hard instead of polynomially solvable? And conversely, in which cases does the problem become polynomially solvable instead of being NP-hard?

4 Feasibility

In this section, we discuss the complexity of finding a feasible solution for the CFLP-CP. Recall that simply finding a feasible solution for the single-source capacitated facility location problem (CFLP) is strongly NP-hard. This extends to the CFLP-CP, considering that the CFLP is a special case where all customers are indifferent between all facilities. The complexity of finding a feasible solution for the CFLP arises from the underlying partition problem, where customers represent with their demand the elements and each facility is allowed to serve the same share of the total demand. The partition structure is trivialized if we consider uniform customer demands, which makes finding a feasible solution for the CFLP polynomially solvable in that case. We show that this also holds for the CFLP-CP with uniform demands. However, before we prove this claim, we also show that strict preferences destroy the underlying partition structure. That is, strict preferences allow to find feasible solutions in polynomial time - even for non-uniform demands.

We start by considering the special case of the CFLP-CP with strict preferences, i.e., either $j <_i k$ or $k <_i j$ holds for all $i \in I$, $j, k \in J$.

Theorem 1. *A feasible solution for an instance of the CFLP-CP where each customer has a strict preference ordering of the potential facilities can be found in $\mathcal{O}(|I| \cdot |J| \cdot \log(|J|))$ steps. If we are given a list of facilities sorted by preference for each customer, then we only need $\mathcal{O}(|J| \cdot |I|)$ steps.*

Proof. If not already given, we construct for every customer a sorted list of facilities, requiring a total of $\mathcal{O}(|I| \cdot |J| \cdot \log(|J|))$ steps.

We start by considering a solution in which all facilities are open, i.e., $F = J$. Since each customer has a strict preference ordering of the facilities, the assignment of customers is easy: assign each customer to their most-preferred open facility. If this corresponds to a feasible solution, we are done.

Suppose this solution is infeasible. Then there exists an open facility $j \in F$ which has to serve more demand than its capacity offers. Since all facilities are open and each customer is assigned to their most-preferred facility, we cannot open another facility to lower the demand assigned to j. Moreover, closing other facilities will either keep j's assigned demand constant or increase it. Hence, there is no feasible solution in which facility j is open. We close facility j and reassign all customers that were assigned to j. If the solution is still infeasible, then we repeat the previous step until either the set of open facilities is empty, i.e., there is no feasible solution, or until we obtain a solution in which the capacity of each open facility is respected.

Note that reassigning a customer can be done in constant time, as we can assign them to the next facility in the preference list. While doing so, we update the aggregated demand that is assigned to the new facility and check whether this demand violates the capacity. Since there are at most $|I|$ customers to be reassigned, the process of reassignment and checking for violated capacities can be done in $\mathcal{O}(|I|)$ steps. In the worst case, $|J|$ facilities are closed until infeasibility is proven. This amounts to $\mathcal{O}(|J| \cdot |I|)$ steps. □

The procedure in the proof above yields a polynomial time algorithm for computing a feasible solution for the CFLP-CP with strict customer preferences. Note that we close exactly those facilities that cannot be open in any feasible solution. Hence, any feasible set of facilities is contained in the set of facilities computed by the above procedure.

Next, we consider the CFLP-CP with uniform demands. We can assume without loss of generality that all demands are equal to one. Otherwise, we apply scaling and rounding to demands and capacities. We can determine feasibility of this problem in polynomial time too.

Proposition 1. *We can compute a feasible solution, or determine that no solution exists, for the CFLP-CP with uniform demands in polynomial time.*

Proof. We introduce an extended facility set $\tilde{J} = \cup_{j \in J}\{j_1, \ldots, j_{Q_j}\}$ to be the set containing a number of copies of each facility according to its capacity, and we set the capacity of each such facility to one, i.e., $Q_j = 1$ for each $j \in \tilde{J}$. Then, we modify the preferences of each customer so that they are indifferent between facilities $j_1, \ldots, j_{Q_j}$ for each $j \in J$. It is evident that the CFLP-CP with uniform demands and this extended problem are equivalent.

Determining feasibility of this extended problem is equivalent to determining feasibility of the *envy-free house allocation problem*. In the envy-free house allocation problem, m houses and n agents with preferences over the houses are given, and the goal is to assign each agent to exactly one house so that every agent likes their house at least as much as any other assigned house. If such an assignment is possible, feasibility is shown. Otherwise, some houses must be removed from the set of potential houses in order to determine

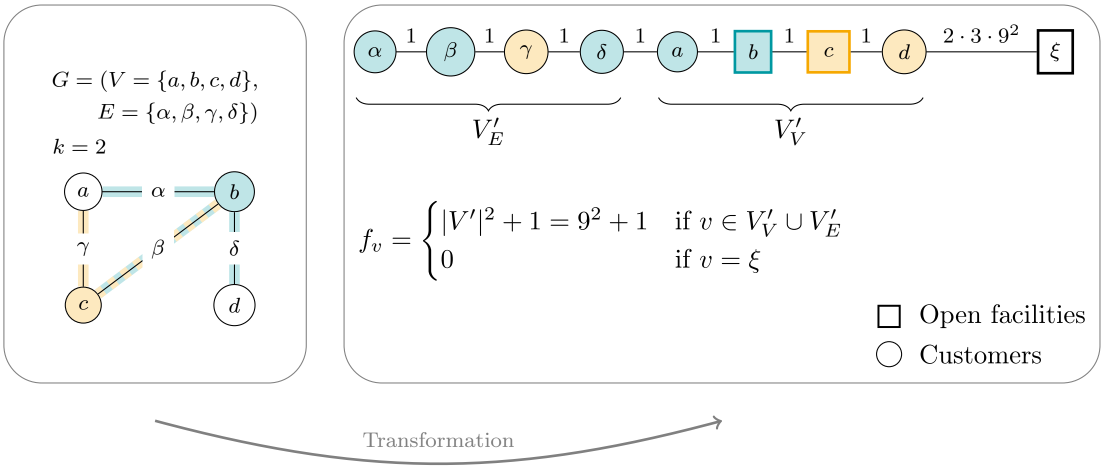

Figure 1: Polynomial transformation of a solution for a given minimum vertex cover instance into a solution for an instance of the UFLP-SCP on paths.

feasibility. Again, the 1-on-1 correspondence between solutions for the extended problem and instances of the envy-free house allocation problem is evident: The set of customers corresponds to the set of agents, the set of open facilities corresponds to the set of houses, and there is an edge between an agent and a house if the corresponding facility is one of the most preferred facilities of the corresponding customer. Gan et al. (2019) show that computing a feasible solution for the envy-free house allocation problem, or determining that no solution exists, can be done in polynomial time. This concludes the proof. □

This section supports our observation that preference constraints do not increase the complexity of assigning customers to facilities.

5 Optimality for strict preferences

In the previous section, we showed that feasibility for the CFLP-CP can be determined in polynomial time if either preferences are strict or customer demands are uniform. These findings yield hope for special cases of the CFLP-CP in which an optimal solution can be computed efficiently. However, in this section, we show that computing an optimal solution for the CFLP-CP is already strongly NP-hard if the structure of the instance corresponds to a path. We show this by proving the stronger result that the uncapacitated version, i.e., the uncapacitated facility location problem with strict customer preferences (UFLP-SCP) is strongly NP-hard on paths. Since this result is independent of the capacities, it further emphasises the computational impact of customer preferences on decisions regarding facility locations.

Theorem 2. *The uncapacitated facility location problem with strict customer preferences (UFLP-SCP) is already strongly NP-hard if the instance is defined on a path graph.*

Proof. We prove the theorem via a reduction from the strongly NP-hard minimum vertex cover problem (Garey and Johnson, 1979). In the minimum vertex cover problem, we are given a graph $G = (V, E)$ and a number $k \in \mathbb{N}$. The goal is to determine a subset of the nodes $S \subseteq V$ with size of at most k so that at least one end node of each edge in graph G is element of set S.

We transform any vertex cover instance $\mathcal{I} = (G = (V, E), k)$ into an instance of the UFLP-SCP on a path as follows; cf. Figure 1. Introduce one node v'_i for each $i \in \{1, \dots, |V|\}$ in the vertex cover instance, and denote this set of nodes with V'_V. Furthermore, introduce one node e'_j for each $j \in \{1, \dots, |E|\}$ in graph G, and denote this set of nodes with V'_E. Finally, introduce one dummy node ξ. Set $V' = V'_V \cup V'_E \cup \{\xi\}$

denotes the nodes in the underlying graph of the transformed UFLP-SCP instance. Set the opening costs of facility ξ to 0 and the opening costs of facilities in set $V'_V \cup V'_E$ to $|V'|^2 + 1$. Connect the nodes in V' such that they form a path in the order $(e'_1, \dots, e'_{|E|}, v'_1, \dots, v'_{|V|}, \xi)$. Assign to each edge in the path costs of 1, except for edge $\{v'_{|V|}, \xi\}$ to which we assign cost of $k \cdot 3 \cdot |V'|^2$. The edge costs are defined so that it is favorable to assign customers in $V'_V \cup V'_E$ to facilities in $V'_V \cup V'_E$ as well as the customer at ξ to the facility at ξ. Otherwise, customers need to cross the expensive edge $\{v'_{|V|}, \xi\}$.

It remains to define the preferences of each customer. The customer at ξ prefers the facility at ξ most. The order of their remaining preferences is arbitrary. Consider customers located in V'_V. Each customer in V'_V prefers the facility at their own node, followed by facilities at nodes of the same type. That is, customers in V'_V prefer facilities in V'_V. Apart from that, the order of preferences is arbitrary. Consider a customer $e' \in V'_E$ who corresponds to an edge $e = \{u, w\} \in E$ in the vertex cover instance. Denote the nodes corresponding to u, w with u', w' in the CFLP-CP instance. We define the most preferred facilities of e' to be those in $\{u', w'\}$, followed by the facility at ξ. The preference order of the remaining facilities is also arbitrary. Assuming that the facility at ξ is open, the preferences of e' ensure that we have to pay high assignment costs by crossing the expensive edge $\{v'_{|V|}, \xi\}$ if we do not open a facility corresponding to one of the end nodes of edge e.

We will show that this yields a correspondence between open facilities in V'_V and vertex covers in V. In particular, we prove that there exists a vertex cover $S \subseteq V$ of size $|S| \leq k$ *if and only if* there exists a solution to the CFLP-CP instance with cost of at most $(k+1) \cdot |V'|^2 + k$.

Assume there exists a vertex cover $S \subseteq V$ with $|S| \leq k$. Let $V'_S \subseteq V'_V$ be the nodes in the CFLP-CP instance corresponding to those in the vertex cover. Then we construct a solution by opening all facilities at nodes in $V'_S \cup \{\xi\}$. This yields opening costs of 0 for the facility at ξ and at most $k \cdot (|V'|^2 + 1)$ for those in V'_S. We can assign the customer at ξ to their own facility, which yields no assignment costs. Furthermore, we assign all customers in $V'_V \cup V'_E$ to facilities in V'_S, which yields assignment cost of at most $(|V| + |E|)^2$ since $|V| + |E|$ customers need to be assigned and every assignment costs at most $|V| + |E|$. The assignment is feasible as customers in $V'_V \cup \{\xi\}$ prefer facilities at nodes of their own type, and the most preferred facilities of customers in V'_E are located in V'_S since S is a vertex cover. Overall, we obtain a feasible solution of cost less than $k \cdot (|V'|^2 + 1) + (|V| + |E|)^2$. Since $(|V| + |E|)^2 \leq (1 + |V| + |E|)^2 = |V'|^2$ holds, we obtain total costs of at most $k \cdot (|V'|^2 + 1) + |V'|^2 = (k+1) \cdot |V'|^2 + k$. Hence, the constructed solution to the CFLP-CP instance has cost of at most $(k+1) \cdot |V'|^2 + k$.

Now, assume we are given a solution to the CFLP-CP instance of cost at most $(k+1) \cdot |V'|^2 + k$. Then the facility at ξ is open. Otherwise, the customer at ξ would need to cross edge $\{v'_{|V|}, \xi\}$ to their assigned facility, which amounts to assignment costs of at least $k \cdot 3 \cdot |V'|^2 \geq (k+2) \cdot |V'|^2 > (k+1) \cdot |V'|^2 + k$.

We can assume without loss of generality that customers in V'_E are not assigned to the facility at ξ. Otherwise, any such customer has to cross edge $\{v'_{|V|}, \xi\}$, and the assignment cost of such a customer to ξ already exceeds bound $(k+1) \cdot |V'|^2 + k$ as shown in the previous observation. Then, for each customer at $e' \in V'_E$ corresponding to edge $e = \{u, w\} \in E$ either a facility at u' or w' has to be open in V'_V. In this case, all customers at nodes in V'_V are served at facilities at nodes in V'_V. In conclusion, each customer in $V'_V \cup V'_E$ incurs assignment cost of at most $|V'|$. Hence, the overall assignment cost incurred by all customers is at most $|V'|^2$.

Recall that opening a facility in V'_V incurs cost of $|V'|^2 + 1$, opening a facility at ξ incurs cost of 0. Then, opening k facilities at nodes in V'_V and a facility at ξ yields overall cost of at most $k \cdot (|V'|^2 + 1) + |V'|^2 = (k+1) \cdot |V'|^2 + k$. Opening one more facility at a node in V'_V leads to overall opening cost of at least

$$(k+1) \cdot (|V'|^2 + 1) = (k+1) \cdot |V'|^2 + k + 1 > (k+1) \cdot |V'|^2 + k$$

and violates the bound on the overall cost. Hence, in a solution to the given CFLP-CP instance of cost at most $(k+1) \cdot |V'|^2 + k$, at most k facilities at nodes in V'_V are open. Denote this set of open facilities with $S' \subseteq V'_V$.

We construct a solution for the minimum vertex cover instance by defining the vertex cover as the set of nodes corresponding to nodes in set S'. Denote this set with $S \subseteq V$. Since each customer located at a node $e' \in V'_E$ is served at a facility $v' \in S'$, this implies that the corresponding edge $e \in E$ is covered by the corresponding node in $v \in S$. Since all customers at nodes in V'_E are served by the facilities in S' and $|S'| \leq k$, the corresponding edges are covered by all nodes in S with $|S| \leq k$. Hence, S is a vertex cover of size $|S| = |S'| \leq k$, which completes the proof. □

It is easy to see that this result also extends to the CFLP-CP. In conclusion, solving both the UFLP-SCP and the CFLP-CP with strict preferences is already strongly NP-hard on basic graph classes, regardless of the actual demands and capacities. Theorem 2 supports our claim that preference constraints increase the complexity of locating facilities: Recall that the traditional UFLP on trees can be solved in polynomial time (Mirchandani and Francis, 1990). Furthermore, the result emphasizes the importance of an underlying structure of the customer preferences.

6 Closest assignments as preferences

In the previous section, we used a rather artificially designed preference pattern to show strong NP-completeness of the CFLP-CP. We now focus on more structured preference constraints. Büsing et al. (2022) show that the single-source capacitated facility location problem with closest assignments (CFLP-CA) is solvable in polynomial time on paths and cycles. On stars, however, it is already already strongly NP-complete to compute a feasible solution. This is because customers may be indifferent between all facilities but the central one, leaving us with an NP-complete partitioning problem. In order to remove the partition-structure from our problem, we now investigate the impact of either strict preferences or uniform demands on the computational complexity of the CFLP-CA. In Section 6.1, we show that the formerly strongly NP-complete problem on stars becomes weakly NP-complete when considering strict preferences. In Section 6.2, we show that the problem on stars can even be solved in polynomial time when considering uniform demands.

6.1 Strict preferences on stars

In the following, we prove that computing an optimal solution for the CFLP-CA with strict preferences on stars is weakly NP-hard. We will refer to *strict preferences* as *distinct assignment costs* since preferences are now defined by assignment costs.

Theorem 3. *The CFLP-CA with distinct assignment costs which are given by an underlying star graph is at least weakly NP-hard.*

Proof. We prove our claim via a reduction from the weakly NP-hard *knapsack problem* (Garey and Johnson, 1979). Consider a knapsack instance consisting of a capacity $B \in \mathbb{N}$ and a set of items N, where each item $i \in N$ is assigned a profit $p_i \in \mathbb{N}$ and a weight $w_i \in \mathbb{N}$. The goal is to find a subset $S \subseteq N$ of items so that the total weight $\sum_{i \in S} w_i$ is at most B and the total profit $\sum_{i \in S} p_i$ is maximal.

We transform a given knapsack instance into an instance of our problem as follows. Introduce node z as the center node in the star graph. The demand, capacity, and opening costs of z are defined as $d_z = 1$, $Q_z = B + 1$, and $f_z = 0$. Introduce for each item in set N a leaf node on the star. Denote the set of leaves with $L = \{1, 2, \ldots, |N|\}$. Each node $i \in L$ has both a demand and a capacity equal to the weight of the corresponding item in the knapsack instance, i.e., $d_i = Q_i = w_i$. We set the assignment cost of node $i \in L$ to the center node to $c_{iz} = i$. This choice ensures that each customer has distinct assignment cost to all other nodes. We define the cost of opening a facility at $i \in L$ so that we save cost of p_i when assigning i to z instead of opening a facility at i, i.e., $f_i = p_i + c_{iz} = p_i + i$. We prove that there exists a knapsack solution of value at least P *if and only if* there exists a facility location solution of cost at most $\sum_{i \in N}(p_i + i) - P$.

Suppose that there exists a feasible solution $S \subseteq N$ for a given knapsack instance with $P \leq \sum_{i\in S} p_i$. We construct the corresponding facility location solution by opening a facility at z as well as all facilities at $\{i \in L | i \notin S\}$. Due to the closest assignment constraints, we need to assign all customers to their own facility if it is open. This assignment meets the capacity constraints on the leaves, as we have $d_i = w_i = Q_i$ for all $i \in L$. The customers in $\{i \in L | i \in S\}$ need to be assigned to the facility at z. Since S is a feasible knapsack solution, this assignment meets the capacity constraint of the facility at z as we have $d_z + \sum_{i\in S} d_i = 1 + \sum_{i\in S} w_i \leq 1 + B = Q_z$. The total opening cost of our constructed solution is $f_z + \sum_{i\notin S} f_i = 0 + \sum_{i\notin S}(p_i + i)$. The total assignment cost is $\sum_{i\in S} c_{iz} = \sum_{i\in S} i$. Since inequality $P \leq \sum_{i\in S} p_i$ holds for the considered knapsack solution, inequality

$$\sum_{i\notin S}(p_i + i) + \sum_{i\in S} i = \sum_{i\in N}(p_i + i) - \sum_{i\in S} p_i \leq \sum_{i\in N}(p_i + i) - P$$

holds for the overall cost of the constructed CFLP-CA-instance.

Conversely, consider a solution to the constructed CFLP-CA instance of cost at most $\sum_{i\in N}(p_i + i) - P$, and denote the set of nodes on which facilities are opened with $F \subseteq L \cup \{z\}$. First, note that $z \in F$ holds since no other facility is capable of serving z in addition to the customer at the respective node. Hence, the customers in $L \setminus F$ are assigned to facility z, while all other customers are assigned to their own facility. We define $S = \{i \in N | i \in L \setminus F\}$ as the corresponding solution to the knapsack instance. Due to the capacity constraint of facility z, we have $\sum_{i\in S} w_i = \sum_{i\in L\setminus F} d_i \leq Q_z - 1 = B$. Hence, S is a feasible knapsack solution. The cost $\sum_{i\in L\setminus F} c_{iz} + \sum_{i\in F\cap L} f_i \leq \sum_{i\in N}(p_i + i) - P$ of our facility location solution implies

$$\begin{aligned}\sum_{i\in S} p_i &\geq \sum_{i\in S} p_i + \left(\sum_{i\in L\setminus F} c_{iz} + \sum_{i\in F\cap L} f_i\right) - \left(\sum_{i\in N}(p_i + i) - P\right)\\ &= \sum_{i\in L\setminus F}(f_i - c_{iz}) + \sum_{i\in L\setminus F} c_{iz} + \sum_{i\in F\cap L} f_i - \sum_{i\in L} f_i + P = P,\end{aligned}$$

which completes the proof □

In the following, we present a pseudo-polynomial algorithm solving the CFLP-CA with distinct assignment costs on stars. Combined with the previous result, we derive that the CFLP-CA with distinct assignment costs on stars is exactly weakly NP-hard.

Consider a star graph $T = (V, E)$, and denote the center node with z. We start by describing the structure of a solution for our problem. Afterwards, we give an algorithm for solving this problem.

Anatomy of a solution

Denote the set of nodes on which facilities are opened with $F \subseteq V$.

Unique closest facility to the center node and always-closed facilities. In any feasible solution, there is a unique open facility which has lowest assignment costs to the center node. Suppose this closest open facility to the center node is located at node $\xi \in V$, i.e., $c_{\xi z} < c_{iz}$ for any open facility $i \in F \setminus \{\xi\}$. Note that z might be the closest open facility with $c_{zz} = 0$. Given ξ, we can restrict the potential open facilities as follows.

Lemma 1. *Let the closest open facility to the center node be located at node $\xi \in V$. Then, all facilities located at nodes in set $C(\xi) = \{p \in V : c_{pz} < c_{\xi z}\} \cup \{p \in V : d_p > Q_p\}$ have to be closed. This results in a demand to be served at ξ of at least $D(\xi) = d_\xi + \sum_{p\in C(\xi)} d_p$.*

Proof. Set $C(\xi)$ contains all locations which are strictly closer to the center than ξ or are incapable of serving the demand of the customer sharing the same node. Since ξ is the closest facility to the center, any facility located at a $p \in V$ with $c_{pz} < c_{\xi z}$ is closed. Furthermore, since each customer prefers the facility at their own node most, each open facility has to be able to serve the demand of the customer at their location. □

Always-open facilities. Next, we analyse the set of open facilities. Based on the location of ξ, we can immediately derive locations on which facilities have to be opened in any optimal solution.

Lemma 2. *Denote the location of the closest facility to the center with ξ. In any optimal solution, set $F(\xi) = \{\xi\} \cup \{i \in V \setminus C(\xi) : f_i \leq c_{i\xi}\}$ describes the nodes on which facilities* have *to be opened.*

Proof. Note that every facility in $V \setminus C(\xi)$ can be opened. Since $f_i \leq c_{i\xi}$ holds for any $i \in F(\xi) \setminus \{\xi\}$, the objective value does not increase when opening a facility at $i \in F(\xi)$ instead of assigning the customer at i to ξ. Moreover, opening a facility at i frees up capacity at ξ for other customers, while i only serves their own customer. Hence, opening all facilities in set $F(\xi)$ is advantageous regarding capacity. $\square$

Undecided facilities. Lastly, we consider the set of facility locations for which it is yet undecided whether a facility is opened. Denote this set with $U(\xi) = V \setminus (C(\xi) \cup F(\xi))$.

In the following, we utilise our knowledge of the anatomy of a solution in order to construct an optimal solution.

Solution approach

The previous observations show that opening a facility at node $\xi \in V$ implies the assignment of all customers in set $C(\xi) \cup F(\xi)$. The only decision to be made is regarding the assignment of customers located at nodes in set $U(\xi)$; that is, are they served at ξ or at a facility located on their own node? We save cost of $f_i - c_{i\xi}$ when assigning customer $i \in U(\xi)$ to facility ξ instead of opening a facility at i. Hence, an optimal choice $U'(\xi) \subseteq U(\xi)$ of customers that are assigned to ξ is given by the following knapsack problem

$$k(\xi) := \max_{U' \subseteq U(\xi)} \left\{ \sum_{i \in U'} f_i - c_{i\xi} \,\middle|\, \sum_{i \in U'} d_i \leq Q'_\xi \right\}, \tag{1}$$

which can be solved via dynamic programming in $\mathcal{O}(|U(\xi)| \cdot Q'_\xi)$ (Korte and Vygen, 2018). The value of an optimal solution for a given ξ is then $\sum_{i \in F(\xi) \cup U(\xi)} f_i + \sum_{i \in C(\xi)} c_{i\xi} - k(\xi)$. To compute an optimal solution for the original problem, we only need to compute $F(\xi), C(\xi), U(\xi)$, and $U'(\xi)$ for all $\xi \in V$ and choose the one with the best objective value.

Proposition 2. *The CFLP-CA with distinct assignment costs which are given by an underlying star graph can be solved in pseudo-polynomial time $\mathcal{O}(\sum_{\xi \in V} |V| \cdot Q_\xi)$.*

Proof. It is easy to see that computing sets $C(\xi)$, $F(\xi)$, $U(\xi)$ takes $\mathcal{O}(|V|)$ steps. Together with the dynamic program for computing $U'(\xi)$, we have $\mathcal{O}(|V| \cdot Q_\xi)$ steps for each possible $\xi \in V$. $\square$

By combining the lower and upper bound on the complexity from Theorem 3 and Proposition 2, we obtain the following result.

Corollary 1. *The CFLP-CA with distinct assignment costs which are given by an underlying star graph is exactly weakly NP-hard.*

To the best of our knowledge, an extension to general trees is not straight-forward − if there is one. An important question to be answered by algorithms for trees is how to efficiently evaluate which of the customers located in subtrees neighbouring nodes $i \in U(\xi)$ are also served at ξ.

Finally, note that in case of uniform demands $d \equiv 1$, we have $Q_\xi \leq |V|$, which yields the following result.

Corollary 2. *The CFLP-CA with distinct assignment costs and uniform customer demands can be solved in $\mathcal{O}(|V|^3)$ steps if the underlying graph corresponds to a star.*

6.2 Uniform demands on stars

In the following, we show that the CFLP-CA with uniform customer demands on stars is solvable in polynomial time, even if customers are indifferent between several facilities. That is, we extend Corollary 2. We assume without loss of generality that all demands are equal to one. Otherwise, we apply scaling and rounding to the demands and capacities.

Anatomy of a solution

Consider an CFLP-CA instance defined on a star $T = (V, E)$, and denote the center node with z. Let $R = \{r_1, \dots, r_k\}$ be the set of distinct distances with $r_1 < r_2 < \dots < r_k$ with $r_1 = c_{zz} = 0$ and $r_k = \max_{i \in V}\{c_{iz}\}$. Since preferences are defined by assignment costs, each customer has to be served at an open facility with lowest assignment cost. Hence, a customer is either served at the facility at their node or at an open facility with lowest assignment cost to the center node.

Suppose the facility with lowest assignment cost to the center node has assignment cost of $r \in R$ to the center. Like in the previous section, there are nodes on which facilities have to stay closed and others on which we have to open facilities.

- *Always-closed facilities.* Facilities located at nodes $p \in V$ have to stay closed if either the assignment cost to the center node is strictly lower than r, i.e., $c_{pz} < r$, or the demand of the customer located on this node exceeds the capacity of this node's facility, i.e., $Q_p < 1$; cf. Lemma 1. Denote this set of nodes with $C(r) = \{p \in V : c_{pz} < r\} \cup \{p \in V : Q_p < 1\}$.
- *Always-open facilities.* Facilities located at nodes $p \in V \setminus C(r)$ are open in a cost-minimising solution if opening a facility at p is at most as expensive as assigning the customer at p to another open facility with lowest assignment cost to the center node, i.e., $f_p \leq c_{pz} + r$; cf. Lemma 2. Denote this set of locations with $F(r) = \{p \in V \setminus C(r) : f_p \leq c_{pz} + r\}$.

The set of *undecided facilities* differs from before. Conversely to the previous section, now the set of undecided facilities is defined as the set of potential open facilities with assignment costs to the center node equal to r, i.e., $U(r) = \{p \in V \setminus C(r) : c_{pz} = r\}$. In order to understand this claim, let us first discuss the assignment of remaining customers located at nodes $W(r) = V \setminus (C(r) \cup F(r) \cup U(r))$. Each customer in $W(r)$ is either served at a facility in $U(r)$ or at the facility at their location. It is evident that for each customer $p \in W(r)$ the former decision is economically better than the latter one. Yet, due to potential capacity restrictions of open facilities in $U(r)$, such an assignment might not be possible. However, if we are given a set of open facilities in $U(r)$, then the assignment of customers in $W(r)$ is immediately clear as the next lemma shows.

Lemma 3. *Let $U'(r) \subseteq U(r)$ be a set of open facilities in a solution for the CFLP-CA with uniform demands. Due to the uniform demands, we can determine the optimal assignment of customers in $W(r)$ in a polynomial number of steps by assigning customers $i \in W(r)$ with greatest values $f_i - (c_{iz} + r)$ to the facilities in $U'(r)$. That is, we solve* $\arg\max_{W'(r) \subseteq W(r)}\{\sum_{i \in W'(r)} f_i - (c_{iz} + r) \mid |W'(r)| = \sum_{i \in U'(r)} Q_i - |C(r) \cup U'(r)|\}$.

Proof. The open facilities in $U'(r)$ have to serve all customers in $C(r) \cup U(r)$. Denote with $B = \sum_{i \in U'(r)} Q_i - \sum_{i \in C(r) \cup U(r)} d_i = \sum_{i \in U'(r)} Q_i - |C(r) \cup U(r)| \geq 0$ the leftover capacity after this assignment. Since $f_p > c_{pz} + r$ holds for any $p \in W(r)$, we can reduce the overall objective value most by assigning B customers in $W(r)$ with highest value $f_p - (c_{pz} + r)$ to the facilities in $U'(r)$. The remaining customers in $W(r)$ are served at the facility at their location. It is evident that this procedure corresponds to solving a knapsack problem in which each customer has a demand of one, and it can be solved in polynomial time. □

Hence, the assignment of customers located at nodes in set $W(r)$ depends on the choice of open facilities located at nodes in set $U(r)$. The assignment can be done efficiently if the set of open facilities in $U(r)$ is given. In the following, we utilise our knowledge of the anatomy of a solution in order to construct an optimal solution.

Solution approach

The previous observations show that the choice of open facilities located at nodes in $U(r)$ determines the optimal objective value for the CFLP-CA on stars with uniform demands where the closest facility to the center has a distance of r to z. Denote with $U'(r) \subseteq U(r)$ the set of open facilities and with $B = \sum_{i \in U'(r)} Q_i - |C(r) \cup U(r)| \in \{0, \dots, \sum_{i \in U(r)} Q_i - |C(r) \cup U(r)|\}$ the leftover capacity after assigning all customers in $|C(r) \cup U(r)|$ to the open facilities in $U'(r)$. For a given amount B of leftover capacity at open facilities, an optimal choice $U'(r, B) \subseteq U(r)$ of facilities that are opened is given by the following min-knapsack problem

$$k(r, B) := \min_{U' \subseteq U(r)} \left\{ \sum_{i \in U'} (f_i - 2 \cdot r) \middle| \sum_{i \in U'} Q_i \geq |C(r) \cup U(r)| + B \right\}, \tag{2}$$

which can be solved via dynamic programming in $\mathcal{O}(|U(r)| \cdot (|C(r) \cup U(r)| + B)) = \mathcal{O}(|V|^2)$. The value of an optimal solution for given r, B is then

$$\sum_{i \in F(r)} f_i + \sum_{i \in C(r) \cup U(r)} (c_{iz} + r) + k(r, B) + \left[\sum_{i \in W(r)} f_i - \max_{W'(r) \subseteq W(r)} \left\{ \sum_{i \in W'(r)} f_i - (c_{iz} + r) \middle| |W'(r)| = B \right\} \right].$$

Note that the former half of the formula represents the cost of opening facilities in $F(r)$ and $U(r)$ so that B customers from $W(r)$ can be served at facilities in $U(r)$. The second half of the formula, i.e., the part in square brackets, represents the cost of assigning B customers in $W(r)$ to open facilities in $U(r)$ so that the saved cost is maximised. To compute an optimal solution for the original problem, we only need to compute $F(r), C(r), U(r), W(r)$, and $U'(r, B)$ for all $B = 0, \dots, \sum_{i \in U(r)} Q_i - |C(r) \cup U(r)|$ and $r \in R$, and choose the one with best objective value.

Proposition 3. *The CFLP-CA with uniform customer demands on a star graph can be solved in $\mathcal{O}(|V|^4)$ steps.*

Proof. It is easy to see that computing $C(r), F(r), U(r)$ takes $\mathcal{O}(|V|)$ steps. Together with the dynamic program for computing $U'(r, B)$, we have $\mathcal{O}(|V|^2)$ steps for each possible $B \in \{1, \dots, |V|\}$ and each $r \in R$. □

Note that the consideration of indifferences within customer preferences increases the number of steps by a factor $|V|$; cf. Corollary 2. We conclude that both distinct assignment costs and uniform demands ease the computational complexity of the CFLP-CA on star graphs. While the CFLP-CA with distinct assignment costs is exactly weakly NP-hard on stars, the restriction to uniform demands allows to solve the CFLP-CA on stars in polynomial time.

7 CFLP-CP with revenue constraints

In this section, we consider a generalisation of the CFLP-CP. In order to model economically operating facilities that require revenue, we introduce lower bounds $\ell_j \in \mathbb{N}_0$ on the total demand to be served at open facilities $j \in J$. These *revenue constraints* are in addition to capacity constraints. We refer to the considered generalisation as the *CFLP-CP with revenue constraints.*

We first show that a feasible assignment within the CFLP-CP with revenue constraints can be computed efficiently if each customer has uniform demand and the set of open facilities is known. Afterwards, we prove that determining feasibility is strongly NP-complete if the decision on the open facilities is yet to be made.

Theorem 4. *Given a set of open facilities, a feasible assignment of the customers to these facilities can be computed in polynomial time for the CFLP-CP with revenue constraints and uniform demands.*

Proof. We can assume without loss of generality that all demands are equal to one. Otherwise, we apply scaling and rounding to demands and capacities.

Let F be the set of open facilities. We reduce the problem of assigning customers I to facilities F to a bipartite matching problem by introducing Q_j copies for each open facility $j \in F$. We denote these copies with $\tilde{F} = \bigcup_{j \in F} \{j_1, \ldots, j_{Q_j}\}$. We construct a bipartite graph $G = (I \cup \tilde{F}, E)$ with edges $E = \{\{i, j_k\} | i \in I, j \in p_i(F), k \leq Q_j\}$, i.e., $i \in I$ is connected to the copies of $j \in F$ if j is contained in the most preferred facilities $p_i(F)$ of i. It is evident that a feasible assignment of customers exists if and only if there is a matching $M \subseteq E$ that matches every customer $i \in I$ and at least ℓ_j many copies of j for all $j \in F$. To compute such a matching, we first compute a matching that meets the lower bounds on the number of served customers and then augment it to a matching covering all customers.

Assume there exists a matching $M \subseteq E$ with the desired properties. Then there exists a subset $M' \subseteq M$ that matches exactly ℓ_j many copies of j for all $j \in F$. We can assume without loss of generality that M' matches the first ℓ_j copies of j. Then M' is also a matching in the induced subgraph $G' = G\left[I \cup \tilde{F}'\right]$ with $\tilde{F}' = \bigcup_{j \in F} \{j_1, \ldots, j_{\ell_j}\}$. Since $|M'| = |\tilde{F}'|$ holds, every maximum matching in G' matches all nodes in $\tilde{F}'$. We can compute a maximum matching in G' in polynomial time (Korte and Vygen, 2018). Let M'' be such a matching. If M'' already matches all customers in I, then M'' corresponds to a feasible solution of the assignment problem. Otherwise, M'' is not a maximum matching in G due to $|M''| < |I| = |M|$. According to Berge's Theorem (Korte and Vygen, 2018), there exists an augmenting M''-path P in G, i.e., a path alternating between edges in $E \setminus M''$ and M'' that starts and ends in different exposed nodes. Finding an augmenting path P can be done in polynomial time via breadth-first search when starting in exposed nodes and traversing edges in an alternating fashion. Note that the symmetric difference $M'' \triangle P$ is a matching of size $|M''|+1$. Moreover, every node that is matched by M'' is also matched by $M'' \triangle P$. In particular, $M'' \triangle P$ still matches the first ℓ_j copies of j for every facility $j \in F$. We can iteratively augment M'' until it is a maximum matching in G. Then we have $|M''| = |M| = |I|$, and thus every customer in I is matched, while matching at least ℓ_j copies of j for every facility $j \in F$. □

Contrary to the results in Section 4, which show that feasible solutions for the CFLP-CP with distinct assignment costs or uniform demands can be computed in polynomial time, it is no longer possible to compute feasible solutions for the CFLP-CP with revenue constraints in polynomial time, unless $P = NP$, as we show in the next Theorem. In the following, we prove that this already holds if preferences are defined by assignment costs.

Theorem 5. *Determining feasibility for the CFLP-CA with revenue constraints and uniform demands is strongly NP-complete. This even holds if assignment costs meet the triangle inequality and exactly two customers have to be served at any open facility.*

Proof. The problem is in NP, as we can check in polynomial time whether a set of open facilities and an assignment of customers to facilities implies a feasible solution. We prove strong NP-hardness via a reduction from the strongly NP-complete problem $(3, B2)$-*SAT* (Berman et al., 2003).

We denote the clauses in our $(3, B2)$-SAT formula with $C_1, \ldots, C_k$ and the Boolean variables with $x_1, \ldots, x_n$. We refer to these variables as *positive literals* and to their negations $\bar{x}_i$ as *negative literals*. In contrast to the 3-SAT problem, $(3, B2)$-SAT instances have the additional property that each literal occurs exactly two times in the formula.

We transform an instance X of $(3, B2)$-SAT into an instance T of the CFLP-CA with revenue constraints. We will define T on an underlying undirected graph $G = (V, E)$, where each node corresponds to both a customer and a facility, i.e., $I = J = V$. Since assignment costs are given by the distance of a shortest path between nodes, they immediately fulfill the triangle inequality. Figure 2 depicts an exemplary construction from a $(3, B2)$-SAT instance to the corresponding graph G. The graph G consists of two layers, called *Layer 1* and *Layer 2*, which are constructed as follows.

Layer 1 consists of one subgraph, named L_1^i, for each variable x_i in X. Subgraph L_1^i contains a cycle $(b_i, y_i^+, y_i^-, e_i, h_i, l_i)$ with edge weights $(1, 2, 1, 2, 3, 2)$ as well as nodes a_i, g_i, which are connected to the cycle via edges $\{a_i, b_i\}$ and $\{e_i, g_i\}$ with weight 3, each.

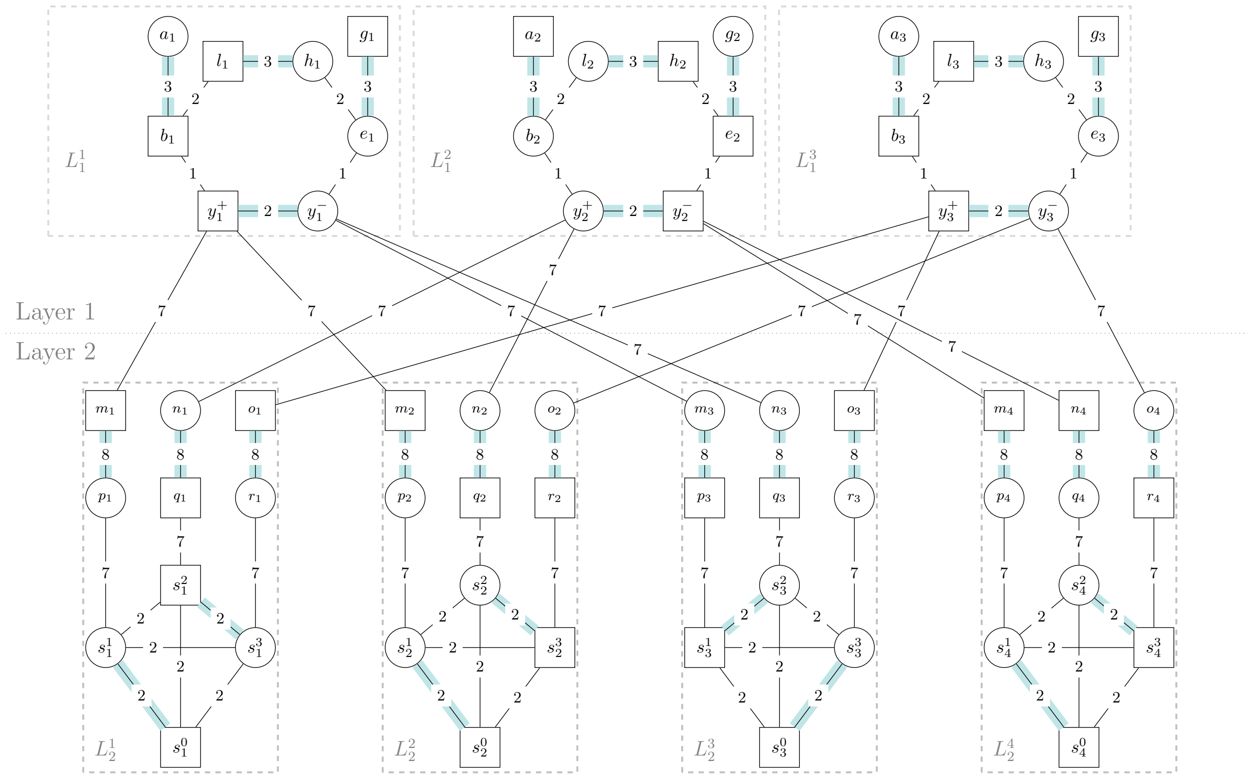

Figure 2: Transformation of a $(3, B2)$-SAT instance $(x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee x_2 \vee \bar{x}_3) \wedge (\bar{x}_1 \vee \bar{x}_1 \vee x_3) \wedge (\bar{x}_2 \vee \bar{x}_2 \vee \bar{x}_3)$ with solution $x_1 = x_3 = 0$ and $x_2 = 1$ to a solution of the considered instance of the CFLP-CA with revenue constraints.

Layer 2 consists of one subgraph, named L_2^j, for each clause C_j in X. Subgraph L_2^j contains a complete graph with four nodes $s_j^0, s_j^1, s_j^2, s_j^3$ in which each edge has weight 2. There are six further nodes $m_j, n_j, o_j, p_j, q_j, r_j$ with edges $\{m_j, p_j\}, \{n_j, q_j\}, \{o_j, r_j\}$ with weight 8, each, and edges $\{p_j, s_j^1\}, \{q_j, s_j^2\}, \{r_j, s_j^3\}$ with weight 7, each.

Layers 1 and 2 are connected via nodes y_i^+, y_i^- and m_j, n_j, o_j. Nodes y_i^+ and y_i^- correspond to the positive and negative literal of variable x_i, respectively. Nodes m_j, n_j, o_j correspond to the three literals in clause C_j. Accordingly, we introduce an edge from y_i^+ to m_j, n_j, or o_j if x_i is the first, second, or third literal in clause C_j, respectively. Analogously, we connect y_i^- if $\bar{x}_i$ is a literal in clause C_j. Each of these edges has weight 7. Note, since X is an instance of $(3, B2)$-SAT, each literal occurs in exactly two clauses and, therefore, each y_i^+ and y_i^- is connected with exactly two subgraphs in Layer 2.

In summary, this yields a graph with $|V| = 8n+10k$ nodes and $|E| = 8n+15k$ edges. For the customers $i \in I = V$ on these nodes, we define uniform demands $d_i = 1$. For the facilities $j \in J = V$, we define opening costs $f_j = 0$ as well as lower and upper bounds $\ell_j = Q_j = 2$. That is, every open facility serves exactly two customers. In conclusion, the constructed instance is of polynomial size.

It remains to prove that there is a feasible solution for the $(3, B2)$-SAT instance X *if and only if* there is a feasible solution for the constructed instance T of the CFLP-CA with revenue constraints. For this, we use the following claims on the anatomy of solutions, which we will prove afterwards in Lemmas 4 and 5.

Claim 1 In every feasible solution, we either open a facility at node y_i^+ or y_i^- for $i = 1, 2, \ldots, n$. Moreover, each customer is served within their respective subgraph L_1^i or L_2^j.

Claim 2 In every feasible solution, at least one facility has to be opened among nodes p_j, q_j, r_j for $j = 1, 2, \dots, k$.

Consider a feasible solution of $(3, B2)$-SAT instance X. We transform this solution into a solution of T as follows. In Layer 1, we open facilities at nodes a_i, y_i^-, e_i, h_i if variable x_i is set to TRUE. If variable x_i is set to FALSE, we instead open facilities at nodes g_i, y_i^+, b_i, l_i. Given this choice, observe that there exists a unique closest facility for each node in L_1^i. In the corresponding assignment, each facility serves exactly two customers.

In Layer 2, we open facilities at nodes m_j, n_j, o_j if the facility at their respective neighbouring node y_i^+ or y_i^- in Layer 1 is opened. This ensures that no customer in m_j, n_j, o_j has an incentive to deviate to a facility in Layer 1. If the neighbouring node y_i^+ or y_i^- is not opened, then we open facilities at the corresponding nodes p_j, q_j, r_j. Since we are given a feasible truth assignment, there is at least one TRUE literal in each clause. Hence, among nodes m_j, n_j, o_j, at least one node is not neighbouring an open facility in Layer 1. Therefore, we open a facility at at least one node from p_j, q_j, r_j in each subgraph L_2^j. Without loss of generality, suppose we open a facility at node p_j. This allows us to open a facility at its neighbouring node s_j^1 without giving the customer at p_j an incentive to go to facility s_j^1. Hence, we open facilities at nodes s_j^0, s_j^1, who serve customers at $s_j^0, \dots, s_j^3$, while customers at m_j, p_j and n_j, q_j as well as o_j, r_j are served at their respective facility. Observe that each customer is assigned to one of their most preferred facility and that each facility serves exactly two customers. Hence, we constructed a feasible solution for T.

Conversely, consider a feasible solution to instance T. In our $(3, B2)$-SAT instance X, we set variable x_i to TRUE if facility y_i^- is opened and to FALSE if facility y_i^+ is opened. This is feasible, since we either open a facility at node y_j^+ or y_j^-, as stated in Claim 1.

Suppose there exists a clause C_j in which all literals are FALSE. In this case, nodes m_j, n_j, o_j are all neighbouring an open facility in Layer 1. In order to prevent customers at these nodes from deviating to Layer 1, we have to open facilities at m_j, n_j, o_j. However, due to Claim 2, at least one facility has to be opened among nodes p_j, q_j, r_j; then, their neighbouring facility at m_j, n_j, o_j does not meet its lower bound on its demand - a contradiction. Hence, there is at least one TRUE literal in each clause. Therefore the constructed assignment is a solution to the $(3, B2)$-SAT instance X, which proves the theorem. □

It remains to prove that Claims 1 and 2 in the previous proof are true. We start by proving Claim 1.

Lemma 4. *Let instance T be as constructed in the proof of Theorem 5. In every feasible solution to T, we either open a facility at node y_i^+ or y_i^- for $i = 1, 2, \dots, n$. Moreover, each customer is served within their respective subgraph L_1^i or L_2^j.*

Proof. We first show that at least one facility is opened in each subgraph L_1^i, L_2^j.

Suppose there is a feasible solution in which all facilities in subgraph L_1^i are closed. Also suppose that there exists an open facility in a neighbouring subgraph L_2^j. Due to the large cost of jumping between layers, all eight customers from L_1^i prefer a facility in one of the neighbouring subgraphs L_2^j. Since T is constructed from a $(3, B2)$-SAT instance in which each literal occurs exactly twice, there are only four edges leaving L_1^i. At least one of these edges is used by at least $8/4 = 2$ customers from L_1^i. Without loss of generality, suppose that at least two customers enter a neighbouring subgraph L_2^j via node m_j. Then, the facility at m_j must be closed since otherwise m_j would serve at least three customers. Analogously, facilities at p_j and s_j^1 must be closed. Therefore, we need at least five potential facilities with equal distance to s_j^1 in L_2^j which serve the two customers from L_1^i as well as the customers at m_j, p_j, and s_2^j. Such facilities do not exist in L_2^j. Hence, if no facility in L_1^i exists, then there cannot be a facility in any neighbouring subgraph L_2^j.

Suppose there is a feasible solution in which all facilities in subgraph L_2^j are closed. Also suppose that there exists an open facility in a neighbouring subgraph L_1^i. Analogue to above, the ten customers from L_2^j are served at a facility in a neighbouring subgraph L_1^i. Since only three edges leave L_2^j, there exists a neighbouring subgraph L_1^i in which at least $\lceil 10/3 \rceil = 4$ customers are served. However, there exist no four nodes with equal distance from the nodes y_i^+, y_i^- via which the customers from L_2^j may enter. Hence, if no facility in L_2^j exists, then there cannot be a facility in any neighbouring subgraph L_1^i.

In summary, if there exists a subgraph in which no facility is opened then we can propagate the arguments above such that there exists no facility at any node that can be reached from the considered subgraph. Hence, the customers cannot be served, contradicting the feasibility of the solution.

Since at least one facility is opened in each subgraph L_1^i for $i = 1, 2, \ldots, n$, the distance from nodes y_i^+ and y_i^- to the nearest facility is less than 6 – the distance between Layer 1 and Layer 2. Hence, customers at nodes y_i^+ and y_i^- are served at facilities in their own subgraph L_1^i. Since y_i^+ or y_i^- lie on every path from L_1^i to every other subgraph, it immediately follows that no customer in L_1^i deviates to another subgraph. Accordingly, we open at least $8/2 = 4$ facilities in L_1^i, since any open facility serves at most two customers. We show next that we open exactly four facilities and that there are only two possible combinations within L_1^i.

Suppose we open a facility at node a_i. Then, we have to ensure that a_i serves at least two customers. If we open a facility at b_i, l_i, or y_i^+ then all customers except a_i prefer these facilities over a_i, and thus a_i only serves itself. Hence, the facility at b_i, l_i, y_i^+ must be closed. Note that due to the symmetry of L_1^i, opening g_i implies that we cannot open facilities at e_i, h_i, y_i^-. Thus, if we open a_i and g_i simultaneously, then all other facilities in L_1^i must be closed, contradicting that we open at least four facilities. Hence, we cannot open a_i and g_i simultaneously. Thus, if facility a_i is open, then we also open facilities y_i^-, e_i, h_i.

Suppose we open no facility at node a_i. Then we must open a facility at node b_i. Otherwise, customers a_i, b_i would both be served at the same facility, thus breaking their capacity, since the distances from b_i to any other facility in L_1^i are distinct. We also need to open facilities at nodes l_i, y_i^+, since their customers would otherwise prefer facility b_i and break their capacity together with customer a_i. Assume that facility g_i is closed. Due to symmetry between a_i and g_i, we then need to open facility e_i. Due to symmetry between b_i and e_i, this implies that facilities h_i and y_i^- need to be open. However, if we open $b_i, l_i, h_i, e_i, y_i^-, y_i^+$, then facilities l_i, h_i can only serve themselves. Hence, facility g_i needs to be open if a_i is closed.

In summary, we either open facilities at a_i, y_i^-, h_i, e_i or at g_i, y_i^+, l_i, b_i. In particular, we either open a facility in node y_i^+ or y_i^-, which shows the first part of the lemma.

We already showed that all customers in Layer 1 are served at their own subgraph. Conversely, if a customer would deviate from their subgraph in Layer 2, then they would be served at a neighbouring subgraph in Layer 1. However, this would break capacities of facilities in Layer 1, since they can serve exactly the customers in their own subgraph. This shows the second part of the lemma. □

Last but not least, we prove Claim 2 from the proof of Theorem 5.

Lemma 5. *Let instance T be as constructed in the proof of Theorem 5. In any feasible solution, at least one facility has to be opened among nodes p_j, q_j, r_j for $j = 1, 2, \ldots, k$.*

Proof. Suppose that all facilities at nodes p_j, q_j, r_j in L_2^j are closed. Then the distance from m_j, n_j, o_j to a different facility in L_2^j is at least 15. From Lemma 4, we know that we open a facility either at y_i^+ or at y_i^- in any subgraph in Layer 1. Hence, the distance from each customer at m_j, n_j, o_j to their closest open facility in Layer 1 is at most 9. In order to stop customers m_j, n_j, o_j from deviating to Layer 1, we have to open facilities at m_j, n_j, o_j.

In order to meet the capacity and demand constraints of the facilities at m_j, n_j, and o_j, they have to serve customers at p_j, q_j, and r_j, respectively. In order to serve the whole demand in L_2^j, we have to open two facilities among nodes $s_j^0, \ldots, s_j^3$ in addition to the facilities at m_j, n_j, o_j. Then, we have to open at least one facility among nodes s_j^1, s_j^2, s_j^3. Opening a facility at one of these three locations encourages their neighbouring customer at p_j, q_j, r_j to deviate from their serving facility at m_j, n_j, o_j to their neighbouring facility at s_j^1, s_j^2, s_j^3. This leaves at least one facility in m_j, n_j, o_j with serving only themselves – a contradiction to the solution being feasible. □

Theorem 5 also holds true for the CFLP-CA with revenue constraints and distinct assignment costs. To see this, note that we never used indifferences in the preference of customers the proofs above. Furthermore, Theorem 5 confirms our observation that preference constraints increase the computational complexity of the location problem. Recall that the complexity of determining feasibility of the CFLP (with revenue constraints) arises from the strong NP-completeness of the corresponding assignment problem, and we can determine feasibility of the CFLP (with revenue constraints) in polynomial time if customer demands are uniform. In contrast, while the assignment problem of the CFLP-CA with revenue constraints and uniform demands can be solved in polynomial time (Theorem 4), determining locations for the facilities and, thereby, determining feasibility is strongly NP-complete.

8 Conclusion

In this paper, we investigate the computational complexity of the single-source capacitated facility location problem with customer preferences (CFLP-CP). While the complexity of (single-source capacitated) facility location problems (CFLP/UFLP) is already well studied, less research has been conducted on the CFLP-CP. Our results indicate that customer preferences can ease the strong NP-completeness of assigning customers to facilities. More specifically, we prove that determining feasibility of the CFLP-CP with strict preferences can be done in polynomial time (Theorem 1), and computing a feasible solution for the CFLP-CP with strict preferences defined by assignment costs is exactly weakly NP-hard (Theorem 3 and Proposition 2). Both cases are strongly NP-complete in the CFLP. Furthermore, we show that the complexity of the assignment problem with uniform customer demands, which is known to be polynomially solvable for the CFLP, seems to retain its computational complexity under the consideration of preference constraints; cf. Propositions 1 and 3 and Theorem 4. Additionally, our results indicate that preference constraints can increase the computational complexity of locating facilities. More specifically, we prove that solving the uncapacitated facility location problem with strict customer preferences (UFLP-SCP) on paths is strongly NP-hard (Theorem 2), conversely to solving the UFLP on paths. Furthermore, we prove that determining feasibility of the CFLP-CP with revenue constrains, uniform demands, and preferences defined by assignment costs is strongly NP-complete (Theorem 5) in contrast to determining feasibility of the CFLP with revenue constraints and uniform demands.

Future work may focus on various directions. First, the polynomial algorithm that determines feasibility of the CFLP-CP with strict preferences can be used in heuristics starting from feasible solutions. This raises the question whether heuristics for the traditional UFLP or the UFLP-SCP can be adapted to our problem. Second, the result that arbitrary preference constraints have a strong impact on the computational complexity of the UFLP-SCP raises the question whether there are further preference patterns besides closest assignments for which there are polynomially solvable or weakly NP-hard special cases. Third, the weak NP-hardness on star graphs for the CFLP-CA with strict preferences as well as the polynomial time algorithm for the CFLP-CA with uniform demands on star graphs raises the question whether it is possible to extend these results to general trees. The consideration of special trees, such as spiders or caterpillars, might help in investigating this question.

Acknowledgments

The authors would like to warmly thank Mariia Anapolska for carefully reading this paper, her valuable comments, and her patience while enduring a lot of misplaced commas. This research has been partially supported by the German Federal Ministry of Education and Research (grant no. 05M16PAA) within the project "HealthFaCT - Health: Facility Location, Covering and Transport", by the Freigeist-Fellowship of the Volkswagen Stiftung, and by the German research council (DFG) Research Training Group 2236 UnRAVeL.

References

K. Aardal, Y. Pochet, and L. A. Wolsey. Capacitated facility location: Valid inequalities and facets. *Mathematics of Operations Research 20*, page 562–82, 1995.

F. Ahmad, A. Iqbal, I. Ashraf, M. Marzband, and I. Khan. Optimal location of electric vehicle charging station and its impact on distribution network: A review. *Energy Reports*, 8:2314–2333, 2022.

A. Ahmadi-Javid, P. Seyedi, and S. S. Syam. A survey of healthcare facility location. *Computers & Operations Research*, 79:223–263, 2017.

P. Avella and M. Boccia. A cutting plane algorithm for the capacitated facility location problem. *Computational Optimization and Applications*, 43:39–65, 2009.

P. Avella, M. Boccia, S. Mattia, and F. Rossi. Weak flow cover inequalities for the capacitated facility location problem. *European Journal of Operational Research*, 289(2):485–494, 2021. ISSN 0377-2217.

P. Berman, M. Karpinski, and A. Scott. Approximation hardness of short symmetric instances of MAX-3SAT. *Electronic Colloquium on Computational Complexity (ECCC)*, 2003. URL: `http://preprints.ihes.fr/2004/M/M-04-28.pdf`.

C. Büsing, T. Gersing, and S. Wrede. Analysing the complexity of facility location problems with capacities, revenues, and closest assignments. *Proceedings of the 10th International Network Optimization Conference (INOC), Aachen, Germany, June 7–10, 2022*, pages 81–86, 2022.

C. Büsing, M. Leitner, and S. Wrede. Cover-based inequalities for the single-source capacitated facility location problem with customer preferences. *Optimization Online*, 2024. URL `https://optimization-online.org/?p=27479`.

X. Cabezas and S. García. A semi-lagrangian relaxation heuristic algorithm for the simple plant location problem with order. *Journal of the Operational Research Society*, 0(0):1–12, 2022.

H. Calvete, C. Galé, J. Iranzo, J. Camacho-Vallejo, and M. Casas-Ramírez. A matheuristic for solving the bilevel approach of the facility location problem with cardinality constraints and preferences. *Computers & Operations Research*, 124:105066, 2020.

L. Cánovas, S. García, M. Labbé, and A. Marín. A strengthened formulation for the simple plant location problem with order. *Operations Research Letters*, 35(2):141–150, 2007.

M. Casas-Ramírez, J. Camacho-Vallejo, and I. Martínez-Salazar. Approximating solutions to a bilevel capacitated facility location problem with customer's patronization toward a list of preferences. *Applied Mathematics and Computation*, 319:369–386, 2018. Recent Advances in Computing.

D. Celik Turkoglu and M. Erol Genevois. A comparative survey of service facility location problems. *Annals of Operations Research*, 292:399–468, 2020.

I. Espejo, A. Marín, and A. M. Rodríguez-Chía. Closest assignment constraints in discrete location problems. *European Journal of Operational Research*, 219(1):49–58, 2012.

M. Fischetti, I. Ljubić, and M. Sinnl. Benders decomposition without separability: A computational study for capacitated facility location problems. *European Journal of Operational Research*, 253(3):557–569, 2016.

J. Gan, W. Suksompong, and A. A. Voudouris. Envy-freeness in house allocation problems. *Mathematical Social Sciences*, 101:104–106, 2019.

M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$-completeness*. W. H. Freeman & Co., New York, USA, 1979. ISBN 978-0716710455.

R. A. Gerrard and R. L. Church. Closest assignment constraints and location models: Properties and structure. *Location Science*, 4(4):251–270, 1996.

S. Görtz and A. Klose. A simple but usually fast branch-and-bound algorithm for the capacitated facility location problem. *INFORMS Journal on Computing*, 24(4):597–610, 2012.

P. Hanjoul and D. Peeters. A facility location problem with clients' preference orderings. *Regional Science and Urban Economics*, 17(3):451–473, 1987.

B. H. Korte and J. Vygen. *Combinatorial Optimization: Theory and Algorithms*. Springer-Verlag, New York, NY, 2018. ISBN 978-3-662-56039-6.

G. Laporte, S. Nickel, and F. Saldanha da Gama, editors. *Location Science*. Springer, 2. edition, 2019. ISBN 978-3-030-32177-2.

J. M. Y. Leung and T. L. Magnanti. Valid inequalities and facets of the capacitated plant location problem. *Mathematical Programming*, 44:271–291, 1989.

A. M. Mestre, M. D. Oliveira, and A. P. Barbosa-Póvoa. Location–allocation approaches for hospital network planning under uncertainty. *European Journal of Operational Research*, 240(3):791–806, 2015.

P. Mirchandani and R. Francis, editors. *Discrete Location Theory*. Wiley Interscience, 1990. ISBN 0-471-89233-5.

S. Polino, J.-F. Camacho-Vallejo, and J. G. Villegas. A facility location problem for extracurricular workshop planning: bi-level model and metaheuristics. *International Transactions in Operational Research*, 0:1–43, 2023.

P. Rojeski and C. ReVelle. Central facilities location under an investment constraint. *Geographical Analysis*, 2(4):343–360, 1970.

I. Vasilyev, X. Klimentova, and Y. A. Kochetov. New lower bounds for the facility location problem with clients' preferences. *Computational Mathematics and Mathematical Physics*, 49:2010–2020, 2010.

I. Vasilyev, X. Klimentova, and M. Boccia. Polyhedral study of simple plant location problem with order. *Oper. Res. Lett.*, 41:153–158, 2013.

J. L. Wagner and L. M. Falkson. The optimal nodal location of public facilities with price-sensitive demand. *Geographical Analysis*, 7(1):69–83, 1975.